Mathematische Annalen

Band 283 Heft 1 1989

Begründet 1868 durch Alfred Clebsch · Carl Neumann
Fortgeführt durch Felix Klein · David Hilbert
Otto Blumenthal · Erich Hecke · Heinrich Behnke

Springer-International

208 Math. Ann. ISSN 0025-5831 MAANA3 283 (1) 1—176 (1989) Januar 1989
Printed on acid-free paper

Mathematische Annalen

Manuscripts must be prepared according to the **Instructions to Authors** printed in every volume. Papers should only be submitted to the editor whose field is nearest to that of the paper, but **not** to the Managing Editor.

Subscription Information. Volumes 282–285 (4 issues each) will appear in 1989.
North America. Annual subscription rate: Approx. US $ 1729.00, single issue price: approx. US $ 129.00, including carriage charges. Subscriptions are entered with prepayment only. Orders should be addressed to: Springer-Verlag New York Inc., Service Center Secaucus, 44 Hartz Way, Secaucus, NJ 07094, USA, Tel. 201/348-4033, Telex 023 125994, FAX (201)348-4505.
All other countries. Annual subscription rate: DM 2992.00, plus carriage charges (Federal Republic of Germany: DM 35.95, including value added tax; for all other countries the price is DM 69.60, except for the following to which SAL (Surface Airmail Lifted) delivery is mandatory: Japan, DM 153.60; India, DM 116.00; Australia/New Zealand, DM 173.60. Airmail delivery to all other countries is available upon request). Volume price: DM 748.00; single issue price: DM 224.40, plus carriage charges. Orders can either be placed via a bookdealer or sent directly to: Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33, Tel. 030/8207-0, Telex 01 83319, FAX (0)30/8214091.
Changes of addresses: Allow six weeks for all changes to become effective. All communications should include both old and new addresses (with Postal Codes) and should be accompanied by a mailing label from a recent issue.
Back volumes: Prices are available on request.
Microform. Microform editions are available from: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106, USA.

Production: Journal Production Department III, Springer-Verlag, Postfach 10 52 80, D-6900 Heidelberg 1, Federal Republic of Germany, Tel. 06221/487-433, Telex 04 61723, FAX (0)622143982

Responsible for Advertisements: E. Lückermann, Heidelberger Platz 3, D-1000 Berlin 33, Telephone: 0 30/8207-0, Telex 01 85411, FAX (0)30/8214091
Printers: Brühlsche Universitätsdruckerei, Giessen — © Springer-Verlag Berlin Heidelberg 1989, Printed in Germany
Springer-Verlag GmbH & Co. KG, D-1000 Berlin 33

Mathematische Annalen

Note to Subscribers

In the last 12 months the number of high-quality manuscripts submitted to our Journal has increased considerably. Our responsibility to authors to ensure that their papers, once accepted, are published as quickly as possible, makes it necessary to announce 4 volumes (16 issues) for 1989 instead of the usual 3 volumes (12 issues).

We are fully aware of the fact that the additional volume represents a financial burden for our subscribers. For this reason, it was with great hesitation that we reached this decision. We therefore ask for your understanding in this matter and would greatly appreciate your support. In 1990 we will return to the usual publication schedule of 3 volumes (12 issues).

For the Editorial Board

Heinz Bauer
Managing Editor

Springer International

Math. Ann. 283, 1–11 (1989)

Mathematische Annalen

Group Actions on Strongly Monotone Dynamical Systems

Janusz Mierczyński[1] and Peter Poláčik[2]

[1] Institute of Mathematics, Technical University of Wrocław, Wybrzeże Wyspiańskiego 27, PL-50-370 Wrocław, Poland

[2] Institute of Applied Mathematics, Comenius University, Mlynská dolina, CS-842 15 Bratislava, Czechoslovakia

It is well known (see [8–10, 13, 14]) that parabolic partial differential equations and systems admitting the strong comparison principle define strongly monotone dynamical systems. If the domain and the coefficients in such an equation exhibit a symmetry then this reflects in the dynamical system being equivariant, i.e. the flow commutes with the action Γ of some group G (see Sect. 3 for examples and e.g. [6, 18, 20] for the general background).

As shown by Hirsch [9, 10] trajectories in strongly monotone dynamical systems have strong tendency to be nonchaotic: almost all of them are quasiconvergent, that is, their ω-limit sets (limit sets, in short) consist of equilibria. More precisely, points which are not quasiconvergent (see Subsect. 1.2 for the precise definition) but have compact trajectory closures form a meagre set.

Our purpose in this paper is to show that if the action Γ is monotone [the homeomorphism $\Gamma(g): X \to X$ is monotone for each $g \in G$] then, loosely speaking, "symmetry is included in nonchaotic behaviour". For instance, almost all trajectories (in the sense as above) eventually symmetrize (their limit sets consist of symmetric equilibria).

To be more specific, suppose that X is a strongly ordered metric space with order relation $\leqq$, Φ is a strongly monotone flow on X and $\Gamma: G \to \operatorname{Hom} X$ is a monotone representation of a compact connected metrizable group G (see Sect. 1 for definitions). Assume (for simplicity) that all trajectories are relatively compact. Then our results assert that:

1) every equilibrium stable from above (or from below) is symmetric: $\Gamma(g)(x) = x$ for all $g \in G$,

2) if X is a separable Banach space then the set Y of points whose limit sets consist of symmetric equilibria is residual in X,

3) if the flow is order-compact then Y is open and dense in X.

Note that we do not impose any smoothness conditions on the flow (so that, unlike [11], the principle of linearized stability is not assumed to hold).

The following fact appears to be crucial in our reasoning. If $x \in X$ and $g \in G$ then one cannot have $\Gamma(g)(x) \leqq x$ unless $\Gamma(g)(x) = x$ (see Proposition 1.3). Therefore in any neighbourhood of a nonsymmetric equilibrium x one can find another one

$\Gamma(g)(x)$ which is not in the relation $<$ to x. This prevents x from being stable. For such an equilibrium even more is true: the group orbit Gx is not stable.

One of fundamental technical tools made use of is Hirsch's Limit Set Dichotomy (see Theorem 1.1). We also use other results of Hirsch and Matano to get more precise information about the behaviour of a flow.

The paper is organized as follows. In Sect. 1 we collect definitions and basic facts concerning strongly monotone dynamical systems and monotone group actions. Section 2 contains main results. In Sect. 3 we give some examples of semilinear and quasilinear parabolic equations for which our abstract theorems from Sect. 2 apply.

1. Preliminaries

1.1. Strongly Ordered Metric Spaces

Let X be a metric space with metric d. By an *ordered space* we mean X endowed with a closed partial order relation $R \subset X \times X$. We write

$x \leqq y$ if $(x, y) \in R$,

$x < y$ if $x \leqq y$ and $x \neq y$,

$x \ll y$ if $(x, y) \in \operatorname{int} R$ and $x \neq y$, where int denotes the interior of a set.

For two subsets $A, B \subset X$ we write $A \leqq B$ ($A < B$, $A \ll B$ respectively) if $x \leqq y$ ($x < y$, $x \ll y$ respectively) for all $x \in A$, $y \in B$.

The reversed relation signs are used in the usual way.

Following [10, Sect. 1] we say that an ordered space X is *strongly ordered* if every open set $U \subset X$ satisfies the following:

(*SO*) If $x \in U$ then there exist a, $b \in U$ such that $a \ll x \ll b$.

We define the *closed order interval*

$$[a, b] := \{x \in X : a \leqq x \leqq b\},$$

and the *open order interval*

$$[[a, b]] := \{x \in X : a \ll x \ll b\}.$$

More generally, for two subsets $A, B \subset X$ we introduce the notation

$$[[A, B]] := \{x \in X : A \ll x \ll B\}.$$

A set $U \subset X$ is *order-bounded* if we have $U \subset [[A, B]]$ for some compact nonempty sets $A, B \subset X$. $U \subset X$ is *order-convex* if it contains $[x, y]$ whenever $x, y \in U$.

The space X can be topologized by taking the collection of all open order intervals as the neighbourhood base. This topology is called the *order topology* (for a set $U \subset X$, $U^{\wedge}$ will stand for U endowed with the relative order topology). The identity map $\mathrm{id}: X \to X^{\wedge}$ is continuous (the order topology is not finer than the original one). Clearly, if K is compact then $K = K^{\wedge}$.

We say that a metric $d^{\wedge}$ for $X^{\wedge}$ is *ordered* if

$$d^{\wedge}(a, b) \leqq d^{\wedge}(u, v)$$

provided that $a < b$, $u < v$ and $[a, b] \subset [u, v]$. For technical reasons, our standing assumption will be that there is an ordered metric $d^{\wedge}$ for $X^{\wedge}$ (it will be clear, however, that many of our results hold without this hypothesis).

When X is an open subset of a Banach space V with norm $\|\cdot\|$, it is always understood that V is strongly ordered by a cone with nonempty interior, d is the metric induced by $\|\cdot\|$ and $d^\wedge$ is the (ordered) metric induced by the order-unit norm $\|\cdot\|_u$ for some $u \gg 0$ (see [1] or [10]).

1.2. Strongly Monotone Dynamical Systems

A map $f: X \to Y$ between ordered spaces is called *monotone* if $x \leqq y$ implies $f(x) \leqq f(y)$, and *strongly monotone* if $x < y$ implies $f(x) \ll f(y)$.

By a *dynamical system* we understand a pair (X, Φ) consisting of a metric space X and a continuous map $\Phi: D(\Phi) \to X$ such that

i) The domain $D(\Phi)$ is an open set in $[0, \infty) \times X$ containing $\{0\} \times X$.
ii) For every $x \in X$, $\Phi(0, x) = x$.
iii) For every $x \in X$, $s \geqq 0$, $t \geqq 0$, we have

$$\Phi(s, \Phi(t, x)) = \Phi(t + s, x),$$

where the equality sign is to be understood in the sense that if one side is defined then so is the other and the equality holds.

We call Φ the *flow*. The map Φ_t is defined as

$$\Phi_t(x) := \Phi(t, x).$$

The set $J_x := \{t \geqq 0 : x \in D(\Phi_t)\}$ is a half-open interval $[0, \tau_x)$, $0 < \tau_x \leqq \infty$, where τ_x is called the *escape time* of x.

We often write $x \cdot t$ instead of $\Phi(t, x)$. By the *trajectory* of $x \in X$ we mean the image of the map

$$\Phi(\cdot, x) : J_x \to X.$$

A set $K \subset X$ is said to be *invariant* if it contains the trajectories of all its members, and *totally invariant* if all its members have infinite escape times and $\Phi_t(K) = K$ for all $t \geqq 0$.

The *limit set* of $x \in X$ is

$$\omega(x) := \{y \in X : \text{there is a sequence } t_k \to \tau_x \text{ such that } x \cdot t_k \to y\}.$$

Its members are called *limit points* of x. If the trajectory of x is *relatively compact* (i.e. its closure is compact) then $\omega(x)$ is nonempty, compact, connected and totally invariant.

An *equilibrium* is a point $p \in X$ such that $p \cdot t = p$ for all $t \geqq 0$. The set of equilibria is denoted by E. If $x \cdot t \to p$ as $t \to \tau_x$ then $p \in E$ and $\tau_x = \infty$. In this case x (and its trajectory) are called *convergent* (we say also that the trajectory of x *converges* to p). A point $x \in X$ is called *quasiconvergent* if its trajectory is relatively compact and $\omega(x) \subset E$.

Let $x, y \in E$. A *trajectory connection* from x to y is given by a continuous map $c: \mathbb{R} \to X$ such that

a) $\lim_{t \to -\infty} c(t) = x$,
b) $\lim_{t \to \infty} c(t) = y$,
c) $c(s + t) = c(s) \cdot t$ for all $s \in \mathbb{R}$, $t \geqq 0$.

A dynamical system (X, Φ) (or its flow Φ) is *strongly monotone* if for each $t > 0$, Φ_t is a strongly monotone map. From now on, it is assumed that every flow is strongly monotone.

We say that Φ is *compact* (resp. *order-compact*) if $\Phi_t(B)$ has compact closure whenever $t>0$ and $B\subset D(\Phi_t)$ is bounded (resp. order-bounded). If Φ is compact (resp. order-compact) then every bounded (resp. order-bounded) trajectory has compact closure and is therefore *global* (i.e. its escape time is infinity).

An equilibrium x is said to be *stable from above* if for every $\varepsilon>0$ there is a $\delta>0$ such that if $y\geqq x$, $d(y,x)<\delta$ and $0\leqq t<\tau_y$ then $d(y\cdot t,x)<\varepsilon$. An invariant set K is said to be *stable* if for every $\varepsilon>0$ there is a $\delta>0$ such that if

$$d(y,K):=\inf\{d(y,x):x\in K\}<\delta$$

and $0\leqq t<\tau_y$ then $d(y\cdot t,K)<\varepsilon$.

If one replaces in the above definitions the metric d by the metric $d^\wedge$, then one obtains the definitions of an equilibrium *order-stable from above* and of an *order-stable* invariant set, respectively. Stability (and order-stability) from below are defined in an analogous way. (Note that our order-stability from above is called upper stability in [10].)

If there exists an ordered metric for $X^\wedge$ (as is the case for X an open subset of a strongly ordered Banach space, see the preceding subsection), then it is straightforward that stability implies order-stability. The notions of stability and order-stability coincide if the flow Φ is order-compact (see [9, p. 47]).

Finally we state without proof a result which will be extensively used in the sequel.

Theorem 1.1 (Limit Set Dichotomy, [10, Theorem 6.8]). *Assume that $x<y$ and that their trajectories are relatively compact. Then either*

$$\omega(x)\ll\omega(y),$$

or else

$$\omega(x)=\omega(y)\subset E.$$

In the latter case for any sequence $t_k\to\infty$ and any $p\in E$ we have $x\cdot t_k\to p\Leftrightarrow y\cdot t_k\to p$.

1.3. Equivariant Group Actions on Strongly Monotone Dynamical Systems

From now on G is a metrizable group with unit element e.

We begin by stating a useful fact.

Lemma 1.2. *Let G be compact. Then for every $g\in G$ there exists a sequence $\{n_k\}$ of positive integers, $n_k\to\infty$ as $k\to\infty$, such that $g^{n_k}\to e$ as $k\to\infty$.*

Proof. By [15, Sect. 1.22] there is a right-invariant metric ϱ on G. Suppose to the contrary that for some $\varepsilon>0$ one has $\varrho(e,g^n)>\varepsilon$ for all n. This implies that $\varrho(g^n,g^m)>\varepsilon$ for all positive integers $n\neq m$. We have a discrete infinite subset $\{g^n:n\in\mathbb{Z}_+\}$ of the compact space G, a contradiction. $\square$

We say that G *acts* on a topological space Y if there is a group homomorphism $\Gamma:G\to\operatorname{Hom}Y$ into the group $\operatorname{Hom}Y$ of homeomorphisms of Y such that the map

$$\gamma:G\times Y\to Y,\qquad \gamma(g,y):=\Gamma(g)(y),$$

is jointly continuous. We will call Γ (or γ) a *group action*. It is well known (see [5, Theorem 1]) that if G is a Baire space (in particular, if G is compact) and Y is metrizable then it suffices to verify only separate continuity in the definition above.

For $y \in Y$ the *orbit* of y is the set

$$Gy := \{\Gamma(g)(y) : g \in G\}.$$

(Note that we use the word "orbit" in connection with the action of the group G, whereas "trajectory" refers to the flow.) By the continuity of γ, if G is compact and/or connected then so is Gy, for each $y \in Y$.

The set $I_y := \{g \in G : \Gamma(g)(y) = y\}$ is called the *isotropy subgroup* of $y \in Y$.

Proposition 1.3. *Let a compact group G act monotonely on an ordered space Y (i.e. for every $g \in G$ the map $\Gamma(g)$ is monotone). Then for every $g \in G$, $y \in Y$, the relation $\Gamma(g)(y) \leqq y$ implies $\Gamma(g)(y) = y$.*

Proof. Assume that $\Gamma(g)(y) \leqq y$ for some $g \in G$, $y \in Y$. Since $\Gamma(g)$ is monotone, we have

$$\ldots \leqq \Gamma(g^{n+1})(y) \leqq \Gamma(g^n)(y) \leqq \ldots \leqq \Gamma(g)(y) \leqq y.$$

By Lemma 1.2 and the continuity of γ,

$$y \geqq \gamma(g, y) \geqq \gamma(g^{n_1}, y) \geqq \ldots \geqq \gamma(g^{n_k}, y) \geqq \gamma(g^{n_{k+1}}, y) \to y.$$

Because $\leqq$ is closed, $y \geqq \gamma(g, y) \geqq y$, so $y = \Gamma(g)(y)$. □

We say that the triple (X, Φ, G) satisfies *Axiom* (GO) if (X, Φ) is a strongly monotone dynamical system, G is a compact connected group, and the following holds.

$(GO1)$ G acts monotonely on X.

$(GO2)$ $\Gamma(g)(x \cdot t) = \Gamma(g)(x) \cdot t$ for all $x \in X$, $g \in G$ and $0 \leqq t < \min(\tau_x, \tau_{\Gamma(g)(x)})$.

The condition $(GO2)$, referred to as *equivariance*, implies that $\tau_x = \tau_{\Gamma(g)(x)}$ for each $x \in X$, $g \in G$. Indeed, if this were not true, then we would have for some $x \in X$ and some $g \in G$

$$\tau_x > \tau_{\Gamma(g)(x)} =: T.$$

But for $s \in [0, T)$, $\Gamma(g)(x) \cdot s = \Gamma(g)(x \cdot s)$, hence, by continuity, $\Gamma(g)(x) \cdot s$ has limit as $s \to T$. From the definition of the flow we deduce that $T < \tau_{\Gamma(g)(x)}$, a contradiction.

By (GO) and the continuity of $\Gamma(g)$, the trajectory of $\Gamma(g)(x)$ is relatively compact (resp. quasiconvergent, convergent) if and only if so is the trajectory of x.

For (X, Φ, G) satisfying Axiom (GO), a point $x \in X$ is said to be *symmetric* if $\Gamma(g)(x) = x$ for any $g \in G$. Otherwise it is called *nonsymmetric*. The set of quasiconvergent points whose limit sets contain only symmetric equilibria is denoted by H.

2. Main Results

In the present section our standing assumption is that (X, Φ, G) satisfies Axiom (GO).

Since in this section we consider a fixed group action, we suppress Γ notationally: we write simply gx instead of $\Gamma(g)(x)$.

Let X_C denote the set of all points having compact trajectory closures, and let Q denote the set of all quasiconvergent points.

Lemma 2.1. *Let $x \in E$. Then for any compact totally invariant set K such that $K > x$ one has $K \gg Gx$.*

Proof. By strong monotonicity and total invariance, $K \gg x$. Consider the set $U := \{z \in Gx : z \ll K\}$. U is nonempty (since $x \in U$) and open in the relative topology of Gx. For each $z \in \operatorname{cl} U \subset E$ (where cl denotes the closure) we have $z \leqq K$, and, by Proposition 1.3, $z < K$, and, again owing to strong monotonicity and total invariance, $z \ll K$. Therefore U is open and closed in the connected space Gx, so $U = Gx$. □

Proposition 2.2. *Let $x \in E$ be nonsymmetric. Then x is isolated in $M^\wedge$ (hence in M), where*

$$M := \bigcup \{\omega(z) : z \in X_C,\ z \geqq x\}.$$

Proof. Assume to the contrary that x is not isolated in $M^\wedge$. Then there is a sequence $y_n \in M \setminus \{x\}$ such that x is its limit in the order topology. By the Limit Set Dichotomy (Theorem 1.1) and Lemma 2.1, for any $z \in X_C$, $z > x$, one has either $\omega(z) = \{x\}$ or $\omega(z) \gg Gx$. Therefore $Gx \ll y_n$ for all n, and, because of the closedness of the relation $\leqq$ in $X^\wedge \times X^\wedge$ (see [10]), $Gx \leqq x$. Proposition 1.3 yields $Gx = \{x\}$, a contradiction. □

Proposition 2.3. *Let $x \in E$ be nonsymmetric. Then there do not exist three points $z_1, z_2, z_3 \in X_C$, $z_1 < z_2 < z_3$, such that $x \in \omega(z_1) \cap \omega(z_3)$.*

Proof. Suppose that such three points exist. By the Limit Set Dichotomy $\omega(z_1) = \omega(z_2) = \omega(z_3)$. Without loss of generality assume $z_1 \ll z_2 \ll z_3$ (if not, replace them by $z_1 \cdot 1$, $z_2 \cdot 1$, $z_3 \cdot 1$, respectively).

Let $t_k \to \infty$ be a sequence such that $z_2 \cdot t_k \to x$. Set $P := \{g \in G : gz_2 \in [[z_1, z_3]]\}$. P is a neighbourhood of e. It follows from the Limit Set Dichotomy that $gz_2 \cdot t_k \to x$ for all $g \in P$ (recall that $gz_2 \in X_C$). On the other hand, from equivariance and continuity of the group action we deduce that $gz_2 \cdot t_k \to gx$. So $P \subset I_x$. Thus I_x has nonempty interior, this implies that I_x equals the connected component of e in G. By connectedness, $I_x = G$, so x is symmetric, a contradiction. □

Theorem 2.4. *Let $x \in E$ be nonsymmetric. Assume that x is not isolated in*

$$\{z \in X_C : z \geqq x\}.$$

Then x is not stable from above. Furthermore, Gx is not stable.

If x is not isolated in

$$\{z \in X_C : z \geqq x\}^\wedge,$$

then the statement holds for order-stability.

Proof. Suppose that a nonsymmetric equilibrium x is stable from above. Due to strong monotonicity, the hypothesis enables us to construct a sequence $z_n \in X_C$, $x \ll z_{n+1} \ll z_n$, $z_n \to x$. From Proposition 2.3 and the stability of x it follows that arbitrarily close to x there is a $y \gg x$, y being a limit point for some $z_n \in X_C$, contrary to Proposition 2.2.

Now suppose that Gx is stable. A reasoning similar to the above one assures us of the existence of a sequence $y_n \gg x$, $y_n \in \omega(z_n)$ for some $z_n \in X_C$, satisfying $d(y_n, Gx) \to 0$ as $n \to \infty$. By Lemma 2.1, $y_n \gg Gx$ for all n. Since Gx is compact, by passing to a subsequence if necessary we may assume that $y_n \to y$ for some $y \in Gx$. This contradicts Proposition 2.2 (with x replaced with y).

The proof for order-stability is quite similar. □

Corollary 2.5. *If the flow Φ is compact (resp. order-compact), then no nonsymmetric equilibrium is stable (resp. order-stable) from above.*

Proof. If $x \in E$ is stable from above then for any $z \gg x$, z near x, the trajectory of z is bounded, hence relatively compact. Similarly for order-stability. □

Remark. Needless to say, all the above results have their analogues for the reversed inequality sign.

Corollary 2.6. *Assume that X is an open order interval in a strongly ordered Banach space V, and that the flow Φ is order-compact. Let $x \in E$ be nonsymmetric. If*

$$S^+(x) := \{z \in E : z > x\}$$

is nonempty, then there exists a symmetric $v \in E$ such that

(i) $v \gg Gx$,

(ii) *v is least in $S^+(y)$ for each $y \in Gx$,*

(iii) *for every $y \in Gx$ there is a trajectory connection from y to v.*

Proof. First observe that the assumptions imply that for all $B \subset X$ and all $t > 0$, $\Phi_t(B)$ has compact closure (so, in particular, $X = X_C$).

A theorem of Matano [13, Theorem 5] asserts that there exists a continuous map $c: (-\infty, 0] \to X$ such that

a) $c(t) \gg x$ for each $t \in (-\infty, 0]$,

b) $c(t) \cdot s = c(t+s)$ for $t \leqq 0$, $s \geqq 0$ with $t + s \leqq 0$,

c) $\lim_{t \to -\infty} c(t) = x$.

From a) and b) we deduce that

$$c(T) \ll c(0) \quad \text{for some} \quad T < 0. \tag{2.1}$$

According to the convergence criterion for strongly monotone flows [10, Theorem 6.4], (2.1) together with relative compactness of trajectories implies that $\omega(c(0)) = \{v\}$ for some $v \in E$, $v \gg c(0) \gg x$. We have obtained a trajectory connection from x to v. Because $c(T) \ll c(0) \ll v$ and these points have $\{v\}$ as their common limit set, from Proposition 2.3 it follows that v is symmetric.

Again by a) and c), for any $z > x$ we have $\omega(z) \geqq v$. Therefore v is least in $S^+(x)$.

Let for some $y \in Gx$ (say $y = gx$) w be a point obtained as above. We have

$$w = g^{-1}w > g^{-1}y = g^{-1}gx = x,$$

so $w > x$, hence $w \geqq v$. Interchanging x and y we get $w \leqq v$, so $w = v$. This concludes the proof of parts (ii) and (iii). Part (i) follows from Lemma 2.1. □

Recall that H denotes the set of all quasiconvergent points whose limit sets consist of symmetric equilibria.

Theorem 2.7. *Let $L \subset X_C$ be simply ordered. Then for each $r \in (L \cap Q) \setminus H$ there is a neighbourhood U_r of r in L such that for every $u \in U_r \setminus \{r\}$ its trajectory converges to a symmetric equilibrium. In particular, the set $(L \cap Q) \setminus H$ is discrete.*

Proof. For any $r \in (L \cap Q) \setminus H$ take a nonsymmetric equilibrium $x \in \omega(r)$. By Proposition 2.3 and the Limit Set Dichotomy, for every $u \in L$, $u \neq r$, we have either $\omega(u) \ll \omega(r)$, or $\omega(u) \gg \omega(r)$. Proposition 2.2 assures us that there exist $y_1, y_2, y_1 \ll x \ll y_2$ such that

$$\omega(u) \notin [[y_1, y_2]] \quad \text{for all} \quad u \in L, u \neq r.$$

Since $x \in \omega(r)$, there is a $T > 0$ such that $r \cdot T \in [[y_1, y_2]]$. By continuity, we can find a neighbourhood $\tilde{U}_r$ of r such that $u \cdot T \in [[y_1, y_2]]$ for all $u \in \tilde{U}_r$.

Define $U_r := \tilde{U}_r \cap L$. If $u \in U_r$ and $u < r$, then for some $s > 0$ we have

$$u \cdot (T+s) \ll y_1 \ll u \cdot T.$$

Since u has compact trajectory closure, the convergence criterion for strongly monotone flows [10, Theorem 6.5] implies that there exists a $v \in E$ such that $\omega(u) = \{v\}$ and $v \ll u \cdot t$ for all $t > T$. Then there are three points $v \ll u \cdot (T+s) \ll u \cdot T$ having $\{v\}$ as their common limit set, hence by virtue of Proposition 2.3 v is symmetric. The argument for $u \in U_r$, $u > r$ is quite similar. □

Recall that a set is said to be *meagre* if it is contained in a countable union of closed nowhere dense sets.

Theorem 2.8. *Let X be an open subset of a strongly ordered separable Banach space V. Then the set $X_C \setminus H$ is meagre.*

Proof. For any simply ordered $L \subset X_C$, $L \setminus Q$ is countable by [10, Theorem 7.3]. The preceding theorem asserts that $(L \cap Q) \setminus H$ is discrete, hence countable. But a subset of a strongly ordered separable Banach space is meagre provided that all its simply ordered subsets are countable [10, Lemma 7.4]. □

Remarks. (a) Note that by [10, Lemma 7.7] one can prove that under the assumptions of Theorem 2.8, $\mu(X_C \setminus H) = 0$ for any Gaussian measure μ on V.

(b) The conclusion of Theorem 2.8 holds if, instead of separability, the ambient Banach space V satisfies any of the hypotheses (a), (b) in [10, Theorem 7.3].

Theorem 2.9. *If, in addition to the hypotheses of Theorem 2.8, $X = X_C$ and the flow is order-compact, then H contains an open dense subset of X.*

Proof. By [10, Theorem 8.12 and Proposition 9.5], the set $\operatorname{int} Q$ is dense in X. We shall prove that $\operatorname{int} H$ is dense in X. Take any open nonempty set $U \subset X$. We have

$$B := U \cap \operatorname{int} Q \neq \emptyset.$$

If $B \subset H$ we are through. Otherwise, let $r \in B \setminus H \subset Q \setminus H$. By an argument similar to that used in the proof of Theorem 2.7 we find an open neighbourhood $V \subset B$ of r such that the trajectory of any member of the nonempty open set

$$D := V \cap \{y \in X : y > r\} \subset U$$

converges to a symmetric equilibrium. Hence $D \subset H$. □

Remark. If for some strongly monotone dynamical system it is known that there exists an open dense subset $X_1 \subset X$ such that for any $x \in X_1$ its trajectory converges to a stable equilibrium (as is the case for some smooth dynamical systems [11, 17]), then Theorem 2.9 follows immediately from Theorem 2.4.

3. Examples from Parabolic Equations

In this section we give examples of second order parabolic partial differential equations which generate strongly monotone dynamical systems satisfying Axiom (GO).

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with smooth boundary $\partial\Omega$. Consider the initial-boundary value problem:

$$u_t(t,x) = \mathscr{F}(u(t,x)) \quad \text{for} \quad t>0, x \in \bar{\Omega} := \Omega \cup \partial\Omega, \tag{3.1}$$

$$\mathscr{B}u(t,x) = 0 \quad \text{for} \quad t>0, x \in \partial\Omega, \tag{3.2}$$

$$u(0,x) = u_0(x) \quad \text{for} \quad x \in \bar{\Omega}, \tag{3.3}$$

where u takes on values in $\mathbb{R}^N$, $N \geqq 1$, $\mathscr{F}: (C^2(\bar{\Omega}))^N \to (C^0(\bar{\Omega}))^N$ is an autonomous strongly elliptic partial differential operator of second order (semilinear or quasilinear), and $\mathscr{B}$ is a boundary operator which is for each component u_i of u either of Dirichlet type:

$$u_i(t,x) = 0 \quad \text{for} \quad t>0, x \in \partial\Omega,$$

or of Neumann type:

$$\partial u_i(t,x)/\partial \nu = 0 \quad \text{for} \quad t>0, x \in \partial\Omega$$

(here ν is the unit normal vector field on $\partial\Omega$ pointing out of Ω).

Under appropriate smoothness conditions imposed on functions included in the operator $\mathscr{F}$ the problem (3.1)–(3.3) defines a flow on a closed subspace X (corresponding to the boundary conditions) of a Sobolev-Slobodeckiĭ space (see [19]) $W_p^\sigma := W_p^\sigma(\Omega, \mathbb{R}^N) \cong (W_p^\sigma(\Omega))^N$, $n < p < \infty$, $1 + n/p < \sigma < 2$ (see [2, 3, 4] or, in the case of semilinear equations, [7]). For semilinear equations one can also get a flow on a subspace of $(C^1(\bar{\Omega}))^N$ (see [16]). Some results of Amann and Mora are collected in [10]. In both cases X is continuously and densely embedded into the product of N spaces, each of them being either $C^0(\bar{\Omega})$ or

$$C_0^1(\bar{\Omega}) := \{u \in C^1(\bar{\Omega}) : u(x) = 0 \text{ for } x \in \partial\Omega\}.$$

The latter two spaces are strongly ordered [10, Sect. 1], so X is strongly ordered by the ordering:

$$u \leqq v \quad \text{if} \quad u_i(x) \leqq v_i(x) \quad \text{for all} \quad x \in \Omega \quad \text{and} \quad i \in \{1, \ldots, N\}$$

(see [10, Corollary 1.12]).

Assume that G is a compact connected subgroup of the group $SO(n)$ of all orthogonal orientation-preserving linear transformations of $\mathbb{R}^n$ whose action leaves the domain $\bar{\Omega}$ invariant:

$$g\bar{\Omega} = \bar{\Omega} \quad \text{for all} \quad g \in G.$$

For a function $v: \bar{\Omega} \to \mathbb{R}^N$ let $\Gamma(g)(v)$ be defined by the formula:

$$\Gamma(g)(v)(x) := v(gx) \quad \text{for} \quad g \in G, x \in \bar{\Omega}. \tag{3.4}$$

It is clear that the map $v \mapsto \Gamma(g)(v)$ does not influence the boundary conditions (3.2). Furthermore, it is a linear isometry on $(C^1(\bar{\Omega}))^N$ as well as on W_p^σ. Appealing to a theorem due to Chernoff and Marsden [5, Theorem 1], we will show that (3.4) defines an action of G on X if we prove that the map $g \mapsto \Gamma(g)(v)$ is continuous for each fixed $v \in X$. It is straightforward that for each $v \in (C^2(\bar{\Omega}))^N$ this is continuous as a map into $(C^2(\bar{\Omega}))^N$, hence as a map into X. Now we use the density of $(C^2(\bar{\Omega}))^N \cap X$ in X (see [19]).

For each $g \in G$, $\Gamma(g)$ is positive, hence monotone. So the triple (X, Φ, G), where Φ is the flow induced on X by (3.1)–(3.3) and G acts on X according to (3.4), satisfies Axiom (GO), provided that Φ is strongly monotone and equivariance ($GO2$) holds. The latter is the case if $\mathscr{F}$ commutes with the action of G:

$$\mathscr{F}(\Gamma(g)(u)(x)) = \Gamma(g)(\mathscr{F}(u(x))) \quad \text{for} \quad g \in G, u \in (C^2(\bar{\Omega}))^N.$$

Now we give examples of such $\mathscr{F}$'s.

Example 1 (Scalar equation). Let $N=1$ and

$$\mathscr{F}(u(x)) := a(x, u, \nabla u)\Delta u + f(x, u, \nabla u),$$

where $a, f: \bar{\Omega} \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ are sufficiently smooth (in general C^2 is needed) and $a(\cdot) \geqq \alpha$ for some positive constant α. By [2, 3, 4] there is a solution flow for (3.1)–(3.3) in a subspace of W_p^σ with σ, p as above. For a independent of u and ∇u one can make use of the results contained in [7] or [16] as well. In this case the resulting flow is compact, and order-compact when in addition f does not depend on ∇u.

The strong comparison principle guarantees that the flow is strongly monotone [10]. As to equivariance it is enough for $h := (a, f)$ to satisfy:

$$h(gx, u, g^{-1}z) = h(x, u, z) \quad \text{for} \quad g \in SO(n), x \in \bar{\Omega}, u \in \mathbb{R}, z \in \mathbb{R}^n,$$

for instance $h = h(r, u, u_r)$ in the polar coordinates.

Example 2 (Strongly cooperative system). Let $N \geqq 1$ and

$$\mathscr{F}(u(x)) := a(u)\Delta u + f(u),$$

where $a(\cdot)$ is an $N \times N$ diagonal matrix function with all entries greater than some positive constant, $f: \mathbb{R}^N \to \mathbb{R}^N$, a and f are sufficiently smooth and $f = (f_1, \ldots, f_N)$ satisfies the strong cooperativity condition: f_i is strictly increasing in u_j for all $i, j \in \{1, \ldots, N\}$, $i \neq j$. Using the theory presented in [2, 3, 4] (or in [7] if a is a constant matrix) one obtains a solution flow for (3.1)–(3.3). If a is constant the flow is order-compact. Strong cooperativity in conjunction with the strong comparison principle implies the strong monotonicity of the flow (cf. [12]). Equivariance is obvious.

Remark. Observe that applying results contained in [3, 4] one can extend our theory to the case of quasilinear parabolic equations under nonlinear boundary conditions.

Acknowledgements. We are grateful to Morris W. Hirsch for sending a preprint of his paper [10], and to André Vanderbauwhede for useful suggestions.

References

1. Amann, H.: Fixed point equations and nonlinear eigenvalue problems in ordered Banach spaces. SIAM Rev. **18**, 620–709 (1976)
2. Amann, H.: Quasilinear evolution equations and parabolic systems. Trans. Am. Math. Soc. **293**, 191–227 (1986)
3. Amann, H.: Quasilinear parabolic systems under nonlinear boundary conditions. Arch. Rat. Mech. Anal. **92**, 153–192 (1986)

4. Amann, H.: Parabolic evolution equations and nonlinear boundary conditions. J. Differ. Equations **72**, 201–269 (1988)
5. Chernoff, P., Marsden, J.E.: On continuity and smoothness of group actions. Bull. Am. Math. Soc. **76**, 1044–1049 (1970)
6. Golubitsky, M., Schaeffer, D.: Singularities and groups in bifurcation theory. Applied Mathematical Sciences 51. Berlin Heidelberg New York: Springer 1986
7. Henry, D.: Geometric theory of semilinear parabolic equations. Lecture Notes Mathematics, Vol. 840. Berlin Heidelberg New York: Springer 1981
8. Hirsch, M.W.: Differential equations and convergence almost everywhere in strongly monotone flows. Contemp. Math. **37**, 267–282 (1983)
9. Hirsch, M.W.: The dynamical systems approach to differential equations. Bull. Am. Math. Soc. , New Ser. **11**, 1–64 (1984)
10. Hirsch, M.W.: Stability and convergence in strongly monotone dynamical systems. J. Reine Angew. Math. **383**, 1–53 (1988)
11. Lions, P.L.: Structure of the set of steady-state solutions of semilinear heat equations. J. Differ. Equations **53**, 362–386 (1984)
12. Martin, Jr., R.H.: A maximum principle for semilinear parabolic systems. Proc. Am. Math. Soc. **74**, 66–70 (1979)
13. Matano, H.: Existence of nontrivial unstable sets for equilibriums of strongly order-preserving systems. J. Fac. Sci., Univ. Tokyo, Sect. I A **30**, 645–673 (1984)
14. Matano, H.: Strongly order-preserving local semidynamical systems – theory and applications. In: Semigroups, theory and applications, Vol. I, Trieste 1984. (Pitman Research Notes Mathematics Series 141, pp. 178–185) Harlow: Longman 1986
15. Montgomery, D., Zippin, L.: Topological transformation groups. Interscience Tracts in Pure and Applied Mathematics 1. New York: Interscience 1955
16. Mora, X.: Semilinear problems define semiflows on C^k spaces. Trans. Am. Math. Soc. **278**, 21–55 (1983)
17. Poláčik, P.: Convergence in smooth strongly monotone flows defined by semilinear parabolic equations. J. Differ. Equations (to appear)
18. Sattinger, D.H.: Bifurcation and symmetry breaking in applied mathematics. Bull. Am. Math. Soc. , New Ser. **3**, 779–819 (1980)
19. Triebel, H.: Interpolation theory, function spaces, differential operators. Berlin: Deutscher Verlag der Wissenschaften 1978
20. Vanderbauwhede, A.: Local bifurcation and symmetry. Research Notes Mathematics, Vol. 75. Boston: Pitman 1982

Received November 10, 1987

Math. Ann. 283, 13–52 (1989)

Mathematische Annalen

Kähler Spaces and Proper Open Morphisms

Jean Varouchas

Université de Nancy I, Faculté des Sciences, Département de Mathématiques, B.P. n° 239, F-54506 Vandœuvre les Nancy Cedex, France

Contents

Introduction

Several years ago, Hironaka [17] raised the following two problems:

Problem A. Let X be a Kähler space. Is the Douady space of X Kähler?

and its weaker version

Problem B. Let $\pi: X \to X'$ be a proper flat surjective morphism of complex spaces. If X is Kähler, is X' Kähler?

Actually problems A and B were raised for compact X but we will consider the non-compact case as well.

Problem A seems inaccessible for the moment.

Problem B was solved affirmatively in [23] for smooth X, X'. The aim of the present paper is to generalize the result to singular spaces. It appears that the flatness hypothesis on π is too strong, so it will be replaced by a less restrictive property which we call *geometric flatness*.

A closely related problem, raised by Lieberman [18] is

Problem C. Let X be a Kähler space and $\mathbf{B}_m(X)$ the Barlet space of compact complex m-cycles of X. Is $\mathbf{B}_m(X)$ Kähler?

A solution to problem C would imply one to problem B for geometrically flat π and reduced X'.

Finally a problem which is of fundamental importance in the theory of complex cycles is

Problem D. Let X be a complex space and $\xi \in H^m(X, \Omega_X^m)$. Is the function $F_\xi : c \mapsto (c \cdot \xi)$ holomorphic on $\mathbf{B}_m(X)$?

Our results can be summarized as follows: Problems B and C are reduced to problem D; problem D has a solution (for fixed X, m) if every compact m-dimensional complex-analytic subset of X has a smoothly embeddable neighborhood (Chap. I, Proposition 3.5.4).

In order to formulate our results completely, and as long as problem D remains unsolved in its full generality, we are led to introduce the notion of *weakly Kähler spaces*. The most useful properties of geometrically flat morphisms and weakly Kähler spaces are

(i) A geometrically flat morphism is proper open surjective with pure dimensional fibers and reduced base. The converse is true if the morphism is flat or the base is normal.

(ii) If G is a finite group of automorphisms of a reduced space X then the canonical projection $X \to X/G$ is geometrically flat.

(iii) Kähler spaces are weakly Kähler. Subspaces of weakly Kähler spaces are weakly Kähler. X is weakly Kähler iff X_{red} is weakly Kähler. A weakly normal space is weakly Kähler iff it is Kähler.

(iv) A compact space is weakly Kähler iff its weak normalization is Kähler.

Now we may enumerate our main results:

(i) If $\pi : X \to X'$ is geometrically flat with m-dimensional fibers, then problem B has a solution for π if problem D has a solution for X, m. Otherwise all we can say is that X' is weakly Kähler. But this is enough to ensure that X' is Kähler if it is normal (Chap. IV, Theorem 3–Corollary 1.2).

(ii) Problem C has a solution for X, m if problem D has a solution for X, m. Otherwise all we can say is that $\mathbf{B}_m(X)$ is weakly Kähler. But this is enough to ensure that, if X is compact, the weak normalization of $\mathbf{B}_m(X)$ is Kähler (Chap. IV, Theorem 4–Corollary 2.2).

(iii) The solution of problem B for normal X' implies that any reduced compact complex space in Fujiki's class $\mathscr{C}$ (holomorphic image of a compact Kähler space) is bimeromorphically equivalent to a compact Kähler space

(Chap. IV, Theorem 5). An alternative proof of this was given in [24] using the solution of problem C for smooth X.

Our paper is organized as follows:

In Chap. I we give a rapid discussion of the sheaf $\mathscr{C}_X^\infty$ in the sense we choose for a complex space X. $\mathscr{C}_X^\infty$ is *not* a subsheaf of the sheaf $\mathscr{C}_X$ of continuous complex-valued functions on X; there is only a canonical morphism $\varphi \mapsto [\varphi]$ from $\mathscr{C}_X^\infty$ to $\mathscr{C}_X$. This is important for the formulation of a smoothing lemma (2.5) for continuous strongly plurisubharmonic (p.s.h.) functions which is essentially due to Richberg [21]. We also remind some of the main properties of the Barlet space $\mathbf{B}_m(X)$ which we will use. Geometric flatness is defined in 3.3.

In Chap. II we define the notions of Kähler metrics, classes, spaces, and morphisms and prove *Theorem 1* (valid on any complex space) according to which, a space is Kähler if it admits an open covering $\mathscr{U}$ with 0-cochain $\varphi=(\varphi_\alpha)$ of *continuous* strongly p.s.h. functions and a 1-cocycle $h=(h_{\alpha\beta})$ of pluriharmonic functions such that $\delta\varphi=[h]$ in $C^1(\mathscr{U},\mathscr{C}_X)$. (The cocycle condition on h is redundant only for X reduced). As a consequence we solve problem B for *finite* $\pi: X\to X'$ such that either π is flat and X' arbitrary (not necessarily reduced) or π is geometrically flat and X' reduced. If X is a Kähler space and G a finite group of automorphisms of X, X/G is Kähler. In particular, $\mathrm{Sym}^k(X)$ is Kähler for any $k\geqq 1$ (Corollary 3.2.1). Finally we define weakly Kähler spaces in 4.1.

Chapter III is entirely devoted to the proof of *Theorem 2*: if X is a complex space and $m\geqq 0$ an integer, then there are open sets $U_\alpha\subset X$ and $U^j_{\alpha\beta}\subset U_\alpha\cap U_\beta$ such that any compact m-dimensional complex-analytic subset of X (resp. $U_\alpha\cap U_\beta$) is contained in some U_α (resp. $U^j_{\alpha\beta}$). Moreover, if ω is a Kähler form on X, then there are (m,m)-forms $\chi_\alpha=\bar\chi_\alpha$ on U_α, $\tau^j_{\alpha\beta}$ on $U^j_{\alpha\beta}$ such that $\omega^{m+1}|U_\alpha=i\partial\bar\partial\chi_\alpha$, $\bar\partial\tau^j_{\alpha\beta}=0$, $(\chi_\alpha-\chi_\beta)|U^j_{\alpha\beta}=\tau^j_{\alpha\beta}+\bar\tau^j_{\alpha\beta}$ and the $\bar\partial$-cohomology class of $\tau^j_{\alpha\beta}$ lies in the image of the canonical morphism $H^m(U^j_{\alpha\beta},\Omega^m)\to H^{m,m}_{\bar\partial}(U^j_{\alpha\beta})$.

Theorem 2 is the main original element of this paper. It relies on Barlet's result [6] according to which m-dimensional compact complex-analytic subsets admit m-complete neighborhoods. For smooth X, Theorem 2 can be easily deduced from this [23, Lemma 3.6] and [24, 2.8] using the Dolbeault isomorphism. For singular X, this is considerably more difficult. Our method can be described as follows: When a complex of sheaves $(\mathscr{L}^\cdot, D)$ fails to be exact, we replace it by the single complex associated to the double complex (δ, D) where δ is the Čech differential with respect to some open covering. We call this new complex the *Čech transform* of $(\mathscr{L}^\cdot, D)$ and apply it to the $\partial\bar\partial$-complex $\mathscr{L}^\cdot_m$ (defined in 3.1). The key step is the existence of a cocycle Φ_{m+1} of degree $2m+2$ (defined in 4.3) of the Čech transform of the complex $\mathscr{L}^\cdot_{m+1}$ whose final component is ω^{m+1}. Using an elementary lemma of algebra (Lemma 2.2) we prove that Φ_{m+1} bounds near every m-dimensional compact complex-analytic subset of X, so ω^{m+1} is $\partial\bar\partial$-exact there. The last part of Theorem 2 relies on two morphisms β and γ (defined in 3.5) connecting the $\partial\bar\partial$-complex to the direct sum of the Dolbeault complex and its conjugate. Chapter III is self-contained.

Finally Chap. IV proves the main results we obtain as consequences of Theorems 1 and 2, namely Theorems 3–5 and corollaries.

List of Symbols

I. Ω_X^m, $\mathscr{C}_X$, $\mathscr{F}_{X,\mathbb{R}}$, $\mathscr{F}(U,\mathbb{R})$

1.1. $\mathscr{C}_X^\infty$, A_X^m, $A_X^{k,l}$, PH_X

1.2. $[\varphi]$, $[\mathscr{C}_X^\infty]$

2.1. P_X^0, SP_X^0, P_X^∞, SP_X^∞, $[P_X^\infty]$, $[SP_X^\infty]$

2.4. $SP^{0,\infty}(U,V)$

2.6. SP_π^∞

3.1. $\mathrm{Sym}^k(X)$, $\sum\{x_j\}$

3.2. $\mathbf{B}_m(X)$

3.3. $\mathbf{D}_m(X)$, $c(Y)$

3.4. $F_\varphi(c)$, $(c\cdot\xi)$

3.5. $\pi_*\varphi$, $\mathbf{B}_m(X)^{(0)}$

II.

1.1. $\mathscr{H}_X^1$, $\mathscr{H}_{X,\mathbb{R}}^1$, $\partial\bar{\partial}\kappa$

1.2. $\hat{c}_1$, c_1, $\tilde{c}_1$

3.1. $\mathrm{Tr}^{(c)}_{X/X'}$, $\mathrm{Tr}^{(h)}_{X/X'}$

4.1. $\mathscr{W}_X$, WPH_X, $\hat{X}$

III.

1.1. $\underline{X}$, $F:\underline{X}\to\underline{Y}$, $\underline{U}\ll\underline{X}$, $\underline{U}_1\cap\underline{U}_2$, ε, δ, $\varphi|\underline{U}$, T

1.2. $\varphi\cdot\psi$

2.1. $\check{C}^q(\underline{X};\mathscr{F},\mathscr{L}^\cdot)$, Δ, $\check{Z}^q(\underline{X};\mathscr{F},\mathscr{L}^\cdot)$, $\check{H}^q(\underline{X};\mathscr{F},\mathscr{L}^\cdot)$

3.1. $\mathscr{L}_m^r$, D

3.3 φ^*, $\mathscr{L}_{m,\mathbb{R}}^r$

3.4. μ

3.5. $\mathscr{G}_m^q$, $\hat{d}$, β, γ

3.6. $\mathscr{E}_m^q(\underline{X})$, $\mathscr{E}_m^q(\underline{X},[\mathbb{R}])$, $\mathscr{E}_m^q(\underline{X},\mathbb{R})$

4.2. $\Phi_1(f,\varphi)$

4.3. $\mathscr{K}^m(\underline{X})$, $\mathscr{K}(\underline{X})$, $\mathscr{K}^m(\underline{X},[\mathbb{R}])$, $\mathscr{K}(\underline{X},[\mathbb{R}])$, $\mathscr{K}^m(\underline{X},\mathbb{R})$, $\mathscr{K}(\underline{X},\mathbb{R})$

4.4. $\Phi\times\Psi$

5.4. $\mathscr{D}_m^q(\underline{X})$, $\hat{\Delta}$

IV.

3.1. $\mathscr{C}$, $\mathscr{C}^*$

I. Preliminaries

X will always denote a complex space, not necessarily reduced unless explicitly stated. $X_{\mathrm{red}}\to X$ denotes the reduction of X. $\mathscr{O}_X=\Omega_X^0$ is the structure sheaf of X and Ω_X^m the sheaf of holomorphic m-forms on X. $\mathscr{C}_X$ is the sheaf of continuous functions on the topological space underlying to X. If $\mathscr{F}=\mathscr{F}_X$ is any sheaf on X, $\mathscr{F}(U)$ will denote $\Gamma(U,\mathscr{F}_X)$. If $\mathscr{F}_X$ is a sheaf of $\mathbb{C}$-vector spaces with a natural $\mathbb{C}$-antilinear involution, $\mathscr{F}_{X,\mathbb{R}}$ will denote the subsheaf of elements left fixed by the involution and $\mathscr{F}(U,\mathbb{R}):=\Gamma(U,\mathscr{F}_{X,\mathbb{R}})$. We always assume X countable at infinity.

1. $\mathscr{C}^\infty$ Forms and Functions on Complex Spaces

There are two inequivalent definitions of $\mathscr{C}_X^\infty$ in the literature. The first, which we call the "old" one [5, 10, 21] defines $\mathscr{C}_X^\infty$ as the subsheaf of $\mathscr{C}_X$ consisting of local restrictions of $\mathscr{C}^\infty$ functions under smooth embeddings. So $\mathscr{C}^\infty_X = \mathscr{C}^\infty_{X_{\text{red}}}$ in this sense. The second which we will call thee "modern" one [8, 12] is the one we give below.

1.1. Definitions. We define on X the sheaves $\mathscr{C}^\omega_X$ of real-analytic functions, PH_X of pluriharmonic functions, $\mathscr{C}^\infty_X = A^0_X$ of $\mathscr{C}^\infty$ functions, A^m_X (resp. $A^{k,l}_X$) of $\mathscr{C}^\infty$ m-forms [resp. (k, l)-forms] as follows: For smooth X, they are well defined. Now suppose $X \to D$ is an embedding of X in a domain D of $\mathbb{C}^n$ and $\mathscr{I}_X \subset \mathcal{O}_D$ is the corresponding coherent ideal sheaf. Set

$$\mathscr{I}^\omega_X := (\mathscr{I}_X + \bar{\mathscr{I}}_X)\mathscr{C}^\omega_D, \qquad \mathscr{I}^\infty_X := (\mathscr{I}_X + \bar{\mathscr{I}}_X)\mathscr{C}^\infty_D$$

and

$$\mathscr{C}^\omega_X := \mathscr{C}^\omega_D / \mathscr{I}^\omega_X, \qquad \mathscr{C}^\infty_X := \mathscr{C}^\infty_D / \mathscr{I}^\infty_X, \qquad A^m_X := A^m_D / (\mathscr{I}^\infty_X A^m_D + d\mathscr{I}^\infty_X A^{m-1}_D)$$

$$A^{k,l}_X := \text{the image of } A^{k,l}_D \text{ under the canonical morphism } \quad A^{k+l}_D \to A^{k+l}_X .$$

It is clear that these sheaves are independent of the choice of the embedding $X \to D$ so they extend to arbitrary X. There are canonical morphisms

$$\mathcal{O}_X \to \mathscr{C}^\omega_X \to \mathscr{C}^\infty_X \to \mathscr{C}_X .$$

1.2. Elementary Properties and Conventions. (i) The canonical morphisms $\mathcal{O}_X \to \mathscr{C}^\omega_X$ and $\mathscr{C}^\omega_X \to \mathscr{C}^\infty_X$ are injective. (The first is elementary and the second is a consequence of the fact that $\mathscr{C}^\infty_D$ is a faithfully flat $\mathscr{C}^\omega_D$-module by Malgrange [19, Chap. VI, Corollary 1.12].) They will be considered as inclusions

$$\mathcal{O}_X \subset \mathscr{C}^\omega_X \subset \mathscr{C}^\infty_X$$

and so we may define $PH_X := \mathcal{O}_X + \bar{\mathcal{O}}_X \subset \mathscr{C}^\omega_X$.

(ii) In $\mathscr{C}^\omega_X$ we have $\mathcal{O}_X \cap \bar{\mathcal{O}}_X = \mathbb{C}$ and there is a commutative diagram with exact rows

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathbb{R} & \longrightarrow & \mathcal{O}_X & \xrightarrow{-2\,\mathrm{Im}} & PH_{X,\mathbb{R}} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow{\scriptstyle\lambda} & & \downarrow & & \\
0 & \longrightarrow & \mathbb{C} & \longrightarrow & \mathcal{O}_X \oplus \bar{\mathcal{O}}_X & \xrightarrow{(i\ \ -i)} & PH_X & \longrightarrow & 0,
\end{array}$$

where $\lambda(f) = (f, \bar{f})$ and the unspecified morphisms are the canonical inclusions.

(iii) The canonical morphism $\varrho : \mathscr{C}^\infty_X \to \mathscr{C}_X$ is not injective in general even for X reduced; for $fg = 0$ in $\mathcal{O}_X$ does not imply $f\bar{g} = 0$ in $\mathscr{C}^\omega_X$. However, for X *reduced and locally irreducible*, ϱ is injective. (It is elementary that the restriction of ϱ to $\mathscr{C}^\omega_X$ is injective; we deduce that ϱ is injective by Malgrange [19, Chap. VI, Theorem 3.10].)

We write $[\varphi] := \varrho(\varphi)$, $[\mathscr{C}^\infty_X] := \varrho(\mathscr{C}^\infty_X)$. So $[\mathscr{C}^\infty_X]$ is the $\mathscr{C}^\infty$ sheaf of the "old" theory. For normal X the two theories coincide, by the above remark.

(iv) The kernel of the canonical morphism $PH_X \to \mathscr{C}_X$ is $\mathscr{N}_X + \bar{\mathscr{N}}_X$ where $\mathscr{N}_X$ is the sheaf of nilpotent sections of $\mathcal{O}_X$. In particular, for reduced X, PH_X may be considered as a subsheaf of $\mathscr{C}_X$.

(v) If $f : X \to Y$ is a morphism of complex spaces, $\varphi \in \mathscr{C}(Y)$ and $\psi \in \mathscr{C}^\infty(Y)$, write $\varphi \circ f \in \mathscr{C}(X)$ and $f^*\psi \in \mathscr{C}^\infty(X)$ for the corresponding induced elements. Write $\psi \circ f$ instead of $[\psi] \circ f$, so that $[f^*\psi] = \psi \circ f$ in $\mathscr{C}(X)$.

(vi) The canonical morphisms $\Omega_X^m \to A_X^{m,o}$ are injective and will be considered as inclusions.

(vii) The inclusions $A_X^{k,l} \subset A_X^{k+l}$ give a direct sum decomposition $A_X^m = \bigoplus_{k+l=m} A_X^{k,l}$ and $A_X^{\cdot}$ is a bigraded algebra with respect to the wedge product. The natural involution $\varphi \mapsto \bar{\varphi}$ applies $A_X^{k,l}$ on to $A_X^{l,k}$.

(viii) There is a canonical morphism $d = \partial + \bar{\partial}: A_X^m \to A_X^{m+1}$ satisfying the usual identities $d^2 = \partial^2 = \bar{\partial}^2 = \partial\bar{\partial} + \bar{\partial}\partial = 0$. However, none of the resulting complexes (Dolbeault, De Rham, etc. ...) is an exact sequence of sheaves in general.

(ix) Any morphism $f: X \to Y$ of complex spaces gives rise to a linear $f^*: A^m(Y) \to A^m(X)$ which is compatible with the wedge product, bigraduation and the operators $d, \partial, \bar{\partial}$. We have $(fg)^* = g^* f^*$.

2. *Strongly Plurisubharmonic Functions*

We write p.s.h. for plurisubharmonic.

2.1. Definitions. We define on X the sheaves of real convex cones P_X^0 (resp. SP_X^0) of continuous p.s.h. (resp. strongly p.s.h.) functions, P_X^∞ (resp. SP_X^∞) of $\mathscr{C}^\infty$ p.s.h. (resp. strongly p.s.h.) functions as the subsheaves of $\mathscr{C}_{X,\mathbb{R}}$ (resp. $\mathscr{C}_{X,\mathbb{R}}^\infty$) consisting of elements induced by corresponding functions on open sets of $\mathbb{C}^n$ under local embeddings. Also define $[P_X^\infty] := \varrho(P_X^\infty)$, $[SP_X^\infty] := \varrho(SP_X^\infty)$ where $\varrho: \mathscr{C}_X^\infty \to \mathscr{C}_X$ is the canonical morphism.

2.2. Examples. (i) On the subspace X of $\mathbb{C}^2$ defined by $z_1 z_2 = z_2^2 = 0$, set $\varphi_t(z_1, z_2) := z_1 \bar{z}_1 + t z_2 \bar{z}_2$ for real t. Then $[\varphi_t] \in SP^0(X)$ is independent of t, $\varphi_t \in \mathscr{C}^\infty(X, \mathbb{R})$ for all t but $\varphi_t \in P^\infty(X)$ only for $t \geqq 0$ and $\varphi_t \in SP^\infty(X)$ only for $t > 0$.

(ii) On the subspace X of $\mathbb{C}^2$ defined by $z_1 z_2 = 0$, set (for real t) $\varphi_t(z_1, z_2) := z_1 \bar{z}_1 + t(z_1 \bar{z}_2 + z_2 \bar{z}_1) + z_2 \bar{z}_2$. Then $[\varphi_t] \in SP^0(X)$ is independent of t, $\varphi_t \in \mathscr{C}^\infty(X, \mathbb{R})$ but $\varphi_t \in P^\infty(X)$ only for $|t| \leqq 1$ and $\varphi_t \in SP^\infty(X)$ only for $|t| < 1$.

(iii) On $\mathbb{C}^n$ set $\varphi(z_1, \dots, z_n) := \sum_{j=1}^n |t_j|^2$ where $t_1, \dots, t_n$ are the roots of $X^n - z_1 X^{n-1} + \dots + (-1)^n z_n$. Then $\varphi \in SP^0(\mathbb{C}^n)$.

2.3. The Cone $SP^{0,\infty}(U, V)$. This is an auxiliary notion introduced to give a meaning to smoothing lemmas of strongly p.s.h. functions. For U, V open in X, $SP^{0,\infty}(U, V)$ is defined as the set of pairs $\varphi = (\varphi^0, \varphi^\infty) \in SP^0(U) \times SP^\infty(U \cap V)$ such that $[\varphi^\infty] = \varphi^0{}_{|U \cap V}$. We set $[\varphi] := \varphi^0$. The following are obvious

(i) $SP^{0,\infty}(U, V) = SP^{0,\infty}(U, U \cap V)$.

(ii) $SP^{0,\infty}(U, \emptyset) \cong SP^0(U)$ canonically.

(iii) $SP^{0,\infty}(U, X) \cong SP^\infty(U)$ canonically.

(iv) For fixed V, $U \mapsto SP^{0,\infty}(U, V)$ is a sheaf on X.

(v) For $\varphi = (\varphi^0, \varphi^\infty) \in SP^{0,\infty}(U, V)$ and $h \in PH(U, \mathbb{R})$, the element $\varphi + h := (\varphi^0 + [h], \varphi^\infty + h_{|U \cap V})$ is in $SP^{0,\infty}(U, V)$.

The following is a slight improvement of a result of Richberg [21, Satz 4.1]. For $X = \mathbb{C}^n$, a complete proof is in [23].

2.4. Richberg's Lemma. *Let U, V, W be open in X with $U \subset\subset W$. Let $\varphi \in SP^{0,\infty}(W, V)$. Then there is a compact S such that $U \subset S \subset W$ and an ele-*

ment $\psi \in SP^{0,\infty}(W, U \cup V)$ *such that* $\varphi|_{W\setminus S} = \psi|_{W\setminus S}$ *in* $SP^{0,\infty}(W\setminus S, V) = SP^{0,\infty}(W\setminus S, U \cup V)$.

Sketch of Proof. Take a finite number of open sets $U_k \subset\subset V_k \subset\subset W_k$ $(1 \leqq k \leqq m)$ such that $U = \bigcup_{k=1}^{m} U_k$ and each W_k is embedded in an open subset D_k^0 of $\mathbb{C}^{n_k}$ such that $[\varphi]|_{W_k}$ is induced by an element of $SP^0(D_k)$. Using the method of [23] one can construct inductively elements $\varphi_k \in SP^{0,\infty}(W, U_1 \cup \ldots \cup U_k \cup V)$ such that $\varphi_k|_{W\setminus \bar{V}_k} = \varphi_{k-1}|_{W\setminus \bar{V}_k}$. Then set $S = \bar{V}_1 \cup \ldots \cup \bar{V}_m$ and $\psi = \varphi_m$.

2.5. The Fornaess-Narasimhan Theorem [10, Theorem 5.3.1]. *Let* $\varphi \in \mathscr{C}(X, \mathbb{R})$. *Suppose that for any holomorphic* $f: \Delta \to X$, *where* Δ *is the unit disc of* $\mathbb{C}$, $\varphi \circ f$ *is subharmonic on* Δ. *Then* $\varphi \in P^0(X)$.

2.6. The Cone $SP_\pi^\infty(X)$. Let $\pi: X \to Y$ be a morphism of complex spaces. Let $\varphi \in \mathscr{C}^\infty(X, \mathbb{R})$. We say that φ is *strongly p.s.h. relatively to* π and write $\varphi \in SP_\pi^\infty(X)$ if for any $x \in X$ there are open subsets $U \subset X$, $V \subset Y$ and $\psi \in SP^\infty(V)$ such that $x \in U \subset \pi^{-1}(V)$ and $(\varphi + \pi^*\psi)|_U \in SP^\infty(U)$.

3. Barlet's Space of Analytic Cycles

3.1. Symmetric Powers of Complex Spaces. If $k \geqq 1$ is an integer, let $\mathrm{Sym}^k(X) := X^k/\mathscr{S}_k$ be the quotient of X^k under the action of the symmetric group permuting components. Denote by $\sum_{j=1}^{k} \{x_j\}$ the image of $(x_1, \ldots, x_k)$ in $\mathrm{Sym}^k(X)$ under the canonical projection.

3.2. Analytic Families of Complex Cycles. $\mathbf{B}_m(X)$. Let X be reduced and $m \geqq 0$ an integer. A *compact complex-analytic m-cycle* (or briefly *m-cycle*) of X is a formal finite sum

$$c = \sum_{i \in I} n_i Y_i,$$

where $n_1 \geqq 1$ are integers and Y_i are compact irreducible m-dimensional complex-analytic subsets of X. $|c| := \bigcup_{i \in I} Y_i$ is called the *support* of c.

Let c be as above and $\sigma: V \to U \times B$ an embedding of an open set $V \subset X$ into a connected open set $U \times B$ of $\mathbb{C}^N = \mathbb{C}^m \times \mathbb{C}^{N-m}$. We say that $\mathscr{V} = (\sigma, V, U \times B)$ is a *well-adapted chart with respect to c* if σ extends to an embedding $\sigma_1: V_1 \to U_1 \times B_1$ such that $V \subset\subset V_1 \subset X$, $U \subset\subset U_1 \subset \mathbb{C}^m$, $B \subset\subset B_1 \subset \mathbb{C}^{N-m}$ and $\sigma_1(|c|) \cap (\bar{U} \times \partial B) = \emptyset$.

If we set $Z_i := \sigma(V \subset Y_i) \subset U \times B$, then the projection $U \times B \to U$ restricted to each Z_i is a branched covering $\pi_i: Z_i \to U$ of finite degree k_i and defines as such a morphism $\psi_i: U \to \mathrm{Sym}^{k_i}(B)$. Set $k := \sum n_i k_i$, $\psi := \sum n_i \psi_i: U \to \mathrm{Sym}^k(B)$, $\deg(c, \mathscr{V}) := k$.

Now let S be a reduced complex space and $(c_s)_{s \in S}$ a family of m-cycles of X parametrized by S. We say that (c_s) is an *analytic family of cycles* if for any $s_0 \in S$ and for *any* well-adapted chart $\mathscr{V}$ with respect to c_{s_0}, there is a neighborhood T of s_0 in S such that

(i) $\mathscr{V}$ is well-adapted with respect to c_s for all $s \in T$.
(ii) $\deg(c_s, \mathscr{V}) = k$ is independent of $s \in T$.
(iii) The resulting map $\psi: U \times T \to \mathrm{Sym}^k(B)$ is holomorphic.

The Barlet space $\mathbf{B}_m(X)$ of m-cycles of X is a reduced complex space, constructed in [3], whose points are the m-cycles of X forming a tautological analytic family and such that for any analytic family $(c_s)_{s\in S}$ of m-cycles of X, there is a unique morphism of complex spaces $H: S \to \mathbf{B}_m(X)$ such that

$$H(s) = c_s \quad \text{for all} \quad s \in S.$$

For X not necessarily reduced, we set

$$\mathbf{B}_m(X) := \mathbf{B}_m(X_{\mathrm{red}}).$$

3.3. Proper Open Morphism. Geometric Flatness. Let $\mathbf{D}_m(X)$ be the Douady space [9] of compact subspaces of pure dimension m of X. In [3, Chap. 5], Barlet constructed a canonical morphism

$$c: (\mathbf{D}_m(X))_{\mathrm{red}} \to \mathbf{B}_m(X).$$

If Y is a point of $\mathbf{D}_m(X)$ (a subspace of X) then $c(Y) = \sum n_i Y_i$ where Y_i are the irreducible components of Y_{red} and $n_i \geqq 1$ integers called multiplicities. If Y is generically reduced, all n_i are equal to 1.

Now suppose that $\pi: X \to X'$ is a morphism of complex spaces such that, for some fixed $m \geqq 0$

$$\text{(3.3)} \qquad \begin{aligned} &\text{(i) } \pi \text{ is proper open and surjective,}\\ &\text{(ii) all fibers of } \pi \text{ are of pure dimension } m,\\ &\text{(iii) } X' \text{ is reduced.}\end{aligned}$$

[If X, X' are pure dimensional, then (i) implies (ii).]

We will say that π is *geometrically flat* if there is a morphism of complex spaces

$$H: X' \to \mathbf{B}_m(X)$$

such that $H(x') = c(\pi^{-1}(x'))$ generically on X'. We call H the *classifying morphism of* π. The domain of validity of the equality $H(x') = c(\pi^{-1}(x'))$ is the dense Zariski open set U' of points of flatness of π (Frisch [11]).

3.3.1. Proposition. *Suppose* $\pi: X \to X'$ *satisfies* (3.3). *Then:*
(i) *If* π *is flat, then it is geometrically flat.*
(ii) *If* X' *is normal, then* π *is geometrically flat.*
(iii) *If* π *is geometrically flat, then* H *defines an isomorphism of* X' *onto a subspace of* $\mathbf{B}_m(X)$.

Proof. (i) If π is flat, then there is a morphism $X' \to \mathbf{D}_m(X)$, factoring through $(\mathbf{D}_m(X))_{\mathrm{red}}$ since X' is reduced, taking the value $\pi^{-1}(x')$ at x'. Composing with $c: (\mathbf{D}_m(X))_{\mathrm{red}} \to \mathbf{B}_m(X)$, we obtain the required H.
(ii) This is part of Theorem 1 of [3].
(iii) This is shown in [24, Appendix, p. 259].

3.3.2. Examples. (i) Let X be the union of two planes defined by $z_1z_2 = z_1z_4 = z_2z_3 = z_3z_4 = 0$ in $\mathbb{C}^4$, Y' the union of two lines defined by $x_1x_2 = 0$ in $\mathbb{C}^2$, $\pi: X \to X' = \mathbb{C}^2$ and $\varrho: Y' \to X'$ defined by $\pi(z_1, z_2, z_3, z_4) = (z_1 + z_2, z_3 + z_4)$ and $\varrho(x_1, x_2) = (x_1, 0)$

$$\begin{tikzcd} X & Y = X \times_{X'} Y' \arrow[l] \arrow[d, "\pi_1"] \\ X' \arrow[u, leftarrow, "\pi"'] & Y' \arrow[l, "\varrho"'] \end{tikzcd}$$

Then π is geometrically flat be 3.3.1 (ii) but π_1 is not since Y consists of one triple line over one branch of Y' and two single lines over the other. π is not flat.

(ii) Let X be the union of two single lines and one double line defined by $z_1z_2 = z_2^2 - z_3^2 = 0$ in $\mathbb{C}^2$ and X' the union of two lines $z_1z_2 = 0$ (as Y' above).

If $\pi(z_1, z_2, z_3) = (z_1, z_2)$, then $\pi: X \to X'$ is flat, X' is reduced but if $r: X_{\mathrm{red}} \to X$ is the reduction of X then $\pi r: X_{\mathrm{red}} \to X'$ is not geometrically flat.

3.4. Integration of Differential Forms. If $\varphi \in A^{m,m}(X)$ and $c = \sum n_i Y_i \in \mathbf{B}_m(X)$, define

$$F_\varphi(c) := \int_c \varphi = \sum n_i \int_{Y_i} \varphi .$$

If $\pi: X \to X'$ is geometrically flat with m-dimensional fibers and φ is a above, define

$$\pi_* \varphi := F_\varphi \circ H_\pi .$$

We have the following:

3.4.1. Proposition [4, 5, 23]. *With the above notations.*

(i) *F_φ (resp. $\pi_*\varphi$) is continuous on $\mathbf{B}_m(X)$ (resp. X').*

(ii) *If $d\varphi = 0$, then F_φ and $\pi_*\varphi$ are locally constant.*

(iii) *If $\varphi = \bar{\varphi}$ and $i\partial\bar{\partial}\varphi \geqq 0$ then F_φ and $\pi_*\varphi$ are p.s.h.*

(iv) *If $\varphi = \bar{\varphi}$ and $i\partial\bar{\partial}\varphi \gg 0$ then F_φ and $\pi_*\varphi$ are strongly p.s.h.*

(v) *If $\bar{\partial}\varphi = 0$ then F_φ and $\pi_*\varphi$ are weakly holomorphic; if moreover X is smooth, they are holomorphic.*

3.4.2. Remark. Case (iii) above needs the Fornaess-Narasimhan theorem if we look at the proof of Proposition 1 of [5].

3.4.3. Definition. A $\bar{\partial}$-closed $\tau \in A^{m,m}(X)$ is said to *represent* an element $\xi \in H^m(X, \Omega_X^m)$ (or to be a *$\bar{\partial}$-closed representative of* ξ) if the class of τ in $H_{\bar{\partial}}^{m,m}(X)$ is the image of ξ under the canonical morphism $H^m(X, \Omega_X^m) \to H_{\bar{\partial}}^{m,m}(X)$. In that case we define $F_\xi(c) := F_\tau(c)$ for $c \in \mathbf{B}_m(X)$ and also write $(c \cdot \xi)$ for $F_\xi(c)$ (since it depends on ξ alone).

3.5. m-Complete and m-Admissible Neighborhoods. By the Andreotti-Grauert theorem [1], if X is a m-complete complex space, then for any coherent analytic sheaf $\mathscr{F}$ on X and any $q > m$ we have $H^q(X, \mathscr{F}) = 0$. We will use

3.5.1. Proposition. *Let Y be a compact m-dimensional complex-analytic subset of X. Then*

(i) *Y admits in X a fundamental system of m-complete neighborhoods (Barlet [6]).*

(ii) *Y admits in X a fundamental system of neighborhoods V such that $H^k(V, \mathbb{R})=0$ for $k>2m$* [23, *Lemma 3.5*].

3.5.2. Definition. An open $U \subset X$ is said to be *m-admissible* if

(i) U is m-complete.

(ii) There is an open V such that $U \subset V \subset X$ and $H^k(V, \mathbb{R})=0$ for all $k>2m$.

3.5.3. Remark. If X is a Kähler manifold with a Kähler form ω and $U \subset X$ is 0-admissible, then one easily sees that $\omega|_U = i\partial\bar{\partial}\varphi$ for some $\varphi \in SP^\infty(U)$. This is the most trivial particular case of our Theorem 2.

3.5.3. Proposition. .(i) *If $U \subset X$ is m-admissible and $k>2m$, then the canonical morphism $H^k(X, \mathbb{R}) \to H^k(U, \mathbb{R})$ is zero.*

(ii) *Any compact m-dimensional complex-analytic subset of X admits a fundamental system of m-admissible neighborhoods.*

Proof. (i) Is obvious by the definitions and (ii) is a restatement of 3.5.1.

3.5.4. Proposition. *Let $\mathbf{B}_m(X)^{(o)}$ be the open set of $\mathbf{B}_m(X)$ consisting of cycles whose support admits in X a smoothly embeddable neighborhood. Let $\xi \in H^m(X, \Omega_X^m)$. Then F_ξ is holomorphic on $\mathbf{B}_m(X)^{(o)}$.*

Sketch of Proof. For $c \in \mathbf{B}_m(X)^{(o)}$, $|c|$ admits a smoothly embeddable neighborhood V therefore by 3.5.1 a neighborhood U with an embedding $\sigma: U \to U_1$ in a smooth m-complete U_1.

If $\mathcal{N}$ is the coherent sheaf on U_1 defined by the exact sequence

$$0 \to \mathcal{N} \to \Omega_{U_1}^m \to \sigma_* \Omega_U^m \to 0,$$

then $H^{m+1}(U_1, \mathcal{N})=0$ and hence $\xi|_U$ is induced by some $\xi_1 \in H^m(U_1, \Omega_{U_1}^m)$. By 3.4.1(v), F_{ξ_1} is holomorphic on $\mathbf{B}_m(U_1)$ so F_ξ is holomorphic near c.

3.5.5. Corollary. *If $\pi: X \to X'$ is geometrically flat with m-dimensional fibers and U' is the set of $x' \in X'$ such that $\pi^{-1}(x')$ admits in X smoothly embeddable neighborhoods then for any $\xi \in H^m(X, \Omega_X^m)$, $\pi_* \xi|_{U'}$ is holomorphic.*

3.6. Note Added in Proof. After having submitted the manuscript, the author together with D. Barlet solved problem D of the Introduction. Proposition 3.5.4 and Corollary 3.5.5 above are now true with $\mathbf{B}_m(X)$ instead of $\mathbf{B}_m(X)^{(0)}$. The notion of a weakly Kähler space loses its importance and Theorems 3 and 4 below (Ch. IV) become

Theorem 3′. *If $\pi: X \to X'$ is geometrically flat with X Kähler and X' reduced, then X' is Kähler.*

Theorem 4′. *If X is Kähler then $\mathbf{B}_m(X)$ is Kähler.*

II. Theorem 1 and its First Consequences

1. Kähler Spaces and Kähler Metrics

Let X be a complex space.

1.1. The Sheaf $\mathscr{K}_X^1$. Define

$$\mathscr{K}_X^1 := \mathscr{C}_X^\infty / PH_X, \qquad \mathscr{K}_{X,\mathbb{R}}^1 := \mathscr{C}_{X,\mathbb{R}}^\infty / PH_{X,\mathbb{R}},$$
$$\mathscr{K}^1(X) := H^0(X, \mathscr{K}_X^1), \qquad \mathscr{K}^1(X,\mathbb{R}) := H^0(X, \mathscr{K}_{X,\mathbb{R}}^1).$$

A section $\kappa \in \mathscr{K}^1(X)$ corresponds by definition to an open covering (U_α) of X together with elements $\varphi_\alpha \in \mathscr{C}^\infty(U_\alpha)$ such that $\varphi_\alpha - \varphi_\beta \in PH(U_\alpha \cap U_\beta)$. We write $\kappa = \{(U_\alpha, \varphi_\alpha)\}$. We have

$$\{(U_\alpha, \varphi_\alpha)\} = \{(V_j, \psi_j)\} \quad \text{iff} \quad (\varphi_\alpha - \psi_j)|_{U_\alpha \cap V_j} \in PH(U_\alpha \cap V_j).$$

For such κ, we set $\partial\bar{\partial}\kappa := \omega \in A^{1,1}(X)$ where

$$\omega|_{U_\alpha} = \partial\bar{\partial}\varphi_\alpha.$$

Of course, ω is well-defined and $d\omega = 0$. We say that κ is *represented* by the φ_α.

1.2. Kähler Metrics, Kähler Classes. A *Kähler metric* on X is by definition an element $\kappa \in \mathscr{K}^1(X,\mathbb{R})$ represented by a system of sections of SP_X^∞. The *Kähler form* of (X,κ) is $\omega := i\partial\bar{\partial}\kappa$ $(i = \sqrt{-1})$. We will often write (X,ω) instead of (X,κ), although ω does not determine κ unless X is smooth.

Similarly, if $\pi: X \to Y$ is a morphism of complex spaces, a *relative Kähler metric for* π is an element κ_π of $\mathscr{K}^1(X,\mathbb{R})$ represented by sections of SP_π^∞.

To any element $\kappa \in \mathscr{K}^1(X)$ we associate three cohomology classes as follows:

From the exact sequence $0 \to PH_X \to \mathscr{C}_X^\infty \to \mathscr{K}_X^1 \to 0$, we deduce a canonical morphism

$$\hat{c}_1 : \mathscr{K}^1(X) \to H^1(X, PH_X) \tag{1.2.1}$$

which obviously sends $\mathscr{K}^1(X,\mathbb{R})$ into $H^1(X, PH_{X,\mathbb{R}})$. From the diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathbb{R} \arrow[r] \arrow[d] & \mathcal{O}_X \arrow[r, "-2\mathrm{Im}"] \arrow[d] & PH_{X,\mathbb{R}} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \mathbb{C} \arrow[r] & \mathcal{O}_X \oplus \bar{\mathcal{O}}_X \arrow[r, "(i\ \ -i)"] \arrow[dr, "(d\ \ o)"'] & PH_X \arrow[r] \arrow[d, "-i\partial"] & 0 \\
 & & & d\mathcal{O}_X &
\end{tikzcd} \tag{1.2.2}$$

we deduce canonical morphisms $H^1(X, PH_X) \to H^2(X, \mathbb{C})$ and $H^1(X, PH_X) \to H^1(X, d\mathcal{O}_X)$ and, composing with $\hat{c}_1$, we obtain

$$\begin{aligned} c_1 &: \mathscr{K}^1(X) \to H^2(X, \mathbb{C}), \\ \tilde{c}_1 &: \mathscr{K}^1(X) \to H^1(X, d\mathcal{O}_X). \end{aligned} \tag{1.2.3}$$

Of course c_1 sends $\mathscr{K}^1(X,\mathbb{R})$ into $H^2(X,\mathbb{R})$. $d\mathcal{O}_X$ is the subsheaf of Ω_X^1 consisting of locally exact holomorphic 1-forms. Sometimes we will replace $\tilde{c}_1(\kappa)$ by its image in $H^1(X, \Omega_X^1)$.

So we have a diagram

$$
(1.2.4)\qquad
\begin{tikzcd}
H^{1,1}_{\bar\partial}(X) & & H^1(X,\Omega^1_X) \arrow[ll] & & H^1(X,d\mathcal{O}_X) \arrow[ll] \\
Z^{1,1}_d(X) \arrow[u] \arrow[dr] & & \mathcal{K}^1(X) \arrow[ll, "i\partial\bar\partial"'] \arrow[urr, "\tilde{c}_1"] \arrow[rr, "\hat{c}_1"] \arrow[dr, "c_1"] & & H_1(X,PH_X) \arrow[u] \arrow[dl] \\
& H^2_d(X) & & H^2(X,\mathbb{C}) \arrow[ll] &
\end{tikzcd}
$$

which is commutative (see 4.2 of Chap. III). This means that if κ is a Kähler metric on X and $\omega=i\partial\bar\partial\kappa$ the corresponding Kähler form, then ω is a d-closed represenattive of $c_1(\kappa)$ in $H^2(X,\mathbb{R})$ and also a $\bar\partial$-closed representative of $\tilde{c}_1(\kappa)$ in $H^1(X,\Omega^1_X)$.

In [15] Grauert proved that if κ is a Kähler metric on a normal compact space X such that $c_1(\kappa)$ lies in the canonical image of $H^2(X,\mathbb{Q})$ in $H^2(X,\mathbb{R})$, then X is a projective variety.

1.3. Kähler Spaces, Kähler Morphisms. X is said to be a *Kähler space* if there exists a Kähler metric on X.

A morphism $\pi: X\to Y$ is a *Kähler morphism* if there exists a relative Kähler metric κ_π for π.

We have the following elementary properties:

1.3.1. Proposition. (i) *Subspaces of Kähler spaces are Kähler.*

(ii) *Smooth Kähler spaces are Kähler manifolds in the usual sense.*

(iii) *$X\to\{y\}$ is a Kähler morphism iff X is a Kähler space.*

(iv) *Kähler morphisms are preserved by composition and base-change* [8].

(v) *Projective morphisms (for example: finite morphisms and blow-ups) are Kähler* [8, 12].

(vi) *If $\pi: X\to Y$ is a Kähler morphism, and Y a Kähler space then any open $U\subset\subset X$ is Kähler. More precisely: If κ_Y is a Kähler metric on Y and κ_π a relative Kähler metric for π, then for any $U\subset\subset X$ there is a constant $c_0>0$ such that for any $c>c_0$, $(\kappa_\pi+c\pi^*\kappa_Y)_{|U}$ is a Kähler metric on U* [8, 12].

On the other hand,

1.3.2. Proposition. (i) *It is not always true that a reduced compact space is Kähler if its normalization is Kähler.*

(ii) *It is not always true that a compact space X is Kähler if X_{red} is Kähler. A counterexample [8, II] is given by an infinitesimal neighborhood of a K3 surface in its space of moduli.*

(iii) *It is not always true that a normal compact space is Kähler if the complement of a point is Kähler* [15, 20].

(iv) *It is not always true that small deformations of compact Kähler spaces are Kähler* [20].

(v) *It is not always true that a normal compact space that is both Moĭezon and Kähler is projective* [20].

2. *Theorem 1*

2.1. Statement. Let X be a complex space. Suppose it admits an open covering $(U_\alpha)_{\alpha\in A}$ and a system of *continuous* strongly p.s.h. functions $\varphi_\alpha \in SP^0(U_\alpha)$ together with pluriharmonic functions $h_{\alpha\beta}\in PH(U_\alpha \cap U_\beta, \mathbb{R})$ such that

$$\text{(2.1.1)} \qquad \begin{aligned} &\text{(i)}\ \ \varphi_\alpha - \varphi_\beta = [h_{\alpha\beta}] \quad \text{in} \quad \mathscr{C}(U_\alpha \cap U_\beta, \mathbb{R}),\\ &\text{(ii)}\ \ h_{\alpha\beta} - h_{\alpha\gamma} + h_{\beta\gamma} = 0 \quad \text{on} \quad U_\alpha \cap U_\beta \cap U_\gamma . \end{aligned}$$

Then there are elements $\psi_\alpha \in SP^\infty(U_\alpha)$ such that

$$\text{(2.1.2)} \qquad \psi_\alpha - \psi_\beta = h_{\alpha\beta} \quad \text{in} \mathscr{C}^\infty(U_\alpha \cap U_\beta, \mathbb{R}).$$

In particular, X is a Kähler space.

2.2. Remark. By Lemma 1.2(iv) of Chap. I, the cocycle condition (ii) is redundant for X reduced. For smooth X, Theorem 1 is proven in [23] and the proof we give there is valid for X reduced and locally irreducible. We will use the conventions stated in 2.4 of Chap. I.

2.3. Proof. Since X is paracompact, it admits two locally finite open coverings (V_k), (W_k) $(k\in\mathbb{N})$ such that $V_0=\emptyset$ and $V_k \subset\subset W_k \subset U_{\alpha_k}$ for each k. Set $T^k_{\alpha\beta} := U_\alpha \cap U_\beta \cap (V_1 \cup \ldots \cup V_k)$.

We will define inductively elements

$$\varphi^k_\alpha \in SP^{0,\infty}(U_\alpha, V_1 \cup \ldots \cup V_k)$$

such that

(i) For some compact S_k, $V_k \subset S_k \subset W_k$,

$$\varphi^k_\alpha|_{U_\alpha\setminus S_k} = \varphi^{k-1}_\alpha|_{U_\alpha \setminus S_k}$$

in $SP^{0,\infty}(U_\alpha\setminus S_k, V_1\cup\ldots\cup V_k) = SP^{0,\infty}(U_\alpha \setminus S_k, V_1 \cup \ldots \cup V_{k-1})$

$$\text{(2.3.1)} \qquad \begin{aligned} &\text{(ii)}\ \ [\varphi^k_\alpha] - [\varphi^k_\beta] = [h_{\alpha\beta}] \quad \text{in} \quad \mathscr{C}(U_\alpha \cap U_\beta, \mathbb{R}),\\ &\text{(iii)}\ \ (\varphi^k_\alpha - \varphi^k_\beta)|_{T^k_{\alpha\beta}} = h_{\alpha\beta}|_{T^k_{\alpha\beta}} \quad \text{in} \quad \mathscr{C}^\infty(T^k_{\alpha\beta}, \mathbb{R}). \end{aligned}$$

We start by taking $\varphi^0_\alpha := \varphi_\alpha$ the initial data.

Suppose φ^{k-1}_α is defined for all α.

Apply Richberg's lemma to $X = W_k$,

$$U = V_k, \qquad V = V_1 \cup \ldots \cup V_{k-1}, \qquad \varphi = \varphi^{k-1}_{\alpha_k}|_{W_k}.$$

We obtain an element

$\psi \in SP^{0,\infty}(W_k, V_1 \cup \ldots \cup V_k)$

and a compact S_k, $V_k \subset S_k \subset W_k$ such that

$$\psi|_{W_k\setminus S_k} = \varphi^{k-1}_{\alpha_k}|_{W_k \setminus S_k}.$$

Now we set

$$\text{(2.3.2)} \qquad \varphi^k_\alpha := \begin{cases} \varphi^{k-1}_\alpha & \text{on} \quad U_\alpha \setminus S_k \\ \psi + h_{\alpha\alpha_k} & \text{on} \quad U_\alpha \cap W_k , \end{cases}$$

where the last expression is defined in 2.4(v) of Chap. I.

By the induction hypothesis, (2.3.1) is valid for the rank $k-1$, hence definition (2.3.2) is consistent. But this implies (2.3.1) for the rank k as well. Indeed, (i) is obvious. (ii) and (iii) can be easily checked on W_k by the cocycle condition (2.1.1)(ii) and outside S_k by the induction hypothesis. So (2.3.1) is valid.

Now since $S_k \subset W_k$, (S_k) is locally finite and, for fixed α, $(\varphi_\alpha^k)_{k\in\mathbb{N}}$ is locally stationary. We may set

$$\psi_\alpha := \lim_{k\to\infty} \varphi_\alpha^k \in SP^\infty(U_\alpha)$$

and the conclusion of Theorem 1 is satisfied.

2.4. Corollary. *The "old" and "modern" definition of a reduced Kähler space coincide.*

Proof. By 1.2(iv) of Chap. I, if X is reduced, PH_X can be identified to a subsheaf of $\mathscr{C}_X$. A Kähler metric in the "old" sense is a section of $\mathscr{C}_{X,\mathbb{R}}/PH_{X,\mathbb{R}}$ represented locally by sections of $[SP_X^\infty]$. Since $[SP_X^\infty] \subset SP_X^0$, Theorem 1 applies.

3. Application to Finite Morphisms

Theorem 1 implies that images of Kähler spaces under certain finite morphisms are Kähler. This solves a problem raised by Lieberman at the end of [18].

3.1. Traces of Continuous and Holomorphic Functions

3.1.1. Definitions. If X is reduced, $k \geqq 1$ an integer and $\varphi \in \mathscr{C}(X)$, then

$$\tilde{\varphi}: \sum_{j=1}^{k} \{x_j\} \mapsto \sum_{j=1}^{k} \varphi(x_j)$$

defines a continuous function on $\mathrm{Sym}^k(X)$. On the other hand, for arbitrary X we have

$$\mathbf{B}_0(X) = \coprod_{k\geqq 1} \mathrm{Sym}^k(X_{\mathrm{red}}).$$

Now suppose $\pi: X \to X'$ is a finite open surjective morphism with connected base X'. We examine the following two situations:

(1) X' is reduced and π is geometrically flat;

(2) X' is arbitrary and π is flat.

In the first case, there is an integer $k = k_\pi \geqq 1$ called the (geometric) *degree of* π such that the classifying morphism $H: X' \to \mathbf{B}_0(X)$ factors through $\mathrm{Sym}^k(X_{\mathrm{red}})$. We have for generic $x' \in X'$ (on the points of flatness of π)

$$H(x') = \sum_{x\in\pi^{-1}(x')} \{x\},$$

where the sum takes account of multiplicities.

Define a *continuous trace* morphism

$$\mathrm{Tr}_{X/X'}^{(c)}: \pi_* \mathscr{C}_X \to \mathscr{C}_{X'}$$

by $\varphi \mapsto \tilde{\varphi} \circ H$.

In the second case, there is an integer $r=r_\pi \geqq 1$ called the (algebraic) *degree of* π such that $\pi_* \mathcal{O}_X$ is a locally free $\mathcal{O}_{X'}$-module of rank r. Define the *holomorphic trace* morphism

$$\mathrm{Tr}^{(h)}_{X/X'}: \pi_* \mathcal{O}_X \to \mathcal{O}_{X'}$$

by $f \mapsto$ trace of the linear map $\{g \mapsto fg\}$.

r_π is preserved by base change and, if X' is reduced, coincides with k_π. For general X', define $\bar{\pi}$, Y, $\bar{Y}$ by the cartesian diagram

$$\begin{tikzcd} X \arrow[d, "\pi"'] & Y := X \times_{X'} Y' \arrow[l] \arrow[d, "\bar{\pi}"] \\ X' & Y' := X'_{\mathrm{red}} . \arrow[l] \end{tikzcd}$$

Then we have $r_\pi = r_{\bar{\pi}} = k_{\bar{\pi}}$. We define

$$\mathrm{Tr}^{(c)}_{X/X'} := \mathrm{Tr}^{(c)}_{Y/Y'}$$

since $\mathscr{C}_X = \mathscr{C}_Y$ and $\mathscr{C}_{X'} = \mathscr{C}_{Y'}$. The two trace morphisms so defined are compatible, i.e. the diagram

$$\begin{tikzcd} \pi_* \mathcal{O}_X \arrow[r, "\varrho"] \arrow[d, "\mathrm{Tr}^{(h)}_{X/X'}"'] & \pi_* \mathscr{C}_X \arrow[d, "\mathrm{Tr}^{(c)}_{X/X'}"] \\ \mathcal{O}_{X'} \arrow[r, "\varrho"] & \mathscr{C}_{X'} \end{tikzcd}$$

is commutative, where $\varrho: f \mapsto [f]$ is the canonical morphism. The holomorphic trace morphism is obviously extended to $\pi_* PH_X \to PH_{X'}$.

We write $\pi_* \varphi$ for $\mathrm{Tr}^{(c)}_{X/X'} \varphi$ or $\mathrm{Tr}^{(h)}_{X/X'} \varphi$ indifferently.

3.1.2. Lemma [5, 23]. *If φ is p.s.h., strongly p.s.h., holomorphic or pluriharmonic on X, then $\tilde{\varphi}$ (resp. $\pi_* \varphi$) has the corresponding properties on* $\mathrm{Sym}^k(X_{\mathrm{red}})$ *(resp. X').*

3.1.3. Remark. (i) For the "p.s.h." part of the above lemma, the Fornaess-Narasimhan theorem is needed.

(ii) It is not true in general that $\pi_* \varphi$ is $\mathscr{C}^\infty$ if φ is $\mathscr{C}^\infty$ even if X and X' are smooth.

3.2. Theorem. *Let X be a Kähler space and $\pi: X \to X'$ a finite open surjective morphism such that either*

(i) *X' is reduced and π is geometrically flat or*

(ii) *π is flat.*

Then X' is Kähler.

Proof. It results from 3.1.2 and Theorem 1 (exactly as Proposition 2.1 of [23]).

3.2.1. Corollary. *If X is a reduced Kähler space and G a finite group of automorphisms of X, then X/G is Kähler. In particular* $\mathrm{Sym}^k(X)$ *is Kähler.*

Proof. It is clear that the canonical projection $X \to X/G$ is geometrically flat. For X smooth and G having isolated fixed points, this is shown by Fujiki [13, Proposition 1].

3.2.2. Corollary. *If $\pi: X \to X'$ is finite surjective with X Kähler and X' normal then X' is Kähler.*

4. Weakly Kähler Metrics

Because of the impossibility to solve (for the moment) problem 3.6 (Chap. I) we are forced to introduce the notion of weakly Kähler spaces.

4.1. Definitions. If X, Y are reduced spaces, a function $f: X \to Y$ is *weakly holomorphic* if it is *continuous* and generically holomorphic. Let $\mathscr{W}_X$ be the sheaf of weakly holomorphic complex-valued functions on X. Define the sheaf WPH_X of *weakly pluriharmonic* functions by $WPH_X := \mathscr{W}_X + \overline{\mathscr{W}_X}$. X is *weakly normal* iff $\mathscr{W}_X = \mathcal{O}_X$. The *weak normalization* of X is a weakly normal space $\hat{X}$ [2] together with a holomorphic homeomorphism $n: \hat{X} \to X$ such that $n_* \mathcal{O}_{\hat{X}} = \mathscr{W}_X$. If X is not reduced, define the weak normalization $\hat{X} \to X$ as that of X_{red} followed by the reduction $X_{\text{red}} \to X$.

A *weakly Kähler metric* on X is a section of the quotient sheaf $\mathscr{C}_{X,\mathbb{R}}/WPH_{X,\mathbb{R}}$ represented by a system of sections of SP_X^0. X is *weakly Kähler* if X_{red} admits a weakly Kähler metric. We have (for X, Y, Z reduced spaces):

4.1.1. Lemma. (i) *If $f: X \to Y$ and $g: Y \to Z$ are weakly holomorphic, then $g \circ f: X \to Z$ is weakly holomorphic.*

(ii) *If $f: X \to Y$ is weakly holomorphic and $h \in WPH(Y)$, then $h \circ f \in WPH(X)$.*

(iii) *X is weakly normal, iff every local irreducible component of X is normal.*

The Fornaess-Narasimhan theorem implies:

4.1.2. Lemma. (i) *If $f: X \to Y$ is weakly holomorphic and $\varphi \in P^0(Y)$, then $\varphi \circ f \in P^0(X)$.*

(ii) *$WPH_X \subset P_X^0$ (weakly pluriharmonic functions are p.s.h.).*

(iii) *$WPH_X SP_X^0 \subset SP_X^0$ [a consequence of (ii)].*

4.1.3. Lemma. *Let $n: X \to \hat{X}$ be the weak normalization of X. For $\varphi \in \mathscr{C}(\hat{X})$, set $n_* \varphi := \varphi \circ n^{-1} \in \mathscr{C}(X)$. Then*

(i) *If $\varphi \in P^0(\hat{X})$, then $n_* \varphi \in P^0(X)$.*

(ii) *If $\varphi \in SP^0(\hat{X})$, then $n_* \varphi \in SP^0(X)$.*

(iii) *If $\varphi \in \mathcal{O}(\hat{X})$, then $n_* \varphi \in \mathscr{W}(X)$.*

(iv) *If $\varphi \in PH(\hat{X})$, then $n_* \varphi \in WPH(X)$.*

4.2. Relation with Kähler Metrics

4.2.1. Lemma. *If X is weakly Kähler and weakly normal, then X is Kähler.*

Proof. Since $WPH_X = PH_X$, Theorem 1 applies.

4.2.2. Lemma. *If $\pi: X \to Y$ is a Kähler morphism and Y a weakly Kähler space, then any open $U \subset\subset X$ is weakly Kähler.*

Proof. By an elementary argument similar to 1.3.1 (vi).

4.2.3. Proposition. *Let X be a complex space and $n: \hat{X} \to X$ its weakly normalization. Then*

(i) *If $\hat{X}$ is Kähler, then X is weakly Kähler.*

(ii) *If X is weakly Kähler, then every open $U \subset\subset \hat{X}$ is Kähler.*

Proof. (i) Is a consequence of Lemma 4.1.3 above.

(ii) Since n is finite, it is Kähler morphism by 1.3.1 (v). We apply 4.2.2 and 4.2.1 to conclude.

4.2.4. Corollary. *If X is compact, then $\hat{X}$ is Kähler iff X is weakly Kähler.*

III. Theorem 2

1. Čech Spaces and Čech Open Sets

1.1. Definitions. A (topological or complex-analytic) *Čech space* will be by definition a pair

$$\underline{X}=(X,\mathscr{X}),$$

where X is a (topological or complex) space and $\mathscr{X}$ an open covering of X. We call X the *space underlying to* $\underline{X}$ and always denote both by the same letter. We will deal only with complex-analytic Čech spaces. If $\mathscr{X}=(X_\lambda)_{\lambda\in\Lambda}$, the X_λ will be called the *elementary* open sets of X.

Suppose $\underline{X}=(X,(X_\lambda)_{\lambda\in\Lambda})$ and $\underline{Y}=(Y,(Y_\mu)_{\mu\in M})$ are two Čech spaces. A morphism

$$F:\underline{X}\to\underline{Y}$$

will be a pair $F=(f,\mu)$ where $f:X\to Y$ is a morphism in the ordinary sense and $\mu:\Lambda\to M$ a map such that

$$X_\lambda\subset f^{-1}(Y_{\mu(\lambda)}) \tag{1.1.1}$$

for all $\lambda\in\Lambda$. We call f the *morphism underlying to* F. We will say that F is an *open inclusion* if f is one.

A *Čech open set* $\underline{U}\leqslant\underline{X}$ will be a Čech space chose underlying space is an open subset of X together with an open inclusion

$$j:\underline{U}\to\underline{X}.$$

Or course, j is not uniquely determined by $\underline{U}$.

If $\underline{U}_1=(U_1,(U_{1,\alpha})_{\alpha\in A_1})$ and $\underline{U}_2=(U_2,(U_{2,\beta})_{\beta\in A_2})$ are two Čech open sets of $\underline{X}$, define

$$\underline{U}_1\cap\underline{U}_2:=(U_1\cap U_2,(U_{1,\alpha}\cap U_{2,\beta})_{(\alpha,\beta)\in A_1\times A_2}). \tag{1.1.2}$$

Notice that there are two open inclusions

$$j_1,j_2:\underline{U}_1\cap\underline{U}_2\to\underline{X}$$

each factoring through $\underline{U}_1$ and $\underline{U}_2$, respectively.

If $\underline{X}=(X,\mathscr{X})$ is a Čech space and $\mathscr{F}$ a sheaf of abelian groups on X, write

$$C^q(\underline{X},\mathscr{F}),\ Z^q(\underline{X},\mathscr{F}),\ H^q(\underline{X},\mathscr{F})$$

for the groups of Čech contains, cocycles and cohomology classes of degree q of the covering $\mathscr{X}$ with coefficients in $\mathscr{F}$. Denote by

$$\varepsilon:H^0(X,\mathscr{F})\to C^0(\underline{X},\mathscr{F}) \tag{1.1.3}$$

the canonical inclusion and by

$$\delta: C^{q-1}(\underline{X}, \mathscr{F}) \to C^q(\underline{X}, \mathscr{F})$$

the Čech differential given by the usual formula

$$(1.1.4) \qquad (\delta\varphi)_{\lambda_0 \dots \lambda_q} := \sum_{r=0}^{q} (-1)^r \varphi_{\lambda_0 \dots \hat{\lambda}_r \dots \lambda_q}|_{X_{\lambda_0} \cap \dots \cap X_{\lambda_q}}.$$

If $\underline{U} \ll \underline{X}$ is a Čech open set with an open inclusion $j: \underline{U} \to \underline{X}$, denote by

$$j^*: C^q(\underline{X}, \mathscr{F}) \to C^q(\underline{U}, \mathscr{F})$$

the obvious morphism. We will write

$$(1.1.5) \qquad \varphi|_{\underline{U}} := j^*(\varphi)$$

if there is no ambiguity about j.

Now suppose there are two open inclusions

$$j_1, j_2: \underline{U} \to \underline{X}$$

with $\underline{U} = (U, (U_\alpha)_{\alpha \in A})$, $\underline{X} = (X, (X_\lambda)_{\lambda \in \Lambda})$.

There is a homotopy operator

$$T: C^{q+1}(\underline{X}, \mathscr{F}) \to C^q(\underline{U}, \mathscr{F})$$

defined by

$$(1.1.6) \qquad (T\varphi)_{\alpha_0 \dots \alpha_q} := \sum_{r=0}^{q} (-1)^r \varphi_{\lambda_0 \dots \lambda_r \mu_r \dots \mu_q}|_{U_{\alpha_0} \cap \dots \cap U_{\alpha_q}},$$

where

$$U_{\alpha_r} \subset X_{\lambda_r} \quad \text{by } j_1$$

and

$$U_{\alpha_r} \subset X_{\mu_r} \quad \text{by } j_2.$$

T is extended by 0 on $C^0(\underline{X}, \mathscr{F})$ and $H^0(X, \mathscr{F})$. The following is obvious

$$(1.1.7) \qquad \delta T + T\delta = j_2^* - j_1^*.$$

1.2. Cup-Products of Čech Cochains. Now suppose that $\mathscr{F}, \mathscr{G}, \mathscr{H}$ are sheaves of differential forms ($\mathscr{C}^\infty$, holomorphic or antiholomorphic) such that

$$\mathscr{F} \wedge \mathscr{G} \subset \mathscr{H}.$$

We define the cup-product

$$C^q(\underline{X}, \mathscr{F}) \times C^r(\underline{X}, \mathscr{G}) \to C^{q+r}(\underline{X}, \mathscr{H})$$

by the identity

$$(1.2.1) \qquad (\varphi \cdot \psi)_{\alpha_0 \dots \alpha_{q+r}} := (\varphi_{\alpha_0 \dots \alpha_q} \wedge \psi_{\alpha_q \dots \alpha_{q+r}})|_{X_{\lambda_0} \cap \dots \cap X_{\lambda_{q+r}}}.$$

As an immediate consequence we have

$$(1.2.2) \qquad \delta(\varphi \cdot \psi) = (\delta\varphi) \cdot \psi + (-1)^q \varphi \cdot \delta\psi,$$

$$(1.2.3) \qquad T(\varphi \cdot \psi) = (T\varphi) \cdot j_2^* \psi + (-1)^q (j_1^* \varphi) \cdot T\psi.$$

1.3. m-Complete and m-Admissible Čech Open Sets. We extend the notion of m-admissible open sets (3.5.3 of Chap. I) to Čech open sets.

1.3.1. Definitions. A Čech space $\underline{X}$ is said to be *m-complete* if for any coherent analytic sheaf $\mathscr{F}$ on X and any $q>m$, we have $H^q(\underline{X},\mathscr{F})=0$.

A sufficient condition for this is that the underlying space X be m-complete and the elementary open sets of $\underline{X}$ be Stein.

If $\underline{U}\ll\underline{X}$ is a Čech open set of $\underline{X}$, we will say that $\underline{U}$ is *m-admissible in X* if

(i) $\underline{U}$ is m-complete.

(ii) There is a Čech open set $\underline{V}$ such that $\underline{U}\ll\underline{V}\ll\underline{X}$ and $H^k(\underline{V},\mathbb{R})=0$ for all $k>2m$.

Of course, if the above are satisfied, then the canonical morphism $H^k(\underline{X},\mathbb{R})\to H^k(\underline{U},\mathbb{R})$ vanishes for $k>2m$, since it factors through $H^k(\underline{V},\mathbb{R})=0$.

This may be expressed as follows:

1.3.2. Lemma. *If $\underline{U}\ll\underline{X}$ is m-admissible, $k>2m$ and $a\in Z^k(\underline{X},\mathbb{R})$, then there is an element $b\in C^{k-1}(\underline{U},\mathbb{R})$ such that $a|_{\underline{U}}=\delta b$.*

1.3.3. Proposition. *Let $\underline{X}$ be a Čech space and $U\subset X$ an open set (in the ordinary sense) that is m-admissible. Then U is underlying to some m-admissible Čech open set $\underline{U}\ll\underline{X}$.*

Proof. By definition U is m-complete and there is an open V such that $U\subset V\subset X$ and $H^k(V,\mathbb{R})=0$ for all $k>2m$. If we take a sufficiently fine Leray open covering of V with respect to the constant sheaf such that $\underline{V}\ll\underline{X}$ and then a sufficiently fine Stein open covering of U such that $\underline{U}\ll\underline{V}$, it is clear that $\underline{U}\ll\underline{X}$ is m-admissible.

2. *Čech Transform of a Complex of Sheaves*

2.1. Definitions. Let $\underline{X}$ be a Čech space and

$$0\to\mathscr{F}\xrightarrow{j}\mathscr{L}^0\xrightarrow{D}\mathscr{L}^1\xrightarrow{D}\dots\xrightarrow{D}\mathscr{L}^m\xrightarrow{D}\dots \tag{2.1.1}$$

a complex of sheaves of abelian groups on the underlying space X. We do not suppose it to be an exact sequence of sheaves.

The *Čech transform* of the complex (2.1.1) over $\underline{X}$ will be the single complex associated to the double complex

$$\begin{array}{ccccccc}
H^0(X,\mathscr{F}) & \to & H^0(X,\mathscr{L}^0) & \to & H^0(X,\mathscr{L}^1) & \to & \dots\\
\downarrow & & \downarrow & & \downarrow & & \\
C^0(\underline{X},\mathscr{F}) & \to & C^0(\underline{X},\mathscr{L}^0) & \to & C^0(\underline{X},\mathscr{L}^1) & \to & \dots\\
\downarrow & & \downarrow & & \downarrow & & \\
C^1(\underline{X},\mathscr{F}) & \to & C^1(\underline{X},\mathscr{L}^0) & \to & C^1(\underline{X},\mathscr{L}^1) & \to & \dots\\
\downarrow & & \downarrow & & \downarrow & & \\
\vdots & & \vdots & & \vdots & &
\end{array}$$

More precisely, we define for $q\geqq 0$

$$\check{C}^q(\underline{X};\mathscr{F},\mathscr{L}^\cdot):=C^q(\underline{X},\mathscr{F})\oplus\left\{\bigoplus_{k=1}^{q}C^{q-k}(\underline{X},\mathscr{L}^{k-1})\right\}\oplus H^0(X,\mathscr{L}^q). \tag{2.1.2}$$

An element of $\check{C}^q(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot)$ has the form

$$\Phi = (f; \varphi^0, \ldots, \varphi^{q-1}; \eta^q),$$

where

$$f \in C^q(\underline{X}, \mathscr{F}),$$

$$\varphi^{k-1} \in C^{q-k}(\underline{X}, \mathscr{L}^{k-1}) \quad \text{for} \quad k = 1, \ldots, q,$$

$$\eta^q \in H^0(X, \mathscr{L}^q).$$

We will call f the *head* of Φ, φ^{k-1} the *k-th component* of Φ and η^q the *tail* of Φ. Define the differential

$$\Delta: \check{C}^q(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot) \to \check{C}^{q+1}(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot)$$

by

(2.1.3) $$\Delta := \delta + (-1)^{q+1} D$$

where

$$\delta\Phi := (\delta f; \delta\varphi^0, \ldots, \delta\varphi^{q-1}, \varepsilon\eta^q; 0),$$

$$D\Phi := (0; jf, D\varphi^0, \ldots, D\varphi^{q-1}; D\eta^q).$$

Sometimes we will change the sign convention

$$\Delta = \delta + (-1)^{q+1} D \quad \text{to} \quad \Delta = \delta + (-1)^q D.$$

We then define

(2.1.4) $$\check{Z}^q(\underline{X}, \mathscr{F}, \mathscr{L}^\cdot) := \operatorname{Ker}\{\check{C}^q(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot) \xrightarrow{\Delta} \check{C}^{q+1}(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot)\}$$

and the Čech hypercohomology groups

(2.1.5) $$\check{H}^q(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot) := \check{Z}^q(X; \mathscr{F}, \mathscr{L}^\cdot)/\Delta\check{C}^{q-1}(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot).$$

We will use the following.

2.2. Lemma. *Let $r^q: H^q(\underline{X}; \mathscr{F}, \mathscr{L}^\cdot) \to H^q(\underline{X}, \mathscr{F})$ be the canonical morphism.*

(i) *If $H^{q-k}(\underline{X}, \mathscr{L}^{k-1}) = 0$ for $k = 1, \ldots, q-1$ then r^q is injective.*

(ii) *If $H^{q-k}(\underline{X}, \mathscr{L}^k) = 0$ for $k = 0, \ldots, q-1$ then r^q is surjective.*

Proof. It is an immediate consequence of the following elementary property of double complexes: If $M^\cdot$ is the single complex associated to a double complex $K^{\cdot,\cdot} = (K^{i,j})_{i,j \geq 0}$, then the canonical morphism $H^q(M^\cdot) \to H^q(K^{\cdot,0})$ is injective if $H^{q-j}(K^{\cdot,j}) = 0$ for $j = 1, \ldots, q-1$ and surjective if $H^{q-j}(K^{\cdot,j+1}) = 0$ for $j = 0, \ldots, q-1$. This is to be applied for

$$K^{i,j} = \begin{cases} 0 & \text{if} \quad i = j = 0 \\ H^0(X, \mathscr{L}^{j-1}) & \text{if} \quad j > i = 0 \\ C^{i-1}(\underline{X}, \mathscr{F}) & \text{if} \quad i > j = 0 \\ C^{i-1}(\underline{X}, \mathscr{L}^{j-1}) & \text{if} \quad i, j > 0. \end{cases}$$

Part (i) of the above lemma is equivalent to

2.3. Corollary. *If a cocycle $\Phi \in \check{Z}^q(X; \mathcal{F}, \mathcal{L}^\cdot)$ has a head that is δ-exact and if $H^{q-k}(\underline{X}, \mathcal{L}^{k-1})=0$ for $k=1,\dots,q-1$, then Φ is Δ-exact and, in particular, the tail of Φ is D-exact.*

2.4. Remark. Definition 3.4.3 of Chap. I can be restated as follows: A $\bar\partial$-closed form $\tau \in A^{k,l}(X)$ is said to represent an element of $H^l(X, \Omega^k)$ if there is a cocycle of degree l

$$c \in \check{Z}^l(\underline{X}; \Omega^k, A^{k,\cdot})$$

of the Čech transform of the Dolbeault complex whose tail is τ, for some open covering of X.

3. *The $\partial\bar\partial$-Complex $\mathcal{L}^\cdot_m$*

Let X be a complex space. For any pair (p,q) of natural integers, there is a complex of sheaves on X of the form

$$\begin{array}{ccccccccccc} 0 \longrightarrow \mathbb{C} \xrightarrow{\lambda} \mathcal{L}^0_{p,q} \xrightarrow{D} \dots \xrightarrow{D} & \mathcal{L}^{p+q-1}_{p,q} & \xrightarrow{D} & \mathcal{L}^{p+q}_{p,q} & \xrightarrow{D} \dots \\ & \| & & \| & \\ & A_X^{p-1,q-1} & \xrightarrow{\partial\bar\partial} & A_X^{p,q} & \end{array}$$

defined in [7].

We will deal exclusively with the case $p=q$, so we write $\mathcal{L}^r_m$ for $\mathcal{L}^r_{m,m}$. The complex $\mathcal{L}^\cdot_m$ defined as follows (the suffix X will be omitted).

3.1. Definitions.

$$\mathcal{L}^r_m := \begin{cases} \Omega^r \oplus A^{r-1} \oplus \bar\Omega^r & \text{if } r<m \\ A^{m-1,r-m} \oplus \dots \oplus A^{r-m,m-1} & \text{if } m \leqq r < 2m \\ A^{r-m,m} \oplus \dots \oplus A^{m,r-m} & \text{if } r \geqq 2m. \end{cases} \tag{3.1}$$

Define $j: \mathbb{C} \xrightarrow{\binom{1}{1}} \Omega^0 \oplus \bar\Omega^0 = \mathcal{L}^0_m$ and

(i) For $0 \leqq r < m-1$,

$$\begin{array}{ccc} \mathcal{L}^r_m & \xrightarrow{\quad D \quad} & \mathcal{L}^{r+1}_m \\ \| & & \| \\ \Omega^r \oplus A^{r-1} \oplus \bar\Omega^r & \xrightarrow{\begin{pmatrix} d & 0 & 0 \\ (-1)^{r+1} & d & (-1)^r \\ 0 & 0 & d \end{pmatrix}} & \Omega^{r+1} \oplus A^r \oplus \bar\Omega^{r+1}. \end{array}$$

(ii) For $r=m-1$,

$$\begin{array}{ccc} \mathcal{L}^{m-1}_m & \xrightarrow{\quad D \quad} & \mathcal{L}^m_m \\ \| & & \| \\ \Omega^{m-1} \oplus A^{m-2} \oplus \bar\Omega^{m-1} & \xrightarrow{((-1)^m \;\; d \;\; (-1)^{m-1})} & A^{m-1}. \end{array}$$

(iii) For $m \leqq r < 2m-1$,

$$\begin{array}{ccc}
\mathscr{L}_m^r & \xrightarrow{\quad D \quad} & \mathscr{L}_m^{r+1} \\
\| & \begin{pmatrix} \partial & \bar\partial & & & 0 \\ 0 & \partial & \ddots & & \\ & & \ddots & \ddots & \\ & & & \bar\partial & 0 \\ 0 & & & \partial & \bar\partial \end{pmatrix} & \| \\
A^{m-1,r-m} \oplus \ldots \oplus A^{r-m,m-1} & \longrightarrow & A^{m-1,r-m+1} \oplus \ldots \oplus A^{r-m+1,m-1}.
\end{array}$$

(iv) For $r = 2m-1$,

$$\begin{array}{ccc}
\mathscr{L}_m^{2m-1} & \xrightarrow{D} & \mathscr{L}_m^{2m} \\
\| & & \| \\
A^{m-1,m-1} & \xrightarrow{\partial\bar\partial} & A^{m,m}.
\end{array}$$

(v) For $r \geqq 2m$,

$$\begin{array}{ccc}
\mathscr{L}_m^r & \xrightarrow{D} & \mathscr{L}_m^{r+1} \\
\| & & \| \\
A^{r-m,m} \oplus \ldots \oplus A^{m,r-m} & \xrightarrow{d} & A^{r-m+1,m} \oplus \ldots \oplus A^{m,r-m+1}.
\end{array}$$

Actually, the part $\mathbb{C} \to \mathscr{L}_m^0 \to \ldots \to \mathscr{L}_m^{2m-1}$ is the single complex associated to the truncated double complex

$$\begin{array}{ccccccc}
\mathbb{C} & \xrightarrow{\quad 1 \quad} & \Omega^0 & \xrightarrow{\quad d \quad} \ldots \xrightarrow{\quad d \quad} & \Omega^{m-1} \\
{\scriptstyle 1}\downarrow & & {\scriptstyle -1}\downarrow & & {\scriptstyle (-1)^m}\downarrow \\
\bar\Omega^0 & \xrightarrow{\quad 1 \quad} & A^{0,0} & \xrightarrow{\quad \partial \quad} \ldots \xrightarrow{\quad \partial \quad} & A^{m-1,0} \\
{\scriptstyle d}\downarrow & & {\scriptstyle \bar\partial}\downarrow & & {\scriptstyle \bar\partial}\downarrow \\
\vdots & & \vdots & & \vdots \\
{\scriptstyle d}\downarrow & & {\scriptstyle \bar\partial}\downarrow & & {\scriptstyle \bar\partial}\downarrow \\
\bar\Omega^{m-1} & \xrightarrow{(-1)^{m-1}} & A^{0,m-1} & \xrightarrow{\quad \partial \quad} \ldots \xrightarrow{\quad \partial \quad} & A^{m-1,m-1}
\end{array}$$

with the indicated sign conventions, and a similar observation may serve to define

$$\mathbb{C} \to \mathscr{L}_{p,q}^0 \to \ldots \mathscr{L}_{p,q}^{p+q-1}.$$

3.2. Proposition (Bigolin [7]). *For smooth X, $(\mathscr{L}_{p,q}^{\cdot}, D)$ is an exact sequence of sheaves.*

3.3. The Involution on $\mathscr{L}_m^{\cdot}$. A $\mathbb{C}$-antilinear involution $\varphi \mapsto \varphi^*$ is defined on $\mathscr{L}_m^{\cdot}$ as follows:

(i) For $(g^r, \psi^{r-1}, \bar h^r)$ in $\mathscr{L}_m^r = \Omega^r \oplus A^{r-1} \oplus \bar\Omega^r$ $(r < m)$

$$(g^r, \psi^{r-1}, \bar h^r)^* := (h^r, -\bar\psi^{r-1}, \bar g^r).$$

(ii) For ψ^{r-1} in $\mathscr{L}_m^r \subset A^{r-1}$, $(\psi^{r-1})^* := -\bar\psi^{r-1}$ $(m \leqq r < 2m)$.

(iii) For ψ^r in $\mathscr{L}_m^r \subset A^r$, $(\psi^r)^* := \bar\psi^r$ $(r \geqq 2m)$.

It is obvious that $(D\varphi)^* = D(\varphi^*)$.

We denote by $\mathscr{L}_{m,\mathbb{R}}^{\cdot}$ the sub-complex of $\mathscr{L}_m^{\cdot}$ of fixed points under $(\cdot)^*$. We set $\mathrm{Re}\,\varphi := \frac{1}{2}(\varphi + \varphi^*)$. Note that a self-conjugate element of $\mathscr{L}_m^r$, for $r < 2m$ has pure imaginary $\mathscr{C}^\infty$ components.

3.4. The Morphism $\mu:\mathscr{L}^{\cdot}_{m+1}\to\mathscr{L}^{\cdot}_m$. A morphism $\mu=\mu^r_m:\mathscr{L}^r_{m+1}\to\mathscr{L}^r_m$ is defined by

(i) For $r<m$, $\mathscr{L}^r_{m+1}=\Omega^r\oplus A^{r-1}\oplus\bar{\Omega}^r=\mathscr{L}^r_m$ and $\mu^r_m:=\mathrm{id}$. We define $\mu=\mathrm{id}$ on $\mathbb{C}$ as well.

(ii) For $m\leqq r<2m$, $\mathscr{L}^r_m$ is a direct summand of $\mathscr{L}^r_{m+1}$ and μ^r_m is defined as the canonical projection.

(iii) For $r=2m$,

$$\begin{tikzcd}
\mathscr{L}^{2m}_{m+1} \arrow[r, "\mu^{2m}_m"] \arrow[d, equal] & \mathscr{L}^{2m}_m \arrow[d, equal] \\
A^{m,m-1}\oplus A^{m-1,m} \arrow[r, "\frac{1}{2}(-\partial\ \ \bar{\partial})"] & A^{m,m}.
\end{tikzcd}$$

(iv) For $r=2m+1$,

$$\begin{tikzcd}
\mathscr{L}^{2m+1}_{m+1} \arrow[r, "\mu^{2m+1}_m"] \arrow[d, equal] & \mathscr{L}^{2m+1}_m \arrow[d, equal] \\
A^{m,m} \arrow[r, "\frac{1}{2}\begin{pmatrix}-\partial\\ \partial\end{pmatrix}"] & A^{m+1,m}\oplus A^{m,m+1}.
\end{tikzcd}$$

(v) For $r>2m+1$, $\mathscr{L}^r_{m+1}$ is a direct summand of $\mathscr{L}^r_m$ and μ^r_m is defined as the canonical inclusion.

3.4.1. Lemma. *The above morphism μ commutes with D and the involution $(\cdot)^*$.*

3.5. Relation with the $(\bar{\partial}\oplus\partial)$-Complex. The $(\bar{\partial}\oplus\partial)$-complex $(\mathscr{G}^{\cdot}_m,\hat{d})$ is the direct sum of the Dolbeault complex and its conjugate

$$\begin{tikzcd}
0 \arrow[r] & \mathscr{G}^{-1}_m \arrow[r, "j"] \arrow[d, equal] & \mathscr{G}^0_m \arrow[r, "\hat{d}"] \arrow[d, equal] & \cdots \arrow[r, "\hat{d}"] & \mathscr{G}^q_m \arrow[r, "\hat{d}"] \arrow[d, equal] & \cdots \\
 & \Omega^m\oplus\bar{\Omega}^m \arrow[r, "\begin{pmatrix}1&0\\0&1\end{pmatrix}"] & A^{m,0}\oplus A^{0,m} \arrow[r, "\begin{pmatrix}\bar{\partial}&0\\0&\partial\end{pmatrix}"] & \cdots \arrow[r, "\begin{pmatrix}\bar{\partial}&0\\0&\partial\end{pmatrix}"] & A^{m,q}\oplus A^{q,m} \arrow[r, "\begin{pmatrix}\bar{\partial}&0\\0&\partial\end{pmatrix}"] & \cdots
\end{tikzcd}$$

We define on $\mathscr{G}^{\cdot}_m$ the involution $(\varphi,\psi)\mapsto(\varphi,\psi)^*:=(\bar{\psi},\bar{\varphi})$. It is related to the $\partial\bar{\partial}$-complex by a homotopy operator $\beta:\mathscr{L}^{m+q+1}_{m+1}\to\mathscr{G}^q_m$ and a morphism of complexes $\gamma:\mathscr{L}^{m+q}_m\to\mathscr{G}^q_m$.

3.5.1. The Homotopy Operator $\beta:\mathscr{L}^{m+1+\cdot}_{m+1}\to\mathscr{G}^{\cdot}_m$. It is defined by

(i) For $q=-1$,

$$\begin{tikzcd}
\mathscr{L}^m_{m+1} \arrow[r, "\beta"] \arrow[d, equal] & \mathscr{G}^{-1}_m \arrow[d, equal] \\
\Omega^m\oplus A^{m-1}\oplus\bar{\Omega}^m \arrow[r, "\begin{pmatrix}-1&0&0\\0&0&1\end{pmatrix}"] & \Omega^m\oplus\bar{\Omega}^m.
\end{tikzcd}$$

(ii) For $0\leqq q<m$,

$$\begin{tikzcd}
\mathscr{L}^{m+q+1}_{m+1} \arrow[r, "\beta"] \arrow[d, equal] & \mathscr{G}^q_m \arrow[d, equal] \\
A^{m,q}\oplus\ldots\oplus A^{q,m} \arrow[r, "(-1)^{m-q}\begin{pmatrix}1&0\ldots0&0\\0&0\ldots0&1\end{pmatrix}"] & A^{m,q}\oplus A^{q,m}.
\end{tikzcd}$$

(iii) For $q=m$,

$$\begin{tikzcd}
\mathscr{L}_{m+1}^{2m+1} \arrow[r, "\beta"] \arrow[d, equal] & \mathscr{G}_m^m \arrow[d, equal] \\
A^{m,m} \arrow[r, "\frac{1}{2}\binom{1}{1}"] & A^{m,m}\oplus A^{m,m}.
\end{tikzcd}$$

(iv) For $q>m$, $\beta:\mathscr{L}_{m+1}^{m+q+1}\to\mathscr{G}_m^q$ is defined by 0.

3.5.2. The Morphism of Complexes $\gamma:\mathscr{L}_m^{m+\cdot}\to\mathscr{G}_m^{\cdot}$. It is defined by

(i) For $q=-1$,

$$\begin{tikzcd}
\mathscr{L}_m^{m-1} \arrow[r, "\gamma"] \arrow[d, equal] & \mathscr{G}_m^{-1} \arrow[d, equal] \\
\Omega^{m-1}\oplus A^{m-2}\oplus\bar{\Omega}^{m-1} \arrow[r, "\begin{pmatrix} d & 0 & 0 \\ 0 & 0 & -d \end{pmatrix}"] & \Omega^m\oplus\bar{\Omega}^m.
\end{tikzcd}$$

(ii) For $0\leqq q<m$,

$$\begin{tikzcd}
\mathscr{L}_m^{m+q} \arrow[r, "\gamma"] \arrow[d, equal] & \mathscr{G}_m^q \arrow[d, equal] \\
A^{m-1,q}\oplus\ldots\oplus A^{q,m-1} \arrow[r, "(-1)^{m-q}\begin{pmatrix} \partial & 0\ldots0 & 0 \\ 0 & 0\ldots0 & \bar{\partial} \end{pmatrix}"] & A^{m,q}\oplus A^{q,m}.
\end{tikzcd}$$

(iii) For $q\geqq m$,

$$\begin{tikzcd}
\mathscr{L}_m^{m+q} \arrow[r, "\gamma"] \arrow[d, equal] & \mathscr{G}_m^q \arrow[d, equal] \\
A^{q,m}\oplus\ldots\oplus A^{m,q} \arrow[r, "\begin{pmatrix} 0 & 0\ldots0 & 1 \\ -1 & 0\ldots0 & 0 \end{pmatrix}"] & A^{m,q}\oplus A^{q,m}.
\end{tikzcd}$$

In particular, for $q=m$, $\gamma(\alpha^{m,m})=(\alpha^{m,m}, -\alpha^{m,m})$.

The following can be easily checked.

3.5.3. Lemma. (i) $\hat{d}\beta+\beta D=\gamma\mu$.

(ii) $\hat{d}\gamma=\gamma D$.

(iii) *If* $\eta^{m,m}$ *and* $\zeta^{m,m}$ *are* (m,m)*-forms, then* $\beta(\eta^{m,m})+\gamma(\zeta^{m,m})=(\varrho^{m,m},\sigma^{m,m})$ *where* $\varrho^{m,m}+\sigma^{m,m}=\eta^{m,m}$.

(iv) β *and* γ *anticommute with the involutions* $(\cdot)^*$

$$\begin{tikzcd}
 & \mathscr{L}_m^{m+q-1} \arrow[dd, "D"] \arrow[dr, "\gamma"] & \\
\mathscr{L}_{m+1}^{m+q} \arrow[rr, "\beta"] \arrow[dr, "\mu"] \arrow[dd, "D"] & & \mathscr{G}_m^{q-1} \arrow[dd, "\hat{d}"] \\
 & \mathscr{L}_m^{m+q} \arrow[dr, "\gamma"] & \\
\mathscr{L}_{m+1}^{m+q+1} \arrow[rr, "\beta"] & & \mathscr{G}_m^q .
\end{tikzcd}$$

3.6. The Čech Transform of the $\partial\bar\partial$-Complex. For any Čech space $\underline{X}$, we denote by $\mathscr{E}_m^{\cdot}(\underline{X})$, $\mathscr{E}_m^{\cdot}(\underline{X},[\mathbb{R}])$ and $\mathscr{E}_m^{\cdot}(\underline{X},\mathbb{R})$ the Čech transforms of the complexes

$$0\to\mathbb{C}\to\mathscr{L}_m^0\to\mathscr{L}_m^1\to\dots,$$
$$0\to\mathbb{R}\to\mathscr{L}_m^0\to\mathscr{L}_m^1\to\dots,$$
$$0\to\mathbb{R}\to\mathscr{L}_{m,\mathbb{R}}^0\to\mathscr{L}_{m,\mathbb{R}}^1\to\dots,$$

respectively. So we set

$$\text{(3.6.1)}\qquad \begin{aligned} &\text{(i)} && \mathscr{E}_m^q(\underline{X}):=\check{C}^q(\underline{X};\mathbb{C},\mathscr{L}_m^{\cdot}),\\ &\text{(ii)} && \mathscr{E}_m^q(\underline{X},[\mathbb{R}]):=\check{C}^q(\underline{X};\mathbb{R},\mathscr{L}_m^{\cdot}),\\ &\text{(iii)} && \mathscr{E}_m^q(\underline{X},\mathbb{R}):=\check{C}^q(\underline{X};\mathbb{R},\mathscr{L}_{m,\mathbb{R}}^{\cdot}). \end{aligned}$$

Of course, $\mathscr{E}_m^q(\underline{X},\mathbb{R})\subset\mathscr{E}_m^q(X,[\mathbb{R}])\subset\mathscr{E}_m^q(X)$. Elements of $\mathscr{E}_m^q(\underline{X})$ will be written in a matrix form. For example an element of $\mathscr{E}_m^{2m}(\underline{X})$ will be written as

$$\text{(3.6.2)}\qquad \Phi=\begin{array}{|c|ccc|c|}\hline a & g^0 & \cdots & g^{m-1} & \multicolumn{1}{c}{} \\ \hline \bar h^0 & \varphi^{0,0} & \cdots & \varphi^{m-1,0} & \multicolumn{1}{c}{} \\ \vdots & \vdots & & \vdots & \multicolumn{1}{c}{} \\ \bar h^{m-1} & \varphi^{0,m-1} & \cdots & \varphi^{m-1,m-1} & \multicolumn{1}{c}{} \\ \hline \multicolumn{4}{c|}{} & \eta^{m,m} \\ \cline{5-5} \end{array}$$

where

$$a\in C^{2m}(\underline{X},\mathbb{C})$$
$$g^k\in C^{2m-k-1}(\underline{X},\Omega^k)$$
$$\bar h^l\in C^{2m-l-1}(\underline{X},\bar\Omega^l)$$
$$\varphi^{k,l}\in C^{2m-k-l-2}(\underline{X},A^{k,l})$$
$$\eta^{m,m}\in H^0(X,A^{m,m})$$

a is the head and $\eta^{m,m}$ the tail of Φ.

$$\Phi\in\mathscr{E}_m^{2m}(\underline{X},[\mathbb{R}])\quad\text{iff}\quad a=\bar a\in C^{2m}(\underline{X},\mathbb{R})$$
$$\Phi\in\mathscr{E}_m^{2m}(\underline{X},\mathbb{R})\quad\text{iff}\quad a=\bar a,$$

$g^k=h^k$, $\varphi^{k,l}+\bar\varphi^{l,k}=0$ and $\eta^{m,m}=\bar\eta^{m,m}$.

If we apply $\Delta:\mathscr{E}_m^{2m}(\underline{X})\to\mathscr{E}_m^{2m+1}(\underline{X})$ we obtain

$$\Delta\Phi=\begin{array}{|c|ccc|c|c|}\hline b & u^0 & \cdots & u^{m-1} & \multicolumn{2}{c}{} \\ \hline \bar v^0 & \psi^{0,0} & \cdots & \psi^{m-1,0} & \multicolumn{2}{c}{} \\ \vdots & \vdots & & \vdots & \multicolumn{2}{c}{} \\ \bar v^{m-1} & \psi^{0,m-1} & \cdots & \psi^{m-1,m-1} & \multicolumn{2}{c}{} \\ \hline \multicolumn{4}{c|}{} & \psi^{m,m} & \lambda^{m+1,m} \\ \cline{5-6} \multicolumn{4}{c|}{} & \lambda^{m,m+1} & \multicolumn{1}{c}{} \\ \cline{5-5} \end{array}$$

where

$$
(3.6.3)\qquad
\begin{array}{rl}
\text{(i)} & b=\delta a \\
\text{(ii)} & u^0=\delta g^0-a \\
\text{(iii)} & u^k=\delta g^k-dg^{k-1} \quad \text{for} \quad 1\leqq k<m \\
\text{(iv)} & \bar{v}^0=\delta \bar{h}^0-a \\
\text{(v)} & \bar{v}^l=\delta \bar{h}^l-d\bar{h}^{l-1} \quad \text{for} \quad 1\leqq l<m \\
\text{(vi)} & \psi^{0,0}=\delta\varphi^{0,0}+g^0-\bar{h}^0 \\
\text{(vii)} & \psi^{k,0}=\delta\varphi^{k,0}+(-1)^k g^k-\partial\varphi^{k-1,0} \quad \text{for} \quad 1\leqq k<m \\
\text{(viii)} & \psi^{0,l}=\delta\varphi^{0,l}-\bar{\partial}\varphi^{0,l-1}+(-1)^{l-1}\bar{h}^l \quad \text{for} \quad 1\leqq l<m \\
\text{(ix)} & \psi^{k,l}=\delta\varphi^{k,l}-\bar{\partial}\varphi^{k,l-1}-\partial\varphi^{k-1,l} \quad \text{for} \quad 1\leqq k,l<m \\
\text{(x)} & \psi^{m,m}=\varepsilon(\eta^{m,m})-\partial\bar{\partial}\varphi^{m-1,m-1} \\
\text{(xi)} & \lambda^{m+1,m}=\partial\eta^{m,m} \\
\text{(xii)} & \lambda^{m,m+1}=\bar{\partial}\eta^{m,m}.
\end{array}
$$

In the next section we construct, for any Kähler space (X,ω), an open covering $\mathscr{X}$ such that on the resulting Čech space $\underline{X}$ and for any integer $m>0$, there is a cocycle in $\mathscr{E}_m^{2m}(\underline{X},[\mathbb{R}])$ whose tail is ω^m.

4. *The Čech Cochains Associated to a Kähler Metric*

We first note that, if X is a Kähler space, it admits by definition an open covering (U_α) such that there are elements $\varphi_\alpha\in SP^\infty(U_\alpha)$ such that $\varphi_\alpha-\varphi_\beta$ is *locally* the real part of a holomorphic function on $U_\alpha\cap U_\beta$. We show that "locally" can be omitted.

4.1. Covering Lemma. *Let X be a paracompact topological space and $(U_\alpha)_{\alpha\in A}$ an open covering of X such that, for every $\alpha,\beta\in A$, $(U^j_{\alpha\beta})_{j\in J_{\alpha\beta}}$ is an open covering of $U_\alpha\cap U_\beta$. Let $J=\bigcup\limits_{\alpha,\beta} J_{\alpha\beta}$.*

Then there exists a refinement

$$\mathscr{X}=(X_\lambda)_{\lambda\in\Lambda}$$

of (U_α) together with two maps

$$\alpha:\Lambda\to A$$

$$j:\Lambda\times\Lambda\to J$$

such that

$$
(4.1.1)\qquad
\begin{array}{rl}
\text{(i)} & X_\lambda\subset U_{\alpha(\lambda)} \\
\text{(ii)} & X_\lambda\cap X_\mu\subset U^{j(\lambda,\mu)}_{\alpha(\lambda)\alpha(\mu)}.
\end{array}
$$

Proof. Since X paracompact, (U_α) admits a refinement $(\bar{V}_\alpha)_{\alpha\in A}$ indexed by the same set A such that $\bar{V}_\alpha \subset U_\alpha$ and $(\bar{V}_\alpha)$ is locally finite. Let Λ be the set of all multi-indices

$$\lambda = (\alpha_0, \ldots, \alpha_s; j_0, \ldots, j_s) \quad (s \in \mathbb{N}) \tag{4.1.2}$$

such that the α_r are pairwise distinct elements of A and $j_r \in J_{\alpha_0\alpha_r}$ for $0 \leqq r \leqq s$. Set

$$X_\lambda := V_{\alpha_0} \cap \bigcap_{r=0}^{s} U_{\alpha_0\alpha_r}^{j_r} \Big\backslash \bigcup_{\beta \neq \alpha_0, \ldots, \alpha_s} \bar{V}_\beta . \tag{4.1.3}$$

X_λ is open since $(\bar{V}_\beta)$ is locally finite.

Define $\alpha(\lambda) := \alpha_0$. Then obviously $X_\lambda \subset U_{\alpha(\lambda)}$.

Now suppose that

$$\mu = (\beta_0, \ldots, \beta_t; k_0, \ldots, k_t)$$

is a multi-index in Λ such that $X_\lambda \cap X_\mu \neq \emptyset$. Then β_0 must be equal to one (and only one) of the α_r for otherwise $\bar{V}_{\beta_0} \cap X_\lambda$ would be empty by construction of X_λ. If $\beta_0 = \alpha_r$, set

$$j(\lambda, \mu) := j_r .$$

It is clear that

$$X_\lambda \cap X_\mu \subset U_{\alpha_0\alpha_r}^{j_r} = U_{\alpha_0\beta_0}^{j_r} = U_{\alpha(\lambda)\alpha(\mu)}^{j(\lambda,\mu)}$$

as required. Finally it is true that the X_λ $(\lambda \in \Lambda)$ cover X; for if $x \in X$ is arbitrary, take $\alpha \in A$ such that $x \in V_\alpha$. The set S of $\beta \in A$ such that $x \in \bar{V}_\beta$ is finite containing α [since $(\bar{V}_\beta)$ is locally finite]; let

$$S = \{\alpha_0, \ldots, \alpha_s\} \quad \text{with} \quad \alpha_0 = \alpha .$$

For all

$$r \in \{0, \ldots, s\}, \qquad x \in V_{\alpha_0} \cap \bar{V}_\alpha \subset U_{\alpha_0} \cap U_{\alpha_r}$$

hence $x \in U_{\alpha_0\alpha_r}^{j_r}$ for some $j_r \in J_{\alpha_0\alpha_r}$. So we obtain a multi-index $\lambda \in \Lambda$ with $x \in X_\lambda$. Since $x \in X$ was arbitrary, the proof is complete.

4.1.1. Corollary. *Let X be a Kähler space with a fixed Kähler metric κ. Then X admits an open covering $\mathscr{X} = (X_\lambda)$ in which is represented by elements*

$$\varphi_\lambda \in SP^\infty(X_\lambda)$$

such that

$$\varphi_\lambda - \varphi_\mu = f_{\lambda\mu} + \bar{f}_{\lambda\mu}, \qquad f_{\lambda\mu} \in \mathscr{O}(X_\lambda \cap X_\mu).$$

Proof. By definition, there is an open covering (U_α) together with $\psi_\alpha \in SP^\infty(U_\alpha)$ such that $\psi_\alpha - \psi_\beta \in PH(U_\alpha \cap U_\beta, \mathbb{R})$. This means that $U_\alpha \cap U_\beta$ admits an open covering $(U_{\alpha\beta}^j)_{j \in J_{\alpha\beta}}$ such that

$$(\psi_\alpha - \psi_\beta)|_{U_{\alpha\beta}^j} = g_{\alpha\beta}^j + \bar{g}_{\alpha\beta}^j, \qquad g_{\alpha\beta}^j \in \mathscr{O}(U_{\alpha\beta}^j).$$

Apply the Covering Lemma above to obtain an open covering (X_λ) of X with $X_\lambda \subset U_{\alpha(\lambda)}$ and $X_\lambda \cap X_\mu \subset U_{\alpha(\lambda)\alpha(\mu)}^{j(\lambda,\mu)}$. Then if we set

$$\varphi_\lambda := \psi_{\alpha(\lambda)}|_{X_\lambda}$$

$$f_{\lambda\mu} := g_{\alpha(\lambda)\alpha(\mu)}^{j(\lambda,\mu)}|_{X_\lambda \cap X_\mu}$$

these elements satisfy the required conditions.

4.2. Kähler-Čech Pairs. It will be convenient to multiply the above elements φ_λ and $f_{\lambda\mu}$ by $i=\sqrt{-1}$ to obtain

$$\text{(4.2.1)}\qquad \begin{aligned} &\text{(i)} && \varphi_\lambda-\varphi_\mu=f_{\lambda\mu}-\bar{f}_{\lambda\mu}\\ &\text{(ii)} && -i\varphi_\lambda\in SP^\infty(X_\lambda)\\ &\text{(iii)} && \partial\bar\partial\varphi_\lambda=\omega|_{X_\lambda}\,. \end{aligned}$$

So the Kähler metric of X is

$$\kappa=\{(X_\lambda,-i\varphi_\lambda)\}\,.$$

A pair (f,φ) with $f\in C^1(\underline{X},\Omega^0)$ and $\varphi\in C^0(\underline{X},A^0)$ satisfying (4.2.1) will be called a *Kähler-Čech pair* and $\underline{X}=(X,\mathscr{X})$ will be called a *Kähler-Čech space.* Since $(\delta\varphi)_{\lambda\mu}=\varphi_\mu-\varphi_\lambda$, we have the identities

$$\text{(A)}\qquad \begin{aligned} &(1) && \delta\varphi=\bar f-f\\ &(2) && \delta f=\delta\bar f\\ &(3) && d\delta f=0\\ &(4) && \partial\delta\varphi=-df\\ &(5) && \bar\partial\delta\varphi=d\bar f\\ &(6) && \partial\bar\partial\varphi=\varepsilon(\omega)\\ &(7) && d\omega=0\,. \end{aligned}$$

Identity (A2) shows that $\delta f\in Z^2(\underline{X},\mathbb{R})$. The diagram

$$\text{(4.2.2)}\qquad \begin{tikzcd} & & -i\delta\varphi \arrow[dd, "-i\partial"'] & i\varphi \arrow[l, "\delta"'] \arrow[dd, "-i\partial"'] \arrow[dr, "i\partial\bar\partial"] & & \\ \delta f & f \arrow[l, "\delta"'] \arrow[ur, "-2\,\mathrm{Im}"] \arrow[dr, "d"'] & & & \partial\bar\partial\varphi & \omega \arrow[l, "\varepsilon"'] \\ & & df & \partial\varphi \arrow[l, "\delta"'] \arrow[ur, shift left, "d"] \arrow[ur, shift right, "\bar\partial"'] & & \end{tikzcd}$$

shows that $-2\,\mathrm{Im} f=-i\delta\varphi$ represents the Kähler class $\hat{c}_1(\kappa)$ of (X,κ) in $H^1(X,PH_{X,\mathbb{R}})$, δf represents $c_1(\kappa)\in H^2(X,\mathbb{R})$ and df represents $\tilde{c}_1(\kappa)\in H^1(X,\Omega^1_X)$. Moreover (4.2.2) confirms that ω is a d-closed representative of $c_1(\kappa)$ and a $\bar\partial$-closed representative of $\tilde{c}_1(\kappa)$, i.e. that diagram (1.2.4) of Chap. II is indeed commutative.

In terms of the $\partial\bar\partial$-complex $\mathscr{L}^{\cdot}_{1,\mathbb{R}}$ given by

$$\text{(4.2.3)}\qquad \begin{tikzcd} 0 \arrow[r] & \mathbb{R} \arrow[r, "\binom{1}{1}"] & \mathscr{L}^0_{1,\mathbb{R}} \arrow[r, "D"] \arrow[d, equal] & \mathscr{L}^1_{1,\mathbb{R}} \arrow[r, "D"] \arrow[d, equal] & \mathscr{L}^2_{1,\mathbb{R}} \arrow[r, "D"] \arrow[d, equal] & \mathscr{L}^3_{1,\mathbb{R}} \arrow[d, equal] \\ & & (\Omega^0\oplus\bar\Omega^0)_{\mathbb{R}} \arrow[r, "(-1\ \ 1)"] & (A^{0,0})_{i\mathbb{R}} \arrow[r, "\partial\bar\partial"] & A^{1,1}_{\mathbb{R}} \arrow[r, "d"] & (A^{1,2}\oplus A^{2,1})_{\mathbb{R}} \end{tikzcd}$$

[where $(\cdot)_{\mathbb{R}}$ denotes self-conjugate elements and $(\cdot)_{i\mathbb{R}}$ anti-self-conjugate elements] we constructed an element

$$(4.2.4)\qquad \Phi_1(f,\varphi):=\begin{array}{|c|c|c}\hline \delta f & f & \\ \hline \bar{f} & \varphi & \\ \hline \multicolumn{2}{c|}{} & \multicolumn{1}{c|}{\omega} \\ \cline{3-3}\end{array}\ \in \mathscr{E}_1^2(\underline{X},\mathbb{R})$$

(with the notations of 3.6) and relations (A) mean precisely that $\Delta\Phi_1(f,\varphi)=0$.

4.3. Generalization to Higher Powers. We now construct the announced element

$$\Phi_m(f,\varphi)\in \check{Z}^{2m}(X;\mathbb{R},\mathscr{L}_m^{\cdot})$$

whose head is $(\delta f)^m$ and tail ω^m, and whose existence is the key step in the proof of Theorem 2. Actually, if we set

$$\text{(i)}\quad \check{\mathscr{K}}^m(\underline{X}):=\check{Z}^{2m}(\underline{X};\mathbb{C},\mathscr{L}_m^{\cdot}),\qquad \check{\mathscr{K}}(\underline{X}):=\bigoplus_{m\geqq 0}\check{\mathscr{K}}^m(\underline{X})$$

$$(4.3.1)\text{(ii)}\quad \check{\mathscr{K}}^m(\underline{X},[\mathbb{R}]):=\check{Z}^{2m}(\underline{X};\mathbb{R},\mathscr{L}_m^{\cdot}),\qquad \check{\mathscr{K}}(\underline{X},[\mathbb{R}]:=\bigoplus_{m\geqq 0}\check{\mathscr{K}}^m(\underline{X},[\mathbb{R}])$$

$$\text{(iii)}\quad \check{\mathscr{K}}^m(\underline{X},\mathbb{R}):=\check{Z}^{2m}(\underline{X};\mathbb{R},\mathscr{L}_{m,\mathbb{R}}^{\cdot}),\qquad \check{\mathscr{K}}(\underline{X},\mathbb{R}):=\bigoplus_{m\geqq 0}\check{\mathscr{K}}^m(\underline{X},\mathbb{R})$$

then there is an associative product law on $\check{\mathscr{K}}(\underline{X})$ with respect to which it is a graded $\mathbb{C}$-algebra admitting $\check{\mathscr{K}}(\underline{X},[\mathbb{R}])$ as a $\mathbb{R}$-subalgebra, but not $\check{\mathscr{K}}(\underline{X},\mathbb{R})$. Then $\Phi_m(f,\varphi)$ is simply the m-th power of $\Phi_1(f,\varphi)$ in $\check{\mathscr{K}}(\underline{X},\mathbb{R})$.

$\Phi_m(f,\varphi)$ is defined by

$$(4.3.2)\quad \Phi_m(f,\varphi):=\begin{array}{|c|ccc|c}\hline a_m & g_m^0 & \cdots & g_m^{m-1} & \\ \hline \bar{h}_m^0 & \varphi_m^{0,0} & \cdots & \varphi_m^{m-1,0} & \\ \vdots & \vdots & & \vdots & \\ \bar{h}_m^{m-1} & \varphi_m^{0,m-1} & \cdots & \varphi_m^{m-1,m-1} & \\ \hline \multicolumn{4}{c|}{} & \multicolumn{1}{c|}{\eta_m^{m,m}} \\ \cline{5-5}\end{array}\ \in\check{\mathscr{K}}^m(\underline{X},[\mathbb{R}]),$$

where $a_m\in C^{2m}(\underline{X},\mathbb{R})$, $g_m^k\in C^{2m-k-1}(\underline{X},\Omega^k)$,

$$\bar{h}_m^l\in C^{2m-l-1}(\underline{X},\bar{\Omega}^l),\qquad \varphi_m^{k,l}\in C^{2m-k-l-2}(\underline{X},A^{k,l}),\qquad \eta_m^{m,m}\in H^0(X,A^{m,m})$$

are given by the relations (B) below. Recall that $\delta f=\delta\bar{f}$ by (A2). We use the cup-product of Čech cochains as defined in 1.2.

(B)

(1) $a_m=(\delta f)^m$

(2) $g_m^k=(-1)^k(df)^k\cdot f\cdot(\delta f)^{m-k-1}$

(3) $\bar{h}_m^l=(\delta f)^{m-l-1}\cdot\bar{f}\cdot(d\bar{f})^l$

(4) $\varphi_m^{k,l}=(-1)^{k+l}(df)^k\cdot f\cdot(\delta f)^{m-k-l-2}\cdot\bar{f}\cdot(d\bar{f})^l$ for $k+l<m-1$

(5) $\varphi_m^{k,l}=(-1)^{m-l-1}(df)^{m-l-1}\cdot\delta\varphi\cdot(d\bar{f})^{m-k-2}\cdot\bar{\partial}\varphi\wedge\omega^{k+l-m+1}$

for $k<m-1\leqq k+l$

(6) $\varphi_m^{m-1,l}=(-1)^{m-l-1}(df)^{m-l-1}\cdot\varphi\wedge\omega^l$

(7) $\eta_m^{m,m}=\omega^m$.

Domains of validity of formulae (B)

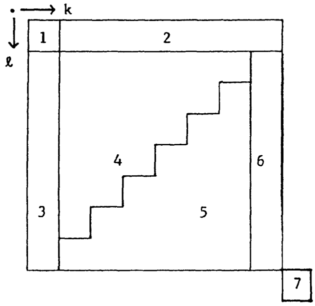


Fig. 1

Before proving that $\Delta\Phi_m(f,\varphi)=0$ we mention

4.4. Relation Between $\Phi_m(f,\varphi)$, $\Phi_n(f,\varphi)$, and $\Phi_{m+n}(f,\varphi)$. A formal consequence of identities (A) and (B) is the following:

(C)

(1) $a_{m+n}=a_m\cdot a_n$

(2) $g_{m+n}^k=g_m^k\cdot a_n$ for $0\leqq k<m$

(3) $\quad=(-1)^m dg_m^{m-1}\cdot g_n^{k-m}$ for $m\leqq k<m+n$

(4) $\bar{h}_{m+n}^l=a_m\cdot\bar{h}_n^l$ for $0\leqq l<n$

(5) $\quad=\bar{h}_m^{l-n}\cdot d\bar{h}_n^{n-1}$ for $n\leqq l<m+n$

(6) $\varphi_{m+n}^{k,l}=(-1)^l g_m^k\cdot\bar{h}_n^l$ for $0\leqq k<m$, $0\leqq l<n$

(7) $\quad=(-1)^m dg_m^{m-1}\cdot\varphi_n^{k-m,l}$ for $m\leqq k<m+n$, $0\leqq l<n$

(8) $\quad=(-1)^n\varphi_m^{k,l-n}\cdot d\bar{h}_n^{n-1}$ for $n\leqq l<m+n-k-1$

(9) $\quad=(-1)^{m-1}\delta\varphi_m^{m+n-l-1,l-n}\cdot\bar{\partial}\varphi_n^{k+l-m-n+1,n-1}$
for $l\geqq n$, $m+n-1\leqq k+l<m+2n-1$

(10) $\quad=\varphi_m^{k-n,l-n}\wedge\eta_n^{n,n}$ for $k+l\geqq m+2n-1$

(11) $\eta_{m+n}^{m+n,m+n}=\eta_m^{m,m}\wedge\eta_n^{n,n}$.

Domains of validity of formulae (C)

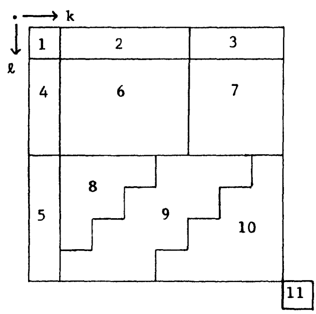


Fig. 2

Actually, the identities (C) define the announced product law

$$\tilde{\mathscr{K}}^m(\underline{X}) \times \tilde{\mathscr{K}}^n(\underline{X}) \to \tilde{\mathscr{K}}^{m+n}(\underline{X}).$$

It is true (the proof be omitted) that the law in question is associative (it will be denoted by the symbol $\times$) and, if Φ and Ψ are Δ-closed, $\Phi \times \Psi$ is also Δ-closed. However it is not compatible with the involution defined in 3.3 and this is the reason for which we work in $\tilde{\mathscr{K}}^m(\underline{X}, [\mathbb{R}])$ instead of $\tilde{\mathscr{K}}^m(\underline{X}, \mathbb{R})$.

Identities (C) will be used to prove

4.5. The Relation $\Delta\Phi_m(f, \varphi) = 0$. In order to prove that the element $\Phi_m(f, \varphi)$ defined in (4.3.3) is Δ-closed, we must prove according to (3.6.3) the relations

$$
\text{(D)} \qquad
\begin{array}{ll}
(1) & \delta a_m = 0 \\
(2) & \delta g_m^0 = a_m \\
(3) & \delta g_m^k = d g_m^{k-1} \quad \text{for} \quad 1 \leqq k < m \\
(4) & \delta \bar{h}_m^0 = a_m \\
(5) & \delta \bar{h}_m^l = d \bar{h}_m^{l-1} \quad \text{for} \quad 1 \leqq l < m \\
(6) & \delta \varphi_m^{0,0} = -g_m^0 + \bar{h}_m^0 \\
(7) & \delta \varphi_m^{k,0} = (-1)^{k-1} g_m^k + \partial \varphi_m^{k-1,l} \quad \text{for} \quad 1 \leqq k < m \\
(8) & \delta \varphi_m^{0,l} = \bar{\partial} \varphi_m^{0,l-1} + (-1)^l \bar{h}_m^l \quad \text{for} \quad 1 \leqq l < m \\
(9) & \delta \varphi_m^{k,l} = \bar{\partial} \varphi_m^{k,l-1} + \partial \varphi_m^{k-1,l} \quad \text{for} \quad 1 \leqq k, l < m \\
(10) & \varepsilon(\eta_m^{m,m}) = \partial \bar{\partial} \varphi_m^{m-1,m-1} \\
(11) & d\eta_m^{m,m} = 0 .
\end{array}
$$

Proof of (D1). It is obvious.

Proof of (D2). $\delta g_m^0 = \delta(f \cdot (\delta f)^{m-1}) = (\delta f)^m = a_m$.

Proof of (D3).

$$
\begin{aligned}
\delta g_m^k &= \delta((-1)^k (df)^k \cdot f \cdot (\delta f)^{m-k-1}) = (df)^k \cdot (\delta f)^{m-k} \\
&= d((-1)^{k-1} (df)^{k-1} \cdot f \cdot (\delta f)^{m-k}) = d g_m^{k-1} .
\end{aligned}
$$

Proof of (D4).

$$
\begin{aligned}
\delta \bar{h}_m^0 &= \delta((\delta f)^{m-1} \cdot \bar{f}) = (\delta f)^m \quad \text{by (A2)} \\
&= a_m .
\end{aligned}
$$

Proof of (D5).

$$
\begin{aligned}
\delta \bar{h}_m^l &= \delta((\delta f)^{m-l-1} \cdot \bar{f} \cdot (d\bar{f})^l) = (\delta f)^{m-l} \cdot (d\bar{f})^l \quad \text{by (A2)} \\
&= d((\delta f)^{m-l} \cdot \bar{f} \cdot (d\bar{f})^{l-1}) = d \bar{h}_m^{l-1} .
\end{aligned}
$$

Proof of (D6).

$$\delta\varphi_m^{0,0}=\delta(f\cdot(\delta f)^{m-2}\cdot\bar{f})=(\delta f)^{m-1}\cdot\bar{f}-f\cdot(\delta f)^{m-1}=\bar{h}_m^0-g_m^0.$$

Proof of (D7). *Case 1.* $k<m-1$

$$\begin{aligned}\delta\varphi_m^{k,0}&=\delta((-1)^k dg_k^{k-1}\cdot\varphi_{m-k}^{0,0}) \quad \text{by (C7)}\\ &=dg_k^{k-1}\cdot\delta\varphi_{m-k}^{0,0}=dg_k^{k-1}\cdot(-g_{m-k}^0+\bar{h}_{m-k}^0) \quad \text{by (D6)}\\ &=-dg_k^{k-1}\cdot g_{m-k}^0+\partial(g_k^{k-1}\cdot\bar{h}_{m-k}^0)\\ &=(-1)^{k-1}g_m^k+\partial\varphi_m^{k-1,0} \quad \text{by (C3) and (C6)}.\end{aligned}$$

Case 2. $k=m-1$

$$\begin{aligned}\delta\varphi_m^{m-1,0}&=\delta((-1)^{m-1}(df)^{m-1}\cdot\varphi)=(df)^{m-1}\cdot\delta\varphi=(df)^{m-1}\cdot(-f+\bar{f}) \quad \text{by (A1)}\\ &=-(df)^{m-1}\cdot f+\partial((-1)^{m-2}(df)^{m-2}\cdot f\cdot\bar{f})\\ &=(-1)^m g_m^{m-1}+\partial\varphi_m^{m-2,0} \quad \text{by (B2) and (B5)}.\end{aligned}$$

Proof of (D8). *Case 1.* $l<m-1$

$$\begin{aligned}\delta\varphi_m^{m-1,0}&=\delta((-1)^{m-1}(df)^{m-1}\cdot\varphi)=(df)^{m-1}\cdot\delta\varphi=(df)^{m-1}\cdot(-f+\bar{f}) \quad \text{by (A1)}\\ &=-(df)^{m-1}\cdot f+\partial((-1)^{m-2}(df)^{m-2}\cdot f\cdot\bar{f})\\ &=(-1)^m g_m^{m-1}+\partial\varphi_m^{m-2,0} \quad \text{by (B2) and (B5)}.\end{aligned}$$

Proof of (D8). *Case 1.* $l<m-1$

$$\begin{aligned}\delta\varphi_m^{0,l}&=\delta((-1)^l\varphi_{m-l}^{0,0}\cdot d\bar{h}^{l-1}) \quad \text{by (C8)}\\ &=(-1)^l\delta\varphi_{m-l}^{0,0}\cdot d\bar{h}_l^{l-1}=(-1)^l(-g_{m-l}^0+\bar{h}_{m-l}^0)\cdot d\bar{h}_l^{l-1} \quad \text{by (D6)}\\ &=\bar{\partial}((-1)^{l-1}g_{m-l}^0\cdot\bar{h}^{l-1})+(-1)^l\bar{h}_{m-l}^0\cdot d\bar{h}_l^{l-1}\\ &=\bar{\partial}\varphi_m^{0,l-1}+(-1)^l\bar{h}_m^l \quad \text{by (C6) and (C5)}.\end{aligned}$$

Case 2. $l=m-1$

$$\begin{aligned}\delta\varphi_m^{0,m-1}&=\delta(\delta\varphi\cdot(d\bar{f})^{m-2}\cdot\bar{\partial}\varphi)=(-1)^{m-1}\delta\varphi\cdot(d\bar{f})^{m-1} \quad \text{by (A5)}\\ &=(-1)^m(f-\bar{f})\cdot(d\bar{f})^{m-1} \quad \text{by (A1)}\\ &=\bar{\partial}((-1)^{m-2}f\cdot\bar{f}\cdot(d\bar{f})^{m-2})+(-1)^{m-1}\bar{f}\cdot(d\bar{f})^{m-2}\\ &=\bar{\partial}\varphi_m^{0,m-2}+(-1)^{m-1}\bar{h}_m^{m-1} \quad \text{by (B4) and (B5)}.\end{aligned}$$

Proof of (D9). *Case1.* $k+l<m-1$.

We can write $m=r+s$ with $r>k$ and $s>l$. Then

$$\begin{aligned}\delta\varphi_m^{k,l}&=\delta\varphi_{r+s}^{k,l}=\delta((-1)^l g_r^k\cdot\bar{h}_s^l) \quad \text{by (C6)}\\ &=(-1)^l\delta g_r^k\cdot\bar{h}_s^l+(-1)^{k+l-1}g_r^k\cdot\delta\bar{h}_s^l\\ &=(-1)^l dg_r^{k-1}\cdot\bar{h}_s^l+(-1)^{k+l-1}g_r^k\cdot d\bar{h}_s^{l-1} \quad \text{by (D3) and (D5)}\\ &=\partial((-1)^l g_r^{k-1}\cdot\bar{h}_s^l)+\bar{\partial}((-1)^{l-1}g_r^k\cdot\bar{h}_s^{l-1})\\ &=\partial\varphi_m^{k-1,l}+\bar{\partial}\varphi_m^{k,l-1} \quad \text{by (C6)}.\end{aligned}$$

Case 2. $k+l=m-1$

$$\begin{aligned}
\delta\varphi_m^{k,l} &= \delta((-1)^k(df)^k\cdot\delta\varphi\cdot(d\bar{f})^{l-1}\cdot\bar{\partial}\varphi) \quad \text{by (B5)}\\
&=(-1)^l(df)^k\cdot\delta\varphi\cdot(d\bar{f})^l \quad \text{by (A5)}\\
&=(-1)^{l-1}(df)^k\cdot f\cdot(d\bar{f})^l+(-1)^l(df)^k\cdot\bar{f}\cdot(d\bar{f})^l \quad \text{by (A1)}\\
&=(-1)^{k+l-1}g_{k+1}^k\cdot d\bar{h}_l^{l-1}+(-1)^l dg_k^{k-1}\cdot\bar{h}_{l+1}^l \quad \text{by (B2) and (B3)}\\
&=\bar{\partial}((-1)^{l-1}g_{k+1}^k\cdot\bar{h}_l^{l-1})+\partial((-1)^l g_k^{k-1}\cdot\bar{h}_{l+1}^l)\\
&=\bar{\partial}\varphi_m^{k,l-1}+\partial\varphi_m^{k-1,l} \quad \text{by (C6)}.
\end{aligned}$$

Case 3. $k<m-1<k+l$

$$\begin{aligned}
\delta\varphi_m^{k,l} &=(-1)^{m-k-1}(df)^{m-l-1}\cdot\delta\varphi\cdot(d\bar{f})^{m-k-1}\wedge\omega^{k+l-m+1} \quad \text{by (B5) and (A5)}\\
&=(-1)^{m-k-1}(df)^{m-l-1}\cdot\delta\varphi\cdot(d\bar{f})^{m-k-1}\cdot\partial\bar{\partial}\varphi\wedge\omega^{k+l-m} \quad \text{by (A6)}\\
&=\bar{\partial}((-1)^{m-l}(df)^{m-l}\cdot\delta\varphi\cdot(d\bar{f})^{m-k-2}\cdot\bar{\partial}\varphi\wedge\omega^{k+l-m})\\
&\quad+\partial((-1)^{m-l-1}(df)^{m-l-1}\cdot\delta\varphi\cdot(d\bar{f})^{m-k-1}\cdot\bar{\partial}\varphi\wedge\omega^{k+l-m}) \quad \text{by (A4), (A5), (A6)}\\
&=\bar{\partial}\varphi_m^{k,l-1}+\partial\varphi_m^{k-1,l} \quad \text{by (B5)}.
\end{aligned}$$

Case 4. $k=m-1$

$$\begin{aligned}
\delta\varphi_m^{m-1,l} &=(df)^{m-l-1}\cdot\delta\varphi\wedge\omega^l \quad \text{by (B6)}\\
&=\bar{\partial}((-1)^{m-l}(df)^{m-l}\cdot\varphi\wedge\omega^{l-1})+\partial((-1)^{m-l-1}(df)^{m-l-1}\cdot\delta\varphi\cdot\bar{\partial}\varphi\wedge\omega^{l-1})\\
&=\bar{\partial}\varphi_m^{m-1,l-1}+\partial\varphi_m^{m-2,l} \quad \text{by (B5) and (B6)}.
\end{aligned}$$

Finally, (D10) and (D11) are obvious since $\varphi_m^{m-1,m-1}=\varphi\omega^{m-1}$ and $\eta_m^{m,m}=\omega^m$. Therefore the proof of the relation $\Delta\Phi_m(f,\varphi)=0$ is complete.

4.5.1. Remark. There are several alternative ways of proving $\Delta\Phi_m(f,\varphi)=0$. For example, identities (C) written only for $n=1$ give a relation between $\Phi_m(f,\varphi)$ and $\Phi_{m+1}(f,\varphi)$, and the relation $\Delta\Phi_m(f,\varphi)$ can be proven by induction on m. Otherwise, one can prove directly that $\Delta\Phi_m=\Delta\Phi_n=0$ implies $\Delta(\Phi_m\times\Phi_n)=0$ using (A), (B), and (C) but the calculations would be longer than the above (30 verifications are needed).

5. *Theorem 2*

5.1. Statement of Theorem 2. Let (X,ω) be a Kähler space and $m\geqq 0$ an integer. Then there exist open sets $U_\alpha\subset X$ $(\alpha\in A)$ and $U_{\alpha\beta}^j\subset U_\alpha\cap U_\beta$ $(j\in J_{\alpha\beta})$ depending on X and m alone such that

(i) Any compact m-dimensional complex-analytic subset of X is contained in some U_α.

(ii) Any compact m-dimensional complex-analytic subset of $U_\alpha\cap U_\beta$ is contained in some $U_{\alpha\beta}^j$.

(iii) There exist elements $\chi_\alpha\in A^{m,m}(U_\alpha,\mathbb{R})$ such that

$$\omega^{m+1}|_{U_\alpha}=i\partial\bar{\partial}\chi_\alpha.$$

(iv) There exist elements $\tau^j_{\alpha\beta}\in A^{m,m}(U^j_{\alpha\beta})$ such that

$$\bar\partial\tau^j_{\alpha\beta}=0 \quad\text{and}\quad (\chi_\alpha-\chi_\beta)|_{U_{\alpha\beta}}=\tau^j_{\alpha\beta}+\bar\tau^j_{\alpha\beta}\,.$$

(v) The $\tau^j_{\alpha\beta}$ are $\bar\partial$-closed representatives of elements $\xi^j_{\alpha\beta}\in H^m(U^j_{\alpha\beta},\Omega^m)$.

5.2. *Proof of (i) and (ii).* We take an open covering $\mathscr{X}$ of X such that $\underline{X}=(X,\mathscr{X})$ is a Kähler-Čech space with a Kähler-Čech pair (f,φ) as in 4.2.

The U_α are taken as the m-admissible open sets of X and the $U^j_{\alpha\beta}$ as the m-admissible open sets of $U_\alpha\cap U_\beta$. Parts (i) and (ii) of Theorem 2 are restatements of Lemma 3.5.4 of I. By Proposition 1.3.3, each U_α is underlying to some m-admissible $\underline{U}_\alpha\ll\underline{X}$ and each $U^j_{\alpha\beta}$ to some m-admissible $\underline{U}^j_{\alpha\beta}\ll\underline{U}_\alpha\cap\underline{U}_\beta$.

5.3. *Proof of (iii).* We use the element

$$\Phi_{m+1}(f,\varphi)\in\check{\mathscr{K}}^{m+1}(X,[\mathbb{R}])=\check{Z}^{2m+2}(\underline{X};\mathbb{R},\mathscr{L}^{\cdot}_{m+1})$$

which is Δ-closed in the Čech transform of the complex

$$0\longrightarrow\mathbb{R}\longrightarrow\mathscr{L}^0_{m+1}\longrightarrow\dots\longrightarrow\mathscr{L}^{2m+1}_{m+1}\xrightarrow{\partial\bar\partial}\mathscr{L}^{2m+2}_{m+1}\longrightarrow\dots\,.$$

Take the restriction [in the sense of (1.1.5)]

(5.3.1) $$\Phi_{m+1,\alpha}:=\Phi_{m+1}(f,\varphi)|_{\underline{U}_\alpha}\in\check{\mathscr{K}}^{m+1}(\underline{U}_\alpha,[\mathbb{R}])\,.$$

Since $\underline{U}_\alpha$ is m-complete, we have

$$H^{2m-k+1}(\underline{U}_\alpha,\mathscr{L}^k_{m+1})=0\quad\text{for}\quad 0\leqq k\leqq 2m\,.$$

Indeed, for $k\leqq m$ this is due to the m-completeness of $\underline{U}_\alpha$ and the fact that $\mathscr{L}^k_{m+1}=\Omega^k\oplus A^{k-1}\oplus\bar\Omega^k$; for $k>m$, it is due to the fact that $\mathscr{L}^k_{m+1}$ is a fine sheaf.

Corollary 2.3 applies and $\Phi_{m+1,\alpha}$ is Δ-exact if its head is δ-exact, since the canonical morphism

$$\check{H}^{2m+2}(\underline{U}_\alpha;\mathbb{R},\mathscr{L}^{\cdot}_{m+1})\to H^{2m+2}(\underline{U}_\alpha,\mathbb{R})$$

is injective. But the head of $\Phi_{m+1,\alpha}$ is $(\delta f)^{m+1}|_{\underline{U}_\alpha}$ whose class in $H^{2m+2}(\underline{U}_\alpha,\mathbb{R})$ is 0 by Lemma 1.3.2, since $\underline{U}_\alpha\ll\underline{X}$ is m-admissible. Therefore

(5.3.2) $$\Phi_{m+1,\alpha}=\Delta\Theta_{m+1,\alpha}$$

for some $\Theta_{m+1,\alpha}\in\mathscr{E}^{2m+1}_{m+1}(\underline{U}_\alpha,[\mathbb{R}])$. In particular, if $\psi_\alpha\in A^{m,m}(U_\alpha)$ is the tail of $\Theta_{m+1,\alpha}$, we have

(5.3.3) $$\omega^{m+1}|_{U_\alpha}=\partial\bar\partial\psi_\alpha\,.$$

It is then sufficient to set

(5.3.4) $$\chi_\alpha:=\frac{i}{2}(\bar\psi_\alpha-\psi_\alpha)$$

to satisfy condition (iii) of Theorem 2.

5.4. *Proof of (iv) and (v).* Take a fixed $\underline{U}=\underline{U}^j_{\alpha\beta}\ll\underline{U}_\alpha\cap\underline{U}_\beta$.

There are open inclusions of Čech open sets

$$\begin{tikzcd} \underline{U}_\alpha \arrow[d] & \underline{U} \arrow[l, "i_\alpha"'] \arrow[r, "i_\beta"] \arrow[dl, "j_\alpha"] \arrow[dr, "j_\beta"'] & \underline{U}_\beta \arrow[d] \\ \underline{X} & & \underline{X}. \end{tikzcd}$$

We may then apply the operator T of (1.1.6) relatively to $j_\alpha, j_\beta: \underline{U} \to \underline{X}$ and set

(5.4.1) $$\tilde{\Theta}_{m+1} := T\Phi_{m+1}(f, \varphi) \in \mathscr{E}^{2m+1}_{m+1}(\underline{U}, [\mathbb{R}]).$$

This element satisfies the conditions

(5.2.1) (i) $\Delta\tilde{\Theta}_{m+1} = j_\beta^*\Phi_{m+1}(f, \varphi) - j_\alpha^*\Phi_{m+1}(f, \varphi)$

(ii) The tail of $\tilde{\Theta}_{m+1}$ is 0.

Indeed, (i) is a consequence of (1.1.7) and (ii) of the fact that T induces 0 on 0-cochains and global sections. Now set

(5.4.3) $$\Theta_{m+1} := i_\alpha^*(\Theta_{m+1,\alpha}) - i_\beta^*(\Theta_{m+1,\beta}) + \tilde{\Theta}_{m+1} \in \mathscr{E}^{2m+1}_{m+1}(\underline{U}, [\mathbb{R}]).$$

This element satisfies, by (5.3.2) and (5.4.2)

(5.4.4) (i) $\Delta\Theta_{m+1} = 0$

(ii) The tail of Θ_{m+1} is $\psi := (\psi_\alpha - \psi_\beta)|_{\underline{U}}$.

We notice that Lemma 2.2(i) does not apply to the canonical morphism

$$\check{H}^{2m+1}(\underline{U}; \mathbb{R}, \mathscr{L}^{\cdot}_{m+1}) \to H^{2m+1}(\underline{U}, \mathbb{R})$$

for among the groups $H^{2m-k}(\underline{U}, \mathscr{L}^k_{m+1})$ there is $H^m(\underline{U}, \mathscr{L}^m_{m+1}) = H^m(\underline{U}, \Omega^m \oplus \bar{\Omega}^m)$ which is not 0 in general. So we apply the operator μ defined in 3.4 to obtain $\mu\Theta_{m+1} \in \mathscr{E}^{2m+1}_m(\underline{U}, [\mathbb{R}])$.

Since μ commutes with D (and δ), $\mu\Theta_{m+1}$ is Δ-closed. This time the canonical morphism

$$\check{H}^{2m+1}(\underline{U}; \mathbb{R}, \mathscr{L}^{\cdot}_m) \to H^{2m+1}(\underline{U}, \mathbb{R})$$

is injective since the groups $H^{2m-k}(\underline{U}, \mathscr{L}^k_m)$ are all 0 for $0 \leq k \leq 2m-1$. Indeed, for $k < m$ this is due to the m-completeness of $\underline{U}$ and, for $k \geq m$, to the fact that $\mathscr{L}^k_m$ is a fine sheaf. So, by Corollary 2.3, $\mu\Theta_{m+1}$ is Δ-exact if its head is δ-exact in $C^{\cdot}(\underline{U}, \mathbb{R})$. But the head of $\mu\Theta_{m+1}$ is equal to the head of Θ_{m+1} which is of the form $c_{m+1}|_{\underline{U}}$ with

$$c_{m+1} \in Z^{2m+1}(\underline{U}_\alpha \cap \underline{U}_\beta, \mathbb{R}).$$

Since $\underline{U} \ll \underline{U}_\alpha \cap \underline{U}_\beta$ is m-admissible, $c_{m+1}|_{\underline{U}}$ is δ-exact (Lemma 1.4.2) and therefore

(5.4.5) $$\mu\Theta_{m+1} = \Delta Z_m$$

for some $Z_m \in \mathscr{E}^{2m}_m(\underline{U}, \mathbb{R})$.

Now we use the operators β and γ defined in 3.5. Denote by $\mathscr{D}^{\cdot}_m(\underline{U})$ the Čech transform of the $(\bar{\partial} \oplus \partial)$-complex over $\underline{U}$, i.e.

(5.4.6) $$\mathscr{D}^q_m(\underline{U}) := \check{C}^q(\underline{U}; \Omega^m \oplus \bar{\Omega}^m, \mathscr{G}^{\cdot}_m)$$

with differential

$$(5.4.7)\qquad \hat{\Delta}:=\delta+(-1)^{m+q+1}\hat{d}:\mathscr{D}^q_m(\underline{U})\to\mathscr{D}^{q+1}_m(\underline{U}).$$

Notice that this sign convention differs from (2.1.3).

Diagram (3.5.3) becomes

$$(5.4.8)\qquad \begin{tikzcd}
 & \mathscr{E}^{2m}_m(\underline{U}) \arrow[dr, "\gamma"] \arrow[dd, "\Delta=\delta-D"] & \\
\mathscr{E}^{2m+1}_{m+1}(\underline{U}) \arrow[rr, "\beta"] \arrow[dr, "\mu"] \arrow[dd, "\Delta=\delta+D"'] & & \mathscr{D}^m_m(\underline{U}) \arrow[dd, "\hat{\Delta}=\delta-\hat{d}"] \\
 & \mathscr{E}^{2m+1}_m(\underline{U}) \arrow[dr, "\gamma"] & \\
\mathscr{E}^{2m+2}_{m+1}(\underline{U}) \arrow[rr, "\beta"] & & \mathscr{D}^{m+1}_m(\underline{U}).
\end{tikzcd}$$

By Lemma 3.5.3 and the sign convention (5.4.7) on $\hat{\Delta}$ we have on $\mathscr{E}^{2m+1}_{m+1}(\underline{U})$

$$(5.4.9)\qquad \begin{aligned}\beta\Delta-\hat{\Delta}\beta&=\beta(\delta+D)-(\delta-\hat{d})\beta=(\beta\delta-\delta\beta)+(\beta D+\hat{d}\beta)\\ &=\beta D+\hat{d}\beta=\gamma\mu.\end{aligned}$$

On the other hand, we have on $\mathscr{E}^{2m}_m(\underline{U})$

$$(5.4.10)\qquad \gamma\Delta=\hat{\Delta}\gamma.$$

If we apply (5.4.9) to Θ_{m+1} and (5.4.10) to Z_m, we get

$$-\hat{\Delta}\beta\Theta_{m+1}=(\beta\Delta-\hat{\Delta}\beta)\Theta_{m+1}=\gamma\mu\Theta_{m+1}=\gamma\Delta Z_m=\hat{\Delta}\gamma Z_m$$

which means that the element

$$(5.4.11)\qquad \Lambda_m:=\beta\Theta_{m+1}+\gamma Z_m\in\mathscr{D}^m_m(\underline{U})$$

satisfies

$$\hat{\Delta}\Lambda_m=0.$$

The tail of Λ_m has the form

$$(\varrho^{m,m},\sigma^{m,m})\in A^{m,m}(U)\oplus A^{m,m}(U)$$

with $\bar{\partial}\varrho^{m,m}=\partial\sigma^{m,m}=0$ (since $\hat{\Delta}\Lambda_m=0$) and

$$(5.4.12)\qquad \varrho^{m,m}+\sigma^{m,m}=\psi$$

by Lemma 3.5.3(iii).

The fact that Λ_m is a $\hat{\Delta}$-cocycle means precisely that $\varrho^{m,m}$ and $\bar{\sigma}^{m,m}$ represent elements of $H^m(\underline{U},\Omega^m)$. So if we set

$$(5.4.13)\qquad \tau^j_{\alpha\beta}:=\frac{i}{2}(\bar{\sigma}^{m,m}-\varrho^{m,m})$$

it is clear that conditions (iv) and (v) of Theorem 2 are satisfied.

5.5. *Remark.* (1) We did not use the positivity of ω in the proof of Theorem 2. The result we can actually prove by our method is the following: If U_α and $U^j_{\alpha\beta}$ are the open sets of Theorem 2, then conditions (i) and (ii) remain unchanged. If moreover $\kappa_0, \ldots, \kappa_m$ are arbitrary elements of $\mathscr{K}^1(X)$ and $\omega_q := \partial\bar{\partial}\kappa_q$ for $0 \leqq q \leqq m$, then

(iii) There are elements $\psi_\alpha \in A^{m,m}(U_\alpha)$ such that $(\omega_0 \wedge \ldots \wedge \omega_m)|_{U_\alpha} = \partial\bar{\partial}\psi_\alpha$.

(iv) There are elements $\varrho^j_{\alpha\beta}, \sigma^j_{\alpha\beta} \in A^{m,m}(U^j_{\alpha\beta})$ such that $\bar{\partial}\varrho^j_{\alpha\beta} = \partial\sigma^j_{\alpha\beta} = 0$ and $(\psi_\alpha - \psi_\beta)|_{U^j_{\alpha\beta}} = \varrho^j_{\alpha\beta} + \sigma^j_{\alpha\beta}$.

(v) $\varrho^j_{\alpha\beta}$ and $\bar{\sigma}^j_{\alpha\beta}$ represent cohomology classes of $H^m(U^j_{\alpha\beta}, \Omega^m)$.

(2) The proof we gave was a reasoning on $\mathscr{E}^{\cdot}_m(\underline{X}, [\mathbb{R}])$. We could have chosen $\mathscr{E}^{\cdot}_m(\underline{X}, \mathbb{R})$ as well, replacing $\Phi_{m+1}(f, \varphi)$ by

$$\mathrm{Re}(\Phi_{m+1}(f, \varphi)) = \tfrac{1}{2}(\Phi_{m+1}(f, \varphi) + \Phi_{m+1}(f, \varphi)^*)$$

and using Lemma 3.5.3(iv).

IV. The Main Results

1. Stability Theorems

We are now in position to prove that some proper images of Kähler spaces are Kähler.

1.1. Theorem 3. *Let $\pi: X \to X'$ be a geometrically flat morphism of complex spaces with m-dimensional fibers (π is proper surjective and X' reduced by definition). Suppose X is Kähler. Then X' is weakly Kähler.*

If moreover there is a discrete $D' \subset X'$ such that for any $x' \in X' \setminus D'$, either

(i) *X' is weakly normal at x' or*

(ii) *$\pi^{-1}(x')$ admits in X a smoothly embeddable neighborhood*

then X' is Kähler.

Proof. With the notations of Theorem 2, set

$$\begin{aligned}
V'_\alpha &:= \{x' \in X' \mid \pi^{-1}(x') \subset U_\alpha\} \\
V_\alpha &:= \pi^{-1}(V'_\alpha) \\
V'^j_{\alpha\beta} &:= \{x' \in X' \mid \pi^{-1}(x') \subset U^j_{\alpha\beta}\} \\
V^j_{\alpha\beta} &:= \pi^{-1}(V'^j_{\alpha\beta}) \\
\psi_\alpha &:= \pi_*(\chi_\alpha|_{V_\alpha}) \\
g^j_{\alpha\beta} &:= \pi_*(\tau^j_{\alpha\beta}|_{V^j_{\alpha\beta}}) .
\end{aligned}$$

Since π is surjective, the sets V'_α cover X' and, for fixed α, β, the $V'^j_{\alpha\beta}$ cover $V'_\alpha \cap V'_\beta$. By Proposition 3.4.1 of Chap. I, $\psi_\alpha \in SP^0(V'_\alpha)$, $g^j_{\alpha\beta} \in \mathscr{W}(V'^j_{\alpha\beta})$ and, since $(\psi_\alpha - \psi_\beta)|_{V'^j_{\alpha\beta}} = g^j_{\alpha\beta} + \bar{g}^j_{\alpha\beta}$, $\psi_\alpha - \psi_\beta \in WPH(V'_\alpha \cap V'_\beta, \mathbb{R})$. So X' is weakly Kähler. Now if conditions (i) and (ii) are fulfilled, then $g^j_{\alpha\beta}$ is holomorphic on $V'^j_{\alpha\beta} \setminus D'$ and $\psi_\alpha - \psi_\beta$ pluriharmonic on $V'_\alpha \cap V'_\beta \setminus D'$. If we take a refinement (W'_λ) of (V'_α) such that each point of D' belongs at most to one W'_λ, then it is clear that Theorem 1 applies and X' is Kähler.

1.2. Corollary. *Let $\pi: X \to X'$ be a proper open surjective morphism. Suppose X is Kähler and X' normal. Then X' is Kähler.*

Many other consequences may be formulated. For example

1.3. Corollary. *Let $\pi: X \to X'$ be a flat projective morphism. Suppose X is Kähler and X' reduced. Then X' is Kähler.*

Proof. The fibers of a projective morphism have smoothly embeddable neighborhoods by construction of $\mathbb{P}(\mathscr{F})$ for a coherent sheaf $\mathscr{F}$.

1.4. Remark. Conditions (i) and (ii) of Theorem 3 are actually unnecessary. See note 3.6 of Chap. I.

2. The Space of Cycles of a Kähler Space

We use the notations of Chap. I, 3.

2.1. Theorem 4. *Let X be a Kähler space and $m \geqq 0$ an integer. Then the Barlet space $\mathbf{B}_m(X)$ of m-cycles of X is weakly Kähler. Moreover, the open subset $\mathbf{B}_m(X)^{(0)}$ of $\mathbf{B}_m(X)$ is Kähler.*

Proof. By an argument similar to the above, set

$$W_\alpha := \{c \in \mathbf{B}_m(X) \mid |c| \subset U_\alpha\}$$

$$W^j_{\alpha\beta} := \{c \in \mathbf{B}_m(X) \mid |c| \subset U^j_{\alpha\beta}\}$$

$$\Phi_\alpha := F_{\chi_\alpha}, \qquad G^j_{\alpha\beta} := F_{\tau^j_{\alpha\beta}},$$

Then $\Phi_\alpha \in SP^0(W_\alpha)$, $G^j_{\alpha\beta}$ is weakly holomorphic on $W^j_{\alpha\beta}$ and holomorphic on $W^j_{\alpha\beta} \cap \mathbf{B}_m(X)^{(0)}$, $(\Phi_\alpha - \Phi_\beta)|_{W^j_{\alpha\beta}} = G^j_{\alpha\beta} + \bar{G}^j_{\alpha\beta}$ and the result follows.

2.2. Corollary. *Let X be a Kähler space. Then the weak normalization of $\mathbf{B}_m(X)$ is Kähler.*

Proof. By a well-known result [5, 12, 18] every connected component of $\mathbf{B}_m(X)$ is compact and, by Theorem 4 above, weakly Kähler. The result follows from Proposition 4.2.4 of Chap. II.

3. Fujiki's Class $\mathscr{C}$

3.1. Definition (Fujiki [12]). A reduced compact complex space X is said to belong to class $\mathscr{C}$ if it is a holomorphic image of a compact Kähler space.

By Hironaka's resolution of singularities it is sufficient to take holomorphic images of compact Kähler manifolds.

Let us define for the moment the class $\mathscr{C}^*$ of reduced compact spaces bimeromorphically equivalent to compact Kähler manifolds, i.e. admitting compact Kähler modifications.

It is then true that $\mathscr{C}$ is stable under holomorphic images and subspaces; but it seems difficult to prove, for example, that a reduced subspace of a space in $\mathscr{C}^*$ is in $\mathscr{C}^*$. Of course, $\mathscr{C}^* \subset \mathscr{C}$.

On the other hand, several important results are valid for compact manifolds in $\mathscr{C}^*$. For example:

(i) If X is a manifold in $\mathscr{C}^*$ and $H^0(X, \Omega_X^2)=0$ then X is Moišezon [14].

(ii) If X is a manifold in $\mathscr{C}^*$, $n=\dim X$ and $\pi: X \to S$ a surjective morphism of X on a complex space S, then $R^q\pi_*(\Omega_X^n)=0$ for all $q > \dim X - \dim S$ (Takegoshi [22]). It seems difficult to prove such results with the hypothesis $X \in \mathscr{C}$. But we have

3.2. Theorem 5. $\mathscr{C}=\mathscr{C}^*$.

Proof. Let X be a compact complex space in $\mathscr{C}$. By definition there is a compact Kähler space X_1 and a surjective morphism $\varrho: X_1 \to X$. By Hironaka's flattening theorem [16], there is a commutative diagram

$$\begin{array}{ccc} X_1 & \xleftarrow{\sigma_1} & Y_1 \\ {\scriptstyle\varrho}\downarrow & & \downarrow{\scriptstyle\pi} \\ X & \xleftarrow{\sigma} & Y, \end{array}$$

where σ, σ_1 are projective modifications and π is flat. Since σ_1 is a Kähler morphism and X_1 a compact Kähler space, Y_1 is Kähler. Moreover Y can be chosen to be normal, since flatness is preserved by base-change. If we apply Corollary 1.2 to $\pi: Y_1 \to Y$, then we deduce that Y is Kähler and $X \in \mathscr{C}^*$ as required.

References

1. Andreotti, A., Grauert, H.: Théorèmes de finitude pour la cohomologie des espaces complexes. Bull. Soc. Math. France **90**, 193–259 (1962)
2. Andreotti, A., Norguet, F.: La convexité holomorphe dans l'espace analytique des cycles d'une variété algébrique. Ann. Sc. Norm. Super. Pisa Cl. Sci., IV. Ser. **21**, 31–82 (1967)
3. Barlet, D.: Espace analytique réduit des cycles analytiques complexes compacts d'un espace analytique complexe de dimension finie – Séminaire F. Norguet. (Lecture Notes Mathematics, Vol. 482, pp. 1–158.) Berlin Heidelberg New York: Springer 1975
4. Barlet, D.: Familles analytiques de cycles et classes fondamentales relatives – Séminaire F. Norguet. (Lecture Notes Mathematics, Vol. 807, pp. 1–24.) Berlin Heidelberg, New York: Springer 1977–79
5. Barlet, D.: Convexité de l'espace des cycles. Bull. Soc. Math. France **106**, 373–397 (1978)
6. Barlet, D.: Convexité au voisinage d'un cycle – Séminaire F. Norguet. (Lecture Notes Mathematics, Vol. 807, pp. 102–121.) Berlin Heidelberg New York: Springer 1977–79
7. Bigolin, B.: Osservazioni sulla coomologia del $\partial\bar{\partial}$. Ann. Sc. Norm. Super. Pisa, Cl. Sci., IV. Ser. **24**, 571–583 (1970)
8. Bingener, J.: Deformations of Kähler spaces. I. Math. Z. **182**, 505–535 (1983); II. Arch. Math. **41**, 517–530 (1983)
9. Douady, A.: Le problème des modules pour les sous-espaces analytiques complexes d'un espace analytique donné. Ann. Inst. Fourier, Grenoble **16**, 1–95 (1966)
10. Fornaess, J.E., Narasimhan, R.: The Levi problem on complex spaces with singularities. Math. Ann. **248**, 47–72 (1980)
11. Frisch, J.: Points de platitude d'un morphisme d'espaces analytiques complexes. Invent. Math. **4**, 118–138 (1967)
12. Fujiki, A.: Closedness of the Douady space of compact Kähler spaces. Pub. Res. Inst. Math. Sci. Kyoto **14**, 1–52 (1978)
13. Fujiki, A.: Kählerian normal complex surfaces. Tohôku. Math. J. **35**, 101–117 (1983)

14. Fujiki, A.: On a complex manifold in $\mathscr{C}$ without holomorphic 2-forms. Publ. Res. Inst. Math. Sci. Kyoto **19**, 193–202 (1983)
15. Grauert, H.: Über Modifikationen und exzeptionelle analytische Mengen. Math. Ann. **146**, 331–368 (1962)
16. Hironaka, H.: Flattening theorem in complex-analytic geometry. Am. J. Math. **97**, 503–547 (1975)
17. Hironaka, H.: Fundamental problems on Douady spaces. Report at the Symposium at Kinosaki (1977) pp. 253–261
18. Lieberman, D.: Closedness of the Chow scheme – Séminaire F. Norguet (Lecture Notes Mathematics, Vol. 670, pp. 140–185.) Berlin Heidelberg New York: Springer 1976
19. Malgrange, B.: Ideals of differentiable functions. Oxford University Press, 1966
20. Moišezon, B.: Singular Kählerian spaces – Proceedings of the international conference on manifolds, Tokyo (1973) pp. 343–351
21. Richberg, R.: Stetige streng pseudokonvexe Funktionen. Math. Ann. **175**, 257–286 (1968)
22. Takegoshi, K.: Relative vanishing theorems in analytic spaces. Duke Math. J. **52**, 273–279 (1985)
23. Varouchas, J.: Stabilité de la classe des variétés kählériennes par certains morphismes propres. Invent. Math. **77**, 117–127 (1984)
24. Varouchas, J.: Sur l'image d'une variété kählérienne compacte – Fonctions de plusieurs variables complexes *V*. (Lecture Notes Mathematics, Vol. 1188, pp. 245–259.) Berlin Heidelberg New York: Springer 1985
25. Varouchas, J., Barlet, D.: Fonctions holomorphes sur l'espace des cycles (to appear)

Received December 1, 1987

Math. Ann. 283, 53–64 (1989)

Mathematische Annalen

Schichten von Matrizen sind rationale Varietäten

Klaus Bongartz

Gesamthochschule Wuppertal, Fachbereich 7 Mathematik, Gaußstrasse 20,
D-5600 Wuppertal 1, Bundesrepublik Deutschland

Einleitung

Operiert eine algebraische Gruppe auf einer Varietät, so bilden die Bahnen einer festen Dimension jeweils eine lokal-abgeschlossene Menge, deren irreduzible Komponenten man *Schichten* nennt [1, 2].

Von besonderem Interesse ist das Studium der Schichten halbeinfacher oder reduktiver Lie-Algebren unter der adjungierten Operation der entsprechenden Gruppen (siehe etwa [1] und die dort angegebene umfangreiche Literatur).

Hier betrachten wir nur den Fall der vollen linearen Gruppe GL_n, die via Konjugation auf der Menge gl_n aller Matrizen operiert. Dieser Fall ist so wichtig und besitzt eine so vollständige und elegante Lösung, daß wir über die von uns erzielten neuen Ergebnisse hinaus auch ältere, zum Teil schwer zugängliche Resultate mit in die Arbeit aufgenommen haben. Dabei sind vor allem einige zentrale Teile aus Petersons reichhaltiger Thesis zu nennen, deren Studium auch den Ausgangspunkt zu dieser Arbeit bildete. Um den Kreis der möglichen Leser nicht unnötig einzuschränken, benutzen wir keinerlei Lie-Theorie, sondern nur fundamentale Ergebnisse aus der algebraischen Geometrie und der linearen Algebra, insbesondere aus der Theorie der Elementarteiler. Dieser Standpunkt wird durch folgende, auf Peterson und Ringel (unveröffentlicht) zurückgehende Beschreibung der Schichten von gl_n ermöglicht:

Zu einer *Partition* $p=(p_1, p_2, \ldots, p_r)$ von n, d.h. einer Folge natürlicher Zahlen $p_1 \geqq p_2 \geqq \ldots \geqq p_r \geqq 1$ mit $p_1+p_2+\ldots+p_r=n$, betrachtet man die Menge $S(p)$ aller Matrizen, deren Elementarteiler $e_1, e_2, \ldots, e_r$ der Bedingung Grad $e_i=p_i$ für $1 \leqq i \leqq r$ genügen. Die Abbildung $p \mapsto S(p)$ ist eine Bijektion zwischen den Partitionen von n und den Schichten von gl_n.

Wir geben im zweiten Paragraphen einen Beweis für dieses Resultat, der sich an Petersons Vorgehen in [5] orientiert. Zuvor rekapitulieren wir im ersten Abschnitt die Elementarteilertheorie der Matrizen, soweit wir sie benötigen, und vertiefen danach im Kernstück der Arbeit einige Aspekte der Theorie. Dies ermöglicht uns dann im dritten Teil kurze Beweise für die Rationalität und Glattheit der Schichten, wobei letztere in Charakteristik 0 bereits auf völlig verschiedenem Weg von

Peterson in [5] sowie Kraft und Luna (unveröffentlicht) gezeigt worden war. Im letzten Paragraphen geben wir zu jeder Partition $p=(p_1, p_2, \dots, p_r)$ einen affinen transversalen Querschnitt $Q(p)$ in $S(p)$ an. Das soll bedeuten, daß $Q(p)$ ein affiner Teilraum von $S(p)$ ist, der jede GL_n-Bahn auf $S(p)$ genau einmal trifft, und zwar so, daß sich in keinem Punkt von $Q(p)$ die Tangentialräume an $Q(p)$ und die Bahn des Punktes echt schneiden. Im Gegensatz zu Petersons Konstruktion aus [5], die in der Einleitung von [1] leicht verständlich dargestellt ist, funktioniert unser Verfahren auch in positiver Charakteristik. Wir wollen es im folgenden kurz beschreiben:

Die Matrizen aus $Q(p)$ bestehen aus r quadratischen Diagonalblöcken $D_1, D_2, \dots, D_r$ jeweils vom Format p_i. Dabei ist D_r die Begleitmatrix (siehe 1.4 für die Definition) eines beliebigen normierten Polynoms vom Grad p_r. Die restlichen D_i's werden rekursiv nach folgender einfacher Vorschrift gebildet. Ist $q_i := p_i - p_{i+1} = 0$, so ist $D_i = D_{i+1}$. Im anderen Fall nimmt man die Begleitmatrix E eines beliebigen normierten Polynoms vom Grad q_i und setzt

$$D_i = \left[\begin{array}{c|c} E & \\ \hline \; 1 & D_{i+1} \end{array}\right].$$

Für die Partition (4, 2, 2, 1) zum Beispiel besteht $Q(p)$ gerade aus allen Matrizen der Form

$$\begin{bmatrix} 0 & d & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & c & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & b & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & a & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & b & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & a & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & a & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a \end{bmatrix}$$

mit beliebigen a, b, c, d aus dem Grundkörper.

In der gesamten Arbeit sei k ein algebraisch abgeschlossener Körper beliebiger Charakteristik. Daß wir uns mit GL_n und gl_n beschäftigen, hat gegenüber dem üblicherweise betrachteten Fall der Matrizen mit Determinante 1 und Spur 0 lediglich einige Vorteile in den Notationen. Natürlich gelten alle unsere Ergebnisse mit den eventuell notwendigen offensichtlichen Modifikationen auch für diesen Fall. Alle topologischen Aussagen beziehen sich stets auf die Zariski-Topologie.

Mein Dank gilt W. Borho, der mir im Anschluß an eine Vorlesung über Schichten die Lektüre von Petersons Arbeit empfahl. Die Rationalität der Schichten ist bei Peterson als offenes Problem formuliert.

1. Elementarteiler

1.1 Jede Matrix $A \in gl_n$ definiert via $X \cdot v = Av$ eine $k[X]$-Modulstruktur M_A auf dem Vektorraum k^n aller Spalten mit n Zeilen. Umgekehrt liefert eine $k[X]$-

Modulstruktur auf k^n eine Matrix, indem man die Darstellungsmatrix der Multiplikation mit X bezüglich der kanonischen Basis $v_1, v_2, \dots, v_n$ von k^n betrachtet. Dabei entspricht die Isomorphie von Moduln der Konjugiertheit der entsprechenden Matrizen. Der aus der linearen Algebra wohlbekannte Elementarteilersatz für Matrizen besagt nun folgendes:

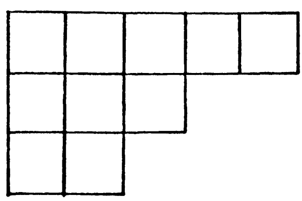
Satz. *Sei* $A \in gl_n$.

a) *Es gibt normierte nicht-konstante Polynome* $e_1, e_2, \dots, e_r$ *in* $k[X]$ *mit* $e_i = r_i e_{i+1}$ *für* $i < r$ *und geeignete* $r_i \in k[X]$, *derart daß* M_A *zur direkten Summe der* $k[X]/(e_i)$ *isomorph ist.*

b) *Falls* M_A *isomorph zu einer direkten Summe zyklischer Moduln der Form* $k[X]/(f_i)$ *mit normierten nicht-konstanten Polynomen* $f_1, f_2, \dots, f_s$ *ist, derart daß* $f_i = s_i f_{i+1}$ *für* $i < s$ *gilt, so folgt* $r = s$ *und* $e_i = f_i$ *für* $1 \leqq i \leqq r$.

Die in dem Satz auftretenden eindeutig bestimmten Polynome $e_i = e_i(A)$ nennt man die *Elementarteiler* von A. Besonders einfach ist die Situation, wenn M_A zyklisch ist. Dann hat nämlich A nur einen Elementarteiler, der zudem mit dem Minimalpolynom und dem charakteristischen Polynom von A übereinstimmt.

1.2. Mit Hilfe der Elementarteiler ordnet man jeder Matrix $A \in gl_n$ eine *Partition* $p(A)$ von n zu, nämlich die Folge der Grade der Elementarteiler. Wie gewöhnlich veranschaulichen wir eine Partition $p = (p_1, p_2, \dots, p_r)$ durch ihr zugehöriges *Young-Diagramm*, indem wir ein Schema zeichnen mit p_1 Kästchen in der ersten Zeile, p_2 in der zweiten, usw. Das Young-Diagramm zur Partition (5, 3, 2) hat also folgende Gestalt.



Durch Vertauschen von Zeilen und Spalten im Young-Diagramm zu p erhält man das Young-Diagramm zur *dualen Partition* $\hat{p}$. So ist z.B. (3, 3, 2, 1, 1) die duale Partition zu (5, 3, 2).

Zu jeder Partition p betrachten wir die Menge $S(p)$ aller Matrizen mit $p(A) = p$. Wie bereits bemerkt, erhalten wir auf diese Art und Weise gerade alle Schichten von gl_n. Hier überlegen wir uns zunächst nur, daß alle Matrizen aus $S(p)$ Bahnen gleicher Dimension liefern. Dazu ordnen wir einer Partition $p = (p_1, p_2, \dots, p_r)$ die Zahl $p^2 = p_2^1 + p_2^2 + \dots + p_r^2$ zu.

Lemma. *Sei* p *eine Partition. Dann gilt für alle* $A \in S(p)$ *die Gleichung* $\dim \operatorname{End} M_A = \hat{p}^2$.

Beweis. Sei $M_A \xrightarrow{\sim} \bigoplus_{i=1}^{r} k[X]/(e_i)$, also $\operatorname{End} M_A$ isomorph zu $\bigoplus_{i,j} \operatorname{Hom}(k[X]/(e_i), k[X]/(e_j))$. Wegen der für alle normierten Polynome g, h gültigen Vektorraumisomorphismen $\operatorname{Hom}(k[X]/(gh), k[X]/(g)) \xrightarrow{\sim} k[X]/(g) \xrightarrow{\sim} (h)/(gh) \xleftarrow{\sim} \operatorname{Hom}(k[X]/(g), k[X]/(gh))$ folgt aus den Teilbarkeitseigenschaften der Elementarteiler jedenfalls

$\dim \operatorname{End} M_A = \sum_{i,j} \min(p_i, p_j)$. Diese Zahl hängt also nicht von A ab, sondern nur von $p(A)$. Um sie wirklich zu berechnen, betrachten wir p_1 verschiedene Körperelemente $\alpha_1, \alpha_2, \dots, \alpha_{p_1}$ und setzen $e_i = \prod_{j=1}^{p_i} (X - \alpha_j)$. Dem Modul $\bigoplus_{i=1}^{r} k[X]/(e_i)$ entspricht dann eine diagonalisierbare Matrix D aus $S(p)$. Trägt man in die erste Spalte des Young-Diagramms zu p stets α_1 ein, dann in die zweite α_2 usw., so erkennt man $M_D \simeq \bigoplus_{j=1}^{p_1} (k[X]/(X-\alpha_j))^{\hat{p}_j}$.

Nun ist $\dim \operatorname{Hom}(k[X]/(X-\alpha), k[X]/(X-\beta)) = \delta_{\alpha\beta}$. Also folgt $\dim \operatorname{End} M_D = \hat{p}^2$.

1.3. Für die nachfolgende einfache Tatsache konnten wir in der Literatur keinen geeigneten Nachweis finden, so daß wir der Vollständigkeit halber einen Beweis anführen.

Lemma. *Sei $A \in S(p)$ mit $M_A = \bigoplus_{j=1}^{r} k[X]/(e_j) m_j$. Dann gilt für jede natürliche Zahl $i \geqq 1$, daß ein von i Elementen erzeugter Untermodul N höchstens $\sum_{j=1}^{i} p_j$ als Dimension hat. Falls dabei Gleichheit auftritt, so ist N isomorph zu $\bigoplus_{j=1}^{i} k[X]/(e_j)$.*

Beweis. Man führt leicht beide Aussagen auf den Fall zurück, wo A nur einen Eigenwert hat, den man sogar noch als 0 annehmen darf. A ist also nilpotent, und die Elementarteiler haben die einfache Gestalt $e_j = X^{p_j}$.

Für $i=1$ ist die Proposition wahr, weil X^{p_1} das Minimalpolynom von A ist. Im Induktionsschritt unterscheiden wir zwei Fälle.

1. Fall. X^{p_1-1} annuliert N.

Sei s die Anzahl der p_j mit $p_1 = p_j$. Dann liegt N schon in dem von $Xm_1, Xm_2, \dots, Xm_s, m_{s+1}, \dots, m_r$ erzeugten Untermodul von M_A. Per Induktion über $\dim M_A$ gilt also $1 + \dim N \leqq \sum_{j=1}^{i} p_j$. Insbesondere tritt in diesem Fall nicht Gleichheit auf.

2. Fall. X^{p_1-1} annulliert N nicht.

Dann seien Erzeugende $n_1, n_2, \dots, n_i$ von N so gewählt, daß $X^{p_1-1} n_1 \neq 0$. Folglich liegt für ein j mit $p_j = p_1$ die Projektion von n_1 auf den direkten Summanden $k[X]/(e_j)m_j$ nicht in dem von Xm_j erzeugten Untermodul. Ohne Einschränkung sei $j=1$. Dann wird durch die Vorschriften $\alpha m_1 = n_1$ und $\alpha m_j = m_j$ für $2 \leqq j \leqq r$ ein Automorphismus α von M_A definiert. Man betrachte nun das folgende kommutative Diagramm mit exakten Zeilen:

$$\begin{tikzcd}
0 \arrow[r] & k[X]\cdot n_1 \arrow[r, "\varepsilon"] \arrow[d, "\beta"] & N \arrow[r] \arrow[d, "\gamma"] & \bar{N} \arrow[r] \arrow[d, "\delta"] & 0 \\
0 \arrow[r] & k[X]/(e_1)m_1 \arrow[r, "\varphi"] & M_A \arrow[r] & \bigoplus_{j=2}^{r} k[X]/(e_j)m_j \arrow[r] & 0
\end{tikzcd}$$

Dabei sind die horizontalen Abbildungen die kanonischen, während γ die Komposition der Inklusion $N \subset M_A$ mit α^{-1} ist. Per Konstruktion faktorisiert $\gamma\varepsilon$

also in der angegebenen Weise. Ferner ist β bijektiv, also nach dem Schlangenlemma δ injektiv. Da $\bar{N}$ von $i-1$ Elementen erzeugt wird, gilt per Induktion $\dim \bar{N} \leqq \sum_{j=2}^{i} p_j$, also auch $\dim N \leqq \sum_{j=1}^{i} p_j$. Gleichheit erzwingt $\dim \bar{N} = \sum_{j=2}^{i} p_j$, also per Induktion $\bar{N} \xrightarrow{\sim} \bigoplus_{j=2}^{i} k[X]/(e_j)$. Da φ per Konstruktion ein Schnitt und β bijektiv ist, ist auch ε ein Schnitt, d.h. $N \xrightarrow{\sim} \bigoplus_{j=1}^{i} k[X]/(e_j)$.

1.4. Zunächst sei der Leser daran erinnert, daß man zu einem normierten nicht konstanten Polynom $f = X^m + a_{m-1}X^{m-1} + \dots + a_0$ die zugehörige Begleitmatrix $B(f) \in gl_m$ definiert als

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & -a_2 \\ & & \cdot & & & \\ \cdot & & & \cdot & & \cdot \\ & & & \cdot & & \\ \cdot & & & & & \cdot \\ 0 & 0 & 0 & \dots & 1 & -a_{m-1} \end{bmatrix}$$

Durch $B(f)$ wird k^m zu einem zyklischen $k[X]$-Modul, so daß f der Elementarteiler von $B(f)$ ist. Ähnlich definieren wir zu je zwei natürlichen Zahlen s, t und zu jedem Polynom $h = b_s X^s + b_{s-1}X^{s-1} + \dots + b_0$ vom Grad $\leqq s$ die $(s+1)xt$ – Matrix $C(s+1, t, h)$ als

$$\begin{bmatrix} 0 & \dots\dots\dots & 0 & b_0 \\ \cdot & & \cdot & b_1 \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ 0 & \dots\dots\dots & 0 & b_s \end{bmatrix}$$

Statt $C(s+1, t, h)$ schreiben wir im folgenden kurz $C(h)$, wenn das Format der Matrix aus dem Zusammenhang heraus klar ist.

Zu vorgegebener Partition p von n spielen später die drei im folgenden definierten Teilmengen $X(p) \supset Y(p) \supset Z(p)$ von gl_n eine entscheidende Rolle. Dabei gehört eine Matrix A zu $X(p)$, falls es für alle $i, j = 1, 2, \dots, r$ normierte Polynome f_i vom Grad p_i gibt und beliebige Polynome h_{ij} vom Grad $\leqq p_i - 1$, so daß A sich wie folgt in Blöcke zerlegt.

$$\left[\begin{array}{c|c|c|c} B(f_1) & C(h_{12}) & \cdots & C(h_{1r}) \\ \hline C(h_{21}) & B(f_2) & \cdots & C(h_{2r}) \\ \hline \vdots & & \ddots & \vdots \\ \hline C(h_{r1}) & C(h_{r2}) & \cdots & B(f_r) \end{array}\right]$$

Natürlich sind bei einer derartigen Matrix die f_i's und h_{ij}'s eindeutig durch A bestimmt. Für die Partition $p=(4,3,1)$ besteht also $X(p)$ gerade aus allen Matrizen der Gestalt

$$\begin{bmatrix} 0 & 0 & 0 & a & 0 & 0 & i & q \\ 1 & 0 & 0 & b & 0 & 0 & j & r \\ 0 & 1 & 0 & c & 0 & 0 & k & s \\ 0 & 0 & 1 & d & 0 & 0 & l & t \\ 0 & 0 & 0 & e & 0 & 0 & m & u \\ 0 & 0 & 0 & f & 1 & 0 & n & v \\ 0 & 0 & 0 & g & 0 & 1 & o & w \\ 0 & 0 & 0 & h & 0 & 0 & p & x \end{bmatrix}$$

mit beliebigen Koeffizienten aus k.

Eine Matrix A aus $X(p)$ gehört zu $Y(p)$, falls folgende drei Bedingungen $Y1$, $Y2$ und $Y3$ erfüllt sind:

$Y1$: Für $i>j$ gilt $h_{ij}=0$

$Y2$: Für $i<r$ ist f_{i+1} ein Teiler von f_i.

$Y3$: Für alle $i<j$ ist f_j ein Teiler von h_{ij}.

Schließlich gehört eine Matrix aus $Y(p)$ zu $Z(p)$, falls alle h_{ij} verschwinden.

Wegen der Blockdiagonalform und der Eigenschaft $Y2$ ist klar, daß eine Matrix A aus $Z(p)$ als Elementarteiler gerade die Polynome $f_1, f_2, \ldots, f_r$ hat, also in $S(p)$ liegt. Offenbar bilden die Matrizen aus $Z(p)$ sogar ein Repräsentantensystem für die Bahnen von GL_n auf $S(p)$.

Während $X(p)$ ein affiner Teilraum von gl_n ist, sind $Y(p)$ und $Z(p)$ nicht durch lineare Gleichungen definiert. Immerhin gilt aber:

Lemma. *$Y(p)$ und $Z(p)$ sind abgeschlossene Teilmengen von gl_n, die zu affinen Räumen isomorph sind.*

Beweis. Statt den Leser durch einen von den Notationen her aufwendigen Beweis zu verwirren, geben wir nur den einfachen Grund für die Korrektheit des Lemmas an.

Man identifiziere die Menge U aller Polynome vom Grad $\leqq s$ mit dem $(s+1)$-dimensionalen affinen Raum $\mathbb{A}^{s+1}$, die Menge V aller normierten Polynome vom Grad $t \leqq s$ mit $\mathbb{A}^t$. Dann ist in $U \times V$ die Menge W aller Paare (f,g), bei denen g ein Teiler von f ist, abgeschlossen und isomorph zu $\mathbb{A}^{s+1}$. Schreibt man nämlich für beliebiges $(f,g) \in U \times V$ nach dem Euklidischen Algorithmus $f=gh+k$ mit Grad $k<t$, so sind die Koeffizienten von h und k polynomial (mit universellen über $\mathbb{Z}$ definierten Polynomen) in den Koeffizienten von f und g. Also ist W abgeschlossen. Identifiziert man auch noch die Menge aller Polynome vom Grad $\leqq s-t$ mit $\mathbb{A}^{s+1-t}$, so liefert $(g,h) \mapsto (gh,g)$ einen bijektiven Morphismus $\mathbb{A}^t \times A^{s+1-t} \to W$, der nach obiger Bemerkung über die Koeffizienten von h sogar einen inversen Morphismus besitzt.

1.5. Wir behalten die in 1.4 eingeführten Bezeichnungen bei.

Proposition. *Sei $p=(p_1, p_2, \ldots, p_r)$ eine Partition von n. Dann gilt $Y(p)=X(p) \cap S(p)$. Ferner hat eine Matrix A aus $Y(p)$ gerade $f_1, f_2, \ldots, f_r$ als Elementarteiler.*

Beweis. Zur Abkürzung setzen wir

$$w_1=v_1, w_2=v_{p_1+1}, \dots, w_r=v_{p_1+p_2+\dots+p_{r-1}+1} \ .$$

Der Nachweis der Inklusion $Y(p)\subset X(p)\cap S(p)$ und des Zusatzes ist einfach. Sei dazu $A\in Y(p)$ gegeben. Per Konstruktion hat dann der Modul M_A obige w_j's als Erzeugende. Wir definieren neue Erzeugende $\bar{w}_j$'s durch $\bar{w}_1=w_1$ und durch $\bar{w}_j=w_j-\sum_{i=1}^{j-1} r_{ij}w_i$ für $j\geqq 2$. Dabei sind die r_{ij} durch die nach $Y3$ gültigen Gleichungen $h_{ij}=r_{ij}f_j$ gegeben. Eine kurze Rechnung, bei der auch die Eigenschaft $Y1$ benutzt wird, zeigt nun, daß $f_j\bar{w}_j=0$ für $1\leqq j\leqq r$ gilt. Deshalb ist M_A Faktormodul von $\bigoplus_{i=1}^{r} k[X]/(f_i)$. Aus Dimensionsgründen gilt sogar $M_A\simeq \bigoplus_{i=1}^{r} k[X]/(f_i)$, und wegen der Bedingungen aus $Y2$ sind die f_i's dann die Elementarteiler von A. Somit liegt A in $S(p)$.

Umgekehrt sei $A\in X(p)\cap S(p)$ mit Elementarteilern $e_1, e_2, \dots, e_r$ gegeben. Für jedes i bezeichne N_i den von $w_1, w_2, \dots, w_i$ erzeugten Untermodul von M_A. Wir beweisen nacheinander, daß A die Bedingungen $Y1$, $Y2$ und $Y3$ erfüllt.

A erfüllt Y1: Nach 1.3 gilt nämlich $\dim N_i\leqq \sum_{j=1}^{i} p_j$. Indem man spaltenweise von links nach rechts vorgeht, entnimmt man nun dem Aussehen der Matrizen in $X(p)$, daß h_{ij} für alle $i>j$ verschwindet. Also ist A eine obere Blockdreiecksmatrix, und man hat für jedes $1<i<r$ eine exakte Sequenz der folgenden Form:

$$(*) \qquad 0\to N_{i-1}\to N_i\to R_i=k[X]/(f_i)\to 0 \ .$$

A erfüllt Y2: Zunächst liest man an der Form von A die Gleichung $\dim N_i=\sum_{j=1}^{i} p_j$ für alle i ab. Nach 1.3 ist also N_i isomorph zu $\bigoplus_{j=1}^{i} k[X]/(e_j)$. Wir zeigen nun per Induktion nach i, daß $e_i=f_i$ gilt, woraus sofort $Y2$ folgt.

Der Induktionsanfang ist wegen $N_1 \xrightarrow{\sim} k[X]/(f_1)$ klar. Zum Induktionsschritt berechnen wir das charakteristische Polynom g der zu N_i gehörenden Matrix (siehe 1.1). Aus $N_i \xrightarrow{\sim} \bigoplus_{j=1}^{i} k[X]/(e_j)$ folgt $g=e_1e_2\dots e_i$. Andererseits zeigt die exakte Sequenz $(*)$ $g=e_1e_2\dots e_{i-1}f_i$, was $e_i=f_i$ impliziert.

A erfüllt Y3: Wir haben schon $N_i \xrightarrow{\sim} \bigoplus_{j=1}^{i} k[X]/(f_i)$ für alle i bewiesen. Dies erzwingt, daß die exakte Sequenz $(*)$ für jedes i spaltet. Aus Dimensionsgründen ist nämlich die induzierte Sequenz

$$0\to \operatorname{Hom}(R_i, N_{i-1})\to \operatorname{Hom}(R_i, N_i)\to \operatorname{Hom}(R_i, R_i)\to 0$$

ebenfalls exakt. Somit besitzt die Projektion $\pi: N_i\to R_i$ einen Schnitt. In der Faser $\pi^{-1}(\bar{1})=\{w_i+w \mid w\in N_{i-1}\}$ des kanonischen Erzeugenden $\bar{1}$ von R_i gibt es also ein Element w_i+w, das von f_i annulliert wird. Also gilt für alle $2\leqq i\leqq r$ die Beziehung:

$$f_iw_i=\sum_{j=1}^{i-1} h_{ji}w_j\in f_iN_{i-1} \ .$$

Es genügt also per Induktion nach $i \geqq 2$ folgende Behauptung zu zeigen: Liegt $\sum_{j=1}^{i-1} g_j w_j$ für irgendwelche Polynome g_j in $f_i N_{i-1}$, so teilt f_i alle g_j mit $1 \leqq j \leqq i-1$.

Zum Beweis setzen wir $f_j = r_{j+1} f_{j+1}$ für $1 \leqq j < r$, was wegen $Y2$ möglich ist. Sei also zuerst $i=2$. Dann ist also $g_1 w_1 \in f_2 N_1$ vorausgesetzt. Multiplikation mit r_2 liefert $r_2 g_1 w_1 \in r_2 f_2 N_1 = f_1 N_1 = 0$. Also liegt $r_2 g_1$ im Annulator von N_1, d.h. $f_1 = r_2 f_2$ teilt $r_2 g_1$, d.h. f_2 teilt g_1.

Im Induktionsschritt ist $\sum_{j=1}^{i-1} g_j w_j \in f_i N_{i-1}$ vorausgesetzt. Multiplikation mit r_i liefert

$$\sum_{j=1}^{i-1} r_i g_j w_j \in r_i f_i N_{i-1} = f_{i-1} N_{i-1} \subseteq f_{i-1} N_{i-2} ,$$

wobei die Inklusion wegen der soeben hergleiteten Beziehung $f_{i-1} w_{i-1} \in f_{i-1} N_{i-2}$ gilt. Folglich annuliert $r_i g_{i-1}$ den vom Bild von w_{i-1} erzeugten Faktormodul N_{i-1}/N_{i-2}, dessen Annulator von f_{i-1} erzeugt wird. Somit ist $r_i f_i$ ein Teiler von $r_i g_{i-1}$, d.h. f_i ein Teiler von g_{i-1}. Ferner ergibt sich nun $\sum_{j=1}^{i-2} r_i g_j w_j \in f_{i-1} N_{i-2}$ erneut wegen $f_{i-1} w_{i-1} \in f_{i-1} N_{i-2}$. Per Induktion teilt also $f_{i-1} = r_i f_i$ alle $r_i g_j$ mit $1 \leqq j \leqq i-2$, d.h. f_i teilt alle g_j mit $1 \leqq j \leqq i-1$.

Der Beweis der Proposition ist damit beendet.

2. Die Schichten von gl_n

2.1. Zu einer Partition p und einer natürlichen Zahl $t \geqq 1$ definieren wir $s(t,p)$ als die Anzahl der Kästchen in den ersten t Spalten des Young-Diagramms zu p. Weiter sei Z_t die Menge der Paare (A,v) aus $gl_n \times k^n$, derart daß die Dimension des von v erzeugten Untermoduls von M_A höchstens t ist. Dies bedeutet gerade, daß der Rang der von den Spalten $v, Av, \ldots, A^{n-1}v$ gebildeten Matrix $\leqq t$ ist, so daß Z_t abgeschlossen in $gl_n \times k^n$ ist. Wir betrachten nun die kanonische Projektion $\pi_t : Z_t \to gl_n$ sowie den Nullschnitt $\sigma_t : gl_n \to Z_t$ mit $\pi_t(A,v) = A$ und $\sigma_t A = (A,0)$.

Lemma. *Für* $A \in S(p)$ *gilt* $\dim \pi_t^{-1}(A) = s(t,p)$.

Beweis. Sei $A \in S(p)$ mit $M_A = \bigoplus_{i=1}^{r} k[X]/(f_i)$ vorgelegt. Für $m = (m_1, m_2, \ldots, m_r)$ mit $m_i \in k[X]/(f_i)$ ist der normierte Erzeuger h_i des Annulators von m_i jeweils ein Teiler von f_i. Ferner wird der Annulator von m gerade vom kleinsten gemeinsamen Vielfachen der h_i's erzeugt. Also gilt

$$\pi_t^{-1}(A) = \cup \bigoplus_{i=1}^{r} \operatorname{Kern} h_i ,$$

wobei $\operatorname{Kern} h_i$ als Untermodul von $k[X]/(f_i)$ aufzufassen ist, und die Vereinigung sich über all diejenigen Tupel $h = (h_1, h_2, \ldots, h_r)$ von normierten Teilern h_i von f_i erstreckt, deren kleinstes gemeinsames Vielfaches Grad $\leqq t$ hat.

Offenbar gilt $\dim \operatorname{Kern} h_i \leqq \min(t,p_i)$, also auch $\dim \bigoplus_{i=1}^{r} \operatorname{Kern} h_i \leqq s(t,p) = \sum_{i=1}^{r} \min(t,p_i)$ für jedes Tupel h mit obigen Eigenschaften. Da es nur endlich viele derartige Tupel gibt, folgt die Ungleichung $\dim \pi_t^{-1}(A) \leqq s(t,p)$. Umgekehrt genügt es wegen der Irreduzibilität von $\bigoplus_{i=1}^{r} \operatorname{Kern} h_i$ ein Tupel h anzugeben mit $\dim \bigoplus_{i=1}^{r} \operatorname{Kern} h_i = s(t,p)$. Dazu wählt man einfach ein Polynom g so, daß g alle f_i mit $\operatorname{grad} f_i \geqq t$ teilt und von allen f_i mit $\operatorname{grad} f_i \leqq t$ geteilt wird. Wegen der Teilbarkeitsbeziehungen zwischen den f_i's existiert ein derartiges g, und man nimmt dann jeweils für h_i den größten gemeinsamen Teiler von f_i und g.

2.2. Auf der Menge der Partition ist durch "$p \leqq q \Leftrightarrow s(t,p) \leqq s(t,q)$ für alle $t \in \mathbb{N}$" eine Ordnungsrelation definiert. Bezüglich dieser Ordnungsrelation sind zwei Partitionen $p < q$ genau dann benachbart (das heißt $p \lneqq p' \leqq q$ impliziert $p' = q$), wenn das zu q gehörende Young-Diagramm E aus dem zu p gehörenden D dadurch entsteht, daß ein unterstes Kästchen einer Spalte von D an die nächstmögliche links davon stehende Spalte unten angefügt wird (siehe z.B. [5]). Daher hat $p \lneqq q$ auch $\hat{p}^2 \lneqq \hat{q}^2$ zur Folge.

Im folgenden bezeichne gl_n^d für eine natürliche Zahl d die Menge aller Matrizen A, deren Bahn die Dimension $n^2 - d$ hat, d.h. deren Isotropiegruppe H in GL_n die Dimension d hat. Da H gerade die Gruppe der Einheiten in $\operatorname{End} M_A$ ist, gilt nach 1.2 also $gl_n^d = \{A \mid \widehat{p(A)}^2 = d\}$.

Lemma [5]. *Sei p eine Partition mit $d = n^2 - \hat{p}^2$. Dann liegt der Abschluß $\overline{S(p)}$ von $S(p)$ in $\bigcup_{q \geqq p} S(q)$. Insbesondere ist $S(p) = \overline{S(p)} \cap gl_n^d$.*

Beweis. Da $\sigma_t(A)$ für jedes t jede irreduzible Komponente des Kegels $\pi_t^{-1}(A)$ trifft, folgt aus dem Halbstetigkeitssatz von Chevalley ([3]), daß für alle t und l die Mengen $X(t,l) = \{A \in gl_n \mid \dim \pi_t^{-1}(\mathrm{A}) \geqq l\}$ abgeschlossen sind. Nach 2.1 ist nun $S(p)$ in der abgeschlossenen Menge $C = \bigcap_t X(t, s(t,p))$ enthalten. Andererseits gilt nach dem Elementarteilersatz $gl_n = \bigcup_q S(q)$, so daß aus 2.1 auch folgt $C \subseteq \bigcup_{q \geqq p} S(q)$. Wie eben bemerkt, trifft $S(q)$ für $p \lneqq q$ nicht gl_n^d, so daß das Lemma vollständig bewiesen ist.

Im allgemeinen ist $S(p)$ eine echte Teilmenge der nach dem Lemma abgeschlossenen Menge $\bigcup_{q \geqq p} S(q)$. Das einfachste Beispiel tritt für $n = 4$ und die Partition $(2,2)$ auf.

2.3. **Proposition** [5]. *Die Abbildung $p \mapsto S(p)$ liefert eine Bijektion zwischen den Partitionen von n und den Schichten von gl_n.*

Beweis. Als Bild der nach 1.4 irreduziblen Varietät $GL_n \times Z(p)$ ist $S(p)$ irreduzibel. Für jedes d ist nach dem Elementarteilersatz und 1.2 gl_n^d die disjunkte Vereinigung der $S(p)$ mit $\hat{p}^2 = d$, die nach 2.2 abgeschlossen sind in gl_n^d. Daher sind die $S(p)$'s sogar die Zusammenhangskomponenten von gl_n^d. Der Rest ist nun klar.

Bemerkung. Aus dieser expliziten Beschreibung ergeben sich einige interessante unmittelbare Folgerungen für die Schichten von gl_n, die bereits vor Peterson von anderen Mathematikern mit völlig anderen Methoden erhalten worden waren. Es gilt:
- Die Schichten von gl_n sind disjunkt (Dixmier).
- Jede Schicht enthält diagonalisierbare Matrizen (Ozeki-Wakimoto und Tauvel).
- Jede Schicht enthält bis auf Konjugation genau eine nilpotente Matrix (Johnston-Richardson).

Literaturhinweise auf die entsprechenden Arbeiten findet der Leser etwa in [4].

3. Rationalität der Schichten

3.1. Sei im folgenden $p=(p_1, \dots, p_r)$ eine Partition von n. Wir betrachten dann zu einer Matrix $A \in gl_n$ die Matrix $T=T(p, A)$, deren Spalten gerade gebildet werden von $w_1, Aw_1, \dots, A^{p_1-1}w_1, w_2, Aw_2, \dots, A^{p_2-1}w_2, \dots, w_r, Aw_r, \dots, A^{p_r-1}w_r$. Dabei setzen wir zur Abkürzung $w_1=v_1$, $w_2=v_{p_1+1}, \dots, w_r=v_{p_1+p_2+\dots+p_{r-1}+1}$. Offenbar ist die Menge $U(p)$ aller Matrizen A, für die $T(p, A)$ invertierbar ist, eine offene Teilmenge von gl_n. Für $A \in Z(p)$ (siehe 1.4) ist sogar $T(p, A)$ die Einheitsmatrix, so daß $U(p)$ eine offene Umgebung von $Z(p)$ ist. Um diese Umgebung explizit zu beschreiben, benötigen wir neben der in 1.4 eingeführten Menge $X(p)$ noch die Untergruppe $G(p)$ aller Matrizen A aus GL_n, die der Bedingung $Aw_i=w_i$ für $i=1, 2, \dots, r$ genügen.

Lemma. *Die Operation von GL_n auf gl_n via Konjugation induziert einen Isomorphismus Φ zwischen $G(p) \times X(p)$ und $U(p)$.*

Beweis. Sei zunächst $(C, D) \in G(p) \times X(p)$. Man überprüft direkt anhand der Definitionen, daß gilt $T(p, CDC^{-1})=C$. Daher liegt $\Phi(C, D)=CDC^{-1}$ überhaupt in $U(p)$.

Setzt man nun $\psi(A)=(T(p, A), T(p, A)^{-1}AT(p, A))$ für A aus $U(p)$, so liegt $\psi(A)$ nach der bekannten Transformationsformel für Matrizen bei Basiswechsel in $G(p) \times X(p)$. Offenbar sind Φ und ψ zueinander inverse Morphismen.

3.2. Wir behalten die in 3.1 und 1.4 eingeführten Bezeichnungen bei. Aus 3.1 und 1.5 ergibt sich nun unmittelbar das Schlüsselresultat dieser Arbeit.

Satz. *Die Konjugation induziert einen Isomorphismus zwischen $G(p) \times Y(p)$ und $U(p) \cap S(p)$.*

Aus dem Satz erhält man direkt zwei wichtige Folgerungen.

Korollar 1. *Die Schichten von gl_n sind rationale Varietäten.*

Beweis. Offenbar ist $G(p)$ eine offene Menge im affinen Raum $\mathbb{A}^m$ mit $m=n(n-r)$. Die Behauptung folgt also aus dem Satz und aus 1.4.

Korollar 2. *Die Schichten von gl_n sind glatte Varietäten.*

Beweis. Nach dem Satz und 1.4 sind alle Punkte von $U(p) \cap S(p)$ glatt. Da $U(p) \cap S(p)$ offene Umgebung des Repräsentantensystems $Z(p)$ ist, ist $S(p)$ glatt.

4. Querschnitte

4.1. Zunächst ziehen wir noch eine einfache Folgerung aus Satz 3.2. Dabei sei $p=(p_1, p_2, \dots, p_r)$ eine Partition von n.

Lemma. *Die Koeffizienten der Elementarteiler sind reguläre Funktionen auf $S(p)$. Versieht man $Z(p)$ mit der trivialen GL_n-Operation, so liefern sie die Komponenten einer GL_n-äquivarianten Retraktion von $S(p)$ auf $Z(p)$. Insbesondere ist die Menge der Bahnen $S(p)/GL_n$ bezüglich der Quotiententopologie homöomorph zu einem affinen Raum.*

Beweis. Auf $U(p) \cap S(p) \xrightarrow{\sim} G(p) \times Y(p)$ sind die Koeffizienten der Elementarteiler nach 1.5 gerade gewisse Komponentenfunktionen auf $Y(p)$, also reguläre Funktionen. Da $U(p) \cap S(p)$ eine offene Umgebung des Repräsentantensystems $Z(p)$ ist, und da die Koeffizienten der Elementarteiler GL_n-invariante Funktionen sind, sind sie überall regulär. Die übrigen Aussagen sind nun nach 1.4 und 1.5 klar.

Die erste Aussage des Lemmas findet sich bereits bei Peterson, die letzte bei Kraft und Peterson.

4.2. Aus unseren bisherigen Untersuchungen folgt leicht, daß $Z(p)$ ein transversaler Querschnitt im Sinne der zu Beginn vereinbarten Definition ist. Die Transversalität ergibt sich nämlich aus der Existenz der GL_n-äquivarianten Retraktion. Nun ist $Z(p)$ zwar isomorph zu einem affinen Raum, hat aber gegenüber Petersons Querschnitt noch den Nachteil, kein affiner Teilraum von gl_n zu sein. Wir werden daher $Z(p)$ durch einen affinen Teilraum $Q(p)$ ersetzen, der zudem noch ein transversaler Querschnitt ist. Dabei gehen wir ein wenig formaler vor als in der Einleitung.

Sind $f_1, f_2, \dots, f_s$ normierte Polynome vom Grad $\geqq 1$, so definieren wir $D(f_1, f_2, \dots, f_s)$ als die Matrix

$$\begin{bmatrix} B(f_1) & & & \\ 1 & B(f_2) & & \\ & 1 & \ddots & \\ & & 1 & B(f_s) \end{bmatrix}$$

mit Nullen außerhalb der besetzten Stellen. Dabei ist wieder $B(f)$ die Begleitmatrix des Polynoms f (siehe 1.4).

Lemma. *In obiger Situation hat die Matrix $D=D(f_1, f_2, \dots, f_s)$ nur einen Elementarteiler, und zwar $f_1 f_2 \dots f_s$.*

Beweis. Per Konstruktion wird der zu D gehörige $k[X]$-Modul vom ersten kanonischen Basisvektor erzeugt. Deshalb hat D nur einen Elementarteiler, nämlich das charakteristische Polynom g. Wegen der unteren Blockdreiecksform von D ist g das Produkt der charakteristischen Polynome der Diagonalblöcke, d.h. g ist das Produkt der f_i's.

4.3. Sei jetzt $p=(p_1, p_2, \ldots, p_r)$ eine Partition von n. Für $i=1,2,\ldots,r$ sei $V(i)$ der affine Raum der normierten Polynome vom Grad $p_i - p_{i+1}$, wobei $p_{r+1}=0$ gesetzt sei. Dann definieren wir eine injektive Abbildung ψ von $\underset{i=1}{\overset{r}{\times}} V(i)$ nach gl_n durch

$$\psi(g_1, g_2, \ldots, g_r) = \begin{bmatrix} \boxed{D_1} & & & \\ & \boxed{D_2} & & \\ & & \ddots & \\ & & & \boxed{D_r} \end{bmatrix}$$

mit $D_i = D(g_{j_1}, g_{j_2}, \ldots, g_{j_r})$, wobei $(j_1, j_2, \ldots, j_s)$ die Teilfolge von $(i, i+1, \ldots, r)$ derjenigen Indizes j ist, für die $g_j \neq 1$ ist. Das Bild von ψ bezeichnen wir mit $Q(p)$.

Proposition. *$Q(p)$ ist ein affiner Teilraum von gl_n, der jede Bahn von GL_n auf $S(p)$ genau einmal trifft. Versieht man $Q(p)$ mit der trivialen GL_n-Operation, so gibt es eine GL_n-äquivariante Retraktion von $S(p)$ auf $Q(p)$.*

Beweis. Per Konstruktion ist $Q(p)$ ein affiner Teilraum. Ist $A=\psi(g_1, g_2, \ldots, g_r)$ aus $Q(p)$, so sind nach 4.2 und 1.5 die Produkte $\prod_{j=1}^{r} g_j$ für $1 \leqq i \leqq r$ die Elementarteiler von A. Offenbar erhält man so jedes mögliche Tupel $(f_1, f_2, \ldots, f_r)$ von Elementarteilern genau einmal. Schließlich lassen sich aus den Koeffizienten der Elementarteiler die Koeffizienten der g_i's, d.h. die Koeffizienten der Matrizen aus $Q(p)$, polynomial berechnen, was im Beweis des Lemmas 1.4 ausführlich begründet ist. Dies liefert die gewünschte Retraktion.

Literatur

1. Borho, W.: Über Schichten halbeinfacher Lie-Algebren. Invent. Math. **65**, 283–317 (1981)
2. Borho, W., Kraft, H.: Über Bahnen und deren Deformationen bei linearen Aktionen reduktiver Gruppen. Comment. Math. Helv. **54**, 61–104 (1979)
3. Dieudonné, J., Grothendieck, A.: Eléments de géométrie algébrique III, IV. Publ. Math. IHES
4. Kraft, H.: Parametrisierung von Konjugationsklassen in sl_n. Math. Ann. **234**, 209–220 (1978)
5. Peterson, D.: Geometry of the adjoint representation of a complex semisimple Lie algebra. Harvard Thesis 1978

Eingegangen am 10. Dezember 1987

Math. Ann. 283, 65–86 (1989)

Mathematische Annalen

A Probabilistic Proof and Applications of Wiener's Test for the Heat Operator

Kôhei Uchiyama

Department of Mathematics, Hiroshima University, Hiroshima, 730 Japan

1. Introduction

In this paper we study a thinness criterion for the coparabolic operator $\frac{1}{2}\Delta+\partial/\partial t$ in stead of the parabolic (heat) operator $\frac{1}{2}\Delta-\partial/\partial t$. The translation of the results and proofs in one context into those in the other is straightforward. The criterion which is a precise analogue of Wiener's test for the Laplace operator will be stated first in a probabilistic way and then in purely analytic terms.

Let $(\Omega, \mathscr{F}, P)$ be a probability space, $E[\cdot]$ the integration by P, and $W(t)$ $(t\geqq 0)$ a standard N-dimensional Wiener process – the Brownian motion starting at zero standardized by $E[W_i(t)\,W_j(t)]=\delta_{ij}t$ – defined on $(\Omega, \mathscr{F}, P)$. Let A be an analytic set of R^{N+1} and let $h(\xi; A)$, $\xi \in R^{N+1}$ denote the hitting probability of A by the space-time Wiener process starting at $\xi=(s, \mathbf{x})$:

$$h(s, \mathbf{x}; A)=P[(s+t, \mathbf{x}+W(t))\in A \quad \text{for some} \quad t>0]$$

where $s\in R$ and $\mathbf{x}\in R^N$. Here and hereafter, abusing the notation, we write simply $f(s,\mathbf{x})$ for $f((s,\mathbf{x}))$ where f is a function of $\xi\in R^{N+1}$. Let us write $x=|\mathbf{x}|$ (the Euclidean length of $\mathbf{x}$), and define

$$p(t,x)=\begin{cases}(2\pi t)^{-N/2}e^{-x^2/2t} & \text{if} \quad t>0\\ 0 & \text{if} \quad t\leqq 0.\end{cases}$$

The function $g(\xi):=p(t,x)$, $\xi=(t,\mathbf{x})$, is a Green (density) function of the space-time Wiener process $(t, W(t))$ which is a density function of the measure $\mu(A) = \int_0^\infty P[(t, W(t))\in A]\,dt$ relative to the Lebesgue measure of R^{N+1}. By means of the Green function are defined coparabolic balls centered at the origin $O:=(0,\mathbf{0})$, which here we parametrize by $a>0$ as follows

$$B(a):=\{(t,\mathbf{x})\in R^{N+1}: p(t,x)>(2\pi a)^{-N/2}\},$$

so that $B(a)$ becomes large together with a. For A a subset of R^{N+1} we say that A is *hit i.o.* (infinitely often) as $t\uparrow\infty$ $[t\downarrow 0]$ if there exists an increasing [resp. a decreasing] sequence $\{t_n\}_{n=1}^\infty$ such that $t_n\to\infty$ [resp. $t_n\to 0$] as $n\to\infty$ and $(t_n, W(t_n))\in A$ for every n.

Theorem 1. *Let A be an analytic set of $[0,\infty)\times R^N$ and set*

$$A_n = A\cap[B(2^{n+1})\setminus B(2^n)].$$

i) *$P[A$ is hit i.o. as $t\uparrow\infty]=0$ or 1 according as $\sum_{n=0}^{\infty} h(O;A_n)$ converges or diverges;*

ii) *$P[A$ is hit i.o. as $t\downarrow 0]=0$ or 1 according as $\sum_{n=0}^{\infty} h(O;A_{-n})$ converges or diverges.* □

The Green kernel $g(\xi,\eta):=g(\eta-\xi)$ is the fundamental solution for $\frac{1}{2}\Delta+\frac{\partial}{\partial t}$ and this theorem accordingly can be stated in potential theoretic terms. We need the fact that if A is bounded and analytic then there exists a finite Borel measure e_A on R^{N+1} (coparabolic equilibrium measure) such that

$$h(s,\mathbf{x};A)=\int_{R^{N+1}} p(t-s,|\mathbf{y}-\mathbf{x}|)\,e_A(dt\,d\mathbf{y}). \tag{1.1}$$

The (coparabolic) capacity of A is the total measure of e_A, which we denote by $\mathrm{Cap}(A)$: $\mathrm{Cap}(A)=e_A(R^{N+1})$. It is then clear that for A_n in Theorem 1

$$(2\pi 2^{n+1})^{-N/2}\,\mathrm{Cap}(A_n)\leqq h(O;A_n)\leqq(2\pi 2^n)^{-N/2}\,\mathrm{Cap}(A_n), \tag{1.2}$$

so that the convergence conditions of the series in Theorem 1 are equivalent to those of $\sum_{n=0}^{\infty} 2^{\mp Nn/2}C(A_{\pm n})$.

Let A be an analytic set that has the origin O as its cluster point. A is said to be (coparabolic) *thin* at O if there is a supercoparabolic function u defined on an open neighbourhood of O such that each of its cluster values at O along A is greater than $u(O)$ Doob [2, p. 309]. It is known that A is thin at O if and only if $P[A$ is hit i.o. as $t\downarrow 0]=0$ [2, p. 656]. Thus the second half of Theorem 1 is paraphrased in purely analytic terms as follows.

Theorem 2. *An analytic set A of R^{N+1} is thin at the origin if and only if $\sum_{n=0}^{\infty} 2^{Nn/2}\,\mathrm{Cap}(A_{-n})<\infty$.* □

To make a similar rewording of the first half of Theorem 1 we consider the (coparabolic) Martin boundary of the upper half space $t\geqq 0$. Among others there is a minimal Martin boundary point, denoted by ζ_0, corresponding to the limit of $t\to\infty$ and $x/t\to 0$, at which the Martin function is identically one [2, p. 374]. An analytic set A of $[0,\infty)\times R^N$ is (coparabolic) minimal thin at ζ_0 if and only if $P[A$ is hit i.o. as $t\uparrow\infty]=0$ [2, p. 731]. Therefore Theorem 1 yields

Theorem 2′. *At the Martin boundary point ζ_0 an analytic set A of $[0,\infty]\times R^N$ is minimal thin if and only if $\sum_{n=0}^{\infty} 2^{-Nn/2}\,\mathrm{Cap}(A_n)<\infty$.* □

Theorem 2 has been established by Evans and Gariepy [4]. The method for their proof is potential theoretic, whereas our approach is very probabilistic and has the advantage of being relatively simple not only in its formal presentation but also in the guiding idea which is intuitive and easy to grasp.

If A lies inside the paraboloid $x^2=ct$ for some $c>0$ Theorem 1 is almost as worthless (because dispensable) as its proof is uninvolved, whereas it works very efficiently if A_n are outside such paraboloids with $c=c_n$ tending infinity as $|n|$ becomes large. The latter fact is conceptually because for large $|n|$ the hittings of A_n in the latter case are very rare events which are relatively strongly positively-correlated with one another. In practice, however, the applicability of Theorem 1 is greatly due to the explicit and simple expression of the distribution v_a of the place at which $(t, W(t))$ hits the coparabolic sphere $S(a):=\partial B(a)$ [the boundary of $B(a)$, $a>0$]. It is shown by Bauer [1] that the parabolic counterpart of v_a agrees with the measure that Fulks [7] introduced as a measure with which the parabolic function is characterized as having local average property over parabolic spheres [this result implies in particular that $R^{N+1}\setminus B(a)$ is thin at the origin]. In the present context it is convenient to parametrise points of $S(a)$ by the ordinate t and the spherical component θ of the abscissa $\mathbf{x}$ (i.e., $\theta=\mathbf{x}/x$); from v_a this parametrization induces the measure

$$\hat{v}_a(dt\,d\theta)=\tfrac{1}{2}x^N p(t,x)t^{-1}\,dt\,d\theta \qquad (0<t\leqq a, \theta\in\Theta) \tag{1.3}$$

with $\hat{v}(\{0\}\times\Theta)=0$ where $x=\sqrt{-Nt\log t/a}$, $\Theta=\{\theta\in R^N:|\theta|=1\}$ and $d\theta$ is a surface element of $(N-1)$-unit-sphere Θ if $N\geqq 2$ and a discrete measure which charges each point of $\Theta=\{-1,1\}$ with unit mass if $N=1$. [To be precise $\hat{v}_a$ is the probability measure on $[0,a]\times\Theta$ induced from v_a by the mapping $(t,\mathbf{x})\to(t,\mathbf{x}/x)$.] Then a little reflection would convince us that Theorem 1 could verify the Kolmogorov's test. In fact not only this is true, but also we can strengthen the harder half of the test as follows.

Theorem 3. *Let $f=f(t,\theta)$ be a positive Borel function of $t\geqq 0$ and $\theta=\mathbf{x}/x$ and G denote the graph of f: $G=\{(t,\mathbf{x}):f(t,\mathbf{x}/x)=x\}$. If*

$$\int_{\substack{0<t<1\\ [t>1]}} t^{-1}\,dt \int_{\Theta} f(t,\theta)^N p(t,f(t,\theta))\,d\theta=\infty\,,$$

then $P[G$ is hit i.o. as $t\downarrow 0$ [resp. $t\uparrow\infty$]$]=1$ or, what is the same, G is not thin at the origin [resp. minimal thin at ζ_0]. □

The transformation $\psi:(t,\mathbf{x})\to(1/t,\mathbf{x}/t)$ which transforms $W(t)$ into $\tilde{W}(t):=tW(1/t)$ does not change the probability law of W, i.e.,

$$\tilde{W}(t) \text{ also is a standard Wiener process under } (\Omega,\mathcal{F},P) \tag{1.4}$$

[a *projective invariance* of $W(\cdot)$]. Applying Theorem 1 with $\psi(A)$ and $\tilde{W}(t)$ in place of A and $W(t)$, we accordingly get that $P[A$ is hit i.o. as $t\downarrow 0$ $[t\uparrow\infty]]=0$ or 1 according as $\sum_{n=0}^{\infty[-\infty]} h(O;A_n')$ converges or diverges where

$$A_n'=\psi(\psi(A)_n)=\{(t,\mathbf{x})\in A:(2\pi 2^{n+1})^{-N/2}<p(1/t,\mathbf{x}/t)\leqq(2\pi 2^n)^{-N/2}\}\,.$$

Note that, A_n' being quite different from A_n, the criterion thus obtained is in appearance not the same as that in Theorem 1.

To make our probabilistic approach self-contained we shall in Appendix give direct probabilistic proofs of (1.1, 1.3). The proof of Theorem 1 will be given in Sect. 2. In Sect. 4 several applications of Theorem 1 will be made to obtain easily computable criteria for the thinness. In particular the Kolmogorov's test and Theorem 3 will be deduced. In Sect. 3 Theorem 1 will be refined in a way as a preparation for the proof of Theorem 3.

2. Proof of Theorem 1

We shall prove only the first half of Theorem 1. The proof for the other half is similar (though not completely parallel). In the rest of the paper n will denote a non-negative integer unless the contrary is explicitly stated. We shall assume that A lies above the ordinate level $t=1$, i.e., $A \subset [1,\infty)\times R^N$. The case $\sum h(O;A_n)<\infty$ is trivially disposed of by the Borel-Cantelli lemma. Since $P[|W(t)|>t$ i.o. as $t\uparrow\infty]=0$ [consider $\tilde{W}(t)=tW(1/t)$], it therefore suffices to show that

$$\text{if } \sum_{n=0}^{\infty} h(O,A_n)=\infty, \quad \text{then} \quad P[A_n \text{ is hit for i.m. } n]=1. \tag{2.1}$$

Here the expression under P means that there exist infinitely many n s.t. $(t, W(t))\in A_n$ for some $t>1$.

The geometry of coparabolic balls $B(a)$ or spheres $S(a):=\partial B(a)$ is fundamental in our proof. Put

$$X_a(t)=\sqrt{-Nt\log t/a}, \qquad 0\leqq t\leqq a.$$

Then $(t,\mathbf{x})\in B(a)$ if and only if $x:=|\mathbf{x}|<X_a(t)$. We see also

$$\frac{d}{dt}X_a(t)=-\frac{N}{2}\,\frac{1+\log t/a}{X_a(t)}. \tag{2.2}$$

Therefore $B(a)$ is an egg-shaped body with the top at $(a,\mathbf{0})$ and its bottom (the center) at the origin $O=(0,\mathbf{0})$, and broadest in the abscissa direction at the ordinate level $t=a/e$. We make another observation of interest: for each $\varepsilon>0$ the mapping $\varphi^{\varepsilon}:(t,\mathbf{x})\to(\varepsilon t,\sqrt{\varepsilon}\,\mathbf{x})$, which does not change the law of the Wiener process [the *scaling invariance*, i.e., the process $W(\varepsilon t)/\sqrt{\varepsilon}$, $t\geqq 0$ has the same distribution as $W(t)$, $t\geqq 0$], transforms $B(a)$ into $B(\varepsilon a)$, i.e., $\varphi^{\varepsilon}(B(a))=B(\varepsilon a)$; in particular

$$h(\varepsilon s,\sqrt{\varepsilon}\,\mathbf{x};\varphi^{\varepsilon}(A_n))=h(s,x;A_n) \quad \text{and} \quad \varphi^{2^{-n}}(A_{n+k})\subset B(2^{k+1})\backslash B(2^k), \tag{2.3}$$

where A_n is supposed to be a subset of $B(2^{n+1})\backslash B(2^n)$. The following inequality immediate from (1.1) and being a trivial extension of (1.2) also plays a key role in our proof: for each $(s,\mathbf{x})\in R^{N+1}$ and each analytic set A of $B(a)\,(a>0)$

$$h(s,\mathbf{x};A)\leqq(2\pi a)^{N/2}\,h(O,A)\sup_{(t,\mathbf{y})\in A} p(t-s,|\mathbf{y}-\mathbf{x}|). \tag{2.4}$$

For later (i.e., beyond the proof of Theorem 1) as well as present needs here is noted that if $s<1/2$, then

$$\Delta x:=X_2(s)-X_1(s)=\frac{sN\log 2}{X_2(s)+X_1(s)}=\frac{\log 2}{2}\cdot\frac{\sqrt{sN}(1+\varepsilon(s))}{\sqrt{-\log s/2}} \tag{2.5}$$

with $0 \leqq \varepsilon(s) \leqq \text{const}(-\log s)^{-1}$: this in particular shows that Δx is much smaller than $X_1(s)$ if s is small.

The next lemma is a result of a simple application of the inequality (2.4).

Lemma 4. *Let $c, s, \delta > 0$, $0 < a < a'$ and A an analytic subset of $D \cap [B(a') \setminus B(a)]$ where $D := \{(t, \mathbf{y}) \in R^{N+1} : t > (1+\delta)s, y > t\}$. Then it follows that for all* $\mathbf{x}$ $h(s, \mathbf{x}; A) \leqq e^x (1 + 1/\delta) a'/a]^{N/2} h(O; A)$. □

Proof. The inequality (2.4) reduces the problem to proving that for $a \leqq b \leqq a'$

$$\sup_{(t,\mathbf{y}) \in D \cap S(b)} p(t-s, |\mathbf{y}-\mathbf{x}|) \leqq e^x \left[1 + \frac{1}{\delta}\right]^{N/2} (2\pi b)^{-N/2}. \tag{2.6}$$

To evaluate the supremum above we express p as follows

$$p(t-s, |\mathbf{y}-\mathbf{x}|) = (2\pi b)^{-N/2} \left(1 - \frac{s}{t}\right)^{-N/2} e^{H(t,\mathbf{y})} \quad \text{for} \quad t > s, (t, \mathbf{y}) \in S(b), \tag{2.7}$$

where $H(t, \mathbf{y}) = \frac{1}{2}(y^2/t - |\mathbf{y}-\mathbf{x}|^2/(t-s))$. Then (2.6) is ready from

$$H(t, \mathbf{y}) < \frac{1}{2t}\{y^2 - |\mathbf{y}-\mathbf{x}|^2\} \leqq \frac{xy}{t} - \frac{x^2}{2t}.$$

The main task for the proof of Theorem 1 is performed by proving

Lemma 5. *There exists two constants $\gamma < 1$ and $C > 0$ such that if $\{A_k\}$ is a sequence of analytic sets the k-th member of which is contained in $B(2^{k+1}) \setminus B(2^k)$, then for $\xi \in \overline{B(1)}$* [the closure of $B(1)$] *and $m = 1, 2, \ldots$*

$$h\left(\xi; \bigcup_{k=1}^{m} A_k\right) \leqq \gamma + C \sum_{k=1}^{m} h(O; A_k). \quad \square \tag{2.8}$$

Proof. Step 1. As the first step we prove that

for each $c > 0$ there exists constants $\gamma < 1$ and C depending only on c and N such that (2.8) *holds for $\xi \in K_c := \{(t, \mathbf{y}) \in \overline{B(1)} : y^2 \leqq ct\}$.* (2.9)

[It may be noted that $S(1)$ and the paraboloid $y^2 = ct$ intersect at an exactly one level of ordinate $t = e^{-N/c}$.] Let $F = \{(t, \mathbf{y}) : y > t > 3/2 \text{ or } t \leqq 3/2\} \setminus B(2)$. Then there exists $\gamma < 1$ such that $\sup^* h(\xi; F) \leqq \gamma$, where the sup* denotes the supremum taken over all $\xi \in K_c \cap S(1)$. On the other hand by Lemma 4 $\sup^* h\left(\xi; \bigcup_{k=1}^{m} A_k \setminus F\right)$ $\leqq C \sum_{k=1}^{m} h(O; A_k)$. Thus (2.8) holds for $\xi \in K_c \cap S(1)$. As for $\xi \in K_c \setminus S(1)$ we have $\varphi^\varepsilon(\xi) \in K_c \cap S(1)$ with some $\varepsilon > 1$ and in view of the scaling relation (2.3) the same argument as above proves (2.8) with a smaller γ and the same C, proving (2.9).

Step 2. Owing to the strong Markov property of $W(\cdot)$ we have only to show (2.8) for $\xi \in S(1)$. By (2.9) it therefore suffices to prove that there is a (large) constant c such that if $\xi = (s, \mathbf{x})$ satisfies

$$(s, \mathbf{x}) \in S(1) \quad (\text{i.e., } x^2 = -Ns\log s) \quad \text{and} \quad x^2 > cs, \tag{2.10}$$

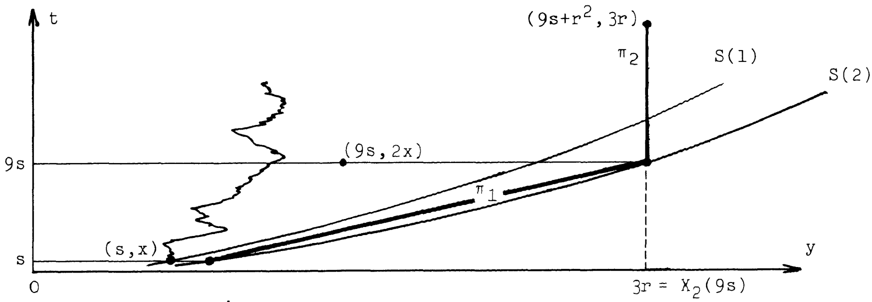


Fig. 1

then (2.8) holds. Let (2.10) hold. Consider two fences in R^{N+1}, say π_1 and π_2, obtained by rotating around the t-axis two line segments lying on the (t, x_1)-plane (x_1 is the first coordinate of R^N-variable $\mathbf{x}$) whose end points have coordinates $(s, X_2(s))$ and $(9s, X_2(9s))$ for π_1; $(9s, X_2(9s))$ and $(9s+[X_2(9s)/3]^2, X_2(9s))$ for π_2, i.e.,

$$\pi_1 := \left\{(t, \mathbf{y}) : \frac{y - X_2(s)}{t-s} = \frac{X_2(9s) - X_2(s)}{8s},\ s \leqq t \leqq 9s\right\}$$

$$\pi_2 := \{(t, \mathbf{y}) : y = 3r, 9s < t < 9s + r^2\},$$

where $r := X_2(9s)/3$. From (2.9) it follows that $-\log s \geqq \frac{c}{N}$, and then we see for large c

$$X_2(9s) = \sqrt{9}\left\{\frac{-\log 9s/2}{-\log s/2}\right\}^{1/2} \cdot X_2(s) \geqq 3X_2(s), \tag{2.11}$$

which inequality we assume to hold in below. Clearly π_1 lies inside $S(2)$. We can suppose that π_2 also lies inside $S(2)$. In fact, since $X_2(t)$ is increasing up to $t = 2/e$, it suffices to have $9s + X_2(9s)^2/9 \leqq 2/e$, which is clearly possible by taking c large enough. Now setting

$$J_1 = P[\pi_1 \cup \pi_2 \text{ is not hit at all} \mid W(s) = \mathbf{x}]$$

$$J_2 = \sup_{|\mathbf{z}| \leqq 3r} h\left(9s + r^2, \mathbf{z}; \bigcup_{k=1}^{m} A_k\right)$$

and applying the strong Markov property of $W(\cdot)$, we have $h\left(s, \mathbf{x}; \bigcup_{k=1}^{m} A_k\right) \leqq 1 - J_1(1 - J_2)$. (Here and hereafter $P[\cdot \mid \mathfrak{A}]$ denotes a conditional law given an event $\mathfrak{A}$.) We may suppose $(9s + r^2, \mathbf{z}) \in K_9$ for $z < 3r$ so as to be able to apply (2.9) with $c = 9$ to obtain $J_2 \leqq \gamma + C \sum_{k=1}^{m} h(O, A_k)$. Therefore

$$h\left(s, \mathbf{x}; \bigcup_{k=1}^{m} A_k\right) \leqq 1 - J_1 + J_1\gamma + J_1 C \sum_{k=1}^{m} h(O, A_k),$$

and if we can show that $J_1 = J_1(s, \mathbf{x}) \geqq p$ with a positive constant p, then (2.8) holds with γ replaced by $\gamma' = 1 - p(1 - \gamma) < 1$.

Step 3. Now it suffices to prove that J_1 is bounded below by a positive constant. We set

$$T=s/(\Delta x)^2,\ L=x/\Delta x,\ \kappa=(X_2(9s)-X_2(s))\cdot\Delta x/8s,$$

where $\Delta x=X_2(s)-X_1(s)$. By applying first Markov property of the Wiener process and then the scaling invariance and the rotation invariance of it, it follows that

$$\begin{aligned} J_1 \geqq P[|W(t)|\leqq L+1+\kappa(t-T) \quad \text{for} \quad T\leqq t\leqq 9T \\ \text{and} \quad |W(9T)|<2L|\ |W(T)|=L] \\ \times \inf_{0\leqq u\leqq 2x} P[\pi_2 \text{ is not hit at all} \mid |W(9s)|=u]. \end{aligned}$$

Since (2.11) implies $x\leqq r$, the second factor on the right-hand side is greater than or equal to the probability under the infimum with $u=2r$, which agrees with

$$p_0:=P[|W(t)|\leqq 3 \quad \text{for} \quad 1\leqq t\leqq 2|\ |W(1)|=2]>0.$$

Let us evaluate the first factor, which we name J_3. To this end single out the first event from the two; and also note that, by (2.5, 2.11), $\kappa\geqq(N\log 2)/8$ if c is large enough. Then the conditional probability of it given $|W(T)|=L$ is greater than

$$p_1:=P[|W(t)|\leqq 1+((N\log 2)/8)t \quad \text{for} \quad t>0]>0,$$

as is clear by first replacing the upside-down frustum $\{(t,y): y\leqq L+1+\kappa(t-T),\ T\leqq t\leqq 9T\}$ [which $W(t)$ is to stay within] by another one $\{(t,\mathbf{y}): |\mathbf{y}-W(T)|\leqq 1+\kappa(t-T),\ T\leqq t\leqq 9T\}$ [which is contained in it] and then applying the translation invariance of the Brownian motion. As for the second event, noticing $T/L^2=s/x^2\leqq 1/c$, we apply the scaling invariance of W again to see that if c becomes large the conditional probability of $|W(9T)|\leqq 2L$ given $|W(T)|=L$ approaches 1; hence it can be made greater than $1-p_1/2>0$. Accordingly $1-J_3\leqq 1-p_1+p_1/2$, i.e., $J_3\geqq p_1/2$, so that $J_1\geqq p_1p_0/2$. The proof of Lemma 5 is complete.

We lastly prepare the following relation

$$\inf_{0<s\leqq 1} P[|W(t)|<X_2(t) \quad \text{for all} \quad 0<t<s|\ |W(s)|=X_1(s)]>0. \tag{2.12}$$

The probability under the infimum is not less than

$$P[|W(t)|<X_2(t)-(X_1(s)/s)t,\, 0<t<s\,|\,W(s)=\mathbf{0}].$$

By conditioning on $W(s/2)$, replacing $X_2(t)$ by a linear function for $s/2\leqq t<s$, and then scaling $W(\cdot)$ with the constant $\sqrt{s}$ for $0\leqq t\leqq s/2$ and with $\Delta x=(X_2(s)-X_1(s))$ for $s/2\leqq t\leqq s$, this probability is bounded below by

$$\begin{aligned} \int_{R^N} P[W(1/2)\in d\mathbf{y}\,|\,W(1)=0]\cdot P[|W(t)|\leqq f(t),\, 0<t\leqq 1/2\,|\,W(1/2)=\mathbf{y}] \\ \times P[|W(t)|\leqq g(t),\, 0<t<s/2\Delta x^2\,|\,W(s/2\Delta x^2)=\sqrt{s}\mathbf{y}/\Delta x] \end{aligned}$$

where $f(t)=[X_2(st)-X_1(s)t]/\sqrt{s}$ and

$$g(t)=\begin{cases} 1+(\Delta x/s)[X_1(s)-2(X_2(s)-X_2(s/2))]t & \text{if} \quad s\leqq 1/5 \\ X_2(1/5)-X_1(1/5) & \text{if} \quad s>1/5. \end{cases}$$

It is easily seen that $f(t) \geqq (1-1/\sqrt{2})X_2(t)$ for $0 \leqq t \leqq 1/2$, and that if $s \leqq 1/5$, $g(t) \geqq 1+\lambda t$ with some positive constant λ independent of s. It holds that $P[|W(t)| \leqq cX_2(t), 0<t<1/2]>0$ for every $c>0$ (see the first part of Appendix). Either by applying the projective invariance (1.4) to obtain conditioning-free expressions for two probabilities in the integral above or by observing the monotonicity of them in y, we, from these bounds for f and g, conclude (2.12).

Proof of Theorem 1. With Lemma 5 and (2.12) having been prepared the proof of Theorem 1 is rather standard. Let A and A_n be as in Theorem 1. We can assume $\sum_{k=1}^{\infty} h(O; A_{2k}) = \infty$ and $\lim_{k\to\infty} h(O; A_{2k}) = 0$, which will entail no loss od generality. Let $\mathfrak{A}_k$ denote the event $\{$there exists $t>0$ such that $(t, W(t)) \in A_k$ and $(s, W(s)) \notin B(2^{k+2})$ for all $0<s<t\}$. By the Blumentahl's $0-1$ law we have only to show that $P[\mathfrak{A}_{2k}$ occurs for i.m. $k]>0$. By considering the time-reversed motion $W(2^{2k+1}-t)$ and applying its strong Markov property it is seen that $P(\mathfrak{A}_k) \geqq qh(O; A_k)$ where q denotes the infimum in (2.12). Therefore for every $\delta>0$ and every sufficiently large integer L there exists $M>L$ such that

$$\delta \leqq \sum_{n=L}^{M} P[\mathfrak{A}_{2n}] \leqq \sum_{n=L}^{M} h(O; A_{2n}) \leqq 2\delta/q. \tag{2.13}$$

Then in view of (2.3) Lemma 5 shows that there exist constants $\delta>0$ and $\gamma'<1$ such that (2.13) implies $h\left(s, \mathbf{x}; \bigcup_{j=n+1}^{M} A_{2j}\right) \leqq \gamma'$ for all $(s, \mathbf{x}) \in \bar{A}_{2n}$ and $L \leqq n < M$. Since under $\mathfrak{A}_{2n}$ the event $\mathfrak{A}_{2j}$, $j>n$, can occur only if A_{2j} is hit after the first hitting of A_{2n}, by the strong Markov property this inequality yields $P\left[\bigcup_{j=n+1}^{M} \mathfrak{A}_{2j} \middle| \mathfrak{A}_{2n}\right] \leqq \gamma'$, or, what is the same,

$$P\left[\mathfrak{A}_{2n} \setminus \bigcup_{j=n+1}^{M} \mathfrak{A}_{2j}\right] \geqq (1-\gamma') P[\mathfrak{A}_{2n}].$$

The events, for $n=L, \ldots, M$, measured on the left-hand side are mutually disjoint and the union of them coincides with the union of $\mathfrak{A}_{2n}$. Thus, by the first inequality of (2.13), $P\left[\bigcup_{n=L}^{M} \mathfrak{A}_{2n}\right] \geqq (1-\gamma')\delta$, which shows that $\lim_{L\to\infty} P\left[\bigcup_{n\geqq L} \mathfrak{A}_{2n}\right] = P[\mathfrak{A}_{2n}$ occur for i.m. $n]>0$ as required.

A Remark to the Proof of Theorem 1. If A is contained in a parabolic body $x^2<ct$, the assertion of Theorem 1 may directly proved by a simple method. Indeed one sees, by (2.7), that if $k \geqq 2$ and A_{n+k} lies within such a paraboloid, then

$$\sup_{(s,\mathbf{x}) \in B(a_{n+k})} \sup_{(t,\mathbf{y}) \in A_{n+k}} p(t-s, |\mathbf{y}-\mathbf{x}|) \leqq \text{const}\, e^{c/2} (a_{n+k})^{-N/2},$$

where $a_n = 2^n$, and accordingly can follow both Lamperti [9] and Ito and McKean [8]. If one take a_n, in stead of 2^n, such that a_{n+1}/a_n tends to infinity rapidly enough, this method can be applied to prove corresponding results (without the above constraint on A), but they are of course not so useful as Theorem 1.

3. A Refinement of Theorem 1

In this section will be given a refinement of Theorem 1 where each shell $B(2^{n+1})\setminus B(2^n)$ is sliced with abscissa hyperplanes to form an infinite sequence of annuli into which the set A whose thinness is in question is to be partitioned. Given $a>0$, set $X_a(t)=(-Nt\log t/a)^{1/2}$ as before and let X_a^{-1} be the inverse function of X_a restricted to $0<t\leqq a/e$. Define a sequence $\{t_k\}_{k=-1}^{\infty}$ by $t_{-1}=1$, $t_0=1/e$ for the first two entries and then inductively by

$$x_k=X_1(t_k),\ t_{k+1}=X_2^{-1}(x_k) \quad \text{for} \quad k\geqq 0 \tag{3.1}$$

and set

$$D_k=\{(t,\mathbf{x})\in B(2)\setminus B(1): t_{k+1}<t\leqq t_k\}, \qquad k=-1,0,1,\ldots.$$

These D_k's together then constitute a partition of $B(2)\setminus B(1)$. Recalling the mapping $\varphi^{\varepsilon}:(t,\mathbf{x})\to(\varepsilon t,\sqrt{\varepsilon}\,\mathbf{x})$ together with the fact stated in (2.3) we finally define

$$D_k^{(n)}=\varphi^{2^n}(D_k).$$

Thus we have a partition $\{D_k^{(n)}\}_{k,n}$ of $(0,\infty)\times R^N$ which is much finer than $\{B(2^{n+1})\setminus B(2^n)\}$. The next proposition therefore is stronger than Theorem 1 (and Theorem 2).

Proposition 6. *An analytic set A of R^{N+1} is minimal thin at ζ_0 [thin at the origin] if and only if*

$$\sum_{\substack{n\geqq 0\\ [n\leqq 0]}} 2^{-Nn/2}\sum_{k=-1}^{\infty}\operatorname{Cap}(A\cap D_k^{(n)})<\infty,$$

or equivalently, $\sum_{n,k} h(O;A\cap D_k^{(n)})<\infty$. □

Before proceeding to the proof of Proposition 6 we point out the following relation

$$\Delta x_k := x_k-x_{k-1}\sim\tfrac{1}{2}\sqrt{N\log 2}\,\sqrt{t_k-t_{k+1}} \quad \text{as} \quad k\to\infty \tag{3.2}$$

("$\sim$" means that the ratio of two quantities which hold it between approaches 1), which we shall use both in proving and in applying Proposition 6. The verification of (3.2) is ready from (2.5) if one observes

$$t_k-t_{k+1}=t_k\left(1-\frac{\log t_k}{\log t_{k+1/2}}\right)\sim(\log 2)\frac{-t_k}{\log t_k} \quad \text{as} \quad k\to\infty. \tag{3.3}$$

Proof of Proposition 6. We consider only the criterion of thinness at infinity. We have merely to show that if $\sum_{k,n} h(O;A\cap D_k^{(n)})=\infty$, then $P[A$ is hit i.o. as $t\uparrow\infty]=1$, for the converse is ready from the Borel-Cantelli lemma. In view of Theorem 1 this implication follows from

Lemma 7. *Set $A_n=A\cap[B(2^{n+1})\setminus B(2^n)]$ and $A_{n,k}=A\cap D_k^{(n)}$. If $h(O;A_n)\to 0$ as $n\to\infty$, then there exists a constant C which may depend on A but not on n such that* $\sum_{k=-1}^{\infty} h(O;A_{n,k})\leqq C\cdot h(O;A_n)$ *for all n.* □

Proof. By arguing as in the proof of Theorem 1 the problem is reduced to showing

$$1-h\left(s,\mathbf{x};\bigcup_{j=-1}^{k-1}A_{n,j}\right)>q \quad \text{if} \quad (s,\mathbf{x})\in A_{n,k+1} \tag{3.4}$$

for all $k\geqq 0$ and for all sufficiently large n where q is a positive constant which may depend on A but not on k nor on n. By the strong Markov property of $W(\cdot)$ the left-hand side of (3.4) is greater than

$$P[|W(2^n t_k)|<r\,|\,W(s)=\mathbf{x}]\left\{1-\sup_{y\leqq r}h\left(2^n t_k,\mathbf{y};\bigcup_{j=-1}^{k-1}A_{n,j}\right)\right\}$$

where $r=X_{2^{n-1}}(2^n t_k)$. That the first factor is bounded below by a positive constant follows from (3.2) and the relation

$$x-r\leqq 2^{n/2}x_k-r=2^{n/2}(X_1(t_k)-X_{1/2}(t_k))\sim\mathrm{const}\,2^{n/2}(x_{k-1}-x_k)$$

as $k\to\infty$ uniformly in n; for n large enough the same is true for the second factor as an application of Lemma 5 (with $m=1$) verifies. Accordingly we get (3.4). This completes the proof of Lemma 7 and hence that of Proposition 6.

4. Applications

1. As a simplest application of Theorem 1 we prove the Kolmogorov's test which was stated in Lévy's book [10, p. 266] without proof and has been proved in various ways by Petrovskii [12], Erdös [3], Feller [5], and Motoo [11] (among these only Petrovskii's proof is purely analytic; the others' proofs are probabilistic but still quite different from one another). Let $f(t)$ be a non-negative Borel function of $t\geqq 0$ and set $F=\{(t,\mathbf{x})\in R^{N+1}\colon t\geqq 0,\ |\mathbf{x}|\geqq f(t)\}$. Then the Kolmogorov's test reads as follows:

Assume $f(t)/\sqrt{t}\uparrow\infty$ as $t\downarrow 0$ [as $t\uparrow\infty$]. Then the set F is thin at the origin [resp. minimal thin at the Martin boundary point ζ_0] if and only if

$$\int_{\substack{0<t<1\\ [t>1]}} f(t)^N p(t,f(t))\frac{dt}{t}<\infty\,. \tag{4.1}$$

Proof. The proof is carried out only in the case where the thinness at ζ_0 is concerned. The other case is reduced to it by using the projective invariance (1.4) (the same comment may be applied throughout this section, though the direct modification of proof also may be possible in some places). First we prove the sufficiency of (4.1). For this part we do not apply Theorem 1, but make use of the explicit expression (1.3) of the hitting distribution ν_a. The assumed monotonicity of $f(t)/\sqrt{t}$ implies that the boundary ∂F crosses $S(a)$ at exactly one ordinate level $t=t(a)$ (if a is sufficiently large). [This is seen either by comparing $f(t)$ with the function $c\sqrt{t}$ or by noting that for it to be true it is sufficient that the lower derivative of $f(t)^2/t$ on the right is greater than $-1/t$.] Let $t_k=2^{2k/N}$ and a_k the value of a for which ∂F crosses $S(a)$ at the level $t=t_k$; if there are many such a's take the infimum of them. Let B_k denote the part of $S(a_k)$ below the level $t=t_{k+1}$. Let τ_a be the first hitting time to $S(a)$ by $(t, W(t))$. Then from (1.3, 2.2) it follows that for t/a small and $0 \leqq s < t$

$$\begin{aligned} P[s<\tau_a \leqq t] &= \int_\Theta d\theta \int_s^t (2\pi a)^{-N/2} X_a(u)^{N+1} \frac{-1}{Nu(1+\log u/a)} dX_a(u) \\ &= \left(\int_\Theta d\theta/N\right)(2\pi a)^{-N/2}(X_a(t)^N - X_a(s)^N)(1+o(1)), \end{aligned} \tag{4.2}$$

where $o(1)\to 0$ as $t/a\to 0$. From $\lim f(t)/\sqrt{t}=\infty$, it follows that $t_k/a_k \to 0$ and $X_{a(k)}(t_{k+1})/X_{a(k)}(t_k)\to 2^{1/N}$ as $k\to\infty$ where $a(k)=a_k$. Now keeping these in mind apply (4.2) and then, noticing $X_{a(k)}(t) \geqq f(t)$ for $t<t_k$, use (1.3) to see that for large k

$$\begin{aligned} h(O;B_k) = P[\tau_{a(k)} \leqq t_{k+1}] &\leqq 4P[t_{k-1} < \tau_{a(k)} \leqq t_k](1+o(1)) \\ &\leqq \text{const} \int_{t_{k-1}}^{t_k} f(t)^N p(t, f(t)) t^{-1}\, dt. \end{aligned}$$

Let F_k be the part of F between two levels $t=t_k$ and $t=t_{k+1}$. Then $h(O;F_k) \leqq h(O;B_k)$ since $R^{N+1}\setminus B(a)$ is thin at O. Accordingly (4.1) implies $\sum h(O;F_k)<\infty$ and hence $P[F$ is hit i.o. as $t\uparrow\infty]=0$ or, equivalently, F is minimal thin at ζ_0.

The proof of the converse part is similar, but this time t_k is defined as the ordinate at which level F intersects $S(2^k)$. Let A_k and A'_k be the parts of $S(2^k)$ and $S(2^{k+1})$, respectively, between two levels $t=t_k$ and $t=t_{k+1}$. Note that A'_k is included in $F'_k := F\cap(B(2^{k+1})\setminus B(2^k))$, whereas A_k lies outside F. By (1.3) and by writing $a(k)=2^k$ we as above see

$$\begin{aligned} h(O;A'_k \geqq v_{a(k+1)}(A'_k) &= C_N \int_{t_k}^{t_{k+1}} X_{a(k+1)}(t)^N (2\pi a(k+1))^{-N/2} t^{-1}\, dt \\ &\geqq 2^{-1-N/2} v_{a(k)}(A_k) \\ &\geqq 2^{-2-N/2} C_N \int_{t_k}^{t_{k+1}} f(t)^N p(t, f(t)) t^{-1}\, dt \end{aligned}$$

for k large enough where C_N is a positive constant. Thus the divergence of the definite integral in (4.1) implies $\sum h(O;F_k)=\infty$; hence by Theorem 1 $P[F$ is hit i.o. as $t\uparrow\infty]=1$. The proof is complete.

2. For the necessity of the condition (4.1) in the Kolmogorov's test a slight modification of its proof dispenses the monotonicity assumption on $f(t)/\sqrt{t}$. In the next theorem not only this assumption is removed, but also the conclusion is strengthened. (F is replaced by the graph of f).

Theorem 8. *Let $f=f(t,\theta)$ be a positive Borel function of $t \geqq 0$ and of $\theta=\mathbf{x}/x$ and G denote the graph of f: $G:=\{(t,\mathbf{x}): f(t,\mathbf{x}/x)=x\}$. Let $\psi(x)$ be a bounded positive function of $x>0$ which is non-increasing and slowly varying as $x\downarrow 0$ and satisfies $\int_0^1 \psi(x)x^{-1}\,dx<\infty$; if $N=2$ it is further assumed that $\psi(x)(-\log x)$ is bounded. Put*

$\Psi_1(x)=1_{(0,1)}(x)$ and $\Psi_N(x)=x^{N-2}\psi(x)1_{(0,1)}(x)$ for $N\geqq 2$ (1_Δ denotes the indicator function of a set Δ). For G to be not thin at the origin [minimal thin at ζ_0] it is sufficient that

$$\int\limits_{\substack{0<t<1\\ [t>1]}}\frac{dt}{t}\int\limits_{\Theta}\{\Psi_N(f(t,\theta)/\sqrt{t})+f(t,\theta)^N p(t,f(t,\theta))\}\,d\theta=\infty\,. \quad \square$$

Remark. A positive function $\psi(x)$ is slowly varying as $x\downarrow 0$ if $\psi(\kappa x)/\psi(x)\to 1$ as $x\downarrow 0$ for every $\kappa>0$. For every $\delta>0$ the function $\min\{1,|\log x|^{-1-\delta}\}$ can serve as ψ in Theorem 8. Of the integrand parenthesized in the inner integral above the first term becomes dominant to the second as $f(t,\theta)/\sqrt{t}$ approaches zero. By comparison Ψ_N obviously can be deleted from the integrand to obtain Theorem 3 of Introduction.

For the proof we prepare two lemmas. Given a positive continuous function $h(t)$ of $t\in[\alpha,\beta]$, let us consider a "projection" π_h which is defined by

$$\pi_h(t,\mathbf{x})=(t,h(t)\mathbf{x}/x),\qquad (t,\mathbf{x})\in[\alpha,\beta]\times R^N.$$

Lemma 9. *Let h and π_h be as above. Let G be a graph of a Borel function f as in Theorem 8 and A a Borel subset of G. If $\alpha\leqq t\leqq\beta$ and $x\geqq h(t)$ for all $(t,\mathbf{x})\in A$, then* $\mathrm{Cap}(\pi_h A)\leqq e^{\gamma/2}\,\mathrm{Cap}(A)$ *where*

$$\gamma:=\sup_{\alpha\leqq s<t\leqq\beta}\frac{(h(t)-h(s))^2}{t-s}.\quad\square$$

Proof. We shall make use of the formula

$$\mathrm{Cap}(A)=\sup\{\mu(A):\mu\text{ is supported by }A\text{ and }g\mu\leqq 1\},\tag{4.3}$$

where, for μ a finite Borel measure on R^{N+1}, $g\mu(\xi)=\int g(\xi,\eta)\,\mu(d\eta)=\int p(t-s,|\mathbf{y}-\mathbf{x}|)\,\mu(dt\,d\mathbf{y})$ $(\xi=(s,\mathbf{x}))$ (cf. Watson [13]; also Doob [2, p. 243]). For a measure μ supported by A, let μ' denote a measure on $A'=\pi_h A$ induced by π_h from μ. In view of (4.3) it suffices to show that if $g\mu'\leqq 1$, then $g\mu\leqq e^{\gamma/2}$, because the mapping: $\mu\to\mu'$ is reversible. Let us write $\mathbf{x}^t=h(t)\mathbf{x}/x$ etc. We shall actually prove that for $\alpha\leqq s\leqq\beta$

$$g\mu(s,\mathbf{x})\leqq e^{\gamma/2}g\mu'(s,\mathbf{x}^s)\quad\text{if}\quad x\geqq h(s)$$

which is enough for our end by virtue of the maximum principle. Since

$$g\mu'(s,\mathbf{x}^s)=\int p(t-s,|\mathbf{y}^t-\mathbf{x}^s|)\,\mu(dt\,d\mathbf{y}),$$

we have only to show that for $\alpha\leqq s<t\leqq\beta$

$$|\mathbf{y}^t-\mathbf{x}^s|^2\leqq|\mathbf{y}-\mathbf{x}|^2+(h(t)-h(s))^2\quad\text{if}\quad x\geqq h(s)\quad\text{and}\quad y\geqq h(t).\tag{4.4}$$

If $\mathbf{x}^s\cdot\mathbf{y}^t\leqq\min\{|\mathbf{x}^s|^2,|\mathbf{y}^t|^2\}$, then is true this inequality even with the second term on the right side discarded. If $\mathbf{x}^s\cdot\mathbf{y}^t\geqq|\mathbf{y}^t|^2$, then

$$|\mathbf{y}^t-\mathbf{x}^s|^2\leqq\min_{b>0}|\mathbf{x}^s-b\mathbf{y}|^2+(h(t)-h(s))^2$$

which clearly yields (4.4). The remaining case, where we have $\mathbf{x}^s\cdot\mathbf{y}^t\geqq|\mathbf{x}^s|^2$, can be treated by interchanging the roles of $\mathbf{x}^s$ and $\mathbf{y}^t$ in the above. Thus Lemma 9 has been proved.

Lemma 10. *Let G and $\Psi_N(x)$ be as in Theorem 8. Then there is a constant C which may depend on Ψ_N but not on G such that if A is a Borel subset of $G \cap \{(t, \mathbf{x}): 0 \leqq t \leqq 1, x \leqq 1\}$, then*

$$\operatorname{Cap}(A) \geqq C \int_{(t,\mathbf{y}) \in A} \Psi_N(f(t,\theta))\, dt\, d\theta \tag{4.5}$$

where $\mathbf{y}$ stands for $f(t,\theta)\theta$. □

Remark. When $N=1$ (4.5) is surpassed by

$$\operatorname{Cap}(A) \geqq \sqrt{2\pi}\, |\pi_0 A| \left(\sup_{0 \leqq s \leqq 1} \int_{\pi_0 A \cap (s,1]} dt / \sqrt{t-s} \right)^{-1}, \tag{4.6}$$

where $\pi_0 A = \{t: (t, \mathbf{y}) \in A \text{ for some } \mathbf{y}\}$ and $|\cdot|$ denotes the Lebesgue measure on R. This inequality may afford a better sufficient condition for thinness than that of Theorem 8.

Proof of Lemma 10. Let μ be the measure on A whose value of A' a Borel subset of A is given by the integral on the right-hand side of (4.5) but with A' in place of A. It suffices to prove that $g\mu(t, \mathbf{x}) \leqq C^{-1}$ for $0 \leqq s \leqq 1$, $x \leqq 1$. Let $x \leqq 1$ and $0 \leqq s \leqq 1$. Noting that if $|\mathbf{y}-\mathbf{x}| > x/2$ then $|\mathbf{y}-\mathbf{x}| > y/3$, we see

$$\begin{aligned} g\mu(s, \mathbf{x}) \leqq & \int_{\substack{|\mathbf{y}-\mathbf{x}| < x/2 \\ (t,\mathbf{y}) \in A}} p(t-s, |\mathbf{y}-\mathbf{x}|)\, \Psi_N(f(t,\theta))\, dt\, d\theta \\ & + \int_s^1 dt \int_\Theta p(t-s, f(t,\theta)/3)\, \Psi_N(f(t,\theta))\, d\theta. \end{aligned} \tag{4.7}$$

The case $N=1$ is readily disposed of and omitted. Let $N \geqq 2$. To get an upper bound of the first integral in (4.7) apply the projection π_h with $h \equiv x/2$ and make a comparison as in Lemma 9. Since $x^{N-1} d\theta$ equals a constant multiple of a surface element on $(N-1)$-dimensional sphere of radius $x/2$, we then see that it is bounded above by

$$C_1 \frac{\psi(x)}{x} \int_0^1 dt \int_0^x p(t,u) u^{N-2} du = C_1' \frac{\psi(x)}{x} \int_0^x du \int_u^\infty e^{-y^2/2} y^{N-3} dy,$$

which is dominated by a constant multiple of $\psi(x)$ if $N \geqq 3$ and that of $\psi(x)\,|\log x|$ if $N=2$. As for the second integral in (4.7) we have only to use the inequality

$$p(u, y/3) y^{N-2} \psi(y) \leqq C_2 \psi(\sqrt{u})/u$$

to bound it above by $C_2' \int_0^1 \psi(\sqrt{u}) u^{-1} du = 2C_2' \int_0^1 \psi(u) u^{-1} du$. By arguing in each of the two cases $y < \sqrt{u}$ and $y \geqq \sqrt{u}$ the inequality above is ready from the monotonicity of ψ and the fact that the slowly varying function is always of the form $\psi(x) = a(x) \exp\left(\int_x^1 \varepsilon(u) u^{-1} du \right)$ where $a(x) \to 1$ and $\varepsilon(x) \to 0$ as $x \to 0$ (cf. Feller [6, p. 274]). Consequently $g\mu$ is bounded. Lemma 10 has been proved.

Proof of Theorem 8. Proposition 6 in the previous section will be of use in this proof. Let $G_{n,k} = G \cap D_{n,k}$. Then applying Lemma 9 with $h(t) = X_{2^n}(t)$,

$2^n t_{k+1} \leqq t \leqq 2^n t_k$ we see that if $k \geqq 0$

$$\sqrt{2}^N h(O; G_{n,k}) \geqq (2\pi 2^n)^{-N/2} \operatorname{Cap}(G_{n,k}) \geqq (2\pi 2^n)^{-N/2} e^{\gamma/2} \operatorname{Cap}(\pi_h G_{n,k}),$$

where

$$\gamma := \sup \left\{ \frac{(X_1(t) - X_1(s))^2}{t-s} : t_{k+1} \leqq t-s \leqq t_k,\ k=0,1,2,\dots \right\}.$$

From (3.2, 3.3, 2.2) it follows that $\gamma < \infty$. Since $x^N p(t,x)$ is decreasing with respect to x in the region $x^2 > Nt$ $(t>0)$ which contains $D_{n,k}$ if $k \geqq 0$, and since the right-most member of the inequalities above equals $e^{\gamma/2} h(O, \pi_h G_{n,k}) = e^{\gamma/2} v_{2^n}(\pi_h G_{n,k})$, an application of (1.3) deduces that if $k \geqq 0$

$$h(O; G_{n,k}) \geqq 2^{-N/2} e^{\gamma/2} \int_{(t,\mathbf{y}) \in G_{n,k}} f(t,\theta)^N p(t, f(t,\theta)) \frac{1}{t} \, dt d\theta,$$

where $\mathbf{y} = f(t,\theta)\theta$. Finally Lemma 10 together with the scaling relation (2.3) proves

$$h(O; G_{n,-1}) \geqq C \int_{(t,\mathbf{y}) \in G_{n,-1}} \Psi_N(f(t,\theta)/\sqrt{t}) t^{-1} \, dt d\theta.$$

The assertion of Theorem 8 follows from Proposition 6 and the last two inequalities.

3. The criterion given by Theorem 8 may be fairly accurate if $f(t,\theta)/\sqrt{t}$ is large, but not if small. To see the latter consider the tube $T_{s,x} = \{(t,\mathbf{y}) : 0 < t < s,\ |\mathbf{y}| = x\}$. Replacing $\Psi_N(x)$ in (4.5) by x^{N-2} or $|\log x|^{-1}$ according as $N \geqq 3$ or $N = 2$, we make a computation similar to that in the proof of Lemma 10 to see that as $x \downarrow 0$

$$\operatorname{Cap}(T_{1,x}) \asymp x^{N-2} \quad \textit{or} \quad |\log x|^{-1} \ \textit{according as } N \geqq 3 \textit{ or } = 2, \tag{4.8}$$

where "$\asymp$" means that the ratio of two quantities which hold this symbol between is bounded away from zero and infinity. (For getting the upper bound it may be noted that $g\mu_1 \leqq g\mu_2$ implies $\mu_1(R^{N+1}) \leqq \mu_2(R^{N+1})$.) By (4.8) together with the scaling relation

$$\operatorname{Cap}(\varphi^\varepsilon A) = \varepsilon^{N/2} \operatorname{Cap}(A).$$

Theorem 1 yields the Dvoretsky-Erdös test [Spitzer's test] which may read as follows: if $f(t,\theta) = f(t)$ a function of t only and $h(t) := f(t)/\sqrt{t}$ decreases to zero as $t \uparrow \infty$, then G is not thin at ζ_0 if and only if $\int^{\infty} h(t)^{N-2} t^{-1} dt = \infty$ in the case $N \geqq 3$ $\left[\text{resp. } \int^{+\infty} |\log h(t)|^{-1} t^{-1} dt = \infty \text{ in the case } N = 2\right]$, which condition is somewhat weaker than the divergence condition given in Theorem 8.

Complementary to (4.8) we have $\operatorname{Cap}(T_{\Delta t, x}) \asymp \sqrt{\Delta t}\, x^{N-1}$ for $x^2 > \Delta t > 0$, $N \geqq 1$. It follows from (2.5) and (3.2) that $x_k \sqrt{t_k - t_{k+1}}/t_k \sim$ const.; then from (1.2) that for $t \geqq \Delta t > 0$

$$\begin{aligned} J(t, \Delta t, x) &:= P[|W(s)| = x \text{ for some } s \in (t, t+\Delta t)] \\ &\asymp \sqrt{\Delta t}\, x^{N-1} p(t,x) \text{ as long as } \frac{\sqrt{\Delta t}}{x}, \frac{x}{t}\sqrt{\Delta t} \leqq C, \end{aligned} \tag{4.9}$$

where C is any constant, which the constants involved in "$\asymp$" may depend on (if $N=1$, then $\sqrt{\Delta t}/x \leqq C$ may be deleted, so that $x=0$ is allowed). The relation in (4.9), which may be rewritten $(x\sqrt{\Delta t/t})J(t,\Delta t,x) \asymp x^N p(t,x)\Delta t/t$, shows that the ratio of the probability $P[W(s)\in S(a)$ for some $s\in[t,t+\Delta t)$ but not for any $s\in(0,t)]$ to the probability $J(t,\Delta t,X_a(t))$ stays off zero or does not according as $X_a(t)\sqrt{\Delta t/t}$ does or does not. This observation suggests that the divergence condition in Theorem 8 fails to be critical if the heights of $G\cap D_{n,k}$ are much smaller than those of $D_{n,k}$. As an example let $f(t)=\sqrt{2ct\log\log t}$ if $2^n\leqq t<2^n(1+n^{-\gamma})$, $n=1,2,\ldots$ and $f(t)=\infty$ otherwise. Then $P[|W(t)|=f(t)$ for some $t\in[2^n,\ 2^{n+1})] \asymp J(2^n,\ 2^n n^{-\gamma},\ f(2^n)) \asymp n^{-\gamma/2-c}\sqrt{\log n}^{N-1}$, (as $n\to\infty$), (1) where for the first relation we employed Lemma 9. This shows that $P[|W(t)|=f(t)$ i.o. as $t\uparrow\infty]=1$ if and only if $0\leqq\gamma/2\leqq 1-c$, whereas a sufficient condition provided by Theorem 8 is $0\leqq\gamma\leqq 1-c$.

4. Let $q(r)$, $r\geqq 0$ be a non-negative non-decreasing function. If $N=3$, it is supposed that $q(0+)=0$ and $\lim_{r\to 0} r|\log q(r)|/q(r)=0$. Let θ_0 be a fixed unit vector of R^N. Put

$$f(t,\theta)=\sqrt{2t\log(|\log t|\vee 1)}(1+q(|\theta-\theta_0|)), \qquad t>0,$$

and $F=\{(t,\mathbf{x}): t>0, x\geqq f(t,\mathbf{x}/x)\}$. Then *for F to be thin at the origin it is necessary and sufficient that*

$$q(0)>0 \qquad \textit{if} \quad N=1 \quad \textit{or} \quad 2$$

$$\int_{0+} \frac{q(u)^{-3/2}}{-\log u}\, dq(u)<\infty \qquad \textit{if} \quad N=3, \tag{4.10}$$

$$\int_{0+} u q(u)^{-2} dq(u)<\infty \qquad \textit{if} \quad N=4, \tag{4.11}$$

$$\int_{0+} u^{N-1} q(u)^{-N/2-2} dq(u)<\infty \qquad \textit{if} \quad N\geqq 5. \tag{4.12}$$

(The point $u=0$ is not contained in the range of the integration for these integrals.) The same criteria are applied to the minimalthinness at ζ_0.

Before starting the proof it is remarked that for the minimalthinness at ζ_0 is necessary the convergence of the series

$$\sum_{n=1}^{\infty} 2^{-nN/2} \int_{0+} X_{2^n}(t)|\{\mathbf{x}:(t,\mathbf{x})\in F\cap S(2^n)\}|\frac{dt}{t}, \tag{4.13}$$

($|\cdot|$ is the $(N-1)$-dimensional surface area) which follows from Theorem 1 simply by taking into account only the frontal part of $F_n := F\cap(B(2^{n+1})\setminus B(2^n))$ in computation of $h(O,F_n)$. It turns out that this necessary condition is also a sufficient one if either $N=4$ and $q(r)=r^\gamma$ or $N\geqq 5$, but far from that if $N\leqq 3$.

We treat only the thinness at infinity. The case $N=1$ or 2 are easy and omitted. The case $N=3$ or 4 is somewhat delicate. We begin with the case $N\geqq 5$. Let $a(n)=2^n$. Let T_n and $k_n(t)$, $e<t<T_n$, be determined by $\sqrt{2T_n\log\log T_n}=X_{a(n)}(T_n)$ and by $X_{a(n)}(t)=\sqrt{2t\log\log t}(1+k_n(t))$, respectively. Let $L_n=\log\log T_n$ and

$T_n' = T_n(\log T_n)^{-1/N}$. Then by a simple computation

$$\frac{N}{5L_n}\log\frac{T_n}{t} \leqq k_n(t) \leqq \frac{N}{3L_n}\log\frac{T_n}{t} \quad \text{for} \quad T_n' \leqq t \leqq T_n, \tag{4.14}$$

for n large enough. Fixing n let $a=a(n)$, $T=T_{n+1}$, $T'=T_n'$, $L=L_n$, $k^0(t)=k_n(t)$ and $T^0=T_n$. Then

$$h(O;F_n)=\frac{1}{2}\int_{T'}^{T} X_a^N p(t,X_a)\frac{dt}{t}\int h(t,X_a\theta;F_n)d\theta+\delta_n, \tag{4.15}$$

where $X_a=X_a(t)$ and $\delta_n \leqq \hat{v}_a((0,T']\times\Theta)$. It follows that $\Sigma\delta_n<\infty$, which may be seen directly or by noting that for $f(t)=\sqrt{3t\log\log t}$ the integral in the Kolmogorov test converges and $f(T_n)/X_{a(n)}(T_n')\to 1$ as $n\to\infty$. Put $A_{n,t}=\{\theta:|\theta-\theta_0|>q^{-1}(k_n(t))\}$ for $t\leqq T^0$ and $A_{n,t}=\theta$ for $t>T^0$, where $q^{-1}(u)=\sup\{r\geqq 0: q(r)\leqq u\}$ with the convention $\sup\emptyset=0$. We devide the inner integral appearing in (4.15) into the integral on $A_{n,t}$ and that on $\Theta\setminus A_{n,t}$, which we denote by $I_n(t)$ and $II_n(t)$, respectively. The contribution from $II_n(t)$ to the double integral in (4.15), which corresponds to the n-th summand of (4.13), is relatively easy to estimate. Since $h(t,X_a\theta;F_n)=1$ for $\theta\notin A_{n,t}$, it is bounded above and below by constant multiples of

$$a^{-N/2}\int_{T'}^{T^0} X_a(t)^N[q^{-1}(k^0(t))]^{N-1}\frac{dt}{t}. \tag{4.16}$$

Applying (4.14) $X_a(t)\leqq\sqrt{-Nt\log T'/a}=\sqrt{3tL}$ $(t>T')$ and $X_a(t)\geqq\sqrt{2tL}$ $(t<L)$ and changing the variable according to $u=N(\log T^0/t)/3L$ we bound (4.16) from above and below by constant multiples of

$$\frac{1}{N}L^{N/2+1}(T^0/a)^{N/2}\int_0^{1/3}[q^{-1}(u)]^{N-1}e^{-3Lu/2}du.$$

Since $T_{n+1}-T_n\asymp T_n$, the convergence of the sum of the last quantity over n is equivalent to that of the integral of it by $d(\log T^0)=dT^0/T^0$; and since $(T^0/a)^{N/2}d(\log T^0)\sim d(\log\log T^0)$, the latter integral equals $1/N$ times

$$\int_1^\infty L^{N/2+1}dL\int_0^{1/3}[q^{-1}(u)]^{N-1}e^{-3Lu/2}du=\text{const}\int_0^{1/3}[q^{-1}(u)]^{N-1}u^{-N/2-2}du.$$

This proves especially the necessity of (4.12).

To get an upper bound of $I_n(t)$ we further put $R=X_{a(n)}(T_{n+1})$, $r=r_n(t)=X_{a(n)}(t)$, $k(t)=k_{n+1}(t)$ and $b=a(n+1)$. Let $W_1^x(\cdot)$ and $Y^x(\cdot)$ be respectively the 1-dimensional Brownian motion and the $(N-1)$-dimensional Bessel process both starting at x which are mutually independent and defined on the same probability space as so far used. Then, by writing $A=A_{n,t}$,

$$I_n(t)\leqq\int_A P\left[W_1^r(s)>r+\frac{R-r}{T-t}s,\ Y^{l(r\theta)}(s)\leqq X_b(t+s)q^{-1}(k(t))\right.$$
$$\left.\text{for some}\quad 0<s<T-t\right]d\theta,$$

where $l(r\theta)$ is the distance of the point $r\theta$ from the line $u\theta_0$, $u\in R$. Since $(R-r)/(T-t)\geqq(dr/dt)|_{t=T}\geqq L/3R$ if n is large enough, the integral on the right-hand side is at most

$$\int_A P\left[W_1^0(s)>\frac{Lr}{3R}s,\ Y^{l(\theta)}(s)<\frac{1}{r}X_b(t+r^2s)\,q^{-1}(k(t))\quad\text{for some}\quad 0<s<\frac{T-t}{r^2}\right]d\theta.$$

Since the law of the last leaving time of $W_1^0(t)$ from the line ct, $t\geqq 0$ is given by $c\sqrt{t}\,e^{-tc^2/2}t^{-1}\,dt$ and since $X_b(t+r^2s)/r\leqq 2\sqrt{T/t}$, if we write $c=Lr/3R$ and $y=2\sqrt{T/t}\,q^{-1}(k(t))$ and denote by τ_y^r the first hitting time to y by $Y^r(\cdot)$, the last integral is dominated by

$$\text{const}\int_{q^{-1}(k^0(t))}^{1} r^{N-2}\,dr\int_0^\infty \frac{c\sqrt{s}}{s}e^{-sc^2/2}P[\tau_y^r<s]\,ds.$$

Consequently, applying the scaling relation: $P[\tau_y^r<s]=P[\tau_{cy}^{cr}<c^2s]$, we obtain

$$I_n(t)\leqq \text{const}\,c^{-N+1}\int_{cq^{-1}(k^0(t))}^{c} r^{N-2}\,dr\int_0^\infty\frac{1}{\sqrt{s}}e^{-s}P[\tau_{cy}^r<2s]\,ds. \tag{4.17}$$

By dominating $P[\tau_{cy}^r<s]$ by the hitting time distribution of the 1-dimensional Brownian motion when $r\geqq 2cy$, the right-hand side of (4.17) is at most a constant multiple of

$$c^{-N+1}\int_{2cy}^\infty r^{N-2}\,dr\int_0^\infty\frac{r-cy}{\sqrt{u^3}}e^{-(r-cy)^2/2u}\,du\int_u^\infty\frac{1}{\sqrt{s}}e^{-s/2}\,ds$$
$$+c^{-N+1}\int_{cq^{-1}(k^0(t))}^{2cy} r^{N-2}(cy/r)^{N-3}\,dr,$$

which, after a bit of calculation, we can bound by $\text{const}(c^{-N+1}+y^{N-1})$ $=\text{const}(\sqrt{T/t})^{N-1}\{L^{-N+1}+[q^{-1}(k(t))]^{N-1}\}$. Then, as before,

$$\sum_n\int_{T'_n}^{T_{n+1}} r_n^N p(t,r_n)\,I_n(t)\frac{dt}{t}\leqq\text{const}\left\{\int_0^1[q^{-1}(u)]^{N-1}u^{-N/2-2}\,du+C_N\right\},$$

where $C_N:=\int_1^\infty L^{-N/2+1}\,dL$ is finite if $N\geqq 5$. This shows the sufficiency of (4.12).

For $N=3$ or 4 the convergence in (4.10) or (4.11) implies that of the sum of the quantities in (4.16) over n, so that it suffices to evaluate the sum corresponding to $I_n(t)$. We make use of the formula

$$\varphi_\varepsilon^r(\lambda):=\int_0^\infty e^{-\lambda t}P[\tau_\varepsilon^r<t]\,dt=\frac{1}{\lambda}\frac{K_\nu(\sqrt{2\lambda}r)/r^{-\nu}}{K_\nu(\sqrt{2\lambda}\varepsilon)/\varepsilon^{-\nu}},\qquad \nu=\frac{N-3}{2}.$$

(K_ν is a modified Bessel function in the standard notation), together with $K_{1/2}(z)$ $=\sqrt{\pi}/\sqrt{8z}+O(z^{3/2})$ $(z\downarrow 0)$ and $K_0(z)=-\log z+\text{const}+o(z)$ $(z\downarrow 0)$. By the Tchebychev inequality $P[\tau_\varepsilon^r<s]\leqq e^s\varphi_\varepsilon^r(1)$. Applying this for $0<s<1$ and the scaling relation mentioned before, we have

$$\int_0^\infty \frac{1}{\sqrt{s}}e^{-s}P[\tau_\varepsilon^r<2s]\,ds\leqq 3\varphi_{\varepsilon/\sqrt{2}}^{r/\sqrt{2}}(1).$$

If $N=4$, by (4.17) and by $K_\nu(z)=O(e^{-z}/\sqrt{z})$ as $z\to\infty$,

$$I_n(t)\leqq c^{-N-1}(cy)^{N-3}\operatorname{const}\int_0^\infty r^{N-2}(e^{-r}\wedge r^{-N+3})\,dr=\operatorname{const}c^{-2}y.$$

Repeating the same argument as made through several lines from (4.16) down to the end of the paragraph which contains it we would obtain a corresponding upper bound which proves the sufficiency of (4.11). The lower bound can be computed in a similar way (disregard the part of F_n above the level $t=T_n$). If $N=3$, $I_n(t)\leqq\operatorname{const}c^{-2}/|\log cb|$ whenever $cb<1$. This time we put $T_n'=T_n/L_n^4$, which still entails that in (4.15) $\sum\delta_n<\infty$. Then, by proceeding as before, the convergence of $\sum_n\int_{T'}^{T}X_a^Np(t,X_a)I_n(t)t^{-1}\,dt$ follows from that of

$$\int^{+\infty}L^{1/2}\,dL\int_0^{4(\log L)/L}e^{-Lu/2}|\log Lq^{-1}(u)|^{-1}\,du.$$

If the condition $q^{-1}(u)=o(u/|\log u|)$ is employed, after changing the order and variable of integration one sees that the latter convergence in turn follows from (4.10). Thus the sufficiency of (4.10) has been proved. The necessity is similar and easier as in the other cases.

5. Appendix

This appendix consists of four parts. In the third part we shall give a probabilistic proof of the result by Bauer [1] stated in Introduction. Our proof is based on the fact that $R^{N+1}\backslash B(a)$ is thin at the origin and on a result of Fulks [7], which will be proved in the first and the second part, respectively. The last part is devoted to showing the existence of the equilibrium measure e_A in (1.1).

1. For $c>0$ put $f(t)=\sqrt{-ct\log t}$ $(0<t<1)$ and $F=\{(t,\mathbf{x}):x\geqq f(t),0<t<1\}$. We here prove that F is thin at O for every c, which in particular implies the thinness of $R^{N+1}\backslash B(a)$ at O. This claim of course follows from Kolmogorov's test, but in our deduction of it we made use of it granted that ν_a defined via (1.3) is the hitting distribution to $S(a)$, which fact we are going to prove. Since the probability of the event $\mathfrak{A}_k:=\{|W(e^{-k+1})|\geqq f(e^{-k})\}$ is not less than 1/2 times the probability of the event {the part of F between $t=e^{-k}$ and $t=e^{-k+1}$ is ever hit} for $k>1$, the asserted thinness follows from

$$\sum_k P[\mathfrak{A}_k]=\sum_k P[W(e)\geqq e^{k/2}f(e^{-k})]\leqq\operatorname{const}\sum_k\sqrt{k}^{N-2}e^{-(c/2e)k}<\infty.$$

2. For reader's convenience the Fulks derivation of an explicit form of v_a is briefly given in a way relevant to the present context. Let $\mathscr{L}$ and $\mathscr{L}^*$ be respectively the parabolic and coparabolic operators: $\mathscr{L}=\frac{1}{2}\Delta-(\partial/\partial t)$, $\mathscr{L}^*=\frac{1}{2}\Delta+(\partial/\partial t)$. Let $D_\varepsilon=\{(t,\mathbf{y})\in B(a): t>\varepsilon\}$, $\varepsilon>0$. Given a function v *which is defined and satisfies* $\mathscr{L}^*v=0$ *in a neighbourhood of* $B(a)$, we apply the divergence theorem to integrate the expression $u\mathscr{L}^*v-v\mathscr{L}u=\frac{1}{2}\nabla\cdot(u\nabla v-v\nabla u)+\partial(uv)/\partial t$ over D_ε with $u=g$ [recall $g(t,\mathbf{x})=p(t,x)$]. Since $\mathscr{L}g=0$ for $t>0$ and $g=(\sqrt{2\pi a})^{-N}$ on $S(a)$, this after rearrangement yields

$$\int\limits_{x<X_a(\varepsilon)} gv\,d\mathbf{x}=-\frac{1}{2}\int\limits_{S_\varepsilon} v\nabla g\cdot n_1\,d\sigma+\frac{1}{\sqrt{2\pi a}^N}\int\limits_{S_\varepsilon}\left(\frac{1}{2}\nabla v\cdot n_1+vn_2\right)d\sigma$$

where $S_\varepsilon=S(a)\cap\partial D_\varepsilon$, n_1 and n_2 are respectively the abscissa and the ordinate components of the outward unit normal vector to S_ε, and $d\sigma$ is a surface element of S_ε. Taking $u\equiv 1$ in place of g, which results in the same equality as above but with 1 replacing g, one sees that the second term on the right-hand side of this relation vanishes as $\varepsilon\downarrow 0$. The left-hand side equals $E[v(W(\varepsilon)); |W(\varepsilon)|<X_a(\varepsilon)]$, which converges to $v(O)$ since $P[|W(\varepsilon)|\geqq X_a(\varepsilon)]=P[|W(1)|\geqq\sqrt{-N\log\varepsilon/a}]\to 0$. Consequently

$$v(O)=-\tfrac{1}{2}\int\limits_{S(a)} v\nabla g\cdot n_1\,d\sigma. \tag{A.1}$$

But, by introducing a variable $\theta=\mathbf{x}/x$, we have $-\nabla g\cdot n_1=(xg/t)|n_1|$ and $|n_1|\,d\sigma=x^{N-1}\,d\theta\,dt$ at $(t,\mathbf{x})\in S(a)$, so that the relation (A.1) can be written

$$v(O)=\int\limits_{(0,a]\times\Theta} v(t,X_a(t)\theta)\,\hat{v}_a(dt\,d\theta), \tag{A.2}$$

or, what is the same, $v(O)=v_a(v):=\int\limits_{S(a)} v\,dv_a$, where

$$\hat{v}_a(dt\,d\theta)=\frac{1}{2}(2\pi a)^{-N/2}\,X_a(t)^N\frac{dt}{t}\,d\theta, \tag{A.3}$$

and v_a is a measure on $S(a)$ induced from $\hat{v}_a$ by the mapping $(t,\theta)\to(t,X_a(t)\theta)\in S(a)$.

3. Here we shall prove that v_a *defined above via* (A.3) *agrees with the hitting distribution to* $S(a)$ *for* $(t, W(t))$. For $\xi=(s,\mathbf{x})$ let P_ξ denote the probability law of $(t+s, W(t)+\mathbf{x})$ the space-time Wiener process starting at ξ induced on the space of continuous functions $C([0,\infty), R^{N+1})$ equipped with the cylindrical Borel field and β_t [or sometimes $\beta(t)$] a generic element of this space. The expectation by P_ξ is denoted by E_ξ. Let T be the first hitting time to $S(a)$ by β_t: $T=\inf\{t>0:\beta_t\in S(a)\}$. Given f a continuous function on $S(a)$, we put

$$u(\xi)=E_\xi[f(\beta_T)],$$

where by convention $f(\beta_T)=0$ if $T=\infty$. Our aim is to show that $u(O)=v_a(f)$. Let us claim that for a sequence ξ_n in $B(a)$ which converges to the origin it holds that

$$\limsup_{\delta\downarrow 0}\,\limsup_{n\to\infty} P_{\xi_n}[T<\delta]=0 \quad\text{implies}\quad u(O)=\lim_{n\to\infty}u(\xi_n). \tag{A.4}$$

Indeed if $\xi \in \overline{B(a)}$, the Markov property shows

$$E_\xi[f(\beta_T)] = E_\xi[E_{\beta_\delta}[f(\beta_T)]; T > \delta] + E_\xi[f(\beta_T); T < \delta],$$

from which it follows that

$$|E_\xi[f(\beta_T)] - E_\xi[u(\beta_\delta)]| \leqq 2\|f\|_\infty P_\xi[T < \delta].$$

Applying this inequality with $\xi = 0$ and $\xi = \xi_n$ and rewriting $E_\xi[u(\beta_\delta)]$ as $E_O[u(\beta_\delta + \xi)]$, we see that the premiss of (A.4) implies

$$\limsup_{n\to\infty} |u(\xi_n) - u(O)| \leqq \limsup_{\delta\to 0} \limsup_{n\to\infty} |E_O[u(\beta_\delta + \xi_n)] - E_O[u(\beta_\delta)]|,$$

where the thinness of $S(a)$ at O also is applied. Since $S(a)$ is not thin at every point of it except the origin and $(a, \mathbf{0})$, $u(\xi)$ is continuous off these two points. Therefore we get the conclusion of (A.4).

As being well established in the theory of Markov processes $\mathscr{L}^* u = 0$ off $S(a)$. Let $\xi_\varepsilon = (\varepsilon, \mathbf{0})$ for $\varepsilon > 0$. Since

$$(\partial/\partial a)\{X_a(a-t)\}^2/N = -\log(1 - t/a) - t/a > 0, \qquad 0 < t < a,$$

if $b < a - \varepsilon$ the shifted ball $B(b) + \xi_\varepsilon$ is in the interior of $B(a)$, so that by the Fulks result (A.2) $u(\xi_\varepsilon) = v_b(u(\cdot + \xi_\varepsilon))$. Let T' be the first hitting time to $S(a/2)$. Then $P_{(\varepsilon,\mathbf{0})}[T < \delta] \leqq P_O[T' < \delta]$ for $\varepsilon < a/2$. Since $R^{N+1} \setminus B(a/2)$ is thin at the origin, the right-hand side of this inequality vanishes as $\delta \to 0$; hence by (A.4) $u(\xi_\varepsilon)$ converges to $u(O)$. On the other hand since $u(\xi)$ is continuous at every point except the top and bottom points of $S(a)$ and coincides with f on $S(a)$ except for these two and since v_a has no point mass, by the dominated convergence theorem $\lim_{\varepsilon \to 0} v_b(u(\cdot + \xi_\varepsilon)) = v_b(u)$ and owing to the explicit form of v_b provided by (A.3) $\lim_{b \uparrow a} v_b(u) = v_a(u)$. Consequently $u(O) = v_a(f)$ as required.

4. In this part we shall continue to use notations of previous three parts but does not make use of results obtained in them in any essential way. Let us consider the "dual" space-time process $(-t+s, W(t)+\mathbf{x})$, $t \geqq 0$, starting at $\xi = (s, \mathbf{x})$, and mark * on the corresponding objects. Thus P^*_ξ is the law of the dual process, $S^*(a) = -S(a)$ and $g^*(\xi, \eta) = g(\eta, \xi)\,(= g(\xi - \eta))$, etc. ($\beta_t$ is not marked because it is merely a sample path; and not also the first hitting times defined by means of it.) *Given a bounded analytic set A of R^{N+1}, we take $\xi_0 \in R^{N+1}$ and $a > 0$ such that $\bar{A}$* (the closure of A) $\subset B^* := B^*(a) + \xi_0$. *Let T_A and T_{S^*} be the first passage times of β_t to A and to $S^* := \partial B^*$, respectively, and put*

$$\mu^*(\cdot) = (2\pi a)^{N/2} P^*_{\xi_0}[\beta(T_{S^*}) \in \cdot],$$

and

$$e_A(\cdot) = \int P^*_\eta[\beta(T_A) \in \cdot]\, \mu^*(d\eta). \tag{A.5}$$

Then e_A is the equilibrium measure of A, i.e. it satisfies (1.1).

Let $t > s$, $\xi = (s, \mathbf{x})$ and $\eta = (t, \mathbf{y})$. First we prove that for a.a. $\mathbf{y}$

$$E_\xi[g(\beta(T_A), \eta)] = P[(u, W^{\mathbf{x};\mathbf{y}}_{t-s}(u)) \in A \quad \text{for some} \quad 0 < u < t-s]\, g(\xi, \eta) \tag{A.6}$$

where $W_t^{\mathbf{x},\mathbf{y}}(u) := W(u) + \{(t-u)\mathbf{x} + u(\mathbf{y} - W(t))\}/t$, $0 \leqq u \leqq t$, a Brownian bridge starting at $\mathbf{x}$ and ending at $\mathbf{y}$. For the proof of (A.6) we can assume $\xi = O$. Let τ_A be the first hitting time to A by $(t, W(t))$. Then, by recalling that $g(\xi, \eta) = p(t-s, |\mathbf{y}-\mathbf{x}|)$ is a transition density of $W(\cdot)$, for a bounded continuous function $\varphi(\mathbf{y})$ we see

$$\begin{aligned}\int_{R^N} E_O[g(\beta_{T_A}, (t, \mathbf{y}))]\, \varphi(\mathbf{y})\, d\mathbf{y} &= E[E[\varphi(W(t-u) + \mathbf{x})]_{u=\tau_A, \mathbf{x}=W(\tau_A)}; \tau_A < t] \\ &= E[\varphi(W(t)); \tau_A < t] \\ &= \int_{R^N} P[\tau_A < t \,|\, W(t) = \mathbf{y}]\; p(t, \mathbf{y})\; \varphi(\mathbf{y})\, d\mathbf{y}\end{aligned}$$

(where the strong Markov property is applied for the second equality), proving (A.6). The process $W_t^{\mathbf{x},\mathbf{y}}(t-u)$, $0 \leqq u \leqq t$ is also a Brownian bridge but starting at $\mathbf{y}$ and ending at $\mathbf{x}$. Keeping this in mind compare the right-hand side of (A.6) with that of its dual version, then you can see that for a.a. (ξ, η)

$$E_\xi[g(\beta(T_A), \eta)] = E_\eta^*[g^*(\beta(T_A), \xi)]\,. \tag{A.7}$$

Since the both sides of this equality are continuous in (ξ, η) off $\bar{A} \times \bar{A}$, it holds outside $\bar{A} \times \bar{A}$. Let us prove that (A.7) is true for all ξ and η. To this end we can assume that $A \subset \{(t', \mathbf{x}') \in R^{N+1} : s < t' < t\}$. Let $\xi \notin \bar{A}$, $\eta \in \bar{A}$ and $A_n = \{(t', \mathbf{x}) \in A : t' < t - 1/n\}$. Then $T_{A_n} \downarrow T_A$ a.s. P_ξ, and $P_\xi[T_{A_n} = T_A \,|\, T_{A_n} < \infty] = 1$. The latter relation implies $E_\xi[g(\beta(T_{A_n}), \xi)] \leqq E_\xi[g(\beta(T_A), \eta)]$; and application of Fatou's lemma then shows that the expectation on the left-hand side of this inequality converges to that on the right-hand side as $n \to \infty$. As for the convergence of $E_\eta^*[g(\beta(T_{A_n}), \xi)]$ we can simply apply the bounded convergence theorem. Consequently we have (A.7). Repeating the same argument we see that (A.7) is true for $\xi, \eta \in \bar{A}$; hence, by symmetry, for all ξ and η.

We now integrate both sides of (A.7) by $\mu^*(d\eta)$. For the moment let $\xi_0 = O$. Then $\mu^*(\sqrt{2\pi a})^N v_a^*$ and if $\xi \notin \overline{B^*(a)}$ it holds that $v_a^* g^*(\xi) = g^*(\xi)$, since then $\mathscr{L}^*\{g^*(\cdot, \xi)\} = 0$ in a neighbourhood of $\overline{B^*(a)}$. This shows that $\mu^* g^* = g\mu^* = 1$ on $S^* \setminus \{\xi_0\}$ and hence on B^*. Now the integral of the left-hand side equals $E_\xi[g\mu^*(\beta(T_A))] = P_\xi[T_A < \infty] = h(\xi, A)$, while by the very definition of e_A that of the right-hand side is ge_A. Thus our claim is proved.

Acknowledgement. I would like to extend my hearty thanks to Professor H. Bauer: a discussion that I had with him about his result on Fulks measure motivated my verifying the Wiener's test in my ignorance of the work of Evans and Gariepy [4]. I am also indebted to a referee and T. Komatsu for their comments on the present work.

References

1. Bauer, H.: Heat balls and Fulks measures. Ann. Acad. Sci. Fen. Ser. A.I Math. **10**, 67–82 (1985)
2. Doob, J.L.: Classical potential theory and its probabilistic counterpart. Berlin Heidelberg New York: Springer 1984
3. Erdös, P.: On the law of the iterated logarithm. Ann. Math. **43**, 419–436 (1942)
4. Evans, L.C., Gariepy, R.F.: Wiener's criterion for the heat equation. Arch. Rat. Mech. Anal. **78**, 293–314 (1982)
5. Feller, W.: On the general form of the so called law of iterated logarithm. Trans. Amer. Math. Soc. **54**, 373–402 (1943)

6. Feller, W.: An introduction to probability theory and its applications. Vol. II, 2nd edn. New York: Wiley 1971
7. Fulks, W.: A mean value theorem for the heat equation. Proc. Amer. Math. Soc. **17**, 6–11 (1966)
8. Ito, K., McKean Jr., H.P.: Diffusion processes and their sample paths. Berlin Heidelberg New York: Springer 1965
9. Lamperti, J.: Wiener's test and markov chains. J. Math. Anal. Appl. **6**, 58–66 (1963)
10. Lévy, P.: Théorie de l'addition des variables aléatoires. Paris: Gautier-Villars 1937
11. Motoo, M.: Proof of the law of iterated logarithm through diffusion equation. Ann. Inst. Statis. Math. **10**, 21–28 (1959)
12. Petrovskii, I.: Zur ersten Randwertaufgabe der Wärmeleitungsgleichung. Compos. Math. **1**, 383–419 (1935)
13. Watson, N.A.: Thermal capacity. Proc. Lond. Math. Soc. **37**, 342–362 (1978)

Received June 19, 1987; in revised form March 25, 1988

Math. Ann. 283, 87–95 (1989)

Mathematische Annalen

The Open Mapping and Closed Range Theorems

B. Rodrigues[1] and S. Simons[2]

[1] Department of Mathematical Sciences, Loyola University, New Orleans, LA 70118, USA
[2] Department of Mathematics, University of California, Santa Barbara, CA 93106, USA

0. Introduction

This paper originated from a study of Browder's version [3] of Banach's closed range theorem for (not necessarily continuous) linear operators between locally convex spaces. It seems that the most reasonable way to proceed is to perform a simultaneous development of the closed range theorem and the open mapping theorem. Insofar as the open mapping theorem is concerned, our results largely parallel those of Köthe [5, 7], but our proofs are less technical and (see Theorem 11) we can dispense with the local convexity of the spaces concerned. Furthermore (see Remark 19) our analysis throws light on the result obtained in [7] that, in the locally convex case, every open map is weakly singular. Insofar as the closed range theorem is concerned, our results represent improvements of the results of Browder [3] and Baker [2]. See the discussion in Sect. 8.

Our presentation leans heavily on the properties of seminorms. We use the characterization of barreled spaces in terms of lower semicontinuous seminorms (see Sect. 1) and our other main tools are the concepts of *Mackey seminorm* (see Sect. 1), the quotient of a seminorm by a linear map (see Sect. 2) and *adequate map* (see Definition 7 and Remark 8). These tools obviate our having to deal with the more abstract concept of a quotient topological vector space.

We introduce two definitions of the adjoint of a linear map. The "small adjoint," introduced in Sect. 3, enables us to give the most succinct statements for the results in Sects. 3–5. The "large adjoint," introduced in Sect. 6, corresponds to the definition usually made.

In Sect. 7 we specialize to the case when local convexity is assumed for the range space (and sometimes also for the domain space).

A number of the results in this paper assume that the range of the operator under consideration is metrizable. We discuss weakenings of this condition in Sect. 8.

In order to keep our treatment short and simple, we have not touched on of the many other interesting concepts dealt with by some authors (e.g., Köthe [5, 7], Baker [2], Mennicken and Sagraloff [8]) such as nearly open operators, weakly open operators, etc.

Other authors who have contributed to the literature include Dieudonné and Schwartz [4], Mochizuki [9], Pták [10], Schaefer [12], and Treves [13, 14].

1. Preliminaries

If E is a real topological vector space, we write E^* for the algebraic dual of E and E' for the topological dual of E. If P is a seminorm on E, we write

$$E_P^* := \{a : a \in E^*, a \leqq P \text{ on } E\}$$

and

$$E'_P := \{a : a \in E', a \leqq P \text{ on } E\}.$$

From the Banach-Alaoglu theorem, E_P^* is $w(E^*, E)$-compact.

We say that E is *barreled* if every lsc (lower semicontinuous) seminorm on E is continuous (see [6, 21.2(2)–(3), p. 257]). (We do not assume that E is necessarily Hausdorff or locally convex.) The authors are grateful to the referee for pointing out that if E is ultrabarreled (in the sense of [11]) then E is barreled and, from [11, Sect. 7, p. 256] the converse is not true even if E is locally convex.

Lemma 1. *Let E be barreled and Q be any seminorm on E. Then E'_Q is $w(E', E)$-compact.*

Proof. Define $R: E \to \mathbb{R}$ by $R := \sup\{a : a \in E'_Q\}$. R is a lsc seminorm on E and $R \leqq Q$. Since E is barreled, R is continuous. The result follows since $E'_Q = E_R^*$.

We shall say that P is a *Mackey seminorm on* E if $E_P^* \subset E'$. [If E is Hausdorff and locally convex, then P is Mackey $\Leftrightarrow P$ is continuous with respect to the Mackey topology $\tau(E, E')$. However, our definition does *not* require that E be either Hausdorff or locally convex.]

Lemma 2. *Any Mackey seminorm is lsc.*

Proof. This follows since, from the Hahn-Banach theorem, any seminorm is the supremum of the linear functionals that it dominates.

Lemma 3. *Let E be pseudometrizable. Then any Mackey seminorm on E is continuous. (This is the counterpart of the result that any metrizable locally convex space has the Mackey topology.)*

Proof. Suppose, on the contrary, that P is a discontinuous Mackey seminorm on E. Then $\exists \{x_n\}_{n \geqq 1} \subset E$ such that $x_n \to 0$ and, $\forall n \geqq 1$, $P(x_n) \geqq 1$. Let $\{U_n\}_{n \geqq 1}$ be a decreasing neighborhood base at 0. Then $\exists$ a subsequence $\{y_n\}_{n \geqq 1}$ of $\{x_n\}_{n \geqq 1}$ such that, $\forall n \geqq 1$, $y_n \in 3^{-n} U_n$. Thus, writing $z_n := 2.3^n y_n$, $z_n \to 0$ and $P(z_n) \geqq 2.3^n$.

We define $n_1 := 1$ and choose $b_1, n_2, b_2, \ldots$ inductively as follows: (using the Hahn-Banach theorem) $b_k \in E_P^* \subset E'$ such that

$$\langle z_{n_k}, b_k \rangle = P(z_{n_k}) \geqq 2.3^k$$

and (since $z_n \to 0$) $n_k > n_{k-1}$ such that

$$\left| \left\langle z_{n_k}, \sum_{j=1}^{k-1} 3^{-j} b_j \right\rangle \right| < \frac{1}{2}.$$

Write $b := 2 \sum_{j=1}^{\infty} 3^{-j} b_j \in E_P^* \subset E'$. Then, $\forall k \geqq 1$,

$$\langle z_{n_k}, b\rangle \geqq 2.3^{-k}\langle z_{n_k}, b_k\rangle + \sum_{j=k+1}^{\infty} 2.3^{-j}\langle z_{n_k}, b_j\rangle - 2\cdot\frac{1}{2}$$

$$\geqq \left(2.3^{-k} - \sum_{j=k+1}^{\infty} 2.3^{-j}\right) P(z_{n_k}) - 1 \geqq 1\,,$$

which is a contradiction since $z_{n_k} \to 0$.

2. Semi-Open Linear Maps

We shall suppose for the rest of this paper that E and F are real Hausdorff topological vector spaces and $T: E \to F$ is linear (but not necessarily continuous). We do not assume that E or F is locally convex until Sect. 7.

If P is a seminorm on E we define the seminorm P/T on $T(E)$ by

$$P/T(y) := \inf P(T^{-1}y)\,(y \in T(E))\,.$$

Lemma 4. *Let $b \in T(E)^*$. Then $b \in T(E)^*_{P/T} \Leftrightarrow b \circ T \in E^*_P$.*

Proof. $(\Rightarrow)$ $\forall x \in E$, $\langle x, b\circ T\rangle = \langle Tx, b\rangle \leqq P/T(Tx) = \inf P(T^{-1}Tx) \leqq P(x)$.

$(\Leftarrow)$ $\forall y \in T(E)$ and $x \in T^{-1}y$, $\langle y, b\rangle = \langle Tx, b\rangle = \langle x, b\circ T\rangle \leqq P(x)$. Taking the infimum over x, $\langle y, b\rangle \leqq P/T(y)$.

Definition 5. We shall say that T is *semi-open* if P/T is a continuous seminorm on $T(E)$ whenever P is a continuous seminorm on E.

Lemma 6. (6.1) $\Rightarrow$ (6.2) $\Rightarrow$ (6.3) $\Rightarrow$ (6.4). *If $T(E)$ is barreled then* (6.1)–(6.4) *are equivalent. If $T(E)$ is metrizable then* (6.1)–(6.3) *are equivalent.*

(6.1) *T is semi-open.*

(6.2) *If $b \in T(E)^*$ and $b \circ T \in E'$ then $b \in T(E)'$.*

(6.3) *If P is a continuous seminorm on E then P/T is Mackey on $T(E)$.*

(6.4) *If P is a continuous seminorm on E then P/T is lsc on $T(E)$.*

Proof. ((6.1) $\Rightarrow$ (6.2)). Let $b \in T(E)^*$ and $b \circ T \in E'$. Since T is semi-open, setting $P := |\langle\cdot, b\circ T\rangle|$, $P/T = |\langle\cdot, b\rangle|$ is continuous on $T(E)$, hence $b \in T(E)'$. Thus (6.2) is true.

((6.2) $\Rightarrow$ (6.3)). Let $b \in T(E)^*_{P/T}$. From Lemma 4, $b \circ T \in E^*_P \subset E'$. From (6.2), $b \in T(E)'$. Hence P/T is Mackey.

It is immediate from Lemma 2 that (6.3) $\Rightarrow$ (6.4).

If $T(E)$ is barreled then it is immediate that (6.4) $\Rightarrow$ (6.1). If $T(E)$ is metrizable then it is immediate from Lemma 3 that (6.3) $\Rightarrow$ (6.1).

3. Connections with the Range of the Small Adjoint

We define $\Delta := \{\delta : \delta \in \overline{T(E)}', \delta \circ T \in E'\} \subset \overline{T(E)}'$ and $T^\tau : \Delta \to E'$ by

$$T^\tau \delta := \delta \circ T \;(\delta \in \Delta)\,.$$

[It is understood in this that $\overline{T(E)}$ has the topology induced from the topology of F.] We define $G := \{(x, Tx) : x \in E\} \subset E \times F$.

Definition 7. We shall say that T is *adequate* if

$$x \in E, a \in E', \langle x, T^\tau(\Delta)\rangle = \{0\} \quad \text{and} \quad \langle T^{-1}(0), a\rangle = \{0\} \Rightarrow \langle x, a\rangle = 0.$$

Remark 8. T is clearly adequate if Δ separates the points of $T(E)$ or $T^{-1}(0)$ is dense in E – in particular if F is locally convex and T is continuous. The condition (8.1) below also ensures that T is adequate.

$$(x, 0) \in G^{00} \Rightarrow x \in T^{-1}(0)^{00}, \tag{8.1}$$

where ^{0} stands for the operation of polarity (orthogonality). To see this, suppose that $\langle x, T^\tau(\Delta)\rangle = \{0\}$ and $\langle T^{-1}(0), a\rangle = \{0\}$. If $(c, b) \in G^0 \subset E' \times F'$ then $b \circ T = -c \in E'$ hence $\delta := b|_{\overline{T(E)}} \in \Delta$ and $c = -\delta \circ T = -T^\tau\delta$. Thus

$$\langle (x, 0), (c, b)\rangle = \langle x, c\rangle = -\langle x, T^\tau\delta\rangle = 0,$$

that is to say, $(x, 0) \in G^{00}$. From (8.1), $x \in T^{-1}(0)^{00}$ hence $\langle x, a\rangle = 0$.

If F is locally convex then the above argument can be reversed and gives that (8.1) $\Leftrightarrow$ T is adequate. If E and F are both locally convex then, from the above and the bipolar theorem, T is adequate $\Leftrightarrow$ T is weakly singular (see Remark 19). The authors are grateful to the referee for pointing out this last fact.

Lemma 9. *We consider the condition:*

$$T^\tau(\Delta) \quad \textit{is} \quad w(E', E)\textit{-closed}. \tag{9.1}$$

Then (6.2) $\Rightarrow$ (9.1). *If T is adequate then* (6.2) $\Leftrightarrow$ (9.1).

Proof. ((6.2) $\Rightarrow$ (9.1)). Let $a \in E'$ and suppose that $\exists$ a net $\delta_\alpha \in \Delta$ such that $T^\tau\delta_\alpha \to a$ in $w(E', E)$: we shall prove that $a \in T^\tau(\Delta)$. Let $x \in T^{-1}(0)$. Then

$$\langle x, a\rangle = \lim_\alpha \langle x, T^\tau\delta_\alpha\rangle = \lim_\alpha \langle Tx, \delta_\alpha\rangle = \lim_\alpha \langle 0, \delta_\alpha\rangle = 0.$$

Thus we have proved that if $x \in T^{-1}(0)$ then $\langle x, a\rangle = 0$. It follows algebraically that $\exists b \in T(E)^*$ such that $b \circ T = a$. From (6.2), $b \in T(E)'$ and, from standard extension arguments, $\exists \delta \in \overline{T(E)}'$ such that $\delta|_{T(E)} = b$, hence $\delta \circ T = a$. Then $\delta \in \Delta$ and $T^\tau\delta = a$, from which $a \in T^\tau(\Delta)$, as required.

((9.1) $\Rightarrow$ (6.2)) Let $b \in T(E)^*$ and $b \circ T \in E'$. Suppose that $x \in E$ and $\langle x, T^\tau(\Delta)\rangle = \{0\}$. Since T is adequate and $\langle T^{-1}(0), b \circ T\rangle = \{0\}$, $\langle x, b \circ T\rangle = 0$. Thus we have proved that

$$x \in E \text{ and } \langle x, T^\tau(\Delta)\rangle = \{0\} \Rightarrow \langle x, b \circ T\rangle = 0.$$

From (9.1) and the separation theorem, $b \circ T \in T^\tau(\Delta)$, hence $\exists \delta \in \Delta$ such that $T^\tau\delta = b \circ T$. Then $b = \delta|_{T(E)} \in T(E)'$. Thus (6.2) is true, as required.

Theorem 10. *Let T be adequate. If $T(E)$ is barreled then* (6.1)–(6.4) *and* (9.1) *are equivalent. If $T(E)$ is metrizable then* (6.1)–(6.3) *and* (9.1) *are equivalent.*

Proof. This is immediate from Lemmas 6 and 9.

4. The Semi-Open Mapping Theorem

We recall that E is a *Pták space* if a subspace L of E' is $w(E', E)$-closed whenever, $\forall$ continuous seminorms P on E, $L \cap E_P^*$ is $w(E', E)$-compact. (We do not assume that E is necessarily locally convex. See Remarks 15 for further comments on Pták spaces.)

Theorem 11 (Semi-Open Mapping Theorem). *If E is a Pták space, T is adequate and $T(E)$ is barreled then T is semi-open and $T^\tau(\Delta)$ is $w(E', E)$-closed.*

Proof. Let P be a continuous seminorm on E. From standard extension arguments and Lemma 4,

$$T^\tau(\Delta) \cap E_P^* = \{b \circ T : b \in T(E)', b \circ T \leqq P\} = \{b \circ T : b \in T(E)'_{P/T}\};$$

from Lemma 1 and the $w(T(E)', T(E)) - w(E^*, E)$ continuity of the map $b \to b \circ T$, this set is $w(E^*, E)$-compact. Since E is a Pták space, $T^\tau(\Delta)$ is $w(E', E)$-closed. The result now follows from Theorem 10.

5. Connections with the Range of T

If $z \in \overline{T(E)}$ we define $\Delta_z := \{\delta : \delta \in \Delta, \langle z, \delta \rangle = 0\}$.

Lemma 12. *If T is semi-open and $z \in \overline{T(E)}$ then, for all continuous seminorms P on E, $T^\tau(\Delta_z) \cap E_P^*$ is $w(E', E)$-compact.*

Proof. Let P be a continuous seminorm on E. Since T is semi-open P/T is (uniformly) continuous on $T(E)$. From standard extension arguments, $\exists$ a continuous seminorm R on $\overline{T(E)}$ such that $R|_{T(E)} = P/T$. We now prove that

$$(12.1) \qquad T^\tau(\Delta_z) \cap E_P^* = \{\delta \circ T : \delta \in \overline{T(E)}_R^*, \langle z, \delta \rangle = 0\}.$$

Let $a \in T^\tau(\Delta_z) \cap E_P^*$. Then $\exists \delta \in \Delta$ such that $\langle z, \delta \rangle = 0$ and $T^\tau \delta = a$. From Lemma 4 ($\Leftarrow$), $\delta|_{T(E)} \in T(E)_{P/T}^*$. Since both δ and R are continuous on $\overline{T(E)}$, $\delta \in \overline{T(E)}_R^*$. This establishes ($\subset$) in (12.1). Conversely, let $\delta \in \overline{T(E)}_R^*$ and $\langle z, \delta \rangle = 0$. Since $\delta|_{T(E)} \in T(E)_{P/T}^*$, from Lemma 4 ($\Rightarrow$), $\delta \circ T \in E_P^* \subset E'$, from which $\delta \in \Delta$, hence $\delta \in \Delta_z$. This establishes ($\supset$) in (12.1). The result now follows from (12.1), the Banach-Alaoglu theorem and the $w(\overline{T(E)}^*, \overline{T(E)}) - w(E^*, E)$ continuity of the map $\delta \to \delta \circ T$.

Lemma 13. *If Δ separates the points of $\overline{T(E)}$ and, $\forall z \in \overline{T(E)}$, $T^\tau(\Delta_z)$ is $w(E', E)$-closed then $T(E)$ is closed in F.*

Proof. Let $z \in \overline{T(E)} \setminus T(E)$. Clearly $z \neq 0$. Since Δ separates the points of $\overline{T(E)}$, $\exists \delta_0 \in \Delta$ such that $\langle z, \delta_0 \rangle = 1$. If now $\delta \in \Delta_z$ then $\langle z, \delta \rangle \neq \langle z, \delta_0 \rangle$ hence [since $z \in \overline{T(E)}$ and both δ_0 and δ are continuous on $\overline{T(E)}$] $\delta|_{T(E)} \neq \delta_0|_{T(E)}$, from which $T^\tau \delta \neq T^\tau \delta_0$. Thus $T^\tau \delta_0 \notin T^\tau(\Delta_z)$. By hypothesis and the separation theorem, $\exists x \in E$ such that $\langle x, T^\tau \delta_0 \rangle = 1$ and

$$(13.1) \qquad \delta \in \Delta_z \Rightarrow \langle x, T^\tau \delta \rangle = 0.$$

Let $\delta \in \Delta$. Since $\langle z, \delta \rangle \delta_0 - \delta \in \Delta_z$, from (13.1) and the fact that $\langle x, T^\tau \delta_0 \rangle = 1$,

$$0 = \langle x, T^\tau(\langle z, \delta \rangle \delta_0 - \delta) \rangle = \langle z - Tx, \delta \rangle.$$

Thus we have proved that

$$\forall \delta \in \Delta, \langle z - Tx, \delta \rangle = 0.$$

Since Δ separates the points of $\overline{T(E)}$, $z = Tx \in T(E)$. This contradiction of the assumption that $z \notin T(E)$ establishes that $\overline{T(E)} \setminus T(E) = \emptyset$, hence $T(E)$ is closed, as required.

Theorem 14 (First Half of the Closed Range Theorem). *Let E be a Pták space and Δ separate the points of $\overline{T(E)}$. Then*

$$T \textit{ is semi-open} \Rightarrow T(E) \textit{ is closed in } F.$$

If, further, $T(E)$ is barreled or metrizable then

$$T^{\tau}(\Delta) \textit{ is } w(E', E)\textit{-closed} \Rightarrow T \textit{ is semi-open}.$$

Proof. This is immediate from Theorem 10 and Lemmas 12 and 13.

Remarks 15. The argument of this section is related to Grothendieck's completeness theorem, but the argument of Theorem 14 does really establish that $T(E)$ is *closed.* We do not know whether, under the conditions of Theorem 14, the conclusion can be strengthened to "$T(E)$ is complete." We observe that it is only the topology of $T(E)$ (as a subspace of F) that is at issue in Theorem 14. We do not otherwise have to be concerned with the topology of F. Theorem 14 complements [7, 37.5 (4), p. 104], which requires that E be locally convex.

The authors are grateful to the referee for the following observation. Let $E = l^{1/2}$, $F = l^1$, and $T: E \to F$ be the canonical inclusion. Then $T(E)$ is metrizable and (being dense) is not closed in F. Further, $\Delta = F' = l^{\infty}$ and $T^{\tau}(\Delta) = E'$ which is $w(E', E)$-closed in E'. Thus, from Theorem 14, E is *not* a Pták space. This example shows that a complete metrizable topological vector space is not necessarily a Pták space. Combining this with [1, 10 (7), p. 54] we see that a B-complete space (in the sense of [1]) is not necessarily a Pták space. On the other hand, if $E' = \{0\}$ then E is a Pták space; consequently, it is easy to give an example of a Pták space that is not complete. Combining this with [1, 10 (9), p. 56] we see that a Pták space (in the sense of [1]) is not necessarily B-complete.

6. The Range of the Large Adjoint

Remark 16. The map T^{τ} with domain Δ can be thought of as an adjoint map for T. This definition is obviously the appropriate one for the analysis of Sects. 3–5. It is, however, more usual to define the domain of the adjoint by

$$D := \{d : d \in F', d \circ T \in E'\} \subset F'$$

and the adjoint $T^t : D \to E'$ by

$$T^t d := d \circ T \quad (d \in D).$$

It is this distinction that prompted us to use the phrases "small adjoint" and "large adjoint" in the section headings for Sects. 3 and this section. If the following condition is satisfied:

(16.1) If $\delta \in \overline{T(E)}'$ and $\delta \circ T \in E'$ then $\exists d \in F'$ such that $d|_{\overline{T(E)}} = \delta$

then $T^t(D) = T^t(\Delta)$. In this case, (9.1) is equivalent to:

(16.2) $$T^t(D) \quad \text{is} \quad w(E', E)\text{-closed}.$$

For the analysis of the next section, it is sufficient to observe that (16.1) is satisfied if F is locally convex.

7. The Locally Convex Case

In this section we suppose that F is locally convex. We shall also suppose for much of this section that E is locally convex.

Lemma 17. *We consider the condition*

(17.1) *T is an open map from E onto $T(E)$.*

Then (17.1) $\Rightarrow$ (6.1). *If E is locally convex then* (17.1) $\Leftrightarrow$ (6.1).

Proof. This follows from the identity

$$\{y : y \in T(E), P/T(y) < 1\} = T\{x : x \in E, P(x) < 1\},$$

which is immediate from the definition of P/T.

Theorem 18 (On the Equivalence of Conditions). *Let E be locally convex and T be adequate. If $T(E)$ is either barreled or metrizable then* (6.2), (6.3), (16.2), *and* (17.1) *are equivalent.*

Proof. This is immediate from Theorem 10 and Lemma 17.

Remark 19 (On Weak Singularity). T is said to be *weakly singular* (see [7, 36.1 (7), p. 81]) if

$$(x, 0) \in \bar{G} \Rightarrow x \in \overline{T^{-1}(0)}.$$

We now show that if (6.4) is satisfied and E is locally convex then T is weakly singular. Suppose that $(x, 0) \in \bar{G}$. Then $\exists$ a net x_α in E such that $x_\alpha \to x$ and $Tx_\alpha \to 0$. Let P be a continuous seminorm on E. Then, eventually, $P(x_\alpha - x) \leqq 1/2$ hence $P/T(Tx_\alpha - Tx) \leqq 1/2$. From (6.4), P/T is lsc on $T(E)$. Thus, since $Tx_\alpha \to 0$, $P/T(-Tx) \leqq 1/2$. Hence

$$\exists w \in -T^{-1}Tx \quad \text{such that} \quad P(w) \leqq 1.$$

Let $u = x + w$. Then $u \in T^{-1}(0)$ and $P(u - x) \leqq 1$. Since this holds for all P and E is locally convex, $x \in \overline{T^{-1}(0)}$.

It follows from these considerations that if E is locally convex and Prák and F is barreled then the implication (6.4) $\Rightarrow$ (17.1) can also be deduced from [7, 37.5 (5), p. 104]. Furthermore, the implications (17.1) $\Rightarrow$ (6.1) $\Rightarrow$ (6.2) $\Rightarrow$ (6.3) $\Rightarrow$ (6.4) [which hold even if $T(E)$ is not barreled or metrizable] throw light on the result of [7, 37.4 (1), p. 100] that, in the locally convex case, every open map is weakly singular.

Theorem 20 (Open Mapping and the Other Half of the Closed Range Theorem). *If E is a locally convex Pták space, T is adequate ($=$ weakly singular) and $T(E)$ is barreled then T is an open map from E onto $T(E)$ and $T^t(D)$ is $w(E', E)$-closed.*

Proof. This is immediate from Theorem 11 and Lemma 17.

Theorem 21 (Surjective Open Mapping Theorem). *If E is a locally convex Pták space, F is barreled T is adequate (= weakly singular) and $T(E)=F$ then T is an open map.*

Proof and remarks. This is immediate from Theorem 20. This is a special case of [7, 37.5 (5), p. 104]. If T is continuous we obtain Pták's homomorphism theorem (see [12, IV.8.3, Corollary 1, p. 164]), since a continuous linear map into a locally convex space is adequate.

Theorem 22 (Semi-Open Mapping and Closed Range Theorem). *Let E be a Pták space and D separate the points of $\overline{T(E)}$. If $T(E)$ is barreled or metrizable then*

$$T^t(D) \text{ is } w(E', E)\text{-closed} \Rightarrow T \text{ is semi-open} \Rightarrow T(E) \text{ is closed in } F.$$

If every closed subspace of F is barreled then

$$T(E) \text{ is closed in } F \Leftrightarrow T \text{ is semi-open} \Rightarrow T^t(D) \text{ is } w(E', E)\text{-closed}.$$

Proof and Remarks. This is immediate from Theorem 11 and Theorem 14. When we compare this result with [3, Theorem 2.1, p. 65] and [2, Theorem 8, p. 283], we observe that we do not require E to be locally convex and that Lemma 3 has enabled us to avoid considering "condition (t)." See the discussion in the next section.

8. More General Versions of our Results

The proof of Lemma 3 shows that, in all cases in this paper where we have assumed that $T(E)$ is metrizable, we could equally well have assumed the weaker condition [akin to that of $T(E)$ being bornological]: if Q is a seminorm on $T(E)$ such that $\{Q(z_n): n \geqq 1\}$ is bounded whenever $\{z_n\}_{n \geqq 1} \subset T(E)$ and $z_n \to 0$ then Q is continuous on $T(E)$.

We can obtain a more radical generalization of the first half of the closed range theorem by modifying Lemma 12 rather than Lemma 3. To this end, let us say that a subspace H of F is a *μ-subspace* if every Mackey seminorm on H can be extended to a Mackey seminorm on $\bar{H}$. We can now prove the following modification of Lemma 12: if (6.3) is satisfied, $T(E)$ is a μ-subspace of F and $z \in \overline{T(E)}$ then, for all continuous seminorms P on E, $T^\tau(\Delta_z) \cap E_P^*$ is $w(E', E)$-compact. This leads to the following modification of Theorem 14: let E be a Pták space, Δ separate the points of $\overline{T(E)}$ and $T(E)$ be a μ-subspace of F. Then

$$T^\tau(\Delta) \text{ is } w(E', E)\text{-closed} \Rightarrow T(E) \text{ is closed in } F,$$

with a corresponding restatement in terms of $T^t(D)$ if F is locally convex. In seminorm terms, *F satisfies condition* (t) (see [2] and [3]) if every Mackey seminorm on every subspace of F can be extended to a Mackey seminorm on F. In particular, if F satisfies condition (t) then *every* subspace of F is a μ-subspace. Consequently, the above result generalizes the relevant part of [2, Theorem 8, p. 283] and [3, Theorem 2.1, p. 65]. Of course, it is requiring much less to ask for an extension from $T(E)$ to $\overline{T(E)}$ than it is to ask for an extension from every subspace of F to F. Furthermore, in view of the close relationship between a subspace and its closure, there is some hope of a reasonable internal characterization of μ-subspace.

References

1. Adasch, N., Ernst, B., Keim, D.: Topological vector spaces. Lecture Notes Math., Vol. 639. Berlin Heidelberg New York: Springer 1978
2. Baker, J.W.: Operators with closed range. Math. Ann. **174**, 278–284 (1967)
3. Browder, F.E.: Functional analysis and partial differential equations, I. Math. Ann. **138**, 55–79 (1959)
4. Dieudonné, J., Schwartz, L.: La dualité dans les espaces (F) et (LF). Ann. Inst. Fourier **1**, 61–101 (1949)
5. Köthe, G.: General linear transformations of locally convex spaces. Math. Ann. **159**, 309–328 (1965)
6. Köthe, G.: Topological vector spaces, I. Berlin Heidelberg New York: Springer 1969
7. Köthe, G.: Topological vector spaces, II. Berlin Heidelberg New York: Springer 1979
8. Mennicken, R., Sagraloff, B.: Characterizations of nearly-openness. J. Reine Angew. Math. **313**, 105–115 (1980)
9. Mochizuki, N.: On fully complete spaces. Tôhoku Math. J., II. Ser., **13**, 485–490 (1961)
10. Pták, V.: Completeness and the open mapping theorem. Bull. Soc. Math. France **86**, 41–74 (1958)
11. Robertson, W.: Completions of topological vector spaces. Proc. London Math. Soc. **8**, 242–257 (1958)
12. Schaefer, H.H.: Topological vector spaces. Berlin Heidelberg New York: Springer 1986
13. Treves, F.: Locally convex spaces and linear partial differential equations. Berlin Heidelberg New York: Springer 1967
14. Treves, F.: Topological vector spaces. Distributions and kernels. New York: Academic Press 1967

Received September 25, 1987; in revised form April 5, 1988

Math. Ann. 283, 97–119 (1989)

Mathematische Annalen

Asymptotics for some Green Kernels on the Heisenberg Group and the Martin Boundary

H. Hueber and D. Müller

Fakultät für Mathematik, Universität Bielefeld, Postfach 8640, D-4800 Bielefeld, Federal Republic of Germany

Introduction

The Heisenberg group and especially the sub- (or Kohn-) Laplacian Δ_K on this group has been studied extensively by many authors during at least the last decade, and it would be beyond the scope of this article to even try to list the possible applications and results related to this study. Let us only note that Δ_K is a prototype of a so-called "subelliptic" operator, a notion which is perhaps not well-defined but commonly used for operators such as the "sum of squares" studied by Hörmander [10], or for the so-called "Rockland operators" studied by Rockland [13], Helffer-Nourrigat [7] and others. Therefore, it is desirable to have as much explicit information about Δ_K as possible, and quite a bit is in fact known by now. For example, Folland was the first to calculate an explicit fundamental solution for Δ_K, and Gaveau (see [6]), using stochastic integration, as well as Hulanicki [11] and Cygan [3], making use of the representation theory of the Heisenberg group, derived a formula for the fundamental solution of the corresponding heat operator $\partial/\partial s-\Delta_K$. However, although this formula suffices to give a canonical way of calculating a fundamental solution for Δ_K, it has one decisive drawback: it is explicit only up to the partial Fourier transform along the center of the Heisenberg group, and it seems very unlikely that one might be able to carry through this Fourier transform explicitly. Therefore the best one might hope for is to be able to describe the asymptotic behaviour of this fundamental solution. A partial solution to this problem has already been given by Gaveau ([6]; compare also Theorem 1.1), but his results did not cover regions which are in some sense "close" to the center of the Heisenberg group.

In this article, we shall present a complete picture of the asymptotics for the fundamental solution of the heat operator. Moreover, we shall apply these results resp. methods in order to solve the analogous problem for the operator $\Delta_K-\mu$, μ a positive real number. Finally, in Sect. 3 we shall use these descriptions of the asymptotic behaviour in order to prove that the so-called Martin boundary corresponding to Δ_K-2 of the Heisenberg group is homeomorphic to the closed unit disk in $\mathbb{C}$. Moreover, we can show that the minimal Martin boundary, that is the space of extremal rays of the cone of all positive solutions h of the equation

$(\Delta_K - 2)h = 0$, is homeomorphic to that part of the Martin boundary that corresponds to the unit circle. These results are in sharp contrast to the classical situation of the Laplace operator on $\mathbb{R}^n$, where the Martin boundary corresponding to $\Delta - 2$ is homeomorphic to the unit sphere S^{n-1} in $\mathbb{R}^n$. A consequence of our result is that each positive eigenfunction of Δ_K corresponding to a positive eigenvalue is independent of the central variable of the Heisenberg group.

0. Preliminaries

(a) In order to avoid a few additional technical problems which arise for higher-dimensional Heisenberg groups and to fix the ideas, we shall only deal with the 3-dimensional Heisenberg group H_1.

H_1 is the 2-step nilpotent Lie group whose underlying real manifold is $\mathbb{C} \times \mathbb{R}$, and where multiplication is given by

$$(z, u)(z', u') = (z + z', u + u' + 2\,\mathrm{Im}\, z \cdot \bar{z}')$$

for (z, u), $(z', u') \in \mathbb{C} \times \mathbb{R}$. We introduce real coordinates (x, y, u) for H_1 by writing $z = x + iy$. Let X, Y, U denote the left-invariant vector fields on H_1 whose values at the neutral element 0 are given by $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial u$ respectively, i.e.:

$$X = \frac{\partial}{\partial x} + 2y\frac{\partial}{\partial u}, \qquad Y = \frac{\partial}{\partial y} - 2x\frac{\partial}{\partial u}, \qquad U = \frac{\partial}{\partial u}.$$

Then we have $[X, Y] = -4U$, and hence it follows for example from [10] that the sub-Laplacian

$$\Delta_K = X^2 + Y^2$$

is hypoelliptic. With respect to the coordinates (x, y, u), Δ_K is explicitly given by

$$\Delta_K = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + 4\left[y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y}\right]\frac{\partial}{\partial u} + 4(x^2 + y^2)\frac{\partial^2}{\partial u^2}.$$

The group of automorphisms of H_1 contains two canonical subgroups of outermorphisms, namely the maximal compact subgroup $SU(2) \simeq \mathbb{T}$, which acts on H_1 by ${}^{\varphi}(z, u) := (e^{i\varphi}z, u)$, $0 \leqq \varphi < 2\pi$, and the group $\mathbb{R}^+$, which acts by dilations $D_r(z, u) := (rz, r^2u)$, $r > 0$. It is easily seen that Δ_K commutes with the action of $SU(2)$, and that

$$\Delta_K(f \circ D_r) = r^2(\Delta_K f) \circ D_r$$

for all $f \in C^2(H_1)$ and $r > 0$. This shows especially that, by a suitable dilation, the operator $\Delta_K - 2$ can be transformed into a multiple of the operator $\Delta_K - \mu$ for any $\mu > 0$. Therefore, it will be no restriction that in Sect. 2 we shall only consider the operator $\Delta_K - 2$ instead of $\Delta_K - \mu$.

(b) We shall frequently use the following standard notation from asymptotic analysis:

If X is a locally compact Hausdorff space, and if A is a subset of X, then we say that two complex functions f and g on X are *asymptotically equivalent* for x in A as x tends to infinity, if for any $\varepsilon > 0$ there exists a compact subset B_ε of X, such that $|f(x) - g(x)| \leqq \varepsilon \min(|f(x)|, |g(x)|)$ for every $x \in A \setminus K$. In this case we write

$$f(x) \sim g(x)$$

for $x \in A$ as $x \to \infty$. Note that if $f(x)$ and $g(x)$ are nonzero for sufficiently large x, then this just means that the quotient $f(x)/g(x)$ tends to 1 as $x \in A$ tends to infinity. In some statements, we shall pose some additional restrictions on the range of validity for an asymptotic equivalence by demanding that certain functions of x tend to certain limits as x tends to infinity.

Note that in case of the Heisenberg group $X = H_1$, there is a natural topology on H_1 which is defined by any homogeneous norm (in the sense of [5]) on H_1, for example by

$$|(z,u)| = (|z|^2 + |u|)^{1/2} \quad \text{or by} \quad \|(z,u)\| = (|z|^4 + u^2)^{1/4}.$$

This topology agrees with the Euclidean topology of the space $\mathbb{C} \times \mathbb{R}$. Let us mention here that the "norm" $\|\cdot\|$ is of special importance, because Folland's fundamental solution for Δ_K is just given by $-\frac{1}{4\pi}\|(z,u)\|^{-2}$.

1. Asymptotic Estimates for the Heat-Semigroup

Let us denote by $p_s(z,u)$ the positive fundamental solution of the heat operator $\frac{\partial}{\partial s} - \frac{1}{2}\Delta_K$ on $\mathbb{R} \times H_1$. It is well-known (cf. [6] or [11]) that p_s is given explicitly (for $s > 0$) by

$$p_s(z,u) = \frac{1}{2(2\pi s)^2} \int_{\mathbb{R}} \exp\left(i\frac{xu}{2s} - \frac{|z|^2}{2s} x \coth x\right) \frac{x}{\sinh x}\, dx,$$

and that $\{p_s\}_{s>0}$ is a probability semigroup on H_1.

If we set $p := p_1$, we have especially

$$p_s(z,u) = \frac{1}{s^2}\, p\left(\frac{z}{\sqrt{s}}, \frac{u}{s}\right). \tag{1.1}$$

We set $R := |z|^2$, and

$$h(R,u) := \int_{\mathbb{R}} e^{iux - Rx\coth x} \frac{x}{\sinh x}\, dx, \tag{1.2}$$

hence

$$p(z,u) = \frac{1}{2(2\pi)^2}\, h\left(\frac{R}{2}, \frac{u}{2}\right). \tag{1.3}$$

We shall first describe the asymptotic behaviour of $p(z,u)$ as $|z|^2 + |u|$ tends to infinity for the case, that the ratio $\omega = |u|/|z|^2$ stays bounded from above.

As in [6] we define a function

$$\theta :]-\pi, \pi[\to \mathbb{R} \tag{1.4}$$

by $\theta(y) = (2y - \sin 2y)/(2\sin^2 y)$. Obviously, θ is an odd function, and in [6] it has been shown that θ is a strictly increasing diffeomorphism. Let $\tau : \mathbb{R} \to]-\pi, \pi[$ be the inverse function of θ, i.e.

$$\tau = \theta^{-1}. \tag{1.5}$$

The following theorem follows by a somewhat technical, but straight-forward modification of the proof of Theorem 2 (3°) in [6] (and a correction by a factor 1/2):

Theorem 1.1. *If the ratio* $\omega=|u|/|z|^2$ *tends towards a limit* $\omega_\infty<\infty$, *then* $p(z,u)$ *is asymptotically given by*

$$p(z,u)\sim\Phi(\omega_\infty)\frac{1}{|z|}e^{-\gamma^2\left(\frac{|u|}{|z|^2}\right)\frac{|z|^2}{2}}$$

as $|z|^2+|u|\to\infty$.

Here, γ *and* Φ *are given by*

$$\gamma(\omega)=\tau(\omega)/\sin\tau(\omega),$$

$$\Phi(\omega)=\frac{1}{2(2\pi)^{3/2}}|\tau(\omega)|\left[\frac{\sin\tau(\omega)}{\sin\tau(\omega)-\tau(\omega)\cos\tau(\omega)}\right]^{1/2}, \tag{1.6}$$

if $\omega\neq0$, *and* $\gamma(0)=1$, $\Phi(0)=\dfrac{3^{1/2}}{2(2\pi)^{3/2}}$.

Now we turn to the case where the ratio $|u|/|z|^2$ tends to infinity, or, equivalently, where

$$\delta:=\sqrt{R/(\pi|u|)} \tag{1.7}$$

tends to zero.

Note, that we may assume $u>0$, since $p(z,u)=p(z,-u)$, and that $\delta\to0$ and $R+|u|\to\infty$ imply $|u|\to\infty$.

So let us assume that $u\to+\infty$, and $\delta\to0$. In this case, the proof of Theorem 2 (3°) in [6] breaks down for the following reason: The proof is based on the stationary phase method. However, in case that $\delta\to0$, the critical point in the complex plane of the phase $iuz-Rz\coth z$ of the integral (1.2) approaches the point $i\pi$, where this phase has a simple pole. But, as we shall see, it is possible to transform the integral (1.2), modulo an error term, into another one, to which the stationary phase method is again applicable.

In order to prepare the next theorem, let us introduce the following functions:

For complex $z\in\mathbb{C}$ with $|z|<1$ let

$$r(z):=1+\frac{1}{z}-(z+1)\pi\cot\pi z, \tag{1.8}$$

if $z\neq0$, and $r(0)=0$. Since $\pi\cot\pi z$ has only one simple pole with residue 1 at $z=0$ in the open disc $|z|<1$, r is holomorphic in this unit disc. We can even describe the Taylor series of r around $z=0$:

The function $\pi\cot\pi z$ has the well-known Laurent-expansion

$$\pi\cot\pi z=\frac{1}{z}-\sum_{n=1}^{\infty}d_n' z^{2n-1},$$

where

$$d_n'=\frac{(2\pi)^{2n}}{(2n)!}B_{2n}=2\sum_{k=1}^{\infty}\frac{1}{k^{2n}},$$

B_j denoting the j-th Bernoulli number (cf. [2]). Especially one has

$$2 < d'_n < 4 \tag{1.9}$$

for all $n \in \mathbb{N}$. This implies

$$(z+1)\pi \cot \pi z = 1 + \frac{1}{z} - \sum_{n=1}^{\infty} d'_n z^{2n-1} - \sum_{n=1}^{\infty} d'_n z^{2n},$$

hence, for $|z| < 1$,

$$r(z) = \sum_{n=1}^{\infty} d_n z^n, \tag{1.10}$$

where

$d_n = d'_{[(n+1)/2]}$, $[x]$ denoting the integer part of x. Note that consequently (1.9) implies

$$|r(z)| \leqq 4|z|/(1-|z|), \tag{1.11}$$

if $|z| < 1$.

Next, for any real ε with $0 \leqq \varepsilon < 1/6$, define the holomorphic function q_ε on $|z| < 1$ by

$$q_\varepsilon(z) = \cosh z + \frac{\varepsilon}{2} r(-\varepsilon e^{-z}). \tag{1.12}$$

We shall sometimes also write $q(\varepsilon, z)$ instead of $q_\varepsilon(z)$.

Lemma 1.2. *Let* $0 \leqq \varepsilon < 1/6$.

(i) *The restriction of* q_ε *to the real numbers* $\mathbb{R}$ *is a real valued function.*

(ii) *There exists exactly one critical point* $\sigma = \sigma(\varepsilon)$ *of* q_ε *in* $\{|z| < 10\varepsilon^2\}$. *If* ε *is sufficiently small, then* $\sigma(\varepsilon)$ *is real, and the function* $\varepsilon \to \sigma(\varepsilon)$ *is smooth.*

Proof. (i) is clear by the definition of r. In order to prove (ii), we consider the derivative

$$q'_\varepsilon(z) = \sinh z + \frac{\varepsilon^2}{2} e^{-z} r'(-\varepsilon e^{-z}).$$

Now, by (1.10),

$$|r'(z)| \leqq \sum_{n=1}^{\infty} n d_n |z|^{n-1} \leqq 4 \sum_{n=1}^{\infty} n|z|^{n-1}$$
$$= \frac{4}{(1-|z|)^2}.$$

This implies

$$|q'_\varepsilon(z) - \sinh z| \leqq 2\varepsilon^2 e^{|z|}/(1-\varepsilon e^{|z|})^2 \leqq 6\varepsilon^2,$$

if $|z| = 10\varepsilon^2$. On the other hand, one easily shows that $|\sinh z| \geqq 8\varepsilon^2$ for $|z| = 10\varepsilon^2$. Therefore

$$|q'_\varepsilon(z) - \sinh z| < |\sinh z| + |q'_\varepsilon(z)|$$

for $|z|=10\varepsilon^2$, and so Rouché's theorem implies that q_ε' has, similarly as sinh, exactly one zero in $|z|<10\varepsilon^2$.

The last statement of (ii) follows easily from the implicit function theorem, applied to the real function $(\varepsilon, t)\to q_\varepsilon'(t)$ near $(\varepsilon, t)=(0,0)$. □

Let finally I_0 denote the modified Bessel function

$$I_0(x)=\frac{1}{\pi}\int_0^\pi e^{-x\cos\varphi}d\varphi \quad \text{(cf. [4])}.$$

Theorem 1.3. (i) *If* $\delta=\sqrt{|z|^2/(\pi|u|)}$ *tends to zero, and if* $|u|\,|z|^2\to+\infty$, *then* $p(z,u)$ *is asymptotically given by*

$$p(z,u)\sim\frac{1}{4(2\pi)^{1/2}(\pi|u|\,|z|^2)^{1/4}}e^{-\frac{|z|^2}{2}-\frac{\pi}{2}|u|+\varrho(\delta)\sqrt{\pi|u|\,|z|^2}}$$

as $|z|^2+|u|\to\infty$, *where (with* σ *as in Lemma 1.2)*

$$\varrho(\delta)=q(\delta,\sigma(\delta)).$$

Moreover, $\varrho(\delta)=1+O(\delta^2)$.

(ii) *If* $|z|^2/|u|$ *tends to zero, and if* $|u|\,|z|^2\leqq C$ *for some positive constant C, then for* $|z|^2+|u|\to\infty$

$$p(z,u)\sim\tfrac{1}{4}I_0(\sqrt{\pi|u|\,|z|^2})e^{-\frac{|z|^2}{2}-\frac{\pi}{2}|u|}.$$

Remark. There is an explicit formula for p if $z=0$, namely

$$p(0,u)=\tfrac{1}{16}[\cosh\tfrac{\pi}{2}u]^{-2}$$

(cf. [6], Theorem 2 (2°), where a factor 2 had been omitted). The proof of this theorem is based on

Lemma 1.4. *If* $0\leqq R<u$, *then*

$$h(R,u)=\pi e^{-R-\pi u}\int_{-\pi}^{\pi}e^{2\sqrt{\pi R u}\cos t+Rr(-\delta e^{it})}\varphi_\delta(t)dt+g(R,u),$$

where δ *is given by* $\delta=\sqrt{R/(\pi u)}$, *and*

$$\varphi_\delta(t)=\frac{\pi\delta e^{it}}{\sin(\pi\delta e^{it})}(1-\delta e^{it}),$$

and where g can be estimated by

$$|g(R,u)|\leqq\frac{40\pi}{1+\sqrt{R}}e^{-\frac{3\pi}{2}u}.$$

Proof. Let $f(z)=e^{iuz-Rz\coth z}\dfrac{z}{\sinh z}$.

f has exactly one (essential) singularity within the region $0<\operatorname{Im}z<3\pi/2$, namely at $z=\pi i$. Therefore, the theorem of residues easily implies

$$h(R,u)=\int_{\mathbb{R}}f(x)dx=\int_{\mathbb{R}}f\left(x+\frac{3\pi}{2}i\right)dx+2\pi i\operatorname{Res}_{z=\pi i}f.$$

Let $g(R,u)=\int\limits_{\mathbb{R}} f\left(x+\frac{3\pi}{2}i\right)dx$. Since

$$\left|f\left(x+\frac{3\pi}{2}i\right)\right| = e^{-\frac{3\pi}{2}u}\left|\frac{x+3\pi i/2}{\cosh x}e^{iux-R\left(x+\frac{3\pi}{i}i\right)\tanh x}\right|$$
$$\leqq e^{-\frac{3\pi}{2}u}\frac{|x|+3\pi/2}{\cosh x}e^{-Rx\tanh x},$$

we have for $R\leqq 1$

$$|g(R,u)|\leqq e^{-\frac{3\pi}{2}u}\int\limits_{\mathbb{R}}\frac{|x|+3\pi/2}{\cosh x}dx\leqq 4\pi e^{-\frac{3\pi}{2}u}.$$

And, if $R>1$, we get

$$|g(R,u)|\leqq 4\pi e^{-\frac{3\pi}{2}u}\int\limits_{\mathbb{R}} e^{-Rx\tanh x}dx$$
$$\leqq 4\pi e^{-\frac{3\pi}{2}u}\left[\int\limits_0^1 e^{-Rx^2/3}dx+\int\limits_1^\infty e^{-Rx/2}dx\right],$$

since for $0<x<1$ one has

$$\frac{\sinh x}{\cosh x}\geqq 2\frac{x}{2e^x}\geqq\frac{x}{3},$$

and for $x>1$ one has $\tanh x>1/2$. This implies

$$|g(R,u)|\leqq 4\pi e^{-\frac{3\pi}{2}u}\left[\sqrt{\frac{3}{2R}\pi}+\frac{2}{3}\right]$$

if $R>1$. Combination of the estimates of $g(R,u)$ for the two cases $R\leqq 1$ and $R\geqq 1$ yields the desired estimate for $g(R,u)$.

Next, in order to calculate the residue of f at $z=\pi i$, we substitute z by $\pi iz+\pi i$ and get

$$\mathrm{Res}_{z=\pi i}f(z)=\pi i\,\mathrm{Res}_{z=0}f(\pi i(z+1))$$
$$=-\pi ie^{-\pi u}\,\mathrm{Res}_{z=0}\left[e^{-\pi uz-R(z+1)\pi\cot(\pi z)}\frac{\pi(z+1)}{\sin(\pi z)}\right].$$

Now, by the definition of the function r, we have

$$-\pi uz-R(z+1)\pi\cot\pi z=-R-\pi uz-\frac{R}{z}+Rr(z).$$

We intend to calculate the residue by integrating along a circle with center $z=0$. The radius δ of this circle has to be less then 1, since $(z+1)\cot(\pi z)$ has a pole at $z=1$, and, because of the estimate (1.11) for $r(z)$, it is clear that in the case $u\gg R$ only the terms $-\pi uz$ and $-\frac{R}{z}$ from above have a strong contribution to this integral. Therefore, we choose δ so, that for $|z|=\delta$ we have $|\pi uz|=|R/z|$. This implies

$$\delta=\sqrt{R/(\pi u)}.$$

Thus, if $R>0$, integration along the circle δe^{it}, $0\leqq t<2\pi$, yields

$$2\pi i \operatorname{Res}_{z=\pi i} f = -\pi i e^{-\pi u} \int_0^{2\pi} \exp\left[-R-2\sqrt{\pi R u}\cos t + Rr(\delta e^{it})\right] \times \left[i\frac{\pi\delta e^{it}}{\sin(\pi\delta e^{it})}(1+\delta e^{it})\right]dt.$$

Substituting t by $t-\pi$ yields the desired formula for $h(R,u)-g(R,u)$.

Finally, the case $R=0$ follows from the case $R>0$ by continuity. □

Proof of Theorem 1.3. ad*(i)*: We may assume $u>0$, and set $R=|z|^2$. We set $\kappa=2\sqrt{\pi R u}$, so that $\kappa\to+\infty$. Let

$$H(R,u)=\int_{-\pi}^{\pi} e^{2\sqrt{\pi R u}\cos t + Rr(-\delta e^{it})}\varphi_\delta(t)dt,$$

so that by Lemma 1.4

$$h(R,u)=\pi e^{-R-\pi u}H(R,u)+g(R,u).$$

For $|z|<1$ and $0\leqq\varepsilon<1/6$ we set

$$\tilde{q}_\varepsilon(z)=\cos z+\frac{\varepsilon}{2}r(-\varepsilon e^{iz}).$$

Then, for $\varepsilon=\delta:=\sqrt{R/(\pi u)}$, we have

$$2\sqrt{\pi R u}\cos t+Rr(-\delta e^{it})=\kappa\tilde{q}_\delta(t),$$

hence

$$H(R,u)=\int_{-\pi}^{\pi} e^{\kappa\tilde{q}_\delta(t)}\varphi_\delta(t)dt.$$

Now, we have

$$\tilde{q}_\varepsilon(z)=q_\varepsilon(-iz),$$

and so Lemma 1.2 implies that $\tilde{q}_\varepsilon$ has a unique critical point in $\{|z|<10\varepsilon^2\}$, namely the point $i\sigma(\varepsilon)$. Moreover, if δ is small enough, the function $z\to e^{\kappa\tilde{q}_\delta(z)}\varphi_\delta(z)$ is holomorphic in the strip $|\mathrm{Im}\, z|<20\delta^2$. Therefore, we choose a path $\gamma:[-\pi,\pi]\to\mathbb{C}$ such that $\gamma(-\pi)=-\pi$, $\gamma(\pi)=\pi$, $|\mathrm{Im}\,\gamma(t)|\leqq 19\delta^2$ for all $t\in[-\pi,\pi]$, and

$$\gamma(t)=t+i\sigma(\delta)$$

for $t\in[-\frac{\pi}{2},\frac{\pi}{2}]$, and clearly have

$$H(R,u)=\int_{-\pi}^{\pi} e^{\kappa(\tilde{q}_\delta\circ\gamma)(t)}\varphi_\delta\circ\gamma(t)\gamma'(t)dt.$$

If we set $F_\varrho(t)=\tilde{q}_\delta\circ\gamma(t)-\tilde{q}_\delta\circ\gamma(0)$, we obtain

$$\begin{aligned} H(R,u)&=e^{\kappa q_\delta(\sigma(\delta))}\int_{-\pi}^{\pi} e^{\kappa F_\delta(t)}\varphi_\delta\circ\gamma(t)\gamma'(t)dt\\ &=\frac{e^{\kappa\varrho(\delta)}}{\sqrt{\kappa}}\int_{-\pi\sqrt{\kappa}}^{\pi\sqrt{\kappa}} e^{\kappa F_\delta(s/\sqrt{\kappa})}\psi_\delta(s/\sqrt{\kappa})ds,\end{aligned}$$

where
$$\psi_\delta(t)=\varphi_\delta\circ\gamma(t)\gamma'(t).$$

One verifies easily that
$$F_\delta(0)=F'_\delta(0)=0,$$
$$F''_\delta(0)=-\cos(i\sigma(\delta))+O(\delta^2),$$

and
$$\left|F_\delta(t)-F''_\delta(0)\frac{t^2}{2}\right|\leqq Ct^3,$$

where C does not depend on δ.

Together, this implies
$$\lim_{\substack{\kappa\to\infty\\ \delta\to 0}}\kappa F_\delta(s/\sqrt{\kappa})=-s^2/2$$

for every $s\in\mathbb{R}$. Also, since $\gamma'(0)=1$, we have
$$\lim_{\substack{\kappa\to\infty\\ \delta\to 0}}\psi_\delta(s/\sqrt{\kappa})=1.$$

Moreover, it is not difficult to show that, for δ sufficiently small, one has
$$|e^{\kappa F_\delta(s/\sqrt{\kappa})}\psi_\delta(s/\sqrt{\kappa})|\leqq Ce^{-s^2/10}$$

for $|s|\leqq\pi\sqrt{\kappa}$, uniformly in δ and κ. So, the dominated convergence theorem implies
$$\lim_{\substack{\kappa\to\infty\\ \delta\to 0}}\int_{-\pi\sqrt{\kappa}}^{\pi\sqrt{\kappa}} e^{\kappa F_\delta(s/\sqrt{\kappa})}\psi_\delta(s/\sqrt{\kappa})ds=\int_{-\infty}^{\infty} e^{-s^2/2}ds=\sqrt{2\pi},$$

hence
$$H(R,u)\sim(2\pi)^{1/2}\frac{e^{\kappa\varrho(\delta)}}{\sqrt{\kappa}}$$

as $\kappa\to\infty$ and $\delta\to 0$. This implies
$$h(R,u)\sim\frac{\pi(2\pi)^{1/2}}{2^{1/2}(\pi Ru)^{1/4}}e^{-R-\pi u+2\sqrt{\pi Ru}\varrho(\delta)}+g(R,u)$$

as $\kappa\to\infty$ and $\delta\to 0$.

Moreover, (1.12) and Lemma 1.2 imply
$$\varrho(\delta)=q(\delta,\sigma(\delta))=1+O(\delta^2).$$

This implies
$$-R-\pi u+2\sqrt{\pi Ru}\,\varrho(\delta)=-\pi[1+\delta^2-2\delta\varrho(\delta)]u=-\pi(1+O(\delta))u.$$

Since $u\to\infty$ and $\frac{R}{u}\to 0$, this and the estimate of $g(R,u)$ in Lemma 1.4 show that $g(R,u)$ is negligible for the asymptotics of $h(R,u)$, and by (1.3), we finally get
$$p(z,u)\sim\frac{\pi(2\pi)^{1/2}}{2(2\pi)^2(\pi Ru)^{1/4}}e^{-\frac{R}{2}-\frac{\pi}{2}u+\sqrt{\pi Ru}\varrho(\delta)},$$

if $u>0$ and $R=|z|^2$.

ad *(ii)*: We adopt the notation of the proof of (i), and write

$$H(R,u)=\int_{-\pi}^{\pi} e^{2\sqrt{\pi Ru}\cos t}\psi(t)dt,$$

where

$$\psi(t)=e^{Rr(-\delta e^{it})}\varphi_\delta(t).$$

For $\delta<1/2$, estimate (1.11) implies

$$|Rr(-\delta e^{it})|\leqq 8R\delta=\frac{8}{\sqrt{\pi}}\left(\frac{R^3}{u}\right)^{1/2}\leqq\frac{8C^{3/2}}{\sqrt{\pi}}\frac{1}{u^2},$$

since $Ru\leqq C$. And, if $R/u\to 0$ and $R+u\to\infty$, then $u\to\infty$, and so $Rr(-\delta e^{it})$ tends to zero uniformly. Moreover, since $\delta\to 0$, φ_δ converges uniformly to the constant function 1 on $[-\pi,\pi]$. Consequently, $\psi\to 1$ uniformly on $[-\pi,\pi]$ as $R+u\to\infty$ and $R/u\to 0$.

But, since $p(z,u)$ and consequently $h(R,u)$ is positive, we have

$$h(R,u)=\pi e^{-R-\pi u}\int_{-\pi}^{\pi} e^{2\sqrt{\pi Ru}\cos t}\operatorname{Re}\psi(t)dt+\operatorname{Re}g(R,u),$$

where also $\operatorname{Re}\psi\to 1$ uniformly on $[-\pi,\pi]$.

But, since $e^{2\sqrt{\pi Ru}\cos t}$ is positive, this clearly implies

$$h(R,u)\sim\pi e^{-R-\pi u}\int_{-\pi}^{\pi} e^{2\sqrt{\pi Ru}\cos t}dt+\operatorname{Re}g(R,u)$$

as $R+u\to\infty$, $R/u\to 0$, and $Ru\leqq C$.

Moreover

$$\int_{-\pi}^{\pi} e^{2\sqrt{\pi Ru}\cos t}dt>\int_{-\pi/2}^{\pi/2} e^0 dt=\pi,$$

and so the estimate of $g(R,u)$ in Lemma 1.4 shows that the term $\operatorname{Re}g(R,u)$ is negligible for the asymptotic behaviour of $h(R,u)$. Thus in combination with (1.3), we get

$$p(z,u)\sim\frac{1}{8\pi}e^{-\frac{R}{2}-\frac{\pi}{2}u}\int_{-\pi}^{\pi} e^{\sqrt{\pi Ru}\cos t}dt.$$

Since

$$\int_{-\pi}^{0} e^{x\cos t}dt=\int_{0}^{\pi} e^{x\cos t}dt,$$

hence

$$\int_{-\pi}^{\pi} e^{x\cos t}dt=2\pi I_0(-x)=2\pi I_0(x),$$

we obtain

$$p(z,u)\sim\tfrac{1}{4}I_0(\sqrt{\pi Ru})e^{-\frac{R}{2}-\frac{\pi}{2}u},\quad \text{q.e.d.}\ \square$$

Remark 1.5. The proof of Theorem 1.3 (i) shows that the exponent in the asymptotic formula for $p(z,u)$ is just the value of the phase $\zeta \to iu\zeta - R\eta \coth\zeta$ of the integral (1.2) at the (unique) critical value $\zeta_0 \in \mathbb{C}$ of this phase. The same is true for the exponent for the asymptotics of $p(z,u)$ in Theorem 1.1 (compare the proof of Theorem 2 (3°) in [6]). Therefore we have

$$-\frac{|z|^2}{2} - \tfrac{\pi}{2}|u| + \varrho(\delta)\sqrt{\pi|u|\,|z|^2} = -\gamma^2\left(\frac{|u|}{|z|^2}\right)\frac{|z|^2}{2}. \tag{1.13}$$

Moreover, a careful analysis of the function $\phi(\tau)$ near $\tau = \pi$ and of the function $\tau(\omega)$ for $\omega \to \pm\infty$ shows that, with $\omega = |u|/|z|^2$,

$$\frac{\phi(\tau(\omega))}{|z|} \sim \frac{1}{4(2\pi)^{1/2}(\pi|u|\,|z|^2)^{1/4}}$$

as

$$\delta = \sqrt{|z|^2/(\pi|u|)} \to 0.$$

Thus, in the case of Theorem 1.3 (i) we have the same asymptotics

$$p(z,u) \sim \frac{\phi(\omega)}{|z|} e^{-\gamma^2\left(\frac{|u|}{|z|^2}\right)\frac{|z|^2}{2}} \tag{1.13}$$

as in the case of Theorem 1.1.

However, the right-hand side of (1.13) does no longer describe the asymptotics of p in the case of Theorem 1.3 (ii), and moreover the formula of Theorem 1.3 (i) is more informative than (1.13).

2. Asymptotic Estimates for the Fundamental Solution of $-\Delta_K + 2$

For $(z,u) \in H_1 \setminus \{0\}$ let

$$K(z,u) := \tfrac{1}{2}\int_0^\infty p_s(z,u)e^{-s}ds. \tag{2.1}$$

Since p is a Schwartz-class function (see e.g. [5]), it is clear that this integral converges for every $(z,u) \neq 0$. In fact, the same observations in combination with the homogeneity of p_s as expressed by (1.3) even imply that K vanishes at infinity. Obviously K is positive and integrable, since $\int_{H_1} p_s(z,u)dzdu = 1$. K is a fundamental solution for $-\Delta_K + 2$. This is indicated by the following formal calculation, which can easily be made precise in the sense of distributions:

$$\Delta_K K = \frac{1}{2}\int_0^\infty (\Delta_K p_s)e^{-s}ds = \int_0^\infty \frac{\partial p_s}{\partial s} e^{-s}ds$$

$$= p_s e^{-s}\Big|_0^\infty + \int_0^\infty p_s e^{-s}ds;$$

but $p_0 = \delta_0$ is the Dirac distribution at the origin, and so

$$\Delta_K K = -\delta_0 + 2K,$$

hence

$$(-\Delta_K+2)K=\delta_0. \tag{2.2}$$

Of course, all this is well-known.

Moreover, by a result of Hervé ([9], p. 142), K is in fact the unique positive fundamental solution of $-\Delta_K+2$ which vanishes at infinity.

We are now going to describe the asymptotic behaviour of $K(z,u)$, and again we start with the case where $\omega=|u|/|z|^2$ stays bounded. We adopt the notation of the preceding paragraph.

Theorem 2.1. *If the ratio $\omega=|u|/|z|^2$ tends towards a limit $\omega_\infty<\infty$, then $K(z,u)$ is asymptotically given by*

$$K(z,u)\sim\Psi(\omega_\infty)\frac{1}{|z|^2}e^{-\sqrt{2}\gamma\left(\frac{|u|}{|z|^2}\right)|z|}$$

as $|z|^2+|u|\to\infty$, where

$$\Psi(\omega)=\frac{1}{8\pi}\left[\frac{\sin^3\tau(\omega)}{\sin\tau(\omega)-\tau(\omega)\cos\tau(\omega)}\right]^{1/2}.$$

Proof. Assume that $\lim|u|/|z|^2=\omega_\infty<\infty$ as $|z|^2+|u|\to\infty$. Then, by Theorem 1.1, for every $\varepsilon>0$, there exist $N(\varepsilon)>0$ and $\delta(\varepsilon)>0$, such that

$$\left|1-\frac{p(z,u)}{\tilde{p}(z,u)}\right|<\varepsilon, \tag{2.3}$$

if $|z|^2+|u|>N(\varepsilon)$ and $\left|\frac{u}{|z|^2}-\omega_\infty\right|<\delta(\varepsilon)$; here $\tilde{p}(z,u)$ denotes the function

$$\tilde{p}(z,u)=\Phi(\omega_\infty)\frac{1}{|z|}e^{-\gamma^2\left(\frac{|u|}{|z|^2}\right)\frac{|z|^2}{2}}. \tag{2.4}$$

Now, fix $\varepsilon>0$, and assume that

$$|z|^{2/3}>N(\varepsilon)+1, \qquad \left|\frac{|u|}{|z|^2}-\omega_\infty\right|<\delta(\varepsilon).$$

Then (2.3) implies, since p and $\tilde{p}$ are positive,

$$(1-\varepsilon)\tilde{p}\left(\frac{z}{\sqrt{s}},\frac{u}{s}\right)\leqq p\left(\frac{z}{\sqrt{s}},\frac{u}{s}\right)\leqq(1+\varepsilon)\tilde{p}\left(\frac{z}{\sqrt{s}},\frac{u}{s}\right)$$

if $s<|z|^{4/3}$, for in this case

$$\left|\frac{z}{\sqrt{s}}\right|^2+\left|\frac{u}{s}\right|\geqq\frac{|z|^2}{s}=\frac{|z|^{4/3}}{s}\cdot|z|^{2/3}>|z|^{2/3}>N(\varepsilon);$$

and of course $\left|\frac{|u/s|}{|z/\sqrt{s}|^2}-\omega_\infty\right|=\left|\frac{|u|}{|z|^2}-\omega_\infty\right|<\delta(\varepsilon)$.

Set

$$K_I(z,u)=\tfrac{1}{2}\int_0^{|z|^{4/3}} p_s(z,u)e^{-s}ds,$$

$$K_{II}(z,u)=\tfrac{1}{2}\int_{|z|^{4/3}}^{\infty} p_s(z,u)e^{-s}ds,$$

and form $\tilde{K}_I$ and $\tilde{K}_{II}$ analogously by replacing p by $\tilde{p}$. Then clearly

$$(1-\varepsilon)\tilde{K}_I(z,u)\leqq K_I(z,u)\leqq(1+\varepsilon)\tilde{K}_I(z,u). \tag{2.5}$$

Moreover, if we set $C_1=\sup\{|p(z,u)|:(z,u)\in H_1\}$, then

$$\begin{aligned}2K_{II}(z,u)&=\int_{|z|^{4/3}}^{\infty} p\left(\frac{z}{\sqrt{s}},\frac{u}{s}\right)\frac{e^{-s}}{s^2}ds\\&\leqq C_1\int_{|z|^{4/3}}^{\infty} e^{-s}ds\\&=C_1e^{-|z|^{4/3}}.\end{aligned}$$

Similarly,

$$\begin{aligned}2\tilde{K}_{II}(z,u)&=\frac{\Phi(\omega_\infty)}{|z|}\int_{|z|^{4/3}}^{\infty} s^{-3/2}e^{-\gamma^2\left(\frac{|u|}{|z|^2}\right)\frac{|z|^2}{2s}-s}ds\\&\leqq\frac{\Phi(\omega_\infty)}{|z|}\int_{|z|^{4/3}}^{\infty} e^{-s}ds\\&\leqq\Phi(\omega_\infty)e^{-|z|^{4/3}}.\end{aligned}$$

Thus, with $2C_2=\max(C_1,\Phi(\omega_\infty))$, we have

$$K_{II}(z,u)\leqq C_2e^{-|z|^{4/3}},\qquad \tilde{K}_{II}(z,u)\leqq C_2e^{-|z|^{4/3}}. \tag{2.6}$$

Next, we can calculate explicitly $\tilde{K}=\tilde{K}_I+\tilde{K}_{II}$: From ([4], p. 82 (23) and p. 10 (42)), it follows that

$$\int_0^\infty s^{-3/2}e^{-\frac{a}{s}-s}ds=\sqrt{\frac{\pi}{a}}e^{-2\sqrt{a}},$$

if $a>0$. Now

$$\tilde{K}(z,u)=\frac{\Phi(\omega_\infty)}{2|z|}\int_0^\infty s^{-3/2}e^{-\gamma^2\left(\frac{|u|}{|z|^2}\right)\frac{|z|^2}{2s}-s}ds,$$

hence

$$\tilde{K}(z,u)=\Psi(\omega_\infty)\frac{1}{|z|^2}\frac{\gamma(\omega_\infty)}{\gamma\left(\frac{|u|}{|z|^2}\right)}e^{-\sqrt{2}\gamma\left(\frac{|u|}{|z|^2}\right)|z|}. \tag{2.7}$$

The estimates in (2.6) and (2.7) show, that there exists an $N'(\varepsilon)>N(\varepsilon)+1$, such that

$$K_{II}(z,u)/\tilde{K}(z,u)<\varepsilon,\qquad \tilde{K}_{II}(z,u)/\tilde{K}(z,u)<\varepsilon,$$

if $|z|^{2/3} > N'(\varepsilon)$. But this, together with (2.5), easily implies

$$(1-2\varepsilon)\tilde{K}(z,u) \leqq K(z,u) \leqq (1+2\varepsilon)\tilde{K}(z,u),$$

if $|z|^{2/3} > N'(\varepsilon)$ and $\left|\frac{|u|}{|z|^2} - \omega_\infty\right| < \delta(\varepsilon)$.

This clearly implies the theorem. $\square$

Finally, we deal with the case $|z|^2/|u| \to 0$:

Theorem 2.2. (i) *If* $\delta = \sqrt{|z|^2/(\pi|u|)}$ *tends to zero, and if* $|z| \to +\infty$, *then* $K(z,u)$ *is asymptotically given by*

$$K(z,u) \sim \tfrac{1}{8} \frac{e^{-[2(\pi|u| + |z|^2 - 2\varrho(\delta)\sqrt{\pi|u|\,|z|^2})]^{1/2}}}{(\pi|u|)^{3/4}|z|^{1/2}}$$

as $|z|^2 + |u| \to +\infty$, *where* ϱ *is defined as in Theorem 1.3.*

(ii) *If* $\delta = \sqrt{|z|^2/(\pi|u|)}$ *tends to zero, and if there exists a constant* $C > 0$, *such that* $|z| \leqq C$, *then*

$$K(z,u) \sim \frac{2^{3/4}\sqrt{\pi}}{8} I_0(\sqrt{2}|z|) \frac{e^{-\sqrt{2\pi|u|}}}{(\pi|u|)^{3/4}}$$

as $|z|^2 + |u| \to \infty$.

Proof. We may again assume that $u > 0$.

We set

$$F(R,u) = \int_0^\infty \frac{1}{s^2}(h-g)\left(\frac{R}{s}, \frac{u}{s}\right) e^{-s} ds,$$

$$G(R,u) = \int_0^\infty \frac{1}{s^2} g\left(\frac{R}{s}, \frac{u}{s}\right) e^{-s} ds,$$

where h and g are defined as in Lemma 1.4. Then (1.2) implies with $R = |z|^2$, that

$$K(z,u) = \frac{1}{4(2\pi)^2}\left[F\left(\frac{R}{2}, \frac{u}{2}\right) + G\left(\frac{R}{2}, \frac{u}{2}\right)\right]. \tag{2.8}$$

We first estimate $G(R,u)$:

Lemma 1.4 implies

$$|G(R,u)| \leqq 40\pi \int_0^\infty \frac{1}{1 + \left(\frac{R}{s}\right)^{1/2}} \frac{1}{s^2} e^{-\frac{3\pi}{2}\frac{u}{s} - s} ds$$

$$\leqq 40\pi \int_0^\infty \frac{1}{s^2} e^{-\frac{3\pi}{2}\frac{u}{s} - s} ds.$$

Moreover, by ([4], p. 82 (23)),

$$\int_0^\infty \frac{1}{s^2} e^{-\frac{a}{s} - s} ds = \frac{2}{\sqrt{a}} K_1(2\sqrt{a}), \quad \text{if} \quad \operatorname{Re} a > 0,$$

where K_1 denotes the modified Hankel function of order 1, and ([4], p. 86 (7)) implies

$$K_1(z)=\sqrt{\frac{\pi}{2z}}\,e^{-z}[1+O(|z|^{-1})]\,,\quad \text{if} \quad \operatorname{Re} z>0\,,$$

which together yields

$$\int_0^\infty \frac{1}{s^2}\,e^{-\frac{a}{s}-s}ds=\frac{\pi^{1/2}}{a^{3/4}}\,e^{-2\sqrt{a}}[1+O(a^{-1/2})]\,, \tag{2.9}$$

if $\operatorname{Re} a>0$.

Thus there exists $C>0$ such that for $|u|>C$

$$|G(R,u)|\leqq 80\pi^{3/2}\frac{e^{-\sqrt{6\pi u}}}{|u|^{3/4}}\,. \tag{2.10}$$

Next, we are going to determine the asymptotics for $F(R,u)$: By Fubini's theorem, we have

$$F(R,u)=\pi\int_{-\pi}^{\pi}\int_0^\infty \frac{1}{s^2}\,e^{-\frac{a(t)}{s}-s}ds\varphi_\delta(t)dt$$

where

$$a(t)=\pi u+R-2\sqrt{\pi R u}\cos t-Rr(-\delta e^{it})\,. \tag{2.11}$$

Since we may assume $u\gg R$, we have $\operatorname{Re} a(t)>0$ for every $t\in[-\pi,\pi]$, and thus (2.9) implies

$$F(R,u)=\pi^{3/2}\int_{-\pi}^{\pi}\frac{e^{-2\sqrt{a(t)}}}{[a(t)]^{3/4}}[1+\xi(t)]\varphi_\delta(t)dt\,, \tag{2.12}$$

where ξ is smooth, depends also on u and R, but where the supremum norm $\|\xi\|_\infty$ of ξ on $[-\pi,\pi]$ satisfies

$$\|\xi\|_\infty=O(|u|^{-1/2})\,. \tag{2.12}$$

For the sequel, it is also important to note, that for the asymptotics considered in Theorem 2.2 we have, uniformly on $[-\pi,\pi]$,

$$a(t)\sim\pi u\,. \tag{2.13}$$

In order to prove (i), let us now assume $R\to+\infty$: The complex critical points of the phase $-2\sqrt{a(z)}$ are given by $a'(z)=0$, or, equivalently, by $\tilde{q}_\delta'(z)=0$, where $\tilde{q}_\delta$ is defined as in the proof of Theorem 1.3 (i). Thus, again the only critical point of this phase in $\{|z|<10\delta^2\}$ is the point $i\sigma(\delta)$. So, introducing a path γ as in the proof of Theorem 1.3 (i) and arguing like there, we obtain

$$F(R,u)=\pi^{3/2}\frac{e^{-2\sqrt{a(i\sigma(\delta))}}}{(\pi u)^{3/4}}\int_{-\pi}^{\pi}e^{\sqrt{R}b(t)}\psi(t)\left[\frac{\pi u}{a(t)}\right]^{3/4}dt\,,$$

where

$$b(t)=[2\sqrt{a(i\sigma(\delta))}-2\sqrt{a\circ\gamma(t)}]/\sqrt{R}\,,$$

$$\psi(t)=[1+\xi\circ\gamma(t)]\varphi_\delta\circ\gamma(t)\gamma'(t)$$

depend also on R and u.

Now, clearly $b(0)=b'(0)=0$, and

$$\begin{aligned} b''(0) &= -\frac{a''(i\sigma(\delta))}{[a(i\sigma(\delta))]^{1/2}\sqrt{R}} \\ &= \frac{-2\cos(i\sigma)+\delta^2 e^{-\sigma}r'(-\delta e^{-\sigma})-\delta^3 e^{-2\sigma}r''(-\delta e^{-\sigma})}{[1+\delta^2-2\delta\cosh\sigma-\delta^2 r(-\delta e^{-\sigma})]^{1/2}}, \end{aligned}$$

where we set $\sigma=\sigma(\delta)$. Since $|\sigma|\leqq 10\delta^2$, this implies, for $\delta\to 0$

$$b''(0)\to -2. \tag{2.14}$$

Similarly, but with a bit more technical effort, one can show that b'' is uniformly bounded on $[-\pi,\pi]$, independently of u and R, if δ is sufficiently small and u sufficiently large. In fact, for $1/\delta$ and u sufficiently large, one even has

$$|b'''(t)|\leqq 4|\sin t|, \quad t\in[-\tfrac{\pi}{2},\tfrac{\pi}{2}].$$

So we can apply similar arguments as in the proof of Theorem 1.3 to show that, if also $R\to\infty$,

$$R^{1/4}\int_{-\pi}^{\pi} e^{\sqrt{R}b(t)}\psi(t)\left[\frac{\pi u}{a(t)}\right]^{3/4}dt=\int_{-\pi R^{1/4}}^{\pi R^{1/4}} e^{\sqrt{R}b(s/R^{1/4})}\psi(s/R^{1/4})\left[\frac{\pi u}{a(s/R^{1/4})}\right]^{3/4}ds$$

tends towards $\int_{-\infty}^{\infty}\exp(-s^2)ds=\sqrt{\pi}$.

Here, one should also note that, as $u\to+\infty$, $1+\xi(t)$ tends towards 1 uniformly on $[-\pi,\pi]$ because of (2.12), and that $\pi u/a(t)$ tends towards 1 uniformly on $[-\pi,\pi]$ because of (2.13). Thus we have shown that

$$F(R,u)\sim\pi^2 e^{-2\sqrt{a(i\sigma(\delta))}}/[(\pi u)^{3/4}R^{1/4}] \tag{2.15}$$

as $u\to+\infty$, $\delta\to 0$ and $R\to+\infty$.

But, since

$$a(i\sigma(\delta))=\pi u+R-\varrho(\delta)2\sqrt{\pi R u},$$

this and (2.10) show that $G(R,u)$ is negligible for the asymptotics of $K(R,u)$, and so (2.8) and (2.15) imply (i).

It remains to prove (ii). So, assume that $|z|$ is bounded by some constant $C>0$. This case is similar to case (ii) of Theorem 1.3, and therefore we abbreviate the argument. Consider formula (2.12). Since

$$|\sqrt{a(t)}-(\pi u+R-2\sqrt{\pi R u}\cos t)^{1/2}|=O\left(\frac{R|r(-\delta e^{it})|}{\sqrt{\pi u}}\right)=O\left(\frac{R^{3/2}}{\sqrt{u}}\right)=O\left(\frac{1}{\sqrt{u}}\right)$$

because of (1.11), and since $\xi(t)\to 0$ and $\varphi_\delta(t)\to 1$ uniformly on $[-\pi,\pi]$ as $u\to+\infty$ and $\delta\to 0$, we easily see that

$$F(R,u)\sim\pi^{3/2}\int_{-\pi}^{\pi}\frac{e^{-2[\pi u+R-2\sqrt{\pi R u}\cos t]^{1/2}}}{(\pi u)^{3/4}}dt,$$

where once again we have also made use of (2.13) and the positivity of K. Moreover,

$$[\pi u+R-2\sqrt{\pi R u}\cos t]^{1/2}=(\pi u)^{1/2}[1+\delta^2-2\delta\cos t]^{1/2}$$
$$=(\pi u)^{1/2}[1-\delta\cos t+O(\delta^2)],$$

and since

$$(\pi u)^{1/2}\cdot\delta^2=\frac{R}{\sqrt{\pi u}}\leqq\frac{C^2}{\sqrt{\pi u}}\to 0,$$

the term $O(\delta^2)$ is negligible.

Thus we have

$$F(R,u)\sim\frac{\pi^{3/2}}{(\pi u)^{3/4}}e^{-2\sqrt{\pi u}}\int_{-\pi}^{\pi}e^{2\sqrt{R}\cos t}dt,$$

hence

$$F(R,u)\sim\frac{2\pi^{5/2}}{(\pi u)^{3/4}}I_0(2\sqrt{R})e^{-2\sqrt{\pi u}}.$$

Again, this shows that $G(R,u)$ may be neglected, and together with (2.8) this implies (ii). □

Remark 2.3. Formula (1.8) shows that the function $-a(t)$ is just the phase function $\zeta\to iu\zeta-R\zeta\cot\zeta$ of (1.2) composed with the function $t\to\pi i(1+\delta e^{i(t-\pi)})$. Therefore, by the proof of Theorem 2.2 (i), it is clear that $-a(i\sigma(\delta))$ is nothing but the value of this phase function at it's critical point, that is $-\gamma^2\left(\frac{|u|}{|z|^2}\right)\frac{|z|^2}{2}$ (compare Remark 1.5). Therefore we have

$$-[2(\pi|u|+|z|^2-2\varrho(\delta)\sqrt{\pi|u|\,|z|^2})]^{1/2}=-\sqrt{2}\gamma\left(\frac{|u|}{|z|^2}\right)|z|. \tag{2.16}$$

Moreover, similarly as in Remark 1.5, one can show that, with $\omega=|u|/|z|^2$,

$$\frac{\psi(\omega)}{|z|^2}\sim\frac{1}{8(\pi|u|)^{3/4}|z|^{1/2}}$$

as $\delta\to 0$, and thus we also have

$$K(z,u)\sim\frac{\psi(\omega)}{|z|^2}e^{-\sqrt{2}\gamma\left(\frac{|u|}{|z|^2}\right)|z|} \tag{2.16}$$

in the case of Theorem 2.2 (i). But again the analog is no longer true in case of Theorem 2.2 (ii).

3. The Martin Boundary

We are now going to determine the Martin compactification of H_1. As references to the notions and results from potential theory which we shall use, we recommend the classical book of Helms [8], where the theory is developped for the Laplacian on $\mathbb{R}^n$ in such a way that it could easily be extended to our situation, and Brelot's book [1]. The latter book presents a more abstract potential theory which is already sufficiently general to cover our situation.

By definition, the *Martin compactification* $M(H_1)$ of H_1 corresponding to the operator $-\Delta_K+2$ is the unique (up to homeomorphisms) compactification of H_1, such that each of the functions

$$F^y(x):=\frac{K(y^{-1}x)}{K(x)}, \quad y\in H_1,$$

which may be considered as a continuous function from H_1 to $[0,\infty]$, can be extended continuously to $M(H_1)$, and such that the set of all those extended functions separates the points of $\Delta:=M(H_1)\backslash H_1$. The set Δ is called the *Martin boundary* of H_1.

Note, that if the extension of F^y to $M(H_1)$ is again denoted by F^y, then the function $F(x,y):=F^y(x)$ is continuous with values in $]0,\infty[$ on $\Delta\times H_1$; moreover, for any fixed $x\in\Delta$, the function $F_x(y):=F(x,y)$ is $(-\Delta_K+2)$-harmonic and positive.

Next, let Δ_1 denote the set of all $x\in\Delta$ such that the function F_x lies on an extremal ray of the cone Γ of all positive $(-\Delta_K+2)$-harmonic functions on H_1. Δ_1 is called the *minimal Martin boundary*. Then the representation theorem of Martin/Choquet ([1] Theorem XIV, 4) states that every $(-\Delta_K+2)$-harmonic function $h\geqq 0$ admits a representation

$$h(y)=\int_{\Delta} F_x(y)d\mu(x), \tag{3.1}$$

where μ is a positive measure on Δ with $\mu(\Delta\backslash\Delta_1)=0$, and μ is uniquely determined by the properties. Note that (3.1) implies that every extremal ray of Γ is of the form $\mathbb{R}^+F_x$ for some $x\in\Delta_1$.

Now we are going to describe a compactification M of H_1 which will turn out to be the Martin compactification.

As a set, let M be the disjoint union of H_1, the complex plane $\mathbb{C}$ and the unit circle S^1 in $\mathbb{C}$, that is $M=H_1\dot\cup\mathbb{C}\dot\cup S^1$. The topology on M is defined as follows:

The neighborhoods of a point in H_1 are just the neighborhoods of the Euclidean topology on H_1.

If ζ_0 is a point in $\mathbb{C}$, then a basis for the system of neighborhoods of ζ_0 is given by the sets

$$U_\varepsilon(\zeta_0)=\left\{(z,u)\in H_1:|z-\zeta_0|<\varepsilon,|u|>\frac{1}{\varepsilon}\right\}\cup\{\zeta\in\mathbb{C}:|\zeta-\zeta_0|<\varepsilon\}\ , \quad \varepsilon>0.$$

If $e^{i\varphi_0}$ is a point in S^1, then a basis of neighborhoods of $e^{i\varphi_0}$ is given by the sets

$$\begin{aligned}V_\varepsilon(e^{i\varphi_0})=&\left\{(re^{i\varphi},u)\in H_1:r>\frac{1}{\varepsilon},\left|e^{i\left[\tau\left(\frac{u}{r^2}\right)+\varphi\right]}-e^{i\varphi_0}\right|<\varepsilon\right\}\\ &\cup\left\{\varrho e^{i\psi}\in\mathbb{C}:\varrho>\frac{1}{\varepsilon},|e^{i\psi}+e^{i\varphi_0}|<\varepsilon\right\}\\ &\cup\{e^{i\varphi}\in S^1:|e^{i\varphi}-e^{i\varphi_0}|<\varepsilon\}\ , \quad \varepsilon>0.\end{aligned}$$

Here τ denotes the diffeomorphism $\tau:\mathbb{R}\to]-\pi,\pi[$ from Sect. 1. Note that

$$\lim_{\kappa\to\pi}\tau^{-1}(\kappa)=+\infty, \qquad \lim_{\kappa\to-\pi}\tau^{-1}(\kappa)=-\infty.$$

It is easy to see that these properties of τ imply that M thus becomes a compact Hausdorff space, and that M is a compactification of H_1. The following lemma shows that M is even much nicer and more symmetric than it might appear from the very definition:

Lemma 3.1. *Let $\tilde{M}'$ denote the "solid torus" $\tilde{M}'=D\times S^1$, where $D=\{z\in\mathbb{C}:|z|\leqq 1\}$ denotes the closed unit disk in $\mathbb{C}$. Introduce the structure of a fibre bundle $\pi:S^1\times S^1\to S^1$ on the boundary $S^1\times S^1$ of $\tilde{M}'$ whose base projection is given by $\pi(e^{i\varphi},e^{i\psi})=e^{i(\varphi-\psi)}$ (note that in fact this defines a principal fibre bundle over the multiplicative group $S^1\subset\mathbb{C}^*$).*

Define a closed equivalence relation R on $\tilde{M}'$ whose classes are the fibres

$$\pi^{-1}(e^{i\varphi})=\{(e^{i(\varphi+t)},e^{it}):t\in\mathbb{R}\}$$

and the one-point sets $\{v\}$, $v\in D^0\times S^1$. Then M is homeomorphic to the quotient space $M'=\tilde{M}'/R$.

Proof. Since $\tilde{M}'$ is compact and R is closed, $M'=\tilde{M}'/R$ is compact too.

Define a mapping $\tilde{\phi}:M\to\tilde{M}'$ by

$$\tilde{\phi}(z,u)=\left(\frac{z}{1+|z|},e^{-i\tau(u/(1+|z|^2))}\right),\qquad (z,u)\in H_1,$$

$$\tilde{\phi}(\zeta)=\left(\frac{\zeta}{1+|\zeta|},-1\right),\qquad \zeta\in\mathbb{C},$$

$$\tilde{\phi}(e^{i\varphi})=(e^{i\varphi},1),\qquad e^{i\varphi}\in S^1,$$

and let ϕ denote the corresponding mapping into $M'=\tilde{M}'/R$. Then it is easy to check that ϕ is bijective and continuous. But, since M is compact, ϕ is necessarily even a homeomorphism. $\square$

Theorem 3.2. (i) *M is the Martin compactification of H_1. Especially, the Martin boundary of H_1 is homeomorphic to the closed unit disk in $\mathbb{C}$, and the minimal Martin boundary is homeomorphic to the unit circle S^1 in $\mathbb{C}$.*

(ii) *If ζ is a point of the part $\mathbb{C}$ of Δ, then*

$$F_\zeta(z,u)=\frac{I_0(\sqrt{2}|z-\zeta|)}{I_0(\sqrt{2}|\zeta|)},$$

and if $e^{i\varphi}$ is a point of the part S^1 of Δ, then

$$F_{e^{i\varphi}}(z,u)=e^{\sqrt{2}\,\mathrm{Re}(e^{i\varphi}\bar{z})}.$$

Proof. First we shall show that for any point η in the boundary $\mathbb{C}\dot{\cup}S^1$ of M and any sequence $\{(z_n,u_n)\}_n$ in H_1 which converges to η in M, we have

$$\lim_{n\to\infty}F_{(z_n,u_n)}(z,u)=F_\eta(z,u) \tag{3.2}$$

pointwise for all $(z,u)\in H_1$. In fact, we shall show that any sequence $\{(z_n,u_n)\}_n$ which converges to η contains a subsequence $\{(z'_m,u'_m)\}_m$ such that

$$\lim_{m\to\infty} F_{(z'_m,u'_m)}(z,u)=F_\eta(z,u). \tag{3.2$'$}$$

This clearly is equivalent to (3.2).

In order to prove (3.2)', assume first that $\eta=\zeta\in\mathbb{C}$. Then we have $z_n\to\zeta$, and $|u_n|\to+\infty$. Hence, the asymptotics of Theorem 2.2 (ii) applies to the sequence $\{(z_n,u_n)\}_n$ as well as to the sequence $\{(z,u)^{-1}\cdot(z_n,u_n)\}_n$, and thus

$$F_{(z_n,u_n)}(z,u)\sim\frac{I_0(\sqrt{2}|z_n-z|)}{I_0(\sqrt{2}|z_n|)}a_n,$$

where

$$a_n=\left[\frac{|u_n|}{|u_n-u-2\operatorname{Im}(z\bar z_n)|}\right]^{3/4}\exp\left[\sqrt{2\pi|u_n|}-\sqrt{2\pi|u_n-u-2\operatorname{Im}(z\bar z_n)|}\right].$$

Since obviously $\lim\limits_{n\to\infty}a_n=1$, and since $z_n\to\zeta$, we hence obtain

$$\lim_{n\to\infty}F_{(z_n,u_n)}(z,u)=\frac{I_0(\sqrt{2}|\zeta-z|)}{I_0(\sqrt{2}|\zeta|)}.$$

Next assume that $\eta=e^{i\varphi}\in S^1$. Then, if $z_n=r_ne^{i\varphi_n}$, we have $r_n\to\infty$, and

$$e^{i[\tau(u_nr_n^{-2})+\varphi_n]}\to e^{i\varphi}.$$

Passing to a subsequence, if necessary, we may either assume that the sequence $\{u_nr_n^{-2}\}_n$ converges to a real number $\omega_\infty\in\mathbb{R}$, or that $|u_n|r_n^{-2}\to+\infty$.

In the first case, Theorem 2.1 applies to the arguments $\{(z_n,u_n)\}_n$ as well as to $\{(z'_n,u'_n)\}_n$, where

$$(z'_n,u'_n)=(z,u)^{-1}\cdot(z_n,u_n)=(z_n-z,u_n-u-2\operatorname{Im}(z\bar z_n)).$$

Then it is clear that

$$F_{(z_n,u_n)}(z,u)\sim e^{\sqrt{2}\Omega_n},$$

where

$$\Omega_n=\gamma\left(\frac{u_n}{|z_n|^2}\right)|z_n|-\gamma\left(\frac{u'_n}{|z'_n|^2}\right)|z'_n|.$$

Now, with $\omega_n=\dfrac{u_n}{|z_n|^2}$ and $\omega'_n=\dfrac{u'_n}{|z'_n|^2}$, we have

$$\gamma(\omega'_n)=\gamma(\omega_n)+\gamma'(\omega_n)(\omega'_n-\omega_n)+O(|\omega'_n-\omega_n|^2),$$

since $\lim\omega_n=\lim\omega'_n=\omega_\infty$.

Then a simple estimate shows that

$$\begin{aligned}\omega'_n-\omega_n&=\frac{u_n-u-2\operatorname{Im}(z\bar z_n)}{|z_n-z|^2}-\frac{u_n}{|z_n|^2}\\&=2\left[\frac{u_n\operatorname{Re}(z_n\bar z)}{|z_n|^4}-\frac{\operatorname{Im}(z\bar z_n)}{|z_n|^2}\right]+O(|z_n|^{-2}).\end{aligned}$$

Especially we have

$$|\omega_n' - \omega_n|^2 = O(|z_n|^{-2}).$$

One can also easily show that

$$|z_n'| = |z_n - z| = |z_n| - \frac{\operatorname{Re}(\bar{z} z_n)}{|z_n|} + O(|z_n|^{-1}).$$

So, together we obtain

$$\Omega_n = \gamma(\omega_n)\frac{\operatorname{Re}(\bar{z} z_n)}{|z_n|} - 2\gamma'(\omega_n)\left[\frac{u_n \operatorname{Re}(\bar{z} z_n)}{|z_n|^4} - \frac{\operatorname{Im}(z\bar{z}_n)}{|z_n|}\right] + O(|z_n|^{-1}).$$

Since $\frac{\bar{z} z_n}{|z_n|} = e^{i\varphi_n}\bar{z}$, and since $e^{i\varphi_n} \to e^{i[-\tau(\omega_\infty)+\varphi]}$, we get

$$\begin{aligned}\lim_{n\to\infty} \Omega_n = {} & \gamma(\omega_\infty)\operatorname{Re}(e^{i[-\tau(\omega_\infty)+\varphi]}\bar{z}) \\ & -2\omega_\infty\gamma'(\omega_\infty)\operatorname{Re}(e^{i[-\tau(\omega_\infty)+\varphi]}\bar{z}) \\ & -2\gamma'(\omega_\infty)\operatorname{Im}(e^{i[-\tau(\omega_\infty)+\varphi]}\bar{z}).\end{aligned}$$

But,

$$\gamma'(\omega) = \frac{\tau'(\omega)}{\sin^2\tau(\omega)}(\sin\tau(\omega) - \tau(\omega)\cos\tau(\omega)),$$

and

$$\theta'(\tau) = 2\frac{\sin\tau - \tau\cos\tau}{\sin^3\tau},$$

hence, since $\tau = \theta^{-1}$,

$$\gamma'(\omega) = \tfrac{1}{2}\sin\tau(\omega).$$

This yields, by definition of θ and τ,

$$\gamma(\omega) - 2\omega\gamma'(\omega) = \cos\tau(\omega),$$

hence

$$\begin{aligned}\lim_{n\to\infty} \Omega_n &= \cos\tau(\omega_\infty)\operatorname{Re}(e^{i[-\tau(\omega_\infty)+\varphi]}\bar{z}) - \sin\tau(\omega_\infty)\operatorname{Im}(e^{i[-\tau(\omega_\infty)+\varphi]}\bar{z}) \\ &= \operatorname{Re}[e^{i\tau(\omega_\infty)}e^{i[-\tau(\omega_\infty)+\varphi]}\bar{z}] \\ &= \operatorname{Re}[e^{i\varphi}\bar{z}].\end{aligned}$$

Thus we obtain

$$\lim_{n\to\infty} F_{(z_n,u_n)}(z,u) = e^{\sqrt{2}\operatorname{Re}(e^{i\varphi}\bar{z})}.$$

It remains the case where $|u_n| r_n^{-2} \to \infty$. Since $\lim_{\kappa\to\pm\infty} \tau(\kappa) = \pm\pi$, we have

$$\lim_{n\to\infty} e^{i\varphi_n} = -e^{i\varphi}.$$

Since in this case the asymptotics of Theorem 2.2 (i) applies, we have

$$F_{(z_n,u_n)}(z,u) \sim b_n e^{\sqrt{2}\Omega'_n},$$

where

$$b_n = \left[\frac{|u_n|}{|u_n - u - 2\,\mathrm{Im}(z\bar{z}_n)|}\right]^{3/4} \cdot \left[\frac{|z_n|}{|z - z_n|}\right]^{1/2},$$

$$\Omega'_n = [\pi|u_n| + |z_n|^2 - 2\varrho(\delta_n)\sqrt{\pi|u_n|\,|z_n|^2}]^{1/2} - [\pi|u'_n| + |z'_n|^2 - 2\varrho(\delta'_n)\sqrt{\pi|u'_n|\,|z'_n|^2}]^{1/2}.$$

Here we set $\delta_n = \sqrt{|z_n|^2/(\pi|u_n|)}$, $\delta'_n = \sqrt{|z'_n|^2/(\pi|u'_n|)}$, and (z'_n, u'_n) is defined as before.

Since $|z_n| \to \infty$ and $\delta_n \to 0$, we also have $|u_n| \to \infty$, and together this implies $b_n \to 1$. Moreover, a tedious, but straight-forward calculation similar to those before shows that

$$\Omega'_n = -\varrho(\delta_n)\frac{\mathrm{Re}(\bar{z}z_n)}{|z_n|} + \xi_n,$$

where $\lim_{n\to\infty} \xi_n = 0$. Since, by Theorem 1.3, $\varrho(\delta_n) = 1 + O(\delta_n^2)$, and since $e^{i\varphi_n} \to -e^{i\varphi}$, this implies

$$\lim_{n\to\infty} F_{(z_n,u_n)}(z,u) = e^{\sqrt{2}\,\mathrm{Re}(e^{i\varphi}\bar{z})}.$$

The proof of (3.2) is now complete.

But, since H_1 is dense in M, it is easy to see that (3.2) implies that the functions $F^{(z,u)}$ [extended to M by (ii)] are continuous with values in $[0, \infty]$ on the whole of M. Moreover, since for different $\eta_1, \eta_2 \in \mathbb{C} \cup S^1$ the functions F_{η_1} and F_{η_2} are obviously different, it is clear that the functions $F^{(z,u)}$, $(z,u) \in H_1$, separate the points of $M \setminus H_1$. Thus we have shown that M is the Martin compactification of H_1. Moreover, it is clear (for example by Lemma 3.1), that $\Delta = M \setminus H_1$ is homeomorphic to the closed unit disk in $\mathbb{C}$. Finally, by (ii) all functions F_η with $\eta \in \Delta$ are independent of the variable u. Thus, if we consider them as functions of the variable z, they are just eigenfunctions of the "classical" Laplacian on $\mathbb{C} \cong \mathbb{R}^2$, and so the well-known classical theory implies that the extremal rays of the cone Γ are just the rays $\mathbb{R}F_{e^{i\varphi}}$, $e^{i\varphi} \in S^1$. This shows that $\Delta_1 = S^1 \subset M$ and concludes the proof of Theorem 3.2. □

The following corollary to Theorem 3.2 is clear by (3.1):

Corollary 3.3. *If $h \geqq 0$ is a $(-\Delta_K + 2)$ harmonic function on H_1, then there exists a unique positive measure μ on $[0, 2\pi]$, such that*

$$h(x,y,u) = \int_0^{2\pi} e^{\sqrt{2}(x\cos\varphi + y\sin\varphi)}\,d\mu(\varphi)$$

for all $(x,y,u) \in H_1$. Especially, h does not depend on the variable u.

The readers who are only interested in a proof of this corollary should note, that one could proceed almost word by word as in [12] to derive this statement. In fact, Margulis' argument covers a large class of harmonic spaces on nilpotent Lie groups. Especially it shows that the immediate generalization of our corollary to sub-Laplacians on stratified Lie groups holds.

References

1. Brélot, M.: On topological boundaries in potential theory. (Lecture notes in mathematics, Vol. 175.) Berlin Heidelberg New York: Springer 1971
2. Courant, R.: Vorlesungen über Differential- und Integralrechnung 1. Berlin Heidelberg New York: Springer 1971
3. Cygan, J.: Fundamental solution of the heat equation on the Heisenberg group. Preprint, Polish academy of mathematics
4. Erdéily, A., Magnus, W., Oberhettinger, F., Tricomi, F.G.: (Higher transcendental functions, Vol. 2). New York: McGraw-Hill 1953
5. Folland, G.B., Stein, E.M.: Hardy spaces on homogeneous groups. Princeton: Princeton University Press 1982
6. Gaveau, B.: Principe de moindre action, propagation de la chaleur et estimées sous elliptiques sur certains groupes nilpotents. Acta Math. **139**, 59–153 (1977)
7. Helffer, B., Nourrigat, J.: Caractérisation des opérateurs hypoelliptiques homogènes invariants à gauche sur un groupe de Lie nilpotent gradué. Commun. Partial Differ. Equations **4**, 899–958 (1979)
8. Helms, L.L.: Introduction to potential theory. (Pure and applied Mathematics, Vol. 22.) New York London Sydney Toronto: Wiley 1969
9. Hervé, R.M., Hervé, M.: Les fonctions sur-harmoniques dans l'axiomatique de M. Brélot associeés à un opérateur elliptique dégénéré. Ann. Inst. Fourier **22**(2), 131–145 (1972)
10. Hörmander, L.: Hypoelliptic second-order differential equations. Acta Math. **119**, 147–171 (1967)
11. Hulanicki, A.: The distribution of energy in the Brownian motion in Gaussian field and analytic hypoellipticity of certain subelliptic operators on the Heisenberg group. Stud. Math. **56**, 165–173 (1976)
12. Margulis, G.A.: Positive harmonic functions on nilpotent groups. Sov. Math. **7**, 241–243 (1966)
13. Rockland, C.: Hypoellipticity for the Heisenberg group. Trans. Am. Math. Soc. **240**, 1–52 (1978)

Received July 1, 1987; in revised form April 26, 1988

Math. Ann. 283, 121–137 (1989)

Mathematische Annalen

Compactifications of $\mathbb{C}^3$. II

Thomas Peternell

Mathematisches Institut, Universität Bayreuth, Postfach 101251, D-8580 Bayreuth, Federal Republic of Germany

1. Introduction

This is the announced second part to the joint paper "Compactifications of $\mathbb{C}^3, I$" with M. Schneider.

In order to state the results of this paper and to put them in connection with the general theory remember that a compactification of $\mathbb{C}^3$ is a compact complex manifold X with a divisor $Y \subset X$ such that $X \setminus Y \simeq \mathbb{C}^3$ biholomorphically. We assume always $b_2(X)=1$ (i.e. Y is irreducible) and X projective (for the non-projective case see [P-S]). Remember that X is a Fano 3-fold of index r, i.e. ω_X^{-1} is ample and there is a generator $\mathscr{L} \in \mathrm{Pic}(X) \simeq \mathbb{Z}$ such that $\mathscr{L}^r \simeq \omega_X^{-1}$. One knows $r \leqq 4$ and the cases $r=4$ (resp. $r=3$) give the classical compactifications $\mathbb{P}^3$ (resp. Q_3, the 3-dimensional quadric) with divisors Y at infinity $\mathbb{P}_2$ resp. the quadric cone.

If $r=2$ Furushima [Fu 1] constructed a new compactification X with two possible divisors at infinity – one normal and one non-normal. By [Fu 1] and [P-S] these are the only ones for $r=2$. Observe that X is rational and has $b_3(X)=0$. Also by [P-S] and [Fu 2], if $r=1$ and Y is normal then X has to be rational with $b_3(X)=0$; i.e. X is a Fano 3-fold of "genus 12". Two such X are known, one constructed by Iskovskij [Is 1], one by Mukai-Umemura [M-U]. Probably these are the only ones (recently proved by Mukai as I understand). Both cases cannot be a compactification with normal divisor at infinity, as proved by Furushima [Fu 3]; the non-normal case is still undecided.

This paper now deals with the case $r=1$ and Y non-normal. The main result is the

Theorem. *Let X be a compactification of $\mathbb{C}^3$ with $b_2(X)=1$ and divisor Y at infinity. Assume that X is a Fano 3-fold of index 1 and Y non-normal.*

Then the genus $g(X) = -\dfrac{c_1(\omega_X)^3}{2} + 1 = 12$, *in particular X is rational and* $b_3(X)=0$.

Even in the case $g(X)=12$ we get some information: the non-normal locus E of Y consists of one or two smooth rational curves meeting transversely in one point;

the conductor ideal is reduced. If $f: \tilde{Y} \to Y$ is the normalization then $\tilde{E}$ – the analytic preimage of E – is reduced too and consists of two smooth rational curves meeting of order two in exactly one point. In particular, $b_3(\tilde{Y})=0$ and hence $\tilde{Y}$ and Y are rational.

In conclusion one can state now the following

Theorem. *Any projective compactification X of $\mathbb{C}^3$ with $b_2(X)=1$ is rational with $b_3(X)=0$.*

The only remaining open problems – besides a problem in the non-algebraic case (cp. [P-S]) – are the questions of existence for the Mukai-Umemura example (normal case) and the Iskovskij and the Mukai-Umemura example (non-normal case) – provided these are the only Fano 3-folds with $r=1$, $b_2(X)=1$, $g(X)=12$.

Some remark to the proof of the main theorem.

Let X be a compactification of $\mathbb{C}^3$ with $b_2(X)=1$ which is a Fano 3-fold of index 1. Let Y be the (irreducible) divisor at infinity. Assume that Y is non-normal. Let $f: \tilde{Y} \to Y$ be the normalization and $\pi: \hat{Y} \to \tilde{Y}$ a desingularization. It is easy to see $\kappa(\hat{Y}) = -\infty$. In order to show $b_3(Y)=0$ we have to do two things. First, we have to control the topology of $\hat{Y}$, namely, we want to prove $b_3(\tilde{Y})=b_3(Y)$. Second, we must prove the rationality of $\hat{Y}$ (i.e. the rationality of Y). Then $b_3(\hat{Y})=b_3(\tilde{Y})=0$, hence $b_3(Y)=0$.

The first problem is solved by analyzing carefully the map f, i.e. the non-normal locus $E \subset Y$ and its analytic preimage $\tilde{E} \subset \tilde{Y}$. The second one is treated by very special hyperplane sections (and by using the Iskovskij classification for X).

2. Preliminaries

(2.1) We recollect some notations and facts from [P-S]. A compactification of $\mathbb{C}^3$ is a pair (X, Y) consisting of a compact complex manifold X and an analytic subset $Y \subset X$ such that $X \setminus Y \simeq \mathbb{C}^3$ (biholomorphically). Necessarily Y is of pure dimension 2. We are only interested here in the case $b_2(X)=1$ which is the same as to say Y is irreducible. We also assume that X is projective. Then X is a Fano 3-fold, i.e. the canonical sheaf ω_X is negative. We treat in this paper the case of index 1, i.e. ω_X generates $\mathrm{Pic}(X) \simeq \mathbb{Z}$ (cp. [P-S], Sect. 0). For more properties of (X, Y) see again [P-S], Sect. 0.

(2.2) For a Fano 3-fold X the genus $g(X)$ is defined by

$$g(X) = -\frac{c_1(\omega_X)^2}{2} + 1 .$$

Iskovskij [Is 1, 2] proved that for a Fano 3-fold X of index 1 with $b_2(X)=1$ one has always $2 \leqq g(X) \leqq 12$ and $g(X) \neq 11$. Moreover these 3-folds can be classified (see [Is 1, 2]). The anticanonical bundle is always very ample except for two cases (cp. 3.14). If ω_X^{-1} is very ample, X is said to be of the principal series.

We need to know the Betti numbers $b_3(X)$, which are given by the following table (see [Is-Šo])

$g(X)$	$\frac{1}{2}b_3(X)$
2	52
3	30
4	20
5	14
6	10
7	5
8	4
9	3
10	2
12	0

(2.3) **Lemma.** *Let X be a purely 1-dimensional reduced projective complex space. Let $X' \subset X$ be an irreducible component with the reduced structure. Then $\omega_{X'}$ is a subsheaf of ω_X (ω denotes the dualizing sheaf), moreover*

$$\omega_X|\omega_{X'} \simeq \mathcal{H}om_{\mathcal{O}_X}(\mathcal{J}_{X'/X}, \omega_X).$$

Proof. It is well known (see e.g. [A-K]) that

$$\omega_{X'} \simeq \mathcal{H}om_{\mathcal{O}_X}(\mathcal{O}_{X'}, \omega_X). \tag{1}$$

Let $\mathcal{J}$ be the ideal sheaf of X' in X. Then from the exact sequence $0 \to \mathcal{J} \to \mathcal{O}_X \to \mathcal{O}_{X'} \to 0$ we obtain:

$$0 \to \mathcal{H}om_{\mathcal{O}_X}(\mathcal{O}_{X'}, \omega_X) \to \omega_X \to \mathcal{H}om_{\mathcal{O}_X}(\mathcal{J}, \omega_X) \to \mathcal{E}xt^1_{\mathcal{O}_X}(\mathcal{O}_{X'}, \omega_X). \tag{*}$$

So by (1) it is sufficient to show:

$$\mathcal{E} = \mathcal{E}xt^1_{\mathcal{O}_X}(\mathcal{O}_{X'}, \omega_X) = 0. \tag{2}$$

Let $\mathcal{O}_X(1)$ be an ample line bundle on X.

Take $v \gg 0$ such that $\mathcal{E} \otimes \mathcal{O}_X(v)$ is globally generated.

Then it is sufficient to show

$$H^0(\mathcal{E}(v)) = 0. \tag{2'}$$

But $H^0(\mathcal{E}(v)) \simeq \mathrm{Ext}^1_{\mathcal{O}_X}(\mathcal{O}_{X'}, \omega_X(v))$.

Since X is 1-dimensional, it is Cohen-Macaulay. So by Serre duality:

$$\mathrm{Ext}^1_{\mathcal{O}_X}(\mathcal{O}_{X'}, \omega_X(v)) \simeq H^0(\mathcal{O}_{X'}(-v)) = 0$$

which proves (2′), hence (2).

(2.4) **Lemma.** *Let X be a purely 1-dimensional projective Cohen-Macaulay space. Let $\mathscr{n}$ be the sheaf of nilpotent functions on X. Then there is an exact sequence*

$$0 \to \omega_{\mathrm{red}\,X} \to \omega_X \to \mathscr{H}om_{\mathcal{O}_X}(\mathscr{n}, \omega_X) \to 0 .$$

Proof. Same as that of (*) and (2) in the proof of (2.3).

3. The Main Result

Let (X, Y) always denote a smooth projective compactification of $\mathbb{C}^3$ with $b_2(X) = 1$. Then X is a Fano 3-fold of genus $g(X)$. We assume that X is of index 1, i.e. ω_X generates $\mathrm{Pic}(X) \simeq \mathbb{Z}$. Moreover let Y be non-normal.

(3.1) **Theorem.** *X has genus $g(X) = 12$, moreover $b_3(X) = 0$ and X is rational.*

The proof will follow from several propositions of this section.

(3.2) We denote by E the non-normal locus of Y equipped with the structure given by the conductor ideal. Let $\tilde{E}$ be the analytic preimage of E with respect to the normalization $f: \tilde{Y} \to Y$. In order to prove (3.1) it will be sufficient (by the Iskovskij classification, 2.2) to show $g(X) > 10$, equivalently $b_3(X) = 0$.

We denote by r (resp. $\tilde{r}$) the number of irreducible components of E (resp. $\tilde{E}$).

(3.3) **Proposition.** 1. *$H^1(\mathcal{O}_E) = H^1(\mathcal{O}_{\mathrm{red}\,E}) = 0$; in particular all components E_i of* red E *are smooth rational curves.*

2) $\omega_{\tilde{E}} \simeq \mathcal{O}_{\tilde{E}}$.

Proof. 1. Using the exact sequence [Mo, 3.34.2]

$$0 \to \mathcal{O}_Y \to f_*(\mathcal{O}_{\tilde{Y}}) \to \omega_E \to 0$$

and $\omega_Y \simeq \mathcal{O}_Y$, we obtain by $H^1(\mathcal{O}_Y) = 0$:

$$0 = H^0(\omega_E) \simeq H^1(\mathcal{O}_E),$$

since E is Cohen-Macaulay (see e.g. [Mo, K-W, S]). Letting $\mathscr{n}$ be the sheaf of nilpotent functions on E and taking cohomology from

$$0 \to \mathscr{n} \to \mathcal{O}_E \to \mathcal{O}_{\mathrm{red}\,E} \to 0$$

we obtain $H^1(\mathcal{O}_{\mathrm{red}\,E}) = 0$.

Last $H^1(\mathcal{O}_{E_i}) = 0$ follows in the same spirit, so $E_i \simeq \mathbb{P}_1$.

2) This follows from $\omega_{\tilde{E}} \mp f^*(\omega_Y) \otimes \mathcal{O}_{\tilde{E}}$ ([Mo, 3.34.1]) and $\omega_Y \simeq \mathcal{O}_Y$.

If $\tilde{E}$ is reduced we can say immediately a lot on the structure of $\tilde{E}$:

(3.4) **Proposition.** *Assume that $\tilde{E}$ is reduced.*

a) *If a connected component of $\tilde{E}$ consists of exactly one irreducible component $\tilde{E}_i$, then $\tilde{E}_i$ is a torus or a singular rational curve with $\omega_{\tilde{E}_i} \simeq \mathcal{O}_{\tilde{E}_i}$, i.e. a cubic in $\mathbb{P}_2$.*

b) *If a connected component of $\tilde{E}$ consists of more than one irreducible component, then all these components $\tilde{E}_i$ are smooth and rational.*

Proof. a) By (3.3, 1) we have $\omega_{\tilde{E}_i} \simeq \mathcal{O}_{\tilde{E}_i}$.

b) By (2.3) $\omega_{\tilde{E}_i}$ is a proper subsheaf of $\omega_{\tilde{E}}$, i.e. of $\mathcal{O}_{\tilde{E}}$. Hence $H^0(\omega_{\tilde{E}_i}) = 0$ and $H^1(\mathcal{O}_{\tilde{E}_i}) = 0$, i.e. $\tilde{E}_i \simeq \mathbb{P}_1$.

(3.5) **Proposition.** $H^1(\mathcal{O}_{\tilde{Y}})=0$; $h^0(\mathcal{O}_E)=h^0(\mathcal{O}_{\tilde{E}})=h^1(\mathcal{O}_{\tilde{E}})=1$.

Proof. Let $g=h^1(\mathcal{O}_{\tilde{Y}})$. By the exact sequence

$$0\to\mathcal{O}_Y\to f_*(\mathcal{O}_{\tilde{Y}})\to\omega_E\to 0$$

we obtain $h^1(\mathcal{O}_{\tilde{Y}})=h^1(\omega_E)-1=h^0(\mathcal{O}_E)-1=\mu-1$ by definition of μ.

Since $\omega_{\tilde{E}}\simeq\mathcal{O}_{\tilde{E}}$, we have $h^0(\mathcal{O}_{\tilde{E}})\simeq h^1(\mathcal{O}_{\tilde{E}})$. Now by the exact sequence [Mo, 3.36.2]

$$0\to\mathcal{O}_E\to f_*(\mathcal{O}_{\tilde{E}})\to\omega_E\to 0$$

we see $h^0(\mathcal{O}_E)=h^0(\mathcal{O}_{\tilde{E}})$ since $h^0(\omega_E)=h^1(\mathcal{O}_E)=0$. Thus:

$$g+1=\mu=h^0(\mathcal{O}_E)=h^0(\mathcal{O}_{\tilde{E}})=h^1(\mathcal{O}_{\tilde{E}})\,. \tag{1}$$

Hence it is sufficient to prove $g=0$. Assume $g>0$, so $\mu>1$. Let $\pi:\hat{Y}\to\tilde{Y}$ be a minimal desingularization of $\tilde{Y}$, $\sigma:\hat{Y}\to Y_m$ a minimal model. Since $\kappa(Y)=-\infty$ and $g>0$, Y_m is a non-rational ruled surface. Let $p:Y_m\to C_m$ be a ruling. Let Z be the exceptional set of π. Then I claim:

$$\dim p\circ\sigma(Z)=0\,. \tag{2}$$

Assume $\dim p\circ\sigma(Z)>0$ and let $S_1,\dots,S_q$ be the irreducible components of Z with $\dim p\circ\sigma(S_i)=1$.

Since $\omega_{\tilde{Y}}$ is a proper subsheaf of $\mathcal{O}_{\tilde{Y}}$ by [Mo, 3.34.3], $H^2(\mathcal{O}_{\tilde{Y}})=0$. Moreover $H^2(\mathcal{O}_{\hat{Y}})=0$. So from the Leray spectral sequence we have

$$H^1(\mathcal{O}_{\hat{Y}})\simeq H^1(\mathcal{O}_{\tilde{Y}})\oplus H^0(R^1\pi_*(\mathcal{O}_{\hat{Y}}))\,. \tag{3}$$

By Riemann-Hurwitz $g(S_i)\geqq\tilde{g}:=g(C_m)$.

Hence $h^1(\mathcal{O}_{S_i})\geqq\tilde{g}$ and consequently $h^1(\mathcal{O}_Z)\geqq q\cdot\tilde{g}$ where Z carries the reduced structure. This last fact is clear by considering the normalization of Z.

Let $\pi(Z)=\{y_1,\dots,y_t\}$. Define λ_i by

$$R^1\pi_*(\mathcal{O}_{\hat{Y}})_{y_i}\simeq\mathbb{C}^{\lambda_i}\,.$$

Then $\sum\limits_{i=1}^{t}\lambda_i\geqq q\cdot\tilde{g}$ because the restriction map $H^1(\mathcal{O}_{\hat{Z}})\to H^1(\mathcal{O}_Z)$, $\hat{Z}$ the completion of Z, is surjective, then use Grauert's comparison theorem. Hence (3) implies: $\tilde{g}=g+\sum\lambda_i\geqq g+q\cdot\tilde{g}$. So $q=0$ since $g>0$, i.e. $\dim p\circ\sigma(Z)=0$ and (2) is proved.

We thus find $x_1,\dots,x_k\in C_m$ such that

$$Z\subset\bigcup_{i=1}^{k}(p\circ\sigma)^{-1}(x_i)\,.$$

Hence from $H^1((p\circ\sigma)^{-1}(x_i),\mathcal{O})=0$, we obtain $H^1(\mathcal{O}_Z)=0$. Since $R^1(p\circ\sigma)_*(\mathcal{O}_{\hat{Y}})=0$, even $H^1((p\circ\sigma)^{-1}(x_i)_\mu,\mathcal{O})=0$ for any infinitesimal neighborhood $(p\circ\sigma)^{-1}(x_i)_\mu$, thus $H^1(\mathcal{O}_{Z_\mu})=0$ for any infinitesimal neighborhood Z_μ of Z, consequently $H^1(\mathcal{O}_{\hat{Z}})=0$.

This proves $R^1\pi_*(\mathcal{O}_{\hat{Y}})=0$ and $\tilde{g}=g$ by (3). We next prove

$$\dim p\circ\sigma(\hat{E})=1\,, \tag{4}$$

where $\hat{E}$ is the strict transform of $\tilde{E}$ in $\hat{Y}$. In fact, otherwise we would get [using (2)] a lot of smooth rational curves l in $Y\backslash E$. Then l is a Cartier divisor in Y, so

$\mathcal{O}_Y(l) \simeq \mathcal{O}_Y(k)$ for some $k \in \mathbb{N}$. Since $\mathrm{Pic}(X) \xrightarrow{\sim} \mathrm{Pic}(Y)$ we can write $l = Y \cap H$ for some hypersurface $H \subset X$. But $(H \cdot E) > 0$, so $H \cap E = \emptyset$ is not possible. This proves (4).

By (4) we find a component, say $\hat{E}_1$, of $\hat{E}$ such $p \circ \sigma(\hat{E}_1) = C_m$. So $g(\hat{E}_1) \geqq \tilde{g} = g$.

Hence $\mathrm{red}\,\tilde{E}$ contains a curve of genus $\geqq g$ implying $h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) \geqq g$. On the other hand by (1): $h^1(\mathcal{O}_{\tilde{E}}) = g + 1$. So we obtain the inequality

$$g + 1 = h^1(\mathcal{O}_{\tilde{E}}) \geqq h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) \geqq g\,. \tag{5}$$

Now consider $\hat{E}$, the strict transform of E, equipped with the reduced structure. If $g > 1$, the component $\hat{E}_1$ described above is unique because otherwise $h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) \geqq 2g$.

(A) So assume $g > 1$ for the moment. Since all other components of $\hat{E}$ are components of fibers $(p \circ \sigma)^{-1}(x)$, we easily get $h^1(\mathcal{O}_{\hat{E}}) = g$.

Exactly the same arguments apply to $\bar{E} = \pi^{-1}(\tilde{E})$ with the reduced structure [use (2)!]. So $h^1(\mathcal{O}_{\bar{E}}) = g$. Now clearly $\mathcal{O}_{\mathrm{red}\,\tilde{E}} \simeq \pi_*(\mathcal{O}_{\bar{E}})$, so we conclude

$$h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = g\,, \tag{6}$$

i.e. (5) becomes a strict inequality and (5) excludes the case $\tilde{E} = \mathrm{red}\,\tilde{E}$.

Assume that $\tilde{E}_1$ is non-reduced. Let $\mathcal{J}$ be the ideal sheaf of $\mathrm{red}\,\tilde{E}$ in $\tilde{E}$. Then by (3.7a) below (for $\mu = 1$) we obtain:

$$h^1(\tilde{E}, \mathcal{O}_{\tilde{E}}/\mathcal{J}^2) \geqq 2g\,,$$

since $h^1(\mathcal{J}/\mathcal{J}^2) \geqq g$. Namely, $\mathcal{J}/\mathcal{J}^2 | \mathrm{red}\,\tilde{E}_1 \simeq \mathcal{O}_{\mathrm{red}\,\tilde{E}_1}$ modulo torsion (3.7).

Hence $h^1(\mathcal{O}_{\tilde{E}}) \geqq 2g$, contradiction.

So $\tilde{E}_1$ is reduced. But then clearly $\omega_{\tilde{E}_1}$ is a subsheaf of $\omega_{\tilde{E}} | \tilde{E}_1 \simeq \mathcal{O}_{\tilde{E}_1}$ [cf. (2.3), $\tilde{E}_1$ is smooth, so $\omega_{\tilde{E}_1}$ is locally free!]

Hence $g(\tilde{E}_1) \leqq 1$, contradiction.

(B) So we are reduced to the case $g = 1$. Then

$$1 \leqq h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) \leqq 2\,.$$

(B_1) First let $h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = 2$.

Letting $\mathcal{O}_{\tilde{E}_{(1)}} = \mathcal{O}_{\tilde{E}}/\mathcal{J}^2$ we have the exact sequence

$$0 \to \mathcal{J}/\mathcal{J}^2 \to \mathcal{O}_{\tilde{E}_{(1)}} \to \mathcal{O}_{\mathrm{red}\,\tilde{E}} \to 0\,.$$

Taking cohomology gives the exact sequence (5):

$$\begin{aligned} 0 &\to H^0(\mathcal{J}/\mathcal{J}^2) \to H^0(\mathcal{O}_{\tilde{E}_{(1)}}) \xrightarrow{\gamma} H^0(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) \to H^1(\mathcal{J}/\mathcal{J}^2) \\ &\to H^1(\mathcal{O}_{\tilde{E}_{(1)}}) \to H^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) \to 0\,. \end{aligned}$$

Since $h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = 2$ and since $h^1(\mathcal{O}_{\tilde{E}}) = 2 \geqq h^1(\mathcal{O}_{E_{(1)}})$ we obtain $h^1(\mathcal{O}_{\tilde{E}_{(1)}}) = 2$.

So γ being surjective, $h^1(\mathcal{J}/\mathcal{J}^2) = 0$.

We see that the components $\tilde{E}_i$ of $\tilde{E}$ of genus 1 (there are exactly one or two!) must be reduced because otherwise again $\mathcal{J}/\mathcal{J}^2 | \mathrm{red}\,\tilde{E}_i \simeq \mathcal{O}_{\mathrm{red}\,\tilde{E}_i}$ modulo torsion by (3.7a), hence $h^1(\mathcal{J}/\mathcal{J}^2) > 0$. But then $\omega_{\tilde{E}_i}$ is again a subsheaf of $\mathcal{O}_{\tilde{E}_i}$. $\tilde{E}_i$ being of genus 1, we conclude by Sect. 2 that $\tilde{E}_i$ is a connected component of $\tilde{E}$.

First assume that there are two elliptic components, say $\tilde{E}_1$ and $\tilde{E}_2$. Since $h^0(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = 2$, we conclude $\tilde{E} = \tilde{E}_1 \cup \tilde{E}_2$, i.e. $\tilde{r} = 2$. So $r \leqq 2$. By the exact sequence

([Ba-Ka, 3.A.7])

$$\begin{array}{ccccccccccc} 0 \to H^2(Y,\mathbb{Z}) & \to & H^2(\tilde{Y},\mathbb{Z}) \oplus H^2(E,\mathbb{Z}) & \xrightarrow{\alpha} & H^2(\tilde{E},\mathbb{Z}) & \to & H^3(Y,\mathbb{Z}) & \to & H^3(\tilde{Y},\mathbb{Z}) \to 0 \\ \downarrow \simeq & & & & \downarrow \simeq & & \downarrow \simeq & & \\ \mathbb{Z} & & & & \mathbb{Z}^r & & \mathbb{Z}^2 & & \end{array}$$

we obtain $2 \leqq b_3(Y) \leqq b_3(\tilde{Y}) + 2 = 4$.

Since $\alpha \neq 0$ (α is the canonical "difference map") we must have $b_3(Y) \leqq 3$, so $b_3(Y) = 2$ since $b_3(Y)$ is even. So $b_3(X) = 2$. But by Iskovskij there is no Fano 3-fold X with $b_3(X) = 2$. Hence there is a unique elliptic component $\tilde{E}_1$. With the same arguments we exclude the case $\tilde{E} = \tilde{E}_1 \cup \tilde{E}_2$, $\tilde{E}_2$ a singular rational cubic in $\mathbb{P}_2$. Since $h^0(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = 2$, there is a second connected component $\tilde{E}'$ consisting of smooth rational curves. If $\tilde{E}'$ is not reduced we conclude by the method of (3.7) that $H^1(\mathcal{J}^\mu/\mathcal{J}^{\mu+1}) = 0$ for all μ. Namely, all components of $\tilde{E}'$ have to be smooth [if some is singular then by $h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}'}) = 1$ it has to be a singular cubic in $\mathbb{P}_2$ whence $h^1(\mathcal{J}/\mathcal{J}^2) > 0$, contradiction]. In fact, since we know $H^1(\mathcal{J}/\mathcal{J}^2) = 0$, and all $\tilde{E}'_i \simeq \mathbb{P}_1$ it is an easy exercise to exclude the only other possible case $H^1(\mathcal{J}^\mu/\mathcal{J}^{\mu+1}) \simeq \mathbb{C}$ (look at the normalization of $\mathrm{red}\,\tilde{E}'$). Then using the higher analogs of (5) (for the infinitesimal neighborhoods of $\mathrm{red}\,\tilde{E}$ in $\tilde{E}$) we get the contradiction $h^0(\mathcal{O}_{\tilde{E}}) \geqq 3$.

(The contradiction can also be derived directly from $h^1(\mathcal{J}^\mu/\mathcal{J}^{\mu+1}) \leqq 1$ using (5) since $h^0(\mathcal{J}^\mu/\mathcal{J}^{\mu+1}) > 0$ as long as $\mathcal{J}^\mu/\mathcal{J}^{\mu+1} \neq 0$.)

So $\tilde{E}'$ is reduced. We know that $H^1(\mathcal{O}_{\tilde{E}'}) \simeq \mathbb{C}$. Let $\bar{E}' := \pi^{-1}(\tilde{E}')$. Then $H^1(\mathcal{O}_{\bar{E}'}) = 0$ since $\dim p \circ \sigma(\bar{E}') = 0$ and $\pi_*(\mathcal{O}_{\bar{E}'}) \simeq \mathcal{O}_{\tilde{E}'}$. This contradicts $H^1(\mathcal{O}_{\tilde{E}'}) \simeq \mathbb{C}$.

(B_2) We are left with the case $h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = 1$. Now the elliptic component of $\tilde{E}$ is uniquely determined. Call it $\tilde{E}_1$.

First let us see that $\tilde{E}$ must be connected. In fact, by (3.7a) we see that $h^0(\mathcal{J}^\mu/\mathcal{J}^{\mu+1}) > 0$ as long as $\mathcal{J}^\mu/\mathcal{J}^{\mu+1} \neq 0$, and that $h^1(\mathcal{J}^\mu/\mathcal{J}^{\mu+1}) \leqq 1$. So by $h^0(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) \geqq 2$ we would obtain [using (5)] that $h^0(\mathcal{O}_{\tilde{E}_{(\nu)}}) = h^0(\mathcal{O}_{\tilde{E}}/\mathcal{J}^{\nu+1}) \geqq 3$ for all ν, in particular $h^0(\mathcal{O}_{\tilde{E}}) \geqq 3$, contradiction.

Next I claim that

$$E_1 = f(\tilde{E}_1) \tag{7}$$

is not reduced.

Assume that E_1 is reduced. Then we find some j such that E_j is non-reduced [otherwise we would find $h^0(\mathcal{O}_E) = 1$, E being connected!]. Hence by [K-W] for general $y_0 \in E_j$ the formal local ring $\mathcal{O}_{Y,y_0}$ is not of the form $(F_1)\mathbb{C}[\![X, Y]\!]/(X \cdot Y)$ and not of the form $(F_2)\mathbb{C}[\![X, Y]\!]/(X^2 - Y^3)$. Now by Iskovskij we find through any y_0 a conic $l \subset X$. Since $(l \cdot Y) = 2$, we conclude $l \subset Y$. Namely, assume $l \not\subset Y$. If Y is irreducible at y_0 (for generic y_0), then f is locally around y_0 a homeomorphism and (by [K-W] and [S, 1.2.20]) we are in situation (F_2). Otherwise, if Y is reducible at y_0, we can locally only have two smooth irreducible components of Y meeting transversely, i.e. $\mathcal{O}_{Y,y_0} \simeq \mathbb{C}\{X, Y, Z\}/(X \cdot Y)$. So we are in situation (F_1).

Thus we have $l \subset Y$ and Y is filled up by conics. The strict transforms $\hat{l}$ of those conics l are contracted by $p \circ \sigma$ (since $g = 1$). By construction the general $\hat{l}$ meets a fixed component $\hat{E}_k$ with $f \circ \pi(\hat{E}_k) = E_j$. So $p \circ \sigma(\hat{E}_k) = C_m$, hence $\hat{E}_k = \hat{E}_1$, contradiction. This proves (7).

The same argument shows that E_1 is the only non-reduced component of E and that only finitely many conics meet E_j, $j \geqq 2$.

Now let $\tilde{E}_0$ be the non-reduced part of $\tilde{E}$; $\tilde{E}_1 \subset \tilde{E}_0$. The components of $\tilde{E}$ not belonging to $\tilde{E}_0$ are smooth rational curves meeting exactly one component of $\operatorname{red} \tilde{E}_0$ transversely in one point (use the exact sequence

$$0 \to \omega_{\tilde{E}_j} \to \mathcal{O}_{\tilde{E}} \to \mathcal{H}om(\mathcal{I}_{\tilde{E}_j}, \mathcal{O}_{\tilde{E}}) \to 0).$$

Let $\mathcal{J}$ be the ideal of $\tilde{E}_0$ in $\tilde{E}$. Then by

$$0 \to \omega_{\tilde{E}_0} \to \mathcal{O}_{\tilde{E}} \to \mathcal{H}om(\mathcal{J}, \mathcal{O}_{\tilde{E}}) \to 0$$

we obtain

$$\begin{array}{c} 0 \to H^0(\omega_{\tilde{E}_0}) \to H^0(\mathcal{O}_{\tilde{E}}) \to H^0(\mathcal{H}om(\mathcal{J}, \mathcal{O}_{\tilde{E}})) \\ \wr\| \\ \mathbb{C}^2 \\ \to H^1(\omega_{\tilde{E}_0}) \to H^1(\mathcal{O}_{\tilde{E}}) \to H^1(\mathcal{H}om(\mathcal{J}, \mathcal{O}_{\tilde{E}})) \to 0. \\ \wr\| \\ \mathbb{C}^2 \end{array}$$

$\operatorname{supp} \mathcal{H}om \not\subseteq (\mathcal{J}, \mathcal{O}_{\tilde{E}}) = \bigcup_{j \in J} \tilde{E}_j$, where $j \in J \Leftrightarrow \tilde{E}_j$ is reduced.

All these $\tilde{E}_j$ are disjoint, and $\mathcal{J} | \tilde{E}_j \simeq \mathcal{O}(-2)$, so

$$h^0(\mathcal{H}om(\mathcal{J}, \mathcal{O}_{\tilde{E}})) = 3r_0, \quad r_0 = \# J$$

and $h^1(\mathcal{H}om(\mathcal{J}, \mathcal{O}_{\tilde{E}})) = 0$.

Now $\tilde{E}_0$ is Cohen-Macaulay, so

$$h^1(\omega_{\tilde{E}_0}) = h^0(\mathcal{O}_{\tilde{E}_0}) = 2,$$

hence the above sequence gives

$$3r_0 \leqq 2, \quad \text{so} \quad r_0 = 0.$$

But $r_0 = 0$ implies:

$$r = 1: \tag{8}$$

assume that there is an reduced irreducible component $E_2 \subset E$. We have seen above (when we proved reducedness of E_j) that through a general point $y \in E_2$ we cannot find a conic in Y. Moreover either Y is a topological manifold around y or $\mathcal{O}_{Y,y} \simeq \mathbb{C}\{X, Y, Z\}/(X \cdot Y)$ (otherwise we would find conics). But then by [5, 1.18, 1.2.20] we can conclude that generically $f^{-1}(E_2)$ is reduced, hence reduced, contradiction. Thus $r = 1$ and (8) is proved.

As seen above, through any point of E there is a conic in Y. The strict transform of the conics in $\hat{Y}$ are contracted by $p \circ \sigma$. Thus the images of the fibers $(p \circ \sigma)^{-1}(x)$ are just the conics in Y.

We want to prove (9): $\tilde{r} = 1$. Assume $\tilde{r} > 1$. Then take $\tilde{E}_2 \subset \tilde{E}_0$, $\tilde{E}_2 \neq \tilde{E}_1$. Since $\dim p \circ \sigma(\hat{E}_2) = 0$ and since $f(\tilde{E}_2) = E_1 = E$ (set-theoretically), we conclude that E_1 is a conic or a line. Now $E_1 = E$ being the non-normal locus of Y, we have for the conormal bundles of $\operatorname{red} E_1$ in Y resp. X:

$$N^*_{\operatorname{red} E_1 | Y} \simeq N^*_{\operatorname{red} E_1 | X}$$

(see [P-S, proof of 2.3]).

By [Is 1] we know

$$N^*_{\operatorname{red} E_1|X} \simeq \begin{cases} \left.\begin{matrix} \mathcal{O}\oplus\mathcal{O} \\ \mathcal{O}(-1)\oplus\mathcal{O}(1) \\ \mathcal{O}(-2)\oplus\mathcal{O}(2) \\ \mathcal{O}(-4)\oplus\mathcal{O}(4) \end{matrix}\right\} & \text{conic case} \\ \left.\begin{matrix} \mathcal{O}(1) \quad \oplus\mathcal{O} \\ \mathcal{O}(2) \quad \oplus\mathcal{O}(-1) \end{matrix}\right\} & \text{line case.} \end{cases}$$

Let $(\operatorname{red} E)_1$ denote the 1st infinitesimal neighborhood of $(\operatorname{red} E)$ in Y.

Then by the exact sequence

$$0 \to H^0(N^*_{\operatorname{red} E|X}) \to H^0(\mathcal{O}_{(\operatorname{red} E)_1}) \to H^0(\mathcal{O}_{(\operatorname{red} E)}) \to 0$$

we get by the table for $N^*: h^0(\mathcal{O}_{(\operatorname{red} E)_1}) \geqq 3$, hence $h^0(\mathcal{O}_{(\operatorname{red} \tilde{E})_1}) \geqq 3$.

Now let $\tilde{\mathcal{J}}$ be the ideal sheaf of $\operatorname{red} \tilde{E}$ in $\tilde{E}$. Then consider

$$\begin{aligned} 0 &\to H^0(\tilde{\mathcal{J}}^\nu/\tilde{\mathcal{J}}^{\nu+1}) \to H^0(\mathcal{O}_{(\operatorname{red} \tilde{E})_\nu} \to H^0(\mathcal{O}_{(\operatorname{red} \tilde{E})_{\nu-1}}) \\ &\to H^1(\tilde{\mathcal{J}}^\nu/\tilde{\mathcal{J}}^{\nu+1}) \to H^1(\mathcal{O}_{(\operatorname{red} \tilde{E})_\nu}) \to H^1(\mathcal{O}_{(\operatorname{red} \tilde{E})_{\nu-1}}) \to 0 . \end{aligned}$$

Here $(\operatorname{red} \tilde{E})_\nu$ denotes the ν-th infinitesimal neighborhood of $\operatorname{red} \tilde{E}$ in $\tilde{E}$. Since $h^1(\tilde{\mathcal{J}}^\nu/\tilde{\mathcal{J}}^{\nu+1}) \leqq 1$ for all ν (3.7) and since $h^0(\mathcal{O}_{\tilde{E}}) = h^1(\mathcal{O}_{\tilde{E}}) = 2$ we conclude $h^0(\mathcal{O}_{(\operatorname{red} \tilde{E})_\nu}) \leqq 2$ for all ν, contradiction and (9) is shown.

So $\tilde{r} = 1$. Similar as in the case $(B1)$ we obtain by the exact sequence [Ba-Ka, 3.A.7]:

$$b_3(Y) = b_3(\tilde{Y}) = 2$$

and a contradiction as in $(B1)$.

This ends the proof of (3.5).

(3.6) **Proposition.** *$\tilde{E}$ is non-reduced iff* $H^1(\mathcal{O}_{\operatorname{red} \tilde{E}}) = 0$.

Proof. If $H^1(\mathcal{O}_{\operatorname{red} \tilde{E}}) = 0$, clearly $\tilde{E} \neq \operatorname{red} \tilde{E}$ since $H^1(\mathcal{O}_{\tilde{E}}) \simeq \mathbb{C}$. So assume $\tilde{E}$ non-reduced. Let $\tilde{n} \subset \mathcal{O}_{\tilde{E}}$ be the sheaf of nilpotent functions on $\tilde{E}$. Then by (2.4) we have the exact sequence

$$0 \longrightarrow \omega_{\operatorname{red} \tilde{E}} \longrightarrow \mathcal{O}_{\tilde{E}} \xrightarrow{\psi} \mathcal{H}om_{\mathcal{O}_{\tilde{E}}}(\tilde{n}, \mathcal{O}_{\tilde{E}}) \longrightarrow 0 .$$

Taking cohomology and using $H^0(\mathcal{O}_{\tilde{E}}) \simeq \mathbb{C}$ (3.5), moreover $H^0(\psi) \neq 0$, it follows $H^0(\omega_{\operatorname{red} \tilde{E}}) = 0$, i.e. $H^1(\mathcal{O}_{\operatorname{red} \tilde{E}}) = 0$, $\operatorname{red} \tilde{E}$ being Cohen-Macaulay.

(3.7) **Proposition.** a) *Let $\tilde{E}_j$ be a non-reduced component of $\tilde{E}$ such that* $\operatorname{red} \tilde{E}_j$ *is smooth. Then, letting $\mathcal{J}$ be the ideal sheaf of* $\operatorname{red} \tilde{E}$ *in* $\tilde{E}$, $(\mathcal{J}/\mathcal{J}^2|\operatorname{red} \tilde{E}_j)/_{\text{torsion}} \simeq \mathcal{O}_{\operatorname{red} \tilde{E}_j}$, *and* $(\mathcal{J}^\mu/\mathcal{J}^{\mu+1}|\operatorname{red} \tilde{E}_j)/_{\text{torsion}}$ *either contains the subsheaf* $\mathcal{O}_{\operatorname{red} \tilde{E}_j}$ *or is* 0.

b) *$\tilde{E}$ is reduced.*

Remark. (3.7, a) will be proved independently of (3.5)!

Proof. a) Denote by $\tilde{E}_1, \ldots, \tilde{E}_s$ the irreducible components of $\tilde{E}$ with the induced structures (so $\tilde{E}_i =$ the biggest subspace of $\tilde{E}$ with underlying reduced space $\operatorname{red} \tilde{E}_i$). By (2.4) there is an exact sequence

$$0 \to \omega_{\operatorname{red} \tilde{E}} \to \omega_{\tilde{E}} \to \mathcal{H}om_{\mathcal{O}_{\tilde{E}}}(\mathcal{J}, \mathcal{O}_{\tilde{E}}) \to 0 . \tag{1}$$

Restricting (1) to $\mathrm{red}\,\tilde{E}_j$ gives

$$\omega_{\mathrm{red}\,\tilde{E}}|\mathrm{red}\,\tilde{E}_j \xrightarrow{\alpha_j} \mathcal{O}_{\mathrm{red}\,\tilde{E}_j} \longrightarrow \mathcal{H}om(\mathcal{I}, \mathcal{O}_{\tilde{E}})|\mathrm{red}\,\tilde{E}_j \longrightarrow 0. \tag{2}$$

If $\tilde{E}$ is reduced, α_j is (generically) injective; if $\tilde{E}_j$ is non-reduced, $\alpha_j = 0$ (observe that $\tilde{E}_j$ then is non-reduced everywhere because $\tilde{E}$ is a Weil divisor on the normal surface Y!).

Now take $\tilde{E}_j$ non-reduced.

We consider the canonical map

$$\phi: \mathcal{H}om(\mathcal{I}, \mathcal{O}_E)|\mathrm{red}\,\tilde{E}_j \to \mathcal{H}om(\mathcal{I}/\mathcal{I}^2|\mathrm{red}\,\tilde{E}_j, \mathcal{O}_{\mathrm{red}\,\tilde{E}_j}).$$

Because of (2), $\mathcal{H}om(\mathcal{I}, \mathcal{O}_{\tilde{E}})|\mathrm{red}\,\tilde{E}_j \simeq \mathcal{O}_{\mathrm{red}\,\tilde{E}_j}$.

Generically $\mathcal{I}/\mathcal{I}^2$ has rank 1, so $\mathcal{H}om(\mathcal{I}/\mathcal{I}^2|\mathrm{red}\,\tilde{E}_j, \mathcal{O}_{\mathrm{red}\,\tilde{E}_j})$ is locally free of rank 1 ($\mathrm{red}\,\tilde{E}_j$ is smooth). Thus ϕ is injective.

Now it is an easy exercise to show that ϕ is also surjective, i.e. any homomorphism $(\mathcal{I}/\mathcal{I}^2|\mathrm{red}\,\tilde{E}_j)_x \to \mathcal{O}_x$ can be lifted locally. Namely, it is sufficient to lift homomorphisms $(\mathcal{I}/\mathcal{I}^2|\mathrm{red}\,\tilde{E}_j)/_{\mathrm{torsion}} \to \mathcal{O}$ locally. The left sheaf being a line bundle, this is clearly possible (for instance lift first to $\tilde{Y}$, then restrict to $\tilde{E}$).

Thus ϕ is an isomorphism, i.e.:

$$\mathcal{H}om(\mathcal{I}, \mathcal{O}_{\tilde{E}})|\mathrm{red}\,\tilde{E}_j \simeq \mathcal{O}_{\mathrm{red}\,\tilde{E}_j}. \tag{3}$$

Let $\tilde{E}_0$ be the union of all non-reduced $\tilde{E}_j$ with the induced structure. Then we obtain also:

$$\mathcal{H}om(\mathcal{I}, \mathcal{O}_{\tilde{E}})|\mathrm{red}\,\tilde{E}_0 \simeq \mathcal{H}om(\mathcal{I}/\mathcal{I}^2|\mathrm{red}\,\tilde{E}_0, \mathcal{O}_{\mathrm{red}\,\tilde{E}_0}) \simeq \mathcal{O}_{\mathrm{red}\,\tilde{E}_0}. \tag{4}$$

This proves the first part of a).

We consider the canonical homomorphism

$$\alpha: S^\mu(\mathcal{I}/\mathcal{I}^2) \to \mathcal{I}^\mu/\mathcal{I}^{\mu+1}.$$

α is an isomorphism on $\mathrm{supp}(\mathcal{I}^\mu/\mathcal{I}^{\mu+1})$ outside a finite set.

By (4), $S^\mu(\mathcal{I}/\mathcal{I}^2|\mathrm{red}\,\tilde{E}_0)/_{\mathrm{torsion}} \simeq \mathcal{O}_{\mathrm{red}\,\tilde{E}_0}$ and via α, for any component $\tilde{E}_j$ of $\tilde{E}_0$, $(\mathcal{I}^\mu/\mathcal{I}^{\mu+1}|\mathrm{red}\,\tilde{E}_j)/_{\mathrm{torsion}}$ contains the subsheaf $\mathcal{O}_{\mathrm{red}\,\tilde{E}_j}$ or 0. This proves the second part of a).

b) Assume that $\tilde{E}$ is non-reduced. Then from a) and $\mathrm{red}\,\tilde{E}_j \simeq \mathbb{P}_1$ for any j (use 3.6) we obtain:

$$h^0(\mathcal{I}^\mu/\mathcal{I}^{\mu+1}) > 0 \quad \text{as long as} \quad \mathcal{I}^\mu/\mathcal{I}^{\mu+1} \neq 0; \tag{5a}$$

$$h^1(\mathcal{I}^\mu/\mathcal{I}^{\mu+1}) = 0 \tag{5b}$$

for any μ.

Some explanation for (5b):

Denote by $\tilde{E}_0(\mu)$ the union of those $\tilde{E}_j$ for which $(\mathcal{I}^\mu/\mathcal{I}^{\mu+1}|\mathrm{red}\,\tilde{E}_j)/_{\mathrm{torsion}} \neq 0$. Then

$$\mathcal{O}_{\mathrm{red}\,\tilde{E}_0(\mu)} \hookrightarrow (\mathcal{I}^\mu/\mathcal{I}^{\mu+1}|\tilde{E}_0(\mu))|_{\mathrm{torsion}}$$

so by $h^1(\mathcal{O}_{\tilde{E}_0(\mu)}) = 0$ (since $h^1(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = 0$) we get our claim (5b).

Let $(\mathrm{red}\,\tilde{E})_\mu$ be the μ-th infinitesimal neighborhood of $\mathrm{red}\,\tilde{E}$ in $\tilde{E}$. Then by (5):

$$h^0(\mathcal{O}_{(\mathrm{red}\,\tilde{E})_\mu}) < h^0(\mathcal{O}_{(\mathrm{red}\,\tilde{E})_{\mu+1}}),$$

as long as $\mathcal{I}^\mu/\mathcal{I}^{\mu+1} \neq 0$, i.e. $(\mathrm{red}\,\tilde{E})_\mu \neq (\mathrm{red}\,\tilde{E})_{\mu+1}$. Since $h^0(\mathcal{O}_{\mathrm{red}\,\tilde{E}}) = 1 = h^0(\mathcal{O}_{\tilde{E}})$ by (3.5), we deduce $\mathrm{red}\,\tilde{E} = \tilde{E}$, a contradiction.

(3.8) **Proposition.** *$\tilde{E}$ consists of two smooth rational curves meeting in exactly one point of order two. Moreover $b_3(Y)=b_3(\tilde{Y})$.*

Proof. We have an exact sequence ([Ba-Ka, 3.A.7])

$$\begin{aligned} 0=H^1(Y,\mathbb{Z})\to H^1(\tilde{Y},\mathbb{Z})\oplus H^1(E,\mathbb{Z})\to H^1(\tilde{E},\mathbb{Z})\to H^2(Y,\mathbb{Z})\\ \to H^2(\tilde{Y},\mathbb{Z})\oplus H^2(E,\mathbb{Z}). \end{aligned}$$

From (3.5) we know $H^1(\tilde{Y},\mathbb{Z})=0$, (via exponential sequence) moreover $H^1(E,\mathbb{Z})=0$ by (3.2).

So $H^1(\tilde{E},\mathbb{Z})=0$.

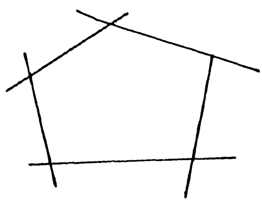
Hence (3.4.a) cannot appear and consequently ($\tilde{E}$ is connected) all irreducible components of $\tilde{E}$ are smooth rational.

Take a component $\tilde{E}_1$. Then we have the exact sequence

$$\begin{array}{l} 0\to\omega_{\tilde{E}_1}\to\omega_{\tilde{E}}\simeq\mathcal{O}_{\tilde{E}}\to\mathcal{H}om(\mathcal{I}_{\tilde{E}_1|\tilde{E}},\mathcal{O}_{\tilde{E}})\to 0 \quad (2.3).\\ \quad\ \ \wr|\\ \mathcal{O}_{\tilde{E}_1}(-2) \end{array}$$

This sequence immediately implies that either $\tilde{E}_1$ meets exactly two components transversely in a point or meets one component in two points transversely or meets one component in one point of order two.

In the first case $\tilde{E}$ must be a cycle:



But then $H^1(\tilde{E},\mathbb{Z})\simeq\mathbb{Z}$, contradiction.

So $\tilde{E}$ is not a cycle, hence clearly $\tilde{r}\leqq 2$ (since $\tilde{E}_1$ is arbitrary in the above considerations). The case $\tilde{r}=1$ is not possible since $\omega_{\tilde{E}}\simeq\mathcal{O}_{\tilde{E}}$. If $\tilde{r}=2$ and the two curves meet in two points (transversely), then $H^1(\tilde{E},\mathbb{Z})\neq 0$. So we are left with $\tilde{r}=2$ and two smooth rational curves meeting in exactly one point of order two. It remains to prove $b_3(Y)=b_3(\tilde{Y})$. To do this we use another part of the above exact sequence:

$$\begin{array}{ccccccccc} 0\to H^2(Y,\mathbb{Z}) & \to & H^2(\tilde{Y},\mathbb{Z})\oplus H^2(E,\mathbb{Z}) & \to & H^2(\tilde{E},\mathbb{Z}) & \to & H^3(Y,\mathbb{Z})\to H^3(\tilde{Y},\mathbb{Z})\to 0\\ \wr| & & \wr| & & \wr| & & \\ \mathbb{Z} & & \mathbb{Z}^r & & \mathbb{Z}^2 & & \end{array}$$

Since $r\leqq 2$, $b_2(\tilde{Y})>0$ and since $b_3(Y)$ and $b_3(\tilde{Y})$ are even ($b_3(Y)=b_3(X)$ and $b_3(\tilde{Y})$ $=b_3(\hat{Y})$!), we obtain $b_3(\tilde{Y})=b_3(Y)$.

In the following we let $\pi:\hat{Y}\to\tilde{Y}$ be a minimal desingularization of $\tilde{Y}$ and let $\sigma:\hat{Y}\to Y_m$ be a minimal model of $\hat{Y}$.

(3.9) **Proposition.** *If* $b_3(X)>0$, Y_m *is a non-rational ruled surface. If* $p: Y_m \to C_m$ *denotes the ruling,* $g(C_m)=\frac{b_3(X)}{2}$.

Proof. It is clear that $\kappa(Y_m)=-\infty$ i.e. Y_m is ruled or $\mathbb{P}_2$. So it suffices to prove $g(C_m) = \frac{b_3(X)}{2}$ in case Y_m is ruled and that $Y_m \neq \mathbb{P}_2$. Let Y_m be ruled. Then $b_3(\tilde{Y})=b_3(\hat{Y}) = b_3(Y_m)=2g(C_m)$. By (3.8), $b_3(\tilde{Y})=b_3(Y)$. Since $b_3(Y)=b_3(X)$, we obtain the equation we want. If $Y_m=\mathbb{P}_2$ the same arguments show $b_3(X)=0$, contradiction.

We first consider the case that X is of the principal series, i.e. the canonical divisor is very ample. We take over all notations of Sect. 2 concerning the genus of X etc. We will make heavily use of

(3.10) **Proposition.** *Let Z be the exceptional set of π. If* $b_3(X)>0$, $\dim p\circ\sigma(Z)=1$.

Proof. Assume $\dim p\circ\sigma(Z)=0$. Then obviously $R^1\pi_*(\mathcal{O}_{\tilde{Y}})=0$ (so $\tilde{Y}$ has only rational singularities). Moreover we know $H^1(\mathcal{O}_{\tilde{Y}})=0$ by (3.5), thus $H^1(\mathcal{O}_{\hat{Y}})=0$. Hence $\hat{Y}$ is rational and $b_3(Y)=b_3(\hat{Y})=0$ (3.8).

(3.11) **Proposition.** $g(X)\geqq 8$ *(i.e.* $g(X)\in\{8, 9, 10, 12\}$*).*

Proof. Remember that X is embedded in $\mathbb{P}_{g+1}$ by the canonical divisor. If we take two smooth hyperplane sections $H, H'\subset X$, the resulting smooth curve $C=H\cap H'$ has genus $g(C)=g(X)=g$. But here we want to consider $C:=Y\cap H$. Since $H\cap S(Y) \neq \emptyset$, C becomes singular, possibly reducible. C being connected (since $H^1(\mathcal{O}_Y(-1))=0$), we have $h^1(\mathcal{O}_C)=g(X)$ (since C is a degeneration of curves of the form $H\cap H'$). Let $C_1\subset C$ be an irreducible component. So $h^1(\mathcal{O}_{C_1})\leqq g(X)$; i.e. $g(C_1)$ (= genus of the normalization) $\leqq g(X)\leqq 7$. Let $\hat{C}_1$ be the strict transform of C_1 in $\hat{Y}$. If H is general, $\dim p\circ\sigma(\hat{C}_1)=1$. Let $\bar{C}_1\to\hat{C}_1$ be the normalization. Then we apply Riemann-Hurwitz to the map $\bar{C}\to C_m$ which has degree say α:

$$2g(\bar{C})-2=\alpha(2g(C_m)-2)+\deg R, \qquad (*)$$

R the ramification divisor.

Now $g(C_m)=\frac{b_3(X)}{2}$ by (3.9), hence by (2.2):

$$g(C_m)\geqq 5.$$

Since $g(\bar{C})\leqq 7$, we obtain from (*): $\alpha=1$ and $g(X)\geqq 5$.

Now for general H, C is irreducible and reduced. Then we obtain $(f\pi(F)\cdot H)=1$ (at least if $C\cap f(S(\tilde{Y}))=\emptyset$). So $f\pi(F)$ is a line in Y. Since $\pi \neq \mathrm{id}$ (otherwise $\dim p\circ\sigma(\hat{E})=1$ and Y would be rational!), by (3.10) all the lines of the form $f\pi(F)$ pass through one fixed point, namely the point $f(\pi(Z_i)$ where $Z_i\subset Z$ is a component with $p\circ\sigma(Z_i)=C_m$.

But every Fano 3-fold X with $g(X)\geqq 4$ (of the principal series) has the property that through any point there are only finitely many lines ([Is 1]), contradiction.

(3.12) **Proposition.** $g(X) \neq 8$.

Proof. Assume the existence of X. Then proceeding as in (3.11) and using the same notations as in (3.11), we obtain now from $g(C_m)=4$ and

$$2g(\bar{C})-2=\alpha(2g(C_m)-2)+\deg R: \qquad (*)$$

$\alpha \leqq 2$ and $\alpha=2$ iff $R=0$, $g(C)=7$.

The case $\alpha=1$ is excluded as in (3.11).

So $\alpha=2$ (which means that $f\pi(F)$ is a conic, hence Y is filled up by conics through a fixed point).

Since $R=0$, $\hat{C}$ is smooth, i.e. $\bar{C}=\hat{C}$. Since $\hat{C} \to C_m$ is unramified, $C'=\sigma(\hat{C})$ is smooth. Let C_0 a section of Y_m with minimal self-intersection; G a fiber of p. Define by $C_0^2=-e$ (cp. [Ha, Chap. V, Sect. 2]). Write for numerical equivalence:

$$C' \sim 2C_0+\beta G.$$

Then the adjunction formula gives:

$$12=2g(C')-2=(-2C_0+(6-e)G \cdot 2C_0+\beta G)+C^2=2\beta-2e+12.$$

Hence $\beta=e$.

On the other hand, for general C, $\hat{C}$ is an ample divisor on $\hat{Y}$, hence $\hat{C}^2>0$. So $C'^2 \geqq \hat{C}^2>0$. But $C'^2=4\beta-4e=0$, contradiction.

(3.13) **Proposition.** $g(X) \neq 9$.

Proof. The proof being similar to (3.14) treating the case $g(X)=10$ (and in fact easier) we will omitt it.

(3.14) **Proposition.** $g(X) \neq 10$.

Proof. Assume $g(X)=10$. Then we will make use of the following construction due to Iskovskij ([Is 1]). Take a sufficiently general line $Z \subset X$. Then there are exactly four lines $Z_1, \ldots, Z_4$ meeting Z. Let $\tau_1: X_1 \to X$ be the blow-up of Z in X. Let $\tau_2: X_2 \to X_1$ be the blow-up of the strict transforms $Z_i^{(1)}$ in X_1. Let $Z^{(2)}$ be the strict transform of $\tau_1^{-1}(Z)$ in X_2, let $Z_i^{(2)}$ be the proper transforms of the $Z_i^{(1)}$.

Let $\mathscr{L}:=\tau_2 * \tau_1 * (\mathcal{O}_X(1)) \otimes \mathcal{O}(-2Z^{(2)}) \otimes \mathcal{O}(-\Sigma Z_i^{(2)})$.

Then $\mathscr{L}$ is globally generated and $h^0(X_2, \mathscr{L})=5$. Let $\phi: X_2 \to \mathbb{P}_4$ be the associated morphism. Then $\phi(X_2)$ is a smooth 3-dimensional quadric Q_3.

Moreover ϕ is birational and contracts exactly S_2 and $Z_i^{(2)}$, where S_2 is the strict transform of the surface $S \subset X$ swept out by all conics in X meeting Z. So far Iskovskij's construction.

Now denote by Y_2 the strict transform of Y in X_2 and let $\phi(Y_2)=Y_0 \subset Q_3$. Since Z is general, Z is not contained in Y. Namely, otherwise Y would be filled up lines. So the strict transforms of the lines in $\hat{Y}$ would have to be contracted by $p \circ \sigma$ (since $g(C_m)=2$ in our case!). But then all the lines would have to pass through one and the same point (because of π!) which is not possible by [Is 1]. So $Z \not\subset Y$. Since $(Z \cdot Y=1)$, we conclude $Z \cap S(Y)=\emptyset$, in particular $Z \cap E=\emptyset$. Hence for any $i: Z_i \not\subset E$. From this we deduce at once: $E \not\subset S$ (otherwise E would be a line or a conic meeting Z).

Going into the construction of Iskovskij we see that $\tau_2 \circ \tau_1 | Y_2 \to Y$ and $\phi | Y_2 \to Y_0$ are birational, moreover the set of indeterminacy of $\phi \circ (\tau_2 \circ \tau_1)^{-1}$ does not contain E. Hence Y_0 is non-normal. Now an easy calculation shows that

$$\deg Y_0 = 6 \text{ (in } \mathbb{P}_4).$$

So Y_0 is the intersection of a quadric (Q_3) and a cubic in $\mathbb{P}_4$. Taking the general quadric and the general cubic and looking at its smooth intersection Y_t, the general smooth hyperplane section C_t of Y_t will have degree 6, hence $g(C_t)=4$ (by adjunction formula).

By degeneration we conclude for the general hyperplane section C_0 of Y_0 (C_0 being singular): $g(C_0) \leqq 3$. Let

$$f_0 : \tilde{Y}_0 \to Y_0$$

be the normalization,

$$\pi_0 : \hat{Y}^0 \to Y_0$$

a minimal desingularization. Let

$$\sigma_0 : \hat{Y}_0 \to Y_{0,m}$$

be a minimal model.

Then $Y_{0,m}$ is a ruled surface over a curve, $C_{0,m}$ of genus 2 (since $g(C_m)=2$), denote by p_0 the projection. Let C_0 be the strict transform of C_0 in $\hat{Y}_0$, and $\bar{C}_0$ its normalization. Apply Riemann-Hurwitz to $\bar{C}_0 \to C_{0,m}$ to obtain:

$$2g(C_0) - 2 = 2\alpha_0 + \deg R_0,$$

R_0 the ramification divisor, α_0 the degree of $\bar{C}_0 \to C_{0,m}$. Now $g(C_0) \leqq 3$, hence either

a) $g(C_0)=3$, $\alpha_0=1$, $\deg R_0=2$
b) $g(C_0)=3$, $\alpha_0=2$, $R_0=0$
c) $g(C_0)=2$, $\alpha_0=1$, $R_0=0$.

a) cannot occur: because of $\alpha_0=1$, $\sigma_0(\hat{C}_0)$ would have to be a section of $Y_{0,m}$, hence smooth. So $\hat{C}_0$ would be smooth, i.e. $\hat{C}_0 = \bar{C}_0$ and $\hat{C}_0 \to \sigma_0(\hat{C}_0)$ would be isomorphic. Hence $R_0=0$. Now assume b). Then we proceed as in (3.12): compute $\sigma_0(C_0)$ in $Y_{0,m}$ for numerical equivalence and conclude $\sigma_0(C_0)^2=0$, which is impossible (argue as in (3.12)).

So we are left with case c). So Y_0 is filled up by lines. Let α be the degree of the images of the fibers $(p \circ \sigma)^{-1}(x)$ in Y. Since Y and Y_0 are non-rational, we deduce that the images of the curves $(p \circ \sigma)^{-1}(x)$ under our birational map $Y \to Y_0$ are just the lines in Y_0. But then we have $\alpha = 1$! Namely, if Z is general, we can achieve $Z_i \not\subset Y$ for all i (since by Iskowskij any line in X meets only finitely many other lines). But then – letting $l = f\pi(p \circ \sigma)^{-1}(x)$ – we conclude

$$\alpha = (c_1(\mathcal{O}_X(1) \cdot l) = (c_1(\mathcal{L}) \cdot l_2) = (c_1(\mathcal{O}_{Q_3}(1)) \cdot \phi(l_2)) = 1$$

for general l (l_2 is the strict transform in X_2).

Conclusion: Y is filled up by lines which have all to pass through a fixed point (since $\pi \neq \mathrm{id}$ as before). This being impossible by Iskovskij the proof is finished.

(3.14) **Conclusion.** We have now proved: If X is of the principal series, then $g(X) \geqq 11$. Since by [Is 1] $g(X) \neq 11$ and $g(X) \leqq 12$, we obtain $g(X)=12$. So it remains to exclude the cases where X is not of the principal series. These are the following ([Is 1])

a) $g(X)=2$ and the anti-canonical map $\phi_{K^{-1}}: X \to \mathbb{P}_3$ is $2:1$ and ramified in a sextic

b) $g(X)=3$ and $\phi_{K^{-1}}: X \to Q_2$ (= smooth quadric in $\mathbb{P}_4$) is $2:1$ and ramified in a surface of degree 8.

(3.15) **Proposition.** *The case* $2 \leqq g(X) \leqq 3$ *and* X *not of the principal series does not occur.*

Proof. We use in principal the same method as in (3.11). We have $g(C_m)=52$ (resp. 30) if $g(X)=2$ (resp. 3). Take $s \in H^0(\mathcal{O}_Y(1))$ general. Then $C=\{s=0\}$ is irreducible and reduced. Moreover $g(C) \leqq 2$ (resp. 3); even $g(C) \leqq 1$ (resp. 2) since C is singular. Considering the map $\hat{C} \to C_m$ we obtain a contradiction to $g(C_m)=52$ (resp. 30).

So theorem (3.1) is proved completely.

We cannot decide here whether a compactification X with $g(X)=12$ (and non-normal Y) exists. But we know something on the structure of X if it exists:

(3.16) **Theorem.** *Assume that* X *is a compactification of* $\mathbb{C}^3$ *with non-normal* Y *such that* X *is a Fano-*3 *fold of the principal series, of index* 1, *with* $g(X)=12$.

Then E *consists either of one smooth rational curve or of two smooth rational curves meeting transversely in one point.* $\tilde{E}$ *consists of two smooth rational curves meeting in one point of order* 2. *Moreover* E *and* $\tilde{E}$ *are reduced. Here we use the notations of* (3.2).

Proof. (3.7), (3.8). The reducedness of E follows from that one of $\tilde{E}$.

(3.17) *Remark.* In the situation of (3.16) one can say move on the singularities of Y and $\tilde{Y}$. Namely, by [K-W] or [S], for general $y \in E$ we have either $\hat{\mathcal{O}}_{Y,y} \simeq \mathbb{C}[\![X, Y]\!]/(X \cdot Y)$ or $\hat{\mathcal{O}}_{Y,y} \simeq \mathbb{C}[\![X, Y]\!]/(X^2+Y^3)$.

Here $\hat{\mathcal{O}}_{Y,y}$ denotes completion of $\mathcal{O}_{Y,y}$.

The first case occurs exactly when E is irreducible, the second when E consists of two components (then f is a homeomorphism).

Moreover the only possible singularity of $\tilde{Y}$ on $\tilde{E}$ is the point where the two components of $\tilde{E}$ intersect ([K-W]). Observe that by [S] Y is weakly normal (sometimes called maximal, cf. [F]). Let us remark that one can show that $Y \backslash E$ is smooth (a priori it could have rational double points), so $\tilde{Y}$ has at most one singularity which must be rational.

(3.18) *Remark.* If Y is assumed normal in (3.12) or (3.13) we can carry out the same construction as in the proof of (3.13) and conclude – with some minor changes in the proof – the non-existence of the compactification (X, Y). This finishes the proof of part I, Theorem 3.5, as promised.

4. A Remark on Compactifications with Index 2

(4.1) This section is joint work with Schneider and gives a supplement to [PS]. We are indebted to Furushima and N. Nakayama for very fruitful discussions.

In [PS] we proved (Theorem 2.4) that two compactifications (X, Y), (X', Y') of $\mathbb{C}^3$ with $b_2(X) = b_2(X') = 1$, where X, X' are Fano 3-folds of index 2 and Y, Y' are either both normal or both non-normal are isomorphic. This means precisely the following: there is a biholomorphic map $\phi: X \to X'$ such that $\phi(Y) = Y'$. As promised in [PS] we present here some details which were omitted in [PS].

Note that it is already clear that X and X' are abstractly isomorphic, namely the Fano 3-fold V_5 of Iskovskij (cf. [Fu 1], [PS]). Moreover Y and Y' are abstractly isomorphic and the structure is well-known (see [PS], Theorem 2.4).

(4.2) Iskovskij constructed a birational morphism from the Fano 3-fold X of type V_5 to a 3-dimensional smooth quadric Q_3. This construction has been modified by Furushima [Fu 1] in the following way.

Take points $p, p_0 \in l$, a line in $Q_3 \subset \mathbb{P}_4$. Take tangent hyperplane sections H, H_0 to p, p_0. Let C be a twisted cubic contained in H_0. Necessarily $p_0 \in C$. Let $\pi: X' \to Q_3$ be the blow-up of C in Q_3. Let $\hat{H}_0$ be the strict transform of H_0 in X'. Then $\hat{H}_0 \simeq \Sigma_2 = \mathbb{P}(\mathcal{O} \oplus \mathcal{O}(-2))$ over $\mathbb{P}_1$ and X' can be blow down along the projection $\hat{H}_0 \to \mathbb{P}_1$. We obtain a modification $X' \to X$ and thus a birational map $X \to Q_3$. σ is the just the blowup of a line $l_0 \subset X$ with $N_{l_0|X} \simeq \mathcal{O}(-1) \oplus \mathcal{O}(1)$. If we set $Y = \sigma\pi^{-1}(C)$, then Y is a non-normal hypersurface in X with non-normal locus l_0 and $X \setminus Y \simeq \mathbb{C}^3$. Moreover *all* compactifications (X, Y) (with X of type V_5) arise in this way. Remark that the strict transform of Y in X' is just Σ_3 and that π contracts exactly the strict transforms of the lines in Y (Y can be described as the surface of lines meeting l_0). The last facts follow from [PS].

Now consider the strict transform A' of H in Y'. Let $A := \sigma(A')$. Then (X, A) is a compactification of $\mathbb{C}^3$ with A normal and all "normal" compactifications arise in this manner [Fu 1].

(4.3) Let (X, Y), (X', Y') be two smooth compactifications of $\mathbb{C}^3$ such X, X' is of type V_5. Assume either both Y and Y' are normal or non-normal. Proving the existence of a biholomorphic map $\phi: X \to X'$ such that $\phi(Y) = Y'$ comes down (by (4.2)) to prove the following.

(4.4) **Theorem.** *Let $(Q, \tilde{Q}, C, l, q)$ be a quintuple consisting of a smooth 3-dimensional quadric $Q \subset \mathbb{P}_4$, a twisted cubic curve $C \subset Q$, the uniquely determined quadric cone $\tilde{Q} \subset Q$ containing C, the uniquely determined line $l \subset Q$ such that $l \cap C$ is the vertex of $\tilde{Q}$ and a point $q \in l$. Let $(Q', \tilde{Q}', C', l', q')$ be another quintuple of this type. Then there exists a biholomorphic map $\phi: Q \to Q'$ such that $\phi(\tilde{Q}) = \tilde{Q}'$, $\phi(C) = C'$, $\phi(l) = l'$, $\phi(q) = q'$.*

Proof. The proof is given in several steps which are well-known and whose proofs are very easy (thus omitted).

1. We may assume $Q = Q'$ and $\tilde{Q} = \tilde{Q}'$ (since there is $\psi: Q \to Q'$ biholomorphic such that $\psi(\tilde{Q}) = \tilde{Q}'$).
2. For any quadric cone $\tilde{Q} \subset \mathbb{P}_3$ and $x \in \tilde{Q}$, $x' \in \tilde{Q}$ there is $\psi \in \mathrm{Aut}(\mathbb{P}_3)$ such that $\psi(x) = x'$, $\psi(C) = C'$.
3. If $C \subset \mathbb{P}_3$ is a twisted cubic, $x \in C$, then there is a uniquely determined quadric cone $\tilde{Q} \subset \mathbb{P}_3$ such that $C \subset \tilde{Q}$ and x is the vertex of $\tilde{Q}$.
4. Put $x :=$ vertex of $\tilde{Q}$ in our situation.

By 2) we find $\psi \in \mathrm{Aut}(\mathbb{P}_3)$ such that $\psi(C)=C'$ and $\psi(x)=x$. By 3) we conclude $\psi(Q)=\tilde{Q}$ since the vertex of $\psi(\tilde{Q})$ is x. Hence we have $\phi \in \mathrm{Aut}(\tilde{Q})$ such that $\phi(C)=C'$. Then automatically $\phi(l)=l'$!

5. Now lift ϕ to an automorphism $\tilde{\phi} \in \mathrm{Aut}(Q)$. This is possible since the restriction map

$$\mathrm{Aut}_{\tilde{Q}}(Q) \to \mathrm{Aut}(\tilde{Q})$$

(from the group of automorphisms of Q fixing $\tilde{Q}$ to $\mathrm{Aut}(\tilde{Q})$) is an isomorphism. In fact, it is sufficient to see $\dim \mathrm{Aut}_{\tilde{Q}}(Q) = \dim \mathrm{Aut}(\tilde{Q}) = 7$ and injectivity of the restriction map.

6. Still we have to see that we can achieve $\phi(q)=q'$. To do this we just mention that any $\psi \in \mathrm{Aut}(C)$ with $\psi(p)=p$ can be liftet to $\tilde{\psi} \in \mathrm{Aut}(\tilde{Q})$ with $\psi(C)=C$, hence to $\tilde{\tilde{\psi}} \in \mathrm{Aut}(Q)$.

Thus the group of automorphisms ϕ constructed in 5) acts transitively on C, q.e.d.

References

[A-K] Altman, A., Kleiman, S.: Introduction to Grothendieck duality theory. (Lecture Notes in Math. Vol. 146). Berlin Heidelberg New York: Springer 1970

[Ba-Ka] Barthel, G., Kaup, L.: Topologie des espaces complexes compactes singulières. Montreal Lectures Notes Vol. 80 (1982)

[F] Fischer, G.: Complex analytic geometry. (Lecture Notes in Math., Vol. 538). Berlin Heidelberg New York: Springer 1976

[Fu 1] Furushima, M.: Singular del Pezzo surfaces and analytic compactifications of $\mathbb{C}^3$. Nagoya Math. J. **104**, 1–28 (1986)

[Fu 2] Furushima, M.: On complex analytic compactifications of $\mathbb{C}^3$. Preprint des MPI Bonn (1987)

[Fu 3] Furushima, M.: On complex analytic compactifications of $\mathbb{C}^3$ (II). Preprint des MPI Bonn (1987)

[Is 1] Iskovskij, V.A.: Fano 3-folds II. Math. USSR Isv. **11**(3), 469–506 (1977)

[Is 2] Iskovskij, V.A.: Algebraic threefolds with special regard to the problem of rationality. Proc. of the international congress of Math. 1983, 733–747 (1986)

[Is-Šo] Iskovskij, V.A., Šokurov, V.V.: Biregular theory of Fano 3-folds. (Lecture Notes in Math., Vol. 732, 171–182). Berlin Heidelberg New York: Springer 1978

[K-W] Kunz, E., Waldi, R.: Der Führer einer Gorensteinvarietät. J. f. d. reine u. angew. Math. **388**, 106–115 (1988)

[Mo] Mori, S.: Threefolds whose canonical bundles are not numerically effective. Ann. Math. **116**, 133–176 (1982)

[P-S] Peternell, Th., Schneider, M.: Compactifications of $\mathbb{C}^3$. I. Math. Ann. **280**, 129–146 (1988)

[S] van Straten, D.: Weakly normal surface singularities and their improvements. Thesis, Leiden (1987)

Received May 27, 1988

Note added in proof. Recently M. Furushima proved that there exists a compactification of $\mathbb{C}^3$ with non-normal boundary at infinity which is a Fano 3-fold of index 1 of "Mukai-Umemura" type.

Math. Ann. 283, 139–149 (1989)

Mathematische Annalen

Über Vorzeichenwechsel einiger arithmetischer Funktionen. I

Bogdan Szydło

Mathematisches Institut der Adam-Mickiewicz-Universität, ul. Matejki 48/49,
PL-60-769 Poznań, Poland

1. Einleitung

Viele interessante, mit der Oszillation des Restgliedes $\Delta_1(x)=\pi(x)-\mathrm{li}(x)$ im Primzahlsatz

$$\lim_{x\to\infty} \pi(x)/\mathrm{li}(x)=1$$

verknüpfte Untersuchungen haben ihren Ursprung in der bahnbrechenden Arbeit von Riemann „Ueber die Anzahl der Primzahlen unter einer gegebenen Größe" [6, S. 145–153]. Durch einige Riemannsche Bemerkungen angeregt, vermutete man beispielsweise, daß

$$\pi(x)<\mathrm{li}(x)$$

stets ist. Im Jahr 1914 zeigte aber Littlewood [5], daß Δ_1 sogar unendlich viele Vorzeichenwechsel hat, und daher erwies sich die oben erwähnte Vermutung als falsch. Die Aufgabe, solch einen Vorzeichenwechsel anzugeben, ist leider bisher ungelöst, vgl. [4].

Die Untersuchung der Größenordnung der Funktion $V_1(X)$, die die Anzahl der Vorzeichenwechsel von Δ_1 im Intervall $(0, X]$ $(X\geqq 0)$ angibt, ist ein für die analytische Zahlentheorie wesentliches Problem. Da man einen Abriß der Geschichte der betreffenden Forschungen z.B. in [2, 3] finden kann, beschränken wir uns hier darauf, die folgende kurze „Zeittafel" der Mathematiker, die bisher die bedeutendsten Ergebnisse erhalten haben, auszuschreiben:

Riemann, Littlewood, Pólya, Ingham, Turán, Knapowski, Pintz, Kaczorowski.

Das beste Resultat, hier in etwas schwächerer Form geschrieben, gehört Kaczorowski [2]:

$$\liminf_{X\to\infty} \frac{V_1(X)}{\log X}>0. \tag{1.1}$$

Zur Grundlage seiner Methode macht er die Anwendung des Operators δ, dessen Einwirkung auf eine Funktion $f:(0,\infty)\to\mathbb{R}$ durch

$$\delta(f)(x)=\int_0^x f(\xi)\frac{d\xi}{\xi} \quad (x>0) \tag{1.2}$$

definiert wird, falls das Integral in der rechten Seite von (1.2) existiert. Genügt nun eine Funktion f gewissen Bedingungen und bezeichnet F ihre Mellinsche Transformation (s. §2), so gilt

$$\delta(f)(x)=\frac{1}{2\pi i}\int_{\sigma_1-i\infty}^{\sigma_1+i\infty} F(s)\frac{x^s}{s}\,ds \qquad (x>0). \tag{1.3}$$

Bezeichnet ferner $V(f,X)$ die Anzahl der Vorzeichenwechsel von f im Intervall $(0,X]$ $(X>0)$, so gibt die Anwendung des Lemmas (§3)

$$V(f,X)\geqq V(\delta(f),X) \qquad (X>0).$$

Statt sich direkt mit f zu beschäftigen, kann man also die gebildete Funktion $\delta(f)$ betrachten, deren oszillatorischer Charakter, durch die gegenseitige Stellung der Pole der Transformation F determiniert, leichter untersucht wird, denn im Integral in der rechten Seite von (1.3) kommt der zusätzliche, die Konvergenz verstärkende Faktor $1/s$ vor. Die geeignete Iteration des Operators δ ist auch eine wesentliche Etappe der Methode von Kaczorowski, die hier nur skizziert wird; vgl. [2, 3].

In vielen für die Theorie der Primzahlverteilung wichtigen konkreten Fällen führt die Anwendung des oben geschilderten Verfahrens zu Ergebnissen vom Typus

$$V(f,X)\geqq c\log X \qquad (X\geqq X_0), \tag{1.4}$$

wo $c>0$ und $X_0\geqq 2$ effektive Konstanten sind (im Gegensatz zu denjenigen, von denen man nur die Existenz aussagt).

Der erste Artikel dieser Reihe präsentiert eine Modifikation der Methode von Kaczorowski. Wir wenden nämlich statt des Operators δ einen allgemeineren $\mathbf{A}$ an:

$$\mathbf{A}(f)(x)=x^{-d-k}\int_0^x f(\xi)\xi^{d+k-1}\,d\xi \qquad (x>0), \tag{1.5}$$

wo d und k Parameter sind, vgl. (1.2).

Die geeignete Einführung dieser Parameter ist nämlich die Hauptneuerung (aber auch die Hauptschwierigkeit), die im Vergleich zu der originalen Ausgangsgestalt der Methode zur Verschärfung (in Hinsicht auf $c>0$) effektiver Resultate vom Typus (1.4) führt. Grob gesagt, geschieht das, weil unsere Modifikation gegebene Information über die Mellinsche Transformation F von f in wirksamerer Weise benutzt.

Das Hauptresultat des Artikels lautet (s. §2 für die Bezeichnungen):

Satz. *Es sei $f:(0,\infty)\to\mathbb{R}$ eine stückweise stetige, in keinem Intervall $(a,b)\subset(0,\infty)$ konstante Funktion. Ferner sei $F=\mathfrak{M}(f)$ ihre Mellinsche Transformation, wobei es ein $\varepsilon>0$ derart gibt, daß*

$$F(s)=\int_0^\infty f(x)x^{-s-1}\,dx$$

für $\sigma>\sigma_1-\varepsilon$ absolut konvergiert. F sei eine in $D\backslash\{\varrho_1,\bar\varrho_1,\ldots,\varrho_r,\bar\varrho_r\}$ holomorphe Funktion, wo

$$D=\{s\in\mathbb{C}:\sigma_0\leqq\sigma\leqq\sigma_1,|t|\leqq h\}\cup\{s\in\mathbb{C}:\sigma\geqq\sigma_1\}:$$

$$\varrho_\nu=\beta_\nu+i\gamma_\nu,\sigma_0<\beta_\nu<\sigma_1,0<\gamma_\nu<h(\nu=1,\ldots,r);$$

$$\varrho_\nu\neq\varrho_{\nu'}(\nu\neq\nu')$$

ist. F habe im Punkt ϱ_ν *einen Pol der Ordnung* $m_\nu \geqq 1$ *mit dem Hauptteil*

$$\sum_{l=-m_\nu}^{-1} a_{\nu,l}(s-\varrho_\nu)^l \qquad (\nu=1,\ldots,r).$$

Es sei

$$|F(s)| \leqq M < +\infty \qquad (s \in \partial D).$$

Es gebe endlich wenigstens einen Pol $\varrho_{\nu_0} = \varrho(=\beta+i\gamma)$ $(\nu_0 \in \{1,\ldots,r\})$ *derart, daß die Bedingung*

$$h^2 > \frac{\sigma_1-\sigma_0}{\beta-\sigma_0}\gamma^2 + (\sigma_1-\beta)(\sigma_1-\sigma_0) \tag{1.6}$$

erfüllt ist. Dann gibt es eine von σ_0, σ_1, h, $\beta_1,\ldots,\beta_r$, $\gamma_1,\ldots,\gamma_r$ *effektiv abhängige Konstante* $c>0$ *und eine von* σ_0, σ_1, h, $\beta_1,\ldots,\gamma_r$, M, $a_{1,-1},\ldots,a_{1,-m_1},\ldots,a_{r,-1},\ldots,a_{r,-m_r}$ *effektiv abhängige Konstante* $X_0 \geqq 2$ *derart, daß*

$$V(f,X) \geqq c \log X$$

für $X \geqq X_0$ *ist.*

Man kann den Satz verallgemeinern, indem man die Existenz reeller Pole der Funktion $F=\mathfrak{M}(f)$ zuläßt. Hier verzichten wir aber darauf, einen geeigneten Satz zu formulieren.

Es ist auch bemerkenswert, daß *die gegenseitige Stellung von nur einem Pol der Funktion* $F=\mathfrak{M}(f)$ *und dem Rechteck* $P=\{s\in\mathbb{C}: \sigma_0<\sigma<\sigma_1, 0<t<h\}$ *ein Ergebnis vom Typus* (1.4) *mit* $c>0$ *determiniert, ohne „das Rechteck P zu überschreiten"*.

Notieren wir die folgende unmittelbare Folgerung aus dem Satz.

Folgerung. *Es sei* $f:(0,\infty)\to\mathbb{R}$ *eine stückweise stetige, in keinem Intervall* $(a,b)\subset(0,\infty)$ *konstante Funktion. Ferner sei* $F=\mathfrak{M}(f)$ *ihre Mellinsche Transformation, die für* $\sigma>\sigma_1^*$ *durch das absolut konvergente Integral*

$$F(s) = \int_0^\infty f(x)x^{-s-1}\,dx$$

definiert ist. F sei in der Halbebene $\sigma>\sigma_0^*$ *meromorph, wo* $-\infty \leqq \sigma_0^* < \sigma_1^*$ *ist. Es gelte*

$$\sup\{\beta^*\in\mathbb{R}: \beta^*>\sigma_0^*, \beta^* \text{ ist ein Pol von } F\} < \sup\{\beta\in\mathbb{R}: \beta>\sigma_0^*, \beta+i\gamma$$
$$\text{ist ein Pol von } F \text{ für gewisses } \gamma\neq 0\}.$$

Dann ist

$$\liminf_{X\to\infty} \frac{V(f,X)}{\log X} > 0.$$

Der Satz verallgemeinert in der bestimmten Richtung ein Resultat von Kaczorowski and Pintz [3, Theorem 2]; die Folgerung ist in der Tat nicht neu, vgl. [3, Theorem 1].

Die beim Beweisen des Satzes genauer dargestellte Methode kann einen methodologischen Ausgangspunkt für mögliche Anwendungen bilden. Einige solche werden schon in Artikel II [7] gezeigt werden, wo ein wichtiger Fall betrachtet werden wird: $f(x)=\Delta(x)=\psi(x)-x (x>0)$, wobei ψ die Tschebyschevsche Funktion bezeichnet. Ohne hier in Einzelheiten einzugehen, können wir sagen, daß in [7] u.a. das geleistet werden wird, was schon der Satz in diesem Spezialfall $f=\Delta$ antizipiert.

Erwähnen wir noch, daß bei der Aufgabe, ein effektives Resultat vom Typus (1.4) für $V(\Delta_1, X)=V_1(X)$ zu geben, die Methode versagt; vgl. jedoch das ineffektive Ergebnis (1.1) [2].

2. Bezeichnungen

$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathbb{R}$	Körper der reellen Zahlen
$\mathbb{C}$	Körper der komplexen Zahlen
$s=\sigma+it$	komplexe Zahl, für die analytische Zahlentheorie kanonisch geschrieben ($\sigma, t\in\mathbb{R}$, $i^2=-1$)
$\mathbb{H}$	$=\{s\in\mathbb{C}: t\geqq 0\}$
∂D	topologischer Rand eines Gebietes $D\subset\mathbb{C}$ oder entsprechende orientierte Kurve
$\mathfrak{M}(f)$	Mellinsche Transformation einer Funktion $f:(0,\infty)\to\mathbb{R}$, die im Punkt $s\in\mathbb{C}$ durch $\int_0^\infty f(x)x^{-s-1}dx$ definiert wird (falls das Integral absolut konvergiert); s. z.B. [1, S. 87ff.] (im Artikel bedeutet $\mathfrak{M}(f)$ eine entsprechende analytische Fortsetzung)
$O(\cdot)$	Landausches Symbol
$\ll$	Winogradovsches Symbol
$\asymp$	$\ll$ und $\gg$ gleichzeitig
$[x]$	größte ganze rationale Zahl, die $\leqq x$ ist ($x\in\mathbb{R}$)
$V(f,X)$	s. §1, unmittelbar nach (1.3)
$\square$	Ende eines Beweises.

3. Lemma

Eine elementare Observation, die eine Grundlage der Kaczorowskischen Methode der Untersuchung der Anzahl der Vorzeichenwechsel arithmetischer Funktionen bildet, lautet:

Lemma [2, I, Lemma 1]. *Es sei $f:(0,\infty)\to\mathbb{R}$ eine stückweise stetige, in keinem Intervall $(a,b)\subset(0,\infty)$ konstante Funktion, für die das Integral $\int_0^a |f(\xi)|d\xi$ für alle $a>0$ endlich ist. Ferner sei $f_1:(0,\infty)\to\mathbb{R}$ durch*

$$f_1(x)=\int_0^x f(\xi)d\xi \qquad (x>0)$$

definiert. Dann gilt für alle $X>0$

$$V(f,X)\geqq V(f_1,X).$$

4. Beweis des Satzes

Fall $r=1$, $m_1=1$. Es genüge f den Voraussetzungen des Satzes. Für $\eta \in \mathbb{R}$ definieren wir die Funktion $A=\mathbf{A}(f)$:

$$A(x)=x^{-\eta}\int_0^x f(\xi)\xi^{\eta-1}d\xi \qquad (x>0).$$

Mit Hilfe des bekannten Fubinischen Satzes stellt man leicht fest, daß

$$\mathfrak{M}(A)(s)=\int_0^\infty A(x)x^{-s-1}dx=\frac{\mathfrak{M}(f)(s)}{s+\eta}=\frac{F(s)}{s+\eta}(\sigma>\sigma_1-\varepsilon, \eta\geqq-\sigma_1+\varepsilon)$$

ist. Führen wir die Bezeichnungen $A_1=A$, $A_n=\mathbf{A}(A_{n-1})$ $(n\geqq 2)$ ein. Induktiv zeigt man, daß

$$\mathfrak{M}(A_n)(s)=\frac{F(s)}{(s+\eta)^n}(n\in\mathbb{N}, \sigma>\sigma_1-\varepsilon, \eta\geqq-\sigma_1+\varepsilon)$$

ist. Bemerken wir nun, daß die Funktionen $A_n(n\geqq 1)$ den Voraussetzungen der Mellinschen Umkehrformel genügen, s. z.B. [1, S. 88f.].

Wir schreiben $n+2 (n\in\mathbb{N})$ statt n und setzen $\eta=d+k$ mit $d, k\in\mathbb{R}$. Dabei sind n, d, k Parameter, die später passend gewählt werden.

Wir nehmen

$$d+k>-\sigma_0 \tag{4.1}$$

an. Dann gilt

$$A_{n+2}(x)=\frac{1}{2\pi i}\int_{\sigma_1-i\infty}^{\sigma_1+i\infty}\frac{F(s)x^s ds}{(s+d+k)^{n+2}}, \tag{4.2}$$

und aus dem Residuensatz folgt

$$A_{n+2}(x)=2\,\mathrm{Re}\frac{a_{-1}x^\varrho}{(\varrho+d+k)^{n+2}}+\frac{1}{2\pi i}\int_{\partial D}\frac{F(s)x^s ds}{(s+d+k)^{n+2}}, \tag{4.3}$$

wo $\varrho=\varrho_1$ ist.

Schreiben wir

$$a_{-1}=|a_{-1}|\exp(i\alpha_1), \varrho+d+k=|\varrho+d+k|\exp(i\alpha_2),$$

wo α_1, $\alpha_2\in\mathbb{R}$. Dann ist

$$2\,\mathrm{Re}\frac{a_{-1}x^\varrho}{(\varrho+d+k)^{n+2}}=2\frac{|a_{-1}|x^\beta}{|\varrho+d+k|^{n+2}}\cos(\gamma\log x+\alpha_1-\alpha_2(n+2)). \tag{4.4}$$

Grob gesagt, werden wenigstens manche Vorzeichenwechsel von f von diesem Term stammen, vgl. das Lemma. Dazu wollte man „das Restglied“ in (4.3) klein genug machen. Weiter wird diese Idee präzisiert und vervollständigt werden.

In den unten angegebenen Abschätzungen werden die in den Symbolen $O(\cdot)$ und $\ll$ implizierten Konstanten effektiv von den entsprechenden Parametern abhängen.

Führen wir die Bezeichnungen ein:

$$\begin{aligned} L_1 &= \{s \in \mathbb{C} : \sigma = \sigma_0, 0 \leqq t \leqq h\}. \\ L_2 &= \{s \in \mathbb{C} : \sigma = \sigma_1, t \geqq h\}, \\ L_3 &= \{s \in \mathbb{C} : \sigma_0 \leqq \sigma \leqq \sigma_1, t = h\} \end{aligned} \tag{4.5}$$

und schätzen (unter geeigneten Voraussetzungen) die entsprechenden Teile vom Integral in der rechten Seite von (4.3) ab:

$$\left|\int_{L_1}\right| \leqq \frac{Mhx^{\sigma_0}}{|\sigma_0+d+k|^{n+2}} \ll \frac{x^{\sigma_0}}{(\sigma_0+d+k)^n},$$

$$\left|\int_{L_2}\right| \leqq \frac{Mx^{\sigma_1}}{|\sigma_1+ih+d+k|^n} \int_h^\infty \frac{dt}{t^2} \ll \frac{x^{\sigma_1}}{|\sigma_1+ih+d+k|^n},$$

$$\left|\int_{L_3}\right| \leqq \frac{M}{|\sigma_0+ih+d+k|^2} \int_{\sigma_0}^{\sigma_1} \frac{x^\sigma d\sigma}{|\sigma+ih+d+k|^n} \ll \int_{\sigma_0}^{\sigma_1} \frac{x^\sigma d\sigma}{|\sigma+ih+d+k|^n}.$$

Die anderen, „symmetrisch gelegenen" Teile des Integrals schätzen wir in derselben Weise ab. Hieraus erhalten wir

$$\left(\frac{x^\beta}{|\varrho+d+k|^n}\right)^{-1} \left|\int_{\partial D} \frac{F(s)x^s ds}{(s+d+k)^{n+2}}\right| \ll x^{\sigma_0-\beta}\left|\frac{\varrho+d+k}{\sigma_0+d+k}\right|^n$$

$$+x^{\sigma_1-\beta}\left|\frac{\varrho+d+k}{\sigma_1+ih+d+k}\right|^n + \int_{\sigma_0}^{\sigma_1} x^{\sigma-\beta}\left|\frac{\varrho+d+k}{\sigma+ih+d+k}\right|^n d\sigma.$$

Ferner bemerken wir, daß

$$\log|1+s| = \sigma - \tfrac{1}{2}(\sigma^2-t^2) + O(|s|^3)$$

für $|s| \leqq 1/2$ gilt. Liegt also s in einer beschränkten Menge, und ist k genügend groß (auch im Vergleich mit d), so erhält man

$$\begin{aligned} \left|\frac{\varrho+d+k}{s+d+k}\right| &= \exp\left(\log\left|1+\frac{\varrho+d}{k}\right| - \log\left|1+\frac{s+d}{k}\right|\right) \\ &= \exp\left\{\frac{\beta-\sigma}{k} + \frac{1}{2k^2}(\gamma^2-t^2+(\sigma-\beta)(\sigma+\beta+2d)) + O\left(\frac{1}{k^3}\right)\right\}. \end{aligned} \tag{4.6}$$

Es werden zwei weitere Parameter λ und a eingeführt:

$$0<\lambda<1 \quad \text{und} \quad a>0.$$

Wir setzen im folgenden voraus

$$X \geqq 2 \quad \text{und} \quad X^\lambda \leqq x \leqq X,$$

und n sei von nun an gegeben durch

$$n := [a \log X] + 1.$$

Setze abkürzend

$$\begin{aligned} \mu_1 &:= \lambda(\sigma_0-\beta) + a\left\{\frac{\beta-\sigma_0}{k} + \frac{1}{2k^2}(\gamma^2+(\sigma_0-\beta)(\sigma_0+\beta+2d)) + O\left(\frac{1}{k^3}\right)\right\}, \\ \mu_2 &:= (\sigma_1-\beta) + a\left\{\frac{\beta-\sigma_1}{k} + \frac{1}{2k^2}(\gamma^2-h^2+(\sigma_1-\beta)(\sigma_1+\beta+2d)) + O\left(\frac{1}{k^3}\right)\right\}. \end{aligned} \tag{4.7}$$

Mit (4.6) folgt

$$x^{\sigma_0-\beta}\left|\frac{\varrho+d+k}{\sigma_0+d+k}\right|^n \ll X^{\mu_1},$$

$$x^{\sigma_1-\beta}\left|\frac{\varrho+d+k}{\sigma_1+ih+d+k}\right|^n \ll X^{\mu_2}.$$

Setze schließlich noch voraus

$$\lambda > a/k$$

und

$$d \geqq -(\sigma_0+\sigma_1)/2. \tag{4.8}$$

Hieraus folgt

$$\max_{\sigma_0\leqq\sigma\leqq\sigma_1}(\sigma-\beta)(\sigma+\beta+2d)=(\sigma_1-\beta)(\sigma_1+\beta+2d).$$

Unter Verwendung von (4.6) folgt dann

$$\int_{\sigma_0}^{\sigma_1} x^{\sigma-\beta}\left|\frac{\varrho+d+k}{\sigma+ih+d+k}\right|^n d\sigma \ll X^{\mu_2},$$

und insgesamt daher

$$\left(\frac{x^\beta}{|\varrho+d+k|^n}\right)^{-1}\left|\int_{\partial D}\frac{F(s)x^s ds}{(s+d+k)^{n+2}}\right| \ll X^{\mu_1}+X^{\mu_2}. \tag{4.9}$$

Nun hat man $\mu_1<0$ und $\mu_2<0$ herbeizuführen. Den Term $O(1/k^3)$ lasse man zunächst unberücksichtigt.

Setze abkürzend

$$\mathsf{Y} := 1+\frac{1}{2k(\beta-\sigma_0)}(\gamma^2+(\sigma_0-\beta)(\sigma_0+\beta+2d)),$$

$$\mathsf{Z} := \frac{\dfrac{\sigma_1-\beta}{k}}{\dfrac{\sigma_1-\beta}{k}+\dfrac{1}{2k^2}(h^2-\gamma^2-(\sigma_1-\beta)(\sigma_1+\beta+2d))}.$$

Man postuliere

$$\lambda > (a/k)\mathsf{Y}, \tag{4.10}$$

$$a/k > \mathsf{Z}. \tag{4.11}$$

Aber auch

$$a/k < \lambda < 1 \tag{4.12}$$

muß erfüllt sein. Damit der Nenner von Z keine Sorgen bereitet, wünscht man sich auch noch

$$h^2-\gamma^2-(\sigma_1-\beta)(\sigma_1+\beta+2d)>0. \tag{4.13}$$

Angenommen es gilt

$$\gamma^2+(\sigma_0-\beta)(\sigma_0+\beta+2d)>0. \tag{4.14}$$

Dann ist

$$\mathsf{Y}>1. \tag{4.15}$$

Aber (4.14) ist gleichbedeutend mit

$$d<-\frac{1}{2}(\sigma_0+\beta)+\frac{1}{2}\frac{\gamma^2}{\beta-\sigma_0}.$$

Wir erinnern uns an (4.8). Dann sei also

$$-\frac{\sigma_0+\sigma_1}{2}\leqq d<-\frac{1}{2}(\sigma_0+\beta)+\frac{1}{2}\frac{\gamma^2}{\beta-\sigma_0}, \tag{4.16}$$

und dies ist möglich. Von nun an erfülle also d diese Bedingung. Um etwas Konkretes vor Augen zu haben, könnte man durchaus $d:=-(\sigma_0+\sigma_1)/2$ wählen.

Als nächstes wird

$$\mathsf{R}:=h^2-\frac{\sigma_1-\sigma_0}{\beta-\sigma_0}\gamma^2+(\sigma_1-\beta)(\sigma_1-\sigma_0)$$

gesetzt. Nach Voraussetzung (1.6) ist

$$\mathsf{R}>0. \tag{4.17}$$

Man rechnet nach, daß (4.13) equivalent mit

$$d<-\frac{1}{2}(\sigma_0+\beta)+\frac{1}{2}\frac{\gamma^2}{\beta-\sigma_0}+\frac{\mathsf{R}}{\sigma_1-\beta}$$

ist. Aber das stimmt wegen (4.16) und (4.17).

Man rechnet nach

$$\mathsf{YZ}<1 \Leftrightarrow (4.17).$$

Nun kann man

$$\mathsf{Z}<a/k<1/\mathsf{Y}$$

wählen. Dann haben wir bereits (4.11), und es ist $(a/k)\mathsf{Y}<1$. Also kann man wählen

$$(a/k)\mathsf{Y}<\lambda<1.$$

Hieraus und aus (4.15) folgen (4.10) und (4.12).

Nun fehlt aber noch $O(1/k^3)$. Dazu braucht man nur $k\geqq k_0$ zu wählen.

Einerseits erhalten wir jetzt aus (4.3), (4.4) und (4.9)

$$V(A_{n+2},X)\geqq(1-\lambda')(\gamma/\pi)\log X \tag{4.18}$$

für gewisses $\lambda'\in(0,1)$, $\lambda'>\lambda$ und $X\geqq X_0\geqq 2$. Andererseits ergibt die Anwendung des Lemmas

$$V(f,X)\geqq V(A_1,X)\geqq\dots\geqq V(A_{n+2},X)\qquad(X>0),$$

was zusammen mit (4.18) zu

$$V(f, X) \geqq c \log X \qquad (X \geqq X_0)$$

führt, wo $c = (1-\lambda')\gamma/\pi$ ist.

Schließlich kann man die postulierte effektive Abhängigkeit der Konstanten c und X_0 durch die Nachprüfung der obigen Überlegungen a posteriori feststellen. □

Fall $r=1$, $m_1 \geqq 2$ *(Skizze des Beweises).* Mit Ausnahme der anderen Gestalt von $\operatorname*{Res}_{s=\varrho} \Phi(s)$, wobei

$$\Phi(s) = \frac{F(s)x^s}{(s+d+k)^{n+2}} \tag{4.19}$$

bezeichnet, führt man den Beweis analog dem Fall $r=1$, $m_1=1$ durch. Man hat nämlich (statt m_1 wird einfach m geschrieben)

$$\operatorname*{Res}_{s=\varrho} \Phi(s) = \frac{x^\varrho}{(\varrho+d+k)^{n+2}} \sum_{l=-m}^{-1} \frac{a_l}{(-l-1)!} \sum_{j=0}^{-l-1} \binom{-l-1}{j}$$
$$\times (\log x)^{-l-1-j} \frac{(-n-2)(-n-3)\dots(-n-2-j+1)}{(\varrho+d+k)^j}.$$

Wählen wir a, d, k, λ und n wie im Fall $m=1$. Für $X^\lambda \leqq x \leqq X$ ist dann

$$\operatorname*{Res}_{s=\varrho} \Phi(s) = \frac{a_{-m}}{(m-1)!} \frac{x^\varrho}{(\varrho+d+k)^{n+2}} \left\{ \left(\log x - \frac{n}{\varrho+d+k} \right)^{m-1} + O((\log X)^{m-2}) \right\}.$$

Wegen $\lambda \log X \leqq \log x \leqq \log X$ und $\gamma > 0$ gilt

$$\left| \log x - \frac{n}{\varrho+d+k} \right| \gtrless \log X, \tag{4.20}$$

und daher

$$\operatorname*{Res}_{s=\varrho} \Phi(s) = \frac{a_{-m}}{(m-1)!} \frac{x^\varrho}{(\varrho+d+k)^{n+2}} \left(\log x - \frac{n}{\varrho+d+k} \right)^{m-1} (1+O((\log X)^{-1})). \tag{4.21}$$

Hieraus ersieht man nun, daß die im Fall $m=1$ durchgeführten Überlegungen mutatis mutandis wiederholt werden können. Das Glied

$$2 \operatorname{Re} \frac{a_{-m}}{(m-1)!} \frac{x^\varrho}{(\varrho+d+k)^{n+2}} \left(\log x - \frac{n}{\varrho+d+k} \right)^{m-1}$$
$$= \frac{2|a_{-m}|}{(m-1)!} \frac{x^\beta}{|\varrho+d+k|^{n+2}} \left| \log x - \frac{n}{\varrho+d+k} \right|^{m-1} \cos(\gamma \log x$$
$$+ \alpha_1 - (n+2)\alpha_2 + (m-1)\alpha(x)), \tag{4.22}$$

wobei

$$\alpha(x) = \operatorname{Arg}\left(\log x - \frac{n}{\varrho+d+k} \right) \qquad (X^\lambda \leqq x \leqq X)$$

und $\alpha_1 = \operatorname{Arg}(a_{-m})$, $\alpha_2 = \operatorname{Arg}(\varrho + d + k)$ gesetzt wird, stellt insbesondere die gewünschte Oszillation her. Der in (4.21) szs. zusätzlich vorkommende Faktor $(\log x - n/(\varrho + d + k))^{m-1}$ ermöglicht uns um so mehr, eine Ungleichung vom Typus (4.9) (mit $\mu_1 < 0$, $\mu_2 < 0$) herzuleiten. Bemerken wir noch, daß die Funktion $\alpha : [X^\lambda, X] \to \mathbb{R}$ stetig und beschränkt ist. Um die gewünschte Anzahl geeigneter Oszillationen von (4.22) zu bekommen, können wir daher die Zwischenwerteigenschaft stetiger Funktionen verwenden. □

Fall $r \geqq 2$ (Skizze des Beweises). Wie im Fall $r = 1$ werden die Funktionen $A_n (n \geqq 2)$ eingeführt. Mit der Bezeichnung (4.19) erhält man

$$A_{n+2}(x) = 2 \sum_{\nu=1}^{r} \operatorname{Re}\left(\operatorname*{Res}_{s=\varrho_\nu} \Phi(s)\right) + \frac{1}{2\pi i} \int_{\partial D} \Phi(s) ds, \tag{4.23}$$

wobei (4.1) angenommen ist (d, $k \in \mathbb{R}$ sind Parameter).

Unter der Voraussetzung

$$X \geqq 2, X^\lambda \leqq x \leqq X \quad \text{und} \quad n = [a \log X] + 1,$$

wo $\lambda \in (0,1)$ und $a > 0$ Parameter sind, die passend zu wählen sind, bekommt man aus (4.20) und (4.21)

$$\left|\operatorname*{Res}_{s=\varrho_\nu} \Phi(s)\right| \begin{smallmatrix}\ll\\\gg\end{smallmatrix} \frac{x^{\beta_\nu}}{|\varrho_\nu + d + k|^n} (\log X)^{m_\nu - 1} \qquad (\nu = 1, \ldots, r). \tag{4.24}$$

Nun sind die Pole ϱ_ν geeignet zu ordnen. Zu diesem Zweck wird zuerst ein Pol $\varrho = \varrho_{\nu_0}$ fixiert, für den die Bedingung (1.6) erfüllt ist. Ferner wird d so gewählt, daß

$$\frac{\gamma^2}{\beta - \sigma_0} - \sigma_0 - \beta < 2d < \frac{h^2 - \gamma^2}{\sigma_1 - \beta} - \sigma_1 - \beta \tag{4.25}$$

ist. Dies ist möglich wegen (1.6). Für s, $s' \in \mathbb{H}$ wird

$$D(s', s) = t^2 - t'^2 - (\sigma - \sigma')(\sigma + \sigma' + 2d)$$

gesetzt, und in $\mathbb{H}$ die Relation $\mathscr{P}$ eingeführt:

$$s' \mathscr{P} s \Leftrightarrow D(s', s) > 0, \quad \text{oder} \quad D(s', s) = 0 \quad \text{und} \quad \sigma' \geqq \sigma \qquad (s, s' \in \mathbb{H}).$$

$\mathscr{P}$ ist eine Ordnung in $\mathbb{H}$ (d.h., sie ist reflexiv, antisymmetrisch, transitiv und hat die Eigenschaft: für beliebige $s, s' \in \mathbb{H}$ ist $s \mathscr{P} s'$ oder $s' \mathscr{P} s$). Insbesondere gibt es in der Menge $\{\varrho_1, \ldots, \varrho_r\} \subset \mathbb{H}$ ein Element $\varrho' = \varrho_{\nu'} (= \beta' + i\gamma')$ derart, daß

$$\varrho' \mathscr{P} \varrho_\nu \tag{4.26}$$

für $\nu = 1, \ldots, r$ ist. Schreibt man $\varrho_0 = \sigma_0$, $\varrho_{r+1} = \sigma_1 + ih$, so hat man aus (4.25): $\varrho \mathscr{P} \varrho_0$ und $\varrho \mathscr{P} \varrho_{r+1}$. Daher gilt (4.26) auch für $\nu = 0$, $r + 1$.

Die Vorzeichenwechsel von A_{n+2} werden von dem gewonnenen Pol ϱ' stammen.

Es gilt (4.8). Unter Voraussetzung (4.12) erhält man aus (4.24)

$$\left(\left|\operatorname*{Res}_{s=\varrho'} \Phi(s)\right|\right)^{-1} \left|\int_{\partial D} \Phi(s) ds\right| \ll X^{\mu'_0} + X^{\mu'_{r+1}},$$

we μ_0', μ_{r+1}' zu μ_1, μ_2 [s. (4.7)] *analoge* Ausdrücke sind. Auch ist

$$\left(\left|\operatorname*{Res}_{s=\varrho'} \Phi(s)\right|\right)^{-1} \left|\operatorname*{Res}_{s=\varrho_\nu} \Phi(s)\right| \ll X^{\mu_\nu'}(\log X)^{m_\nu - m_{\nu'}} \qquad (\nu=1,\ldots,r;\ \nu \neq \nu'),$$

wobei $\mu_\nu'(\nu \neq \nu')$ analog zu setzen sind.

Mit geeigneter Wahl der Parameter a, λ und k führt man nun $\mu_\nu'<0$ $(\nu=0,\ldots,r+1;\ \nu \neq \nu')$ herbei. Dabei kann man sich auf (4.26) stützen und im allgemeinen wie im Beweis des Satzes im Fall $r=1$ vorgehen. Der technische Unterschied besteht darin, daß aktuell der Parameter d anders als im Fall $r=1$ oben gewählt wird [vgl. (4.16) und (4.25)], und das ist zu berücksichtigen.

Man schreibt daher (4.23) in Gestalt

$$A_{n+2}(x)=2\operatorname{Re}\left(\operatorname*{Res}_{s=\varrho'} \Phi(s)\right)(1+O(X^{\mu'})),$$

wo $\mu'<0$ ist. In derselben Weise wie im Fall $r=1$ betrachtet man $\operatorname*{Res}_{s=\varrho'} \Phi(s)$. Unter Verwendung des Lemmas folgt die Behauptung des Satzes. □

Danksagungen. Artikel I und II dieser Reihe bilden eine verbesserte Version eines Teils meiner Doktorarbeit. Herrn Professor Włodzimierz Staś danke ich herzlich für wissenschaftliche Leitung. Mein Dank gilt auch Jerzy Kaczorowski für eine Reihe von wertvollen Bemerkungen. Dem anonymen Referenten, der zahlreiche Verbesserungen vorgeschlagen hat, bin ich besonders verpflichtet.

Literatur

1. Courant, R., Hilbert, D.: Methoden der mathematischen Physik. Bd. I. 2. Aufl. Berlin: Springer 1937
2. Kaczorowski, J.: On sign-changes in the remainder-term of the prime-number formula. I. II. Acta Arith. **44**, 365–377 (1984); ibid. **45**, 65–74 (1985)
3. Kaczorowski, J., Pintz, J.: Oscillatory properties of arithmetical functions. I. Acta Math. Hung. **48** (1–2), 173–185 (1986)
4. Lehman, R.: On the difference $\pi(x)-\operatorname{li}(x)$. Acta Arith. **11**, 397–410 (1966)
5. Littlewood, J.E.: Sur la distribution des nombres premiers. C. R. Acad. Sci. Paris **158**, 1869–1872 (1914)
6. Riemann, B.: Gesammelte mathematische Werke und wissenschaftlicher Nachlass. 2. Aufl. Leipzig: Teubner 1892
7. Szydło, B.: Über Vorzeichenwechsel einiger arithmetischer Funktionen. II. Math. Ann. **283**, 151–163 (1989)

Eingegangen am 26. Oktober 1987; revidierte Fassung am 31. Mai 1988

Math. Ann. 283, 151–163 (1989)

Mathematische
Annalen

Über Vorzeichenwechsel einiger arithmetischer Funktionen. II

Bogdan Szydło

Mathematisches Institut der Adam-Mickiewicz-Universität, ul. Matejki 48/49, PL-60-769 Poznań, Poland

1. Einleitung

In dem vorliegenden Artikel dieser Reihe wenden wir die allgemeine, in [12] eingeführte Methode auf die Untersuchung der Vorzeichenwechsel einer konkreten, für die Primzahltheorie wichtigen Funktion an. Die erwähnte Methode entwickelt das Verfahren von Kaczorowski [7, 8]. Die Hauptneuerung besteht generell darin, daß wir statt des Operators δ von Kaczorowski den allgemeineren A benutzen. Für eine genauere Schilderung der Methode, Bezeichnungen und einige allgemeine Ergebnisse s. [12].

Es sei ψ die Tschebyschevsche Funktion. Für das Restglied im Primzahlsatz in Gestalt

$$\lim_{x\to\infty} \psi(x)/x = 1$$

geschrieben, führen wir die Bezeichnung $\Delta(x) = \psi(x) - x$ und für die Anzahl der Vorzeichenwechsel von Δ im Intervall $(0, X]$ $(X > 0)$ die Bezeichnung $V(\Delta, X)$ ein.

Das Ziel dieses Artikels ist, die folgenden Sätze zu beweisen.

Satz 1. *Für $X \geqq 10^{2250}$ gilt*

$$V(\Delta, X) \geqq 0{,}013 \log X .$$

Satz 2. *Es sei $H \geqq 501{,}5$. Jede nichttriviale Nullstelle $\varrho = \beta + i\gamma$ der Riemannschen Zetafunktion ζ, die im Streifen $|t| < H$ liegt, genüge der Bedingung*

$$\beta = 1/2 .$$

Dann gilt für $X \geqq \exp(0{,}09 \max\{4400, H\})$ die Ungleichung

$$V(\Delta, X) \geqq \left(1 - \frac{3}{H}\right) \frac{\gamma_0}{\pi} \log X ,$$

wo $\gamma_0 = 14{,}13\ldots$ den Imaginärteil der niedrigsten nichttrivialen Nullstelle $\varrho_0 = \beta_0 + i\gamma_0$ von ζ bezeichnet.

Aus Satz 2 ziehen wir hier die unmittelbaren Folgerungen.

Folgerung 1. *Für* $X \geqq \exp(198594)$ *gilt*

$$V(\Delta, X) \geqq 0{,}994 \frac{\gamma_0}{\pi} \log X .$$

Beweis. Wir wenden Satz 2 mit $H = 501{,}5$ an; vgl. [13, S. 331]. □

Folgerung 2. *Für* $X \geqq \exp(9 \cdot 10^{14})$ *gilt*

$$V(\Delta, X) \geqq 0{,}99999997 \frac{\gamma_0}{\pi} \log X .$$

Beweis. Wir wenden Satz 2 mit $H = 10^8$ an; vgl. [9]. □

In dem Satz der Arbeit von Kaczorowski [7, I] finden wir das Resultat

$$V(\Delta, X) \geqq \frac{\gamma_0}{4\pi} \log X \qquad (X \geqq X_0), \tag{1.1}$$

wo X_0 eine effektive, aber dort nicht berechnete Konstante bedeutet. Abgesehen von der explizit berechneten unteren Schranke X_0 für X, ist Satz 1 schwächer und Folgerung 1 besser als (1.1). Im Gegensatz zu (1.1) sind aber diese Ergebnisse ganz „komputerfrei", während in [7, I] man die Lokalisation der nichttrivialen Nullstellen von ζ bis zur Höhe $H = 10^6$ benutzt, was umfangreiche Komputerrechnungen verlangte.

Ferner bemerken wir, daß die Anwendung des Operators δ, wie es in [7, 8] gemacht wird, die Lokalisation von wenigstens einigen tausend nichttrivialen Nullstellen von ζ erfordert, um überhaupt ein Resultat vom Typus

$$V(\Delta, X) \geqq c \log X \qquad (X \geqq X_0)$$

mit effektiven (explizit gegebenen) Konstanten $c > 0$ und $X_0 \geqq 2$ geben zu können.

Die Beweismethode von Satz 1 ist einerseits das erste Beispiel für eine konkrete Wirkung der in Artikel I dieser Reihe präsentierten Methode. Anderseits benutzt man während des Beweises bescheidene Information über die zwei niedrigsten Nullstellen $\varrho_0 = \beta_0 + i\gamma_0$ und $\varrho_1 = \beta_1 + i\gamma_1$ von ζ. Wir werden uns nämlich nur darauf stützen, daß ϱ_0 eine einfache Nullstelle und

$$\beta_0 = 1/2, \tag{1.2}$$

$$14{,}13 \leqq \gamma_0 \leqq 14{,}14, \tag{1.3}$$

$$21 \leqq \gamma_1 \tag{1.4}$$

ist; vgl. [4], auch [2, 11] für die Geschichte der numerischen Bestimmung der Nullstellen von ζ. Es ist zuzugeben, daß Satz 1 gerade das leistet, was bereits der allgemeine Satz aus Artikel I im Spezialfall antizipiert.

Da die Riemannsche Vermutung bis zu sehr großen H verifiziert ist (und wird), wird Satz 2 zur Grundlage solcher Ergebnisse wie Folgerung 2.

Hilfssätze 4, 5 und 6 sind keine vollkommen originalen Resultate, denn in den Beweisen benutzen wir bekannte klassische Argumente, die z. B. in [1, §§ 13, 15, 17] zu finden sind. Die notwendigen Werte der zumeist im Symbol $O(\cdot)$ implizierten numerischen Konstanten sind jedoch unsere Produkte.

Abschließend erwähnen wir, daß bei der Aufgabe, analoge (effektive) Resultate für $V(\Delta_1, X)$ zu geben, wo $\Delta_1(x) = \pi(x) - \mathrm{li}(x)$ $(x > 0)$ das Restglied im Primzahlsatz in Gestalt $\lim_{x \to \infty} \pi(x)/\mathrm{li}(x) = 1$ bezeichnet, die Methode versagt; vgl. jedoch [7].

Bemerkung zur Bezeichnung. In diesem Artikel bezeichnet $\varrho = \beta + i\gamma$ ($\beta, \gamma \in \mathbb{R}$) eine nichttriviale Nullstelle der Riemannschen Zetafunktion ζ; außerdem behalten wir die Bezeichnungen von [12] bei.

2. Hilfssätze

Hilfssatz 1. *Für* $1 < \sigma \leqq 2$ *gilt*

$$-\frac{\zeta'}{\zeta}(\sigma) \leqq \frac{1}{\sigma - 1}. \tag{2.1}$$

Beweis. Für den Fall $1 < \sigma \leqq 1{,}5$ findet sich ein Beweis von (2.1) in [3, S. 184f.]. Bemerken wir, daß die dort angegebene Überlegung auch zum Fall $1{,}5 \leqq \sigma \leqq 2$ verwendbar ist. □

Hilfssatz 2 (Poisson, vgl. [10, S. 181f.]). *Für* $\sigma > 0$ *gilt*

$$\frac{\Gamma'}{\Gamma}(s) = \log s - \frac{1}{2s} - 2\int_0^\infty \frac{\tau\, d\tau}{(\tau^2 + s^2)(e^{2\pi\tau} - 1)}.$$

Hilfssatz 3. *Es gilt*

$$I = \int_0^\infty \frac{\tau\, d\tau}{e^{2\pi\tau} - 1} \leqq \frac{1}{8}.$$

Beweis. Spalte auf

$$I = \left(\int_0^{\tau_0} + \int_{\tau_0}^\infty\right) \frac{\tau\, d\tau}{e^{2\pi\tau} - 1} \leqq \int_0^{\tau_0} \frac{\tau\, d\tau}{2\pi\tau} + \int_0^\infty \frac{3!\,\tau\, d\tau}{(2\pi\tau)^3}$$
$$= \frac{1}{2\pi}\left(\tau_0 + \frac{1}{2\pi^2 \tau_0}\right)$$

und wähle $\tau_0 = \sqrt{6}/(2\pi)$. □

Hilfssatz 4. *Für* $h \geqq 1$ *gilt*

$$\max_{0 \leqq t \leqq h} \left|\frac{\zeta'}{\zeta}(-1 + ih)\right| \leqq \log h + 8.$$

Beweis. Die Funktionalgleichung von ζ, in der asymmetrischen Gestalt [1, S. 73]

$$\zeta(1-s) = 2^{1-s}\pi^{-s} \cos\frac{\pi s}{2}\, \Gamma(s)\, \zeta(s)$$

geschrieben, gibt

$$\frac{\zeta'}{\zeta}(1-s) = \log 2\pi + \frac{\pi}{2}\operatorname{tg}\frac{\pi s}{2} - \frac{\Gamma'}{\Gamma}(s) - \frac{\zeta'}{\zeta}(s).$$

Weiter setzen wir $s = 2 + it$ und benutzen die Dreiecksungleichung:

$$\left|\frac{\zeta'}{\zeta}(-1 + it)\right| = \left|\frac{\zeta'}{\zeta}(-1 - it)\right|$$
$$\leqq \log 2\pi + \left\{-\frac{\zeta'}{\zeta}(2)\right\} + \frac{\pi}{2}\left|\operatorname{tg}\left(\frac{\pi}{2}(2 + it)\right)\right| + \left|\frac{\Gamma'}{\Gamma}(2 + it)\right|. \tag{2.2}$$

Um das letzte Glied in der rechten Seite dieser Ungleichung abzuschätzen, verwenden wir Hilfssätze 2 und 3. Für $t \geqq 1$ erhalten wir

$$\left|\frac{\Gamma'}{\Gamma}(2+it)\right|$$
$$\leqq \log|2+it| + \frac{\pi}{2} + \frac{1}{2|2+it|} + 2\int_0^\infty \frac{\tau\, d\tau}{|\tau^2+s^2|(e^{2\pi\tau}-1)}$$
$$\leqq \log t + \log 5 + \frac{\pi}{2} + \frac{1}{2\sqrt{5}} + \frac{I}{2t} \leqq \log t + 3{,}47,$$

für $t \in [0,1]$ dagegen

$$\left|\frac{\Gamma'}{\Gamma}(2+it)\right| \leqq \log\sqrt{5} + \operatorname{arc\,tg} 0{,}5 + 0{,}25 + \frac{2I}{3} \leqq 2.$$

Aus (2.1) und (2.2) bekommen wir daher für $h \geqq 1$

$$\max_{0 \leqq t \leqq h} \left|\frac{\zeta'}{\zeta}(-1+it)\right| \leqq \log h + 3{,}47 + \log 2 + 1 + \frac{\pi}{2}$$
$$\leqq \log h + 8. \quad \square$$

Hilfssatz 5. *Für* $-1 \leqq \sigma \leqq 2$ *und* $t \geqq 20$ *gilt*

$$\left|\frac{\zeta'}{\zeta}(s)\right| \leqq \left|\sum_{|\gamma - t| \leqq 1/2} \frac{1}{s-\varrho}\right| + 2{,}6 \log t.$$

Beweis. Man hat [1, S. 80ff.]

$$-\frac{\zeta'}{\zeta}(s) = -\frac{\log \pi}{2} + \frac{1}{s-1} + \frac{1}{2}\frac{\Gamma'}{\Gamma}\left(\frac{s}{2}+1\right) - \sum_{\varrho} \frac{1}{s-\varrho}. \tag{2.3}$$

Hieraus folgt

$$\begin{aligned}\left|-\frac{\zeta'}{\zeta}(s)\right| &\leqq \left|-\frac{\zeta'}{\zeta}(s) + \frac{\zeta'}{\zeta}(2+it)\right| + \left|-\frac{\zeta'}{\zeta}(2+it)\right| \\ &\leqq \left|\sum_{|\gamma-t|<1/2} \frac{1}{s-\varrho}\right| + \left|\sum_{|\gamma-t|<1/2} \frac{1}{2+it-\varrho}\right| \\ &\quad + \sum_{|\gamma-t| \geqq 1/2} \left|\frac{1}{s-\varrho} - \frac{1}{2+it-\varrho}\right| + \left|\frac{1}{s-1} - \frac{1}{2+it-1}\right| \\ &\quad + \frac{1}{2}\left|\frac{\Gamma'}{\Gamma}\left(\frac{s}{2}+1\right) - \frac{\Gamma'}{\Gamma}\left(\frac{2+it}{2}+1\right)\right| + \left\{-\frac{\zeta'}{\zeta}(2)\right\}.\end{aligned} \tag{2.4}$$

Für $|\gamma - t| \geqq 1/2$ gilt

$$\left|\frac{1}{s-\varrho} - \frac{1}{2+it-\varrho}\right| \leqq 3\sqrt{3}\,\mathrm{Re}\frac{1}{2+it-\varrho},$$

für $|\gamma - t| < 1/2$ ist auch

$$\left|\frac{1}{2+it-\varrho}\right| \leqq 3\sqrt{3}\,\mathrm{Re}\frac{1}{2+it-\varrho}.$$

Aber aus (2.3) folgt

$$\mathrm{Re}\sum_{\varrho}\frac{1}{2+it-\varrho}$$

$$\leqq\left\{-\frac{\zeta'}{\zeta}(2)\right\}-\frac{\log\pi}{2}+\mathrm{Re}\frac{1}{2+it-1}+\frac{1}{2}\mathrm{Re}\frac{\Gamma'}{\Gamma}\left(\frac{2+it}{2}+1\right),\tag{2.5}$$

und daher bekommt man aus (2.4)

$$\begin{aligned}\left|\frac{\zeta'}{\zeta}(s)\right|\leqq&\left|\sum_{|\gamma-t|<1/2}\frac{1}{s-\varrho}\right|+\frac{3\sqrt{3}}{2}\mathrm{Re}\frac{\Gamma'}{\Gamma}\left(2+\frac{it}{2}\right)\\&+(3\sqrt{3}+1)\left\{-\frac{\zeta'}{\zeta}(2)\right\}-\frac{3\sqrt{3}}{2}\log\pi\\&+\frac{1}{2}\left|\frac{\Gamma'}{\Gamma}\left(\frac{s}{2}+1\right)-\frac{\Gamma'}{\Gamma}\left(2+\frac{it}{2}\right)\right|+\left|\frac{1}{s-1}-\frac{1}{1+it}\right|\\&+3\sqrt{3}\,\mathrm{Re}\frac{1}{1+it}.\end{aligned}\tag{2.6}$$

Mit der Beweismethode von Hilfssatz 1 [3, S. 184f.] und unter Berücksichtigung von [1, S. 81, (10)] erhält man leicht

$$-\frac{\zeta'}{\zeta}(2)\leqq\frac{3}{2}-\frac{1}{2}(\log\pi+C),$$

wobei $C=0{,}577215\ldots$ die Eulersche Konstante bedeutet.

Um die Behauptung des Hilfssatzes zu bekommen, schätzt man jetzt unter Verwendung von Hilfssätzen 2 und 3 die rechte Seite von (2.6) genug scharf ab. □

Hilfssatz 6. *Unter den Voraussetzungen von Satz 2 bezeichne $N^*(T)$ die Anzahl der nichttrivialen Nullstellen von ζ im Streifen $T-1/2<t<T+1/2$, wobei $500\leqq T\leqq H-1/2$ ist. Dann gilt*

$$N^*(T)\leqq\tfrac{5}{6}\log T.$$

Beweis. Wenn $\beta=1/2$ und $|\gamma-T|<1/2$ ist, so gilt

$$\frac{6}{10}\leqq\mathrm{Re}\frac{1}{2+iT-\varrho}.$$

Also erhält man

$$N^*(T)\leqq\frac{10}{6}\mathrm{Re}\sum_{\varrho}\frac{1}{2+iT-\varrho},$$

nach (2.5) ist dies

$$\leqq\frac{10}{6}\left\{-\frac{\zeta'}{\zeta}(2)-\frac{\log\pi}{2}+\mathrm{Re}\frac{1}{1+iT}+\frac{1}{2}\mathrm{Re}\frac{\Gamma'}{\Gamma}\left(2+\frac{iT}{2}\right)\right\}\leqq\frac{5}{6}\log T.\quad\square$$

3. Beweis von Satz 1

Auf dieselbe Weise wie im Beweis des Satzes aus [12] führen wir den Operator A ein, dessen Einwirkung auf eine reelle Funktion $f:(0,\infty)\to\mathbb{R}$ durch

$$\mathrm{A}(f)(x)=x^{-k}\int_0^x f(\xi)\xi^{k-1}\,d\xi \qquad (x>0) \tag{3.1}$$

definiert wird, falls f und $k\geqq 0$ geeigneten Bedingungen genügen.

Setzen wir

$$k=1000 \tag{3.2}$$

und bezeichnen mit $A_n(f)$ das Ergebnis der n-fachen Iteration des Operators A.

Da

$$\mathfrak{M}(\psi)(s)=-\frac{\zeta'}{\zeta}(s)\frac{1}{s}$$

ist, so folgt

$$A_{n+1}(\psi)(x)=\frac{1}{2\pi i}\int_{\sigma_1-i\infty}^{\sigma_1+\infty}\left\{-\frac{\zeta'}{\zeta}(s)\right\}\frac{x^s\,ds}{s(s+k)^{n+1}} \tag{3.3}$$

für $\sigma_1>1$, $n\geqq 1$, $x>0$; vgl. den Anfang des Beweises des Satzes aus [12], insbesondere (4.2).

Es seien

$$\sigma_0=-1,\ \sigma_1=1{,}01,\ h=20{,}5 \tag{3.4}$$

und bezeichne

$$D=\{s\in\mathbb{C}:\sigma_0\leqq\sigma\leqq\sigma_1,|t|\leqq h\}\cup\{s\in\mathbb{C}:\sigma\geqq\sigma_1\}, \tag{3.5}$$

außerdem

$$\alpha_1=\operatorname{Arg}\varrho_0,\ \alpha_2=\operatorname{Arg}(\varrho_0+k). \tag{3.6}$$

Bekanntlich ist [2, S. 66, (1)]

$$\frac{\zeta'}{\zeta}(0)=\log 2\pi. \tag{3.7}$$

Da ϱ_0 eine einfache Nullstelle von ζ ist, schließen wir aus dem Residuensatz und (1.2)–(1.4), (3.3)–(3.7), daß

$$\begin{aligned}A_{n+1}(\Delta)(x)&=-2\,\mathrm{Re}\frac{x^{\varrho_0}}{\varrho_0(\varrho_0+k)^{n+1}}-\frac{\log 2\pi}{k^{n+1}}\\&\quad+\frac{1}{2\pi i}\int_{\partial D}\left\{-\frac{\zeta'}{\zeta}(s)\right\}\frac{x^s\,ds}{s(s+k)^{n+1}}\\&=-2\frac{x^{1/2}}{|\varrho_0|\,|\varrho_0+k|^{n+1}}\{\cos(\gamma_0\log x-\alpha_1-(n+1)\alpha_2)+R_1+R_2\}\end{aligned} \tag{3.8}$$

stattfindet, wobei

$$|R_1|\leqq\frac{\log(2\pi)}{2}|\varrho_0|x^{-1/2}\left|\frac{\varrho_0+k}{k}\right|^{n+1} \tag{3.9}$$

und

$$|R_2| \leqq \frac{|\varrho_0|\,|\varrho_0+k|^{n+1}}{4\pi} \left| \int_{\partial D} \left\{ -\frac{\zeta'}{\zeta}(s) \right\} \frac{x^{s-1/2}\,ds}{s(s+k)^{n+1}} \right| \tag{3.10}$$

ist.

Ferner setzen wir voraus

$$X \geqq 2 \quad \text{und} \quad X^{0{,}997} \leqq x \leqq X\,. \tag{3.11}$$

Wir wählen

$$n = [830 \log X] + 1\,. \tag{3.12}$$

Im folgenden benutzen wir die für $|s| \leqq 0{,}2$ geltende Beziehung:

$$\log|1+s| = \sigma - \tfrac{1}{2}(\sigma^2 - t^2) + \tfrac{1}{3}(\sigma^3 - 3\sigma t^2) + R \tag{3.13}$$

mit

$$|R| \leqq 0{,}3\,|s|^4\,.$$

Zuerst beschäftigen wir uns mit R_1. Auf Grund (1.2), (1.3), (3.2) und (3.13) erhalten wir

$$\begin{aligned} \log\left|1+\frac{\varrho_0}{k}\right| &\leqq \frac{\beta_0}{k} + \frac{1}{2k^2}(\gamma_0^2 - \beta_0^2) + \frac{1}{3k^3}(\beta_0^3 - 3\beta_0\gamma_0^2) \\ &\quad + \frac{0{,}3}{k^4}(\beta_0^2 + \gamma_0^2)^2 \\ &\leqq \frac{0{,}5}{k} + \frac{100}{k^2} = \frac{0{,}6}{k} \leqq 0{,}0006\,. \end{aligned} \tag{3.14}$$

Hieraus und aus (3.11), (3.12) folgt

$$|R_1| \leqq 13{,}1\,X^{-0{,}0005}\,. \tag{3.15}$$

Mit (vgl. [12, (4.5)])

$$\begin{aligned} L_1 &= \{s \in \mathbb{C} : \sigma = \sigma_0,\, 0 \leqq t \leqq h\}\,, \\ L_2 &= \{s \in \mathbb{C} : \sigma = \sigma_1,\, t \geqq h\}\,, \\ L_3 &= \{s \in \mathbb{C} : \sigma_0 \leqq \sigma \leqq \sigma_1,\, t = h\} \end{aligned} \tag{3.16}$$

folgt

$$|R_2| \leqq R_{21} + R_{22} + R_{23}\,, \tag{3.17}$$

wobei

$$R_{2j} := \frac{1}{2\pi} |\varrho_0|\,|\varrho_0+k|^{n+1} \left| \int_{L_j} \left\{ -\frac{\zeta'}{\zeta}(s) \right\} \frac{x^{s-1/2}\,ds}{s(s+k)^{n+1}} \right| \qquad (j=1,2,3) \tag{3.18}$$

gesetzt wird.

Es ist

$$\left|\int_{L_1}\right| \leqq \max_{0 \leqq t \leqq h} \left| -\frac{\zeta'}{\zeta}(-1+it) \right| (1+\log h) \frac{x^{-3/2}}{(k-1)^{n+1}}\,.$$

Analog wie vorher schätzen wir ab:

$$\log\left|1-\frac{1}{k}\right| \geqq -\frac{1}{k}-\frac{1}{2k^2}-\frac{1}{3k^3}-\frac{0,3}{k^4} \geqq -\frac{1}{k}-\frac{0,501}{k^2} \geqq -\frac{1,001}{k}, \tag{3.19}$$

was zusammen mit (3.14) gibt

$$\log\left|\frac{\varrho_0+k}{k-1}\right| \leqq \frac{1,601}{k} \leqq 0,002.$$

Hieraus, aus Hilfssatz 4, zusammen mit (3.11) und (3.12), folgt

$$R_{21} \leqq 110 X^{-0,16667}. \tag{3.20}$$

Aus Hilfssatz 1 folgt weiter

$$\left|\int_{L_2}\right| \leqq \left\{-\frac{\zeta'}{\zeta}(1,01)\right\}\left(\frac{1}{k}\int_h^k \frac{dt}{t}+\int_k^\infty \frac{dt}{t^2}\right)\frac{x^{0,51}}{|\sigma_1+ih+k|^n}$$
$$\leqq \frac{100(1+\log(k/h))}{k}\frac{x^{0,51}}{|\sigma_1+ih+k|^n}.$$

Überdies ist

$$\log\left|1+\frac{\sigma_1+ih}{k}\right| \geqq \frac{\sigma_1}{k}+\frac{1}{2k^2}(h^2-\sigma_1^2)+\frac{1}{3k^3}(\sigma_1^3-3\sigma_1 h^2)$$
$$-\frac{0,3}{k^4}(\sigma_1^2+h^2)^2$$
$$\geqq \frac{1,01}{k}+\frac{209}{k^2} \geqq \frac{1,219}{k},$$

was zusammen mit (3.14) gibt

$$\log\left|\frac{\varrho_0+k}{\sigma_1+ih+k}\right| \leqq -\frac{0,619}{k},$$

also

$$R_{22} \leqq 1150 X^{-0,00377}. \tag{3.21}$$

Es bleibt noch übrig, das Integral $\int\limits_{L_3}$ zu betrachten. Es ist

$$\left|\int_{L_3}\right| \leqq \max_{\sigma_0 \leqq \sigma \leqq \sigma_1}\left|\frac{\zeta'}{\zeta}(\sigma+ih)\right|\frac{1}{h(k-1)}\int_{\sigma_0}^{\sigma_1}\frac{x^{\sigma-1/2}\,d\sigma}{|\sigma+ih+k|^n}.$$

Für $\sigma_0 \leqq \sigma \leqq \sigma_1$ ist

$$\log\left|1+\frac{\sigma+ih}{k}\right| \geqq \frac{\sigma}{k}+\frac{1}{2k^2}(h^2-\sigma^2)+\frac{1}{3k^3}(\sigma^3-3\sigma h^2)$$
$$-\frac{0,3}{k^4}(\sigma^2+h^2)^2$$
$$\geqq \frac{\sigma}{k}+\frac{209}{k^2},$$

was zusammen mit (3.14) gibt

$$\log\left|\frac{\varrho_0+k}{\sigma+ih+k}\right| \leqq \frac{1/2-\sigma}{k} - \frac{109}{k^2}.$$

Es ist leicht nachzusehen, daß für $X \geqq 2$ die Ungleichung

$$X^{0,997} \geqq \exp(n/k)$$

gilt. Außerdem stellen wir auf Grund von Hilfssatz 5 und (1.3), (1.4) fest, daß

$$\max_{\sigma_0 \leqq \sigma \leqq \sigma_1} \left|\frac{\zeta'}{\zeta}(\sigma+ih)\right| \leqq 2{,}6 \log h$$

gilt. Nach diesen Vorbereitungen bekommen wir die Kette von Ungleichungen:

$$\begin{aligned} R_{23} &\leqq \frac{2{,}6|\varrho_0|\,|\varrho_0+k|\log h}{2\pi h(k-1)} \left(\int_{\sigma_0}^{\sigma_1} \left(\frac{x}{\exp(n/k)}\right)^{\sigma-1/2} d\sigma\right) X^{-0{,}109\cdot 0{,}83} \\ &\leqq \frac{2{,}6\cdot 2{,}01|\varrho_0|\,|\varrho_0+k|\log h}{2\pi h(k-1)} \left(\frac{X}{\exp(n/k)}\right)^{0{,}51} X^{-0{,}09047} \\ &\leqq 2X^{-0{,}00377}. \end{aligned} \tag{3.22}$$

Aus (3.15), (3.17), (3.20)–(3.22) erhalten wir schließlich

$$|R_1+R_2| \leqq 13{,}1\,X^{-0{,}0005} + 110\,X^{-0{,}16667} + 1200\,X^{-0{,}00377}.$$

Für $X \geqq 10^{2250}$ ist also

$$|R_1+R_2| \leqq 0{,}99.$$

Auf gleiche Weise wie im Beweis des Satzes aus [12] schließen wir unsere Überlegungen. □

4. Beweis von Satz 2

Der Beweis wird analog dem vorhergehenden verlaufen. Insbesondere führen wir wieder den Operator A ein [s. (3.1)], was verlangt, den Wert des Parameters k zu bestimmen. Zu diesem Zweck setzen wir

$$k = 2h, \tag{4.1}$$

wobei h noch folgendermaßen zu wählen ist:

h liegt im Intervall $(H-1{,}5,\ H-0{,}5)$ und genügt der Bedingung:

$$|h-\gamma| \geqq \frac{1}{2(N^*(H-1)+1)} \tag{4.2}$$

für jede nichttriviale Nullstelle $\varrho = \beta + i\gamma$ von ζ.

Die Bedeutung von $N^*(H-1)$ wurde in Hilfssatz 6 angegeben. Es ist

$$k \geqq 1000. \tag{4.3}$$

Ferner findet die zu (3.8) analoge „explizite Formel“

$$A_{n+1}(\Delta)(x) = \frac{-2x^{1/2}}{|\varrho_0|\,|\varrho_0+k|^{n+1}}\{\cos(\gamma_0 \log x - \alpha_1 - (n+1)\alpha_2) + R_1 + R_2 + R_3\} \tag{4.4}$$

statt, wobei wir die Werte $\sigma_0 = -1$, $\sigma_1 = 1{,}01$, die Bezeichnungen (3.5), (3.6) und (3.16), den Sinn von R_1, R_2 [s. (3.8)–(3.10)], und auch von R_{2j} [s. (3.17), (3.18)], mit den eben eingeführten k und h, beibehalten, und wo

$$|R_3| \leqq |\varrho_0|\,|\varrho_0+k|^{n+1} \sum_{\gamma_1 \leqq \gamma < h} \frac{1}{|\varrho|\,|\varrho+k|^{n+1}} \tag{4.5}$$

ist.

Für $X \geqq 2$ setzen wir

$$n = \left[\frac{5{,}76}{1+99/h}\log X\right] + 1$$

und beschränken den Bereich für x wie folgt:

$$X^{2,9/h} \leqq x \leqq X\,. \tag{4.6}$$

Wegen (4.1) und (4.3) können wir die im vorhergehenden Beweis angegebenen Abschätzungen und Rechnungen auswerten.

Aus (3.14) und (4.1) folgt insbesondere

$$\log\left|1+\frac{\varrho_0}{k}\right| \leqq \frac{0{,}5}{k} + \frac{100}{k^2} = \frac{0{,}25}{h} + \frac{25}{h^2} \leqq 0{,}0006\,. \tag{4.7}$$

Somit ist

$$\begin{aligned} |R_1| &\leqq \frac{\log(2\pi)}{2}|\varrho_0|\exp(0{,}0012) \\ &\quad \cdot \exp\left\{\left(\frac{-1{,}45}{h} + \left(\frac{0{,}25}{h}+\frac{25}{h^2}\right)\frac{5{,}76}{1+99/h}\right)\log X\right\} \\ &\leqq 13{,}1\,X^{-0{,}01/h+1{,}5/(h^2)}\,. \end{aligned} \tag{4.8}$$

Aus (3.19) und (4.7) folgt

$$\log\left|\frac{\varrho_0+k}{k-1}\right| \leqq \frac{0{,}75}{h} + \frac{25{,}2}{h^2} \leqq 0{,}002\,,$$

also

$$\begin{aligned} R_{21} &\leqq \frac{|\varrho_0|(\log h+8)(\log h+1)\exp(0{,}004)}{2\pi} \\ &\quad \cdot \exp\left\{\left(\frac{-1{,}5\cdot 2{,}9}{h} + \left(\frac{0{,}75}{h}+\frac{25{,}2}{h^2}\right)\frac{5{,}76}{1+99/h}\right)\log X\right\} \\ &\leqq 2{,}3(\log h+8)^2\,X^{-0{,}03/h}\,. \end{aligned} \tag{4.9}$$

Im folgenden benötigen wir die für $|\sigma| \leqq 0{,}002$ und $|t| \leqq 0{,}5$ geltende Ungleichung

$$\log|1+s| \geqq \sigma - \tfrac{1}{2}(\sigma^2 - t^2) + \tfrac{1}{3}(\sigma^3 - 3\sigma t^2) - 0{,}3\,|s|^4. \tag{4.10}$$

Unter Verwendung dieser Beziehung bekommt man

$$\begin{aligned}\log\left|1+\frac{\sigma_1+ih}{k}\right| &\geqq \frac{\sigma_1}{k} + \frac{1}{2k^2}(h^2-\sigma_1^2) + \frac{1}{3k^3}(\sigma_1^3 - 3\sigma_1 h^2)\\ &\quad - \frac{0{,}3}{k^4}(\sigma_1^2+h^2)^2\\ &\geqq \frac{0{,}7575}{k} + 0{,}1062.\end{aligned}$$

Hieraus und aus (4.7) folgt

$$\log\left|\frac{\varrho_0+k}{\sigma_1+ih+k}\right| \leqq -\frac{0{,}2525}{k} + \frac{100}{k^2} - 0{,}1062 \leqq -0{,}1062,$$

weil $k \geqq 100/0{,}2525$ ist. Somit ist

$$\begin{aligned}R_{22} &\leqq \frac{100|\varrho_0|\,|\varrho_0+k|\,(1+\log 2)}{2\pi k}\exp\left\{\log X\left(0{,}51 - \frac{0{,}1062\cdot 5{,}76}{1+99/h}\right)\right\}\\ &\leqq 390\,X^{-0{,}0006}.\end{aligned} \tag{4.11}$$

Für $\sigma_0 \leqq \sigma \leqq \sigma_1$ ist

$$\log\left|1+\frac{\sigma+ih}{k}\right| \geqq \frac{\sigma}{2h} - \frac{0{,}12625}{h} + 0{,}1062,$$

was zusammen mit (4.7) gibt

$$\log\left|\frac{\varrho_0+k}{\sigma+ih+k}\right| \leqq \frac{1/2-\sigma}{2h} + \frac{0{,}2}{h} - 0{,}1062.$$

Aus (4.2), Hilfssatz 5 und 6 folgt

$$\max_{\sigma_0 \leqq \sigma \leqq \sigma_1}\left|\frac{\zeta'}{\zeta}(\sigma+ih)\right| \leqq 2{,}3\log^2 H.$$

Für

$$X \geqq \exp(25) \tag{4.12}$$

gilt

$$X^{2{,}9/h} \geqq \exp(n/(2h)).$$

Nach diesen Vorbereitungen, (4.12) vorausgesetzt, erhalten wir

$$\begin{aligned}R_{23} &\leqq \frac{2{,}3\log^2 H\,|\varrho_0|\,|\varrho_0+k|}{2\pi h(k-1)}\int_{\sigma_0}^{\sigma_1}\left(\frac{x}{\exp(n/(2h))}\right)^{\sigma-1/2}d\sigma\\ &\quad\cdot\exp\left\{\log X\left(\frac{0{,}2}{h} - 0{,}1062\right)\frac{5{,}76}{1+99/h}\right\}\\ &\leqq X^{-0{,}0006}.\end{aligned} \tag{4.13}$$

Betrachten wir endlich R_3, s. (4.5). Zu diesem Zweck benötigen wir die folgende, für $\gamma_1 \leqq \gamma < h$ gleichmäßige Abschätzung:

$$\log\left|\frac{1+\varrho_0/k}{1+\varrho/k}\right| \leqq \frac{1}{2k^2}(\gamma_0^2-\gamma^2)+\frac{1}{2k^3}(\gamma^2-\gamma_0^2)$$
$$+\frac{0{,}3}{k^4}((1/4+\gamma^2)^2+(1/4+\gamma_0^2)^2)$$
$$\leqq -\frac{0{,}394}{2k^2}(\gamma^2+1/4),$$

die man aus (1.3), (1.4), (3.13) und (4.10) herleiten kann. Man hat auch [2, S. 159f.]

$$2\sum_{\gamma\geqq\gamma_1}\frac{\beta}{\beta^2+\gamma^2}\leqq 0{,}0182.$$

Also ist

$$|R_3| \leqq |\varrho_0/\varrho_1| \sum_{\gamma_1\leqq\gamma<h}\exp\left\{\frac{-0{,}394n}{8h^2}(\gamma^2+1/4)\right\}$$
$$\leqq \frac{8|\varrho_0|h^2}{0{,}394|\varrho_1|n}2\sum_{\gamma_1\leqq\gamma}\frac{\beta}{\beta^2+\gamma^2}$$
$$\leqq \frac{8(1+99/h)0{,}0182|\varrho_0|}{0{,}394\cdot 5{,}76|\varrho_1|}\frac{h^2}{\log X}\leqq\frac{0{,}06h^2}{\log X}. \tag{4.14}$$

Aus (3.17), (4.8), (4.9), (4.11), (4.13) und (4.14) folgt schließlich

$$|R_1+R_2+R_3| \leqq 13{,}1\,X^{-0{,}01/h+1{,}5/(h^2)}+2{,}3(\log h+8)^2X^{-0{,}03/h}$$
$$+400X^{-0{,}0006}+\frac{0{,}06h^2}{\log X}. \tag{4.15}$$

Setzen wir voraus

$$\log X \geqq 0{,}09\max\{4400, H\}H.$$

Dann ist die Bedingung (4.12) erfüllt, und aus (4.15) erhält man

$$|R_1+R_2+R_3|\leqq 0{,}97.$$

Dies hat zur Folge [s. (4.4) und (4.6)]

$$V(A_{n+1}(\Delta),X)\geqq\left(1-\frac{2{,}9}{h}\right)\frac{\gamma_0}{\pi}\log X-2\geqq\left(1-\frac{3}{H}\right)\frac{\gamma_0}{\pi}\log X.$$

Auf gleiche Weise wie im Beweis des Satzes aus [12] schließen wir unsere Überlegungen. □

Danksagungen. Artikel I und II dieser Reihe bilden eine verbesserte Version eines Teils meiner Doktorarbeit. Herrn Professor Włodzimierz Staś danke ich herzlich für wissenschaftliche Leitung. Mein Dank gilt auch Jerzy Kaczorowski für eine Reihe von wertvollen Bemerkungen. Dem anonymen Referenten, der zahlreiche Verbesserungen vorgeschlagen hat, bin ich besonders verpflichtet.

Literatur

1. Davenport, H.: Multiplicative number theory. 2nd. ed. (revised by H. L. Montgomery). Berlin Heidelberg New York: Springer 1980
2. Edwards, H.M.: Riemann's zeta function. New York London: Academic Press 1974
3. Ellison, W.J. (en collaboration avec M. Mendès France): Les nombres premiers. Paris: Hermann 1975
4. Gram, J.-P.: Sur les zéros de la fonction $\zeta(s)$ de Riemann. Acta Math. **27**, 289–304 (1903)
5. Handbook of mathematical functions. New York: Dover Publications 1972
6. Handbook of tables for mathematics. Revised 4th ed. Cleveland, Ohio: CRS-Press 1975
7. Kaczorowski, J.: On sign-changes in the reminder-term of the prime-number formula. I. II. Acta Arith. **44**, 365–377 (1984), ibid. **45**, 65–74 (1985)
8. Kaczorowski, J., Pintz, J.: Oscillatory properties of arithmetical functions. I. Acta Math. Hung. **48** (1–2), 173–185 (1986)
9. van de Lune, J., te Riele, H.J.J.: On the zeros of the Riemann zeta function in the critical strip. III. Math. Comput. **41**, 759–767 (1983)
10. Nielsen, N.: Handbuch der Theorie der Gammafunktion. Leipzig: Teubner 1906
11. Siegel, C.L.: Über Riemanns Nachlaß zur analytischen Zahlentheorie. Quellen und Studien zur Geschichte der Mathematik, Astronomie und Physik, Abteilung B: Studien **2**, 45–80 (1932)
12. Szydło, B.: Über Vorzeichenwechsel einiger arithmetischer Funktionen. I. Math. Ann. **283**, 139–149 (1988)
13. Titchmarsh, E.C.: The theory of the Riemann zeta function. Oxford: Clarendon Press 1951

Eingegangen am 29. Oktober 1987; revidierte Fassung am 31. Mai 1988

Math. Ann. 283, 165–170 (1989)

Mathematische Annalen

Mazur's Intersection Property for Finite Dimensional Sets

Abderrazzak Sersouri

Equipe d'Analyse, Université Paris 6, 4, Place Jussieu,
F-75252 Paris Cedex 05, France

Introduction

In this paper we consider only real Banach spaces. This is not a restriction since the properties we consider in this paper depend only on the real structure of the space.

In Theorem 1 we give dual characterizations for the properties (I_n) defined by:

> Every convex compact set with *affine* dimension
> at most n, is an intersection of balls.

This will be used to characterize the property:

$$\left.\begin{array}{l}\text{Every finite dimensional convex compact set}\\ \text{is an intersection of balls.}\end{array}\right\}(\mathrm{I}_{f,d})$$

We also prove that for n-dimensional Banach spaces, property (I_{n-1}) implies property (I_n), and examples are constructed to show that this result is the best possible.

At the end we give (without proofs) some stability results for property $(\mathrm{I}_{f,d})$, and we ask whether every Banach space can be renormed to have property (I_1). We also mention an application of Theorem 1 to spaces of compact operators.

Notation

A point x of a Banach space X is said to be an extreme point if $x=0$ or if $x/\|x\|$ is an extreme point of $B(X)$, the unit ball of X. The set of extreme points of X will be denoted by $\mathrm{Ext}(X)$.

For a finite dimensional set C, $\dim C$ will always mean the *affine* dimension of C.

A slice of a bounded set C (in some Banach space X) is a subset of C of the form:

$$S(C,f,\delta)=\left\{x\in C: f(x)>\left(\sup_C f\right)-\delta\right\}$$

for some $f\in X^*$, $\delta>0$.

The closed ball [resp. open ball, resp. sphere] centered at x and with radius r will be denoted by $B(x,r)$ [resp. $\mathring{B}(x,r)$; resp. $S(x;r)$].

Results

Our main result is the following theorem, which is analogous in its spirit to Theorem 1 of [4].

Theorem 1. *For every Banach space X, and every natural number n, the following properties are equivalent:*

(1) *Every compact convex subset C of X with* $\dim C \leqq n$, *is an intersection of balls.*

(2) *For every $f \in X^*$, every $(n+1)$-points $(x_i)_{0 \leqq i \leqq n} \in X$, and every $\varepsilon > 0$, there exists $g \in \operatorname{Ext}(X^*)$ such that $\sup\limits_{0 \leqq i \leqq n} |x_i(f-g)| < \varepsilon$.*

This theorem will be an immediate consequence of the more precise result stated in the next lemma. But we need first to introduce some notation.

For a bounded convex set C, let $\hat{C}$ denote the intersection of the balls containing C, and define $\varrho(C) = \sup\limits_{x \in \hat{C}} \operatorname{dist}(x, C)$.

If X is a Banach space, and n an integer, let $\Lambda = \Lambda_n = \sup\{\varrho(C) : C \subset B(X), \dim C \leqq n\}$. [Hence X has (I_n) if and only if $\Lambda_n = 0$.] We also define

$$\lambda = \lambda_n = \inf \Big\{ \mu > 0 \text{ such that} : \forall f \in S(X^*), \forall C \subset B(X),\ \dim C \leqq n,\ \exists g \in \operatorname{Ext}(X^*) : \sup_C |f-g| \leqq \mu \Big\}.$$

With these notations, a "quantitative version" of Theorem 1 is given by:

Lemma 2. $\dfrac{\lambda}{2} \leqq \Lambda \leqq 2\lambda$.

Proof of $\lambda \leqq 2\Lambda$. It is enough to prove that given $C \subset B(X)$, $\dim C \leqq n$, and $f \in S(X^*)$, there exists for every $\eta > 0$, an element $g \in \operatorname{Ext}(X^*)$ such that $\sup\limits_C |f-g| \leqq 2(\Lambda + \eta)$.

Let C, f, and η be as before.

If $m = \sup\limits_C |f| \leqq 2(\Lambda + \eta)$, take $g = 0$.

If $m \geqq 2(\Lambda + \eta)$, choose $u_0 \in K = cv(\pm C)$ such that $f(u_0) = m$, and let $u = \dfrac{\Lambda + \eta}{m} u_0$. Define also $D = K \cap \ker f$.

It is clear that $D \subset B(X)$, $\dim D \leqq n$, and $\operatorname{dist}(u, D) \geqq \Lambda + \eta$. By definition of Λ, there exists a Ball $B(z, r)$ containing D and not containing u.

Let w be the unique element of $S(z, r) \cap cv[z, u]$, x the norm one vector $\dfrac{w-z}{r}$, and choose an extreme point h of $B(X^*)$ such that $h(x) = 1$.

One can easily check that $0 \leqq \sup\limits_D h \leqq \sup\limits_{B(z,r)} h < h(u)$. So there exists $\alpha > 0$, such that $\sup\limits_K \alpha h = 1$, and from the above inequalities we deduce that $\sup\limits_D \alpha h < \alpha h(u) \leqq \dfrac{\Lambda + \eta}{m}$.

Then, by Phelps' lemma [applied to the Banach space $sp(K)$ with K as a unit ball], we obtain $\left\|\frac{f}{m}\pm\alpha h\right\|_K\leqq 2\,\frac{\Lambda+\eta}{m}$. This concludes the proof of "$\lambda\leqq 2\Lambda$" since both $\pm\alpha mh$ are in $\mathrm{Ext}(X^*)$.

Proof of $\Lambda\leqq 2\lambda$. It is enough to prove that given $C\subset B(X)$, $\dim C\leqq n$, then $x\in X$ can be separated from C by a ball whenever $\mathrm{dist}(x, C)>2\lambda$.

Let C and x be as above. We can suppose that $x\in B(X)$, since if not $B(X)$ separates C and x.

Let $K=\dfrac{C-x}{2}\subset B(X)$, and observe that to separate x and C, it is enough to separate 0 and K.

Since $\mathrm{dist}(x, C)>2\lambda$, we have $\mathrm{dist}(0, K)>\lambda$, which means that $K\cap(\lambda B(X))=\emptyset$. This implies that we can find $f\in S(X^*)$ such that $\inf\limits_K f>\lambda$, and by the definition of λ, we can find $g\in\mathrm{Ext}(B(X^*))$ such that $\inf\limits_K g=3\varepsilon>0$.

Since g is an extreme point of $B(X^*)$, by a well known result due to Choquet, we can find $x\in S(X)$, $\delta>0$ such that:

$$g\in S(B(X^*);x,\delta)\subset\left\{h\in B(X^*):\sup_K|g-h|<\varepsilon\right\}.$$

We are going to prove that there exists an $r>0$ such that K is included in $D_r=B(r\varepsilon x,(r-1)\varepsilon)$. This will conclude the proof, since none of the balls D_r contain 0.

Indeed, if not, by a compacity argument, and since the balls D_r are increasing (with r), the set $L=\bigcap\limits_{r>0}(K\backslash\mathring{D}_r)$ will be non empty.

Take an element $y\in L$, and for every $r>0$, let $g_r\in S(X^*)$ be such that $g_r(r\varepsilon x-y)=\|r\varepsilon x-y\|\geqq(r-1)\varepsilon$. This inequality implies easily that $\lim\limits_{r=\infty}g_r(x)=1$, and so $g_r\in S(B(X^*),x,\delta)$ for r large enough.

On the other hand, it is not difficult to see that $(g-g_r)(y)\geqq 2\varepsilon$, and so $\sup\limits_K|g-g_r|\geqq 2\varepsilon$, for every $r>0$. This conclusion contradicts the preceding one by the choice of x and δ.

This completes the proof of the lemma. □

That Lemma 2 implies Theorem 1 is an immediate consequence of the following geometrical observation: If C is a compact convex set with $\dim C\leqq n$, there exists $(n-1)$-points $(x_i)_{0\leqq i\leqq n}$ such that $C\subset K=cv\{x_i:0\leqq i\leqq n\}$ (and K also satisfies $\dim K\leqq n$).

In the sequel we will list some consequences of Theorem 1.

Corollary 3. *For every Banach space, the following properties are equivalent:*

(1) *Every finite dimensional compact convex subset of X is an intersection of balls.*

(2) *The set* $\mathrm{Ext}(X^*)$, *of extreme points of* X^*, *is* w^**-dense in* X^*.

Proposition 4. *Let X be a Banach space such that for every $n\geqq 1$, X has an equivalent norm $|\cdot|_n$ which satisfies property* (I_n). *Then X has a equivalent norm $[\![\cdot]\!]$ which satisfies property* $(\mathrm{I}_{f,d})$.

Proof. Let us denote by $\|\cdot\|$ the original norm of X. It is easy to see that we can suppose that $|\cdot|_n \geqq \|\cdot\|$ for every n, and hence $|\cdot|_n^* \leqq \|\cdot\|^*$ (on X^*).

Define on X^* an equivalent dual norm by:

$$[\![x^*]\!] = \left(\sum_{n=1}^{\infty} \frac{1}{2^n} |x^*|_n^{*2}\right)^{1/2}$$

and let us prove that its predual norm works.

An easy (and standard) convexity argument shows that $\operatorname{Ext}(X^*_{[\![\cdot]\!]}) \supset \bigcup_{n\geqq 1} \operatorname{Ext}(X^*_{|\cdot|_n})$. This implies by Theorem 1 that $\operatorname{Ext}(X^*_{[\![\cdot]\!]})$ is w^*-dense in X^*, and the conclusion follows by Corollary 3. □

Proposition 5. *Let E be an n-dimensional Banach space with property* (I_{n-1}). *Then E has property* (I_n).

Proof. We will prove that under the hypothesis of Proposition 5, the set of extreme points of $B(E^*)$ is norm dense in $S(E^*)$. This clearly implies the conclusion in view of Theorem 1 (see also [1, 3]).

Let $(e_i)_{1\leqq i\leqq n}$ be a basis of E, then for every $f\in S(E^*)$, and every $\varepsilon>0$, there exists $\delta>0$, such that for every $g\in E^*$, $\|f-g\|<\varepsilon$ whenever $\sup_{1\leqq i\leqq n} |e_i(f-g)|<\delta$.

Assuming (I_{n-1}), for every $f\in S(E^*)$, we can find $g\in \operatorname{Ext}(E^*)$ such that $\sup_{1\leqq i\leqq n} |e_i(f-g)|<\delta$, hence $\|f-g\|<\varepsilon$, and then $\left\|f-\frac{g}{\|g\|}\right\|<2\varepsilon$. □

The result of Proposition 5 cannot be improved as it is shown by the following:

Proposition 6. *For every natural numbers k and n such that $n\geqq k+2$, there exists on $\mathbb{R}^n$ an equivalent norm $|\cdot|_{n,k}$ satisfying* (I_k) *but not* (I_{k+1}).

Remark. The norm $|\cdot|_{n,k}$ (we will define) is given by $|x|_{n,k}=\left(\sum_{i=1}^{k+1} |\tilde{x}_i|^2\right)^{1/2}$, where $(\tilde{x}_i)_{1\leqq i\leqq n}$ is the decreasing rearrangement of $(|x_i|)_{1\leqq i\leqq n}$. But this formula is of no help in proving the proposition.

Proof of Proposition 6. Let n and k be fixed natural numbers such that $n\geqq k+2$.

We need first to introduce some notation and to prove a preliminary result.

Let $(e_i)_{1\leqq i\leqq n}$ be the natural basis of $\mathbb{R}^n$, $\|\cdot\|$ be the Euclidean norm, $\langle\cdot,\cdot\rangle$ the Euclidean scalar product, and S the Euclidean unit sphere.

Let $\mathscr{P}=\{A\subset[1,n]:\operatorname{card}A=k+1\}$, and for every $A\in\mathscr{P}$ define $H_A=sp[e_i: i\in A]$. Define also the sets $\mathscr{E}=\bigcup_{A\in\mathscr{P}}(S\cap H_A)$, and $C=cv(\mathscr{E})$, and let us prove that $\operatorname{Ext}(C)=\mathscr{E}$.

By the Krein-Milman theorem, since $\mathscr{E}$ is closed, it is enough to prove that $\mathscr{E}\subset\operatorname{Ext}(C)$, and to do this, it is again enough to prove that if $x, x_1, \ldots, x_p\in\mathscr{E}$, $\lambda_1,\ldots,\lambda_p\in\mathbb{R}^+$ are such that $x=\sum_{i=1}^p \lambda_i x_i$ and $\sum_{i=1}^p \lambda_i=1$, then $x=x_1=\ldots=x_p$.

Let us prove that the above statement is true. Let $A\in\mathscr{P}$ be such that $x\in H_A$, and denote by P_A the orthogonal projection on H_A.

From $x=\sum_{i=1}^{p}\lambda_i x_i$ we deduce that $x=\sum_{i=1}^{p}\lambda_i P_A(x_i)$ which implies that $x=P_A(x_1)=\ldots=P_A(x_p)$ by the properties of the Euclidean norm.

In particular we have $\|P_A(x_i)\|=1$, for every i, and so $P_A(x_i)=x_i$ (since $\|x_i\|=1$). This proves that $x=x_i=\ldots=x_p$, and concludes the proof of $\mathscr{E}=\mathrm{Ext}(C)$.

Let us return now to the proof of Proposition 6. Since C is convex, symmetric, closed, and with no empty interior, C defined a (dual) norm on $\mathbb{R}^n$, the (pre-) dual of which we will denote by $|\cdot|_{n,k}=|\cdot|$, i.e., $B(\mathbb{R}^n_{|\cdot|})=C^0$ (*the polar set of* C).

But what we have proved in the preliminary part we have that

$$\mathrm{Ext}(\mathbb{R}^n_{|\cdot|*})=\mathbb{R}\cdot\mathscr{E}=\bigcup_{A\in\mathscr{P}} H_A .$$

Let now $f\in\mathbb{R}^n$, $(x_i)_{0\leqq i\leqq k}\in\mathbb{R}^n$, and suppose that $(x_i)_{0\leqq i\leqq l}$ is a *maximal* linearly independent subfamily of $(x_i)_{0\leqq i\leqq k}$. Then choose an $A\in\mathscr{P}$ such that $((x_i)_{0\leqq i\leqq l};(e_j)_{j\notin A})$ is still linearly independent, and find $g\in\mathbb{R}^n$ such that $\langle g,x_i\rangle=\langle f,x_i\rangle$ for $0\leqq i\leqq l$, and $\langle g,e_j\rangle=0$ for every $j\notin A$.

Such a g is in H_A, hence $g\in\mathrm{Ext}(\mathbb{R}^n_{|\cdot|*})$, and also is such that $\sup_{0\leqq i\leqq k}|x_i(f-g)|=0$ (by the "maximality" of the chosen subfamily).

Theorem 1 implies then that $\mathbb{R}^n_{|\cdot|}$ has (I_k).

On the other hand let $f\in\mathbb{R}^n$ be such that $\langle f,e_i\rangle=1$ for $1\leqq i\leqq k+2$. Then there is no element $g\in\mathrm{Ext}(\mathbb{R}^n_{|\cdot|*})$ such that $\sup_{1\leqq i\leqq k+2}|e_i(f-g)|<1$.

Indeed for every $A\in\mathscr{P}$, the set $[1,n]\backslash A$ intersects $[1,k+2]$ (cardinality argument), then for every $g\in H_A$, we have $\sup_{1\leqq i\leqq k+2}|e_i(f-g)|\geqq 1$.

Theorem 1 again implies that $\mathbb{R}^n_{|\cdot|}$ fails (I_{k+1}). □

Remark. Using the same proofs* as in [4], one can obtain the following results:

1) If $T:X\to Y$ is such that T and T^* are injective, and if Y has an equivalent $(\mathrm{I}_{f,d})$-norm then X has also an equivalent $(\mathrm{I}_{f,d})$-norm.

2) Every Banach space has an equivalent $(\mathrm{I}_{f,d})$-norm if and only if the above result is true without the hypothesis "T^* injective".

3) If $(P_\alpha)_{0\leqq\alpha\leqq\mu}$ is a Schauder decomposition for the Banach space X, such that for every α, $0\leqq\alpha\leqq\mu$, the space $(P_{\alpha+1}-P_\alpha)(X)$ has an equivalent $(\mathrm{I}_{f,d})$-norm, then X has an equivalent $(\mathrm{I}_{f,d})$-norm.

Using the Hahn-Banach theorem, one can easily see that for every Banach space X, the set $\mathrm{Ext}(X^*)$ of extreme points of X^* intersects all the *affine*, w^*-closed, 1-codimensional subspaces of X^*.

In view of this one can ask the following:

Problem. Does every Banach space X have an equivalent norm such that $\mathrm{Ext}(X^*)$ intersects all the *affine*, w^*-closed, 2-codimensional subspaces of X^*?

A positive answer to this will imply that every Banach space has an equivalent (I_1)-norm. Up to now it is unknown if this conclusion holds even for Asplund spaces.

---

* We cannot reproduce these proofs because of their length

Remark. In [5], Theorem 1 is used to prove that the spaces $K(X, Y)$ and $X \underset{\varepsilon}{\otimes} Y$ (with their usual norms) never have property (I_2) if $\dim X \geqq 2$ and $\dim Y \geqq 2$, and Theorem 1 is also used to show that the space $K(l_2^2) = l_2^2 \underset{\varepsilon}{\otimes} l_2^2$ has property (I_1).

References

1. Giles, J., Gregory, D., Sims, B.: Characterisation of normed linear space with Mazur's intersection property. Bull. Austr. Math. Soc. **18**, 105–123 (1978)
2. Mazur, S.: Über schwache Konvergenz in den Räumen (L^p). Stud. Math. **4**, 128–133 (1933)
3. Phelps, R.: A representation theorem for bounded convex sets. Proc. Am. Math. Soc. **11**, 976–983 (1960)
4. Sersouri, A.: Mazur property for compact sets. Pac. J. Math. **113**, 185–195 (1988)
5. Sersouri, A.: Smoothness in spaces of compact operators. Bull. Austr. Math. Soc. **37**, 221–225 (1988)
6. Zizler, V.: Renorming concerning Mazur's intersection of balls for weakly compact sets. Math. Ann. **276**, 61–66 (1986)

Received November 16, 1987; in revised form June 22, 1988

Math. Ann. 283, 171–176 (1989)

Mathematische Annalen

Stabilité du fibré tangent des surfaces de Del Pezzo

Rachid Fahlaoui

Mathématiques, Université Paris-Sud, Bâtiment 425, F-91405 Orsay Cedex, France

Introduction

On sait que sur une surface complexe compacte S dont le fibré canonique est positif (resp. négatif), l'existence d'une métrique de Kähler-Einstein entraîne la semi-stabilité du fibré tangent de S par rapport à la polarisation canonique (resp. anticanonique). Notons S_r la surface obtenue à partir de $\mathbb{P}^2$ en éclatant r points en position générale. Il semble probable que pour $3 \leqq r \leqq 8$, la surface S_r admette une métrique de Kähler-Einstein (ce qui entraînerait la semi-stabilité du fibré tangent de S_r). C'est dans cette optique que nous proposons dans cet article une démonstration algébrique (valable en caractéristique quelconque) de la semi-stabilité du fibré tangent des surfaces à fibré canonique négatif.

Enfin je voudrais remercier mon professeur A. Beauville pour ses remarques et ses suggestions qui m'ont aidé à simplifier énormément les calculs, ainsi que pour son aide considérable à la rédaction de cet article.

Notations

Nous considérons des surfaces projectives et lisses sur un corps algébriquement clos K. Soient S une telle surface, D et D' deux diviseurs sur S; on note

$\langle D, D' \rangle$ le produit d'intersection de D et D';

$D \equiv D'$ si D et D' sont linéairement équivalents;

K_S ou K un diviseur canonique, c'est-à-dire un diviseur sur S tel que $O_S(K) = \Omega_S^2$.

$H^i(S, O_S(D))$ ou $H^i(O_S(D))$ les espaces de cohomologie du faisceau $O_S(D)$.

$h^i(D)$ la dimension du K-espace vectoriel $H^i(S, O_S(D))$.

1. Définitions et généralités

Soient $p_1, \ldots, p_r$ $(r \leqq 8)$ des points de $\mathbb{P}^2$ en position générale: cela signifie qu'il n'existe pas de droite qui contienne trois d'entre eux, ni de conique qui contienne six d'entre eux, ni (si $r = 8$) de cubique passant par 7 de ces points et ayant un point double au huitième.

On note S_r l'éclaté des points $p_1, \dots, p_r$ dans $\mathbb{P}^2$ et E_i la droite exceptionnelle image réciproque de p_i. On désigne par E_0 l'image réciproque d'une droite de $\mathbb{P}^2$. Soit e_i $(0 \leqq i \leqq r)$ la classe du diviseur E_i dans $\mathrm{Pic}(S_r)$; le groupe $\mathrm{Pic}(S_r)$ est engendré par $e_0, \dots, e_r$, et la forme d'intersection est donnée par:

$$\langle e_0, e_0 \rangle = 1, \quad \langle e_i, e_i \rangle = -1 \quad \text{pour} \quad i > 0, \quad \langle e_i, e_j \rangle = 0 \quad \text{pour} \quad i \neq j.$$

La classe dans $\mathrm{Pic}(S_r)$ du diviseur canonique est $-3e_0 + \sum_{i=1}^{r} e_i$.

Soit Φ_{-K} l'application rationnelle définie par les sections globales du fibré anticanonique $O_S(-K)$. Pour $0 \leqq r \leqq 6$, Φ_{-K} est un plongement de S_r dans $\mathbb{P}^{9-r}$; Φ_{-2K} est un plongement de S_7 dans $\mathbb{P}^6$, et Φ_{-3K} un plongement de S_8 dans $\mathbb{P}^8$ [D].

Soient S une surface algébrique, E un faisceau localement libre de rang 2 sur S et H la classe dans le groupe de Néron-Severi de S d'un diviseur ample (une telle classe est appelée une *polarisation* sur S).

Définition. On dit que E est stable (resp. semi-stable) par rapport à H si pour tout sous-faisceau localement libre L de rang 1 de E on a

$$\langle L, H \rangle < \tfrac{1}{2} \langle \Lambda^2 E, H \rangle \qquad (\text{resp. } \langle L, H \rangle \leqq \tfrac{1}{2} \langle \Lambda^2 E, H \rangle).$$

Dans la suite, on s'intéressera à la stabilité du fibré tangent $\Omega^1_{S_r}$ par rapport à la polarisation $-K$. Soit L un sous-faisceau de $\Omega^1_{S_r}$; on a donc une section globale s de $\Omega^1_{S_r} \otimes L^{-1}$. Soit Y le schéma des zéros de s; il définit un cycle de dimension $\leqq 1$ qui s'écrit $\sum n_i V_i + \sum m_j P_j$, où les V_i sont des courbes irréductibles et les P_j des points. On pose $(s)_1 = \sum n_i V_i$ et $(s)_0 = \sum m_j P_j$. Soit M le faisceau inversible associé à $(s)_1$; le faisceau $\Omega^1_{S_r} \otimes L^{-1} \otimes M^{-1}$ possède une section s' dont le schéma des zéros Z définit le cycle $(s)_0$. On a alors une suite exacte [R]

$$0 \to O_{S_r} \to \Omega^1_{S_r} \otimes L^{-1} \otimes M^{-1} \to IN \to 0,$$

où N est le faisceau $\Lambda^2(\Omega^1_{S_r} \otimes L^{-1} \otimes M^{-1})$ et I l'idéal du schéma Z. Observons qu'on a

$$c_2(\Omega^1_{S_r} \otimes L^{-1} \otimes M^{-1}) = \deg(Z) = \sum m_j.$$

Si $M = O_S$, on dit que le sous-faisceau L de $\Omega^1_{S_r}$ est saturé. Comme $\langle L, H \rangle \leqq \langle L \otimes M, H \rangle$ pour toute polarisation H, il suffit pour prouver la (semi-)stabilité de $\Omega^1_{S_r}$ de considérer les sous-faisceaux *saturés* de $\Omega^1_{S_r}$.

Lemme 1. *Soient L, M deux sous-faisceaux saturés de $\Omega^1_{S_r}$ qui ne sont pas isomorphes. On a alors $h^0(\omega_{S_r} \otimes L^{-1} \otimes M^{-1}) \geqq 1$.*

Démonstration. D'après ce qui précède, on a une suite exacte

$$0 \to O_{S_r} \to \Omega^1_{S_r} \otimes L^{-1} \to I\omega_{S_r} \otimes L^{-2} \to 0.$$

On en déduit la suite exacte

$$0 \to L \otimes M^{-1} \to \Omega^1_{S_r} \otimes M^{-1} \to I\omega_{S_r} \otimes L^{-1} \otimes M^{-1} \to 0,$$

d'où une suite exacte longue de cohomologie

$$0 \to H^0(L \otimes M^{-1}) \to H^0(\Omega^1_{S_r} \otimes M^{-1}) \to H^0(I\omega_{S_r} \otimes L^{-1} \otimes M^{-1})$$
$$\to H^1(L \otimes M^{-1}) \to \dots.$$

On remarque que l'un des deux espaces $H^0(L \otimes M^{-1})$ ou $H^0(M \otimes L^{-1})$ est nul, sans quoi on aurait $L=M$ contrairement à l'hypothèse. Supposons par exemple $H^0(L \otimes M^{-1})=0$; on en déduit $h^0(I\omega_{S_r} \otimes L^{-1} \otimes M^{-1}) \geqq 1$, ce qui achève la démonstration.

2. Exemples de sous-faisceaux de $\Omega^1_{S_r}$

Les sous-faisceaux inversibles saturés de $\Omega^1_{S_r}$ correspondent biunivoquement aux 1-formes rationnelles sur $\mathbb{P}^2$ qui s'annulent en codimension 2 seulement. Cette correspondance est définie de la façon suivante. Soit ω une section globale de $\Omega^1_{\mathbb{P}^2}(k)$, avec $k \geqq 2$. Si Z désigne le schéma des zéros de ω, on a une suite exacte

$$0 \to O_{\mathbb{P}^2} \to \Omega^1_{\mathbb{P}^2}(k) \to I_Z O_{\mathbb{P}^2}(2k-3) \to 0 .$$

Soient (X, Y, T) des coordonnées homogènes dans $\mathbb{P}^2$; la forme ω s'écrit $PdX+QdY+RdT$, où P, Q, R sont des polynômes homogènes de degré $(k-1)$ vérifiant $XP+YQ+TR=0$.

Identifions les sections globales de $\Omega^1_{\mathbb{P}^2}(k)$ aux 1-formes rationnelles sur $\mathbb{P}^2$ ayant un pôle seulement le long de la droite $T=0$. Notons $\pi: \hat{S} \to \mathbb{P}^2$ l'éclatement de $\mathbb{P}^2$ au point $p=(0,0,1)$ et E la droite exceptionnelle $\pi^{-1}(p)$. Alors $\pi^*\omega$ est une 1-forme rationnelle sur $\hat{S}$ et $O_{\hat{S}}((\pi^*\omega)_1)$ est un sous-faisceau de $\Omega^1_{\hat{S}}$. Il est clair qu'en dehors de E, les zéros ou les pôles de $\pi^*\omega$ correspondent via π à ceux de ω; on a donc $(\pi^*\omega)_1 = \pi^*((\omega)_1)+lE$, où l est un entier que nous allons calculer. Soit U l'ouvert $T \neq 0$ dans $\mathbb{P}^2$; prenons les coordonnées locales (x, y) dans U définies par $x=X/T$, $y=Y/T$. Soit $\hat{U} \subset U \times \mathbb{P}^1$ la sous-variété d'équation $xY'-yX'=0$ (X', Y' étant des coordonnées locales dans $\mathbb{P}^1$). Au voisinage du point (p, ∞) de $\hat{U}$, on peut prendre x et $t=Y'/X'$ comme coordonnées locales; on a

$$\pi^*\omega|_{\hat{U}} = [P(x,tx,1)+tQ(x,tx,1)]\,dx + xQ(x,tx,1)\,dt .$$

Ecrivons $P(x,y,1)=p_m(x,y)+p_{m+1}(x,y)+\dots$, où les p_j sont des polynômes homogènes de degré j en x, y, avec $p_m \neq 0$. L'entier m est par définition la multiplicité de la courbe d'équation $P=0$ au point p.

Ecrivons de même $Q(x,y,1)=q_n(x,y)+q_{n+1}(x,y)+\dots$. On a alors

$$\begin{aligned}\pi^*\omega|_{\hat{U}} = [&x^m(p_m(1,t)+xp_{m+1}(1,t)+\dots)+tx^n(q_n(1,t)\\ &+xq_{n+1}(1,t)+\dots)]\,dx\\ &+x^{n+1}(q_n(1,t)+xq_{n+1}(1,t)+\dots)\,dt .\end{aligned}$$

Il en résulte qu'on a $l=\inf(m,n)$ ou $l=\inf(m,n)+1$, le second cas ne pouvant se produire que si on a $m=n$ et $p_m(1,t)+tq_m(1,t)=0$, c'est-à-dire $xp_m(x,y)+yq_m(x,y)=0$. Compte tenu de la relation $XP+YQ+TR=0$, cette dernière inégalité signifie que R s'annule en p avec multiplicité $\geqq m+2$.

Exemples. 1) $\omega = X\,dY - Y\,dX$.

On a $(\pi^*\omega)_1 = -2E_0+2E$.

2) $\omega=(Y^2T-T^2Y)\,dX+(T^2X-X^2T)\,dY+(X^2Y-Y^2X)\,dT$.

Prenons $p_1=(1,0,0)$, $p_2=(0,1,0)$, $p_3=(0,0,1)$, $p_4=(1,1,1)$. L'argument ci-dessus montre que $(\pi^*\omega)_1$ contient les droites exceptionnelles E_1, E_2, E_3, E_4 avec

multiplicité 2. On a donc
$$(\pi^*\omega)_1 = -4E_0 + 2\sum_{i=1}^{4} E_i,$$

et toute section de $\Omega^1_{\mathbb{P}^2}(4)$ vérifiant cette propriété est proportionnelle à ω. Les zéros de ω sont $p_1, \ldots, p_4$ et les points $(0,1,1)$, $(1,0,1)$ et $(1,1,0)$. Comme il y a au plus 4 de ces points qui soient en position générale, on en conclut qu'il n'existe pas de sous-faisceau de $\Omega^1_{S_r}$ de la forme
$$O\left(-4E_0 + 2\sum_{j=1}^{4} E_j + E_i\right) \quad \text{pour} \quad 1 \leqq i \leqq r.$$

3) Considérons le cas $k=5$. Prenons $p_1=(1,0,0)$, $p_2=(0,1,0)$; si $\pi^*\omega$ s'annule sur E_1 et E_2 avec multiplicité $\geqq 2$, P s'annule triplement en p_1 et Q en p_2; comme T divise $XP+YQ$, on en déduit que T divise P (et Q). Ainsi si $(\pi^*\omega)_1 = -5E_0 + 2E_1 + \ldots + 2E_s$ $(s \leqq r)$, chacune des droites $\langle p_1, p_1 \rangle$, pour $2 \leqq i \leqq s$, est contenue dans la courbe $P=0$. Les points p_i étant en position générale, on conclut qu'on a $s \leqq 5$: *il n'existe pas de sous-faisceau de $\Omega^1_{S_r}$ de la forme $O(-5E_0 + 2E_1 + \ldots + 2E_6)$.*

3. Stabilité du fibré tangent

Théorème. *Soit S une surface dont le fibré anticanonique est ample. Alors le fibré tangent de S est semi-stable par rapport à la polarisation $-K$. Si S n'est pas isomorphe à $\mathbb{P}^1 \times \mathbb{P}^1$ ni à S_1, son fibré tangent est stable par rapport à $-K$.*

Toute surface dont le fibré anticanonique est ample est isomorphe à $\mathbb{P}^1 \times \mathbb{P}^1$ ou à une surface S_r $(0 \leqq r \leqq 8)$. Les cas $S=\mathbb{P}^2$ et $S=\mathbb{P}^1 \times \mathbb{P}^1$ sont immédiats; nous considérons désormais la surface S_r pour $1 \leqq r \leqq 8$.

Soit $L_a = O_{S_r}\left(-kE_0 + \sum_{i=1}^{r} a_i E_i\right)$ un sous-faisceau saturé de $\Omega^1_{S_r}$, avec $k>2$. En utilisant le lemme 1 du par. 1 et l'exemple 1 du par. 2, on voit qu'il existe un diviseur effectif C_a tel que
$$C_a \equiv (k-1)E_0 - (a_1+1)E_1 - \sum_{i=2}^{r} (a_i - 1)E_i.$$
Posons $d = \langle -K, C_a \rangle = 3k - \sum a_i + r - 5$. On a alors $d>0$. Observons que l'inégalité $\langle -K, L_a \rangle < \frac{1}{2}\langle -K, K \rangle$ s'écrit
$$2\sum a_i < 6k + r - 9, \tag{2}$$
soit encore $2d \geqq r$. Nous supposerons donc dans la suite $d < r/2$.

1) Cas $r \leqq 6$

Dans ce cas, Φ_{-K} envoie C_a sur une courbe de degré d dans $\mathbb{P}^{9-r}$, et on a $d \leqq 2$. On a alors $\langle C_a, E_1 \rangle \leqq 2$, d'où $a_1 \leqq 1$; comme les points $p_1, \ldots, p_r$ jouent un rôle symétrique, on en déduit $a_i \leqq 1$ pour tout i. Si $k>2$ cela donne $d \geqq 4$, ce qui contredit l'hypothèse. Pour $k=2$ les seuls sous-faisceaux saturés de $\Omega^1_{S_r}$ sont $O_{S_r}(-2E_0 + 2E_1)$ et $O_{S_r}(-2E_0)$; ceux-ci vérifient (2), sauf le premier dans le cas $r=1$, pour lequel on a egalité dans (2). Ceci démontre le théorème pour $r \leqq 6$.

2) Cas $r=7$

On peut supposer ici $d \leqq 3$. D'autre part puisque L_a est saturé on a $c_2(\Omega^1_{S_r} \otimes L_a^{-1}) \geqq 0$, c'est-à-dire

$$\sum_{i=1}^{r} (a_i - a_i^2) + k^2 - 3k + 10 \geqq 0 .$$

En majorant $\sum a_i^2$ par $\frac{1}{r}(\sum a_i)^2$, on obtient

$$(3k-d+2)^2 \leqq 7(k^2-d+12) . \tag{3}$$

Par ailleurs, sur la surface S_6 obtenue en éclatant $p_1, \ldots, p_6$ dans $\mathbb{P}^2$, le diviseur $(k-1)E_0 - (a_1+1)E_1 - \sum_{i=2}^{6} (a_i-1)E_i$ est effectif, et son image par le plongement Φ_{-K} est de degré $d+a_7-1$. Prenant l'intersection avec E_1 on obtient

$$a_1 + 1 \leqq d + a_7 - 1 . \tag{4}$$

Si $d=1$ on en déduit $a_1 \leqq a_7 - 1$, ce qui est impossible car les a_i jouent un rôle symétrique.

Si $d=2$ on en déduit $a_1 = a_7$, et par suite $a_1 = a_2 = \ldots = a_7$. Mais on a alors $3k = 7a_1$ d'où $k \geqq 7$, ce qui contredit (3).

Supposons enfin $d=3$. Par (4) on obtient $a_i \leqq a_j + 1$ quels que soient i et j. D'autre part on a $a_i \leqq k-2$, et $k \leqq 7$ par (3). Compte tenu de ces inégalités, les seuls L_a possibles sont

$$O(-7E_0 + 3E_1 + \ldots + 3E_6 + 2E_7) ;$$

$$O(-6E_0 + 3E_1 + 3E_2 + 3E_3 + 2E_4 + \ldots + 2E_7) ;$$

$$O(-5E_0 + 2E_1 + \ldots + 2E_7); \qquad O(-4E_0 + 2E_1 + \ldots + 2E_4 + E_5 + E_6 + E_7).$$

Les deux derniers cas sont impossibles par les exemples 2 et 3 du par. 2. Par le lemme 1 (par. 1) et l'exemple 2, il existe un diviseur effectif C tel que

$$C \equiv (k+1)E_0 - \sum_{i=1}^{4} (a_i+1)E_i - \sum_{i=5}^{7} (a_i-1)E_i ,$$

ce qui donne pour les deux premiers cas les courbes suivantes:

$$C_1 = 8E_0 - 4E_1 - \ldots - 4E_4 - 2E_5 - 2E_6 - E_7 ,$$

$$C_2 = 7E_0 - 4E_1 - 4E_2 - 4E_3 - 3E_4 - E_5 - E_6 - E_7 .$$

Notons Q_i $(i=5,6,7)$ la conique de $\mathbb{P}^2$ passant par les points $p_1, \ldots, p_4$ et p_i. Le théorème de Bezout entraîne que chacune des courbes C_1 et C_2 doit contenir doublement Q_5 et Q_6 et simplement Q_7, ce qui est impossible.

3) Cas $r=8$

On peut supposer ici $d \leqq 3$, et aussi $k \geqq 3$ [les seuls sous-faisceaux saturés de $\Omega^1_{S_8}$ avec $k \leqq 2$ sont les $O(-2E_0 + 2E_i)$ et $O(-2E_0)$, qui vérifient l'inégalité (1)]. Numérotons les points p_i de façon que $a_1 \geqq a_i$ pour tout i; comme $d = 3k + 3 - \sum a_i$, on a alors $a_1 \geqq 2$.

Soit D le diviseur $6E_0-3E_1-2\sum_{i=2}^{8} E_i$. On a $\langle D,D\rangle=\langle D,K\rangle=-1$; il s'ensuit par Riemann-Roch que D est linéairement équivalent à une courbe exceptionnelle C (rationnelle et lisse). Considérons la suite exacte

$$0\to O_S(-C)|_C\to\Omega^1_{S_8}|_C\to\omega_C\to 0\,;$$

Par produit tensoriel avec L_a^{-1} on obtient la suite exacte

$$0\to L_a^{-1}(-C)|_C\to(\Omega^1_{S_8}\otimes L_a^{-1})|_C\to\omega_C\otimes L_a^{-1}|_C\to 0\,. \tag{5}$$

On a $\deg_C(L_a^{-1}(-C)|_C)=2d-a_1-5$ et $\deg_C(\omega_C\otimes L_a^{-1}|_C)=2d-a_1-8$, de sorte que ces deux faisceaux sont de degré <0. On déduit alors de la suite exacte de cohomologie associée à (5) l'égalité $H^0(C,\Omega^1_{S_8}\otimes L_a^{-1}|_C)=0$. Cela signifie que toute section globale de $\Omega^1_{S_8}\otimes L_a^{-1}$ s'annule sur C, ce qui contredit le fait que L_a est saturé. Cela achève la démonstration du théorème.

Bibliographie

[D] Demazure, M.: Surfaces de Del Pezzo (exp. IV). Séminaire sur les singularités des surfaces. Lectures Notes in Mathematics, Vol. 777, pp. 50–60. Berlin Heidelberg New York: Springer 1980

[R] Raynaud, M.: Fibrés vectoriels instables – Applications aux surfaces. Surfaces algébriques. Lecture Notes in Mathematics, Vol. 868, pp. 293–314. Berlin Heidelberg New York: Springer 1981

Reçu le 14 juillet 1988

Math. Ann. 283, 177–191 (1989)

Mathematische
Annalen

Singular Moduli, Modular Polynomials, and the Index of the Closure of $\mathbb{Z}[j(\tau)]$ in $\mathbb{Q}(j(\tau))$

David R. Dorman

Department of Mathematics and Computer Science, Middleburg College, Middleburg, VT 05753, USA

1. Introduction

Let τ be an element in the upper half plane and imaginary quadratic over $\mathbb{Q}$. Then τ satisfies an integral quadratic equation $a\tau^2+b\tau+c=0$ with $a, b, c\in\mathbb{Z}$, $(a,b,c)=1$. Denote by $\operatorname{disc}\tau=d=b^2-4ac<0$ the discriminant of τ, $h=h(d)$ the order of the class group of the quadratic order $\mathcal{O}=\mathbb{Z}[(b+\sqrt{d})/2]\subset\mathbb{Q}(\sqrt{d})=K$, and $w=w(d)$ be the number of roots of unity in that order.

The value $j(\tau)$ where j is the elliptic modular function, is called a singular modulus and is an algebraic integer of degree h over $\mathbb{Q}$.

Fix two fundamental negative disriminants d_1 and d_2. Denote by w_i the number of roots of unity in the quadratic order of discriminant d_i, and let h_i denote the class number of those orders.

Three objects of study are:

1. The differences of singular moduli

$$J(d_1,d_2)=\left(\prod_{\substack{[\tau_1][\tau_2]\\ \operatorname{disc}(\tau_1)=d_1\\ \operatorname{disc}(\tau_2)=d_2}}(j(\tau_1)-j(\tau_2))\right)^{4/w_1w_2}. \tag{1.1}$$

Here $[\tau]$ denotes an equivalence class modulo $SL_2(\mathbb{Z})$.

2. The "polynomial"

$$f_d(x)=\left(\prod_{\substack{[\tau]\\ \operatorname{disc}\tau=d/g^2}}(x-j(\tau))\right)^{2/w(d/g^2)}, \tag{1.2}$$

where $[\tau]$ denotes an equivalence class mod $PSL_2[\mathbb{Z}]$. Note that

$$J(d_1,d_2)=\prod_{\substack{[\tau_1]\\ \operatorname{disc}(\tau_1)=d_1}}f_{d_2}(j(\tau_1))^{2/w(d_1)}.$$

3. The m^{th} modular polynomial, $\varphi_m(x,y)\in\mathbb{Z}[x,y]$ defined by

$$\varphi_m(j(z)-j(z'))=\prod_{\substack{\det\gamma=m\\ \operatorname{mod} SL_2(\mathbb{Z})}}(j(z)-j(\gamma z')). \tag{1.3}$$

This product is taken over all equivalence classes of 2×2 matrices of determinant m, modulo the left action of $SL_2(\mathbb{Z})$.

In [8] Gross and Zagier produced a formula for $\mathrm{ord}_\ell(J(d_1, d_2))$ at a finite rational prime ℓ. Using the formula they then studied f_d and φ_m, with the application of finding the index, I, of $\mathbb{Z}[j(\tau)]$ in its integral closure in $\mathbb{Q}(j(\tau))$. These results, with general relatively prime composite discriminants, were later used in a fundamental way in their paper [7]. They gave algebraic and analytic proofs of their results. However, the algebraic proof was only for the case of prime discriminants.

In [4], generalizing the work of Gross and Zagier, we produced a formula for $\mathrm{ord}_\ell(J(d_1, d_2))$ at a finite rational prime ℓ in the case of relatively prime *composite* discriminants and gave a totally algebraic proof of our theorems. In this paper we use our earlier results and the blueprint provided by Gross and Zagier to generalize the study of φ_m and I to composite discriminants. We also produce a formula for the index I in the case of composite discriminant d.

One simply stated result is any prime λ of either $\mathbb{Q}(\sqrt{d_1})$ or $\mathbb{Q}(\sqrt{d_2})$ dividing $\varphi_m(j(\tau_1) - j(\tau_2))$ must have characteristic $\ell < md_1d_2/4$. A complete description is given in Sect. 4.

The results on the index, I, are quite technical and we must introduce some additional notation. Let d, K, and $\mathcal{O}$ be as in Sect. 1, and for non-negative integers n let $R(n)$ (resp. $r_1(n)$) the number of integral ideals (resp. integral ideals in the principal class) of $\mathcal{O}$ having norm n. Extend both functions to $\mathbb{R}$ by setting $R(x) = r_1(x) = 0$ for arguments other than non-negative integers. For each prime p, finite of infinite, let $\varepsilon_p : \mathbb{Q}_p(\sqrt{d}) \to \{\pm 1\}$ be the local character given by class field theory.

Define

$$\varrho_\ell(n) = \begin{cases} 0 & \text{if there exist two primes } p \mid d \text{ such that} \\ & \varepsilon_p((n - |d|)/n) = -1 . \\ 2^{a(n)} & \text{otherwise, where } a(n) = \mathrm{Card}\{p \mid (n, d)\} . \end{cases}$$

(1.4) **Theorem.** *Let ℓ be a rational prime not dividing d. Then*

$$\mathrm{ord}_\ell(I) = \frac{1}{2} \sum_{n \geqq 0} \sum_{k \geqq 1} \varrho_\ell(n) \cdot (R(n) - r_1(n)) \cdot R\left(\frac{|d| - n}{\ell^k}\right).$$

A more precise form of this theorem extending it to all rational primes is found in Sect. 5.

As in our earlier paper the proofs depend on the interplay between the geometry of supersingular elliptic curves in characteristic ℓ and the arithmetic of orders the definite quaternion algebra ramified at ℓ and ∞. These results can be found in detail in [3, 4] and are collected, without proofs, in Sect. 2 for the convenience of the reader and to introduce notation. Section 3 presents the results of [4] that will be generalized in this paper. Sections 4 and 5 contain the main results of this paper, and Sect. 6 gives a simple example of an index computation. Section 7, the appendix, is devoted to a proof of the formula for the discriminant of $\mathbb{Q}(j(\tau))/\mathbb{Q}$. A formula for the discriminant may be well known, however we could not find it in the literature. Dummit, Gold, and Kisilevsky [5] compute the square free part of this discriminant.

2. Preliminaries

In addition to the notation already given let t be the number of distinct prime factors of the discriminant d. H will denote the Hilbert class field of $K=\mathbb{Q}(\sqrt{d})$, $G=\mathrm{Gal}(H/K)\cong \mathrm{Pic}(\mathcal{O})$, where $\mathrm{Pic}(\mathcal{O})$ is the ideal class group of K. Let χ_d be the primitive quadratic character defined mod d extended to $\mathbb{Z}$ in the usual way. For each $p|d$ let χ_p be the associated quadratic character mod p.

Associated to χ_d is an idèle character $\varepsilon: \mathbb{A}_{\mathbb{Q}}^*/\mathbb{Q}^*\mathbb{N}\mathbb{A}_K^* \to \mathrm{Gal}(K/\mathbb{Q})\cong\{\pm 1\}$. Thus for each prime p of $\mathbb{Q}$ there is a local character $\varepsilon_p: \mathbb{Q}_p \to \mathrm{Gal}(K/\mathbb{Q})$ associated to the extension $\mathbb{Q}_p(\sqrt{d})$. Let

$$\mathrm{Frob}\, p=\begin{cases} 1 & \text{if } p \text{ splits in } K \\ -1 & \text{if } p \text{ remains inert in } K. \end{cases}$$

Then ε_p can be evaluated on any integer n by writing $n=u\cdot p^{a_p}$ where $gcd(p,u)=1$ and letting

$$\varepsilon_p(n)=\begin{cases} \mathrm{sgn}(n) & p=\infty \\ \mathrm{Frob}\, p^{a_p} & p\nmid d \\ \chi_p(u)\chi_{d/p}(p)^{a_p} & p|d \end{cases}. \tag{2.1}$$

The ε_p can be thought of as the genus characters of K. In this language the main theorem of genus theory states

$$\prod_{p|d} \varepsilon_p(\mathbb{N}\mathfrak{a})=1 \quad \text{for all ideals } \mathfrak{a} \text{ in } \mathcal{O}. \tag{2.2}$$

Here $\mathbb{N}$ is the absolute norm.

Let ℓ be an inert or ramified prime in K. We now describe maximal orders containing $\mathcal{O}$ and their subrings in the definite quaternion algebra $\mathbb{D}$ defined over $\mathbb{Q}$ ramified only at ℓ and ∞. Details for the following are given in [3].

Assume ℓ is inert in K. Fix a prime q such that for all $p|d\chi_p(-\ell q)=1$. The existence of q is guaranteed by Dirichlet's Theorem. These conditions imply that $q\mathcal{O}=\mathfrak{q}\bar{\mathfrak{q}}$ and that $x^2\equiv -\ell q \bmod p$ has two solutions for each $p|d$. For each p fix one such solution λ_p. By the Chinese Remainder Theorem we fix a *congruence solution* $\lambda\in\mathbb{Z}$ such that $\lambda\equiv\lambda_p \bmod p$. Then $\mathbb{D}=\{d, -\ell q\}$ (see Vignéras [12] p. 2 for the notation) and can be realized as the subalgebra

$$\mathbb{D}=\left\{[\alpha,\beta]=\begin{bmatrix} \alpha & \beta \\ -\ell q\bar{\beta} & \bar{\alpha}\end{bmatrix} : \alpha,\beta\in K\right\}.$$

Here $\alpha\mapsto\bar{\alpha}$ denotes complex conjugation. Note, there is a fixed embedding of K into $\mathbb{D}$ by $\alpha\mapsto[\alpha,0]$.

For any ideal $\mathfrak{a}$ of $\mathcal{O}$ there exists a maximal order $R(\mathfrak{a})$ in $\mathbb{D}$ containing $\mathcal{O}$ optimally, that is $\mathcal{O}\subset R(\mathfrak{a})$ and $R(\mathfrak{a})\cap K=\mathcal{O}$. Namely let $\lambda'_p=-1^{\mathrm{ord}(\mathfrak{a})}\lambda_p$ where $\mathfrak{p}$ is the prime of $\mathcal{O}$ above p. Let λ' be the corresponding congruence solution and let $\mathscr{D}^{-1}$ be the inverse different of $\mathcal{O}$. Then

$$R(\mathfrak{a})=R(\mathfrak{a},\lambda')=\{[\alpha,\beta]\in\mathbb{D}: a\in\mathscr{D}^{-1}, \beta\in\mathfrak{q}^{-1}\mathscr{D}^{-1}\mathfrak{a}^{-1}\bar{\mathfrak{a}};\ \alpha\equiv\lambda'\beta \bmod \mathcal{O}\}. \tag{2.3}$$

Observe that $R(\mathfrak{a})$ admits a filtration

$$R(\mathfrak{a})_n=\{[\alpha,\beta]\in R(\mathfrak{a}): \beta\equiv 0 \bmod \ell^{n-1}\}. \tag{2.4}$$

That $R(\mathfrak{a})$ is maximal is proved in [3] where the following important facts concerning $R(\mathfrak{a})$ are established:

1. Any maximal order in $\mathbb{D}$ containing $\mathcal{O}$ optimally is of the form $R(\mathfrak{a})$ for some ideal $\mathfrak{a}$.

2. The conjugation relation $R(\mathfrak{a})\mathfrak{b}=\mathfrak{b}R(\mathfrak{a}\mathfrak{b})$ holds for integral ideals $\mathfrak{a}$ and $\mathfrak{b}$ of $\mathcal{O}$.

3. Up to conjugation by K^* there are exactly h distinct maximal orders containing $\mathcal{O}$ optimally.

A critical observation is if the ideal $\mathfrak{b}|\langle\sqrt{d}\rangle$ then $\bar{\mathfrak{b}}\mathfrak{b}^{-1}=1$ so $R(\mathfrak{a},\lambda')\mathfrak{b}=\mathfrak{b}R(\mathfrak{a},\lambda'')$ where λ'' differs from λ' only in the choice of sign on the λ_p's. Thus, up to conjugacy by K^* there are 2^{t-1} orders that all look like $R(\mathfrak{a})$ except for the congruence solution. Note, we do not obtain 2^t orders since changing all the signs on the λ_p amounts to conjugating by $\sqrt{d}\in K^*$. This ambiguity will cause an obstruction as pointed out in the proof of Lemma 4.8.

Now assume ℓ is ramified in K so $\ell=p|d$. Choose a prime q such that for all $p'|d$, $p'\neq p$, $\chi_{p'}(q)=1$ and $\chi_p(q)=-1$. Then

$$\mathbb{D}=\left\{[\alpha,\beta]=\begin{bmatrix}\alpha & \beta\\ -q\bar{\beta} & \bar{\alpha}\end{bmatrix}:\alpha,\beta\in K\right\}$$

and

$$\begin{aligned}(2.5)\qquad R(\mathfrak{a})=R(\mathfrak{a},\lambda')=\{[\alpha,\beta]\in\mathbb{D}:\alpha\in\mathfrak{p}\mathcal{D}^{-1},\\ \beta\in\mathfrak{q}^{-1}\mathcal{D}^{-1}\mathfrak{a}^{-1}\bar{\mathfrak{a}};\ \alpha\equiv\lambda'\beta \bmod p',\ p'\neq p\}\end{aligned}$$

$$(2.6)\qquad R(\mathfrak{a})_n=\{[\alpha,\beta]\in E(\mathfrak{a}):\beta\equiv 0 \bmod \mathfrak{p}^n\}.$$

As before $\mathfrak{p}$ is the prime of K over p. The same observations and facts hold in this case as they did previously.

3. Review of Earlier Work

Some of the results and techniques from [4] are used in an essential way in what follows so we review that work with some detail now.

Let $d_1=d$ and d_2 be another fundamental negative discriminant relatively prime to d_1. Let $w_i=w(d_i)$. Fix a finite prime v of H having characteristic ℓ and denote by $A=A_v$ the completion of the maximal, unramified, extension of the ring of v integers in H. Let $W=W_v=A[s]$ where s is a fixed element which satisfies an integral quadratic equation of discriminant d_2. Let e be the ramification index of W/A and π a uniformizer for W.

Let $j_1=j(\tau_1)$. The algebraic integer

$$(3.1)\qquad \alpha=\alpha(\tau_1,d_2)=\Bigg(\prod_{\substack{\operatorname{disc}\tau_2=d_2\\ [\tau_2]}}(j_1-j(\tau_2))\Bigg)^{4/w_1w_2},$$

lies in H, and, in fact, lies in $\mathbb{Q}(j_1)$ when $d_1,d_2\neq -4$. Observe that $\mathbb{N}_{H/\mathbb{Q}(\sqrt{d_1})}(\alpha)=J(d_1,d_2)$. The product is taken over representative classes mod $SL_2(\mathbb{Z})$. One can, of course, try to compute $\operatorname{ord}_v(\alpha)$ for any v in H. Let E be an elliptic curve defined over W with complex multiplication by $\mathcal{O}$ and invariant $j(E)=j_1$. It is easy to show such a curve exists, and, by a theorem of Serre and Tate [11], E has good reduction and is unique up to W isomorphism since the residue field is algebraically closed. Similarly, for each τ_2 of discriminant d_2 let E' denote the elliptic curve defined over W having complex multiplication by $\mathbb{Z}[s]$ and invariant $j(E')=j(\tau_2)$. Then

$$\operatorname{ord}_v(\alpha)=\frac{4}{ew_1w_2}\sum_{\operatorname{disc}(\tau_2)=d_2}\operatorname{ord}_\pi(j(E)-j(E')),$$

which by [8, pp. 196 and 200] gives

$$\operatorname{ord}_v(\alpha)=\frac{4}{ew_1w_2}\sum_{\operatorname{disc}(\tau_2)=d_2}\sum_{n\geqq 1}\frac{1}{2}\operatorname{Card}\{\operatorname{Iso}_{W/\pi^n}(E,E')\}.$$

Thus the problem is reduced to counting isomorphisms $f:E\tilde{\to}E' \bmod \pi^n$, or equivalently, [8, pp. 200–201] endomorphisms $s_f=f^{-1}\circ s\circ f$ of $E \bmod \pi^n$ belonging to the set

$$S_{n,v}=\left\{\begin{matrix}\alpha_0\in\operatorname{End}_{W/\pi^n}(E):\operatorname{Tr}(\alpha_0)=\operatorname{Tr}(s),\ \mathbb{N}(\alpha_0)=\mathbb{N}(s),\\ \alpha_0 \text{ induces multiplication by } s \text{ on } \operatorname{Lie}(E).\end{matrix}\right\}$$

Thus, $\operatorname{ord}_v(\alpha)=\dfrac{2}{ew_1}\displaystyle\sum_{n\geqq 1}\operatorname{Card}\{S_{n,v}\}$.

If $\ell=\operatorname{char}(v)$ splits in K then $\operatorname{ord}_v(\alpha)=0$ since E has ordinary reduction in this case. If ℓ is inert or ramifies in K then one can hope to determine $\operatorname{ord}_v(\alpha)$ with the aid of the $R(\mathfrak{a})_n$. For simplicity we consider the case ℓ inert. An analogous sitation holds for ℓ ramified.

Deuring's theory [2] tells us that in this case there exists an integral ideal $\mathfrak{a}$ of $\mathcal{O}$ such that $\operatorname{End}_{W/\pi^n}(E)=R(\mathfrak{a})_n$ where $R(\mathfrak{a})_n$ is given by (2.4). One then checks that an $\alpha_0\in R(\mathfrak{a})_n$ gives rise to an integer x and an integral ideal $\mathfrak{b}$ of $\mathcal{O}$, $\mathfrak{b}\sim\mathfrak{q}\mathfrak{a}^2$ solving the Diophantine equation $x^2+4\ell^{2n-1}\mathbb{N}\mathfrak{b}=d_1d_2$. However, the converse is not quite true. A solution $(x,\mathfrak{b})$ does indeed give rise to an endomorphism, however it does not necessarily end up in the original $R(\mathfrak{a})_n$ but in one of the 2^{t-1} rings conjugate to $R(\mathfrak{a})_n$ by an ideal $\mathfrak{c}\,|\,\pi$. There is no way to distinguish in which of these 2^{t-1} rings the endomorphism lies. Moreover, a single solution $(x,\mathfrak{b})$ not just one but $\operatorname{Card}\{p\,|\,(x,d_1)\}$ endomorphisms. Thus, while in principle one can factor α in H, the hopes of finding a formula for this factorization seem unobtainable by our present method. The above difficulty is overcome by descending to the subfield L of H fixed by the subgroup of G generated by the elements of order 2. Note $[H:L]=2^{t-1}$ and if u is a prime of L under v then

$$\operatorname{ord}_u(\mathbb{N}_{H/L}(\alpha))=\sum_{v'|u}\operatorname{ord}_{v'}(\alpha).$$

It turns out that a formula for $\operatorname{ord}_u(\mathbb{N}_{H/L}(\alpha))$ can be obtained and the remarkable fact is that it depends only on the genus class of $\mathfrak{q}$ but not of $\mathfrak{a}$. We remark that an analogous situation holds if $\ell=p\,|\,d_1$ but now we obtain a solution $(x,\mathfrak{b})$ to $x^2+4p\mathbb{N}\mathfrak{b}=d_1d_2$ with $x\in\mathbb{Z}$ and $\mathfrak{b}$ an integral ideal of $\mathcal{O}$ in the class of $\mathfrak{p}\mathfrak{q}\mathfrak{a}^2$. Here $\mathfrak{p}\,|\,p$.

Let $r_*(n)$ be the number of integral ideals of $\mathcal{O}$ in the class of $*$ having norm n. Then u has characteristic ℓ and we have

(3.2) **Theorem.** *Let ℓ be a rational prime and H, L, u, $\mathfrak{a}$, $\mathfrak{p}$, and $\mathfrak{q}$ be as above and in Sect. 2. Then*

$$\operatorname{ord}_u(\mathbb{N}_{H/L}(\alpha)) = \begin{cases} 0 & \text{if } \chi_{d_1}(\ell)=1 \\ \dfrac{1}{2}\sum_{x\in\mathbb{Z}}\sum_{n\geqq 1} 2^{a(x)}\cdot r_{\mathfrak{q}\mathfrak{a}^2}\left(\dfrac{d_1d_2-x^2}{4\ell^n}\right) & \text{if } \chi_{d_1}(\ell)=-1 \\ \dfrac{1}{2}\sum_{x\in\mathbb{Z}} 2^{a(x)}\cdot r_{\mathfrak{q}\mathfrak{p}\mathfrak{a}^2}\left(\dfrac{d_1d_2-x^2}{4p}\right) & \text{if } \ell=p\mid d \end{cases}$$

where $a(x)=\operatorname{Card}\{p\mid(x,d_1)\}$.

Proof. See [4, Propositions 3.4, 3.9, 3.11, 3.13, and Theorem 4.1]. □

4. Generalization to Modular Polynomials

From here on we fix one fundamental negative discriminant D having t distinct prime divisors and we assume $j=((D+\sqrt{D})/2)$ is a singular modulus. All notation is as before except for W which we will define later.

Assume $m\in\mathbb{Z}$ is not a perfect square. Then Kronecker's identity relating $\varphi_m(x,x)$ to $f_D(x)$ is

$$\varphi_m(x,x)=\pm\prod_{t\in\mathbb{Z};\,t^2<4m} f_{t^2-4m}(x).$$

Suppose that $m\geqq 1$ is *not* the norm of an element $(a+b\sqrt{D})/2$ in $\mathcal{O}$. Then the value $\varphi_m(j,j)\neq 0$. Thus for ℓ inert and $\mathfrak{a}$ and $\mathfrak{q}$ as defined in Sect. 2, Kronecker's identity can be use to recast the results of Theorem 3.2 as

$$\begin{aligned}\operatorname{ord}_u(\mathbb{N}_{H/L}(\varphi_m(j,j)))^{4/w^2} &= \frac{4}{w^2}\sum_{t^2\leqq 4m}\left\{\frac{1}{2}\sum_{x\in\mathbb{Z}}\sum_{n\geqq 1} 2^{a(x)} r_{\mathfrak{q}\mathfrak{a}^2}\left(\frac{|D|(4m-t^2)-x^2}{4\ell^k}\right)\right\} \\ &= \sum_{n\geqq 0}\sum_{k\geqq 1} 2^{a(n)} r_1(n) r_{\mathfrak{q}\mathfrak{a}^2}\left(\frac{m|D|-n}{\ell^k}\right),\end{aligned}$$

where $n=(x^2-Dt^2)/4$ is the norm of an ideal in the principal class. For an ideal class $\mathfrak{b}$ define $r_{\mathfrak{b}}(0)=1/w$. If $\ell=p\mid D$ the last line above would read

$$=\sum_{n\geqq 0}\sum_{k\geqq 1} 2^{a(n)} r_1(n) r_{\mathfrak{q}\mathfrak{p}^k\mathfrak{a}^2}\left(\frac{m|D|-n}{p^k}\right).$$

This formula can be generalized. Let $\mathfrak{b}$ be an ideal of $\mathcal{O}$ and $m\geqq 1$ an integer which is *not* the norm of an ideal in the class of $\mathfrak{b}$. Denote by $\sigma_{\mathfrak{b}}=\sigma$ be the element of $\operatorname{Gal}(H/K)$ that corresponds to $\mathfrak{b}$ under the Artin map. Then

$$\beta=\varphi_m(j,j^\sigma)^{4/w^2} \tag{4.2}$$

is an algebraic integer and we now produce a formula for $\operatorname{ord}_u(\mathbb{N}_{H/L}\beta)$.

First observe by 1.2

$$\varphi_m(j,j^\sigma)^{4/w^2}=\prod_{j'}(j-j')^{4/w^2}, \tag{4.3}$$

where j' is the modular invariant of a curve E' which is m-isogenous to E^σ. Recall A is the completion of the maximal unramified extension of the v-integers of H. Now let $W=A[j']$ the field obtained by adjoining all the invariants j' of curves that are m-isogenous to E^σ. Let π be a uniformizer for W. Note if $l\nmid m$ then $W=A$ since the j' are all unramified over A. This is the case of greatest interest since $m=1$ in the later application.

Let E be an elliptic curve defined over W having complex multiplication by $\mathcal{O}$ and modular invariant $j(E)=j'$. From the definition of φ_m it follows that the factorization of β depends on determining the cardinality of

$$S_n=\{\mathrm{Iso}_{W/\pi^n W}(E,E');\ E'\ m\text{-isogenous to } E\}.$$

By (1.3) and (4.3) $\mathrm{Card}(S_n)$ is related to the cardinality of the set

$$T_n=\mathrm{Hom}_{W/\pi^n W}(E,E^\sigma)_{\mathrm{degree}\, m}.$$

(4.4) **Proposition.** *With the notation as above* $T_n=S_n$.

Proof. To see this let $f\in S_n$ and φ an m isogeny between E' and E^σ. Then $f\circ\varphi$ is an m isogeny for E to E^σ so $S_n\subset T_n$. On the other hand, if φ is an m isogeny for E^σ to E define $E'=E^\sigma/\ker\varphi$. Then $E\cong E'$ thus $T_n\subset S_n$. □

Since $\mathrm{ord}_v(\beta)=4/w^2\sum_{j'}\mathrm{ord}_v(j-j')$, by Proposition 2.3 in [5] $\mathrm{ord}_v(\beta)$ $=4/w^2\sum_{n\geqq 1}\mathrm{Card}(S_n)/2$. And since $S_n=T_n$ we can use T_n in the above formula.

As mentioned in Sect. 2 the ambiguity in the congruence solution in the definition of the $R(\mathfrak{a})$ will prevent us from determining $\mathrm{ord}_v(\beta)$ in H. Instead descend to the subfield L and let u be a prime of L under v. Then

$$\mathrm{ord}_u(\mathbb{N}_{H/L^\beta})=\frac{4}{w^2}\sum_{n\geqq 1}\sum_{v'|u}\frac{\mathrm{Card}(T_n)}{2}. \tag{4.5}$$

The formula for (4.5) is given by

(4.6) **Theorem.** *Let m,β,L, and u be as defined above and let ℓ be a rational prime not dividing m, and let* $B=\mathbb{N}_{H/L^\beta}$

1. *If* $\chi_D(\ell)=1$ *then* $\mathrm{ord}_u(B)=0$
2. *If* $\chi_D(\ell)=-1$, *then for* $\mathfrak{a}$ *and* $\mathfrak{q}$ *as in Sect. 2*

$$\mathrm{ord}_u(B)=\sum_{n\geqq 0}\sum_{k>1}2^{a(n)}r_{\mathfrak{b}^{-1}}(n)r_{\mathfrak{q}\mathfrak{a}^2}\left(\frac{m|D|-n}{\ell^k}\right).$$

3. *If* $\ell=p|D$, *then for* $\mathfrak{a}$, $\mathfrak{q}$, *and* $\mathfrak{q}$ *as in Sect. 2*

$$\mathrm{ord}_u(B)=\sum_{n\geqq 0}\sum_{k>1}2^{a(n)}r_{\mathfrak{b}^{-1}}(n)r_{\mathfrak{q}\mathfrak{p}^k\mathfrak{a}^2}\left(\frac{m|D|-n}{p^k}\right).$$

The proof of this theorem is broken down into 3 lemmas.

(4.7) **Lemma.** *Assume* $\chi_D(\ell)=1$. *Then* $\mathrm{ord}_u(\mathrm{B})=0$ *for any prime u of L of characteristic* ℓ.

Proof. By During's theory [2] $\mathrm{End}_{W/\pi^n W}E\cong\mathcal{O}$. Let $\mathfrak{b}$ be an ideal in the class of σ. By results of Serre [1] $\mathrm{Hom}_W(E,E^\sigma)\cong\mathcal{O}$ as an $\mathrm{End}_W E\cong\mathcal{O}$ module inside K. Thus

$\mathrm{Hom}_W(E, E^\sigma) = \mathrm{Hom}_{W/\pi^n W}(E, E^\sigma) = \mathcal{O}\mathfrak{b} = \mathfrak{b}$. Now assume φ is an isogeny. Then $\deg\varphi = \mathbb{N}\alpha/\mathbb{N}\mathfrak{b}$ where $\alpha \in \mathfrak{b}$. Since we are looking for isogenies of degree m it follows that $\mathbb{N}\alpha = m\mathbb{N}\mathfrak{b}$. Put $\mathfrak{c} = (\alpha)\mathfrak{b}^{-1}$. Then $\mathbb{N}\mathfrak{c} = m$. Hence $r_{\sigma^{-1}}(m) = r_\sigma(m) > 0$. This is impossible since assuming that m is not the norm of an ideal in the class of $\mathfrak{b}$ means $r_\sigma(m) = 0$. Thus $\mathrm{Card}(T_{n,v'}) = 0$; for all v' and the lemma is proved. □

(4.8) **Lemma.** *Assume $\chi_D(\ell) \neq 1$ and $\ell \nmid Dm$. Then the sum $\sum_{n \geqq 0} \sum_{v'|u} \frac{1}{2} \mathrm{Card}\, T_{n,v'}$ is equal to the number of solutions to the equation $\mathbb{N}\mathfrak{c} + \ell^{2k-1}\mathbb{N}\mathfrak{d} = Dm$ where $\mathfrak{c}$ and $\mathfrak{d}$ are integral ideals of $\mathcal{O}$ in the class of $\mathfrak{b}^{-1}$ and $\mathfrak{q}\mathfrak{b}\mathfrak{a}^2$ respectively. Each solution is counted with multiplicity $2^{a(\mathbb{N}\mathfrak{c})} \cdot \frac{1}{4} \cdot w^2$.*

Proof. Since $\ell \nmid Dm$, Deuring's theory tells us that there exists an ideal $\mathfrak{a}$ of $\mathcal{O}$ such that $R(\mathfrak{a})_n \cong \mathrm{End}_{W/\pi^n W}(E)$. Serre's result in this case shows that $\mathrm{Hom}_{W/\pi^n W}(E, E^\sigma) = \mathrm{End}_{W/\pi^n W}(E)\mathfrak{b}$ as an $\mathrm{End}_{W/\pi^n W}(E)$ module inside $\mathbb{D}$. Explicitly

$$\mathrm{Hom}_{W/\pi^n W}(E, E^\sigma) = \{[\alpha, \beta] \in \mathbb{D} : \alpha \in \mathcal{D}^{-1}\mathfrak{b},\ \beta \in \mathcal{D}^{-1}\mathfrak{q}^{-1}\ell^{n-1}\overline{\mathfrak{b}\mathfrak{a}}\,\mathfrak{a}^{-1};\ \alpha \equiv \beta \bmod \mathcal{O}\}.$$

Thus we must find homomorphisms $[\alpha, \beta]$ with norm equal to $m\mathbb{N}\mathfrak{b}$. Writing $\alpha = \gamma/\sqrt{D}$ with $\gamma \in \mathfrak{b}$ and $\beta = \ell^{n-1}\delta/\sqrt{D}$ where $\delta \in (\mathfrak{q}\mathfrak{a})^{-1} \cdot \overline{\mathfrak{q}\mathfrak{a}}$ the norm condition implies

$$-1(\mathbb{N}\gamma + q\ell^{2n-1}\mathbb{N}\delta)/D = m\mathbb{N}\mathfrak{b}$$

or equivalently

$$\mathbb{N}\gamma + q\ell^{2n-1}\mathbb{N}\delta = |D|\, m\mathbb{N}\mathfrak{b}\,. \tag{4.9}$$

Setting $\mathfrak{c} = (\gamma)/\mathfrak{b}$ and $\mathfrak{d} = (\delta)\mathfrak{q}\mathfrak{a}(\bar{\mathfrak{b}}\bar{\mathfrak{a}})^{-1}$ we obtain a solution $(\mathfrak{c}, \mathfrak{d})$ to

$$\mathbb{N}\mathfrak{c} + \ell^{2n-1}\mathbb{N}\mathfrak{d} = |D|m \tag{4.10}$$

where $\mathfrak{c}$ and $\mathfrak{d}$ are integral ideals of $\mathcal{O}$ in the class of $\mathfrak{b}^{-1}$ and $\mathfrak{q}\mathfrak{b}\mathfrak{a}^2$ respectively.

On the other hand, beginning with a solution $(\mathfrak{c}, \mathfrak{d})$ to (4.10) reversing the steps in the above argument a homomorphism $[\alpha, \beta]$ is constructed in at least one of the 2^{t-1} conjugate right ideals $R(\mathfrak{a}, \lambda)_n$. However, there is no way to determine in which of these ideals the homomorphism lies. Moreover, it is possible for a single solution to contribute more than one homomorphism if $gcd(D, \mathbb{N}\mathfrak{c}) > 1$. For example, if $p = \gcd(D, \mathbb{N}\mathfrak{c})$ then reducing (4.10) mod p shows $\ell^{2n-1}\mathbb{N}\mathfrak{d} \equiv 0 \bmod p$. This implies $\delta \equiv 0 \bmod p$, hence $0 \equiv \mathbb{N}\mathfrak{c} \equiv \pm\lambda_p\delta^2 \bmod p$ is trivially satisfied with both signs on λ_p. Thus 2 homomorphisms are constructed from a single solution. Continuing in this way it is easy to see if $a(\mathbb{N}\mathfrak{c}) = \mathrm{Card}\{p \mid \gcd(D, \mathbb{N}\mathfrak{c})\}$ then $2^{a(\mathbb{N}\mathfrak{c})}$ homomorphisms are obtained from a single solution.

Both of these ambiguities are bypassed by descending to the field L where computing the $\mathrm{ord}_u(B)$ is the same as summing over all the right ideals K^* conjugate to $R(\mathfrak{a})$. This accounts for the inner summation sign.

The $\frac{1}{4} \cdot w^2$ term is 1 if $D < -4$. When $D = -3$ the generator δ of $\mathfrak{d}(\mathfrak{a}\mathfrak{q})^{-1}\overline{\mathfrak{b}\mathfrak{a}}$ and the generator γ of $\mathfrak{c}\mathfrak{b}$ can each be altered by a sixth root of unity giving $9 = \frac{1}{4} \cdot w^2$ times as many homomorphisms as expected for each solution $(\mathfrak{c}, \mathfrak{d})$. Similarly, when $D = -4$, γ and δ can each be altered by a fourth root of unity giving $4 = \frac{1}{4} \cdot w^2$ times the number of expected homomorphisms. □

The final lemma is

(4.11) **Lemma.** *Assume* $\chi_D(\ell) \neq 1$ *and* $\ell = p|D$, *but* $\ell \nmid m$. *Then the sum* $\sum_{n \geq 0} \sum_{v|u} \frac{1}{2}\mathrm{Card}(T_n)$ *is equal to the number of solutions to the equation* $\mathbb{N}\mathfrak{c} + p^{n-1}\mathbb{N}\mathfrak{d} = |D|m$, *where* $\mathfrak{c}$ *and* $\mathfrak{d}$ *are integral ideals of* $\mathcal{O}$ *in the classes of* $\mathfrak{b}^{-1}\mathfrak{p}$ *and* $\mathfrak{q}\mathfrak{b}\mathfrak{p}^n\mathfrak{a}^2$ *respectively. Each solution is counted with multiplicity* $2^{a(\mathbb{N}\mathfrak{c})} \cdot \frac{1}{4} \cdot w^2$.

Proof. For some ideal $\mathfrak{a}$ (4.6) provides a discription of $\mathrm{End}_{W/\pi^n W}(E)$. Combining this with Serre's result gives

$$\begin{aligned}\mathrm{Hom}_{W/\pi^n W}(E, E^\sigma) = \{[\alpha, \beta] \in \mathbb{D} : \alpha \in \mathfrak{p}\mathscr{D}^{-1},\ \beta \in \mathscr{D}^{-1}\mathfrak{q}^{-1}\mathfrak{p}^n\mathfrak{a}^{-1}\overline{\mathfrak{b}\mathfrak{a}};\\ \alpha \equiv \lambda\beta \bmod p',\ p' \neq p\}.\end{aligned}$$

The argument now proceeds as in the last lemma. Note in this case we can obtain an extra factor of 2 in the special case $\mathbb{N}\mathfrak{c} \equiv 0 \bmod D$ which allows a change of sign on the generator of $\mathfrak{d}$ itself. □

The proof of Theorem 4.6 is simply a matter of combining the above three lemmas.

5. An Application

One application of Theorem 4.6 is the computation of the index I of the order $\mathbb{Z}[j]$ in its integral closure in $\mathbb{Q}(j)$. When $m=1$ and $\mathfrak{b}$ is not a principal ideal (4.1) and (4.2) give the prime factorization of $\mathbb{N}_{H/L}(j - j^\sigma)$. Norming this quantity down to K and then taking the product over all calsses $\mathfrak{b} \nsim 1$ gives the discriminant of the monic polynomial, f, of degree h satisfied by j. It is also true that

$$d(\mathbb{Q}(j)) \cdot I^2 = d(f).$$

Here $d(\mathbb{Q}(j))$ is the absolute discriminant of $\mathbb{Q}(j)$ and $d(f)$ is the discriminant of f. Hence the prime factorization of I can be determined once $d(\mathbb{Q}(j))$ is known.

Write $D = D_0 \cdot D_1$ where

$$D_1 = \begin{cases} 1 & \text{if at least 2 primes congruent to } 3 \bmod 4 \text{ divide } D \\ p & \text{if } p \text{ is the unique prime congruent to } 3 \bmod 4 \text{ dividing } D \\ 4 & \text{if } 4 \| D \text{ and no primes congruent to } 3 \bmod 4 \text{ divide } D \\ 8 & \text{if } 8 \| D \text{ and no primes congruent to } 3 \bmod 4 \text{ divide } D. \end{cases}$$

We then have

(5.1) **Proposition.** $d(\mathbb{Q}(j)) = D_0^{1/2h} \cdot D_1^{1/2(h - 2^{t-1})}$.

Proof. See appendix. □

We now determine I.

First let $\ell \nmid D$ so with $m=1$ (4.10) becomes

$$\mathbb{N}\mathfrak{c} + \ell^{2k-1}\mathbb{N}\mathfrak{d} = |D|. \tag{5.2}$$

Now $\mathfrak{c} \sim \mathfrak{b}^{-1}$ and $\mathfrak{d} \sim \mathfrak{q}\mathfrak{b} \bmod \mathrm{Pic}(\mathcal{O})^2$ if and only if $\mathfrak{c} \sim \mathfrak{q}\mathfrak{d} \bmod \mathrm{Pic}(\mathcal{O})^2$, or equivalently,

$$\varepsilon_p(\mathbb{N}\mathfrak{c}\mathfrak{q}\mathfrak{d}) = 1, \tag{5.3}$$

for each genus character ε_p; $p|D$. Write $\mathbb{N}\mathfrak{c}=n$. Then (5.2) and (5.3) show

$$\mathbb{N}(\mathfrak{c}\mathfrak{q}\mathfrak{d})=nq\left(\frac{|D|-n}{\ell^{2k-1}}\right)\equiv\frac{n-|D|}{n}\cdot -q\ell \bmod \mathbb{Q}^{*2}.$$

Moreover, since q was selected so that $-q\ell$ was congruent to a square mod p for all $p|D$ condition (5.3) translates to

$$\varepsilon_p\left(\frac{n-|D|}{n}\right)=1 \quad \text{for all } p|D.$$

Observe that this condition is satisfied if $p\nmid n$ since then $(n-|D|)/n\equiv 1 \bmod p$. Moreover, since $\varepsilon_\infty((n-|D|)/n)=1$ the product formula, (2.2), implies if the above condition fails it must do so for an even number of primes dividing D.

Now assume $\ell=p|D$, so we are working with the equation

$$\mathbb{N}\mathfrak{c}+p^{k-1}\mathbb{N}\mathfrak{d}=|D|. \tag{5.4}$$

There are two cases to consider.

Case 1. Assume $n\neq 0$. Then $\mathfrak{c}\sim\mathfrak{b}^{-1}$ and $\mathfrak{d}\sim\mathfrak{q}\mathfrak{b}\mathfrak{p}^{k-1} \bmod \operatorname{Pic}(\mathcal{O})^2$. Thus $\mathfrak{c}\sim\mathfrak{q}\mathfrak{d}\mathfrak{p}^{k-1} \bmod \operatorname{Pic}(\mathcal{O})^2$, or equivalently $\varepsilon_{p'}(\mathfrak{c}\mathfrak{q}\mathfrak{d}\mathfrak{p}^{k-1})=1$ for all $p'|D$, $p'\neq p$. As above, if this condition fails it can only do so for an even number of primes dividing D.

Case *2*. Assume $n=0$. Equation (5.4) becomes $p\mathbb{N}\mathfrak{d}={}^*D|$. Hence $\mathfrak{d}\sim\mathfrak{q} \bmod \operatorname{Pic}(\mathcal{O})^2$. But we have $\mathfrak{d}\sim\mathfrak{b}\mathfrak{q}\mathfrak{p} \bmod \operatorname{Pic}(\mathcal{O})^2$ so $\mathfrak{b}\sim\mathfrak{q} \bmod \operatorname{Pic}(\mathcal{O})^2$. Since it is required that $\mathfrak{b}\nsim 1$ we have $\mathfrak{p}\nsim 1$. Now if $\mathfrak{q}\sim 1$ then $\varepsilon_{p'}(q)=1$ for all $p'\neq p$. However, since $\chi_{p'}(-q)=1$ it follows that $\mathfrak{q}\sim 1$ if and only if $\chi_{p'}(-1)=1$, i.e., $p'\equiv 1 \bmod 4$.

The above analysis leads to the following definition. For each positive integer n let

$$\varrho_\ell(n)=\begin{cases}0 & \text{if there exist two primes } p|d \text{ so that}\\ & \varepsilon_p((n-|D|)/n)=-1.\\ 2^{a(n)} & \text{otherwise, where } a(n)=\operatorname{Card}\{p\mid \gcd(n,D)\}.\end{cases}$$

and for $n=0$

$$\varrho_\ell(0)=\begin{cases}0 & \text{if } \ p'\equiv 1 \bmod 4 \text{ for all } p'\neq p=\ell,\\ 2^t & \text{otherwise}.\end{cases}$$

Put $m=1$ and $n=\mathbb{N}\mathfrak{c}$ into (4.1) and (4.2) and take norms down to K and sum over all classes $\mathfrak{b}\nsim 1$. The same result is given in both cases, namely, for any prime λ of K of characteristic ℓ

$$\operatorname{ord}_\lambda(d(f))=\sum_{n\geqq 0}\sum_{k\geqq 1}\varrho_\ell(n)\cdot(R(n)-r_1(n))\cdot R\left(\frac{|D|-n}{\ell^k}\right).$$

Combining this with Proposition 5.1 we have proved

(5.5) **Theorem.** *Let λ be a prime of K with* $\operatorname{char}(\lambda)=\ell$. *Then*

$$\operatorname{ord}_\lambda(I)=\frac{1}{2}\sum_{n\geqq 0}\sum_{k\geqq 1}\varrho_\ell(n)\cdot(R(n)-r_1(n))\cdot R\left(\frac{|D|-n}{\ell^k}\right)$$
$$-\frac{1}{2}\left(\frac{h}{2}\cdot\operatorname{ord}_\lambda(D_0)+\left(\frac{h-2^{t-1}}{2}\right)\cdot\operatorname{ord}_\lambda(D_1)\right) \quad \square$$

(5.6) **Corollary.** *Let ℓ be a prime in $\mathbb{Z}$.*

1. *If* $\chi_D(\ell)=1$, *then* $\mathrm{ord}_\ell(I)=0$.
2. *If* $\chi_D(\ell)=-1$, *then*

$$\mathrm{ord}_\ell(I)=\frac{1}{2}\sum_{n\geqq 0}\sum_{k\geqq 1}\varrho_\ell(n)\cdot(R(n)-r_1(n))\cdot R\left(\frac{|D|-n}{\ell^k}\right).$$

3. *If* $\ell=p\,|\,D$ *then*

$$\mathrm{ord}_p(I)=\frac{1}{4}\sum_{n\geqq 0}\sum_{k\geqq 1}\varrho_\ell(n)\cdot(R(n)-r_1(n))\cdot R\left(\frac{|D|-n}{p^k}\right)$$
$$-\tfrac{1}{4}(\tfrac{1}{2}\cdot h\cdot \mathrm{ord}_\lambda(D_0)+\tfrac{1}{2}(h-2^{t-1})\mathrm{ord}_\lambda(D_1)).$$

Proof. 1. is a restatement of Theorem 4.6. 2. follows immediately from the theorem since ℓ is a degree 1 prime and also there is no need for the correction term. For 3. we need only remark that $\mathrm{ord}_p(I)=\frac{1}{2}\mathrm{ord}_\lambda(I)$ in this case. □

(5.7) **Corollary.** *With ℓ as above, if the class number is odd or equal to 2 then*

$$\mathrm{ord}_\ell(I)=\sum_{n\geqq 0}\sum_{k\geqq 1}\varrho_\ell(n)\cdot\tfrac{1}{2}\cdot(R(n)-r_1(n))\cdot R\left(\frac{|D|-n}{\ell^k}\right).$$

Proof. By the above corollary we only need to check the term $\mathrm{ord}_p(I)$ when $n=0$ for $p\,|\,D$.

Case 1. h odd. Remark, this is a theorem of Gross–Zagier [6]. Here $D=-p$ where $-p\equiv 1 \bmod 4$ and so $D_0=1$ and $D_1=p$. Moreover, $2^{t-1}=1$ so

$$-\tfrac{1}{4}(\tfrac{1}{2}\cdot \mathrm{h}\cdot\mathrm{ord}_\lambda(D_0)+\tfrac{1}{2}(h-2^{t-1})\cdot\mathrm{ord}_\lambda(D_1))=-\tfrac{1}{4}(\tfrac{1}{2}(h-1)).$$

Observe that $\frac{1}{4}\varrho_\ell(0)\cdot(R(0)-r_1(0))\cdot R\left(\frac{p}{p}\right)=\frac{1}{4}(\frac{1}{2}(h-1))$. So the term corresponding to 0 in the summation cancels against the correction term for the discriminant of $\mathbb{Q}(j)$ and the corollary is proved in this case.

Case 2. $h=2$. Then $D=-pq$ where $-p\equiv q\equiv 1 \bmod 4$. So $D_0=q$ and $D_1=p$. First consider the prime p. In this situation $\varrho_\ell(0)=0$ so $\frac{1}{4}\varrho_\ell(0)\cdot(R(0)-r_1(0))\cdot R(q)=0$. But note, since $D_1=p$ then $-\frac{1}{4}(\mathrm{ord}_\lambda(q)+\mathrm{ord}_\lambda(p))=0$ so the terms cancel. Finally consider q. Here $\varrho_\ell(0)=4$ so $\frac{1}{4}\varrho_\ell(0)\cdot(R(0)-r_1(0))\cdot R(p)=\frac{1}{2}$ and $-\frac{1}{4}(\mathrm{ord}_\lambda(q)+\mathrm{ord}_\lambda(p))=-\frac{1}{2}$. Again the terms cancel and the corollary is proved. □

6. A Computational Example

A simple example will suffice to reveal the computational techniques involved in Theorem 5.5.

Let $D=-15$. The complete diagram of fields is

$$\begin{tikzcd}[column sep=small] & \mathbb{Q}(\sqrt{-3},\sqrt{5})=H \arrow[dl, dash] \arrow[dr, dash] & \\ \mathbb{Q}(j)=\mathbb{Q}(\sqrt{5}) \arrow[dr, dash] & & \mathbb{Q}(\sqrt{-15}) \arrow[dl, dash] \\ & \mathbb{Q} & \end{tikzcd}$$

Thus $d(\mathbb{Q}(j))=D_0=5$ and $D_1=3$. The inert or ramified primes dividing $|D|-n$ with $0\leqq n\leqq |D|$ are 3, 5, 7, 11, and 13. 2 is the only split prime, consequently by (5.5) $\mathrm{ord}_2(I)=0$.

In the calculation of $\mathrm{ord}_3(I)$ and $\mathrm{ord}_5(I)$ we know the correction term is zero since we are in the class number 2 situation. Nevertheless, we will use Theorem 5.5 directly to illustrate how the computations are made.

Case $p=3$. Consider the table.

n	$\varrho_3(n)$	$R(n)$	$r_1(n)$	$R((15-n)/3^k)$
0	0	1	$\frac{1}{2}$	1
3	2	1	0	3
6	2	2	2	1 for $k=1$ or 2
9	2	2	2	2
12	2	3	0	1
15	0	1	1	1

Since $5\equiv 1 \bmod 4$ we have $\varrho_3(0)=\varrho_3(15)=0$. On the other hand we have $\varrho_3(3)=\varrho_3(6)=\varrho_3(9)=\varrho_3(12)=2$ since $\varepsilon_3((n-|D|)/n)\neq -1$ for at least two primes dividing D in these latter cases. Next observe that the correction term is 0 since $\mathrm{ord}_{\lambda_3}5=0$ and $\left(\frac{h-2^{t-1}}{2}\right)=0$. Note λ_3 is the unique prime of K over 3. Then from the table above and the formula in Theorem 5.5 we see

$$\mathrm{ord}_3(I)=0+\tfrac{1}{4}\cdot 2(3+0+0+0+3)=3\,.$$

Case $p=5$. Consider the table

n	$\varrho_5(n)$	$R(n)$	$r_1(n)$	$R((15-n)/5^k)$
0	4	1	$\frac{1}{2}$	1
5	1	1	0	2
10	1	2	2	1
15	2	1	1	1

In this case $\varrho_5(0)=4$ and $\varrho_5(5)=\varrho_5(10)=2$. The correction term in this case is $\frac{1}{4}(\mathrm{ord}_{\lambda_5}(5))=\frac{1}{2}$. λ_5 is the unique prime above 5. Then the table and (5.5) imply $\mathrm{ord}_5(I)=\frac{1}{4}(4\cdot\frac{1}{2}+2\cdot 1\cdot 2+0)-\frac{1}{2}=1$.

Case $p=7$. The table is

n	$\varrho_7(n)$	$R(n)$	$r_1(n)$	$R((15-n)/7^k)$
1	1	1	1	2
8	1	4	0	1

and $\mathrm{ord}_7(I)=\frac{1}{2}\cdot 4=2$.

Case $p=11$. Our table has one entry

n	$\varrho_{11}(n)$	$R(n)$	$r_1(n)$	$R((15-n)/11^k)$
4	1	3	3	1

Thus, $\mathrm{ord}_{11}(I)=\frac{1}{2}\cdot 0=0$.

Case $p=13$. Our table has one entry

n	$\varrho_{13}(n)$	$R(n)$	$r_1(n)$	$R((15-n)/13^k)$
2	1	2	0	1

Thus, $\mathrm{ord}_{13}(I)=\frac{1}{2}\cdot 2=1$.

So we find $I=3^3\cdot 5\cdot 7^2\cdot 13$.

7. Appendix

We now compute the absolute disriminant of $\mathbb{Q}(j)$ over $\mathbb{Q}$.

To simplify notation set $F=\mathbb{Q}(j)$ and for any extension of fields B/A let $d(B/A)$ denote the discriminant of that extension. If $A=\mathbb{Q}$ it will be dropped from the notation. All other notation will be as in the previous sections.

Consider the diagram of fields

$$\begin{tikzcd}[column sep=small]
& H \arrow[dr, dash, "h_2"] & & \\
F \arrow[ur, dash] \arrow[dr, dash] & & M \arrow[dr, dash, "2^{t-1}"] & \\
& M^+ \arrow[ur, dash] \arrow[dr, dash] & & K \\
& & \mathbb{Q} \arrow[ur, dash] &
\end{tikzcd}$$

Here $h_2=h/2^{t-1}$, M is the genus field of K, and M^+ is the composition of all the real quadratic subfields contained in H. This last condition is equivalent to the fact that M^+ is the totally real subfield of M, hence the $^+$ in the notation. Using the norm-discriminant formula the following equalities can be read from the diagram

$$d(F)^2\mathbb{N}_{F/\mathbb{Q}}(d(H/F))=d(H)=d(K)^h\mathbb{N}_{K/\mathbb{Q}}(d(H/K)).$$

Now $d(K)=D$ and $d(H/K)=1$ since H is unramified over K. Thus the above equations reduce to

$$d(F)^2\mathbb{N}_{F/\mathbb{Q}}(d(H/F))=D^h.$$

Hence we can compute $d(F)$ once we compute $\mathbb{N}_{F/\mathbb{Q}}(d(H/F))$.

Now consider the field M^+. That M^+ is actually contained in F is shown in Rohrlich [9] where he also proves

(7.1) **Proposition.** *The following are equivalent.*

a. *There are at least two primes congruent to* $3 \bmod 4$ *which divide* D;

b. *the extension* M/M^+ *is unramified outside infinity,*

c. *the extension* H/F *is unramified outside infinity.*

Recall $D = D_0 \cdot D_1$ where

$$D_1 = \begin{cases} 1 & \text{if at least 2 primes congruent to } 3 \bmod 4 \text{ divide } D \\ p & \text{if } p \text{ is the unique prime congruent to } 3 \bmod 4 \text{ dividing } D \\ 4 & \text{if } 4 \| D \text{ and no primes congruent to } 3 \bmod 4 \text{ divide } D \\ 8 & \text{if } 8 \| D \text{ and no primes congruent to } 3 \bmod 4 \text{ divide } D \end{cases}$$

and

(5.1). **Proposition.** $d(\mathbb{Q}(j)) = D_0^{1/2h} \cdot D^{1/2(h-2^{t-1})}$.

Proof. Case 1. Assume at least two primes congruent to $3 \bmod 4$ divide D. Rohrlich's proposition shows $d(H/F) = 1$, so $d(F) = D^{1/2h}$ as claimed.

Before embarking on cases 2 and 3 we make some observations. Let p be the unique prime *not* congruent to $3 \bmod 4$ dividing D. By class field theory p splits into $h/2$ prime factors in H, each of degree 2. Corresponding to each of these factors there is an inertia subgroup of $\mathrm{Gal}(H/\mathbb{Q})$ each of order 2. Since H/F is ramified at p at least one of these subgroups corresponds to $\mathrm{Gal}(H/F)$. Let $\langle\sigma\rangle$ be this subgroup. To determine $d(H/F)$ we need to see precisely how many of the other inertia subgroups are equal to $\langle\sigma\rangle$. By Galois theory this is equivalent to finding the order of the normalizer of $\langle\sigma\rangle$ in $\mathrm{Gal}(H/\mathbb{Q})$. We know $\mathrm{Gal}(H/\mathbb{Q}) = \mathrm{Gal}(H/K) \rtimes \langle\sigma\rangle$, with the action of σ given by $\sigma g \sigma^{-1} = g^{-1}$ for all $g \in \mathrm{Gal}(H/K)$. Thus the centralizer of σ in $\mathrm{Gal}(H/\mathbb{Q})$ – which is the same as the normalizer in this case – is

$$\mathscr{Z}(\sigma) = \{\sigma^n g : n = 0, 1 \text{ and } g \in \mathrm{Gal}(H/K),\ g^2 = 1\}.$$

Thus $\mathrm{Card}(\mathscr{Z}(\sigma)) = 2^t$. Thus the number of distinct quadratic subfields of H is $[H:\mathbb{Q}]/\mathrm{Card}(\mathscr{Z}(\sigma)) = 2h/2^t = h/2^{t-1}$. Consequently the number of ramified primes dividing the different $\mathscr{D}_{H/F}$ is $\dfrac{h/2}{h/2^{t-1}} = 2^{t-2}$.

Now onto cases 2 and 3 where we calculate the $\mathrm{ord}_{\mathfrak{p}_i}(\mathscr{D})$ for any of the primes $\mathfrak{p}_i$ dividing $\mathscr{D}^{H/F}$. We have the formula

$$\text{(Serre [10, p. 64]); } \mathrm{ord}_{\mathfrak{p}_i}(\mathscr{D}_{H/F}) = \sum_{n \geqq 0} \mathrm{Card}(\mathrm{Gal}(H/F)_n) - 1.$$

Where $\mathrm{Gal}(H/F)_n$ denotes the n^{th} higher ramification group of $\mathrm{Gal}(H/F)$.

Case 2. Let p be odd. Then H/F is tamely ramified at p so $\mathrm{Card}(\mathrm{Gal}(H/F)_n) = 1$ for $n \geqq 0$. Hence for all $\mathfrak{p}_i$

$$\mathscr{D}_{H/F} = \prod_{i=1}^{2^{t-2}} \mathfrak{p}_i.$$

Consequently $d(H/F) = \prod_{i=1}^{2^{t-2}} \mathfrak{p}_i$, for those $\mathfrak{p}_i | p$.

Since each $\mathfrak{p}_i$ is of degree 2 it follows that $\mathbb{N}_{F/\mathbb{Q}}(d_{H/F}) = (p^2)^{2^{t-2}} = p^{2^{t-1}}$. Thus $d_F^2 = D^h/p^{2^{t-1}}$ hence $d_F = D_0^{1/2h} \cdot p^{1/2(h-2^{t-1})}$ as claimed.

Remark. If $D = p$ we find $d_F = p^{(h-1)/2}$ which was obtained by Gross in [5].

Case 3. Let $p = 2$. Then H/F has wild ramification at 2 and we must compute the higher ramification groups. Since $\mathrm{Gal}(H/F)$ is a subgroup of $\mathrm{Gal}(H/\mathbb{Q})$ we have $\mathrm{Gal}(H/F)_n = \mathrm{Gal}(H/\mathbb{Q})_n \cap \mathrm{Gal}(H/F)$. So the question is reduced to computing the higher ramification groups for $\mathrm{Gal}(H/\mathbb{Q})$. These can be computed using Herbrand's Theorem from the knowledge of $\mathrm{Gal}(H/K)_n$ and $\mathrm{Gal}(K/\mathbb{Q})_n$.

Let $\mathscr{G}=\mathrm{Gal}(H/\mathbb{Q})$, $\mathscr{H}=\mathrm{Gal}(H/K)$, and $G=\mathrm{Gal}(K/\mathbb{Q})$. Since H/K is unramified $\mathscr{H}_n=0$ for $n\geqq 0$. If $4\|D$ then $G=G_0=G_1 \underset{2}{\supset} G_2=0$. And if $8\|D$ the $G=G_0=G_1=G_2 \underset{2}{\supset} G_3=0$. Following the notation of Serre [10, p. 73] we have the transition formula

$$\varphi_{H/K}(u)=\int_0^u (\mathscr{H}_0:\mathscr{H}_t)^{-1}\mathrm{d}t=\int_0^u 1\,dt=u.$$

So in particular $\varphi_{H/K}(n)=n$ for any integer $n\geqq 0$. Herbrand's Theorem gives $G_n=(\mathscr{G}/\mathscr{H})_n=(\mathscr{G}_n\mathscr{H})/\mathscr{H}$. So $\mathrm{Gal}(H/\mathbb{Q})_n=\mathrm{Gal}(K/\mathbb{Q})_n$ for all $n\geqq 0$. Using this fact it is easy to see that

$$\mathscr{D}_{H/F}=\prod_{i=1}^{2^{t-2}} \mathfrak{p}_i^a;\quad \text{where}\quad a=\begin{cases}2 & \text{if } 4\|D\\ 3 & \text{if } 8\|D\end{cases}.$$

A computation similar to that in case 2 reveals $d(\mathbb{Q}(j))=D_0^{1/2h}\cdot D_1^{1/2(h-2^{t-1})}$ as claimed. □

Acknowledgements. It is a pleasure to thank Benedict H. Gross for generously sharing his ideas with me and his encouragement during this investigation.

References

1. Borel, A., Chowla, S., Herz, C.S., Iwasawa, K., Serre, J.-P.: Seminar on Complex Multiplication. (Lecture Notes in Math. Vol. 21). Berlin Heidelberg New York: Springer 1966
2. Deuring, M.: Die Typen der Multiplikatorenringe elliptischer Funktionenkörper. Abh. Math. Semin. Hamb. **14**, 197–272 (1941)
3. Dorman, D.R.: Global orders in definite guaternion algebras as endomorphism rings for reduced CM elliptic curve. Proceedings of Conférence internationale de theorie des nombres (to appear)
4. Dorman, D.R.: Special values of the elliptic modular function and factorization formulae. J. Reine Angew. Math. **383**, 207–220 (1988)
5. Dummit, D.S., Gold, R., Kisilevsky, H.: The field generated by the discriminant of the class invariants of an imaginary quadratic field. Can. Math. Bull. **26**, 280–282 (1983)
6. Gross, B.H.: Arithmetic on elliptic curves with complex multiplication. (Lectures Notes in Math. Vol. 776). Berlin Heidelberg New York: Springer 1980
7. Gross, B.H., Zagier, D.B.: Heegner points and derivatives of L-Series. Invent. Math. **84**, 225–320 (1986)
8. Gross, B.H., Zagier, D.B.: On singular moduli. J. Reine Angew. Math. **355**, 191–220 (1985)
9. Rohrlich, D.E.: Elliptic curves with good reduction everywhere. J. Lond. Math. Soc. **25**, 216–222 (1982)
10. Serre, J.-P.: Local fields. Grad. Texts in Math. 67. Berlin Heidelberg New York: Springer 1979
11. Serre, J.-P., Tate, J.T.: Good reduction of abelian varities. Ann. Math. **88**, 492–517 (1968)
12. Vignéras, M.-F.: Arithmétique des algèbres de quaternions. (Lecture Notes in Math. Vol. 800). Berlin Heidelberg New York: Springer 1980

Received June 16, 1987

Math. Ann. 283, 193–202 (1989)

Mathematische Annalen

Complemented Infinite Type Power Series Subspaces of Nuclear Fréchet Spaces

A. Aytuna[1], J. Krone[2], and T. Terzioğlu[1]

[1] Department of Mathematics, Middle East Technical University, Ankara, Turkey
[2] Fachbereich Mathematik, Bergische Universität, D-5600 Wuppertal, Federal Republic of Germany

The space of rapidly decreasing sequences s plays a prominent role in the theory of nuclear Fréchet spaces. In this article, we prove among other things, that if s is isomorphic to a subspace of a nuclear Fréchet space E, then E has a complemented subspace isomorphic to s. We note that we do not assume the existence of a basis in E. In fact, with the additional assumption that E has a strong finite dimensional decomposition, Holmström [8] obtained the same result.

We deal with the problem in a more general setting by assuming that there is what we call a local imbedding $i:\lambda(A)\rightarrow E$ and an imbedding $j:E\rightarrow\lambda(A)^{\mathbb{N}}$, where $\lambda(A)$ is a stable, nuclear G_∞-space. In this setting we prove that E has a complemented subspace which is isomorphic to $\lambda(A)$. By Vogt's characterization of subspaces of s [23], any subspace of s, which has a regular basis, can be expressed as a nuclear G_∞-space.

Pelczynski's decomposition method [13] has been adopted by Vogt [24] to apply to nuclear, stable power series spaces, so that if a Fréchet space E is isomorphic to a complemented subspace of a stable nuclear power series space $\Lambda_\infty(\alpha)$ and $\Lambda_\infty(\alpha)$ in turn isomorphic to a complemented subspace of E, one concludes that E and $\Lambda_\infty(\alpha)$ are in fact isomorphic. This powerful method has been used extensively by Vogt in [24, 25]. As a corollary of our result we improve this method so that one can reach the same conclusion by only requiring that there is a local imbedding of $\Lambda_\infty(\alpha)$ into E and, as before, that E is isomorphic to a complemented subspace of $\Lambda_\infty(\alpha)$.

By the well-known Komura-Komura imbedding theorem [9], every nuclear Fréchet space is isomorphic to a subspace of $s^{\mathbb{N}}$. Even in the case of a nuclear, stable G_∞-space, the existence of an imbedding $j:E\rightarrow\lambda(A)^{\mathbb{N}}$ can be expressed in terms of the diametral dimension simply as $\Delta(\lambda(A))\subset\Delta(E)$ [16, 17]. In order to apply our version of the decomposition method effectively, in the second section we deal with the problem of the existence of a local imbedding $i:\lambda(A)\rightarrow E$. In [14] Pelczynski asked whether a complemented subspace E of a nuclear Köthe space has a basis. In its generality this is still an open problem. For a complemented subspace E of s, Wagner [31] has proved that if E is isomorphic to $E\times E$, then it has a basis. As an important application of the improved decomposition method, we show that if E

and $E \times E$ have equal diametral dimensions, then E is isomorphic to a power series space $\Lambda_\infty(\alpha)$. For other positive answers to Pelczynski's problem in the case of complemented subspaces of s, we refer to [5] and [11]. Using our result, it is a simple matter to conclude that the space of analytic functions $O(M)$ on a Stein manifold M of dimension d has property (DN) of Vogt [23] if and only if $O(M)$ is isomorphic to the space of entire functions $O(\mathbb{C}^d)$.

0

We use the standard terminology and notation of the theory of locally convex spaces as in [10]. For nuclear spaces we refer to [15]. Throughout E will denote a Fréchet space over the real $\mathbb{R}$ or the complex field $\mathbb{C}$ with a fundamental sequence of seminorms $\| \ \|_0 \leqq \| \ \|_1 \leqq \dots$ and $U_k = \{x \in E : \|x\|_k \leqq 1\}$. $L(E, F)$ is the space of continuous linear maps from E into F and $B(E, F)$ the closed unit ball of $L(E, F)$ provided E and F are Banach spaces.

A Fréchet space E is said to have property (DN) if it has a fundamental sequence of seminorms such that for each k there is a p and $C > 0$ with $\|x\|_k \leqq r \|x\|_0 + (C/r) \|x\|_p$ for all $x \in E$ and $r > 0$. A nuclear Fréchet space has (DN) if and only if it is isomorphic to a subspace of the space of rapidly decreasing sequences s [23]. E has property (Ω) if for every p there is a q such that for every k there is a j and $C > 0$ with

$$U_q \subset C r^j U_k + \frac{1}{r} U_p$$

for all $r > 0$. For nuclear Fréchet spaces, the condition (Ω) characterizes quotient spaces of s [29].

The diametral dimension $\Delta(E)$ of E is the set of all sequences (ξ_n) such that for every k there is a p with $\lim \xi_n d_n(U_p, U_k) = 0$, where $d_n(U_p, U_k)$ denotes the n-th Kolmogorov diameter of U_p with respect to U_k [15, 20]. For the calculation of $\Delta(E)$ in case E has (DN) or (Ω) we shall refer to [22]. In particular if a nuclear Fréchet space E has (DN) and (Ω), for $p = 0$ we find q_0 as in the (Ω) condition and set $\alpha_n = -\log d_n(U_{q_0}, U_0)$. We then have $\Delta(E) = \Delta(\Lambda_\infty(\alpha))$ [22, *Sect.* 3, (2)].

A Köthe space $\lambda(A)$ which satisfies the following conditions is called a G_∞-space [4, 20]

(1) $a_n^0 = 1$ and $a_n^k \leqq a_{n+1}^k$

(2) for every k there is a p with $((a_n^k)^2 / a_n^p) \in \ell_\infty$.

Power series spaces of infinite type are certainly the best known examples of G_∞-spaces. $\lambda(A)$ is nuclear if and only if $(1/a_n^k) \in \ell_1$ for some k [20]. Any subspace of s, which has a regular basis, can be expressed as a G_∞-space [23] (cf. also [18]). A Fréchet space E is called *stable* if it is isomorphic to $E \times E$. Stability of a G_∞-space is simply equivalent to: for each k there is a p with $(a_{2n}^k / a_n^p) \in \ell_\infty$ [21]. We shall use the following equivalent fundamental sequence of norms for a nuclear G_∞-space:

$$|x|_k = \sum_{n=1}^{\infty} |x_n| a_n^k \quad \text{and} \quad \|x\|_k = \left(\sum_{n=1}^{\infty} (|x|_n a_n^k)^2 \right)^{1/2}.$$

The diametral dimension is a complete isomorphic invariant for the class of nuclear G_∞-spaces, since $\Delta(\lambda(A))=\lambda(A)'$ [20].

1

A continuous linear map $i:\lambda(A)\to E$ will be called a *local imbedding* if there is a continuous seminorm $\|\ \|$ on E such that $|x|_0\leqq\|ix\|$ holds. Certainly an imbedding of $\lambda(A)$ is a local imbedding and a local imbedding is one-to-one. The map which sends each $x\in\Lambda_\infty(\alpha)$ to $(x_nR^{\alpha_n})$ for some fixed $R>1$ is a local imbedding of $\Lambda_\infty(\alpha)$ into the finite type power series space $\Lambda_1(\alpha)$ and it is easily seen to be compact. However, a power series space $\Lambda_\infty(\alpha)$ of infinite type is not necessarily isomorphic to a subspace of $\Lambda_1(\alpha)$. Even this simple example shows that a local imbedding can be quite different than an imbedding.

If E is isomorphic to a subspace of a nuclear locally convex space F, we have $\Delta(F)\subset\Delta(E)$ [20]. Here F need not be metrisable. Our first result indicates that in terms of diametral dimension a local imbedding has the same effect as an imbedding.

1.1. Proposition. *If there is a local imbedding of a nuclear G_∞-space $\lambda(B)$ into a nuclear locally convex space F, then $\Delta(F)\subset\Delta(\lambda(B))$.*

Proof. Let $i:\lambda(B)\to F$ be a local imbedding. We have $|x|_0\leqq\|ix\|$ for some continuous semi-norm $\|\ \|$ on F. Let $U=\{y\in i(\lambda(B)):\|y\|\leqq 1\}$, and $(\xi_n)\in\Delta(F)$. Since $\Delta(F)\subset\Delta(i(\lambda(B))$ there is a neighborhood V of $i(\lambda(B))$ with $\lim\xi_n d_n(V,U)=0$. By continuity of i we find k and $C_k>0$ with $P_V(ix)\leqq C_k|x|_k$, $x\in\lambda(B)$. Now if $V\subset dU+L$ where L is a subspace of $i(\lambda(B))$ with dimension not exceeding n and $d>0$, using the fact that i is $1-1$, we find a subspace $\tilde{L}$ of $\lambda(B)$ of dimension not exceeding n, $i(\tilde{L})=L$ and so get

$$U_k\subset C_k dU_0+\tilde{L}.$$

Hence

$$(b_n^k)^{-1}=d_n(U_k,U_0)\leqq C_k d_n(V,U)$$

and therefore $(\xi_n)\in\lambda(B)'=\Delta(\lambda(B))$.

Throughout the rest of this section we let $\lambda(A)$ stand for a *stable, nuclear G_∞-space.* We note that even in case of an imbedding $i:\lambda(A)\to\lambda(A)$, it may happen that $i(\lambda(A))$ is not complemented. For example, we have an exact sequence

$$0\longrightarrow s\xrightarrow{\ i\ }s\longrightarrow s^{\mathbb{N}}\longrightarrow 0$$

[23] and here $i(s)$ certainly not complemented in s.

For the canonical basis (e_n^j) of $\lambda(A)^{\mathbb{N}}$ we have

$$\|e_n^j\|_k=\begin{cases}a_n^k & \text{for } j\leqq k\\ 0 & \text{for } j>k,\end{cases}$$

where $(\|\ \|_k)$ is the sequence of standard Hilbertian seminorms on the product space $\lambda(A)^{\mathbb{N}}$. With the bijection $\beta:\mathbb{N}\times\mathbb{N}\to\mathbb{N}$ defined by $\beta(j,n)=2^{j-1}(2n-1)$ we set $f_{\beta(j,n)}=e_n^j$. Hence $\|f_\nu\|_k\leqq a_\nu^k$. Using the stability of $\lambda(A)$ we find r_k, $D_k>0$ with

$$a_{2^k\nu}^k\leqq D_k a_\nu^{r_k}.$$

So if $\|f\|_k \neq 0$, then $f_\nu = e_n^j$ with $j \leqq k$ and

$$a_\nu^k = a_{2^{j-1}(2n-1)}^k \leqq D_k a_n^{r_k} = D_k \|f_\nu\|_{r_k}.$$

Hence we either have $\|f_\nu\|_k = 0$ or $a_\nu^k \leqq D_k \|f_\nu\|_{r_k}$. These simple observations will be used in the proof of the following result which is crucial in the subsequent development.

1.2. Proposition. *Let $i: \lambda(A) \to \lambda(A)^{\mathbb{N}}$ be a local imbedding. Then there is a complemented subspace G of $\lambda(A)^{\mathbb{N}}$ which is isomorphic to $\lambda(A)$ and contained in $\overline{i(\lambda(A))}$.*

Proof. For each k, we find $m(k)$, $C_k > 0$ so that for all $x \in \lambda(A)$ we have

$$\|ix\|_k \leqq C_k |x|_{m(k)}$$

and

$$|x|_0 \leqq C_0 \|ix\|,$$

where $\|\ \ \|$ is a suitable semi-norm on the nuclear space $\lambda(A)^{\mathbb{N}}$ defined by a scalar product $(\ ,\ )$ and we may initially arrange things so that $\|y\| = \|y\|_{m(0)}$ holds for $y \in \lambda(A)^{\mathbb{N}}$. We note that $\|\ \ \|$ is in fact a norm on $i(\lambda(A))$. Let Q be the projection on $\lambda(A)^{\mathbb{N}}$ with $Q^{-1}(0) = \{y \in \lambda(A)^{\mathbb{N}}: \|y\| = 0\}$. By stability of $\lambda(A)$, the range of Q is isomorphic to $\lambda(A)$ itself.

We now construct a basic sequence (g_n) in $i(\lambda(A))$. We choose $g_1 \in \operatorname{sp}\{ie_1, ie_2\}$ with $(g_1, f_1) = 0$ and $\|g_1\| = 1$. We want to select g_n with the following properties:

(i) $g_n \in \operatorname{sp}\{ie_1, \ldots, ie_{2n}\}$.
(ii) $(g_n, f_\nu) = (g_n, g_j) = 0$ for $\nu = 1, \ldots, n$ and $j = 1, \ldots, n-1$.
(iii) $\|g_n\| = 1$.

Suppose we have already determined $g_1, \ldots, g_{n-1}$. Since $\operatorname{sp}\{ie_1, \ldots, ie_{2n}\}$ is a $2n$-dimensional space, we can find $\tilde{g}_n \neq 0$ in this space with $(\tilde{g}_n, f_\nu) = 0$ for $\nu = 1, \ldots, n$ and $(\tilde{g}_n, g_j) = 0$ for $j = 1, \ldots, n-1$. We simply let $g_n = (1/\|\tilde{g}_n\|)\tilde{g}_n$. Further if

$$g_n = i\left(\sum^{2n} \mu_j^n e_j\right)$$

we have

$$\sum^{2n} |\mu_j^n| = \left|\sum^{2n} \mu_j^n e_j\right|_0 \leqq C_0 \|g_n\| = C_0$$

and therefore

$$\|g_n\|_k \leqq C_k \left|\sum^{2n} \mu_j^n e_j\right|_{m(k)} \leqq C_k \sum^{2n} |\mu_j^n| a_j^{m(k)} \leqq C_k C_0 a_{2n}^{m(k)}.$$

Using the stability of $\lambda(A)$, for each k we determine s_k, $\varrho > 0$ such that the inequality

(iv) $\|g_n\|_k \leqq \varrho a_n^{s_k}$

holds for all $n \in \mathbb{N}$.

At this stage we note that we can replace g_n by Qg_n in (ii), (iii), and (iv), because $(Qg_n, y) = (g_n, y)$ for all $y \in \lambda(A)^{\mathbb{N}}$, and $\|Qg_n\|_k \leqq \|g_n\|_k$.

For $x=\sum x_\nu f_\nu \in \lambda(A)^{\mathbb{N}}$ we have

$$\begin{aligned}|(g_n, x)|\, \|g_n\|_k &\leqq \sum_{\nu>n} |(g_n, f_\nu)|\, |x_\nu|\, \|g_n\|_k \\ &\leqq \varrho \sum_{\nu>n} |x_\nu|\, \|f_\nu\|_{m(0)} a_n^{s_k} \\ &\leqq \varrho \sum_{\nu>n} |x_\nu|\, \|f_\nu\|_{m(0)} a_\nu^{s_k}.\end{aligned}$$

For n fixed, let $I(j)=\{\nu:\nu>n, \|f_\nu\|_j \neq 0\}$. From above we get the estimate

$$|(g_n, x)|\, \|g_n\|_k \leqq \varrho \sum_{\nu\in I(m(0))} |x_\nu|\, a_\nu^{m(0)} a_\nu^{s_k}.$$

Using the fact that $\lambda(A)$ is a nuclear G_∞-space, we now determine $m_k > m(0)$, σ_k such that

$$\varrho a_n^{m(0)} a_n^{s_k} \leqq \sigma_k n^{-2} a_n^{m_k}$$

holds and from $I(m(0)) \subset I(m_k)$ obtain

$$|(g_n, x)|\, \|g_n\|_k \leqq \sigma_k n^{-2} \sum_{\nu\in I(m_k)} |x_\nu|\, a_\nu^{m_k}.$$

If we select r_k and $C>0$ such that $a_j^{m_k} \leqq C\|f_\nu\|_{r_k}$ whenever $\|f_\nu\|_{r_k} \neq 0$, the above estimate yields

(v) $|(g_n, x)|\, \|g_n\|_k \leqq \sigma_k C n^{-2} |x|_{r_k}$.

Hence we can define a continuous operator P on $\lambda(A)^{\mathbb{N}}$ by

$$Px = \sum_{n=1}^{\infty} (g_n, x) g_n.$$

We have $P(g_n)=g_n \in \overline{i(\lambda(A))}$ and so P is a projection. Its range G is contained in $\overline{i(\lambda(A))}$ and (g_n) is a basis of G.

To conclude the proof it remains to show that G is isomorphic to $\lambda(A)$. Since (Qg_n) satisfies (ii), (iii), (iv), and (v), we can define another projection P_0 on $\lambda(A)^{\mathbb{N}}$ by

$$P_0 x = \sum_{n=1}^{\infty} (Qg_n, x) Qg_n$$

so that the range G_0 of P_0 has (Qg_n) as a basis and it is contained in $Q(\lambda(A)^{\mathbb{N}})$, which is isomorphic to $\lambda(A)$. Hence, as a complemented subspace of $\lambda(A)$, G_0 is isomorphic to some G_∞-space $\lambda(B)$ [21, Theorem 3.1]. So $\lambda(A)' = \Delta(\lambda(A)) \subset \Delta(G_0) = \lambda(B)'$. Since $(Qg_n, y)=(g_n, y)$ for all $y\in\lambda(A)^{\mathbb{N}}$, P_0 and P have equal kernels and therefore G_0 is isomorphic to G. Hence it remains to show that G_0 is isomorphic to $\lambda(A)$. For this purpose we must prove $\lambda(B)' = \Delta(G_0) \subset \lambda(A)'$, but this is an immediate consequence of 1.1 Proposition.

We would like to note that if $i:\lambda(A)\to\lambda(A)^{\mathbb{N}}$ were to be an imbedding, then the space G would be contained in $i(\lambda(A))$ and one could show that it is isomorphic to $\lambda(A)$ itself without having to introduce G_0. It should also be noticed how extensively the stability of $\lambda(A)$ is used in the proof.

A Fréchet space E is said to be $\lambda(A,\mathbb{N})$-nuclear if $\lambda(A)' \subset \Delta(E)$. This generalization of $\Lambda_{\mathbb{N}}(\alpha)$-nuclearity [17] was introduced by Ramanujan and Rosenberger [16].

1.3. Theorem. *If E is $\lambda(A, \mathbb{N})$-nuclear and if there is a local imbedding of $\lambda(A)$ into E, then E has a complemented subspace isomorphic to $\lambda(A)$.*

Proof. Let $i: \lambda(A) \to E$ be a local imbedding and $j: E \to \lambda(A)^{\mathbb{N}}$ an imbedding, whose existence is equivalent to the $\lambda(A, \mathbb{N})$-nuclearity of E [16, 17]. Then $ji: \lambda(A) \to \lambda(A)^{\mathbb{N}}$ is a local imbedding and so by 1.2 Proposition there is a subspace G isomorphic to $\lambda(A)$ which is contained in $\overline{ji(\lambda(A))} = \overline{ji(\overline{\lambda(A)})} \subset j(E)$ and G is complemented in $\lambda(A)^{\mathbb{N}}$.

With the additional assumptions that E has a basis and there is an imbedding of $\lambda(A)$ into E, Holmström [7] reached the same conclusion as in 1.3 Theorem. In a subsequent work [8] he obtained the following corollary, by assuming E has a strong finite dimensional decomposition and a subspace isomorphic to s.

1.4. Corollary. *If there is a local imbedding of s into a nuclear Fréchet space E, then E has a complemented subspace which is isomorphic to s.*

Let $\Lambda_\infty(\alpha)$ be a stable nuclear power series space of infinite type. Complemented subspaces of $\Lambda_\infty(\alpha)$ have been characterized by Vogt and Wagner [30] as those $\Lambda_{\mathbb{N}}(\alpha)$-nuclear Fréchet spaces which have the properties (DN) and (Ω). Vogt's decomposition theorem [24] in this case states that if $\Lambda_\infty(\alpha)$ is isomorphic to a complemented subspace of E, where E is a complemented subspace of $\Lambda_\infty(\alpha)$, then E must be isomorphic to $\Lambda_\infty(\alpha)$. An immediate consequence of 1.3 Theorem is the following improvement of Vogt's decomposition method.

1.5. Corollary. *Let $\Lambda_\infty(\alpha)$ be nuclear and stable. Let E be a $\Lambda_{\mathbb{N}}(\alpha)$-nuclear Fréchet space with (DN) and (Ω). If there is a local imbedding of $\Lambda_\infty(\alpha)$ into E, then E is isomorphic to $\Lambda_\infty(\alpha)$.*

2

As we have already pointed out, the existence of an imbedding $j: E \to \lambda(A)^{\mathbb{N}}$ can be expressed in terms of the diametral dimensions of $\lambda(A)$ and E. In order to exploit the decomposition method given in the previous section more fully, we need to know more about the existence of local imbeddings. We start with an example. We believe that this can serve as a model for constructing a local imbedding into a space of functions. In fact, quite a number of Fréchet spaces of functions can be represented as power series spaces [19, 24, 25].

Let M be an irreducible Stein space of dimension d and let $T: \Lambda_\infty(n^{1/d}) \to O(\mathbb{C}^d)$ be an isomorphism [19]. We determine $C > 0$ and $R_0 > 1$ so that the inequality

$$|x|_0 \leqq C \sup\{|Tx(z)| : z \in R_0 \Delta^d\}$$

holds, where $R_0 \Delta^d$ is the polydisc in $\mathbb{C}^d$ around zero with multiradius R_0. Fix $R_1 > R_0$. Choose a regular point $\xi_0 \in M$ and find $f_i \in O(M)$. $i = 1, \ldots, d$, such that $f_i(\xi_0) = 0$ and $F = (f_1, \ldots, f_d): M \to \mathbb{C}^d$ has rank d at ξ_0 [6, p. 209]. By composing F with a linear transformation of $\mathbb{C}^d$ if necessary, we can determine a relatively compact neighborhood U_0 of ξ_0 such that F maps U_0 onto $R_1 \Delta^d$ biholomorphically. We define $F_*: O(\mathbb{C}^d) \to O(M)$ by $F_*(f)(\xi) = f(F(\xi))$, $\xi \in M$. F_* is in fact a continuous algebra homomorphism. Furthermore we have

$$\sup |f(z)| : z \in R_0 \Delta^d\} \leqq \sup\{|g(F(\xi))| : \xi \in U_0\}.$$

Hence $F_*T: \Lambda_\infty(n^{1/d}) \to O(M)$ is a local imbedding.

Let M be now a Stein manifold of dimension d which is always assumed to be *connected*. By the Oka-Cartan theorem [6], $O(M)$ has property (Ω) and $O(M)$ is isomorphic to a subspace of $\Lambda_1(n^{1/d})$ [2]. If $O(M)$ has property (DN), then it is already $\Lambda_{\mathbb{N}}(n^{1/d})$-nuclear [22, 30]. Now that we know the existence of a local imbedding $i: \Lambda_\infty(n^{1/d}) \to O(M)$, 1.5 Corollary yields the following result.

2.1. Proposition. *Let M be a Stein manifold of dimension d. The space of analytic functions $O(M)$ has property (DN) if and only if it is isomorphic to $O(\mathbb{C}^d)$.*

We note that this proposition was already proved in [2] under the additional assumption that $O(M)$ has a basis. In [34] Zaharjuta states that $O(M)$ is isomorphic to $O(\mathbb{C}^d)$ if it satisfies (DN), but in the discussion he also seems to assume that $O(M)$ has a basis. In the case of $d=1$, our proposition was proved in [1] and [11] by different methods. Zaharjuta [32] gave another characterization of $O(\mathbb{C})$. Aytuna and Vogt showed that $O(M)$ has property (DN) if and only if every bounded plurisubharmonic function on M is constant [1]. In case M is an algebraic variety, Mitiagin and Henkin [12] asked whether $O(M)$ is isomorphic to $O(\mathbb{C}^d)$. This was answered positively by Zaharjuta [33], Djakov and Mitiagin [3] and also by Vogt [24] as an application of the decomposition method.

A nuclear Fréchet space E is isomorphic to a complemented subspace of s if and only if it has the properties (DN) and (Ω) [29, 1.10 Satz]. In this case $\Delta(E)$ is equal to the diametral dimension of some $\Lambda_\infty(\alpha)$ [22]. Also, if a complemented subspace of s has a basis, it is isomorphic to some $\Lambda_\infty(\alpha)$ [29, 2.9 Satz]. However, whether a complemented subspace of a nuclear Köthe space has a basis, is an open problem which has been posed by Pelczynski [14]. Wagner [31] has proved that a complemented subspace E of s, which is stable, (i.e. E is isomorphic to $E\times E$), has a basis. Krone [11] reached the same conclusion under the assumption $\Delta(E\times E) = \Delta(E)$ and $\alpha_n \geqq n$ where $\Delta(E) = \Delta(\Lambda_\infty(\alpha))$. In contrast to some kind of stability which is assumed in these and in the following, Dubinsky and Vogt [5] (cf. also [27, 7.2, 7.3] have proved that a complemented subspace of an unstable power series space of infinite type always has a basis. We note that we can obtain 2.1 Proposition also as a corollary of the following theorem.

2.2. Theorem. *Let E be a nuclear Fréchet space with (DN) and (Ω). If $\Delta(E\times E) = \Delta(E)$, then E is isomorphic to some $\Lambda_\infty(\alpha)$.*

Proof. We have $\Delta(E) = \Lambda_\infty(\alpha)'$, $\Delta(E\times E) = \Delta(\Lambda_\infty(\alpha)\times\Lambda_\infty(\alpha))$ and so $\Lambda_\infty(\alpha)$ is stable. Since E is isomorphic to a complemented subspace of $\Lambda_\infty(\alpha)$ [30, 3.5 Satz], from 1.5 Corollary and the following lemma we reach the conclusion.

2.3. Lemma. *Let E be a nuclear space with (DN) and (Ω). If $\Delta(E) \subset \Delta(\lambda(B))$, then there is a local imbedding of the G_∞-space $\lambda(B)$ into E.*

Proof. Without loss of generality we may assume that E has a fundamental sequence of norms $\|\ \|_k$, where each $\|\ \|_k$ is defined by an inner product. Since $\lambda(B)$ has (DN) and E has (Ω), by various results of Vogt [28, 5.1 Theorem 3.3 Lemma, 3.4

Proposition] we have that for every μ there is an m such that for all k and $r>0$ the following holds:

$$L(\lambda(B), E_m) \subset L(\lambda(B), E_k) + rB(\lambda(B)_0, E_\mu).$$

In this condition, called $(\tilde{S}_1)$ by Vogt [26, 2.2 Theorem], E_k is the Hilbert space obtained by completing $(E, \| \ \|_k)$ and $\lambda(B)_0$ the completion of $(\lambda(B), | \ |_0)$. We choose integers (m_k) increasing to infinity with $m_0=0$ such that $(\tilde{S}_1)$ holds for $\mu=m_{k-1}, m=m_k, k=m_{k+1}$ and further, we may arrange things so that for every i we have a j, $C>0$ with

$$U_{m_1} \subset Cr^j U_i + \frac{1}{r} U_0, \ r>0.$$

To simplify notation we set $k=m_k$ and so we have

$$L(\lambda(B), E_k) \subset L(\lambda(B), E_{k+1}) + rB(\lambda(B)_0, E_{k-1})$$

and $\Delta(E) = \Lambda_\infty(\alpha)' = \Delta(\Lambda_\infty(\alpha))$ where $\alpha_n = -\log d_n(U_1, U_0)$. Since $(e^{\alpha_n}) \in \lambda(B)'$ $= \Delta(\lambda(B))$, for some j and $C>0$ the inequality $(1/d_n(U_1, U_0)) \leqq Cb_n^j$ holds for all n. By nuclearity of $\lambda(B)$ we find $C_0>0$ and k_0 with $|x|_0 \leqq C_0 \|x\|_{k_0}$. So there is a k_1 with $b_n^{k_0} \leqq Cb_n^{k_1} d_n(U_1, U_0)$. Since $\lim d_n(U_1, U_0)=0$, the linking map $\varrho_{1,0}: E_1 \to E_0$ is compact and hence it can be written in the form

$$\varrho_{1,0} y = \sum d_n(U_1, U_0)(y|f_n) g_n,$$

where (f_n) and (g_n) are orthonormal sequences in E_1 and E_0 respectively. We define $T_1: \lambda(B) \to E_1$ by

$$T_1 x = \sum b_n^{k_0} (d_n(U_1, U_0))^{-1} x_n f_n.$$

Then $\|T_1 x\|_1 \leqq C \|x\|_{k_1}$ and also

$$|x|_0 \leqq C_0 \|x\|_{k_0} = C_0 (\sum |x_n|^2 (b_n^{k_0})^2)^{1/2} = C_0 \|T_1 x\|_0.$$

We choose $\varepsilon_i > 0$ with $\sum \varepsilon_i \leqq (1/2C_0)$. By $(\tilde{S}_1)$ we choose $T_2: \lambda(B) \to E_2$ such that

$$\|T_1 x - T_2 x\|_0 \leqq \varepsilon_1 |x|_0.$$

Then

$$\left(\frac{1}{C_0} - \varepsilon_1\right) |x|_0 \leqq \|T_2 x\|_0.$$

Applying $(\tilde{S}_1)$ repeatedly, for each k we find $T_k \in L(\lambda(B), E_k)$ such that

$$\|T_k x - T_{k+1} x\|_{k-1} \leqq \varepsilon_k |x|_0$$

and

$$\left(\frac{1}{C_0} - \sum_{i=1}^{k-1} \varepsilon_i\right) |x|_0 \leqq \|T_k x\|_0$$

hold for all $x \in \lambda(B)$. So for each k we have a map $S_k \in L(\lambda(B), E_k)$ defined by $S_k x = \lim_i T_{k+i} x$ and $\varrho_{k+1,k} S_{k+1} = S_k$. Thus we have obtained a continuous linear map $T: \lambda(B) \to E$ such that $\varrho_k T = S_k$, where $\varrho_k: E \to E_k$ is the canonical inclusion.

Further T satisfies

$$\frac{1}{2C_0}|x|_0 \leqq \|Tx\|_0$$

and therefore it is a local imbedding.

References

1. Aytuna, A.: On the linear topological structure of spaces of analytic functions. Doğa Tr. J. Math. **10**, 46–49 (1986)
2. Aytuna, A., Terzioğlu, T.: On certain subspaces of a nuclear power series space of finite type. Stud. Math. **69**, 79–86 (1986)
3. Djakov, P.B., Mitiagin, B.S.: The structure of polynomial ideals in the algebra of entire functions. Stud. Math. **68**, 84–104 (1980)
4. Dragilev, M.M.: On regular bases in nuclear spaces. AMS Trans. **2**, 61–82 (1970)
5. Dubinsky, E., Vogt, D.: Bases in complemented subspaces of power series spaces. Bull. Pol. Acad. Sci. Math. **34**, 65–67 (1986)
6. Gunning, R., Rossi, H.: Analytic functions of several complex variables. Englewood Cliffs: Prentice Hall 1965
7. Holmström, L.: Superspaces of (s) with basis. Stud. Math. **75**, 139–152 (1983)
8. Holmström, L.: Superspaces of (s) with strong finite dimensional decomposition. Arch. Math. **42**, 58–66 (1984)
9. Komura, T., Komura, Y.: Über die Einbettung der nuklearen Räume in $(s)^A$. Math. Ann. **162**, 284–288 (1966)
10. Köthe, G.: Topologische lineare Räume. Berlin Göttingen Heidelberg: Springer 1960
11. Krone, J.: On projections in power series spaces and the existence of bases. To appear in Proc. AMS
12. Mitiagin, B.S., Henkin, G.M.: Linear problems of complex analysis. Russ. Math. Surv. **26**, 99–164 (1971)
13. Pelczynski, A.: Projections in certain Banach spaces. Stud. Math. **19**, 209–228 (1960)
14. Pelczynski, A.: Problem 37. Stud. Math. **38**, 476 (1970)
15. Pietsch, A.: Nuclear locally convex spaces. Berlin Heidelberg New York: Springer 1972
16. Ramanujan, M.S., Rosenberger, B.: On $\lambda(\phi, P)$-nuclearity. Compos. Math. **37**, 113–125 (1975)
17. Ramanujan, M.S., Terzioğlu, T.: Power series spaces $\Lambda_k(\alpha)$ of finite type and related nuclearities. Stud. Math. **53**, 1–13 (1975)
18. Robinson, W.: Some equivalent classes of Köthe spaces. Rocz. Pol. Tow. Math. Ser. I, Mat. Prace **20**, 449–451 (1978)
19. Rolewicz, S.: On spaces of holomorphic functions. Stud. Math. **21**, 135–160 (1962)
20. Terzioğlu, T.: Die diametrale Dimension von lokalkonvexen Räumen. Collect. Math. **20**, 49–99 (1969)
21. Terzioğlu, T.: Stability of smooth sequence spaces. J. Reine Angew. Math. **276**, 184–189 (1975)
22. Terzioğlu, T.: On the diametral dimension of some classes of F-spaces. J. Karadeniz Uni. Ser. Math.-Phys. **8**, 1–13 (1985)
23. Vogt, D.: Charakterisierung der Unterräume von s. Math. Z. **155**, 109–117 (1977)
24. Vogt, D.: Ein Isomorphiesatz für Potenzreihenräume. Arc. Math. **38**, 540–548 (1982)
25. Vogt, D.: Sequence space representations of spaces of test functions and distributions. Advances in functional analysis, holomorphy and approximation theory. G.I. Zapata (ed.). Lect. Notes Pure Appl. Math. Vol. 83. 405–443. New York 1983
26. Vogt, D.: Some results on continuous linear maps between Fréchet spaces. Functional Analysis: Surveys and recent results. III. Bierstedt, K.D., Fuchssteiner, B. (eds). pp. 349–381. North-Holland Math. Studies 90 (1984)
27. Vogt, D.: Kernels of Eidelheit matrices and related topics. Doğa Tr. J. Math. **10**, 232–236 (1986)

28. Vogt, D.: On the functors $\mathrm{Ext}^1(E, F)$ for Fréchet spaces. Stud. Math. **85**, 163–197 (1987)
29. Vogt, D., Wagner, M.J.: Charakterisierung der Quotientenräume von s und eine Vermutung von Martineau. Stud. Math. **67**, 225–240 (1979)
30. Vogt, D., Wagner, M.J.: Charakterisierung der Unterräume und Quotientenräume der nuklearen stabilen Potenzreihenräume von unendlichem Typ. Studia Math. **70**, 63–80 (1981)
31. Wagner, M.J.: Stable complemented subspaces of (s) have a basis. Seminar Lecture. A-G Funktionalanalysis Düsseldorf/Wuppertal, 1985
32. Zaharjuta, V.P.: Spaces of functions of one variable, analytic in open sets and on compacta. Math. USSR Sb. **11**, 75–88 (1970)
33. Zaharjuta, V.P.: Spaces of analytic functions on algebraic varieties in $\mathbb{C}^n$ (Russian). Izv. Sev.-Kavk. Nauchn. Zentra Vysh. Scholy **5** (1977)
34. Zaharjuta, V.P.: Isomorphism of spaces of analytic functions. Sov. Math. Dokl. **22**, 631–634 (1980)

Received January 1, 1988

Math. Ann. 283, 203–209 (1989)

Mathematische Annalen

A Measure for Semialgebraic Sets Related to Boolean Complexity

Gilbert Stengle

Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA

1. Introduction and Basic Definitions

Let X be a real affine or spherical semialgebraic set over a real closed field R. By spherical set we mean a subset of the manifold of the rays through the origin in the affine space R^{N+1}, or equivalently a subset of $R^{N+1}\backslash\{0\}/R^{+}$, defined by homogeneous polynomial relations. It is clear that in this latter space, using equivalence by positive homogeneity, we can define semialgebraic sets by systems of homogeneous polynomial equations and inequalities just as we define projective varieties by systems of homogeneous equations. Spherical geometry enjoys advantages not to be found in affine or projective spaces and in this paper we will show, using natural mappings between spherical and affine spaces, how it can be applied to affine geometry as well. As an application (Proposition 5.1 below) we demonstrate lower bounds for the number of inequalities necessary to define certain affine semialgebraic sets. Our key ingredients are a coarse, highly structured measure for the complexity of affine or spherical semialgebraic sets, a complexity reducing operator for spherical sets, and several complexity-monotonic mappings between spherical and affine spaces.

We introduce our complexity measure and an associated filtered algebraic structure for semialgebraic subsets of X in the following sequence of definitions.

Definition 1.1. Let X be a semialgebraic set. Let $(\gamma(X), +, \cdot)$ be the Boolean algebra of all semialgebraic subsets of X equipped with the operations $Y_1+Y_2=(Y_1 \cup Y_2)\backslash(Y_1 \cap Y_2)$ and $Y_1 \cdot Y_2=Y_1 \cap Y_2$.

This has the structure of a Z_2-algebra well known in measure theory (see Halmos [H]) in which addition is symmetric difference, ϕ is the additive identity and X is the multiplicative identity. For the moment it merely expresses the Boolean structure of $\gamma(X)$ in an alternate notation.

Definition 1.2. Let X be an affine (spherical) semialgebraic set. Let $\mathbb{U}(X)$ be the collection of all principal basic open subsets of X of the form $U=\{g>0\}\cap X$ where g is a polynomial (homogeneous polynomial).

Since the collection $\mathbb{U}(X)$ generates $\gamma(X)$ as a ring, any semialgebraic subset Y of X can be represented as a polynomial in the elements of $\mathbb{U}$. In terms of these representations one can define various measures of the complexity of Y: the minimum number of generators, the miniumum multiplicative or additive complexity among all polynomials representing Y, and so forth. From among these we choose a less obvious notion which, as we will show, is well suited to the semialgebraic category.

Definition 1.3. Let γ_0 be the subring of $\gamma = \gamma(X)$ generated by the algebraic subsets of X.

Obviously $\gamma = \gamma_0[\mathbb{U}]$, that is, every semialgebraic set can be written as a polynomial P in $\gamma_0[\mathbb{U}]$. This allows the following notion of complexity.

Definition 1.4. For $S \in \gamma$ let $\kappa(S)$, the complexity of S, be the minimum total degree of any polynomial representing S as an element of $\gamma_0[\mathbb{U}]$. Let γ_k be the γ_0-module of semialgebraic subsets of X of complexity not exceeding k. For $k<0$ let γ_k be the trivial ring $\{X, \phi\}$.

Thus the lowest level of complexity, complexity 0 in our scale, is represented by sets requiring for their definition, in addition to the defining relations of X, only equations or inequations but no true inequalities.

The author is indebted to Ludwig Bröcker for reading a preliminary version of this paper and suggesting many improvements.

2. Elementary Properties of Complexity

Lemma 2.1. *For $Y, Z, W \in \gamma(X)$*

1. $\kappa(Y^c) = \kappa(Y)$,
2. $\kappa(YZ) \leqq \kappa(Y) + \kappa(Z)$,
3. $\kappa(Y \cup Z) \leqq \kappa(Y) + \kappa(Z)$,
4. *If* $Y \cap Z = \phi$, *then* $\kappa(Y \cup Z) \leqq \max\{\kappa(Y), \kappa(Z)\}$ *and if* $\kappa(Y) \neq \kappa(Z)$, *then equality holds,*
5. $\kappa(Y) = 0 \Leftrightarrow Y \in \gamma_0$,
6. *W algebraic* $\Rightarrow \kappa(WY) \leqq \kappa(Y)$.

Proof. 1. Since $Y^c = Y + X$, if P represents Y, then $P + X$ is a polynomial of the same degree representing Y^c.

2. An obvious property of the total degree of a polynomial over any ring.

3. This follows from 1 and 2 by DeMorgan's law.

4. If $Y \cap Z = \phi$ then $Y \cup Z = Y + Z$ and the inequality again expresses an obvious property of total degree of a polynomial. If inequality holds, say $\kappa(Y) = k$ but $\kappa(Z) < k$, then $Y + Z \in \gamma_{k-1}$ implies $Y = Y + Z + Z \in \gamma_{k-1}$, a contradiction.

5 and 6 follow directly from the definitions.

A few examples illustrate the character of the function κ.

Example 2.2. In R^3 let $Y = \{x_1^2 + x_2^2 + x_3^2 < 1\} \setminus \{x_1 > 0,\ x_2 > 0,\ x_3 = 0\}$. Then $\kappa(Y) = 2$. But $\kappa(\{x_1^2 + x_2^2 + x_3^2 \leqq 1\}) = 1$.

This example shows that complexity is not a generic property. It also shows that the operation of topological closure can reduce complexity. On the other hand it can easily happen that a set consists of disjoint pieces of low complexity, the closures of which intersect in a more complicated way, so that its topological closure has greater complexity.

Example 2.3. In R^3 let $Y=\{x_3\neq 0\}\{x_1>0\}+\{x_3=0\}\{x_2>0\}$. Then

$$\{x_3=0\}\bar{Y}=\{x_3=0\}[\{x_1\geqq 0\}\cup\{x_2\geqq 0\}]=\{x_3=0\}+\{x_3=0\}\{x_1<0\}\{x_2<0\}$$

which has complexity 2. Hence, by Lemma 2.1.6, $\kappa(\bar{Y})\geqq 2>1=\kappa(Y)$.

Proposition 2.4. *For* $Y\in\gamma(R^N)(\gamma(S^N))$, $\kappa(Y)\leqq\dim(Y)\,(\dim(Y)+1)$.

Proof. By induction on $d=\dim(Y)$. If $d=0$ the conclusion is true in the affine case. However in the spherical case the estimate $\kappa(Y)\leqq 1$ cannot be improved since, even if $N=0$, we still need one inequality to separate a point from its antipode. More generally, since the decomposition $Y=Y\{x_0>0\}+Y\{x_0<0\}+Y\{x_0=0\}$ represents any Y by summands, the first two of which can be identified with affine sets and a third which is a subset of S^{N-1}, it follows by induction on N that the asserted affine estimate implies the spherical. If Y is affine it follows by results of Bröcker [B1] [B2] using the abstract theory of the stability index that Y is generically a disjoint union of basic open sets defined by no more than d inequalities. That is, there is an algebraic subset W of dimension less than d such that $\kappa(YW^c)\leqq d$. Applying Lemma 2.1 and the induction hypothesis to the decomposition $Y=YW+YW^c$ completes the proof.

Corollary 2.5. *If X is affine then the ascending sequence $\gamma_k(X)$, $k=0,1,2,\ldots$ gives a filtration of $\gamma(X)$ of length no greater than* $\dim(X)$.

3. An Operation Which Reduces Complexity

Definition 3.1. Let ϱ denote the involution $x\to -x$ of S^N. Let X be a ϱ-invariant semialgebraic set. Denote the operation induced on $\gamma(X)$ by $Y\to Y^\varrho$. Define $\delta\colon\gamma(X)\to\gamma(X)$ by $\delta Y=Y+Y^\varrho$.

From now on we will assume, unless we specify otherwise, that X is a ϱ-invariant subset of S^N.

It is plain from our definitions that $\kappa(\delta Y)\leqq\kappa(Y)$. Moreover it is easy to see that the raw number of Boolean operations needed to define a set is, in general, not reduced by the action of δ. However the following proposition shows that our complexity is strictly reduced by this action. We note that ϱ also acts on R^N but seems not to enjoy any useful properties there. Thus this proposition depends critically upon using spherical rather than affine geometry and upon using our complexity measure.

Proposition 3.2. 1. $\delta^2=\phi$.

2. $\delta\gamma_k(X)\subset\gamma_{k-1}(X)$.

Proof. 1. $\delta^2 Y=Y+Y^\varrho+(Y+Y^\varrho)^\varrho=Y+Y+Y^\varrho+Y^\varrho=\phi$.

2. By induction on k. For $k=0$ any element Y of γ_0 is ϱ-invariant and $\delta Y = Y + Y = \phi \in \gamma_{-1}$. For $k>0$, since δ is an γ_0-homomorphism of modules, it suffices to establish the property for basic open sets defined by k inequalities. Any such set has the form $Y = Z\{g>0\}$ where $Z \in \gamma_{k-1}$. If g is an even form then $\delta Y = \{g>0\}(Z+Z^\varrho\}$ which, by induction hypothesis, belongs to $\{g>0\}\gamma_{k-2} \subset \gamma_{k-1}$. If g is an odd form then

$$\{g>0\}^\varrho = \{g<0\} = X + \{g>0\} + \{g=0\}$$

and

$$\begin{aligned}\delta Y &= Z\{g>0\} + YZ^\varrho[X + \{g>0) + \{g=0\}]\\ &= \{g>0\}(Z+Z^\varrho) + Z^\varrho X + Z^\varrho\{g=0\}.\end{aligned}$$

Again by induction hypothesis this implies

$$\delta Y \subset \gamma_1\gamma_{k-2} + \gamma_{k-1}\gamma_0 + \gamma_{k-1}\gamma_0 \subset \gamma_{k-1}.$$

Remark 3.3. Let $h(X)$ be the length of the filtration $[\gamma_k(X)]$ of $\gamma(X)$. Then the sequence

$$\phi \hookrightarrow \delta\gamma_h \hookrightarrow \gamma_h \xrightarrow{\delta} \gamma_{h-1} \xrightarrow{\delta} \dots \xrightarrow{\delta} \gamma_0 \xrightarrow{\delta} \phi$$

(3.1) forms a complex.

Proposition 3.4. *The sequence* (3.1) *is exact.*

Proof. By induction on the dimension N of the ambient sphere. For $N=0$ (S^0 consists of two points) the assertion is trivial. For $N>0$ suppose that $Y \in \gamma_k$ and $\delta Y = \phi$ or, equivalently, $Y^\varrho = Y$. Then

$$\begin{aligned}Y &= Y\{x_N \neq 0\} + Y\{x_N = 0\}\\ &= Y\{x_N>0\} + Y^\varrho\{x_N<0\} + Y\{x_N=0\}\\ &= \delta(Y\{x_N>0\}) + Y\{x_N=0\}.\end{aligned}$$

The set $Y\{x_N=0\}$ can be identified with a subset Y' of S^{N-1} satisfying $\delta Y' = \phi$. By induction $Y' = \delta Z'$. Morover Z' can be identified with the cylindrical subset defined by the same relations in S^N. Then

$$\begin{aligned}Y &= \delta(Y\{x_N>0\}) + \{x_N=0\}\delta Z'\\ &= \delta[Y\{x_N>0\} + \{x_N=0\}Z'] \in \delta\gamma_{k+1}.\end{aligned}$$

4. Some Mappings Between Affine and Spherical Sets

We next describe some mappings useful in applying the structure $\{\gamma_k, \delta\}$ to affine geometry.

Definition 4.1 (lifting from S^N to R^{N+1}). For $Y \subset S^N$ let $Y^* \subset R^{N+1} \setminus \{0\}$ be the union of the open rays in R^{N+1} parameterized by the points of Y.

Observation 4.2.

1. The mapping $Y \to Y^*$ gives an injection of the semialgebraic subsets of S^N into $R^{N+1} \setminus \{0\}$. If $Y \subset S^N$ is defined by any set $\mathbb{P}$ of homogeneous polynomial relations then Y^* consists of the nonzero solutions of $\mathbb{P}$ regarded as relations in R^{N+1}.
2. $\kappa(Y) = \kappa(Y^*)$.

Definition 4.3 (restriction from R^N (S^N) to $R^{N-1}(S^{N-1})$). For $Y \subset R^N(S^N)$ and λ a nonconstant affine function (linear form), let $Y|_{\lambda=0}$ be the image of the restriction of Y to the hyperplane (hypersphere) $\{\lambda=0\}$ under some fixed isomorphism of $\{\lambda=0\}$ with $R^{N-1}(S^{N-1})$.

Observation 4.4.

1. If Y is defined by any set $\mathbb{P}$ of polynomial relations and λ has the form $x_N - \mu$, then $Y|_{x_N=\mu}$ can be identified with the subset of $R^{N-1}(S^{N-1})$ determined by substituting $x_N = \mu$ into each polynomial in $\mathbb{P}$.
2. $\kappa(Y|_{\lambda=0}) \leqq \kappa(Y)$ since

a) any representation of Y in γ descends to a representation of the restriction of Y of no greater degree upon multiplication by the element $\{\lambda=0\}$ of the ground ring γ_0 and

b) complexity is preserved by an isomorphism.

We also require the following somewhat subtler mapping from affine spaces to spheres. Geometrically it is a mapping from R^N to the generating $(N-1)$-sphere of its tangent cone at a point.

Definition 4.5. Let a be a point of R^N (or ∞). For $Y \in \gamma(R^N)$ let $\tau_a Y$ be the set of rays through a (the origin) which sufficiently near a (ultimately) lie in Y.

Proposition 4.6. *If* $Y \in \gamma(R^N)$ *then*

1. $\tau_a Y \in \gamma(S^{N-1})$,
2. $\kappa(\tau_a Y) \leqq \kappa(Y)$.

Proof. By definition $\tau_a Y$ is a subset of S^{N-1}. We need to show that it is semialgebraic. Consider the case $a=\infty$. Suppose Y is an open set determined by a single polynomial inequality $\{g>0\}$. Let the decomposition of g into its homogeneous parts be $g = g_0 + g_1 + \ldots + g_m$ where g_j is a [degree$(g) - j$]-form. Then $\tau_\infty\{g>0\}$ is given explicitly by

$$\tau_\infty\{g>0\} = \{g_0>0\} + \{g_0=0\}\{g_1>0\} + \ldots + \{g_0 = g_1 = \ldots = g_{m-1} = 0\}\{g_m>0\}.$$

Hence $\tau_\infty\{g>0\}$ is semialgebraic and has complexity not greater than 1.

We next check that τ_∞ is a ring homomorphism by observing that it preserves intersections and complements. It is obvious from its definition that it preserves intersections. Less obvious is $\tau_\infty(Y^c) = (\tau_\infty(Y))^c$. For example, if Y is the spiral $\{r = \log\theta\}$ in R^2, then $\tau_\infty(Y) = \tau_\infty(Y^c) = \phi$. However in the semialgebraic category such pathology cannot occur. For if a ray lies ultimately in $X = Y \cup Y^c$ it can cross between Y and Y^c only finitely many times and hence must lie ultimately in one or the other. Since τ_∞ is a ring homorphism, properties 1 and 2, already established for generators, follow for general Y. For $a \in R^N$ we simply replace the decomposition of g into homogeneous parts by its Taylor expansion around $x=a$.

5. Applications

As an application we use the structure $\{\gamma_k(X), \delta\}$ to obtain lower bounds for the complexity of certain semialgebraic sets. Our reasoning imitates the following very crude argument in the topological category with the structure $\{C_k, \partial\}$ where C_k denotes the Z_2-module of k-dimensional chains and ∂ is the usual boundary operator of algebraic topology. Here it is obvious that: 1) ∂ reduces dimension and 2) restriction to a subchain does not increase dimension. These weak properties suffice, for example, to prove the primitive result that the dimension of S^N is not less than N. For otherwise by 2) $\dim\{x_N \geqq 0\} < N$ and by 1) $\dim \partial\{x_N \geqq 0\} = \dim S^{N-1} < N-1$. Iterating, we find that S^0 is empty, a contradiction.

Proposition 5.1. *In R^N let $X_N = \bigcap_i \{x_i \geqq 1\}$, $Y_N = \bigcup_i \{x_i \leqq 0\}$. If $S \in \gamma(R^N)$ satisfies $X_N \subset S \subset Y_N^c$ then $\kappa(S) \geqq N$.*

Proof. Using coordinates $(x_0, x_1, \ldots, x_{j-1})$ in R^j and the operations defined in Sects. 3 and 4, for $Y \in \gamma(R^j)$ define $\varphi_j Y = (\delta \tau_\infty Y)^*|_{x_{j-1}=1}$. Then the following check of the four indicated operations shows that

$$\varphi_j \gamma_k(R^j) \subset \gamma_{k-1}(R^{j-1}). \tag{5.1}$$

First, if $\kappa(Z) = k$, then, by Proposition 4.6, τ_∞ maps Z into $\gamma_k(S^{j-1})$. Next, by Proposition 3.2, δ reduces the complexity by at least 1. Finally, the lifting $(\ )^*$ from S^j to R^j and restriction $x_{j-1} = 1$ to R^{j-1}, by observations of Sect. 4, do not increase the complexity.

We now derive a contradiction from the assumption that $S \in \gamma_{N-1}(R^N)$ by producing a disagreement between an algebraic and a geometric calculation. For, let $T = \varphi_2 \varphi_3 \ldots \varphi_N S$. Then repeated use of (5.1) shows that $T \in \gamma_0(R)$. But at the level of point sets it is easy to check that $X_{j-1} \subset \varphi_j(X_j)$ and $\varphi_j(Y_j) = Y_{j-1}$. Hence we find $X_1 \subset T \subset Y_1^c$ or $[1, \infty) \subset T \subset (0, \infty)$. Since $\gamma_0(R)$, the ring generated by the algebraic subsets of R, is precisly the ring of finite and cofinite subsets, this is impossible.

Corollary 5.2. *The positive orthant in R^N has complexity exactly N.*

Proof. Obviously N is an upper bound for the complexity. What is less obvious is that N is also a lower bound. But this follows immediately from the preceding proposition.

Other proofs of this fact can be given but these also seem rather complicated by comparison with the very simple conclusion. However it is reasonable that some delicacy is required since, for example, the relation $x + y - (x^2 + y^2)^{1/2} > 0$ defines the positive quadrant with a single inequality of a more general type. A circle of related questions involves the Mostowski separation theorem [BE] [C] [M] and the Mostowski number $m(N)$. This number, roughly speaking, is the largest number of square roots of definite polynomials which must be adjoined to the polynomial ring to obtain a function which separates a disjoint pair of closed semialgebraic sets in a variety of dimension N. Proposition 5.1 can be used to give lower bounds for $m(N)$ along the lines given in [S]. However, since other methods give sharper estimates, we merely give a corollary which shows the principle of the argument.

Corollary 5.3. *For $N \leqq 2$ the Mostowski number satisfies $m(N) \geqq N-1$.*

Proof. By contradiction. If $m(2)=0$, then any disjoint pair of closed semialgebraic sets in the plane can be separated by a polynomial function. In particular (recalling the sets X_N and Y_N defined in the proof of Proposition 5.1) there exists a polynomial function f positive on X_2 and negative on Y_2. But then the set $\{f>0\}$ violates Proposition 5.1. Similarly if $m(3)<2$ then X_3 and Y_3 can be separated by a function of the form $f=a+b\sqrt{g}$ where a and b are polynomials and g is a definite polynomial. Let $D=a^2-b^2g$. Then it is easy to check that $\{f>0\}=\{a>0\}\{D>0\}+\{b>0\}\{D<0\}+\{b>0\}\{a>0\}\{D=0\}$. It follows that $\kappa(\{f>0\})\leqq 2$ and again the set $\{f>0\}$ violates Proposition 5.1.

References

[B1] Bröcker, L.: Minimal generation of basic semialgebraic sets. Rocky Mt. J. Math. **14** (no. 4), 935–938 (1984)

[B2] Bröcker, L.: Minimale Erzeugung von Positivbereichen. Geom. Dedicata **16** (no. 3), 335–350 (1984)

[BE] Bochnak, J., Efroymson, G.: Real algebraic geometry and the 17th Hilbert problem. Math. Ann. **251**, 213–241 (1980)

[C] Coote, M.: Ensembles Semi-algebrique et fonctions de Nash. Seminaire de geometrie reelle de Paris 7, Fascicole no. 18, Ferrier (1981)

[H] Halmos, P.: Measure theory. New York: Van Nostrand 1950

[M] Mostowski, T.: Some properties of the ring of Nash functions. Ann. Sc. Norm. Super., Pisa, Cl. Sci., IV. Ser. Sci. **24**, 597–632 (1970)

[S] Stengle, G.: A lower bound for the complexity of separating functions. Rocky Mt. J. Math. **14** (no. 4), 927–929 (1984)

Received August 13, 1986; in revised form February 16, 1988

Math. Ann. 283, 211–239 (1989)

Mathematische Annalen

Asymptotic Behavior of Fundamental Solutions and Potential Theory of Parabolic Operators with Variable Coefficients

Nicola Garofalo[1] and Ermanno Lanconelli[2]

[1] Department of Mathematics, Northwestern University, Evanston, IL 60208, USA
[2] Dipartimento di Matematica, Università di Bologna, Bologna, Italy

1. Introduction

In $\mathbf{R}^{n+1}$ we consider the second order parabolic differential equation

$$Lu = \operatorname{div}(A(x,t)\nabla_x u) - D_t u = 0, \tag{1.1}$$

where $A(x,t)=(a_{ij}(x,t))$ is a real, symmetric, matrix-valued function on $\mathbf{R}^{n+1}$ with C^∞ entries. We assume there exists $\nu\in(0,1]$ such that for every $(x,t)\in\mathbf{R}^{n+1}$ and every $\xi\in\mathbf{R}^n$

$$\nu|\xi|^2 \leqq \sum_{i,j=1}^{n} a_{ij}(x,t)\xi_i\xi_j \leqq \nu^{-1}|\xi|^2. \tag{1.2}$$

We also assume, as this is not restrictive for our purposes, that there exists a compact set $F_0\subset\mathbf{R}^{n+1}$ containing the origin such that for $i,j=1,\dots,n$

$$a_{ij}(x,t)=\delta_{ij} \quad \text{for} \quad (x,t)\in\mathbf{R}^{n+1}\setminus F_0. \tag{1.3}$$

Then a fundamental solution Γ for (1.1) exists and under the assumptions made on $A(x,t)$ Γ is C^∞ off the diagonal in $\mathbf{R}^{n+1}\times\mathbf{R}^{n+1}$, see e.g., [Fr]. Throughout this paper we will use the letters z, z_0, ζ to denote respectively the points (x,t), (x_0,t_0), (ξ,τ) in $\mathbf{R}^{n+1}$. Then if $z\in\mathbf{R}^{n+1}$ and $r>0$, we set

$$\Omega_r(z)=\{\zeta\in\mathbf{R}^{n+1}\,|\,\Gamma(z;\zeta)>(4\pi r)^{-n/2}\} \tag{1.4}$$

and

$$\Psi_r(z)=\{\zeta\in\mathbf{R}^{n+1}\,|\,\Gamma(z;\zeta)=(4\pi r)^{-n/2}\}^-, \tag{1.5}$$

where for a subset $E\subset\mathbf{R}^{n+1}$, E^- denotes its closure. We call $\Omega_r(z)$ and $\Psi_r(z)$ respectively the *parabolic ball* and the *parabolic sphere* "centered" at z and of radius r. If $A(z)=$ Identity, and hence L in (1.1) is the heat operator on $\mathbf{R}^{n+1}$, $H=\Delta-D_t$, then Γ is given by the Gauss-Weierstrass kernel

$$K(x,t;\xi,\tau)=K(x-\xi;t-\tau)=\begin{cases}(4\pi(t-\tau))^{-n/2}\exp\left(-\dfrac{|x-\xi|^2}{4(t-\tau)}\right), & t>\tau,\\ 0, & t\leqq\tau.\end{cases} \tag{1.6}$$

In this case $\Omega_r(z)$ is a football-shaped domain in $\mathbf{R}^{n+1}$ whose intersection with hyperplanes perpendicular to the time axis are n-dimensional balls

$$|x-\xi|^2 \leqq R_r(t-\tau), \quad t-r<\tau<t, \tag{1.7}$$

where $R_r(t-\tau)=2n(t-\tau)\ln\left(\frac{r}{t-\tau}\right)$. We remark that the "center" z of the parabolic ball $\Omega_r(z)$ lies on the boundary $\Psi_r(z)$ of the ball itself. Throughout the paper dH_n stands for n-dimensional Hausdorff measure. In [GL] we proved

Theorem A. *Let $u \in C^\infty(\mathbf{R}^{n+1})$ and let $z \in \mathbf{R}^{n+1}$. For a.e. $r>0$ we have*

$$\begin{aligned} -\int_{\Psi_r(z)} u(\zeta)A(\zeta)(\nabla_\xi \Gamma(z;\zeta))\cdot \vec{N}_\xi(\zeta)dH_n(\zeta) &= u(z) \\ &+ \int_{\Omega_r(z)} Lu(\zeta)\left[\Gamma(z;\zeta)-(4\pi r)^{-n/2}\right]d\zeta. \end{aligned} \tag{1.8}$$

For every $r>0$ we have

$$\begin{aligned} (4\pi r)^{-n/2}\int_{\Omega_r(z)} u(\zeta)\frac{A(\zeta)(\nabla_\xi \Gamma(z;\zeta))\cdot \nabla_\xi \Gamma(z;\zeta)}{\Gamma^2(z;\zeta)}d\zeta &= u(z) \\ +\frac{n}{2}r^{-n/2}\int_0^r l^{n/2}\int_{\Omega_l(z)} Lu(\zeta)\left[\Gamma(z;\zeta)-(4\pi l)^{-n/2}\right]d\zeta\frac{dl}{l}. \end{aligned} \tag{1.9}$$

In (1.8) $\vec{N}_\xi(\zeta)$ denotes the spatial component of the outer normal $\vec{N}(\zeta) =(\vec{N}_\xi(\zeta), N_\tau(\zeta))$ in ζ to the surface $\Psi_r(z)$.

It appears clear from Theorem A that in order to fully use (1.8), (1.9) we must have as much information as possible about the kernels appearing in them. This leads to the study of the asymptotic behavior for small times of the fundamental solution Γ and of its derivatives. Section 2 in this paper is devoted to this purpose. Our main result there, Theorem 2.1, reads: *if x is sufficiently close to y and if $\Gamma(x,y,t)=\Gamma(x,t;y,0)$, then as $t\to 0^+$*

$$\Gamma(x,y,t)\sim(4\pi t)^{-n/2}\exp\left(-\frac{d^2(x,y,t)}{4t}\right)\sum_{j=0}^\infty t^j u_j(x,y,t). \tag{1.10}$$

A similar result holds for the derivatives of Γ.

For a complete explanation of (1.10) we refer to Sect. 2. We only mention that $d(x,y,t)$ represents the Riemannian distance generated on $\mathbf{R}^n$ at the time t by the metric $g_{ij}(t)dx_i \otimes dx_j$, where $g_{ij}(t)(x)=a^{ij}(x,t)$, and $(a^{ij}(x,t))=A^{-1}(x,t)$. For time-independent parabolic operators (1.10) is a classical result of Minashisundaram and Pleijel, see [BGM]. A different approach based on transmutation formulas is due to Kannai [K]. Theorem 2.1 plays a basic role in this paper. It has also been crucial in our recent work [GL] on Wiener's criterion for the operator L in (1.1) which has extended Evans and Gariepy's previous result for the heat operator [EG]. To provide further motivation for the results in this paper we describe a lemma in [GL] somewhat of a geometric flavor the proof of which ultimately relies on Theorem 2.1. If $K(x,t)=K(x,t;0,0)$ is as in (1.6) and if $E(x,t)=\ln K(x,t)$, then for $t>0$ we have

$$\nabla_x E(x,t)=-\frac{x}{2t}, \quad D_t E(x,t)=\frac{|x|^2}{4t^2}-\frac{n}{2t}.$$

From this it is then immediate to see that given any $\theta>1$ the inequality

$$|\nabla_x E(x,t)|^2 \leqq \theta D_t E(x,t) \tag{1.11}$$

holds iff $t \leqq \delta|x|^2$ where $\delta=\delta(\theta)=(\theta-1)/2n\theta$, i.e., iff (x,t) lies outside a paraboloid with vertex at $(0,0)$ and aperture depending on θ. Let now $\Gamma(x,t)=\Gamma(x,t;0,0)$ be the fundamental solution of (1.1) with pole at $(0,0)$ and let us again set $E(z)=E(x,t)$ $=\ln\Gamma(x,t)$. Then in [GL] we proved the following

Lemma B. *There exist $r_0>0$, and for every $\theta>1$ a $\delta=\delta(\theta)>0$, such that*

$$A(z)(\nabla_x E(z))\cdot\nabla_x E(z) \leqq \theta D_t E(z), \tag{1.12}$$

for every $z=(x,t)\in\{\zeta\in\mathbf{R}^{n+1} | \Gamma(\zeta;0)>(4\pi r)^{-n/2}\}$ with $t\leqq\delta|x|^2$, and every $r\leqq r_0$.

We remark that in the intrinsic notation of Theorem 2.1 below (1.12) can be rewritten as

$$|\nabla_{M_t} E(z)|_t^2 \leqq \theta D_t E(z), \tag{1.13}$$

where ∇_{M_t} is the gradient and $|\cdot|_t$ is the norm in the metric $g_{ij}(t)dx_i\otimes dx_j$. A comparison of (1.11) and (1.13) unravels the deep connection between the geometry induced on $\mathbf{R}^n$ by the matrix $A(x,t)$ in (1.1) and the fundamental solution of the operator L. Whereas the proof of (1.11) is a simple calculation based on the explicit knowledge of the fundamental solution K of the heat operator, the proof of Lemma B is quite delicate and uses the full strength of Theorem 2.1. (1.11) was first observed by Evans and Gariepy who used it as a key step in the proof of a strong version of Harnack's inequality for the heat operator, see Lemma 3.2 in [EG]. Roughly speaking, the latter asserts that if u is nonnegative solution of $Hu=0$ in a heat ball $\Omega_{2r}(0)$, then the infimum of u in a smaller ball "concentric" to $\Omega_{2r}(0)$, say $\Omega_{r/2}(0)$, is strictly positive (and independent of r!) provided that the n-dimensional average of u at the time level $t=-r$ is one. The interest of this result is that it provides a stronger information than the normal Harnack inequality for parabolic equations and cannot be derived from it. This is so because the parabolic ball $\Omega_{r/2}(0)$ contains regions which lie outside any paraboloid with vertex in $0\in\mathbf{R}^{n+1}$ and aperture in the negative time direction. Using Lemma B we proved a similar result for the operator L in (1.1), see Theorem 1.4 in [GL].

Before proceeding with the plan of the paper we pause to provide some historical background. If in (1.8) we take u such that $Lu=0$ we obtain

$$u(z)=-\int_{\Psi_r(z)} u(\zeta)A(\zeta)(\nabla_\xi\Gamma(z;\zeta))\cdot\vec{N}_\xi(\zeta)dH_n(\zeta). \tag{1.14}$$

The ancestor of (1.14) is the well-known formula

$$u(x)=\frac{1}{n\omega_n r^{n-1}}\int_{|x-y|=r} u(y)d\sigma(y), \tag{1.15}$$

valid for any harmonic function u in $\mathbf{R}^n$, any $x\in\mathbf{R}^n$ and $r>0$ (ω_n=volume of n-dimensional unit ball). (1.15) is the keystone of classical potential theory. Immediate consequences of it are the maximum principle, Harnack's inequality, the smoothness of harmonic functions (Weyl's lemma), just to name a few, see e.g., [H] and [F]. Since in (1.14)

$$\vec{N}_\xi(\zeta) = -\frac{\nabla_\xi \Gamma(z;\zeta)}{|(\nabla_\xi \Gamma(z;\zeta), D_\tau \Gamma(z;\zeta))|},$$

if A in (1.1) is the identity matrix we obtain

$$u(z) = \int_{\Psi_r(z)} u(\zeta) \frac{|\nabla_\xi K(z-\zeta)|^2}{|(\nabla_\xi K(z-\zeta), D_\tau K(z-\zeta))|} dH_n(\zeta). \tag{1.16}$$

(1.16) was first discovered and used by Pini in [P 1] – [P 3] for the heat operator in $\mathbf{R}^2$. Fulks [Fu] subsequently extended Pini's result to any number of variables. Watson [W] starting from Fulks' formula (1.16) obtained the following representation

$$u(z) = (4\pi r)^{-n/2} \int_{\Omega_r(z)} u(\zeta) \frac{|x-\xi|^2}{4(t-\tau)^2} d\zeta, \tag{1.17}$$

for any solution u of the $Hu=0$, any $z=(x,t) \in \mathbf{R}^{n+1}$ and $r>0$. If we observe that

$$|\nabla_\xi K(z-\zeta)|^2 = \frac{|x-\xi|^2}{4(t-\tau)^2} K(z-\zeta)^2,$$

we see that (1.17) is just a special case of

$$u(z) = (4\pi r)^{-n/2} \int_{\Omega_r(z)} u(\zeta) \frac{A(\zeta)(\nabla_\xi \Gamma(z;\zeta)) \cdot \nabla_\xi \Gamma(z;\zeta)}{\Gamma^2(z;\zeta)} d\zeta, \tag{1.18}$$

valid for any solution of (1.1), any $z \in \mathbf{R}^{n+1}$ and $r>0$. (1.14) was found by Fabes and one of us in [FG], and (1.18) was established in the same paper using (1.14) and Federer's coarea formula, see [Fe]. Theorem A above extends all previous results and can be used to study parabolic potential theory.

An unfavorable feature displayed by (1.18) consists in the unboundedness of the kernel appearing in it. This can be easily seen in the case of the heat operator [see (1.17)], where the kernel takes the form $(4\pi r)^{-n/2} \frac{|x-\xi|^2}{4(t-\tau)^2}$. In Sect. 3 we prove some new representation formulas for smooth functions on $\mathbf{R}^{n+1}$ containing kernels which are not only bounded, but also possess a degree of regularity arbitrarily high. These new formulas are based on the following observations. If $m \in \mathbf{N}$ and $y \in \mathbf{R}^m$ we define a parabolic operator on $\mathbf{R}^{n+m+1}$ by setting $\hat{L} = L + \Delta_y$, if L is as in (1.1). Then if u is a function on $\mathbf{R}^{n+1}$ and for $y \in \mathbf{R}^m$ we define

$$\hat{u}(x,y,t) = u(x,t)$$

we have

$$\hat{L}\hat{u}(x,y,t) = Lu(x,t).$$

Moreover, if $\hat{\Gamma}$ and Γ are respectively the fundamental solutions of $\hat{L}$ and L, then it turns out that

$$\hat{\Gamma}(x,y,t;\xi,\eta,\tau) = \Gamma(x,t;\xi,\tau) K_m(y-\eta;t-\tau) \tag{1.19}$$

where K_m is the Gauss-Weierstrass kernel in $\mathbf{R}^{m+1}$, see (1.6). With these observations in mind we apply the representation results in [GL] to the function $\hat{u}$ and the operator $\hat{L}$ on $\mathbf{R}^{n+m+1}$ above defined. Because of (1.19) and the fact that the matrix $\hat{A}$ of $\hat{L}$ is given by the $(n+m)\times(n+m)$ block matrix

$$\hat{A}=\begin{bmatrix} A & 0 \\ 0 & I_y \end{bmatrix},$$

something magic happens and the dependence in the added variable $y\in\mathbf{R}^m$ disappears in all the integrals involved. We refer to Theorem 3.1 below for details. We only quote here a special case. For $m\in\mathbf{N}$ we define the *modified parabolic ball* centered at $z=(x,t)\in\mathbf{R}^{n+1}$ and of radius $r>0$

$$\Omega_r^m(z)=\left\{\zeta=(\xi,\tau)\in\mathbf{R}^{n+1}|(4\pi(t-\tau))^{-m/2}\Gamma(z;\zeta)>(4\pi r)^{-\frac{n+m}{2}}\right\}, \tag{1.20}$$

see (1.4). Then if $u\in C^\infty(\mathbf{R}^{n+1})$ is a solution of $Lu=0$ in $\mathbf{R}^{n+1}$ we have

$$u(z)=\int_{\Omega_r^m(z)} u(\zeta)E_r^{(m)}(z;\zeta)d\zeta. \tag{1.21}$$

If on $\Omega_r^m(z)$ we define the function $\zeta\mapsto R_r(z;\zeta)$ by setting

$$R_r^2(z;\zeta)=4(t-\tau)\ln\left[(4\pi r)^{\frac{n+m}{2}}(4\pi(t-\tau))^{-m/2}\Gamma(z;\zeta)\right],$$

then in Sect. 3 we prove that the kernel $E_r^{(m)}$ in (1.21) is given by

$$E_r^{(m)}(z;\zeta)=(4\pi r)^{-\frac{n+m}{2}}\omega_m R_r^m(z;\zeta)\left[\frac{A(\zeta)(\nabla_\xi\Gamma(z;\zeta))\cdot\nabla_\xi\Gamma(z;\zeta)}{\Gamma^2(z;\zeta)}+\frac{m}{m+2}\frac{R_r^2(z;\zeta)}{4(t-\tau)^2}\right], \tag{1.22}$$

where ω_m denotes the volume of the m-dimensional unit ball. We emphasize that if we agree to set $\omega_0=1$, then we obtain from (1.22)

$$E_r^{(0)}(z;\zeta)=(4\pi r)^{-n/2}\frac{A(\zeta)(\nabla_\xi\Gamma(z;\zeta))\cdot\nabla_\xi\Gamma(z;\zeta)}{\Gamma^2(z;\zeta)},$$

and therefore (1.21) gives (1.18) back if $m=0$.

In Sect. 4 using (1.10) and similar expansions for the derivatives of Γ we prove that if $m\in\mathbf{N}$ and $m>2$, then the kernel $E_r^{(m)}$ in (1.21) is bounded by an appropriate power of r on the modified parabolic ball $\Omega_r^m(z)$. This fact is used in Theorem 4.1 to give an elementary proof of Harnack's inequality modelled on the classical proof via mean value formulas for harmonic functions. The result itself is clearly not new as Moser [M] has proved Harnack's inequality for parabolic operators with bounded measurable coefficients. Our point however is to emphasize the elementary and self-contained character of the proof.

In Sect. 5 we investigate several questions in classical potential theory related to the averaging operators $u\mapsto u_r^{(m)}$ where

$$u_r^{(m)}(z)=\int_{\Omega_r^m(z)} u(\zeta)E_r^{(m)}(z;\zeta)d\zeta, \tag{1.23}$$

for $z \in \mathbf{R}^{n+1}$, $r>0$, $m \in \mathbf{N} \cup \{0\}$. To be more specific, we need to introduce some definition. A bounded open set $U \subset \mathbf{R}^{n+1}$ is said to be *L-regular* if for any $\varphi \in C(\partial U)$ there exists a (unique) $H_\varphi^U \in C^\infty(U) \cap C(\bar{U})$ such that $LH_\varphi^U = 0$ in U and for which

$$\lim_{\substack{z \to z_0 \\ z \in U}} H_\varphi^U(z) = \varphi(z_0)$$

for every $z_0 \in \partial U$. Given an open set $D \subset \mathbf{R}^{n+1}$ a function $w: D \to \bar{\mathbf{R}}$ is said to be *L-superparabolic* in D if: (i) $-\infty < w \leqq +\infty$, $w < +\infty$ in a dense subset of D; (ii) w is lower semi-continuous (l.s.c.); (iii) for every L-regular subset $U \subset \bar{U} \subset D$, and every $\varphi \in C(\partial U)$ if $w|_{\partial U} \geqq \varphi$, then $w \geqq H_\varphi^U$ in U. In Proposition 5.1 we show that L-superparabolic functions are characterized by the super mean value property, i.e., a l.s.c. function on $\mathbf{R}^{n+1}$ is L-superparabolic iff for every $z \in \mathbf{R}^{n+1}$ and r sufficiently small

$$u(z) \geqq u_r^{(m)}(z) \quad \text{for any fixed } m \in \mathbf{N} \cup \{0\}.$$

The rest of Sect. 5 is devoted to proving Proposition 5.2 (see also Corollary 5.1) which states that for a L-superparabolic function on $\mathbf{R}^{n+1}$ the averages $u_r^{(m)}(z)$ increase as $r \to 0$ to the value of the function u at z. Moreover, for any $r>0$ small enough the function $z \mapsto u_r^{(m)}(z)$ is itself L-superparabolic in $\mathbf{R}^{n+1}$. The proof of Proposition 5.2 is accomplished in several steps. The crucial one is to prove the following property of the fundamental solution Γ of L. If for $\zeta \in \mathbf{R}^{n+1}$ we set $w = \Gamma(\cdot\,;\zeta)$ and denote by w_r^σ the surface average of w defined as in (1.14), then for every ϱ, r sufficiently small and $z \in \mathbf{R}^{n+1}$ we have

$$(w_r^\sigma)_\varrho(z) \leqq w_r^\sigma(z). \tag{1.24}$$

$(w_r^\sigma)_\varrho(z)$ denotes the solid average defined by the right-hand side of (1.18) of the function $z \mapsto w_r^\sigma(z)$. (1.24) implies that a similar inequality holds for the solid averages of Γ, i.e.,

$$((\Gamma(\cdot\,;\zeta))_r)_\varrho(z) \leqq (\Gamma(\cdot\,;\zeta))_r(z),$$

see Lemma 5.2. The latter inequality easily leads to the conclusion of the proof of Proposition 5.2.

Section 6 contains one of the main results in this paper. It is well-known that a superharmonic function in $\mathbf{R}^n$ can be approximated by an increasing sequence of smooth superharmonic functions. This can be accomplished by the usual device of mollification. The same device can be applied to supertemperatures in $\mathbf{R}^{n+1}$, i.e., supersolutions of the heat operator. Such an approach does not work, however, for operators with variable coefficients. In this general context the problem of approximating a given supersolution with an increasing sequence of (sufficiently) smooth supersolutions is a rather delicate one. For superharmonic functions in $\mathbf{R}^n$ a different approach is based on the use of the averaging operators $u \mapsto u_r$ where

$$u_r(x) = \frac{1}{\omega_n r^n} \int_{|x-y|<r} u(y)dy, \tag{1.25}$$

see [H]. It is well-known that the operator $u \mapsto u_r$ is a smoothing operator which, moreover, preserves superharmonicity, hence one obtains the desired approximation property by successive iterations of (1.25). In Sect. 6 we take up this approach. The main result, Theorem 6.1, reads: *given a L-superparabolic function u in* $\mathbf{R}^{n+1}$ *and* $\nu \in N$ *there exists a sequence* $(u_j)_{j \in N}$ *with* $u_j \in C^\nu(\mathbf{R}^{n+1})$, u_j *L-superparabolic in* $\mathbf{R}^{n+1}$, $u_j \leqq u_{j+1}$ *for any* $j \in \mathbf{N}$, *such that* $u_j \to u$ *as* $j \to \infty$. We mention that this result plays an important role in the proof of the sufficiency of Wiener's condition in [GL]. The proof of Theorem 6.1 is based on some mean value operators which are constructed from (1.20) through a process of superposition. These operators which are introduced in (3.20)–(3.22) of Sect. 3, are reminiscent of those in Weyl's classical proof of smoothness of harmonic functions, see [F]. Besides their independent interest they turn out to be extremely useful since to check their smoothing properties is considerably easier than doing the same for the operators $u_r^{(m)}$ defined by (1.23).

Although the results of this paper do apply to time-independent solutions of the Eq. (1.1), thus providing corresponding results for elliptic equations, we have omitted any reference to such a situation. Concerning Theorem 6.1, however, we mention that we have recently become aware of a very interesting paper by Littman [L 2] (dated 1963) which is concerned with the problem of monotonic approximation of supersolutions of elliptic equations. Although Littman's result is not based on exact formulas, it is very similar in spirit to our approach. In fact, in the case of a divergence form elliptic operator with sufficiently smooth coefficients, the conclusion of Theorem A in [L 2], although not the construction of the smooth approximation sequence, is essentially analogous to the conclusion of Theorem 6.1 in this paper. We wish to thank Prof. A. Devinatz for bringing references [L 1], [L 2] to our attention and Prof. W. Littman for kindly discussing with us the results of his mentioned papers.

Finally, we would like to thank the referee, whose constructive criticism has led us to improve the presentation of the proof of Theorem 2.1 and simplify the proof of Lemma 5.2.

2. Asymptotic Behavior of the Fundamental Solution for Small Times

This section is devoted to obtain an asymptotic expansion of the fundamental solution $\Gamma(x, t; y, s)$ of (1.1) together with its derivatives as $t \to s^+$ and x, y vary in a compact neighborhood of zero. Such an estimate plays a fundamental role in the work [GL] on Wiener's criterion as well as in the subsequent sections of this paper. Our approach is modelled on Minakshisundaram and Pleijel's asymptotic evaluation of the fundamental solution of the heat operator on a compact manifold, see e.g., [BGM]. In what follows K will denote a (sufficiently small) connected compact neighborhood of zero in $\mathbf{R}^n$. For a fixed $t \geqq 0$ we let $A(t)$ denote the matrix-valued function on $\mathbf{R}^n$ $A(t)(x) = (a_{ij}(x, t))$, where $A(x, t) = (a_{ij}(x, t))$ is as in (1.1). Also we set $A(t)^{-1} = (a^{ij}(\cdot, t))$. M_t will denote the Riemannian manifold $(K, g_{ij}(t))$, where $g_{ij}(t) = a^{ij}(\cdot, t)$, $g^{ij}(t) = a_{ij}(\cdot, t)$. $d(x, y, t)$ will indicate the distance between two points x, y on M_t.

Theorem 2.1. *For* $x, y \in K$, $x \neq y$, *we let* $\Gamma(x, y, t) = \Gamma(x, t; y, 0)$ *denote the fundamental solution of L in* (1.1) *with pole at* $(y, 0)$. *Then as* $t \to 0^+$ *we have the asymptotic expansion*

$$\Gamma(x, y, t) \sim (4\pi t)^{-n/2} \exp\left(-\frac{d^2(x, y, t)}{4t}\right) \sum_{j=0}^{\infty} t^j u_j(x, y, t). \tag{2.1}$$

By (2.1) *we mean that there exist a suitably small* $T>0$ *and a sequence* $(u_j)_{j\in N}$, *with* $u_j\in C^\infty(K\times K\times[0,T])$ *such that*

$$\Gamma(x,y,t)-(4\pi t)^{-n/2}\exp\left(-\frac{d^2(x,y,t)}{4t}\right)\sum_{j=0}^{k} t^j u_j(x,y,t)=w_k(x,y,t), \tag{2.2}$$

with

$$w_k(x,y,t)=0\left(t^{k+1-n/2}\exp\left(-\frac{\delta|x-y|^2}{4t}\right)\right), \quad as \quad t\to 0^+, \tag{2.3}$$

uniformly for $x,y\in K$. *In* (2.3) $\delta>0$ *is a number depending on* v *in* (1.2) *and* n. *The function* u_0 *in* (2.1) *can be chosen such that* $u_0(x,x,0)=1$.

An expansion similar to (2.1) *holds for the derivatives of* Γ.

Proof. We fix a small enough neighborhood of the origin, K, and a number $T>0$ sufficient small, so that we can find $\varepsilon_0>0$ with the property that for every $y\in K$ the geodesic ball

$$B^t_{\varepsilon_0}(y)=\{z\in K\,|\,d(z,y,t)<\varepsilon_0\}$$

is a normal, convex neighborhood of y for every $t\in[0,T]$. It is clear that restricting, if needed, T and ε_0, we can determine a cylinder $U_\varepsilon\times(0,T)$, with

$$0<\varepsilon\leqq\varepsilon_0, \qquad U_\varepsilon=\{(x,y)\in K\times K\,|\,d(x,y,t)<\varepsilon\}, \qquad t\in[0,T],$$

for which the function $\exp\left(-\dfrac{d^2(x,y,t)}{4t}\right)$ is in $C^\infty(U_\varepsilon\times(0,T))$. After these preliminaries, for every $t\in[0,T]$ and $y\in K$ fixed we denote by $(r(t),\theta(r(t)))$ the intrinsic geodesic polar coordinates on M_t with pole at y. We have $r(t)(x)=d(x,y,t)$. In what follows we let $r(x,t)=r(t)(x)$. We will need to distinguish between the Euclidean gradient on K and the intrinsic gradient on M_t. The former will be denoted by ∇_x, the latter by ∇_{M_t}. If $g(t)=\det(g_{ij}(t))$, the Laplace-Beltrami operator on M_t is

$$\Delta_{M_t}=\frac{1}{\sqrt{g(t)}}\sum_{i,j=1}^{n} D_j(\sqrt{g(t)}g^{ij}(t)D_i),$$

where $D_i=\dfrac{\partial}{\partial x_i}$. The Euclidean inner product on K will be denoted by $\cdot$, whereas the intrinsic inner product on M_t by $\langle\,,\rangle_t$. The symbol $|u|$ will always stand for the Euclidean length of the vector $u\in\mathbf{R}^n$, whereas $|u|_t=(\langle u,u\rangle_t)^{1/2}$. In the geodesic polar coordinates $(r(t),\theta(r(t)))$ the Laplace-Beltrami operator can be written as (see [BGM])

$$\Delta_{M_t}=D_r^2+D_r(\ln\sqrt{A(r(t))})D_r+\Delta_{S_{r(t)}}, \tag{2.4}$$

where $D_r=\dfrac{\partial}{\partial r}$, $S_{r(t)}=$ geodesic surface of radius $r(t)$ and centered at y, $\sqrt{A(r(t))}=r(t)^{n-1}\theta(r(t))$, and $\Delta_{S_{r(t)}}$ is the Laplace-Beltrami operator on $S_{r(t)}$. Replacing the expression of $\sqrt{A}$ in (2.4) yields

$$\Delta_{M_t}=D_r^2+\left[\frac{n-1}{r}+\frac{D_r\theta(r(t))}{\theta(r(t))}\right]D_r+\Delta_{S_{r(t)}}. \tag{2.5}$$

Now we consider the function $G=(4\pi t)^{-n/2}\exp\left(-\dfrac{r^2}{4t}\right)$, which is in $C^\infty(U_\varepsilon\times(0,T))$, and we set

$$\gamma_k(x,y,t)=G(x,y,t)\sum_{j=0}^{k} t^j u_j(x,y,t), \tag{2.6}$$

where the functions u_j are to be determined. We wish to calculate $L\gamma_k$. To this end we observe that

$$\begin{aligned} L&=\sum_{i,j=1}^{n} D_j(a_{ij}(x,t)D_i)-D_t \\ &=\Delta_{M_t}+\vec{b}(x,t)\cdot\nabla_x-D_t, \end{aligned} \tag{2.7}$$

where $\vec{b}$ is the vector field whose components are

$$b_i(x,t)=-\sum_{j=1}^{n} g^{ij}(x,t)D_j[\ln\sqrt{g(t)(x)}]. \tag{2.8}$$

By (2.5) we obtain

$$\Delta_{M_t}G=\left(\frac{r^2}{4t^2}-\frac{n}{2t}-\frac{1}{2t}\frac{rD_r\theta}{\theta}\right)G. \tag{2.9}$$

Also

$$D_tG=\left(\frac{r^2}{4t^2}-\frac{n}{2t}-\frac{1}{4t}D_t(r^2)\right)G, \tag{2.10}$$

$$\nabla_{M_t}G=-\frac{1}{4t}G\nabla_{M_t}(r^2),\qquad \nabla_xG=-\frac{1}{4t}G\nabla_x(r^2). \tag{2.11}$$

Using (2.7), (2.9), (2.10), and (2.11) we have

$$\begin{aligned} L\gamma_k&=\Delta_{M_t}\gamma_k+\vec{b}\cdot\nabla_x\gamma_k-D_t\gamma_k \\ &=G\Bigg[\left(\frac{r^2}{4t^2}-\frac{n}{2t}-\frac{1}{2t}\frac{rD_r\theta}{\theta}\right)\sum_{j=0}^{k} t^j u_j \\ &\quad+\sum_{j=0}^{k} t^j\Delta_{M_t}u_j+2\sum_{j=0}^{k} t^j\left(-\frac{1}{4t}\right)\langle\nabla_{M_t}(r^2),\nabla_{M_t}u_j\rangle_t \\ &\quad-\frac{1}{4t}\vec{b}\cdot\nabla_x(r^2)\sum_{j=0}^{k} t^j u_j+\sum_{j=0}^{k} t^j\vec{b}\cdot\nabla_x u_j \\ &\quad-\left(\frac{r^2}{4t^2}-\frac{n}{2t}-\frac{1}{4t}D_t(r^2)\right)\sum_{j=0}^{k} t^j u_j-\sum_{j=0}^{k} t^jD_tu_j-\sum_{j=1}^{k} jt^{j-1}u_j\Bigg]. \end{aligned} \tag{2.12}$$

Now we observe that

$$\langle\nabla_{M_t}(r^2),\nabla_{M_t}u_j\rangle_t=2rD_ru_j,$$

therefore (2.12) gives

$$\begin{aligned} L\gamma_k = G\Bigg[& \frac{1}{t}\left(-rD_r u_0 - \frac{rD_r\theta}{2\theta}u_0 + \frac{1}{4}D_t(r^2)u_0 - \frac{1}{4}\vec{b}\cdot\nabla_x(r^2)u_0\right) \\ & + \Bigg(-rD_r u_1 - \frac{rD_r\theta}{2\theta}u_1 + \frac{1}{4}D_t(r^2)u_1 \\ & - \frac{1}{4}\vec{b}\cdot\nabla_x(r^2)u_1 - u_1 + \Delta_{M_t}u_0 + \vec{b}\cdot\nabla_x u_0 - D_t u_0\Bigg) \\ & + t\Bigg(-rD_r u_2 - \frac{rD_r\theta}{2\theta}u_2 + \frac{1}{4}D_t(r^2)u_2 - \frac{1}{4}\vec{b}\cdot\nabla_x(r^2)u_2 \\ & - 2u_2 + \Delta_{M_t}u_1 + \vec{b}\cdot\nabla_x u_1 - D_t u_1\Bigg) + \dots \\ & + t^{k-1}\Bigg(-rD_r u_k - \frac{rD_r\theta}{2\theta}u_k + \frac{1}{4}D_t(r^2)u_k - \frac{1}{4}\vec{b}\cdot\nabla_x(r^2)u_k \\ & - ku_k + \Delta_{M_t}u_{k-1} + \vec{b}\cdot\nabla_x u_{k-1} - D_t u_{k-1}\Bigg) \\ & + t^k(\Delta_{M_t}u_k + \vec{b}\cdot\nabla_x u_k - D_t u_k)\Bigg]. \end{aligned} \tag{2.13}$$

If we set

$$\begin{aligned} \Phi_j = & -rD_r u_j - \frac{rD_r\theta}{2\theta}u_j + \frac{1}{4}D_t(r^2)u_j - \frac{1}{4}\vec{b}\cdot\nabla_x(r^2)u_j \\ & - ju_j + \Delta_{M_t}u_{j-1} + \vec{b}\cdot\nabla_x u_{j-1} - D_t u_{j-1}, \end{aligned} \tag{2.14}$$

for $j=0,1,\dots,k$, with $u_{-1}\equiv 0$, and

$$\Phi_{k+1} = \Delta_{M_t}u_k + \vec{b}\cdot\nabla_x u_k - D_t u_k, \tag{2.15}$$

then we can rewrite

$$L\gamma_k = G\sum_{j=0}^{k+1} t^{j-1}\Phi_j. \tag{2.16}$$

At this point we would like to determine the $k+1$ functions $u_0, u_1, \dots, u_k$ in such a way that

$$L\gamma_k = t^k G\Phi_{k+1} \quad \text{in} \quad U_\varepsilon \times (0,T). \tag{2.17}$$

By (2.14), (2.17) will be true if we can solve the $k+1$ equations

$$rD_r u_j = (q-j)u_j + p_j, \qquad j=0,1,\dots,k, \tag{2.18}$$

where

$$q = -\frac{rD_r\theta}{2\theta} + \frac{1}{4}D_t(r^2) - \frac{1}{4}\vec{b}\cdot\nabla_x(r^2), \tag{2.19}$$

and

$$p_0 \equiv 0, \quad p_j = \Delta_{M_t} u_{j-1} + \vec{b} \cdot \nabla_x u_{j-1} - D_t u_{j-1}, \quad j = 1, \ldots, k. \tag{2.20}$$

We remark that solving (2.18) amounts to show that $\Phi_j \equiv 0, j = 0, 1, \ldots, k$. From the definition (2.19) of $q = q(r, \theta, t)$ it is clear that

$$q(0, \theta, t) = 0 \quad \text{uniformly in } \theta, t. \tag{2.21}$$

This fact plays a crucial role in the following considerations. We now

Claim. *There exist $k+1$ $C^\infty(U_\varepsilon \times [0, T])$ functions, $u_0, u_1, \ldots, u_k$, which solve the Eqs.* (2.18). *Moreover, u_0 can be chosen so that*

$$u_0(0, \theta, t) \equiv 1, \quad \textit{uniformly in } \theta, t. \tag{2.22}$$

In the proof of the claim we follow closely the inductive argument given in [BGM]. We start with u_0. For $j = 0$ (2.18) becomes [see (2.20)]

$$r D_r u_0 = q u_0. \tag{2.23}$$

A solution to (2.23) is provided formally by

$$u_0(r, \theta, t) = u_0(\theta, t) \exp\left(\int_0^r q(\varrho, \theta, t) \frac{d\varrho}{\varrho}\right).$$

We emphasize that the function q defined by (2.19) belongs to $C^\infty(U_\varepsilon \times [0, T])$ since by (2.5) and the fact that $r^2 \in C^\infty(U_\varepsilon \times [0, T])$ it follows that $\frac{r D_r \theta}{\theta} \in C^\infty(U_\varepsilon \times [0, T])$. Moreover, because of (2.21) it is integrable near $r = 0$ with respect to the measure $\frac{d\varrho}{\varrho}$. Since the choice of the initial value $u_0(\theta, t)$ is up to us, we take

$$u_0(\theta, t) = u_0(0, \theta, t) \equiv 1. \tag{2.24}$$

In what follows for a fixed $t \geqq 0$, and for $0 \leqq s \leqq r(t) = d(x, y, t)$, we denote by x_s the point on the unique geodesic in M_t joining x to y and having distance s from y. If $T_y(M_t)$ denotes the tangent space to M_t in y and $\exp_y : T_y(M_t) \to M_t$ is the exponential map, then for $0 \leqq \alpha \leqq 1$ we have

$$x_{\alpha r} = x_{\alpha r(t)} = \exp_y(\alpha \exp_y^{-1}(x)). \tag{2.25}$$

With these remarks in mind and taking (2.24) into account we have

$$\begin{aligned} u_0(x, y, t) &= \exp\left(\int_0^r q(x_\varrho, y, t) \frac{d\varrho}{\varrho}\right) = \exp\left(\int_0^1 q(x_{\varrho r}, y, t) \frac{d\varrho}{\varrho}\right) \\ &= \exp\left(\int_0^1 q(\exp_y(\varrho \exp_y^{-1}(x)), y, t) \frac{d\varrho}{\varrho}\right), \end{aligned} \tag{2.26}$$

where in the last equality we have used (2.25). Because of (2.21) the map

$$(x, y, t, \varrho) \mapsto \frac{1}{\varrho} q(\exp_y(\varrho \exp_y^{-1}(x)), y, t)$$

is C^∞ in $U_\varepsilon\times[0,T]\times[0,1]$. The claim is thus proved in the case $k=0$. We can then start the induction. Now let $j=k$ and consider (2.18). We assume we have determined k functions $u_0,u_1,\dots,u_{k-1}$ in $C^\infty(U_\varepsilon\times[0,T])$ which solve (2.18). We look for a solution u_k of (2.18) of the form

$$u_k(r,\theta,t)=C_k(r,\theta,t)r^{-k}\theta^{-1/2}. \tag{2.27}$$

By the method of variation of constants C_k must satisfy the equation

$$rD_rC_k=\psi C_k+p_kr^k\theta^{1/2},$$

where we have set $\psi=q+\dfrac{rD_r\theta}{2\theta}$, so that by (2.19) $\psi(0,\theta,t)=0$ uniformly in θ,t, and $\psi\in C^\infty(U_\varepsilon\times[0,T])$. Formally, a solution C_k is given by

$$\begin{aligned}C_k(x,y,t)=&\exp\left(\int_0^r\psi(x_r,y,t)\frac{d\tau}{\tau}\right)\int_0^r p_k(x_\sigma,y,t)\theta^{1/2}(x_\sigma,y,t)\sigma^k\\&\times\exp\left(-\int_0^\sigma\psi(x_s,y,t)\frac{ds}{s}\right)\frac{d\sigma}{\sigma}.\end{aligned} \tag{2.28}$$

Performing obvious changes of variables in the integrals involved in (2.28) and using (2.25) we can rewrite

$$\begin{aligned}C_k(x,y,t)=&r^k\exp\left(\int_0^1\psi(x_{\tau r},y,t)\frac{d\tau}{\tau}\right)\int_0^1 p_k(x_{\sigma r},y,t)\theta^{1/2}(x_{\sigma r},y,t))\sigma^k\\&\times\exp\left(-\int_0^\sigma\psi(x_{sr},y,t)\frac{ds}{s}\right)\frac{d\sigma}{\sigma}\\=&r^k\exp\left(\int_0^1\psi(\exp_y(\tau\exp_y^{-1}(x)),y,t)\frac{d\tau}{\tau}\right)\int_0^1 p_k(\exp_y(\sigma\exp_y^{-1}(x)),y,t)\\&\times\theta^{1/2}(\exp_y(\sigma\exp_y^{-1}(x)),y,t)\sigma^k\exp\left(-\int_0^\sigma\psi(\exp_y(s\exp_y^{-1}(x)),y,t)\frac{ds}{s}\right)\frac{d\sigma}{\sigma}\\=&r^k\tilde C_k(x,y,t).\end{aligned}$$

From the definition of $\tilde C_k$, the fact that $\psi\in C^\infty(U_\varepsilon\times[0,T])$, and $\psi(0,\theta,t)\equiv0$, it is immediate to recognize that

$$\tilde C_k(x,y,t)\in C^\infty(U_\varepsilon\times[0,T]).$$

Therefore, from (2.27) we obtain

$$u_k(x,y,t)=\theta^{-1/2}(x,y,t)\tilde C_k(x,y,t),$$

which shows that $u_k\in C^\infty(U_\varepsilon\times[0,T])$. This completes the proof of the claim, from which (2.17) follows.

Now we pick $\chi=C_0^\infty(\mathbf{R}^n\times\mathbf{R}^n)$, $0\leqq\chi\leqq1$, with $\chi\equiv1$ on $U_{\varepsilon/3}$, $\chi\equiv0$ outside $U_{(2/3)\varepsilon}$. We set

$$H_k(x,y,t)=\chi(x,y)\gamma_k(x,y,t), \tag{2.29}$$

where γ_k is defined by (2.6). If we extend γ_k with zero outside $U_\varepsilon\times(0,T)$, we have

$$H_k\in C^\infty(\mathbf{R}^n\times\mathbf{R}^n\times(0,T)).$$

A computation and (2.17) give

$$LH_k = t^k \chi G \Phi_{k+1} + \gamma_k \operatorname{div}(A\nabla_x \chi) + 2(A\nabla_x \chi) \cdot \nabla_x \gamma_k. \tag{2.30}$$

Let $\pi_1 : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ be the map $\pi_1(x, y) = x$. For $\varphi \in C_0^\infty(\pi_1(U_{\varepsilon/3}))$ we set

$$\Gamma_k \varphi(x, t) = \int_{\mathbf{R}^n} [\Gamma(x, y, t) - H_k(x, y, t)] \varphi(y) dy. \tag{2.31}$$

Since Γ is the fundamental solution of L we have $L\left(\int_{\mathbf{R}^n} \Gamma(\cdot, y, \cdot)\varphi(y)dy\right) = 0$. Therefore, (2.30) yields

$$\begin{aligned} L(\Gamma_k \varphi)(x, t) &= -L(H_k \varphi)(x, t) = -\int_{\mathbf{R}^n} LH_k(x, y, t)\varphi(y)dy \\ &= -t^k \int_{\mathbf{R}^n} \chi(x, y) G(x, y, t) \Phi_{k+1}(x, y, t) \varphi(y) dy \\ &\quad - \int_{\mathbf{R}^n} \gamma_k(x, y, t) \operatorname{div}(A\nabla_x \chi)(x, y, t)\varphi(y)dy \\ &\quad -2 \int_{\mathbf{R}^n} [(A\nabla_x \chi) \cdot \nabla_x \gamma_k](x, y, t)\varphi(y)dy \\ &= f_{k,1}(x, t) + f_{k,2}(x, t) + f_{k,3}(x, t) = f_k(x, t). \end{aligned} \tag{2.32}$$

Using the fact that there exist two positive constants α, β such that for every $(x, y, t) \in U_{(2/3)\varepsilon} \times [0, T]$

$$\alpha |x - y|^2 \leqq d^2(x, y, t) \leqq \beta |x - y|^2, \tag{2.33}$$

it is not difficult to see that the following holds: there exist $C > 0$ and $\gamma > 0$ such that

$$\begin{cases} |f_k(x, t)| \leqq C \exp(\gamma |x|^2), & \text{for every } (x, t) \in \mathbf{R}^n \times (0, T); \\ |\Gamma_k \varphi(x, t)| \leqq C \exp(\gamma |x|^2), & \text{for every } (x, t) \in \mathbf{R}^n \times (0, T); \\ \lim_{t \to 0^+} \Gamma_k \varphi(x, t) = 0, & \text{for every } x \in \mathbf{R}^n. \end{cases}$$

This implies, by uniqueness of the solution of Cauchy problem (see [Fr]) for T small enough

$$\begin{aligned} \Gamma_k \varphi(x, t) &= -\int_0^t \int_{\mathbf{R}^n} \Gamma(x, t; \xi, \tau) f_k(\xi, \tau) d\xi d\tau \\ &= -\sum_{i=1}^3 \int_0^t \int_{\mathbf{R}^n} \Gamma(x, t; \xi, \tau) f_{k,i}(\xi, \tau) d\xi d\tau = I_k + II_k + III_k. \end{aligned} \tag{2.34}$$

Now

$$\begin{aligned} I_k &= \int_{\mathbf{R}^n} \varphi(y) dy \left(\int_0^t \int_{\mathbf{R}^n} \Gamma(x, t; \xi, \tau) G(\xi, y, \tau) \tau^k \chi(\xi, y) \Phi_{k+1}(\xi, y, \tau) d\xi d\tau \right) \\ &= \int_{\mathbf{R}^n} \varphi(y) A_{k,1}(x, y, t) dy, \end{aligned} \tag{2.35}$$

the exchange of order of integration in I_k being possible in virtue of Fubini's theorem and the fact that there exists a $\delta > 0$ such that

$$\begin{aligned} A_{k,1}(x, y, t) &= \int_0^t \int_{\mathbf{R}^n} \Gamma(x, t; \xi, \tau) G(\xi, y, \tau) \tau^k \chi(\xi, y) \Phi_{k+1}(\xi, y, \tau) d\xi d\tau \\ &= t^{k+1-n/2} \exp\left(-\delta \frac{|x - y|^2}{4t} \right) w_k(x, y, t), \end{aligned} \tag{2.36}$$

where $w_k(x,y,t)=0(1)$ uniformly for $(x,y,t)\in \bar{U}_{\varepsilon/3}\times[0,T]$. To see this recall that $\operatorname{supp}\chi\subset \bar{U}_{(2\varepsilon/3)}$ and that $\Phi_{k+1}\in C^\infty(U_\varepsilon\times[0,T])$, therefore

$$|A_{k,1}(x,y,t)|\leqq \sup_{\bar{U}_{(2/3)\varepsilon}\times[0,T]}|\Phi_{k+1}|\int_0^t\int_{\mathbf{R}^n}\Gamma(x,t;\xi,\tau)G(\xi,y,\tau)\tau^k d\xi d\tau$$

(by [Fr, Th. 11, p. 24] and (2.33))

$$\leqq C\int_0^t\int_{\mathbf{R}^n}(t-\tau)^{-n/2}\exp\left(-\frac{\delta|x-\xi|^2}{4(t-\tau)}\right)$$

$$\times\tau^{-(n/2-k)}\exp\left(-\frac{\delta|\xi-y|^2}{4\tau}\right)d\xi d\tau$$

$$=C't^{k+1-n/2}\exp\left(-\frac{\delta|x-y|^2}{4t}\right).$$

In the last equality we have used Lemma 3 on p. 15 of [Fr]. By similar arguments it can be recognized that II_k and III_k can be written, respectively, as

$$II_k=\int_{\mathbf{R}^n}\varphi(y)A_{k,2}(x,y,t)dy \quad\text{and}\quad III_k=\int_{\mathbf{R}^n}\varphi(y)A_{k,3}(x,y,t)dy,$$

with $A_{k,i}$, $i=1,2$, satisfying estimates similar to (2.36). Setting $A_k=\sum_{i=1}^{3}A_{k,i}$ from our work above and the fact that φ is an arbitrary function in $C_0^\infty(\pi_1(U_{(2/3)\varepsilon}))$, we conclude that

$$\Gamma(x,y,t)=H_k(x,y,t)+A_k(x,y,t),\qquad (x,y,t)\in U_{(2/3)\varepsilon}\times(0,T),\tag{2.37}$$

with A_k satisfying the estimate

$$A_k(x,y,t)=0\left(t^{k+1-n/2}\exp\left(-\frac{\delta|x-y|^2}{4t}\right)\right),\tag{2.38}$$

for $(x,y,t)\in U_{\varepsilon/3}\times(0,T)$ and a certain $\delta>0$. Recalling now that $H_k=\chi\gamma_k$ and that $\chi\equiv1$ on $\bar{U}_{\varepsilon/3}$ we conclude from (2.37) that

$$\Gamma(x,y,t)=G(x,y,t)\sum_{j=0}^{k}t^j u_j(x,y,t)+A_k(x,y,t)$$

for $(x,y,t)\in U_{\varepsilon/3}\times(0,T)$, with A_k having the asymptotic behavior given by (2.38). This completes the proof of (2.1). Analogous arguments prove that the derivatives of Γ have similar asymptotic expansions.

3. A Class of Well-behaved Mean Value Formulas

In this section we generalize some mean value formulas relative to the operator L in (1.1) first found in [FG] and [GL]. The advantage of these formulas with respect to those found in [FG] and [GL] consists in the fact that the kernel appearing in them is not only bounded, but possesses a degree of regularity which can be made arbitrarily large. Our starting point is the observation that if $u\in C^2(\mathbf{R}^{n+1})$ and we define a new function in $\mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}$ by setting for $(x,t) \in \mathbf{R}^{n+1}$

$$\hat{u}(x,y,t) = u(x,t), \qquad y \in \mathbf{R}^m,$$

then with L as in (1.1) we have

$$(L + \Delta_y)\hat{u}(x,y,t) = Lu(x,t). \tag{3.1}$$

Therefore, we can apply to the function $\hat{u}$ the representation formulas established in [FG] or [GL]. Before stating the results we need to introduce some notation. If Γ is the fundamental solution of L, $z=(x,t)$, $\zeta=(\xi,\tau)$, and $r>0$ we set (cf. [GL])

$$E(z;\zeta) = \frac{A(\zeta)(\nabla_\xi \Gamma(z;\zeta)) \cdot \nabla_\xi \Gamma(z;\zeta)}{\Gamma^2(z;\zeta)}, \qquad E_r(z;\zeta) = (4\pi r)^{-n/2} E(z;\zeta). \tag{3.2}$$

Next we define for a fixed $m \in \mathbf{N}$

$$\Phi(z;\zeta) = (4\pi(t-\tau))^{-m/2}\Gamma(z;\zeta), \tag{3.3}$$

$$R_r^2(z;\zeta) = 4(t-\tau)\ln\left[(4\pi r)^{\frac{n+m}{2}}\Phi(z;\zeta)\right]. \tag{3.4}$$

Finally, we recall the definition (1.20) of the *modified parabolic ball* centered at z and of radius r. Using (3.3) we rewrite (1.20) as

$$\Omega_r^m(z) = \left\{\zeta \in \mathbf{R}^{n+1} \,|\, \Phi(z;\zeta) > (4\pi r)^{-\frac{n+m}{2}}\right\}. \tag{3.5}$$

Theorem 3.1. *Let $u \in C^2(\mathbf{R}^{n+1})$ and let $z \in \mathbf{R}^{n+1}$. Then for every $r>0$ we have*

$$\begin{aligned} &\omega_m (4\pi r)^{-\frac{n+m}{2}} \int_{\Omega_r^m(z)} u(\zeta) R_r^m(z;\zeta)\left[E(z;\zeta) + \frac{m}{m+2}\,\frac{R_r^2(z;\zeta)}{4(t-\tau)^2}\right] d\zeta \\ &= u(z) + \frac{n+m}{m+2} w_m (4\pi)^{-\frac{n+m}{2}} \int_0^r l^{-\left(\frac{n+m}{2}+1\right)} \int_{\Omega_l^m(z)} Lu(\zeta)\frac{R_l^{m+2}(z;\zeta)}{4(t-\tau)}\, d\zeta\, dl. \end{aligned} \tag{3.6}$$

In (3.6) ω_m denotes the measure of the unit ball in $\mathbf{R}^m$.

Proof. Our starting point is formula (1.37) of Theorem 1.6 in [GL] which we now recall.

$$\frac{d}{dr}u_r(z) = \frac{n}{2}(4\pi)^{-n/2} r^{-n/2-1} \int_{\Omega_r(z)} Lu(\zeta)\ln[(4\pi r)^{n/2}\Gamma(z;\zeta)]\,d\zeta, \tag{3.7}$$

where $\Omega_r(z)$ is the parabolic ball defined in (1.4) and we have set [see (3.2)]

$$u_r(z) = \int_{\Omega_r(z)} u(\zeta) E_r(z;\zeta)\, d\zeta. \tag{3.8}$$

By integration $\left[\text{keeping in mind that } \lim_{r\to 0^+} u_r(z) = u(z)\right]$, we obtain from (3.7)

$$u_r(z) = u(z) + \frac{n}{2}(4\pi)^{-n/2}\int_0^r l^{-(n/2+1)} \int_{\Omega_l(z)} Lu(\zeta)\ln[(4\pi l)^{n/2}\Gamma(z;\zeta)]\,d\zeta\, dl. \tag{3.9}$$

We now fix $m \in \mathbf{N}$ and for $y \in \mathbf{R}^m$ we denote by

$$\hat{L} = L + \Delta_y \tag{3.10}$$

the parabolic operator acting on the (x, y, t)-variables in $\mathbf{R}^{n+m+1}$. If $\hat{\Gamma}(x, y, t; \xi, \eta, \tau)$ is the fundamental solution of $\hat{L}$ in (3.10) one easily verifies that

$$\hat{\Gamma}(x, y, t; \xi, \eta, \tau) = \Gamma(x, t; \xi, \tau) K(y-\eta; t-\tau), \tag{3.11}$$

where

$$K(y-\eta; t-\tau) = \begin{cases} (4\pi(t-\tau))^{-m/2} \exp\left(-\frac{|y-\eta|^2}{4(t-\tau)}\right), & t > \tau \\ 0, & t \leqq \tau. \end{cases} \tag{3.12}$$

Let now $\hat{z}$ denote the point (x, y, t) in $\mathbf{R}^{n+m+1}$. If $\hat{u}(\hat{z}) = u(z)$, then by (3.1) $\hat{L}\hat{u}(\hat{z}) = Lu(z)$. Therefore, if we apply the $(n+m+1)$-dimensional version of (3.9) to $\hat{u}$ and $\hat{L}$, we obtain

$$(\hat{u})_r(\hat{z}) = u(z) + \frac{n+m}{2} (4\pi)^{-\frac{n+m}{2}} \int_0^r l^{-\left(\frac{n+m}{2}+1\right)} \times \int_{\hat{\Gamma}(\hat{z};\hat{\zeta}) > (4\pi l)^{-\frac{n+m}{2}}} Lu(\zeta) \ln\left[(4\pi l)^{\frac{n+m}{2}} \hat{\Gamma}(\hat{z};\hat{\zeta})\right] d\hat{\zeta}\, dl, \tag{3.13}$$

where the notation in (3.13) means that the inner integral is performed over the set in $\mathbf{R}^{n+m+1}$

$$\left\{\hat{\zeta} \in \mathbf{R}^{n+m+1} \,|\, \hat{\Gamma}(\hat{z};\hat{\zeta}) > (4\pi l)^{-\frac{n+m}{2}}\right\}.$$

Because of (3.11), keeping in mind (3.3), (3.4), we have

$$\begin{aligned} &\int_{\hat{\Gamma}(\hat{z};\hat{\zeta}) > (4\pi l)^{-\frac{n+m}{2}}} Lu(\zeta) \ln\left[(4\pi l)^{\frac{n+m}{2}} \hat{\Gamma}(\hat{z};\hat{\zeta})\right] d\hat{\zeta} \\ &= \int_{\Omega_l^m(z)} Lu(\zeta) \left(\int_{|y-\eta| < R_l(z;\zeta)} \ln\left[(4\pi l)^{\frac{n+m}{2}} \hat{\Gamma}(\hat{z};\hat{\zeta})\right] d\eta \right) d\zeta \\ &= \int_{\Omega_l^m(z)} Lu(\zeta) \left(\int_{|y-\eta| < R_l(z;\zeta)} \frac{1}{4(t-\tau)} [R_l^2(z;\zeta) - |y-\eta|^2] d\eta \right) d\zeta \\ &= \frac{\omega_m}{2(m+2)(t-\tau)} \int_{\Omega_l^m(z)} Lu(\zeta) R_l^{m+2}(z;\zeta) d\zeta. \end{aligned} \tag{3.14}$$

From (3.8) and (3.2) we have

$$(\hat{u})_r(\hat{z}) = (4\pi r)^{-\frac{n+m}{2}} \int_{\hat{\Gamma}(\hat{z};\hat{\zeta}) > (4\pi r)^{-\frac{n+m}{2}}} u(\zeta) \hat{E}(\hat{z};\hat{\zeta}) d\hat{\zeta}. \tag{3.15}$$

Using again (3.11) we obtain

$$\hat{E}(\hat{z};\hat{\zeta}) = E(z;\zeta) + \frac{|y-\eta|^2}{4(t-\tau)^2}. \tag{3.16}$$

Replacing (3.16) in (3.15) and proceeding along the same lines as above, we end up with

$$(\hat{u})_r(\hat{z}) = \omega_m (4\pi r)^{-\frac{n+m}{2}} \int_{\Omega_r^m(z)} u(\zeta) R_r^m(z;\zeta) \left[E(z;\zeta) + \frac{m}{m+2} \frac{R_r^2(z;\zeta)}{4(t-\tau)^2} \right] d\zeta. \tag{3.17}$$

Inserting (3.14), (3.17) into (3.13) finally gives (3.6).

Remark 3.1. We emphasize that if $m=0$ in Theorem 3.1, then (3.6) reduces to (1.31) of Theorem 1.5 in [GL] if we agree to let $\omega_0=1$.

Remark 3.2. The idea of climbing up in the dimension by adding variables is not new. Kupcov had already employed it in [Ku] to obtain a mean-value formula with a well-behaved kernel for solutions of $Hu=\Delta u-D_t u=0$.

From now on, to avoid clumsy notation we set for $m\in\mathbf{N}$ and $z,\zeta\in\mathbf{R}^{n+1}$

$$E_r^{(m)}(z;\zeta)=(4\pi r)^{-\frac{n+m}{2}}\omega_m R_r^m(z;\zeta)\left[E(z;\zeta)+\frac{m}{m+2}\frac{R_r^2(z;\zeta)}{4(t-\tau)^2}\right]. \tag{3.18}$$

If we denote by $u_r^{(m)}(z)$ the average $(\hat{u})_r(\hat{z})$ appearing in (3.17) [observe that (3.17) says that $(\hat{u})_r(\hat{z})$ is in fact a function of z alone], then (3.17) can be rewritten as

$$u_r^{(m)}(z)=\int_{\Omega_r^m(z)} u(\zeta)E_r^{(m)}(z;\zeta)d\zeta. \tag{3.19}$$

Again, one should note that if $m=0$ (3.19) reduces to (3.8).

We close this section by establishing an average formula for solutions of $Lu=0$ which will turn out to be very useful when we will study the smoothing of superparabolic functions in Sect. 6. Such a formula is easily obtained by superposition from (3.19), by suitably adapting the idea in the proof of H. Weyl's lemma on harmonic functions (c.f., e.g., [F], p. 92). In what follows we choose and fix a function $\varphi\in C_0^\infty(\mathbf{R}^+)$ such that $\varphi\geqq 0$, $\operatorname{supp}\varphi\subset(1,2)$ and $\int_0^{+\infty}\varphi(r)dr=1$. For $m\in\mathbf{N}$ and $z\in\mathbf{R}^{n+1}$ we define

$$J_r^{(m)}(u)(z)=\int_0^{+\infty} u_l^{(m)}(z)\varphi\left(\frac{l}{r}\right)\frac{dl}{r} \tag{3.20}$$

where $u\in L^\infty_{\mathrm{loc}}(\mathbf{R}^{n+1})$ and $u_r^{(m)}$ is as in (3.19). Substituting for $u_r^{(m)}$ its expression given by (3.19) and exchanging the order of integration, we obtain from (3.20)

$$J_r^{(m)}(u)(z)=\int_{\mathbf{R}^{n+1}} u(\zeta)M_r^{(m)}(z;\zeta)d\zeta, \tag{3.21}$$

where we have set for $z=(x,t)$, $\zeta=(\xi,\tau)$

$$M_r^{(m)}(z;\zeta)=\int_{[4\pi\Phi(z;\zeta)^{\frac{2}{n+m}}]^{-1}}^{+\infty} E_l^{(m)}(z;\zeta)\varphi\left(\frac{l}{r}\right)\frac{dl}{r} \tag{3.22}$$

if $t>\tau$, whereas $M_r^{(m)}(z;\zeta)=0$ for $t\leqq\tau$. Since φ is compactly supported in $\mathbf{R}^+$, $M_r^{(m)}(z;\cdot)$ is supported in a parabolic ball $\Omega_{r_0}^m(z)$ centered at z [see (3.5)]. We will show in Sect. 6 that as a function of z the kernel $M_r^{(m)}$ can be made arbitrarily regular by choosing m large enough.

4. An Elementary Proof of Harnack's Inequality

An immediate consequence of Theorem 3.1 is that if u is a (smooth) solution of $Lu=0$ in $\mathbf{R}^{n+1}$, then for every $z\in\mathbf{R}^{n+1}$ and every $r>0$ we have [see (3.19)]

$$u(z)=u_r^{(m)}(z)=\int_{\Omega_r^m(z)} u(\zeta)E_r^{(m)}(z;\zeta)d\zeta, \tag{4.1}$$

where $m \in \mathbf{N}$ is arbitrarily fixed. The advantage of formula (4.1) with respect to (1.18) is that the kernel $E_r^{(m)}(z;\cdot)$ is bounded from above on the parabolic ball $\Omega_r^{(m)}(z)$ by an appropriate power of r, provided that m is large enough. When $m=0$, $E_r^{(0)}(z;\zeta)=E(z;\zeta)$ [see (3.2)] and (4.1) reduces to (1.18). As observed in Sect. 1 $E(z;\cdot)$ is not bounded on $\Omega_r(z)$.

In this section to illustrate the good feature of formula (4.1), we present an elementary proof of Harnack's inequality for parabolic equations which does not use the parabolic BMO machinery developed by Moser in his classical paper [M]. Our approach is direct and imitates the proof of Harnack's inequality for harmonic functions. In what follows we will use the intrinsic notation introduced in Sect. 2. In virtue of Theorem 2.1 for any $\theta>1$ we can find $r_0=r_0(\theta,L)>0$, L as in (1.1), such that for any $r \leqq r_0$, $z \in \mathbf{R}^{n+1}$ and $\zeta \in \Omega_r^{(m)}(z)$, $m \in \mathbf{N}$, we have

$$\theta^{-1}G(z;\zeta) \leqq \Gamma(z;\zeta) \leqq \theta G(z;\zeta). \tag{4.2}$$

where $G(z;\zeta)$ is the generalized Gaussian introduced in Theorem 2.1. Using (4.2) it is easy to check that: *for every $\varepsilon>0$ and a fixed $m \in \mathbf{N}$ there exists $\delta=\delta(L,m,\varepsilon)>0$ such that given $z_0=(x_0,t_0) \in \mathbf{R}^{n+1}$, $r \leqq \frac{r_0}{2}$, and $z=(x,t) \in \Omega_r^m(z_0)$, if $t_0-t \geqq \varepsilon r$ then*

$$\Omega_{\delta r}^m(z) \subset \Omega_{2r}^m(z_0). \tag{4.3}$$

We will use (4.3) in the proof of Theorem 4.1 below.

Theorem 4.1 (Harnack's inequality). *Let $D \subset \mathbf{R}^{n+1}$ be an open set and let $u \geqq 0$ be a solution of $Lu=0$ in D. Given $z_0 \in D$ let $r \leqq \frac{r_0}{4}$ be such that $\Omega_{4r}^m(z_0) \subset D$. If $\varepsilon>0$ let $z \in \Omega_r^m(z_0)$ be such that $t_0-t \geqq \varepsilon r$. Then there exists a constant $C=C(L,m,\varepsilon)>0$ such that*

$$u(z) \leqq Cu(z_0). \tag{4.4}$$

Remark 4.1. In the statement above and throughout the discussion below, m is a fixed integer with $m>2$.

Proof of Theorem 4.1. By Theorem 3.1 we have

$$u(z_0)=\int_{\Omega_{3r}^m(z_0)} u(\zeta)E_{3r}^{(m)}(z_0;\zeta)d\zeta,$$

$E_{3r}^{(m)}$ being defined by (3.18). By the positivity of u and the observation (4.3) we obtain the following inequalities

$$\begin{aligned} u(z_0) &\geqq \int_{\Omega_{\delta r}^m(z)} u(\zeta)E_{3r}^{(m)}(z_0;\zeta)d\zeta \\ &= \int_{\Omega_{\delta r}^m(z)} u(\zeta)\frac{E_{3r}^{(m)}(z_0;\zeta)}{E_{\delta r}^{(m)}(z;\zeta)}E_{\delta r}^{(m)}(z;\zeta)d\zeta. \end{aligned} \tag{4.5}$$

In order to get (4.4) all we have to show is the existence of two numbers $C_i=C_i(L,m,\varepsilon)>0$, $i=1,2$, such that

$$\inf_{\zeta \in \Omega_{\delta r}^m(z)} E_{3r}^{(m)}(z_0:\zeta) \geqq C_1 r^{-(n/2+1)}, \tag{4.6}$$

$$\sup_{\zeta \in \Omega_{\delta r}^m(z)} E_{\delta r}^{(m)}(z;\zeta) \leqq C_2 r^{-(n/2+1)}. \tag{4.7}$$

Recalling (3.18) we have for $\zeta \in \Omega^m_{\delta r}(z)$

$$E^{(m)}_{3r}(z_0;\zeta) \geqq C_m r^{-\frac{n+m}{2}} \frac{R^{m+2}_{3r}(z_0;\zeta)}{(t_0-\tau)^2}, \tag{4.8}$$

for a certain $C_m > 0$. Now using the definition of $R_{3r}(z_0;\zeta)$ [see (3.4)] and the fact that

$$(12\pi r)^{\frac{n+m}{2}} \Phi(z_0;\zeta) \geqq \left(\tfrac{3}{2}\right)^{\frac{n+m}{2}} \quad \text{on } \Omega^m_{2r}(z_0),$$

we obtain for $\zeta \in \Omega^m_{\delta r}(z)$

$$E^{(m)}_{3r}(z_0;\zeta) \geqq C'_m r^{-\frac{n+m}{2}} (t_0-\tau)^{m/2-1}. \tag{4.9}$$

Finally, recalling that for $\zeta \in \Omega^m_{\delta r}(z)$ we have $(t_0-\tau) \geqq (t_0-t) \geqq \varepsilon r$, we obtain (4.6) from (4.9).

Now we look at (4.7). We claim that as a consequence of the results in Theorem 2.1 the following estimate for the kernel E_r holds

$$\frac{A(\zeta)(\nabla\Gamma_\xi(z;\zeta)) \cdot \nabla\Gamma_\xi(z;\zeta)}{\Gamma^2(z;\zeta)} \leqq C\left(\frac{d^2(x,\xi,\tau)}{(t-\tau)^2} + 1\right), \tag{4.10}$$

uniformly for $z=(x,t) \in \mathbf{R}^{n+1}$, $\zeta=(\xi,\tau) \in \Omega^m_r(z)$, and $r \leqq r_0$, where $C>0$ depends only on L. We outline the proof of (4.10). For complete details one should see the proof of Lemma 2.1 in [GL]. Let us fix a $k \in \mathbf{N}$ sufficiently large, for instance $k > \dfrac{m+n}{2} + 10^3$. Then from Theorem 2.1, if $z=(x,t) \in \mathbf{R}^{n+1}$ is fixed and $\zeta=(\xi,\tau) \in \Omega^m_r$, $r \leqq r_0$, we can write

$$\Gamma(z;\zeta) = \gamma_k(z;\zeta) + w_k(z;\zeta), \tag{4.11}$$

$$D_{\xi_j}\Gamma(z;\zeta) = D_{\xi_j}\gamma_k(z;\zeta) + w_{k,j}(z;\zeta), \quad j=1,\dots,n, \tag{4.12}$$

where γ_k is defined as in (2.6) and $w_k(z;\cdot)$, $w_{k,j}(z;\cdot)$, $j=1,\dots,n$, are $(n+1)$ functions in $\Omega^m_r(z)$ such that $w_k(z;\cdot)=0(r)$, $w_{k,j}(z;\cdot)=0(r)$, $j=1,\dots,n$, as $r \to 0$. Now let us observe that in the notation of Sect. 2 (4.10) becomes

$$\frac{|\nabla_{M_\tau}\Gamma(z;\zeta)|^2_\tau}{\Gamma^2(z;\zeta)} \leqq C\left(\frac{d^2(x,\xi,\tau)}{(t-\tau)^2} + 1\right). \tag{4.13}$$

By (4.12) we have

$$|\nabla_{M_\tau}\Gamma|^2_\tau = |\nabla_{M_\tau}\gamma_k|^2_\tau + 2\langle \nabla_{M_\tau}\gamma_k, A(\tau)\vec{w}_k\rangle_\tau + |A(\tau)\vec{w}_k|^2_\tau, \tag{4.14}$$

where we have set $\vec{w}_k = (w_{k,1}, w_{k,2}, \dots, w_{k,n})$. Replacing in (4.14) the expression for γ_k, using the fact that the function u_0 in the expansion (2.1) is 1 uniformly in x, ξ at the initial time $t=\tau$, and (4.2), we finally obtain (4.13), hence (4.10). For more details one can look at the proof of Lemma 2.1 in [GL] or at Sect. 6 below. Next we observe that because of (2.33) by (3.2) and (4.10) we get

$$E_r(z;\zeta) \leqq C r^{-n/2}\left(\frac{|x-\xi|^2}{(t-\tau)^2} + 1\right), \tag{4.15}$$

for every $z \in \mathbf{R}^{n+1}$, $\zeta \in \Omega_r^m(z)$, and $r \leqq r_0$. Also (3.3), (3.4), and (4.2) yield for a suitable $C>0$

$$R_r^2(z;\zeta) \leqq C(t-\tau)\ln\left(\frac{Cr}{t-\tau}\right), \tag{4.16}$$

uniformly in $z \in \mathbf{R}^{n+1}$, $\zeta \in \Omega_r^m(z)$, and $r \leqq r_0$. We now take δr instead of r. Replacing then (4.15), (4.16) in (3.18) we finally obtain

$$E_{\delta r}^{(m)}(z;\zeta) \leqq Cr^{-\frac{n+m}{2}}(t-\tau)^{m/2}\left[\ln\left(\frac{Cr}{t-\tau}\right)\right]^{m/2}\left[1+\frac{1}{t-\tau}\ln\left(\frac{Cr}{t-\tau}\right)\right].$$

The right-hand side of the last inequality is bounded by $Cr^{-(n/2+1)}$. This proves (4.7), and therefore the theorem.

5. Averaging of Superparabolic Functions

In this section we study the averaging operators $u \mapsto u_r^{(m)}$ introduced in Sect. 3 [see (3.19)] and their connection with L-superparabolic functions. In [FG] it was proved that if $u \in C^\infty(\mathbf{R}^{n+1})$ is a solution of $Lu=0$ in $\mathbf{R}^{n+1}$, then for every $z \in \mathbf{R}^{n+1}$ $u(z) \equiv u_r(z)$, for every $r>0$. In virtue of Theorem 3.1 a similar identity holds for the averages $u_r^{(m)}$. Vice-versa, by the maximum principle it can be easily seen that if $u \in C(\mathbf{R}^{n+1})$ and $u(z)=u_r^{(m)}(z)$ for every $z \in \mathbf{R}^{n+1}$ and every $r>0$, then $u \in C^\infty(\mathbf{R}^{n+1})$ and $Lu=0$. In what follows we will use the notion of L-superparabolic function given in Sect. 1.

Proposition 5.1. *Let $u: \mathbf{R}^{n+1} \to \bar{R}$ be a l.s.c. function, and let $m \in \mathrm{N} \cup \{0\}$. The following statements are equivalent:*

(i) *u is L-superparabolic in $\mathbf{R}^{n+1}$;*

(ii) *there exists $r_0 = r_0(L) > 0$ such that for every $z \in \mathbf{R}^{n+1}$ and $r \leqq r_0$*

$$u(z) \geqq u_r^{(m)}(z). \tag{5.1}$$

To prove Proposition 5.1 we will need the following.

Lemma 5.1. *There exists $r_0 = r_0(L) > 0$ such that if $z \in \mathbf{R}^{n+1}$, $r \leqq r_0$, and $\Omega_r(z)$ is the L-parabolic ball* (1.4), *then every point of $\Psi_r(z) \setminus \{z\}$ is L-regular.*

Proof. Because of the assumption (1.3) on L and the asymptotic estimates (4.11) and (4.12) we can find $r_0 = r_0(L) > 0$ small enough such that for $z \in \mathbf{R}^{n+1}$, $r \leqq r_0$ and $\zeta \in \Psi_r(z) \setminus \{z\}$ we have

$$|\nabla_\xi \Gamma(z;\zeta)| \neq 0 \quad \text{or} \quad D_\tau \Gamma(z;\zeta) < 0.$$

This is enough to reach the conclusion.

Proof of Proposition 5.1. That (ii) implies (i) is a consequence of the fact that a function for which (5.1) holds satisfies the strong minimum principle. To show that (i) implies (ii), we first argue in the case $m=0$. In what follows r_0 is fixed as in Lemma 5.1. For $r \leqq r_0$, let $\varphi \in C(\Psi_r(z))$ with $\varphi \leqq u$ on $\Psi_r(z)$. We

Claim.

$$u(z) \geqq \int_{\Psi_r(z)} \varphi(\zeta) Q_r(z;\zeta) dH_n(\zeta),$$

where $Q_r(z;\zeta)$ is the kernel in (1.14). Suppose the claim is true. Taking the supremum over all continuous φ's on $\Psi_r(z)$ such that $\varphi \leqq u$ on $\Psi_r(z)$ we obtain

$$u(z) \geqq \int_{\Psi_r(z)} u(\zeta)Q_r(z;\zeta)dH_n(\zeta). \tag{5.2}$$

Proceeding now as in the proof of Theorem 2 in [FG] we obtain from (5.2) $u(z) \leqq u_r(z)$ for any $r \leqq r_0$, which proves that (i) implies (ii) in the case $m=0$. Let now $m \in \mathbf{N}$. We use the notation of Sect. 3. By thinking of u as a function in $\mathbf{R}^{n+m+1}$, i.e., setting $\hat{u}(x,y,t)=u(x,t)$, then $\hat{u}$ is $\hat{L}$-superparabolic with $\hat{L}$ as in (3.10). Therefore, from the discussion of the case $m=0$ we have

$$\hat{u}(\hat{z}) \geqq (\hat{u})_r(\hat{z}), \tag{5.3}$$

for every $\hat{z} \in \mathbf{R}^{n+m+1}$ and every $r \leqq r_0$. We remark that r_0 does not depend on m, and therefore it can be taken to be the same as in Lemma 5.1. (5.3) is another way to write (5.1), if one observes that $(\hat{u})_r(\hat{z})=u_r^{(m)}(z)$, see the proof of Theorem 3.1.

We are thus left with the proof of the claim. To this end we place a tiny $(n+1)$-dimensional right circular cone on the top of $\Psi_r(z)$ so to cut out the irregular point $z=(x,t)$. The axis of the cone is along the t-axis and the vertex at the point $N_\varepsilon=(x,t+\varepsilon)$, $\varepsilon>0$. More precisely, let

$$B_\varepsilon(z)=\{\zeta \in \mathbf{R}^{n+1} \mid |\xi-x|^2+(\tau-t)^2 \leqq \varepsilon^2\}$$

be the closed $(n+1)$-ball of radius ε centered at z. Let $\Sigma_{r,\varepsilon}=B_\varepsilon(z) \cap \Psi_r(z)$ and $\partial\Sigma_{r,\varepsilon}(z) = \partial B_\varepsilon(z) \cap \Psi_r(z)$. From N_ε we project a cone onto $\partial\Sigma_{r,\varepsilon}(z)$. We denote this cone by $C_{r,\varepsilon}(z)$ and define

$$\tilde{\Psi}_{r,\varepsilon}(z)=[\Psi_r(z) \setminus \Sigma_{r,\varepsilon}(z)] \cup C_{r,\varepsilon}(z).$$

Now for φ chosen as in the claim we let $\tilde{\varphi} \in C(\tilde{\Psi}_{r,\varepsilon}(z))$ be such that $\tilde{\varphi} \equiv \varphi$ on $\Psi_r(z) \setminus \Sigma_{r,\varepsilon}(z)$. As $\tilde{\Psi}_{r,\varepsilon}(z)$ is the boundary of an L-regular bounded open set in $\mathbf{R}^{n+1}$, we let $H_{\tilde{\varphi}}$ denote the solution of the Dirichlet problem for it with boundary datum $\tilde{\varphi}$. Since u is L-superparabolic and $LH_{\tilde{\varphi}}=0$, using (1.8) we obtain

$$\begin{aligned} u(z) \geqq H_{\tilde{\varphi}}(z) &= \int_{\Psi_r(z)} H_{\tilde{\varphi}}(\zeta)Q_r(z;\zeta)dH_n(\zeta) \\ &= \int_{\Sigma_{r,\varepsilon}(z)} H_{\tilde{\varphi}}(\zeta)Q_r(z;\zeta)dH_n(\zeta) + \int_{\Psi_r(z)\setminus\Sigma_{r,\varepsilon}(z)} \varphi(\zeta)Q_r(z;\zeta)dH_n(\zeta). \end{aligned} \tag{5.4}$$

As $\varepsilon \to 0^+$ the first integral in the right-hand side of (5.4) goes to zero, whereas the second integral tends to the right-hand side of (5.2). This proves the claim, and therefore Proposition 5.1.

Proposition 5.2. *Let u be a L-superparabolic function on $\mathbf{R}^{n+1}$. Then if r_0 is as in Lemma 5.1 we have*

(i) $u_\varrho(z) \leqq u_r(z)$, *for every $z \in \mathbf{R}^{n+1}$ and every $r<\varrho \leqq r_0$;*
(ii) $u_r(z) \nearrow u(z)$ *as $r \to 0^+$, for every $z \in \mathbf{R}^{n+1}$;*
(iii) u_r *is L-superparabolic on $\mathbf{R}^{n+1}$ for every $r \leqq r_0$.*

Proof. We begin by remarking that we can assume that u is the L-potential of a compactly supported nonnegative measure μ on $\mathbf{R}^{n+1}$, i.e., $u=\Gamma_\mu$ where

$$\Gamma_\mu(z)=\int_{\mathbf{R}^{n+1}} \Gamma(z;\zeta)d\mu(\zeta),$$

and Γ is the fundamental solution of L in (1.1). In fact, if u is L-superparabolic on $\mathbf{R}^{n+1}$, then $u \in L^1_{\mathrm{loc}}(\mathbf{R}^{n+1})$ and $Lu \leqq 0$ in the sense of $\mathscr{D}'(\mathbf{R}^{n+1})$. Therefore, there exists a nonnegative measure ν on $\mathbf{R}^{n+1}$ such that $Lu = -\nu$. Now, if D is an arbitrary bounded open set in $\mathbf{R}^{n+1}$ and if we set $\mu = \nu|_{\bar{D}}$, then $L(u - \Gamma_\mu) = 0$ in D, and therefore $u = \Gamma_\mu + h$ in D, where $Lh = 0$ in D. From this representation of u, and the fact that h coincides with its parabolic average at every point $z \in D$, we conclude that it is sufficient to prove (i), (ii), and (iii) above in the case in which $u = \Gamma_\mu$.

We begin with (i). Let $z = (x, t) \in \mathbf{R}^{n+1}$ and consider the parabolic spheres centered at z $\Psi_r(z)$, $\Psi_\varrho(z)$. We want to prove that for $r < \varrho \leqq r_0$

$$\int_{\Psi_\varrho(z)} u(\zeta) Q_\varrho(z;\zeta) dH_n(\zeta) \leqq \int_{\Psi_r(z)} u(\zeta) Q_r(z;\zeta) dH_n(\zeta), \tag{5.5}$$

where Q_r has the same meaning as in (5.2). In the sequel we will denote by $u^\sigma_r(z)$ and $u^\sigma_\varrho(z)$ the surface averages appearing respectively in the left-hand side and in the right-hand side of (5.5). Since by a result of H. Bauer every L-superparabolic function is the pointwise limit of a monotone sequence of continuous L-superparabolic functions (see [Ba, Satz 2.5.8]), we may assume without loss of generality that u is continuous. Now for $\varepsilon > 0$ fixed we perform the same cutting and pasting as in the proof of Proposition 5.1. Letting U denote the L-parabolic extension of u to the set $\tilde{\Omega}_{\varrho,\varepsilon}(z)$, for which $\tilde{\Psi}_{\varrho,\varepsilon}(z) = \partial \tilde{\Omega}_{\varrho,\varepsilon}(z)$, we finally obtain

$$\begin{aligned} u^\sigma_r(z) &\geqq \int_{\Psi_r(z)} U(\zeta) Q_r(z;\zeta) dH_n(\zeta) = U(z) \\ &= \int_{\Psi_\varrho(z)} U(\zeta) Q_\varrho(z;\zeta) dH_n(\zeta) = \int_{\Sigma_{\varrho,\varepsilon}(z)} U(\zeta) Q_\varrho(z;\zeta) dH_n(\zeta) \\ &\quad + \int_{\Psi_\varrho(z) \setminus \Sigma_{\varrho,\varepsilon}(z)} u(\zeta) Q_\varrho(z;\zeta) dH_n(\zeta). \end{aligned} \tag{5.6}$$

Letting $\varepsilon \to 0^+$ we obtain (5.5) from (5.6). To complete the proof of (i) we will show that $u^\sigma_\varrho \leqq u^\sigma_r$ implies $u_\varrho \leqq u_r$. From the results in [FG] we obtain for every $r > 0$

$$\begin{aligned} u_r(z) &= \frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} \left(\int_{\Psi_l(z)} u(\zeta) Q_l(z;\zeta) dH_n(\zeta) \right) dl \\ &= \frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} u^\sigma_l(z) dl. \end{aligned} \tag{5.7}$$

Differentiating (5.7) with respect to r we have

$$\frac{d}{dr} u_r(z) = \frac{n}{2} r^{-1} [u^\sigma_r(z) - u_r(z)]. \tag{5.8}$$

On the other hand, since by (5.5) $r \mapsto u^\sigma_r(z)$ is decreasing, we have

$$\begin{aligned} u_r &= \frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} u^\sigma_l(z) dl \geqq \frac{n}{2} r^{-n/2} u^\sigma_r(z) \int_0^r l^{n/2-1} dl \\ &= u^\sigma_r(z). \end{aligned} \tag{5.9}$$

Using (5.9) in (5.8) we conclude that

$$\frac{d}{dr} u_r(z) \leqq 0.$$

This proves (i). The proof of (ii) is a standard consequence of the lower semicontinuity of u, and we omit it. The proof of (iii) is more delicate. We first establish the following.

Lemma 5.2. *With r_0 as in Lemma 5.1 for every $\varrho, r \leqq r_0$ and every $\zeta, z \in \mathbf{R}^{n+1}$ we have*

$$((\Gamma(\cdot;\zeta))_r)_\varrho(z) \leqq (\Gamma(\cdot;\zeta))_r(z). \tag{5.10}$$

In particular, (5.10) *implies that for every $\zeta \in \mathbf{R}^{n+1}$ the function $z \mapsto (\Gamma(\cdot;\zeta))_r(z)$ is L-superparabolic for every $r \leqq r_0$.*

Proof. For $\zeta \in \mathbf{R}^{n+1}$ fixed we set $w = \Gamma(\cdot;\zeta)$. By w_r^σ we will denote the parabolic surface average of w. The proof of (5.10) is based on the following.

Claim. For every $r \leqq r_0$ and every $z \in \mathbf{R}^{n+1}$ such that $\Gamma(z;\zeta) \neq (4\pi r)^{-n/2}$ we have

$$w_r^\sigma(z) = \min\{\Gamma(z;\zeta), (4\pi r)^{-n/2}\}.$$

Let us take the claim for granted and use it to prove the lemma. First, we observe that the function

$$v_r = \min\{\Gamma(\cdot;\zeta), (4\pi r)^{-n/2}\},$$

being the minimum of two L-superparabolic functions, is L-superparabolic in $\mathbf{R}^{n+1}$. Then by Proposition 5.1

$$(v_r)_\varrho \leqq v_r \tag{5.11}$$

for every $\varrho \leqq r_0$. On the other hand the claim gives

$$w_r^\sigma(z) = v_r(z) \quad \text{if} \quad \Gamma(z;\zeta) \neq (4\pi r)^{-n/2}. \tag{5.12}$$

Now using (5.7) for w_r we have from (3.8) for every $z_0 \in \mathbf{R}^{n+1}$, $\varrho, r \leqq r_0$

$$\begin{aligned}
(w_r)_\varrho(z_0) &= \int_{\Omega_\varrho(z_0)} w_r(z) E_\varrho(z_0;\zeta)dz = \int_{\Omega_\varrho(z_0)} \left[\frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} w_l^\sigma(z)dl\right] E_\varrho(z_0;z)dz \\
&= [\text{by } (5.12)] \int_{\Omega_\varrho(z_0)} \left[\frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} v_l(z)dl\right] E_\varrho(z_0;z)dz \\
&= \frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} \left[\int_{\Omega_\varrho(z_0)} v_l(z) E_\varrho(z_0;z)dz\right] dl \\
&= \frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} (v_l)_\varrho(z_0)dl \\
&\leqq [\text{by } (5.11)]\ \frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} v_l(z_0)dl \\
&= [\text{by } (5.12)]\ \frac{n}{2} r^{-n/2} \int_0^r l^{n/2-1} w_l^\sigma(z_0)dl = w_r(z_0).
\end{aligned}$$

We are thus left with the proof of the claim. We use (1.31) in Theorem 1.5 of [GL]. Applied to w the latter says that for every $r \leqq r_0$

$$w_r^\sigma(z_0) = w(z_0) + \int_{\Omega_r(z_0)} Lw(z)\,[\Gamma(z_0;z) - (4\pi r)^{-n/2}]dz, \tag{5.13}$$

provided that $\zeta \notin \overline{\Omega_r(z_0)}$. But $\zeta \notin \overline{\Omega_r(z_0)}$ implies $Lw(z) = L(\Gamma(\cdot;\zeta))(z) = 0$ for every $z \in \Omega_r(z_0)$. On the other hand $\zeta \notin \overline{\Omega_r(z_0)}$ (iff $\Gamma(z_0;\zeta) < (4\pi r)^{-n/2}$ and $\zeta \neq z_0$. (5.13) then gives for $\Gamma(z_0;\zeta) < (4\pi r)^{-n/2}$ and $\zeta \neq z_0$

$$w_r^\sigma(z_0) = w(z_0) = \Gamma(z_0;\zeta) = v_r(z_0).$$

But $z_0 = \zeta$ trivially implies $w_r^\sigma(z_0) = 0 = v_r(z_0)$, since in this case $w \equiv 0$ on $\Psi_r(z_0)$. To complete the proof of the claim we need only to consider the case in which z_0 is such that $\Gamma(z_0;\zeta) > (4\pi r)^{-n/2}$. Recalling that $Lw(z) = -\delta_\zeta(z)$, whereas $z \mapsto \Gamma(z;\zeta)$ is C^∞ in a neighborhood of z_0 that does not include ζ, again from (1.31) of Theorem 1.5 in [GL] we obtain

$$\begin{aligned} w_r^\sigma(z_0) &= w(z_0) + \int_{\mathbf{R}^{n+1}} Lw(z)\chi_{\Omega_r(z_0)}(z)\,[\Gamma(z_0;z) - (4\pi r)^{-n/2}]dz \\ &= w(z_0) - \langle \delta_\zeta(z), \chi_{\Omega_r(z_0)}(z)\,[\Gamma(z_0;z) - (4\pi r)^{-n/2}]\rangle \\ &= w(z_0) - \chi_{\Omega_r(z_0)}(\zeta)\Gamma(z_0;\zeta) + (4\pi r)^{-n/2} = (4\pi r)^{-n/2}. \end{aligned}$$

In the last equality we have used the fact that, since $\Gamma(z_0;\zeta) > (4\pi r)^{-n/2}$ iff $\zeta \in \Omega_r(z_0)$, $\chi_{\Omega_r(z_0)}(\zeta) = 1$. The proof of the claim is thus completed and, with it, the proof of Lemma 5.2.

We then return to the

Proof of Proposition 5.2 (continued). We are left with proving (iii) in the case in which $u = \Gamma_\mu$, where μ is a nonnegative compactly supported measure on $\mathbf{R}^{n+1}$. We observe first that if $\chi_{\Omega_r(z_0)}$ is the characteristic function of the parabolic ball $\Omega_r(z_0)$, we can write

$$\begin{aligned} u_r(z_0) &= \int_{\Omega_r(z_0)} u(z)E_r(z_0;z)dz = \int_{\mathbf{R}^{n+1}} \chi_{\Omega_r(z_0)}(z)u(z)E_r(z_0;z)dz \\ &= \int_{\mathbf{R}^{n+1}} \chi_{\Omega_r(z_0)}(z)\left[\int_{\mathbf{R}^{n+1}} \Gamma(z;\zeta)d\mu(\zeta)\right]E_r(z_0;z)dz \\ &= \int_{\mathbf{R}^{n+1}}\left[\int_{\mathbf{R}^{n+1}} \chi_{\Omega_r(z_0)}(z)\Gamma(z;\zeta)E_r(z_0;z)dz\right]d\mu(\zeta) \\ &= \int_{\mathbf{R}^{n+1}} (\Gamma(\cdot;\zeta))_r(z_0)d\mu(\zeta). \end{aligned}$$

The exchange of order of integration above is justified by Tonelli's theorem. The above identity and a similar exchange of order of integration give

$$\begin{aligned} (u_r)_\varrho(z_0) &= \int_{\mathbf{R}^{n+1}} \chi_{\Omega_\varrho(z_0)}(z)u_r(z)E_\varrho(z_0;z)dz \\ &= \int_{\mathbf{R}^{n+1}} \chi_{\Omega_\varrho(z_0)}(z)\int_{\mathbf{R}^{n+1}} [(\Gamma(\cdot;\zeta))_r(z)d\mu(\zeta)]\,E_\varrho(z_0;z)dz \\ &= \int_{\mathbf{R}^{n+1}}\left[\int_{\Omega_\varrho(z_0)} (\Gamma(\cdot;\zeta))_r(z)\,E_\varrho(z_0;z)dz\right]d\mu(\zeta) \\ &= \int_{\mathbf{R}^{n+1}} ((\Gamma(\cdot;\zeta))_r)_\varrho(z_0)d\mu(\zeta) \\ &\leqq \int_{\mathbf{R}^{n+1}} (\Gamma(\cdot;\zeta))_r(z_0)d\mu(\zeta) = u_r(z_0). \end{aligned} \tag{5.17}$$

In the last inequality above we have used (5.10). (5.17) shows that (ii) of Proposition 5.1 holds for u_r (with $m=0$), and therefore u_r is L-superparabolic. This completes the proof of Proposition 5.2.

Corollary 5.1. *Let u be L-superparabolic on $\mathbf{R}^{n+1}$. Then the conclusions* (i)–(iii) *of Proposition 5.2 hold unchanged if we replace u_r by $u_r^{(m)}$ for any $m \in \mathbf{N} \cup \{0\}$.*

Proof. It follows by the same arguments of the proof of Proposition 5.2 once we observe that $u_r^{(m)}(z)$ coincides with $\hat{u}_r(\hat{z})$ and that $\hat{u}$ is $\hat{L}$-superparabolic in $\mathbf{R}^{n+m+1}$; see Sect. 3.

Before stating the next corollary of Proposition 5.2, we recall the Weyl type formulas (3.20), (3.21).

Corollary 5.2. *Let u be L-superparabolic on $\mathbf{R}^{n+1}$. Then for every $m \in \mathbf{N} \cup \{0\}$ we have:*

(i) $J_\varrho^{(m)} u \leqq J_r^{(m)} u$ *for every* $r < \varrho \leqq \dfrac{r_0}{2}$.

(ii) $J_r^{(m)} u \nearrow u$ *as* $r \to 0^+$.

(iii) $J_r^{(m)} u$ *is L-superparabolic on $\mathbf{R}^{n+1}$ for every* $r \leqq \dfrac{r_0}{2}$.

Proof. By the definition (3.20) of the operators $J_\varrho^{(m)}$ we obtain for every $z \in \mathbf{R}^{n+1}$

$$J_\varrho^{(m)}(u)(z) = \int_0^{+\infty} u_l^{(m)}(z) \varphi\left(\frac{l}{\varrho}\right) \frac{dl}{\varrho} = \int_0^{+\infty} u_{\varrho l/r}^{(m)}(z) \varphi\left(\frac{l}{r}\right) \frac{dl}{r}$$
$$\leqq \int_0^{+\infty} u_l^{(m)}(z) \varphi\left(\frac{l}{r}\right) \frac{dl}{r} = J_r^{(m)}(u)(z).$$

In the inequality above we have used the fact that $\dfrac{\varrho}{r} > 1$ and Corollary 5.1. This proves (i). Since for every $z \in \mathbf{R}^{n+1}$

$$u(z) - J_r^{(m)}(u)(z) = \int_0^{+\infty} [u(z) - u_l^{(m)}(z)] \varphi\left(\frac{l}{r}\right) \frac{dl}{r},$$

from Proposition 5.1 and Corollary 5.1 we have

$$0 \leqq u(z) - J_r^{(m)}(u)(z) \leqq \sup_{0 < l \leqq 2r} [u(z) - u_l^{(m)}(z)]$$
$$= u(z) - u_{2r}^{(m)}(z) \to 0 \quad \text{as} \quad r \to 0^+.$$

Hence (ii) holds. We finally look at (iii). It is immediate to check that $J_r^{(m)}(u)$ is l.s.c. To prove that it is L-superparabolic we show that $J_r^{(m)}(u)$ is super mean valued. In fact, for every $z \in \mathbf{R}^{n+1}$ and $r \leqq \dfrac{r_0}{2}$ we have

$$(J_r^{(m)}(u))_\varrho(z) = \int_{\Omega_\varrho(z)} J_r^{(m)}(u)(\zeta) E_\varrho(z;\zeta) d\zeta$$
$$= \int_{\Omega_\varrho(z)} E_\varrho(z;\zeta) \left[\int_0^{+\infty} u_l^{(m)}(\zeta) \varphi\left(\frac{l}{r}\right) \frac{dl}{r} \right] d\zeta$$
$$= \int_0^{+\infty} \left[\int_{\Omega_\varrho(z)} u_l^{(m)}(\zeta) E_\varrho(z;\zeta) d\zeta \right] \varphi\left(\frac{l}{r}\right) \frac{dl}{r}$$

[by (iii) of Corollary 5.1 and Proposition 5.1]

$$\leqq \int_0^{+\infty} u_l^{(m)}(z) \varphi\left(\frac{l}{r}\right) \frac{dl}{r} = J_r^{(m)}(u)(z).$$

This completes the proof of the corollary.

6. Smoothing of L-Superparabolic Functions

In several questions in potential theory a crucial problem is that of regularizing superharmonic functions. For harmonic functions this can be achieved by means of the usual mollification process. The same procedure can be followed for supertemperatures as the heat operator has constant coefficients. However, for an operator like L in (1.1) usual mollification does not allow to approximate L-superparabolic functions by smooth functions which are still L-superparabolic. In this section we show that this approximation problem can be solved by means of the Weyl type averaging operators introduced in (3.20). We emphasize that the operators $u \mapsto u_r^{(m)}$ defined by (3.19) could also be used, although it would be much more complicated to prove their regularizing properties.

Theorem 6.1. *Let u be a L-superparabolic function in $\mathbf{R}^{n+1}$ and let $v \in \mathbf{N}$ be fixed. Then there exists a sequence of functions $(u_j)_{j\in\mathbf{N}}$ such that*

(i) $u_j \in C^v(\mathbf{R}^{n+1})$, $j \in \mathbf{N}$,

(i)) u_j *is L-superparabolic in* $\mathbf{R}^{n+1}$, $j \in \mathbf{N}$,

(iii) $u_j \leqq u_{j+1}$, $j \in \mathbf{N}$,

(iv) $u_j(z) \to u(z)$ *as* $j \to +\infty$ *for every* $z \in \mathbf{R}^{n+1}$

(v) *if for a given compact $K \subset \mathbf{R}^{n+1}$ $Lu = 0$ in $\mathbf{R}^{n+1} \backslash K$, then for every $\delta > 0$ there exists $j_0 \in \mathbf{N}$ such that $Lu_j = 0$ in $\mathbf{R}^{n+1} \backslash K_\delta$ for every $j \geqq j_0$, where*

$$K_\delta = \{z \in \mathbf{R}^{n+1} | \operatorname{dist}(z, K) \geqq \delta\}.$$

Proof. Let $(r_j)_{j\in\mathbf{N}}$ *be a sequence of positive numbers such that $r_j \to 0$ as $j \to \infty$ and $r_{j+1} \leqq r_j \leqq \frac{r_0}{2}$* for every $j \in \mathbf{N}$ (r_0 is as in Corollary 5.2). For every $j \in \mathbf{N}$ we set

$$u_j = J_{r_j}^{(m)}(u), \tag{6.1}$$

where m is a positive integer to be fixed later on and $J_r^{(m)}$ is defined as in (3.21). By Corollary 5.2 it immediately follows that the sequence $(u_j)_{j\in\mathbf{N}}$ verifies (ii), (iii), and (iv) above. We now prove (v). By Theorem 2.1 we can find two positive numbers $C, \bar{r}$, depending only on L, m, and n, such that

$$\Omega_r^{(m)}(z) \subset \{(\xi, \tau) \in \mathbf{R}^{n+1} \,|\, |x-\xi|^2 \leqq Cr,\ 0 < t-\tau < Cr\} \tag{6.2}$$

for every $z = (x, t) \in \mathbf{R}^{n+1}$ and $0 < r < \bar{r}$. It is therefore clear that given $\delta > 0$ there exists $j_0 \in \mathbf{N}$ such that

$$\Omega_l^{(m)}(z) \subset \mathbf{R}^{n+1} \backslash K$$

for every $l \leqq 2r_j$, $j \geqq j_0$, and $z \in \mathbf{R}^{n+1}$ such that $\operatorname{dist}(z, K) > \delta$. For such z's and l's we then have (see Theorem 3.1)

$$u_l^{(m)}(z) = u(z).$$

By (3.20), recalling that $\operatorname{supp}\varphi \subset (1, 2)$, we obtain

$$u_j(z) = u(z)$$

for every $j \geqq j_0$ and $z \in \mathbf{R}^{n+1}$ with $\operatorname{dist}(z, K) > \delta$. From this (v) immediately follows.

We are left with proving (i). We will show that:

$$\textit{There exists } m \in \mathbf{N} \textit{ such that } J_r^{(m)}(u) \in C^v(\mathbf{R}^{n+1}) \textit{ for every } r,\ 0 < r < \frac{r_0}{2}. \tag{6.3}$$

For ease of notation we set

$$\varphi_r(l) = \omega_m (4\pi l)^{-\frac{n+m}{2}} \frac{1}{r} \varphi\left(\frac{l}{r}\right),$$

where φ is the function appearing in (3.20). Then by (3.18) and (3.22), the kernel $M_r^{(m)}(z;\zeta)$ in (3.21) takes the form

$$M_r^{(m)} = \int_{\left[4\pi\Phi^{\frac{2}{n+m}}\right]^{-1}}^{+\infty} R_l^m \left[E + \frac{m}{m+2} \frac{R_l^2}{4(t-\tau)^2}\right] \varphi_r(l)\,dl. \tag{6.4}$$

In (6.4) and in what follows we simply write $M_r^{(m)}$, R_l, and E, instead of $M_r^{(m)}(z;\zeta)$, $R_l(z;\zeta)$, $E(z;\zeta)$. Moreover, if $z=(x,t)$ and $\zeta=(\xi,\tau)$, the reader should keep in mind that $M_r^{(m)}$ is given by (6.4) if $t>\tau$, whereas $M_r^{(m)} \equiv 0$ if $t \leqq \tau$. Let us now take $m=2h$, for $h \in \mathbf{N}$. From (6.4) and (3.4) we obtain for $t>\tau$

$$\begin{aligned}
M_r^{(2h)} = & \int_{\left[4\pi\Phi^{\frac{2}{n+2h}}\right]^{-1}} (4(t-\tau))^h [\ln((4\pi l)^{n/2+h}\Phi)]^h \\
& \times \left[E + \frac{h}{h+1} \frac{1}{t-\tau} \ln((4\pi l)^{n/2+h}\Phi)\right] \varphi_r(l)\,dl \\
= & \sum_{k=0}^{h} c_{h,k} (t-\tau)^h (\ln\Phi)^{h-k} E \int_{\left[4\pi\Phi^{\frac{2}{n+2h}}\right]^{-1}}^{+\infty} (\ln(4\pi l))^k \varphi_r(l)\,dl \\
& + \sum_{k=0}^{h+1} c'_{h,k} (t-\tau)^{h-1} (\ln\Phi)^{h+1-k} \int_{\left[4\pi\Phi^{\frac{2}{n+2h}}\right]^{-1}}^{+\infty} (\ln(4\pi l))^k \varphi_r(l)\,dl,
\end{aligned} \tag{6.5}$$

where $c_{h,k}$ and $c'_{h,k}$ are suitable constants. For $k=0,1,\ldots,h+1$, we set

$$\psi_k(s) = \begin{cases} (\ln s)^{h-k} \displaystyle\int_{\left[4\pi s^{\frac{2}{n+2h}}\right]^{-1}}^{+\infty} (\ln(4\pi l))^k \varphi_r(l)\,dl, & s>0, \\ 0, & s \leqq 0. \end{cases} \tag{6.6}$$

Since $\operatorname{supp}\varphi_r \subset (r,2r)$, $\psi_k \in C^\infty(\mathbf{R})$ and $\operatorname{supp}\psi_k \subset ((8\pi r)^{-n/2-h}, +\infty)$. If we replace (6.6) in (6.5), letting $c_{h,h+1}=0$, we obtain

$$M_r^{(2h)} = (t-\tau)^{h-1} \sum_{k=0}^{h+1} \psi_k(\Phi)\,[c_{h,k}(t-\tau)E + c'_{h,k}(\ln\Phi)], \tag{6.7}$$

which we think valid over all of $\mathbf{R}^{n+1} \times \mathbf{R}^{n+1}$ if we agree to set $\psi_k(s)\ln s = 0$ for $s \leqq 0$. Using (6.7) we are going to prove the following:

Claim. *For any fixed* $r \in \left(0, \frac{r_0}{2}\right)$, *for any* $v \in \mathbf{N}$ *and any multi-index* $\alpha = (\alpha_1, \ldots, \alpha_{n+1})$, $\alpha_i \in \mathbf{N} \cup \{0\}$ *and* $\alpha_1 + \ldots + \alpha_{n+1} \leqq v$, *there exist* $h = h(v) \in \mathbf{N}$ *and a constant* $C = C(r, \alpha, h) > 0$ *such that*

$$|D_z^\alpha M_r^{(2h)}(z;\zeta)| \leqq C \tag{6.8}$$

for every $z, \zeta \in \mathbf{R}^{n+1}$, *with* $z \neq \zeta$. *In* (6.8)

$$D_z^\alpha = D_{x_1}^{\alpha_1} D_{x_2}^{\alpha_2} \ldots D_{x_n}^{\alpha_n} D_t^{\alpha_{n+1}}.$$

It is clear that from the claim, and (3.21), (6.3) follows, and therefore (i). To see this we differentiate under the integral sign in (3.21), use (6.8), observe that $u \in L^1_{\mathrm{loc}}(\mathbf{R}^{n+1})$ since u is L-superparabolic, and finally recall that for every $z \in \mathbf{R}^{n+1}$ fixed, the support of $M_r^{(m)}(z;\cdot)$ is contained in the parabolic ball $\Omega_{r_0}^{(m)}(z)$ and the latter by (6.2) is contained in the cylinder

$$\{(\xi,\tau)\in \mathbf{R}^{n+1} \mid |x-\xi|^2 \leqq Cr_0, 0<t-\tau \leqq Cr_0\}.$$

We are therefore left with proving the claim. For every $\zeta \in \mathbf{R}^{n+1}$ fixed the support of the function $M_r^{(2h)}(\cdot,\zeta)$ is contained in

$$\Lambda_{r_0}(\zeta)=\{z\in \mathbf{R}^{n+1} \mid \Phi(z;\zeta)>(4\pi r_0)^{-(n/2+h)}\}$$

[we remark that $\Lambda_{r_0}(\zeta)$ is a modified parabolic ball relatively to the operator L^*, adjoint of L]. Hence, it will be enough to prove (6.8) for every $\zeta \in \mathbf{R}^{n+1}$ and $z \in \Lambda_{r_0}(\zeta)$. From Theorem 2.1 [see also (4.11) and (4.12)] we deduce the following estimate. For every multi-index $\alpha=(\alpha_1,\ldots,\alpha_{n+1})$ and for every $s\in \mathbf{N}$ there exist $q=q(\alpha,s)\in \mathbf{N}$ and a constant $C=C(r_0,\alpha,s)>0$ such that

$$|(t-\tau)^q D_z^\alpha \Phi(z;\zeta)| \leqq C(t-\tau)^{-h}[G(z;\zeta)+(t-\tau)^s], \tag{6.9}$$

for every $\zeta \in \mathbf{R}^{n+1}$ and $z\in \Lambda_{r_0}(\zeta)$. In (6.9) $G(z;\zeta)$ is the generalized Gaussian introduced in Sect. 2. If $s \geqq 1$, then on $\Lambda_{r_0}(\zeta)$ $G(z;\zeta)\geqq C(t-\tau)^s$, by suitably modifying the constant C in (6.9). On the other hand, as for (4.2) we have $G(z;\zeta)\leqq C\Gamma(z;\zeta)$ for every $\zeta \in \mathbf{R}^{n+1}$ and $z\in \Lambda_{r_0}(\zeta)$. From these considerations and (6.9) we obtain

$$|(t-\tau)^q D_z^\alpha \Phi(z;\zeta)| \leqq C\Phi(z;\zeta) \tag{6.10}$$

for every $\zeta \in \mathbf{R}^{n+1}$ and $z\in \Lambda_{r_0}(\zeta)$. Observing now that the function $(t-\tau)^{n/2+h}\Phi(z;\zeta)$ is bounded on $\mathbf{R}^{n+1}\times \mathbf{R}^{n+1}$ we obtain from (6.10): for every multi-index $\alpha=(\alpha_1,\ldots,\alpha_{n+1})$ there exist $q=q(\alpha)\in \mathbf{N}$ and a constant $C=C(r_0,\alpha,h)>0$ such that

$$|(t-\tau)^q D_z^\alpha \Phi(z;\zeta)| \leqq C \tag{6.11}$$

for every $\zeta \in \mathbf{R}^{n+1}$, $z\in \Lambda_{r_0}(\zeta)$. We then prove that an estimate similar to (6.11) holds for the functions $\psi_k(\Phi)$ defined through (6.6). To this purpose we observe that $D^\alpha \psi_k(\Phi)$ is a finite sum of terms of the type

$$\mathrm{const}\, \psi_k^{(p)}(\Phi) D^{\beta_{(1)}}\Phi \ldots D^{\beta_{(p)}}\Phi\,,$$

where $p\in \mathbf{N}$ and $p\leqq \alpha_1+\ldots+\alpha_{n+1}$, $\beta_{(1)},\ldots,\beta_{(p)}$ are multi-index whose length does not exceed the length of α, and $\psi_k^{(p)}$ is the p^{th} order derivative of ψ_k. Now from (6.6) one easily obtains the following estimate

$$|\psi_k^{(p)}(s)| \leqq \mathrm{const}(\ln s)^h, \quad \text{for every } s\in ((8\pi r)^{-n/2-h}, +\infty),$$

whereas by (4.2)

$$|\ln \Phi(z;\zeta)| \leqq \ln[C_2(t-\tau)^{-n/2-h}], \quad \text{for} \quad \zeta\in \mathbf{R}^{n+1} \text{ and } z\in \Lambda_{r_0}(\zeta).$$

Using (6.11) we finally have: for every multi-index α there exist $q=q(\alpha)\in \mathbf{N}$ and $C=C(\alpha,r,h)>0$ such that

$$|(t-\tau)^q D_z^\alpha \psi_k(\Phi(z;\zeta))| \leqq C, \qquad \zeta\in \mathbf{R}^{n+1} \text{ and } z\in \Lambda_{r_0}(\zeta). \tag{6.12}$$

In the same way one proves an analogous estimate for the functions $\ln \Phi$ and E. Concerning the derivatives of E, it is enough to observe that by (3.2) we can write

$$E(z;\zeta)=A(\zeta)\,[V_\xi(\ln\Phi(z;\zeta))]\cdot V_\xi(\ln\Phi(z;\zeta)).$$

From (6.7) and (6.12), and analogous estimates for $\ln\Phi$ and E we finally obtain: for every $v\in\mathbf{N}$ and for every multi-index $\alpha=(\alpha_1,\ldots,\alpha_{n+1})$, with $\alpha_1+\ldots+\alpha_{n+1}\leqq v$, there exist $h=h(v)\in\mathbf{N}$ and a constant $C=C(\alpha,r,h)>0$ such that

$$|D_z^\alpha M_r^{(2h)}(z;\zeta)|\leqq C, \qquad \zeta\in\mathbf{R}^{n+1},\ z\in\Lambda_{r_0}(\zeta).$$

This proves the claim, thus completing the proof of Theorem 6.1.

References

[Ba] Bauer, H.: Harmonische Räume und ihre Potentialtheorie (Lecture Notes in Mathematics, Vol. 22). Berlin Heidelberg New York: Springer 1966

[BGM] Berger, M., Gauduchon, P., Mazet, E.: Le Spectre d'une Variété Riemanniene (Lecture Notes in Mathematics, Vol. 194). Berlin Heidelberg New York: Springer 1971

[EG] Evans, L.C., Gariepy, R.F.: Wiener's criterion for the heat equation. Arch. Rat. Mech. Anal. **78**, 293–314 (1982)

[FG] Fabes, E.B., Garofalo, N.: Mean value properties of solutions to parabolic equations with variable coefficients. J. Math. Anal. Appl. **121**, 305–316 (1987)

[Fe] Federer, H.: Geometric measure theory (Die Grundlehren der mathematischen Wissenschaften, Vol. 153). Berlin Heidelberg New York: Springer 1969

[F] Folland, G.B.: Introduction to partial differential equations. Princeton Univ. Press 1976

[Fr] Friedman, A.: Partial differential equations of parabolic type. New York: Prentice-Hall 1964

[Fu] Fulks, W.: A mean value theorem for the heat equation. Proc. Am. Math. Soc. **17**, 6–11 (1966)

[GL] Garofalo, N., Lanconelli, E.: Wiener's criterion for parabolic equations with variable coefficients and its consequences, Trans. Am. Math. Soc. **307** (to appear) (1988)

[H] Helms, L.L.: Introduction to potential theory. New York: Wiley-Interscience 1969

[K] Kannai, Y.: Off diagonal short time asymptotics for fundamental solutions of diffusion equations. Commun. Partial Differ. Equations **2**, 781–830 (1977)

[Ku] Kupcov, L.P.: The mean property and the maximum principle for the parabolic equations of second order. Dokl. Akad. Nauk SSSR **242**, no. 3 (1978); English transl.: Soviet Math. Dokl. **19**, 1140–1144 (1978)

[L 1] Littman, W.: A strong maximum principle for weakly L-subharmonic functions. J. Math. Mech. **8**, 761–770 (1959)

[L 2] Littman, W.: Generalized subharmonic functions: Monotonic approximations and an improved maximum principle. Ann. Sc. Norm. Super. Pisa, Cl. Sci., IV. Ser. **17**, 207–222 (1963)

[M] Moser, J.: A Harnack inequality for parabolic differential equations. Commun. Pure Appl. Math. **17**, 101–134 (1964)

[P 1] Pini, B.: Sulle equazioni a derivate parziali lineari del secondo ordine in due variabili di tipo parabolico. Ann. Mat. Pura e Appl. **32**, 179–204 (1951)

[P 2] Pini, B.: Maggioranti e minoranti delle soluzioni delle equazioni paraboliche, Ann. Mat. Pura Appl., IV. Ser. **37**, 249–264 (1954)

[P 3] Pini, B.: Sulla soluzione generalizzata di Wiener per il primo problema di valori al contorno nel caso parabolico. Rend. Semin. Mat. Univ. Padova **23**, 422–434 (1954)

[W] Watson, N.A.: A theory of temperatures in several variables. Proc. Lond. Math. Soc. **26**, (3) 385–417 (1973)

Received December 28, 1987; in revised form April 25, 1988

Math. Ann. 283, 241–255 (1989)

Mathematische Annalen

An Index Theory and Existence of Multiple Brake Orbits for Star-Shaped Hamiltonian Systems

Andrzej Szulkin*

Department of Mathematics, University of Stockholm, S-11385 Stockholm, Sweden

1. Introduction

Let $H:\mathbf{R}^{2N}\to\mathbf{R}$ be a continuously differentiable function and denote

$$J=\begin{pmatrix}0 & -I\\ I & 0\end{pmatrix},$$

where I is the $N\times N$ identity matrix. Consider the Hamiltonian system of differential equations

$$\dot{x}=JH'(x). \tag{HS}$$

If $x(t)$ satisfies (HS), then $\frac{d}{dt}H(x(t))=0$, so each solution of (HS) must necessarily lie on some energy surface $\{x\in\mathbf{R}^{2N}:H(x(t))=\text{const}\}$.

In this paper we will be concerned with the existence of periodic solutions of (HS) (of a priori unknown period) on a given compact hypersurface $H^{-1}(1)$. Such solutions are called periodic orbits. The problem of existence of at least one periodic orbit on $H^{-1}(1)$ has been studied by several authors, see e.g. Seifert [16], Weinstein [22], Rabinowitz [12], Viterbo [20], and Hofer and Zehnder [8].

Let $x=(p,q)\in\mathbf{R}^N\times\mathbf{R}^N$. A special kind of periodic orbits, henceforth called *brake orbits* (cf. [22]), are those for which the q-component of the solution of (HS) oscillates back and forth between two restpoints. More precisely, the corresponding $p(t)$ and $q(t)$ are T-periodic functions for some $T>0$, p odd and q even about 0 and $\frac{T}{2}$. Recently Rabinowitz [14] has shown that if H is even in p [i.e., $H(-p,q)=H(p,q)$] and $H^{-1}(1)$ bounds a compact star-shaped neighbourhood of $0\in\mathbf{R}^{2N}$ such that $x\cdot H'(x)\neq 0\ \forall x\in H^{-1}(1)$, then (HS) possesses a brake orbit (see also [15] for a more general result).

In [4] Ekeland and Lasry have shown that (HS) has at least N periodic orbits if $H^{-1}(1)$ bounds a convex region and satisfies a certain geometric condition. This

* Supported in part by the Swedish Natural Science Research Council

result has been later generalized by Berestycki et al. [2] to the case of $H^{-1}(1)$ bounding a star-shaped region. Results of the same type, for Hamiltonians having certain symmetry properties, have been obtained by Girardi [5] and van Groesen [19]. The proofs in these papers were carried out by reducing the problem to the one of finding critical points of an appropriate S^1- or $\mathbf{Z}_2$-symmetric functional. Multiple critical points were then obtained by invoking minimax arguments and topological index theories for symmetric sets.

When looking for brake orbits of (HS) it is natural to work in the space of periodic functions $(p(t), q(t))$ such that p is odd and q even in t. Although the functional is no longer symmetric in this space, it has a useful partial $\mathbf{Z}_2$-symmetry as we will see later. Our goal is to show that if $H(-p,q)=H(p,q)$ and $H^{-1}(1)$ bounds a star-shaped region and satisfies a geometric condition similar to that in [2], then (HS) possesses at least N brake orbits on $H^{-1}(1)$. The main tool in the proof is an index theory which we construct in Sect. 2. It is a variant of the relative index introduced by Berestycki et al. [2] and further developed by Tarantello [18] (see also Benci [1] for a related concept of pseudoindex). A special feature of our index is a strong dimension property (see Proposition 2.8 and Remark 2.9). In Sect. 3 we establish an abstract minimax principle which we employ in Sect. 4 to the problem of existence of N brake orbits.

An important role in the proof of dimension property is played by the generalized Borsuk theorem (Lemma 2.10). Its different versions are known in the literature (see e.g. Michalek [10], Nirenberg [11], Wang [21] and the references there). Our version is a simple adaptation of that in [10, 11, 21] to the case of partial $\mathbf{Z}_2$-symmetry. For the reader's convenience we include an appendix containing a proof of Lemma 2.10.

2. An Index Theory

In this section we develop an index theory similar to that in [2, 18]. Although we restrict our attention to the symmetry group $\mathbf{Z}_2$, it is clear that at least for separable spaces one obtains a corresponding theory for other symmetry groups (in particular, for S^1 and $\mathbf{Z}_p$ with p a prime integer).

Let E be a real Hilbert space and T a unitary representation of $\mathbf{Z}_2$ in E. That is, $T_0=I_E$ (the identity mapping on E) and T_1 is a linear isometry such that $T_1=T_1^{-1}$. A subset $A\subset E$ is said to be *T-invariant* (or simply *invariant*) if $T_1 A\subset A$. Let

$$E^G=\{x\in E: T_1x=x\}$$

be the fixed point set of T. To T there corresponds an orthogonal decomposition $E=E^G\oplus F$. It is easy to see that F is invariant and

$$T_1(x+y)=x-y \ \forall x\in E^G,\, y\in F\,.$$

Let

$$\Sigma=\{A\subset E: A \textit{ is closed and invariant}\}\,.$$

For $A\in\Sigma$ we define *the index of* A, denoted $\gamma(A)$, to be the smallest integer k such that there exists a continuous mapping $f: A\to\mathbf{R}^k-\{0\}$ satisfying $f(T_1x)=-f(x)$.

If there is no such k, then $\gamma(A)=\infty$. For the empty set $\emptyset$ we define $\gamma(\emptyset)=0$. Observe that if $A\cap E^G\neq\emptyset$, then $\gamma(A)=\infty$ (because $f(x)=f(T_1x)=-f(x)\ \forall x\in A\cap E^G$). It is easy to verify that γ satisfies the usual properties of index (which may be found e.g. in [13, 17]). In particular, denoting $N_\delta(A)=\{x\in E: d(x,A)\leqq\delta\}$, where $d(x,A)$ is the distance from x to A, we have

2.1. Proposition (Continuity Property). *If $A\in\Sigma$ is compact, then $\gamma(A)=\gamma(N_\delta(A))$ for some $\delta>0$.*

Let Y be a closed subspace of E. Henceforth P_Y will denote the orthogonal projection from E to Y and Y^G will be the set $Y\cap E^G$. A function $\xi: A\to\mathbf{R}$ is said to be *T-invariant* (or *invariant*) if $\xi(T_1x)=\xi(x)\ \forall x\in E$. A mapping $f: A\to E$ is *T-equivariant* (or *equivariant*) if $f(T_1(x))=T_1f(x)\ \forall x\in E$, and f is *compact* if the image of each bounded subset of A is contained in a compact set.

Let now $E=Y\oplus X$, where X, Y are orthogonal to each other and invariant. For $A\in\Sigma$, let $\mathscr{F}_k(A)$ be the set of all continuous mappings $f=(f_1,f_2): A\to Y\times\mathbf{R}^k-\{(0,0)\}$ satisfying the following conditions:

(i) *f is equivariant in the sense that $f_1(T_1x)=T_1f_1(x)$, $f_2(T_1x)=-f_2(x)$,*

(ii) *$f_1=P_Y-K$, where K is compact and $K(A)$ is bounded in Y,*

(iii) *$f_1(x)=x\ \forall x\in A\cap Y^G$ (and $f_2(x)=0$ by equivariance).*

Let $A\in\Sigma$. We define *the index of A relative to X*, denoted $\gamma_r(A,X)$, or shortly $\gamma_r(A)$ when no ambiguity can arise, to be the smallest integer k such that $\mathscr{F}_k(A)\neq\emptyset$. If there is no such k, we set $\gamma_r(A)=\infty$, and we define $\gamma_r(\emptyset)=0$.

It should be noted that the main difference between our index and that in [2, 18] lies in the requirement that K in (ii) above be of bounded range. In the following propositions we collect some basic properties of γ_r.

2.2. Proposition (Mapping Property). *Let $A,B\in\Sigma$ and let $g: A\to B$ be a continuous mapping such that $g(x)=e^{-\xi(x)L}x-K(x)$, where*

(i) *$L: E\to E$ is linear, equivariant, selfadjoint and $LY\subset Y$,*

(ii) *$\xi: A\to\mathbf{R}$ is invariant and $\xi(A)$ is bounded,*

(iii) *$K: A\to E$ is equivariant, compact and $K(A)$ is bounded.*

If $\xi|_{A\cap Y^G}=0$ and $K|_{A\cap Y^G}=0$ (i.e., if $g|_{A\cap Y^G}=I_{A\cap Y^G}$), then $\gamma_r(A)\leqq\gamma_r(B)$.

Proof. The conclusion is trivial if $\gamma_r(B)=\infty$. Assume that $\gamma_r(B)=k<\infty$. Then there exists an $f\in\mathscr{F}_k(B)$, $f=(f_1,f_2)$, $f_1=P_Y-C$. Note that $P_YL=LP_Y$ by selfadjointness of L. Define $\varphi: A\to Y\times\mathbf{R}^k$ by setting

$$\varphi(x)=(\varphi_1(x),\varphi_2(x))=(e^{\xi(x)L}f_1g(x), f_2g(x)).$$

Then φ is equivariant, $\varphi(x)\neq(0,0)\ \forall x\in A$ and

$$\begin{aligned}\varphi_1(x)&=e^{\xi(x)L}(P_Y-C)g(x)=e^{\xi(x)L}(P_Y-C)(e^{-\xi(x)L}x-K(x))\\&=P_Yx-e^{\xi(x)L}(P_YK(x)+C(g(x)))\equiv P_Yx-N(x).\end{aligned}$$

It is easy to see that N is compact and $N(A)$ bounded. If $x\in A\cap Y^G$, then $g(x)=x\in B\cap Y^G$, so $\varphi_1(x)=x$. Hence $\varphi\in\mathscr{F}_k(A)$ and $\gamma_r(A)\leqq k=\gamma_r(B)$. □

2.3. Proposition (Monotonicity). *If $A,B\in\Sigma$ and $A\subset B$, then $\gamma_r(A)\leqq\gamma_r(B)$.*

Proof. Take $g(x)=x$ in the preceding proposition. □

2.4. Proposition (Subadditivity). *If $A, B \in \Sigma$, then $\gamma_r(A \cup B) \leqq \gamma_r(A) + \gamma(B)$.*

Proof. It suffices to consider A, B with $\gamma_r(A) = k < \infty$, $\gamma(B) = m < \infty$. Suppose $f = (P_Y - K, f_2) \in \mathscr{F}_k(A)$ and $g: B \to \mathbf{R}^m - \{0\}$ satisfies $g(T_1 x) = -g(x)$. Since the convex hull of a compact set is compact, there exists a compact extension $\tilde{K}: E \to Y$ of K such that $\tilde{K}(E)$ is bounded. We may assume that $\tilde{K}$ is equivariant (if it is not, replace it by $\frac{1}{2}\tilde{K} + \frac{1}{2}T_1\tilde{K}T_1$). Similarly, there exist extensions $\tilde{f}_2: E \to \mathbf{R}^k$, $\tilde{g}: E \to \mathbf{R}^m$ of f_2 and g such that $\tilde{f}_2(T_1 x) = -\tilde{f}_2(x)$, $\tilde{g}(T_1 x) = -\tilde{g}(x)$. Let now

$$h(x) = (P_Y x - \tilde{K}(x), \tilde{f}_2(x), \tilde{g}(x)) \in Y \times \mathbf{R}^k \times \mathbf{R}^m.$$

Then $h: A \cup B \to Y \times \mathbf{R}^{k+m}$, h is equivariant, $h = (P_Y - \tilde{K}, h_2)$ and $h(x) \neq (0, 0)$ $\forall x \in A \cup B$ because $(P_Y x - \tilde{K}(x), f_2(x)) = (P_Y x - K(x), f_2(x)) \neq (0, 0)$ on A and $\tilde{g}(x) = g(x) \neq 0$ on B. Since $\gamma(B) < \infty$, $B \cap Y^G = \emptyset$. Therefore $(A \cup B) \cap Y^G = A \cap Y^G$, and for $x \in A \cap Y^G$, $P_Y x - \tilde{K}(x) = x$. Hence $h \in \mathscr{F}_{k+m}(A \cup B)$ and $\gamma_r(A \cup B) \leqq k + m$. $\square$

2.5. Proposition. *If $A, B \in \Sigma$ and $\gamma(B) < \infty$, then $\gamma_r(\overline{A - B}) \geqq \gamma_r(A) - \gamma(B)$.*

Proof. It follows from Propositions 2.3 and 2.4 that $\gamma_r(A) \leqq \gamma_r(\overline{A - B} \cup B) \leqq \gamma_r(\overline{A - B}) + \gamma(B)$. $\square$

2.6. Proposition (Intersection Property). *Let $A \in \Sigma$. Suppose that $X = X_0 \oplus X_1$, where X_0, X_1 are orthogonal to each other, invariant and $X_0 \cap E^G = \{0\}$. If* $\dim X_0 = k < \infty$ *and $\gamma_r(A) > k$, then $A \cap X_1 \neq \emptyset$.*

Proof. Suppose that $A \cap X_1 = \emptyset$ and let $f(x) = P_Y x + P_{X_0} x$. Then $f: A \to Y \oplus X_0 - \{0\}$. Since X_0 is invariant and $X_0 \cap E^G = \{0\}$, $T_1 x_0 = -x_0$ $\forall x_0 \in X_0$ (because $x_0 + T_1 x_0 \in X_0 \cap E^G$). So $P_{X_0} T_1 x = T_1 P_{X_0} x = -P_{X_0} x$. It follows that X_0 can be identified with $\mathbf{R}^k$ and f with the equivariant mapping $x \mapsto (P_Y x, f_2(x))$ from A to $Y \times \mathbf{R}^k$ ($f_2(x) \in \mathbf{R}^k$ corresponds to $P_{X_0} x \in X_0$). Hence $\mathscr{F}_k(A) \neq \emptyset$ and $\gamma_r(A) \leqq k$, a contradiction. $\square$

In order to state the next property of γ_r we will need the following geometric condition.

2.7. Definition. A set A is said to satisfy *condition* ($\mathscr{G}$) if for each finite dimensional subspace E_0 of E and each $r > 0$ there exists an $R > 0$ such that if $\|x\| \leqq r$, then $A \cap (x + E_0) \subset B_R$.

Here $x + E_0 = \{x + y \in E : y \in E_0\}$ and $B_R = \{x \in E : \|x\| < R\}$.

2.8. Proposition (Dimension Property). *Let $X_0 \subset X$ be an invariant subspace with* $\dim X_0 = k$ *and $X_0 \cap E^G = \{0\}$. Let U be an open invariant neighbourhood of the origin in $Y \oplus X_0$. If $\bar{U}$ satisfies condition ($\mathscr{G}$), then $\gamma_r(\partial U) = k$, where ∂U is the boundary of U in $Y \oplus X_0$.*

2.9. Remarks. (i) For the index in [2, 18] the conclusion of Proposition 2.8 remains valid if U is bounded but fails in general (see Bögle [3]). In Sect. 4 we will employ the above proposition to sets which satisfy condition ($\mathscr{G}$) and are unbounded. It was the need of a dimension property for unbounded sets that has motivated us to look for an index theory different from already existing ones.

(ii) As we have just observed, our index satisfies a stronger dimension property than that in [2, 18]. On the other hand, our mapping property is weaker (because the mapping K in Proposition 2.2 has bounded range). This has the disadvantage that in order to establish the minimax principle of the next section it seems necessary to modify the standard deformation lemma. Such a modification will be carried out for a class of functionals satisfying a compactness hypothesis which is somewhat stronger than the usual Palais-Smale condition.

In the proof of Proposition 2.8 we will use the following generalized Borsuk theorem (cf. Appendix).

2.10. Lemma. *Let W be an open bounded neighbourhood of $0 \in \mathbf{R}^m \times \mathbf{R}^n$ such that if $(x, y) \in W$, then $(x, -y) \in W$. Let $f=(g, h): \bar{W} \to \mathbf{R}^m \times \mathbf{R}^n$ be a continuous mapping with $g(x, -y)=g(x, y)$, $h(x, -y)=-h(x, y)$ $\forall (x, y) \in \bar{W}$, $f|_{\partial W} \neq 0$ and $f(x, 0)=(x, 0)$ $\forall (x, 0) \in \partial W$. Then the Brouwer degree* $\deg(f, W, 0)$ *is an odd integer.*

Proof of Proposition 2.8. By Proposition 2.6, $\gamma_r(\partial U) \leqq k$ (because ∂U does not intersect the orthogonal complement of X_0 in X). Suppose $\gamma_r(\partial U)<k$. Let $f=(P_Y-K, f_2): \partial U \to Y \times \mathbf{R}^{k-1}$, where $f \in \mathscr{F}_{k-1}(\partial U)$. By replacing $f_2(x)$ with $f_2(x)/(1+|f_2(x)|)$ if necessary we may assume that $f_2(\partial U)$ is bounded. Since X_0 is invariant and $X_0 \cap E^G=\{0\}$, we may also assume (as in the proof of Proposition 2.6) that $f=P_Y-K+f_2: \partial U \to Y \oplus X'_0$, where X'_0 is a $k-1$ dimensional subspace of X_0. Let $\tilde{K}: \bar{U} \to Y$, $\tilde{f}_2: \bar{U} \to X'_0$ be equivariant extensions of K and f_2 such that $\tilde{K}$ is compact and $\tilde{K}(\bar{U})$, $\tilde{f}_2(\bar{U})$ are bounded (cf. the proof of Proposition 2.4). Since $K|_{\partial U \cap Y^G}=0$, we may assume that $\tilde{K}|_{\bar{U} \cap Y^G}=0$. Let r be a number such that $\tilde{K}(\bar{U}) \subset B_r$. By condition ($\mathscr{G}$), there is an $R>0$ such that

$$\bar{U} \cap (y+X_0) \subset B_R \text{ whenever } \|y\| \leqq r. \tag{1}$$

Let $x \in \bar{U}$. Suppose $\|P_Y x\| \leqq r$. Since $x=P_Y x+P_{X_0} x \subset P_Y x+X_0$, $x \in \bar{U} \cap (P_Y x+X_0)$ $\subset B_R$ according to (1). It follows that if $x \in \bar{U} \cap \partial B_R$, then $\|P_Y x\|>r$, and therefore $P_Y x-\tilde{K}(x) \neq 0$. Hence $\tilde{f}(x)=P_Y x-\tilde{K}(x)+\tilde{f}_2(x) \neq 0$ whenever $x \in \partial(U \cap B_R)$ (recall that $\tilde{f}=f \neq 0$ on ∂U). Furthermore, $\tilde{f}=I-C$, where $C=\tilde{K}+P_{X_0}-\tilde{f}_2$ is a compact mapping. It follows that the Leray-Schauder degree (see e.g. [9]) $\deg(\tilde{f}, U \cap B_R, 0)$ is well defined. Since $\tilde{f}(U \cap B_R) \subset Y \oplus X'_0$, $\tilde{f}(x)+te \neq 0$ for any $x \in \overline{U \cap B_R}$, $t>0$ and $e \in X_0-X'_0$. So by the homotopy invariance,

$$\deg(\tilde{f}, U \cap B_R, 0)=0. \tag{2}$$

It follows from the properties of $\tilde{K}$ and $\tilde{f}_2$ that the mapping C is equivariant and $C|_{\bar{U} \cap Y^G}=0$. Given $\varepsilon>0$, there exists a mapping C_ε such that $\|C(x)-C_\varepsilon(x)\|$ $<\varepsilon$ $\forall x \in \overline{U \cap B_R}$ and $C_\varepsilon(U \cap B_R)$ is contained in a finite dimensional space [9, Theorem 4.2.2]. By a slight modification of the proof in [9] we will obtain C_ε which has some additional properties. Since the set $\overline{C(U \cap B_R)}$ is compact, it may be covered by open balls

$$B_\varepsilon(v_i)=\{z \in Y \oplus X_0: \|z-v_i\|<\varepsilon\}, \quad 0 \leqq i \leqq 2N.$$

Moreover, we may assume that $v_0=0$ and $v_{i+N}=-v_i$ for $i=1, \ldots, N$. Let $m_i(x)$ $=\max\{0, \varepsilon-\|C(x)-v_i\|\}$ and

$$\theta_i(x)=m_i(x) \Big/ \sum_{j=0}^{2N} m_j(x), \quad 0 \leqq i \leqq 2N.$$

Define

$$D_\varepsilon(x) = \sum_{i=0}^{2N} \theta_i(x) v_i$$

and $C_\varepsilon(x) = \frac{1}{2} D_\varepsilon(x) + \frac{1}{2} T_1 D_\varepsilon(T_1 x)$. It is easy to see (cf. [9]) that C_ε has the properties mentioned above, it is equivariant and $C_\varepsilon|_{\overline{U \cap B_R} \cap Y^G} = 0$. (Indeed, if $x \in \overline{U \cap B_R} \cap Y^G$, then $C(x) = C(T_1 x) = 0$, so $m_i(x) = m_{i+N}(x)$ for $1 \leq i \leq N$. Therefore $D_\varepsilon(x) = D_\varepsilon(T_1 x) = 0$. It follows that $C_\varepsilon(x) = 0$.) Choose a finite dimensional invariant subspace Y_0 of Y such that $Y_0 \oplus X_0$ contains the range of C_ε. If ε is sufficiently small,

$$\deg(\tilde{f}, U \cap B_R, 0) = \deg(I - C_\varepsilon, U \cap B_R \cap (Y_0 \oplus X_0), 0). \tag{3}$$

By Lemma 2.10, the right-hand side of (3) is an odd integer, a contradiction to (2). Hence $\gamma_r(\partial U)$ cannot be less than k. □

3. A Minimax Principle

Let E be a real Hilbert space with an inner product $\langle\ ,\ \rangle$ and T a unitary representation of Z_2 in E. Let $L: E \to E$ be an equivariant and selfadjoint bounded linear operator. Define $\Phi(x) = \frac{1}{2}\langle Lx, x\rangle$. Then Φ is an invariant functional and $\Phi'(x) = L(x)$ (here we assume via the Riesz representation theorem that $\Phi'(x) \in E$). Let ψ be an invariant functional on E such that $\psi^{-1}(1) \neq \emptyset$. Set $M = \psi^{-1}(1)$. Suppose further that $\psi \in C^{1,1}(E, \mathbf{R})$ and ψ' is a compact mapping which is bounded away from 0 on bounded subsets of M. Then M is an invariant $C^{1,1}$-manifold. For $x \in M$ denote

$$\lambda(x) = \frac{\langle Lx, \psi'(x)\rangle}{\|\psi'(x)\|^2}$$

and observe that $\langle Lx - \lambda(x)\psi'(x), \psi'(x)\rangle = 0$. So $Lx - \lambda(x)\psi'(x)$ is an element of the tangent space $T_x(M)$. In particular, $x \in M$ is a critical point of $\Phi|_M$ if and only if $Lx = \lambda(x)\psi'(x)$. We will need the following compactness hypothesis which is stronger than the usual Palais-Smale condition:

(C^*) *If* $(x_n) \subset M$ *is a sequence such that* $\Phi(x_n) \to c \in \mathbf{R}$ *and* $\dfrac{Lx_n - \lambda(x_n)\psi'(x_n)}{(\|x_n\| + 1)^{1/2}} \to 0$, *then* (x_n) *has a convergent subsequence.*

Let

$$\Phi_c = \{x \in M : \Phi(x) \leq c\} \quad \textit{and} \quad K_c = \{x \in M : \Phi(x) = c, Lx = \lambda(x)\psi'(x)\}.$$

In the proof of the minimax principle we will employ the following deformation lemma.

3.1. Lemma. *Suppose that Φ, ψ and M are as above and $\Phi|_M$ satisfies (C^*). Given $c \in \mathbf{R}$, $\bar{\varepsilon} > 0$ and a neighbourhood U of K_c in M, there exist $\varepsilon \in (0, \bar{\varepsilon})$ and a mapping $\eta: [0,1] \times M \to M$ such that:*

(i) *$\eta(t, \cdot)$ is a homeomorphism* $\forall t \in [0,1]$,

(ii) $\eta(0, x) = x$ $\forall x \in M$,

(iii) $\eta(1,x)=x\ \forall x\in M, \Phi(x)\notin(c-\bar{\varepsilon},c+\bar{\varepsilon})$,

(iv) $\|\eta(1,x)-x\|\leqq 2\|L\|\ \forall x\in M$ *($\|L\|$ is the norm of the operator L),*

(v) $\eta(1,\Phi_{c+\varepsilon}-U)\subset\Phi_{c-\varepsilon}$,

(vi) $\eta(1,x)=e^{-\theta(x)L}x-K(x)$, *where* $\theta(x)\in[0,1]\ \forall x\in M$, *$K$ is compact and $K(M)$ is bounded,*

(vii) $\eta(1,\cdot)$ *is equivariant.*

Proof. Since results of similar type are well known (see e.g. [13, Appendix A]), we will only outline the argument. By (C^*), K_c is a compact set. It follows that if $N_\delta=\{x\in M: d(x,K_c)\leqq\delta\}$, where $d(x,K_c)$ is the distance from x to K_c, then $N_\delta\subset U$ for $\delta>0$ sufficiently small. Thus we may assume that $U=N_\delta$. We claim that there exist $\hat{\varepsilon}\in(0,\bar{\varepsilon})$ and $b>0$ such that

$$\|Lx-\lambda(x)\psi'(x)\|\geqq b(\|x\|+1)^{1/2}\geqq b\ \forall x\in\Phi_{c+\hat{\varepsilon}}-(\Phi_{c-\hat{\varepsilon}}\cup N_{\delta/8}). \tag{4}$$

For if not, we find $b_n\to 0$, $\hat{\varepsilon}_n\to 0$ and $x_n\in\Phi_{c+\hat{\varepsilon}_n}-(\Phi_{c-\hat{\varepsilon}_n}\cup N_{\delta/8})$ such that

$$\|Lx_n-\lambda(x_n)\psi'(x_n)\|\leqq b_n(\|x_n\|+1)^{1/2}.$$

By (C^*), $x_n\to\bar{x}\in K_c$ after passing to a subsequence. This is impossible because $x_n\notin N_{\delta/8}$. So (4) is satisfied for some $\hat{\varepsilon}$ and b. Choose $\varepsilon\in(0,\hat{\varepsilon})$ and let $\chi_1,\chi_2: M\to[0,1]$ be two Lipschitz continuous functions such that $\chi_1(x)=0$ if $\Phi(x)\notin(c-\hat{\varepsilon},c+\hat{\varepsilon})$, $\chi_1(x)=1$ if $\Phi(x)\in[c-\varepsilon,c+\varepsilon]$ and $\chi_2(x)=0$ if $x\in N_{\delta/8}$, $\chi_2(x)=1$ if $x\notin N_{\delta/4}$. Since the sets in the definitions of χ_1 and χ_2 are invariant, we may assume that χ_1, χ_2 are invariant functions.

Let $\chi(x)=\chi_1(x)\chi_2(x)$ and consider the initial value problem

$$\frac{d\eta}{dt}=-\frac{\chi(\eta)}{\|\eta\|+1}(L\eta-\lambda(\eta)\psi'(\eta)),\qquad \eta(0,x)=0, \tag{5}$$

where $x\in M$. Since $\left\|\frac{d\eta}{dt}\right\|\leqq 2\|L\|$ and the vector field in (5) is locally Lipschitz continuous, (5) has a unique solution $\eta(t,x)$ defined for all $t\in\mathbf{R}$. It is therefore clear that (i)–(iv) are satisfied. Since

$$\begin{aligned}\frac{d}{dt}\Phi(\eta(t,x))&=\left\langle L\eta,\frac{d\eta}{dt}\right\rangle=-\frac{\chi(\eta)}{\|\eta\|+1}\langle L\eta,L\eta-\lambda(\eta)\psi'(\eta)\rangle\\&=-\frac{\chi(\eta)}{\|\eta\|+1}\|L\eta-\lambda(\eta)\psi'(\eta)\|^2\leqq 0,\end{aligned}$$

$\Phi(\eta(t,x))$ is nonincreasing as t increases. Furthermore, according to the first inequality in (4), $\frac{d}{dt}\Phi(\eta(t,x))\leqq -b^2$ whenever $\chi(\eta)=1$. Now one can follow the argument in [13, p. 84] (with obvious changes) in order to obtain (v).

Denote $\omega(x)=\frac{\chi(x)}{\|x\|+1}$. As in [13, p. 86] or [14], one sees from (5) that $\eta(1,x)=e^{-\theta(x)L}x-K(x)$, where $\theta(x)=\theta(1,x)$,

$$\theta(t,x)=\int_0^t\omega(\eta(s,x))ds$$

and

$$K(x) = -\int_0^1 [\exp(\theta(t,x)-\theta(1,x))L]\,\omega(\eta(t,x))\,\lambda(\eta(t,x))\,\psi'(\eta(t,x))\,dt. \tag{6}$$

Hence $\theta(x) \in [0,1]$. Using the definitions of ω and λ,

$$\|\omega(x)\lambda(x)\psi'(x)\| \leqq \frac{\|Lx\|}{\|x\|+1},$$

so it follows from (6) that $K(M)$ is bounded. Recall that ψ' is bounded away from 0 on bounded subsets of M. Therefore the mapping $x \mapsto \lambda(x)\psi'(x)$, $x \in M$, is compact, and the same argument as in [13] or [14] (using (iv)) shows that K is compact. So (vi) is verified. Finally, (vii) is a direct consequence of the fact that the vector field in (5) is equivariant. □

Now we are ready to state the main result of this section.

3.2. Theorem. *Suppose that Φ, ψ and M satisfy the assumptions at the beginning of this section and $\Phi|_M$ satisfies (C^*). Let $E = Y \oplus X$, where $Y = X^\perp$, X and Y are invariant and $LY \subset Y$. Define*

$$\Gamma_j = \{A \in \Sigma : A \subset M, \gamma_r(A, X) \geqq j\}$$

and

$$c_j = \inf_{A \in \Gamma_j} \sup_{x \in A} \Phi(x), \qquad j = 1, \ldots, N.$$

Suppose also that there exist finite numbers a, b such that $c_j \in (a,b)$ for $1 \leqq j \leqq N$ and $Y^G \cap \Phi^{-1}([a,b]) = \emptyset$. Then all c_j are critical values of $\Phi|_M$. Furthermore, if $c_j = \ldots = c_{j+p}$ for some j and some $p \geqq 0$, then $\gamma(K_{c_j}) \geqq p+1$.

Proof. It suffices to prove the second assertion. Let $c_j = \ldots = c_{j+p} \equiv c$, where $p \geqq 0$, and suppose that $\gamma(K_c) \leqq p$. Since K_c is compact (by (C^*)), it follows from Proposition 2.1 that $\gamma(K_c) = \gamma(N_\delta(K_c))$ for $\delta > 0$ sufficiently small. Let $\bar{\varepsilon} < \min\{b-c, c-a\}$, $U = N_\delta(K_c)$, and let ε, η be as in the conclusion of Lemma 3.1. Choose $A \in \Gamma_{j+p}$ such that $A \subset \Phi_{c+\varepsilon}$. According to Proposition 2.5, $\gamma_r(\overline{A - N_\delta(K_c)}) \geqq \gamma_r(A) - \gamma(N_\delta(K_c)) \geqq j+p-p = j$. So $\overline{A - N_\delta(K_c)} \in \Gamma_j$. Let $B = \eta(1, \overline{A - N_\delta(K_c)})$. Since $a < c - \bar{\varepsilon}$, $b > c + \bar{\varepsilon}$ and $Y^G \cap \Phi^{-1}([a,b]) = \emptyset$, $\eta(1,x) = x$ $\forall x \in \overline{A - N_\delta(K_c)} \cap Y^G$ by (iii) of Lemma 3.1. It follows therefore from Proposition 2.2 (using (vi) and (vii) of Lemma 3.1) that $\gamma_r(B) \geqq j$. This is a contradiction because $B \subset \Phi_{c-\varepsilon}$ according to (v) of Lemma 3.1. □

3.3. Remark. (i) Note that it does not follow from the theorem that $\Phi|_M$ has more than one critical point. Indeed, if all c_j coincide and if K_{c_j} consists of a single point in E^G, then $\gamma(K_{c_j}) = \infty$. So the conclusion is not violated.

(ii) Suppose that $E^G \cap K_{c_j} = \emptyset$ for $j = 1, \ldots, N$. Then $\Phi|_M$ has at least N distinct pairs of critical points because either all c_j are distinct or $c_j = \ldots = c_{j+p}$ for some j and some $p > 0$. In the latter case $\gamma(K_{c_j}) \geqq 2$, so K_{c_j} is an infinite set.

4. Existence of Brake Orbits

Let $H \in C^2(\mathbf{R}^{2N}, \mathbf{R})$ and suppose that $H(-p,q)=H(p,q)\ \forall (p,q) \in \mathbf{R}^{2N}$. Suppose also that the set $S=H^{-1}(1)$ is compact, bounds a star-shaped neighbourhood of the origin in $\mathbf{R}^{2N}$ and $x \cdot H'(x) \neq 0$ on S (the dot denotes the inner product in $\mathbf{R}^{2N}$). There exist two positive numbers, r and R, such that

$$r \leqq |x| \leqq R \ \forall x \in S. \tag{7}$$

Let ϱ be the largest number for which

$$T_y(S) \cap \{x \in \mathbf{R}^{2N} : |x| < \varrho\} = \emptyset \ \forall y \in S, \tag{8}$$

where $T_y(S)$ is the tangent hyperplane to S at y and $|x|=(x \cdot x)^{1/2}$.

We will be concerned with the existence of brake orbits for the Hamiltonian system

$$\dot{x}=JH'(x), \tag{HS}$$

which lie on the hypersurface S. Our goal is to prove the following

4.1. Theorem. *Let $H \in C^2(\mathbf{R}^{2N}, \mathbf{R})$ be such that:*

(i) $H(-p,q)=H(p,q)\ \forall (p,q) \in \mathbf{R}^{2N}$,

(ii) *The set $\mathscr{A}=\{x \in \mathbf{R}^{2N} : H(x) \leqq 1\}$ is nonempty, compact, star-shaped with respect to the origin and $S=H^{-1}(1)$ is the boundary of $\mathscr{A}$,*

(iii) $x \cdot H'(x) \neq 0\ \forall x \in S$.

If $R^2 < 2\varrho^2$, where R and ϱ are as in (7), (8), *then* (HS) *has at least N distinct brake orbits on S.*

As the first step towards proving the above result we will find a convenient variational formulation of the problem. In doing this we essentially follow [14]. It is known that changing H outside S does not change the orbits (see e.g. [12, Lemma 1.5] or [13, Proposition 6.47]). We may therefore assume that $H(0)=0$ and $H(x)=\alpha(x)^2$ if $x \neq 0$, where $\alpha(x)$ is the unique positive number such that $\dfrac{x}{\alpha(x)} \in S$. Then H is homogeneous of degree two, $H \in C^2(\mathbf{R}^{2N}-\{0\}, \mathbf{R}) \cap C^{1,1}(\mathbf{R}^{2N}, \mathbf{R})$ and $\dfrac{|H'(x)|}{|x|}$ is bounded [12, p. 160]. It is also clear that $H(-p,q) = H(p,q)$. Since

$$r \leqq \frac{|x|}{\alpha(x)} \leqq R$$

according to (7), it follows that

$$\frac{|x|^2}{R^2} \leqq H(x) \leqq \frac{|x|^2}{r^2} \ \forall x \in \mathbf{R}^{2N}. \tag{9}$$

By [14, Lemma 2.3], see also [12, p. 161], there is a one-to-one correspondence between brake orbits for (HS) on S (of unknown period T) and 2π-periodic brake orbits for

$$\dot{x}=\lambda JH'(x), \qquad \lambda>0, \tag{10}$$

on S. The number λ in (10) is unknown and related to T.

Now we introduce a suitable function space. Let $H^{1/2}(S^1, \mathbf{R}^{2N})$ be the Sobolev space of 2π-periodic $\mathbf{R}^{2N}$-valued functions

$$x = \sum_{k \in \mathbf{Z}} c_k e^{ikt}, \quad \textit{where} \quad c_k \in \mathbf{C}^{2N} \quad \textit{and} \quad c_{-k} = \bar{c}_k,$$

such that

$$\sum_{k \in \mathbf{Z}} (1 + |k|)\, |c_k|^2 < \infty .$$

Let

$$E = \{x = (p, q) \in H^{1/2}(S^1, \mathbf{R}^{2N}) : p(-t) = -p(t), q(-t) = q(t) \ \forall t\} .$$

For $x = (p, q) \in E$, set $z = p + iq$. Then

$$z = \sum_{k \in \mathbf{Z}} a_k e^{ikt}, \quad \textit{where} \quad a_k \in \mathbf{C}^N . \tag{11}$$

Since p is odd and q is even, it is easy to see that $Re\, a_k = 0 \ \forall k$. A convenient norm in E, equivalent to the $H^{1/2}$-norm, is given by

$$\|x\|^2 = 2\pi \Big(|a_0|^2 + \sum_{k \neq 0} |k|\, |a_k|^2 \Big) . \tag{12}$$

Note that

$$\|x\|_{L^2}^2 = \int_0^{2\pi} |x|^2 dt = 2\pi \sum_{k \in \mathbf{Z}} |a_k|^2 \leqq \|x\|^2 . \tag{13}$$

Let $E = E^- \oplus E^0 \oplus E^+$ be the orthogonal decomposition of E into the parts corresponding to $k < 0$, $k = 0$ and $k > 0$ in (11). If $e_1, \ldots, e_N$ is the standard basis in $\mathbf{R}^N$, then E^0 is spanned by $(p, q) = (0, e_j)$, $1 \leqq j \leqq N$, and $E^\pm$ by

$$(p, q) = (e_j \sin kt, \mp e_j \cos kt), \qquad 1 \leqq j \leqq N, \qquad 1 \leqq k < \infty .$$

For $x \in E$, let

$$\Phi(x) = \frac{1}{2} \int_0^{2\pi} (-J\dot{x} \cdot x)\, dt$$

and $\Phi'(x) = Lx$ (recall that $\Phi'(x) \in E$ via the Riesz representation theorem). It is easy to see that

$$\Phi(x) = \frac{1}{2} \int_0^{2\pi} (-i\dot{z} \cdot z)\, dt$$

(here $\cdot$ denotes the inner product in $\mathbf{C}^N$), so by (11) and (12), if $x = x^- + x^0 + x^+ \in E^- \oplus E^0 \oplus E^+$, then

$$\Phi(x) = \tfrac{1}{2} \|x^+\|^2 - \tfrac{1}{2} \|x^-\|^2 \quad \textit{and} \quad Lx = x^+ - x^- .$$

Let

$$\psi(x) = \frac{1}{2\pi} \int_0^{2\pi} H(x)\, dt \quad \textit{and} \quad M = \psi^{-1}(1) .$$

By [14, pp. 607–608], each critical point of $\Phi|_M$ corresponds to a 2π-periodic brake orbit for (10) on S, and therefore to a brake orbit for (HS) on S.

By [14, Proposition 2.10], $\psi \in C^{1,1}(E, \mathbf{R})$ and ψ' is compact. Since ψ is homogeneous of degree 2, $\langle \psi'(x), x\rangle = 2\psi(x) = 2\ \forall x \in M$. So M is a $C^{1,1}$-manifold and ψ' is bounded away from 0 on bounded subsets of M. By (9),

$$\|x\|_{L^2}^2 \leqq R^2 \int_0^{2\pi} H(x)dt \leqq 2\pi R^2 \quad \textit{whenever} \quad \psi(x) \leqq 1 . \tag{14}$$

In particular, M is bounded in L^2.

4.2. Lemma. *$\Phi|_M$ satisfies condition (C^*).*

Proof. We slightly modify the argument of [2, Lemma 3.7]. Suppose that $(x_n) \subset M$, $\Phi(x_n) \to c$ and

$$z_n = \frac{Lx_n - \lambda(x_n)\psi'(x_n)}{(\|x_n\| + 1)^{1/2}} \to 0 . \tag{15}$$

Since $\Phi(x_n) = \frac{1}{2}\|x_n^+\|^2 - \frac{1}{2}\|x_n^-\|^2$, there are positive constants a_1, a_2, a_3 such that

$$-a_3 + a_1\|x_n^+\| \leqq \|x_n^-\| \leqq a_2\|x_n^+\| + a_3 . \tag{16}$$

Recall that $\langle \psi'(x_n), x_n\rangle = 2\psi(x_n) = 2$ and $\langle Lx_n, x_n\rangle = 2\Phi(x_n)$. This and (15) imply

$$\lambda(x_n) = \Phi(x_n) - \tfrac{1}{2}(\|x_n\| + 1)^{1/2}\langle z_n, x_n\rangle .$$

So $|\lambda(x_n)| \leqq a_4 + a_5\|x_n\|^{3/2}$. Scalar multiplication of (15) by x_n^+ gives

$$\|x_n^+\|^2 = (\|x_n\| + 1)^{1/2}\langle z_n, x_n^+\rangle + \lambda(x_n)\int_0^{2\pi} H'(x_n)\cdot x_n^+\, dt .$$

Since $\dfrac{|H'(x)|}{|x|}$ is bounded and M is bounded in L^2, the integral above is also bounded. Consequently,

$$\|x_n^+\|^2 \leqq a_6 + a_7\|x_n\|^{3/2} . \tag{17}$$

Since $\|x_n^0\|_{L^2} = \|x_n^0\|$, (x_n^0) is bounded in E. This, (16) and (17) imply that (x_n) is bounded. After passing to a subsequence, $x_n \to \bar{x}$ weakly in E, strongly in L^2, and $\lambda(x_n) \to \bar{\lambda}$. Since ψ' is compact, it follows from (15) that $Lx_n = x_n^+ - x_n^-$ converges strongly. Therefore $x_n \to \bar{x}$ strongly in E. □

4.3. Lemma. *The set $\mathscr{B} = \{x \in E : \psi(x) \leqq 1\}$ satisfies condition $(\mathscr{G})$ of Definition 2.7.*

Proof. Let E_0 and r be given. Since the E- and the L^2-norm are equivalent on E_0, $\|w\| \leqq C\|w\|_{L^2}$, where C is a constant independent of $w \in E_0$. Let $y \in \mathscr{B} \cap (x + E_0)$, $\|x\| \leqq r$. Then $y = x + w$, $w \in E_0$. Using (13) and (14), it follows that

$$\|w\|_{L^2} \leqq \|y\|_{L^2} + \|x\|_{L^2} \leqq \sqrt{2\pi}\, R + r .$$

Therefore

$$\|y\| \leqq \|x\| + \|w\| \leqq r + C\|w\|_{L^2} \leqq (C+1)r + \sqrt{2\pi}\, RC . \quad □$$

Proof of Theorem 4.1. Let $T_0x=x$ and $T_1x(t)=x(t+\pi)$. Then $T=\{T_0, T_1\}$ is a unitary representation of $\mathbf{Z}_2$ in E. The fixed point set E^G of T consists of those $x \in E$ which are π-periodic [i.e., $a_k=0$ for all odd k in the Fourier expansion (11)].

We will use Theorem 3.2. Note that L is equivariant and selfadjoint, $L(E^-\oplus E^0) \subset E^-\oplus E^0$ (because $Lx=x^+-x^-$), and ψ is invariant. Recall that ψ' is compact and bounded away from zero on bounded subsets of M. By Lemma 4.2, $\Phi|_M$ satisfies (C^*). Let $\gamma_r(\cdot)=\gamma_r(\cdot, E^+)$ and

$$c_j=\inf_{A\in\Gamma_j}\sup_{x\in A}\Phi(x), \quad j=1,\dots,N,$$

where $\Gamma_j=\{A\in\Sigma: A\subset M, \gamma_r(A)\geqq j\}$. If $A\in\Gamma_j$, then $\gamma_r(A)\geqq 1$, so $A\cap E^+\neq\emptyset$ according to Proposition 2.6. Let $x\in A\cap E^+\subset M\cap E^+$. Then, using (13) and (9),

$$\Phi(x)=\frac{1}{2}\|x\|^2\geqq\frac{1}{2}\|x\|_{L^2}^2\geqq\frac{1}{2}r^2\int_0^{2\pi}H(x)dt=\pi r^2.$$

So $c_j\geqq\pi r^2$. Let E_1 be the N-dimensional subspace of E^+ corresponding to $k=1$ in (11) (that is, E_1 is spanned by $(p,q)=(e_j\sin t, -e_j\cos t)$, $1\leqq j\leqq N$). Let $A=M \cap(E^-\oplus E^0\oplus E_1)$. Then A is the boundary of the set $\{x\in E^-\oplus E^0\oplus E_1: \psi(x)\leqq 1\}$ in $E^-\oplus E^0\oplus E_1$, and this set satisfies condition ($\mathscr{G}$) according to Lemma 4.3. It follows therefore from Proposition 2.8 with $Y=E^-\oplus E^0$ and $X_0=E_1$ that $\gamma_r(A)=N$. If $x=x^-+x^0+x^+\in A$, then, by (9),

$$\|x^+\|_{L^2}^2\leqq\|x\|_{L^2}^2\leqq R^2\int_0^{2\pi}H(x)dt=2\pi R^2.$$

Observe that $\|x^+\|_{L^2}=\|x^+\|$ whenever $x^+\in E_1$ [cf. (12), (13)]. Hence

$$\Phi(x)=\tfrac{1}{2}\|x^+\|^2-\tfrac{1}{2}\|x^-\|^2\leqq\tfrac{1}{2}\|x^+\|^2=\tfrac{1}{2}\|x^+\|_{L^2}^2\leqq\pi R^2,$$

so $c_j\leqq\pi R^2$.

We have shown that $c_j\in[\pi r^2,\pi R^2]$ for $j=1,\dots,N$. Since $\Phi(x)=\frac{1}{2}\|x^+\|^2-\frac{1}{2}\|x^-\|^2$, $\Phi|_{Y^G}\leqq 0$, where $Y^G=(E^-\oplus E^0)^G$. Now all hypotheses of Theorem 3.2 are verified. Assume for the moment that all $x\in K_{c_j}$, $1\leqq j\leqq N$, have minimal period 2π. Then $E^G\cap K_{c_j}=\emptyset$ and it follows from Remark 3.3(ii) that $\Phi|_M$ possesses at least N pairs of critical points. If x_1 and x_2 are two distinct points in K_{c_j} corresponding to the same brake orbit for (10), then $x_2(t)=x_1(t+\alpha)$, where $\alpha\in(0,2\pi)$, and the Fourier expansion (11) shows that $\alpha=\pi$. So $x_2=T_1x_1$. Therefore there exist at least N distinct brake orbits.

It remains to show that if $x\in K_{c_j}$, $1\leqq j\leqq N$, then indeed x has minimal period 2π. Our argument is close to that in [6, Lemma 6]. Suppose x is $\frac{2\pi}{m}$-periodic, $m\geqq 1$. Then, using (10) and the fact that $H'(x)\cdot x=2H(x)$,

$$\Phi(x)=\frac{1}{2}\int_0^{2\pi}(-J\dot{x}\cdot x)dt=\frac{1}{2}\lambda\int_0^{2\pi}H'(x)\cdot x\,dt=\lambda\int_0^{2\pi}H(x)dt=2\pi\lambda. \tag{18}$$

Let $x(t)=\bar{x}+\tilde{x}(t)$, where $\bar{x}\in E^0$ and $\tilde{x}$ has mean value zero. Then, by the Wirtinger inequality $\|\dot{\tilde{x}}\|_{L^2}\leqq\frac{1}{m}\|\tilde{x}\|_{L^2}$ and (10),

$$\begin{aligned}\Phi(x)&=\frac{1}{2}\int_0^{2\pi}(-J\dot{x}\cdot x)\,dt=\frac{1}{2}\int_0^{2\pi}(-J\dot{\tilde{x}}\cdot\tilde{x})\,dt\leqq\frac{1}{2}\|\dot{\tilde{x}}\|_{L^2}\|\tilde{x}\|_{L^2}\\&\leqq\frac{1}{2m}\|\dot{\tilde{x}}\|^2_{L^2}=\frac{1}{2m}\|\dot{x}\|^2_{L^2}=\frac{1}{2m}\int_0^{2\pi}|\lambda H'(x)|^2\,dt.\end{aligned}$$

Since $|H'(x)|\leqq\frac{2}{\varrho}\ \forall x\in S$ (see e.g. [2, Proof of Theorem 4.10]),

$$\Phi(x)\leqq\frac{4\pi\lambda^2}{m\varrho^2}.$$

Combining this with (18), $\lambda\geqq\frac{1}{2}m\varrho^2$ and $\Phi(x)\geqq\pi m\varrho^2$. Since $c_j\leqq\pi R^2$, it follows from the hypothesis $R^2<2\varrho^2$ that

$$\pi m\varrho^2\leqq\Phi(x)\leqq\pi R^2<2\pi\varrho^2.$$

Hence $m<2$. The minimal period of x is therefore 2π. □

4.4. Remarks. (i) If we remove the assumption that $R^2<2\varrho^2$ in Theorem 4.1, then $\Phi|_M$ will still have a critical value in $[\pi r^2,\pi R^2]$. So there exists at least one brake orbit of (HS) on S, and we recover the main result of [14].

(ii) Using [6, Lemma 6] it is easy to see that the hypothesis $R^2<2\varrho^2$ may be replaced by $R^2<\sqrt{2}r\varrho$.

Appendix

We will prove the following

Generalized Borsuk Theorem. *Let W be an open bounded neighbourhood of $0\in\mathbf{R}^m\times\mathbf{R}^n$ such that if $(x,y)\in W$, then $(x,-y)\in W$. Let $f=(g,h):\bar{W}\to\mathbf{R}^m\times\mathbf{R}^n$ be a continuous mapping with $g(x,-y)=g(x,y)$, $h(x,-y)=-h(x,y)\ \forall(x,y)\in\bar{W}$, $f|_{\partial W}\neq 0$ and $f(x,0)=(x,0)\ \forall(x,0)\in\partial W$. Then the Brouwer degree $\deg(f,W,0)$ is an odd integer.*

Proof. Our argument is essentially an adaptation of Nirenberg's proof in [11] (see also [10] and [21]) to the simpler case of $\mathbf{Z}_2$-symmetry.

Given $\varepsilon>0$, there exists a C^∞ mapping f_ε such that $|f(x,y)-f_\varepsilon(x,y)|\leqq\varepsilon\ \forall(x,y)\in\bar{W}$. We may assume that f_ε is equivariant by replacing it if necessary with $\bar{f}_\varepsilon(x,y)=\frac{1}{2}(g_\varepsilon(x,y),h_\varepsilon(x,y))+\frac{1}{2}(g_\varepsilon(x,-y),-h_\varepsilon(x,-y))$. For $\delta>0$, let $\chi:[0,\infty)\to[0,1]$ be a C^∞ function satisfying $\chi(t)=1$ if $0\leqq t\leqq\frac{\delta}{2}$ and $\chi(t)=0$ if $t\geqq\delta$. Define

$$\tilde{f}(x,y)=\chi(|y|)(x,y)+(1-\chi(|y|))f_\varepsilon(x,y).$$

It is easy to see that if $\delta=\delta(\varepsilon)$ is small enough, $|\tilde{f}(x,y)-f(x,y)|\leqq 2\varepsilon\ \forall (x,y)\in\partial W$. By the continuity property of degree,

$$\deg(f,W,0)=\deg(\tilde{f},W,0) \tag{A.1}$$

whenever ε is sufficiently small. Let

$$V=\left\{(x,y)\in W: |y|<\frac{\delta}{2}\right\}.$$

Since $\tilde{f}|_{\partial V}\neq 0$ and $\tilde{f}|_V$ is the identity mapping,

$$\deg(\tilde{f},W,0)=\deg(\tilde{f},V,0)+\deg(\tilde{f},W-\bar{V},0)=1+\deg(\tilde{f},W-\bar{V},0). \tag{A.2}$$

We will show that $\deg(\tilde{f},W-\bar{V},0)$ is an even integer. Let $A=(a_{ij})$ be an $m\times n$ and $B=(b_{ij})$ an $n\times n$ real matrix. Let $(x,y)=(x_1,\ldots,x_m,y_1,\ldots,y_n)$, $\tilde{f}=(\tilde{g}_1,\ldots,\tilde{g}_m,\tilde{h}_1,\ldots,\tilde{h}_m)$ and

$$\begin{aligned}\bar{g}_i(x,y,A,B)&=\tilde{g}_i(x,y)+\sum_{j=1}^n a_{ij}y_j^2, \quad 1\leqq i\leqq m,\\ \bar{h}_i(x,y,A,B)&=\tilde{h}_i(x,y)+\sum_{j=1}^n b_{ij}y_j, \quad 1\leqq i\leqq n.\end{aligned} \tag{A.3}$$

If A and B are fixed and sufficiently small (in the sense that $|a_{ij}|<\varepsilon_0$, $|b_{ij}|<\varepsilon_0$ for all i,j and an appropriate $\varepsilon_0>0$), then, setting $\bar{f}=\bar{f}_{A,B}=(\bar{g}_1,\ldots,\bar{g}_m,\bar{h}_1,\ldots,\bar{h}_n)$,

$$\deg(\tilde{f},W-\bar{V},0)=\deg(\bar{f},W-\bar{V},0) \tag{A.4}$$

by the continuity property of degree. Assume for the moment that 0 is a regular value of $\bar{f}|_{W-\bar{V}}$. Then $\bar{f}(x,y)=0$ for finitely many $(x,y)\in W-\bar{V}$ and each such (x,y) gives a contribution of $+1$ or -1 to the degree. Since $y\neq 0$ on $W-\bar{V}$ and $\bar{f}$ is equivariant, the zeros of $\bar{f}|_{W-\bar{V}}$ come in pairs $(x,\pm y)$. So $\deg(\bar{f},W-\bar{V},0)$ is the sum of an even number of terms, each equal to $+1$ or -1, and is therefore an even integer. Using this and (A.1), (A.2), (A.4), it follows that $\deg(f,W,0)$ is an odd integer.

It remains to show that A and B may be chosen in such a way that 0 is a regular value of $\bar{f}$. Let M_A and M_B be the spaces of all sufficiently small $m\times n$ and $n\times n$ real matrices. Let

$$F:(W-\bar{V})\times M_A\times M_B\to\mathbf{R}^m\times\mathbf{R}^n$$

be the mapping given by $F(x,y,A,B)=\bar{f}_{A,B}(x,y)$. We claim that the derivative of F is transversal to the origin (i.e. surjective) whenever $F(x,y,A,B)=0$. To see this, let $z=(\xi,\eta)\in\mathbf{R}^m\times\mathbf{R}^n$. We need to find $\bar{x}$, $\bar{y}$, $\bar{A}$, $\bar{B}$ such that

$$(D_xF)\bar{x}+(D_yF)\bar{y}+(D_AF)\bar{A}+(D_BF)\bar{B}=z, \tag{A.5}$$

where D_xF etc.... are the partial derivatives of F at (x,y,A,B). A simple computation using (A.3) shows that (A.5) with $\bar{x}=0$ and $\bar{y}=0$ is equivalent to the system of equations

$$\begin{aligned}\sum_{j=1}^n \bar{a}_{ij}y_j^2&=\xi_i, \quad 1\leqq i\leqq m,\\ \sum_{j=1}^n \bar{b}_{ij}y_j&=\eta_i, \quad 1\leqq i\leqq n.\end{aligned} \tag{A.6}$$

Since $y \neq 0$ in $W - \bar{V}$, $y_j \neq 0$ for some j. So (A.6) can be solved for $\bar{A}$ and $\bar{B}$. Therefore (A.5) is satisfied with $\bar{x} = 0$, $\bar{y} = 0$ and $\bar{A}$, $\bar{B}$ just found. This proves the claim. Finally, it follows from the transversality theorem (see e.g. [7, p. 68]) that 0 is a regular value of $\bar{f}_{A,B}$ for almost all $A \in M_A$ and $B \in M_B$. □

Acknowledgement. I would like to thank H. O. Walther for bringing the results of Bögle [3] to my attention.

References

1. Benci, V.: On critical point theory for indefinite functionals in the presence of symmetries. Trans. Am. Math. Soc. **274**, 533–572 (1982)
2. Berestycki, H., Lasry, J.M., Mancini, G., Ruf, B.: Existence of multiple periodic orbits on star-shaped Hamiltonian surfaces. Commun. Pure Appl. Math. **38**, 253–289 (1985)
3. Bögle, J.G.: Indextheorien zur Existenz von periodischen Lösungen Hamiltonscher Systeme. Thesis, Ludwig-Maximilians-Universität München, 1987
4. Ekeland, I., Lasry, J.M.: On the number of periodic trajectories for a Hamiltonian flow on a convex energy surface. Ann. Math. **112**, 283–319 (1980)
5. Girardi, M.: Multiple orbits for Hamiltonian systems on starshaped surfaces with symmetries. Ann. Inst. Henri Poincaré Nouv. Ser., Sect. B **1**, 285–294 (1984)
6. Girardi, M., Matzeu, M.: Solutions of minimal period for a class of nonconvex Hamiltonian systems and applications to the fixed energy problem. Nonl. Anal., Theory Methods Appl. **10**, 371–382 (1986)
7. Guillemin, V., Pollack, A.: Differential topology. Englewoods Cliffs. Prentice-Hall 1974
8. Hofer, H., Zehnder, E.: Periodic solutions on hypersurfaces and a result by C. Viterbo. Invent. Math. **90**, 1–9 (1987)
9. Lloyd, N.G.: Degree theory. Cambridge: Cambridge University Press 1978
10. Michalek, R.: A $\mathbf{Z}^p$ Borsuk-Ulam theorem and index theory with a multiplicity result. Preprint
11. Nirenberg, L.: Comments on nonlinear problems. Le Matematiche **36**, 109–119 (1981)
12. Rabinowitz, P.H.: Periodic solutions of Hamiltonian systems. Commun. Pure Appl. Math. **31**, 157–184 (1978)
13. Rabinowitz, P.H.: Minimax methods in critical point theory with applications to differential equations. CBMS 65. Am. Math. Soc., Providence, R.I. 1986
14. Rabinowitz, P.H.: On the existence of periodic solutions for a class of symmetric Hamiltonian systems. Nonl. Anal., Theory Methods Appl. **11**, 599–611 (1987)
15. Rabinowitz, P.H.: On a theorem of Hofer and Zehnder. In: Periodic solutions of Hamiltonian systems and related topics. (to appear)
16. Seifert, H.: Periodische Bewegungen mechanischer Systeme. Math. Z. **51**, 197–216 (1948)
17. Szulkin, A.: Critical point theory of Ljusternik-Schnirelmann type and applications to partial differential equations. In: Séminaire de Mathématiques Supérieures, Les Presses de l'Université de Montréal. (to appear)
18. Tarantello, G.: Subharmonic solutions for Hamiltonian systems via a $\mathbf{Z}_p$ pseudoindex theory. Preprint
19. Groesen, E.W.C. van: Existence of multiple normal mode trajectories on convex energy surfaces of even, classical Hamiltonian systems. J. Differ. Equations **57**, 70–89 (1985)
20. Viterbo, C.: A proof of Weinstein's conjecture in $\mathbf{R}^{2n}$. Ann. Inst. Henri Poincaré, Nouv. Sér. Sect. B **4**, 337–356 (1987)
21. Wang, Z.Q.: A $\mathbf{Z}_p$ Borsuk-Ulam theorem. Preprint
22. Weinstein, A.: Periodic orbits for convex hamiltonian systems. Ann. Math. **108**, 507–518 (1978)

Received May 13, 1988

Math. Ann. 283, 257–270 (1989)

Mathematische Annalen

On the Principle of Linearized Stability for Variational Inequalities

Pavol Quittner

Institute of Physics, EPRC, Slovak Academy of Sciences, Dúbravská cesta 9, CS-84228 Bratislava, Czechoslovakia

Introduction

In this paper we study the stability of stationary solutions of parabolic variational inequalities in Hilbert spaces. It is well known that in the case of semilinear parabolic equations the spectrum of "linearized" operator determines the Lyapunov stability of an equilibrium (e.g. [4]). In the case of inequalities this problem is more complicated: the stability of an equilibrium may "substantially" change if we linearize the operator in the inequality (see Example 1). Nevertheless, it is still possible to find some conditions which are independent of the nonlinearity and which are sufficient for the stability or the instability of the equilibrium, respectively. Moreover, these conditions are under some additional assumptions also necessary.

To point out the main results, let us consider the following problem, to which our theory can be applied: let $\Omega \subset \mathbb{R}^N$ be a smooth bounded domain, let L be a second-order elliptic operator of the form $Lu := -\frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right)$ with $a_{ij} \in L^\infty(\Omega)$, $a_{ij}(x)\xi_i\xi_j \geqq \alpha|\xi|^2$, $\alpha > 0$, and let $F(u) := f(x, u, \nabla u)$, where $f = f(x, u, p)$: $\Omega \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ is measurable in x and C^1 in u and p, $f(\cdot, 0, 0) \in L^2(\Omega)$, $\frac{\partial f}{\partial p}$ is bounded and $\left|\frac{\partial f}{\partial u}(x, u, p)\right| \leqq a(x) + C(|u|^\gamma + |p|^{2/N})$ for some $a \in L^N(\Omega)$, $\gamma \leqq 2/(N-2)$ (if $N > 2$) and $C > 0$. Let $\tilde{K}$ be a closed convex set in the Sobolev space $H_0^1(\Omega)$ and let $\tilde{u} \in \tilde{K}$ be a stationary solution of the inequality

$$
\begin{aligned}
&u(t) \in \tilde{K} \\
&\left(\frac{du}{dt} + Lu + F(u), v - u\right) \geqq 0 \quad \forall v \in \tilde{K} \quad \text{a.e. in } (0, T) \qquad (0) \\
&u(0) = u_0 \ ,
\end{aligned}
$$

i.e.

$$
(L\tilde{u} + F(\tilde{u}), v - \tilde{u}) = \int_\Omega \left(a_{ij}(x)\frac{\partial \tilde{u}}{\partial x_j}\frac{\partial (v - \tilde{u})}{\partial x_i} + f(x, \tilde{u}, \nabla \tilde{u})(v - \tilde{u})\right)dx \geqq 0 \quad \forall v \in \tilde{K} \ .
$$

Put

$$Au := Lu + F'(\tilde{u})u$$

$$K := \tilde{K} - \tilde{u} = \{v - \tilde{u};\ v \in \tilde{K}\}$$

$$F_0 := L\tilde{u} + F(\tilde{u})$$

$$\lambda_I := \liminf_{\substack{u \in K \\ \|u\|_{H^1} \to 0}} \frac{(Au + F_0, u)}{(u, u)} .$$

Then the condition $\lambda_I > 0$ is sufficient for the asymptotic stability of the stationary solution $u_0 = \tilde{u}$ of the inequality (0) in the topology of $H_0^1(\Omega)$ (more precisely see Theorem 1).

In order to see that the condition $\lambda_I < 0$ may be sufficient for the instability result, suppose, moreover, that K is a cone with its vertex at zero, $F_0 = 0$ and A is symmetric (i.e. $a_{ij} = a_{ji}$ and f is independent of p) and denote by $\sigma(A)$ the spectrum of the operator A and by $\sigma_K(A)$ the set of all (real) eigenvalues of the inequality

$$u \in K\colon (Au - \lambda u, v - u) \geqq 0 \quad \forall v \in K , \tag{1}$$

i.e. $\sigma_K(A) := \{\lambda \in \mathbb{R};$ the inequality (1) has a nontrivial solution$\}$.

Then we have

$$\lambda_I = \min_{\substack{u \in K \\ \|u\|_{H^1} = 1}} \frac{(Au, u)}{(u, u)} = \min \sigma_K(A) \geqq \min \sigma(A) .$$

Let u_I be any nontrivial solution of (1) with $\lambda = \lambda_I$ and let $\delta > 0$. Then the function $u(t) = \tilde{u} + \delta e^{-\lambda_I t} u_I$ is a solution of the linearized inequality

$$u(t) \in \tilde{K}$$

$$\left(\frac{du}{dt} + Lu + F(\tilde{u}) + F'(\tilde{u})(u - \tilde{u}), v - u\right) \geqq 0 \quad \forall v \in \tilde{K} \quad \text{a.e. in } (0, T)$$

$$u(0) = \tilde{u} + \delta u_I ,$$

which implies that the condition $\lambda_I < 0$ guarantees the instability of the solution $u_0 = \tilde{u}$ for the linearized inequality. If, moreover,

$$\lambda_I = \min_{\substack{u \in K - u_I \\ u \neq 0}} \frac{(Au, u)}{(u, u)} ,$$

then Theorem 2 implies the instability result also for the nonlinear inequality (0) (in the topology of $L^2(\Omega)$).

An application of these results to a more concrete problem is given in Example 3. Example 2 shows that the condition $\operatorname{Re}(\sigma_K(A) \cup \sigma(A)) > 0$ is not sufficient for the stability in the nonsymmetric case and in Example 1 it is shown that the condition $\exists \lambda \in \sigma_K(A), \lambda < 0$, is, in general, not sufficient for the instability result.

The proofs of Theorems 1 and 2, which are the main results of this paper, are based on the existence and regularity results of Brézis [1, 2].

Main Results

We shall suppose that V and H are real Hilbert spaces with norms $\|\cdot\|$ and $|\cdot|$, respectively, and $V \subset H \subset V'$, where the inclusions are dense and compact. By $(\cdot,\cdot)$ we denote the duality between V' and V and also the scalar product in H. We shall study the inequality

$$u(t) \in K$$

$$\left(\frac{du}{dt} + Au + N(u) + F_0, v - u\right) \geqq 0 \quad \forall v \in K \quad \text{a.e. in } (0, T) \tag{2}$$

$$u(0) = u_0 \ ,$$

where

K is a closed convex set in V, $0 \in K$,
$A: V \to V'$ is a continuous linear map, $A = A_1 + A_2$,
$A_1: V \to V'$ is symmetric and coercive
(i.e. $(A_1 u, v) = (A_1 v, u)$ and $(A_1 v, v) \geqq \alpha \|v\|^2$ for some $\alpha > 0$ and any $u, v \in V$),
$A_2: V \to H$ is continuous (i.e. $|A_2 v| \leqq C_1 \|v\|$ for some $C_1 > 0$ and any $v \in V$),
$N: V \to V'$ is a nonlinear map of the form $N = N_1 + N_2$,
N_1 is a gradient of a C^1 convex functional $\Phi: V \to \mathbb{R}$, $\Phi(0) = 0$, $\Phi'(0) = 0$, $\Phi''(0) = 0$,
$N_2: V \to H$ is a continuous map satisfying
$|N_2(u) - N_2(v)| \leqq C_2 \|u - v\|$ for any $u, v \in B$ and some $C_2 > 0$,
$(N_2(u), u) \geqq -\varphi(\|u\|)\,\|u\|^2$ for any $u \in B$,
where $\varphi(t) \to 0$ as $t \to 0+$ and B is a given neighbourhood of zero in V,
$F_0 \in V'$, $(F_0, v) \geqq 0$ for any $v \in K$,
$u_0 \in K$.

By a (strong) *solution* of (2) we mean a function $u \in C([0, T], K)$ such that $u: (0, T) \to H$ is differentiable a.e. and fulfils (2). The main result of this paper is the following

Theorem 1. *Let*

$$\lambda_I := \liminf_{u \in K,\, \|u\| \to 0} \frac{(Au + F_0, u)}{|u|^2} > 0 \ .$$

Then the solution $u_0 = 0$ of (2) is asymptotically stable in the topology of V, i.e. for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any $u_0 \in K$ with $\|u_0\| < \delta$ there exists an unique solution u of (2) on $[0, +\infty)$ and fulfils $\|u(t)\| < \varepsilon$ for any $t \geqq 0$. Moreover, for any $\eta > 0$ we have

$$\frac{du}{dt} \in L^2((0, \infty), H) \cap L^2((\eta, \infty), V)$$

and for any $\lambda < \lambda_I$, $\Lambda \geqq 0$ there exist constants $C = C(\lambda, \Lambda) > 0$ and $\delta = \delta(\lambda, \varepsilon) > 0$ such that $\|u_0\| < \delta$ implies

$$|u(t)| \leqq e^{-\lambda t}|u_0| \quad \textit{for any} \quad t \geqq 0$$

$$\|u(t)\| \leqq Ce^{-\lambda t}(\|u_0\| + \sqrt{(F_0, u_0)}e^{-At}) \quad \textit{for any} \quad t \geqq 0$$

$$\left|\frac{du}{dt}(t)\right| \leqq C\left(1 + \frac{1}{\sqrt{t}}\right)e^{-\lambda t}(\|u_0\| + \sqrt{(F_0, u_0)}e^{-At}) \quad \textit{for a.e.} \ t \geqq 0 \ .$$

If there exists $g_0 \in H$ *such that* $(A_1 u_0 + N_1(u_0) + F_0 - g_0, v - u_0) \geqq 0$ *for any* $v \in K$, *then*

$$\frac{du}{dt} \in L^\infty((0, \infty), H) \cap L^2((0, \infty), V) \ .$$

If, moreover, $N_1 \equiv 0$, $F_0 \in V$ *and* $(I + \mu A_1)^{-1}(K) \subset K$ *for any* $\mu > 0$, *then* $A_1 u(t) \in H$ *for any* $t \geqq 0$ *and*

$$\begin{aligned} |A_1 u(t)| &\leqq Ce^{-\lambda t}(\|u_0\| + |A_1 u_0| \max(1 - t, 0) + \sqrt{(F_0, u_0)}e^{-At}) + \sqrt{2(A_1 F_0, u(t))^-} \\ &\leqq C(e^{-\lambda t}|A_1 u_0| + e^{-\lambda t/2}\|F_0\|) \quad \textit{for any} \quad t \geqq 0 \ , \end{aligned}$$

where $x^- := \max(-x, 0)$.

Proof. By C we shall denote various constants, which may depend on the given constants $\alpha, C_1, C_2, \|A_1\|_{L(V,V')}, \lambda, A$ and the function φ, but do *not* depend on u_0 and F_0. Without loss of generality we may suppose that $B = V$ (otherwise we redefine N_2 outside B) and φ is strictly increasing and continuous.

Let $u_0 \in K$, $\|u_0\| < \delta < 1$. Put

$$\Phi_1(v) := \Phi(v) + \tfrac{1}{2}(A_1 v, v) + |(F_0, v)| \ ,$$

$$\tilde{A}_1(v) := A_1 v + N_1(v) + F_0 \ ,$$

$$\tilde{A}(v) := Av + N(v) + F_0 \ .$$

Then $\tilde{A} : V \to V'$ is pseudomonotone, $(\tilde{A}(v), v) \geqq \frac{\alpha}{2}\|v\|^2 - C|v|^2$ for any $v \in K$, thus the results of Brézis [1, Corollaire II.1, Remarque II.5] imply the existence of a weak solution of (2), i.e. for any $T > 0$ there exists $\hat{u} \in L^2((0, T), K) \cap C([0, T], H)$ such that $\hat{u}(0) = u_0$ and

$$\int_0^T \left(\frac{dv}{dt} + \tilde{A}(\hat{u}), v - \hat{u}\right)e^{-2Ct}dt \geqq -\frac{1}{2}|v(0) - u_0|^2 \quad \forall v \in L^2((0, T), K) \ ,$$

$$\frac{dv}{dt} \in L^2((0, T), V') \ .$$

Put $f(t) := -A_2 \hat{u}(t) - N_2(\hat{u}(t)$. Then $f \in L^2((0, T), H)$ and again the results of Brézis [1, Théorème II.8, the proof of Corollaire II.2] and [2, Proposition 5, Lemme 6] imply that there exists a unique $u \in C([0, T], H)$ such that

$$\begin{aligned} &u(0) = u_0 \\ &\left(\frac{du}{dt} + \tilde{A}_1(u) - f, v - u\right) \geqq 0 \quad \forall v \in K \quad \text{a.e. in } (0, T) \ . \end{aligned} \tag{3}$$

Moreover $u:[0,T]\to H$ is absolutely continuous and differentiable a.e.,

$$\frac{du}{dt}\in L^2((0,T),H)\,,\quad \left(\int_0^T \left|\frac{du}{dt}\right|^2 dt\right)^{1/2} \leqq \left(\int_0^T |f|^2 dt\right)^{1/2} + \sqrt{\Phi_1(u_0)} \tag{4}$$

$$u\in C([0,T],K)$$

$$t\mapsto \Phi_1(u(t)) \quad \text{is absolutely continuous and} \quad \left|\frac{du}{dt}\right|^2 + \frac{d}{dt}\,\Phi_1(u(t)) = \left(f,\frac{du}{dt}\right) \quad \text{a.e.}$$

Since both u and $\hat{u}$ are weak solutions of the inequality (3) (cf. [1, Remarque II.11], the uniqueness result of [1, Théorème II.3, Remarque II.5] implies $u=\hat{u}$.

By putting $v=0$ in (3) we get

$$\left(\frac{du}{dt}+\tilde{A}(u),u\right)\leqq 0 \qquad \text{a.e. in } (0,T)\ . \tag{5}$$

Let $\lambda\in(0,\lambda_I)$ be fixed. Then there exists $\varepsilon_1>0$ such that

$$(Au+F_0,u)\geqq\frac{\lambda+\lambda_I}{2}\,|u|^2 \qquad \text{for any } u\in K,\ \|u\|\leqq\varepsilon_1\ ,$$

$$(Au,u)\geqq\frac{\alpha}{2}\,\|u\|^2 - C|u|^2 \qquad \text{for any } u\in V\ ,$$

$$(N(u),u)\geqq -\varphi(\|u\|)\,\|u\|^2 \qquad \text{for any } u\in V\ .$$

By choosing $\eta:=(\lambda_I-\lambda)/(4\max(\lambda_I,C))$ we obtain

$$(Au+F_0,u)\geqq\frac{\eta\alpha}{2}\,\|u\|^2+\lambda|u|^2+\eta(F_0,u) \qquad \text{for any } u\in K,\ \|u\|\leqq\varepsilon_1\ .$$

By using (5) and putting $\beta:=\eta\alpha/4$ we get

$$\frac{1}{2}\,\frac{d}{dt}\,|u|^2+\lambda|u|^2+\beta\,\|u\|^2+\eta(F_0,u)\leqq 0 \qquad \text{a.e. in } (0,T_\beta)\ , \tag{6}$$

where $T_\beta:=\sup\{t;\ \|u(\tau)\|\leqq\min(\varphi^{-1}(\beta),\varepsilon_1)$ for any $\tau\in[0,t]\}>0$ if δ is small enough. The inequality (6) implies

$$|u(t)|\leqq|u_0|\,e^{-\lambda t} \tag{6a}$$

$$\int_{t_1}^{t_2}\|u\|^2dt\leqq C(|u(t_1)|^2-|u(t_2)|^2)\leqq C|u_0|^2e^{-2\lambda t_1} \tag{6b}$$

$$\int_{t_1}^{t_2}(F_0,u)\,dt\leqq C(|u(t_1)|^2-|u(t_2)|^2)\leqq C|u_0|^2e^{-2\lambda t_1} \tag{6c}$$

for any $t,t_1,t_2\in[0,T_\beta]$. Now it follows from (4) and (6b) that

$$\begin{aligned}\alpha\|u(t)\|^2 &\leqq \Phi_1(u(t)) = \Phi_1(u_0) + \int_0^t \left(\left(f, \frac{du}{dt}\right) - \left|\frac{du}{dt}\right|^2\right) dt \\ &\leqq \Phi_1(u_0) + \frac{1}{2}\int_0^t |f|^2 dt \leqq C\left(\|u_0\|^2 + (F_0, u_0) + \int_0^t \|u\|^2 dt\right) \\ &\leqq C(\|u_0\|^2 + (F_0, u_0) + |u_0|^2) < \alpha(\min(\varphi^{-1}(\beta/2), \varepsilon_1/2))^2 \end{aligned} \tag{6d}$$

for any $t \in [0, T_\beta]$ provided δ is sufficiently small. Hence $T_\beta = T$.

By putting $v := u(t+h)$ in (3) we obtain

$$\left(\frac{du}{dt}(t) + Au(t) + N(u(t)) + F_0, \frac{u(t+h)-u(t)}{h}\right) \geqq 0 \quad \text{a.e. in } (0, T-h) .$$

Similarly,

$$\left(\frac{du}{dt}(t+h) + Au(t+h) + N(u(t+h)) + F_0, \frac{u(t)-u(t+h)}{h}\right) \geqq 0$$
$$\text{a.e. in } (0, T-h) .$$

By adding the last two inequalities we get

$$\left(\left(\frac{du}{dt}\right)_h(t) + Au_h(t) + (N \circ u)_h(t), u_h(t)\right) \leqq 0 \qquad \text{a.e. in } (0, T-h) ,$$

where $w_h(t) := (w(t+h) - w(t))/h$. We have

$$(Au_h, u_h) \geqq \frac{\alpha}{2}\|u_h\|^2 - C|u_h|^2$$

$$((N \circ u)_h, u_h) \geqq -C_2\|u_h\| \cdot |u_h| \geqq -\frac{\alpha}{4}\|u_h\|^2 - C|u_h|^2 ,$$

therefore,

$$\frac{1}{2}\frac{d}{dt}|u_h|^2 + \frac{\alpha}{4}\|u_h\|^2 - C|u_h|^2 \leqq 0 \qquad \text{a.e. in } (0, T-h) . \tag{7}$$

Consequently,

$$|u_h(t_2)| \leqq |u_h(t_1)| e^{C(t_2 - t_1)} \tag{7a}$$

$$\int_{t_1}^{t_2} \|u_h\|^2 dt \leqq C\left(|u_h(t_1)|^2 - |u_h(t_2)|^2 + \int_{t_1}^{t_2} |u_h|^2 dt\right) \tag{7b}$$

for any $t_1, t_2 \in [0, T-h]$.

The last inequality, together with (4), implies that $u : (0, T) \to V$ is (locally) absolutely continuous, $\frac{du}{dt} \in L^2_{\text{loc}}((0, T), V)$ and

$$\int_{t_1}^{t_2} \left\|\frac{du}{dt}\right\|^2 dt \leqq C\left(\left|\frac{du}{dt}(t_1)\right|^2 + \int_{t_1}^{t_2} \left|\frac{du}{dt}\right|^2 dt\right) \tag{7c}$$

for a.e. $t_1 \in (0, T)$ and any $t_2 \in (t_1, T)$.

Choose $\Lambda \geqq 0$, $C_* \geqq 1$ and suppose there exists $t_2^* \in (0, T)$ such that

$$\|u(t_2^*)\| \geqq 10\, C_* \exp(-\lambda t_2^*)(\|u_0\| + \sqrt{(F_0, u_0)}\exp(-\Lambda t_2^*)) .$$

We may suppose

$$t_2^* = \min\{t \in (0, T);\ \|u(t)\| \geqq 10\, C_* \exp(-\lambda t)(\|u_0\| + \sqrt{(F_0, u_0)}\exp(-\Lambda t))\}$$

and we put

$$t_1^* := \min\{t \in (0, t_2^*);\ \|u(\tau)\| \geqq C_* \exp(-\lambda\tau)(\|u_0\| + \sqrt{(F_0, u_0)}\exp(-\Lambda\tau)) \text{ for any } \tau \in [t, t_2^*]\} .$$

Since (6d) implies $\|u(t)\| \leqq C(\|u_0\| + \sqrt{(F_0, u_0)})$ for every t, we have $t_1^* \geqq 3$ (if C_* is sufficiently large). By using (6b) we get

$$C|u_0|^2 e^{-2\lambda t_1^*} \geqq \int_{t_1^*}^{t_2^*} \|u\|^2 dt \geqq \int_{t_1^*}^{t_2^*} C_*^2 e^{-2\lambda t}\|u_0\|^2 dt = \frac{C_*^2\|u_0\|^2}{2\lambda}(e^{-2\lambda t_1^*} - e^{-2\lambda t_2^*}) \tag{8}$$

hence $e^{-2\lambda t_1^*} < (1+\zeta)e^{-2\lambda t_2^*}$, where $\zeta \leqq C/(C_*^2 - C)$. Consequently, for any $\gamma > 0$ fixed we have

$$t_2^* - t_1^* \leqq \min\left(1, \frac{1}{\Lambda}, \frac{1}{\lambda}, \gamma\right) , \tag{9}$$

if $C_* = C_*(\gamma, \Lambda, \lambda)$ is sufficiently large. Now

$$\|u(t_1^*)\| = C_* e^{-\lambda t_1^*}(\|u_0\| + e^{-\Lambda t_1^*}\sqrt{(F_0, u_0)}) \leqq 9\, C_* e^{-\lambda t_2^*}(\|u_0\| + e^{-\Lambda t_2^*}\sqrt{(F_0, u_0)}) ,$$

so that

$$(C_* e^{-\lambda t_2^*}(\|u_0\| + e^{-\Lambda t_2^*}\sqrt{(F_0, u_0)}))^2 \leqq \|u(t_2^*) - u(t_1^*)\|^2 \leqq (t_2^* - t_1^*)\int_{t_1^*}^{t_2^*} \left\|\frac{du}{dt}\right\|^2 dt . \tag{10}$$

Choose $t_0^* \in [t_1^* - 2, t_1^* - 1]$ such that

$$(F_0, u(t_0^*)) \leqq \int_{t_1^*-2}^{t_1^*-1} (F_0, u)dt \tag{11a}$$

and choose $t_1 \in [t_0^*, t_1^*]$ such that

$$\left|\frac{du}{dt}(t_1)\right|^2 \leqq \int_{t_0^*}^{t_1^*} \left|\frac{du}{dt}\right|^2 dt . \tag{11b}$$

Then (10, 9, 7c, 11b, 4, 6b, 11a, 6c, 9) imply

$$(C_* e^{-\lambda t_2^*}(\|u_0\| + e^{-\Lambda t_2^*}\sqrt{(F_0, u_0)}))^2 \leqq \gamma \int_{t_1}^{t_2^*} \left\|\frac{du}{dt}\right\|^2 dt \leqq C\gamma \int_{t_0^*}^{t_2^*} \left|\frac{du}{dt}\right|^2 dt$$

$$\leqq C\gamma\left(\int_{t_0^*}^{t_2^*} |f|^2 dt + \Phi_1(u(t_0^*))\right)$$

$$\leqq C\gamma\left(\int_{t_0^*}^{t_2^*} \|u\|^2 dt + \|u(t_0^*)\|^2 + (F_0, u(t_0^*))\right)$$

$$\leq C\gamma(|u_0|^2 e^{-2\lambda t_0^*} + (10 C_* e^{-\lambda t_0^*}(\|u_0\| + e^{-\Lambda t_0^*}\sqrt{(F_0, u_0)}))^2$$
$$+ |u_0|^2 e^{-2\lambda(t_1^*-2)})$$
$$\leq C\gamma(C_* e^{-\lambda t_2^*}(\|u_0\| + e^{-\Lambda t_2^*}\sqrt{(F_0, u_0)}))^2 ,$$

where the constant C does not depend on γ and C_*. Hence choosing $\gamma < 1/C$ we get

$$\|u(t)\| \leq 10 C_* e^{-\lambda t}(\|u_0\| + e^{-\Lambda t}\sqrt{(F_0, u_0)}) \tag{12}$$

for any t (the constant C_* does not depend on T).

In order to prove the estimate for $\left|\frac{du}{dt}\right|$ let us prove the following inequality

$$(F_0, u(t)) \leq C e^{-2\lambda t}(\|u_0\|^2 + e^{-2\Lambda t}(F_0, u_0)) . \tag{13}$$

By using (4) we get

$$(F_0, u(t)) - \Phi_1(u(t_0)) \leq \Phi_1(u(t)) - \Phi_1(u(t_0))$$
$$\leq \int_{t_0}^{t} |f|^2 dt \leq C \int_{t_0}^{t} \|u\|^2 dt \leq C |u_0|^2 e^{-2\lambda t_0}$$

for any $t_0 < t$, thus

$$(F_0, u(t)) \leq (F_0, u(t_0)) + C e^{-2\lambda t_0}(\|u_0\|^2 + e^{-2\Lambda t_0}(F_0, u_0)) . \tag{14}$$

Now the inequality (13) is obvious for $t \leq 1$. If $t > 1$, then there exists $t_0 \in (t-1, t)$ such that

$$(F_0, u(t_0)) \leq \int_{t-1}^{t} (F, u(t)) dt \leq C |u_0|^2 e^{-2\lambda t}$$

(we have used (6c)), which, together with (14), proves (13).

By using (4, 6b, 12, 13) we obtain further

$$\int_{t_1}^{t_2} \left|\frac{du}{dt}\right|^2 dt \leq C\left(\int_{t_1}^{t_2} \|u\|^2 dt + \|u(t_1)\|^2 + (F_0, u(t_1))\right)$$
$$\leq C e^{-2\lambda t_1}(\|u_0\|^2 + e^{-2\Lambda t_1}(F_0, u_0)) . \tag{15}$$

Choose $t > 0$ such that $\frac{du}{dt}(t)$ exists and put $t_0 = \max(t/2, t-1)$. Then there exists $t_1 \in [t_0, t]$ such that

$$\left|\frac{du}{dt}(t_1)\right|^2 \leq \frac{1}{t-t_0} \int_{t_0}^{t} \left|\frac{du}{dt}\right|^2 dt \leq \frac{C e^{-2\lambda t_0}}{t-t_0} (\|u_0\|^2 + e^{-2\Lambda t_0}(F_0, u_0))$$
$$\leq \frac{C e^{-2\lambda t}}{\min(1, t)} (\|u_0\|^2 + e^{-2\Lambda t}(F_0, u_0))$$

by (15). According to (7a) we get

$$\left|\frac{du}{dt}(t)\right| \leq e^{C(t-t_1)} \left|\frac{du}{dt}(t_1)\right| \leq C\left(1 + \frac{1}{\sqrt{t}}\right) e^{-\lambda t}(\|u_0\| + e^{-\Lambda t}\sqrt{(F_0, u_0)}) . \tag{16}$$

Now let $g_0 \in H$, $(A_1 u_0 + N_1(u_0) + F_0 - g_0, v - u_0) \geqq 0$ for any $v \in K$. Then it follows from [2] (Proposition 5 used for the functional $\Phi(u) = \chi_K(u) + \Phi_1(u) - \Phi_1(u_0) + (g_0, u_0 - u)$, where χ_K is the indicatrix function of the set K)

$$\int_0^T \left|\frac{du}{dt}\right|^2 dt \leqq 4 \int_0^T |f + g_0|^2 dt \leqq C\left(T|g_0|^2 + \int_0^T \|u\|^2 dt\right),$$

which, together with (7a, 12), implies

$$\left|\frac{du}{dt}\right| \leqq C(|g_0| + \|u_0\| + \sqrt{(F_0, u_0)}) \quad \text{a.e. in } (0, \infty) . \tag{17}$$

By using (7c, 16, 17) we get $\frac{du}{dt} \in L^2((0, \infty), V)$.

Finally, let $N_1 \equiv 0$, $F_0 \in V$ and $(I + \mu A_1)^{-1}(K) \subset K$ for any $\mu > 0$. Since $\frac{du}{dt} \in L^2((0, \infty), V)$, we have $\frac{df}{dt} \in L^2((0, \infty), H)$. The results of [1, Lemme II.4] imply that $A_1 u(t) \in H$ for any $t \geqq 0$ and the function $u : [0, \infty) \to H$ is differentiable from the right everywhere. Moreover,

$$\left(\frac{d^+ u}{dt}(t) + A_1 u(t), A_1 u(t)\right) \leqq (f(t) - F_0, A_1 u(t)) \quad \text{for any} \quad t \geqq 0 . \tag{18}$$

It follows from (7a), (16) and (17) that

$$\left|\frac{d^+ u}{dt}(t)\right| \leqq C e^{-\lambda t}(\|u_0\| + |g_0| \max(1 - t, 0) + \sqrt{(F_0, u_0)}\, e^{-At})$$

By using (18) we get

$$\begin{aligned}
|A_1 u(t)|^2 &\leqq C\left(|f(t)|^2 + \left|\frac{d^+ u}{dt}(t)\right|^2\right) - 2(A_1 F_0, u(t)) \\
&\leqq C\left(\|u(t)\|^2 + \left|\frac{d^+ u}{dt}(t)\right|^2\right) - 2(A_1 F_0, u(t)) \\
&\leqq C e^{-2\lambda t}(\|u_0\|^2 + |A_1 u_0|^2 \max(1 - t, 0) + (F_0, u_0) e^{-2At}) \\
&\quad -2(A_1 F_0, u(t)) \\
&\leqq C(e^{-\lambda t}|A_1 u_0| + e^{-\lambda t/2}\|F_0\|)^2 .
\end{aligned}$$

It remains to prove the uniqueness of our solution. Let u_1, u_2 be two (weak) solutions of the inequality (2). It follows from the preceeding considerations that u_1, u_2 are also strong solutions, $\|u_i(t)\| < \varepsilon$ for any $t \geqq 0$, $i = 1, 2$,

$$\left(\frac{du_i}{dt} + Au_i + N(u_i) + F_0, v - u_i\right) \geqq 0 \quad \text{for any } v \in K \quad \text{a.e. in } (0, T) .$$

Choosing $v = u_{3-i}$ in this inequality and adding the resulting inequalities we get

$$\left(\frac{d(u_1 - u_2)}{dt} + A(u_1 - u_2) + N(u_1) - N(u_2), u_1 - u_2\right) \leqq 0 \quad \text{a.e. in } (0, T) ,$$

which implies (cf. the deriving of (7))

$$\frac{1}{2}\frac{d}{dt}|u_1-u_2|^2 \leqq C|u_1-u_2|^2 \quad \text{a.e. in } (0,T) .$$

Now $u_1(0)=u_2(0)$ implies $u_1 \equiv u_2$. □

Theorem 2. *Let, in addition to our general assumptions, K be a cone with its vertex at zero, $F_0=0$, $|N_2(v)| \leqq C\|v\|^2$ for any $v \in B$. Let $u_1 \in K$, $|u_1|=1$, be an eigenvector of the inequality (1) with an eigenvalue $\lambda_1<0$ (i.e. $(Au_1-\lambda_1 u_1, v-u_1) \geqq 0$ for any $v \in K$) and let*

$$(Ay,y) \geqq \lambda_1|y|^2 \quad \text{for any } y \in K-u_1 := \{v-u_1 ; v \in K\} . \tag{A1}$$

Then the zero solution of (2) is unstable in the toplology of H; more precisely, there exists $\varepsilon>0$ such that for any $\delta>0$ the solution u of (2) with $u_0=\delta u_1$ exists on $[0, T_\delta]$ and $|u(T_\delta)| \geqq \varepsilon$.

Proof. Similarly as in the proof of Theorem 1 we get the existence of a strong solution of (2) (with N_2 redefined outside B). Moreover, there exist ε_0, $C>0$ such that $\{u \in V; \|u\| \leqq \varepsilon_0\} \subset B$ and

$$\left.\begin{aligned} &\int_0^t \|u\|^2 dt \leqq C\left(\int_0^t |u|^2 dt + |u_0|^2\right) \\ &\|u(t)\|^2 \leqq C\left(\int_0^t |u|^2 dt + \|u_0\|^2\right) \end{aligned}\right\} \quad \text{for any } t \leqq T_{\max} , \tag{19}$$

where $T_{\max} := \inf\{t; \|u(t)\| \geqq \varepsilon_0\} > 0$ if $\|u_0\|$ is small enough (cf. the derivation of (6, 6d)).

Choose $\varepsilon<\varepsilon_0$, $\delta>0$ and let $u_0=\delta u_1$. By putting $v=0$ and $v=2u$ in (2) we get

$$\left(\frac{du}{dt}+Au+N(u), u\right)=0 ,$$

so that

$$\frac{d}{dt}|u| \leqq -\lambda_1|u|+C\|u\|^2 \quad \text{a.e. in } \{t \in (0, T_{\max}); u(t) \neq 0\} . \tag{20}$$

Choose $\beta \in (0,1)$ fixed and suppose $|u(s)| \leqq (1+\beta)|w(s)|$ for $s \leqq t \leqq T_{\max}$, where $w(s)=\delta u_1 e^{-\lambda_1 s}$ is the solution of the linearized inequality. By using (20, 19) we obtain

$$\begin{aligned} |u(t)| &\leqq e^{-\lambda_1 t}|u_0| + \int_0^t e^{-\lambda_1(t-s)} C\|u(s)\|^2 ds \\ &= \delta e^{-\lambda_1 t} + C\int_0^t \|u(s)\|^2 ds - \lambda_1 C\int_0^t \left(\int_0^s \|u(\tau)\|^2 d\tau\right) e^{-\lambda_1(t-s)} ds \\ &\leqq \delta e^{-\lambda_1 t} + C\left(\int_0^t |u(s)|^2 ds + |u_0|^2\right) + C\int_0^t e^{-\lambda_1(t-s)}\left(\int_0^s |u(\tau)|^2 d\tau + |u_0|^2\right) ds \\ &\leqq |w(t)| + C|w(t)|^2 < \left(1+\frac{\beta}{2}\right)|w(t)| \end{aligned}$$

whenever $|w(t)| \leqq \beta/C_0$, i.e. $t \leqq T_\delta := -\frac{1}{\lambda_1} \log \frac{\beta}{C_0 \delta}$ (where C_0 is some fixed constant). Therefore,

$$|u(t)| \leqq (1+\beta)|w(t)| \quad \text{for any } t \leqq \min(T_\delta, T_{\max}) \ . \tag{21}$$

By putting $y := u - w$ we get

$$\left(\frac{du}{dt} + Au + N(u), y\right) \leqq 0 \ ,$$

$$\left(\frac{dw}{dt} + Aw, y\right) \geqq 0 \ ,$$

hence

$$\left(\frac{dy}{dt} + Ay + N(u), y\right) \leqq 0 \ ,$$

which implies

$$\frac{d}{dt}|y| \leqq -\lambda_1 |y| + C\|u\|^2 \quad \text{a.e. in } \{t \in (0, T_{\max}); y(t) \neq 0\} \ ,$$

$$|y(t)| \leqq \int_0^t e^{-\lambda_1(t-s)} C\|u(s)\|^2 ds \leqq \beta |w(t)| \quad \text{for any } t \leqq \min(T_\delta, T_{\max}) \ , \tag{22}$$

$$|u(t)| \geqq (1-\beta)|w(t)| \quad \text{for any } t \leqq \min(T_\delta, T_{\max}) \ .$$

We have by (19) and (21)

$$\|u(t)\| \leqq C(\|u_0\| + \delta e^{-\lambda_1 t}) \quad \text{for} \quad t \leqq \min(T_\delta, T_{\max}) \ ,$$

thus

$$\|u(t)\| \leqq C(\|u_0\| + \delta e^{-\lambda_1 T_\delta}) = C\left(\|u_0\| + \frac{\beta}{C_0}\right) < \frac{\varepsilon_0}{2}$$

if $\beta < C_0\left(\frac{\varepsilon_0}{2C} - \|u_0\|\right)$ and $t \leqq \min(T_\delta, T_{\max})$, which implies $T_\delta \leqq T_{\max}$. By (22)

$$|u(T_\delta)| \geqq (1-\beta)|w(T_\delta)| \geqq \frac{(1-\beta)\beta}{C_0} \ ,$$

which proves our assertion. □

Remark 1. Theorem 2 holds also if we replace the assumption $|N_2(v)| \leqq C\|v\|^2$ by the assumption $\|N_2(v)\|_{V'} \leqq C\|v\| \cdot |v|$.

Examples

Example 1. Let

$$V = H = \mathbb{R}^2, \ K = \left\{\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \in \mathbb{R}^2; \ u_2 \geqq u_1 \geqq 0\right\} \ ,$$

$$A = \begin{pmatrix} -3 & 0 \\ 5 & -1 \end{pmatrix}, \quad N\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} -(u_2)^2 \\ 0 \end{pmatrix}, \quad F_0 = 0.$$

Then $\sigma(A)=\{-3,\ -1\}$, $\sigma_K(A)=\{-1,1/2\}$, nevertheless the zero solution is stable.

Proof. Let us choose $\varepsilon>0$ ($\varepsilon\ll 1$). We shall show that the solution $u(t)$ of (0) with $u_0=\begin{pmatrix}0\\ \delta\end{pmatrix}$, δ sufficiently small, satisfies $\|u(t)\|\leqq\varepsilon$ for any $t\geqq 0$, which, together with the geometry of K and of the trajectory of $u(t)$, implies our assertion (see Fig. 1).

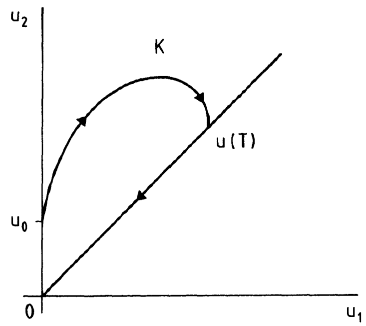


Fig. 1

To prove this let us study the solution $v(t)$ of the corresponding equation

$$\begin{aligned}\dot v_1&=3v_1+v_2^2\\ \dot v_2&=-5v_1+v_2\\ v_1(0)&=0\ ,\qquad v_2(0)=\delta\ ,\end{aligned}\tag{23}$$

where $\begin{pmatrix}\dot v_1\\ \dot v_2\end{pmatrix}=\dfrac{dv}{dt}$.

Since $\dot v_1(t)>0$, $v_1(t)>0$ for any $t>0$, we have $v\equiv u$ on $[0,T]$, where $T=\inf\{t;v_1(t)\geqq v_2(t)\}$. Thus it is sufficient to prove $T<\infty$ and $\|v(t)\|\leqq\varepsilon$ for $t\leqq T$, since $u(t)$ decays exponentially to zero for $t>T$.

Since $v_1\geqq 0$, we have $\dot v_2\leqq v_2$, hence

$$v_2(t)\leqq\delta e^t\ .\tag{24}$$

Now

$$v_1(t)=\int_0^t e^{3(t-s)}v_2^2(s)ds\leqq e^{3t}\delta^2\int_0^t e^{-s}ds\leqq e^{3t}\delta^2\leqq\frac{\delta}{5}$$

if $t\leqq 1$ and δ is sufficiently small. Consequently, $\dot v_2(t)\geqq 0$ and $v_2(t)\geqq\delta$ for $t\leqq 1$, which implies

$$v_1(1)\geqq\int_0^1 v_2^2(s)ds\geqq\delta^2\ .$$

Therefore,

$$v_1(t)\geqq\delta^2 e^{3(t-1)}\qquad\text{for any}\quad t\geqq 1\ .\tag{25}$$

Now (24) and (25) give us the following estimate for T

$$T\leqq T^*\ ,\qquad\text{where}\quad\delta^2 e^{3(T^*-1)}=\delta e^{T^*}\ .$$

Finally,

$$v_1(t) \leqq v_2(t) \leqq \delta e^t \leqq \delta e^T \leqq \delta e^{T^*} = e^{3/2} \sqrt{\delta}$$

for any $t \leqq T$, which proves our assertion. □

Remark 2. In [3, 5] there are given some general assumptions, under which stationary solutions of certain reaction-diffusion systems loose their "linearized stability" when we add unilateral conditions to the system. In an abstract setting we have $\operatorname{Re}\sigma(A) > 0$, $\sigma_K(A) \cap \{\lambda; \lambda < 0\} \neq \emptyset$ and, moreover, the results of [3, 5] imply that there exists $\lambda \in \sigma_K(A)$, $\lambda < 0$, which is also a bifurcation point of the stationary "nonlinear" inequality

$$(Au + N(u) - \lambda u, v - u) \geqq 0 \qquad \forall v \in K \tag{26}$$

(for any "suitable" N). Example 1 is *not* counterexample for the linearization principle in this case, since in this example

(i) we do not have $\operatorname{Re}\sigma(A) > 0$

(ii) $\lambda = -1$ is not a bifurcation point for the inequality (26).

Example 2. Let $V = H = \mathbb{R}^3$, $K = \{u \in \mathbb{R}^3; u_3 = 0\}$, $F_0 = 0$, $N \equiv 0$,

$$A = \begin{pmatrix} -1 & -2 & 16 \\ 2 & -1 & 0 \\ -2 & 0 & 9 \end{pmatrix}.$$

Then $\sigma(A) = \{1, 3 \pm 2i\}$ and $\sigma_K(A) = \emptyset$, since any $\lambda \in \sigma_K(A)$ is an eigenvalue of the operator $B := PA/K$ (where $P: \mathbb{R}^3 \to K$ is the orthogonal projection) and $\sigma(B) = \{-1 \pm 2i\}$. The inequality (2) is equivalent to the equation

$$\frac{du}{dt} + Bu = 0,$$

hence the zero solution is unstable.

Example 3. Let $\Omega = (0, \pi) \subset \mathbb{R}^1$, $V = H_0^1(\Omega)$, $H = L^2(\Omega)$, $F_0 = 0$, $Au = -u'' + \lambda u$, $K = \{u \in V; u(\pi/2) \leqq 0,\ u(2\pi/3) \geqq 0\}$, $N(u) = f(u)$, where $f \in C^1(\mathbb{R}, \mathbb{R})$, $f(0) = f'(0) = 0$.

Then $u_0 = 0$ is a stable solution of the equation

$$\frac{du}{dt} + Au + N(u) = 0$$

$$u(0) = u_0$$

provided $\lambda > -1$ and it is unstable if $\lambda < -1$. Similarly, $u_0 = 0$ is a stable solution of the inequality (2) if $\lambda > -9/4$ and it is unstable if $\lambda < -9/4$.

The stability result follows from Theorem 1, the instability result from the proof of Theorem 2. Note that the assumption (A1) is not satisfied, nevertheless, by putting $u_0 = \delta u_1$, where $u_1(x) = -\sin(3x/2)$ for $x \leqq 2\pi/3$ and $u_1(x) = 0$ for $x \geqq 2\pi/3$, and by using the notation from the proof of Theorem 2 we get $(Ay, y) \geqq \lambda_1 |y|^2$ for $y = u - w$, since $u(t)(2\pi/3) = 0$ for $t \leqq T_\delta$.

Acknowledgement. The author wishes to thank Professor H. Amann for the motivation of this work and for the hospitality during the author's stay at Universität Zürich.

References

1. Brézis, H.: Problèmes unilatéraux. J. Math. Pures Appl. **51**, 1–168 (1972)
2. Brézis, H.: Propriétés régularisantes de certains semi-groupes non linéaires. Isr. J. Math. **9**, 513–534 (1971)
3. Drábek, P., Kučera, M.: Eigenvalues of inequalities of reaction-diffusion type and destabilizing effect of unilateral conditions. Czech. Math. J. **36**, 116–130 (1986)
4. Kielhöfer, H.: Stability and semilinear evolution equations in Hilbert space. Arch. Ration. Mech. Anal. **57**, 150–165 (1974)
5. Quittner, P.: Bifurcation points and eigenvalues of inequalities of reaction-diffusion type. J. Reine Angew. Math. **380**, 1–13 (1987)

Received June 9, 1988

Math. Ann. 283, 271–283 (1989)

Mathematische Annalen

Symmetries of Möbius Ladders

Erica Flapan*

Department of Mathematics, Pomona College, Claremont, CA 91711, USA

Introduction

Chemistry has recently motivated the study of graphs embedded in $\mathbb{R}^3$, and of their symmetries as an extension of knot theory. We are interested in the following question: Given a graph G embedded in $\mathbb{R}^3$ or $S^3 = \mathbb{R}^3 \cup \infty$, what can be said about its symmetries just from the topology of the graph itself? More precisely, we shall let $\mathrm{Sym}(G)$ denote the group of homeomorphisms of G, up to isotopy. If G is embedded in a manifold M, then $\mathrm{Sym}(M, G)$ is the group of diffeomorphisms of M which leave G invariant, up to isotopy respecting G. We are interested in the general question of how an element of $\mathrm{Sym}(G)$ can be represented by an element of $\mathrm{Sym}(S^3, G)$, for some embedding of G in S^3. Of course, not every graph G can be embedded in such a way that a given element of $\mathrm{Sym}(G)$ can be represented by some element of $\mathrm{Sym}(S^3, G)$. In Sect. 1, we will provide an example of a graph G and a particular element of $\mathrm{Sym}(G)$ such that, no matter what the embedding of G in S^3, that element cannot be represented by an element of $\mathrm{Sym}(S^3, G)$.

In the case where each element of $\mathrm{Sym}(G)$ can be represented by an element of $\mathrm{Sym}(S^3, G)$, we are interested in which elements of $\mathrm{Sym}(G)$ can be represented by periodic and/or orientation reversing elements of $\mathrm{Sym}(S^3, G)$. Since not all periodic elements of $\mathrm{Sym}(S^3, G)$ restrict to periodic elements of $\mathrm{Sym}(\mathbb{R}^3, G)$, we consider separately the question of which elements of $\mathrm{Sym}(G)$ can be represented by periodic elements of $\mathrm{Sym}(\mathbb{R}^3, G)$. In Sects. 2 and 3, we completely answer these questions for one class of graphs. Understanding the symmetries of this particular class of graphs also has some applications in chemistry, which we explain below.

It is often important in the field of chemistry to determine whether a molecule is distinct from its mirror image. A molecule which can convert itself to its mirror image is said to be *chemically achiral*, whereas one which cannot is *chemically chiral*. The existence of such a molecular deformation depends on a variety of physical conditions, and thus cannot be completely characterized mathematically. Instead, we abstract the molecule as a graph in space, and ask whether this embedded graph can be deformed in space to its mirror image. A graph which can

* Partially supported by a grant from the ONR

be deformed to its mirror image is *topologically achiral*, while one which cannot be deformed to its mirror image is *topologically chiral* [Wa]. (The property of topological achirality for graphs is analogous to the property of amphicheirality for knots.) A molecule whose associated graph is topologically chiral will necessarily be chemically chiral, hence this concept is of some use to chemists.

One particular molecular graph which is of interest is the "molecular Möbius ladder", which was first synthesized by Walba et al. [WRH]. This is a molecule shaped like a ladder with three rungs which was made to join itself end-to-end with one half twist (see Fig. 1). The sides of the ladder represent a molecular chain while the rungs represent double bonds; hence in the associated molecular graph we distinguish between the edges making up the sides and those making up the rungs. The synthesis of this molecule was a significant achievement in chemistry because of its topologically interesting molecular structure. Walba had conjectured that this molecule was chemically chiral [Wa], however chemical achirality could not be completely ruled out until Simon [Si] proved that its associated embedded molecular graph was topologically chiral.

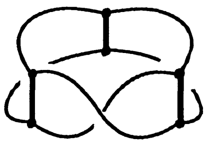


Fig. 1

More generally, let M_n denote the graph illustrated in Fig. 2, with $n \geqq 3$, where the rungs of the ladder are $\alpha_1, \ldots, \alpha_n$ and the sides of the ladder together form the loop K. Observe that the graph M_3 is just the bipartite graph (3, 3) which is one of Kuratowski's non-planar graphs. For all $n > 3$, M_n contains this non-planar graph and hence is itself non-planar.

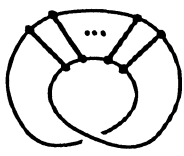


Fig. 2

What Simon showed is that for the embedding of M_n illustrated in Fig. 2, for any $n \geqq 3$, there is no orientation reversing diffeomorphism h of S^3 with $h(M_n) = M_n$ and $h(K) = K$. The chemical motivation for the requirement that $h(K) = K$ is that the loop K represents a molecular chain, which is chemically different from the rungs which represent molecular bonds. We note however, that Simon [Si] has also shown that if $n \geqq 4$ then every automorphism of M_n leaves K setwise invariant. Thus if we restrict our attention to Möbius ladders with at least four rungs, then $h(K) = K$ will follow whenever $h(M_n) = M_n$. So only in the case where $n = 3$ does the hypothesis that $h(K) = K$ make any difference.

Simon's results naturally led to the question of topological chirality for other embeddings of the graph M_n. That is, is it possible to reembed M_n in S^3 in such a way that there is an orientation reversing diffeomorphism h of S^3 with $h(M_n) = M_n$ and $h(K) = K$? For $n \geqq 4$, this is just the question of whether there is any element of $\mathrm{Sym}(M_n)$ which can be represented by an orientation reversing element of

$\mathrm{Sym}(S^3, M_n)$. We answer this particular question by showing that, for any $n \geqq 3$ which is odd, no matter how M_n is embedded there is no such h. On the other hand, for any n which is even there is an embedding of M_n in S^3 and an orientation reversing diffeomorphism h of S^3 with $h(M_n) = M_n$ and $h(K) = K$. In addition, we provide a general analysis of how elements of $\mathrm{Sym}(M_n)$ can be represented by elements of $\mathrm{Sym}(S^3, M_n)$ and $\mathrm{Sym}(\mathbb{R}^3, M_n)$. In particular, the group $\mathrm{Sym}(M_n)$ is generated by those rotations and reflections of K which leave the set of rungs invariant. Each element of $\mathrm{Sym}(M_n)$ can be realized by a periodic orientation preserving element of $\mathrm{Sym}(S^3, M_n)$, as will be illustrated in Figs. 5, 9, and 10. Theorem 4 together with Figs. 6 and 3, analyzes which symmetries of M_n can be realized by periodic orientation preserving elements of $\mathrm{Sym}(\mathbb{R}^3, M_n)$. The question of which elements of $\mathrm{Sym}(M_n)$ can be represented by orientation reversing elements of $\mathrm{Sym}(S^3, M_n)$ is answered by Theorem 2 together with Fig. 4. For periodic orientation reversing elements of $\mathrm{Sym}(S^3, M_n)$ and $\mathrm{Sym}(\mathbb{R}^3, M_n)$ the question is answered by Theorem 3 and Lemma 2. Finally, in Sect. 4, we make some observations about topological chirality for a slightly larger class of graphs which includes the graphs M_n.

1. An Element of Sym(G) which Cannot be Represented by an Element of Sym(S^3, G)

We are interested in whether there is a graph G such that some element of $\mathrm{Sym}(G)$ cannot be realized by any element of $\mathrm{Sym}(S^3, G)$, no matter how G is embedded in S^3. The following theorem shows that K_6 (the complete graph on six vertices) is an example of such a graph.

Theorem 1. *For any embedding of K_6 in S^3, and any labelling of the vertices of K_6 by the numbers one through six, there is no element of* $\mathrm{Sym}(S^3, K_6)$ *which induces the permutation* (1234) *on the vertices of K_6.*

Proof. Choose some labelling of the vertices of K_6 by the numbers one through six. Any unordered set of three such numbers will determine a loop consisting of the three vertices with those numbers together with the edges between them. However, since K_6 only has six vertices, this set also determines a disjoint complementary loop. So we shall let each set of three such numbers represent a pair of disjoint loops. Since each pair of loops can be represented in two different ways by complementary sets of numbers, there are ten pairs of such loops in K_6.

We consider the orbits of these pairs of loops under the permutation (1234). Being careful not to list the same pair in two different ways, we see that the collection of orbits of loop pairs is $\langle 123, 234, 341, 412 \rangle$, $\langle 125, 235, 345, 415 \rangle$, and $\langle 135, 245 \rangle$. The observation we wish to make here is that the set of all loop pairs is partitioned into orbits which each contain an even number of elements.

Now suppose that the graph K_6 is embedded in S^3 in such a way that there is a diffeomorphism $h: S^3 \to S^3$ with $h(K_6) = K_6$ and h induces the permutation (1234) on the vertices of K_6. For each pair of disjoint loops A and B in K_6, let $\omega(A, B)$ denote the mod 2 linking number of A and B in S^3. Since h is a diffeomorphism, $\omega(A, B) = \omega(h(A), h(B))$. Thus all the pairs in a given orbit will have the same mod 2 linking number. Define $\lambda = \sum \omega(A, B)$, where the summation is in $\mathbb{Z}_2$ over all pairs of disjoint loops in K_6. Since every orbit has an even number of pairs in it, $\lambda=0$. However, Conway and Gordon have proved in [CG] that for any embedding of K_6 in S^3, $\lambda=1$. Thus there could not have been such a diffeomorphism h. □

2. Orientation Reversing Symmetries of Möbius Ladders

Any graph which is homeomorphic, as a 1-complex, to the graph in Fig. 2, is a *Möbius ladder* as defined originally by Harary and Guy [HG]. More formally,

Definition. For $n \geqq 3$ we define a *Möbius ladder* M_n to be any graph which is homeomorphic to a $2n$-gon K together with disjoint chords $\alpha_1, \ldots, \alpha_n$ joining opposite pairs of vertices. We will refer to K as the *loop* of M_n and the chords $\alpha_1, \ldots, \alpha_n$ as the *rungs* of M_n.

Lemma 1. *Let M_n be a Möbius ladder which is embedded in S^3 with loop K and rungs $\alpha_1, \ldots, \alpha_n$. Suppose $h: S^3 \to S^3$ is an orientation reversing diffeomorphism with $h(M_n)=M_n$ and $h(K)=K$. Then there are at most two rungs α_i such that $h(\alpha_i)=\alpha_i$.*

Proof. Let X denote the double branched cover of S^3, branched over K. Let $K_1, \ldots, K_n$ be the preimages of $\alpha_1, \ldots, \alpha_n$ respectively. For each i, let k_i denote the simple closed curve consisting of α_i together with some component of $K-\alpha_i$. Let F_i be a Seifert surface for k_i; *and let* S_i denote the preimage of F_i in X. Observe that for each i, K_i is the boundary of S_i.

It is not hard to show that $H_1(X)$ is finite (see [Ro]); and so, by Poincaré duality, $H_2(X)$ is trivial. Thus if S and S' are both surfaces bounded by K_i then the algebraic intersection number of K_j with S must equal the algebraic intersection number of K_j with S'. Hence we can define $\mathrm{lk}(K_i, K_j)$ as the algebraic intersection of K_j with the surface S_i.

Let p be the algebraic intersection number of $\mathrm{Int}(\alpha_j)$ with $F_i - K$. Since α_j meets $F_i \cap K$ at one point, then $\mathrm{lk}(K_i, K_j) = \pm(2p+1)$, depending on orientations. In particular, for all $i \neq j$, we have $\mathrm{lk}(K_i, K_j) \neq 0$. Now suppose $h(\alpha_i)=\alpha_i$ for $i=1,2,3$. Since S^3 has a unique double branched cover over K, any homeomorphism of (S^3, K) will lift to X. So let $g: X \to X$ be one lift of h. Then g is orientation reversing and $g(K_i)=K_i$ for $i=1,2,3$. Give K_1, K_2, K_3 orientations, then suppose that g preserves the orientation of K_1. Since g is orientation reversing and $\mathrm{lk}(K_1, K_2) \neq 0$ and $\mathrm{lk}(K_1, K_3) \neq 0$, it must be that g reverses the orientations of both K_2 and K_3. But this is impossible because $\mathrm{lk}(K_2, K_3) \neq 0$. We obtain a similar contradiction if we suppose that g reverses the orientation of K_1. Thus we could not have had $h(\alpha_i)=\alpha_i$ for $i=1,2,3$. □

Theorem 2. *Let M_n be a Möbius ladder which is embedded in S^3 with loop K, where n is an odd number. Then there is no diffeomorphism $h: S^3 \to S^3$ which is orientation reversing with $h(M_n)=M_n$ and $h(K)=K$.*

Proof. Suppose there were such an h. First we shall consider the case where h reverses the orientation of K. In this case h performs a permutation of order two on the rungs. Since the number of rungs, n, is odd there must be some rung which h maps to itself. By the definition of a Möbius ladder $n \geqq 3$, so we can assume the rungs have been labelled in such a way that $h(\alpha_1)=\alpha_1$, $h(\alpha_2)=\alpha_n$, and $h(\alpha_n)=\alpha_2$. As in Lemma 1, let X be the double branched cover of S^3 with branch set K; and let g be one lift of h and, for each i, let K_i be the preimage of α_i. Then $g(K_1)=K_1$, $g(K_2)=K_n$, and $g(K_n)=K_2$. As in the proof of Lemma 1, the algebraic linking number is well defined and $\mathrm{lk}(K_1,K_2)\neq 0$, $\mathrm{lk}(K_1,K_n)\neq 0$, and $\mathrm{lk}(K_2,K_n)\neq 0$. Since g is orientation reversing and $\mathrm{lk}(K_2,K_n)\neq 0$, we can assume without loss of generality that $g(K_2)=-K_n$ and $g(K_n)=+K_2$. Now suppose that $g(K_1)=+K_1$. Then $g(K_2)=-K_n$ implies that $\mathrm{lk}(K_1,K_2)=\mathrm{lk}(K_1,K_n)$. But $g(K_n)=+K_2$ implies that $\mathrm{lk}(K_1,K_n)=-\mathrm{lk}(K_1,K_2)$. Since these linking numbers are non-zero, this is impossible. Supposing that $g(K_1)=-K_1$ yields a similar contradiction. Thus no such h could exist with $h(K)=-K$.

Now we consider the case where h preserves the orientation of K. We can assume that the rungs $\alpha_1,\ldots,\alpha_n$ have been labelled consecutively. In this case there exists a number p, such that h rotates each α_i to α_{i+p}. Since n is odd, there is some odd number q, such that $h^q(\alpha_i)=\alpha_i$ for all $i=1,\ldots,n$. But h^q is orientation reversing and $n\geqq 3$, hence this contradicts Lemma 1. □

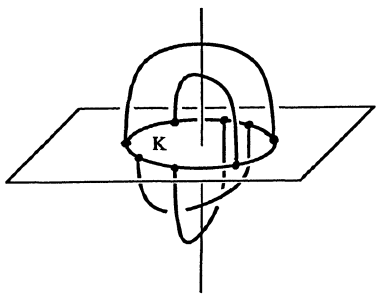


Fig. 3

In Fig. 3, we illustrate an example of an embedded Möbius ladder M_4 with four rungs, which has an orientation reversing diffeomorphism $h\colon S^3\to S^3$ with $h(M_4)=M_4$ and $h(K)=K$. In this example the map h can be seen as the composition of a reflection through the plane containing the loop K followed by a rotation of 90° about an axis perpendicular to that plane. Thus h preserves the orientation of K, and h has order equal to four. For any n which is even we can draw a similar example of a Möbius ladder M_n, with loop K lying in a plane, such that there is an orientation reversing diffeomorphism $h\colon S^3\to S^3$ which is the composition of a reflection though the plane containing K followed by a rotation of 90° about an axis perpendicular to that plane with $h(M_n)=M_n$ and $h(K)=K$. Thus, for any n which is even there is a Möbius ladder M_n and an orientation reversing diffeomorphism $h\colon S^3\to S^3$ with $h(M_n)=M_n$, $h(K)=+K$, and h is of order four.

We will also illustrate a Möbius ladder M_n and an orientation reversing diffeomorphism $h\colon S^3\to S^3$ with $h(M_n)=M_n$ and $h(K)=-K$; however, in order to explain the way we have drawn our example, we first prove the following lemma.

Lemma 2. *Let M_n be a Möbius ladder which is embedded in $\mathbb{R}^3$. Let $h\colon\mathbb{R}^3\to\mathbb{R}^3$ be an orientation reversing diffeomorphism with $h(M_n)=M_n$ and $h(K)=K$. Then h is not of finite order.*

Proof. Suppose h is of finite order. Then $h|K$ is also of finite order and h reverses the orientation of K. Thus, h fixes two antipodal points, x and y, of K. The points x and y separate K into arcs A and B, with $h(A)=B$ and $h(B)=A$. Since h leaves the collection of rungs setwise invariant, it follows from the definition of a Möbius ladder that each rung α_i has one endpoint in A and the other endpoint in B. By Smith Theory [Sm], the fixed point set of an orientation reversing finite order diffeomorphism of $\mathbb{R}^3$ is either one point or a plane. In this case, the fixed point set of h must contain the points x and y, so the fixed point set of h must be a plane P. Now $P \cap K = \{x, y\}$, hence the arc A is contained in one component of $\mathbb{R}^3 - P$, and the arc B is contained in the other component of $\mathbb{R}^3 - P$. Since P separates $\mathbb{R}^3$, each rung α_i must intersect P. But this means that h fixes a point of each rung α_i, which implies that $h(\alpha_i)=\alpha_i$ for all i. By the definition of Möbius ladder $n \geqq 3$, hence this contradicts Lemma 1. □

Suppose that M_n is an example of a Möbius ladder which is embedded in S^3 (i.e. $\mathbb{R}^3 \cup \infty$) in such a way that there is an orientation reversing diffeomorphism $h: S^3 \to S^3$ with $h(M_n)=M_n$ and $h(K)=-K$, and without loss of generality, $h(\infty)=\infty$. Then by the proof of Lemma 2, one of the fixed points of $h|K$ must actually be the point at infinity. Thus if we want to illustrate an example where $h(K)=-K$ and h can be seen easily as the composition of a reflection and a rotation, then we must draw M_n so that the point at infinity is on K.

In Fig. 4, we illustrate a Möbius ladder M_6, which is embedded in S^3 in such a way that the ends of the loop K meet at the point at infinity. We define an orientation reversing diffeomorphism $h: S^3 \to S^3$ which reflects S^3 through the origin. That is, h is the composition of a reflection through the plane drawn in the Fig. 4, followed by a rotation of 180° about K. Thus, $h(M_6)=M_6$, and $h(K)=-K$, and h has order two. For any n which is even we can embed M_n in S^3 in a similar way, so that there is an orientation reversing diffeomorphism $h: S^3 \to S^3$ with $h(M_n)=M_n$, and $h(K)=-K$, and h has order two. By moving K slightly at the point at infinity we obtain an embedding of M_n in $\mathbb{R}^3$, and an orientation reversing diffeomorphism $g: S^3 \to S^3$ with $g(M_n)=M_n$, and $g(K)=-K$. However, by Lemma 2 no such g could be of finite order.

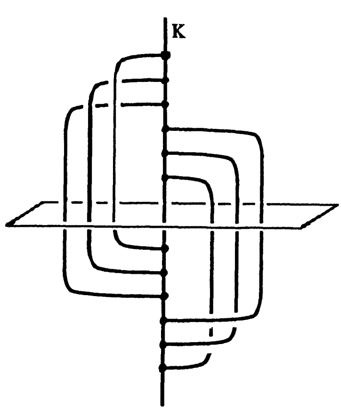


Fig. 4

These examples suggest that we might be able to say something more about finite order diffeomorphisms of Möbius ladders which are embedded in S^3. By Smith theory [Sm] we know that any finite order orientation reversing diffeomorphism of S^3 has either two points or a 2-sphere as its fixed point set; and any finite order orientation preserving diffeomorphism has either the empty set or a simple closed curve as its fixed point set. Let M_n be a Möbius ladder which is embedded in S^3.

Lemma 3. *Let $h: S^3 \to S^3$ be an orientation reversing diffeomorphism which is of finite order, with $h(M_n) = M_n$ and $h(K) = K$. Then the fixed point set of h consists of two points.*

Proof. By Smith Theory [Sm] if the fixed point set of h is not two points then it is a 2-sphere F. So we can pick the point x at infinity to be any fixed point of h which is not on M_n. Then h restricts to an orientation reversing periodic diffeomorphism $g: \mathbb{R}^3 \to \mathbb{R}^3$ with $g(M_n) = M_n$ and $g(K) = K$, and with fixed point set a plane P. Since P separates $\mathbb{R}^3$ into two components and g is orientation reversing, g must switch the two components of $\mathbb{R}^3 - P$. But since $g(K) = K$ the intersection of P and K must be non-empty. Hence $g(K) = -K$; which contradicts Lemma 2. Thus the fixed point set of h could not be a 2-sphere. □

Now we prove Theorem 3, which shows that Figs. 3 and 4 provide examples of the only possible orders for orientation reversing finite order diffeomorphisms of Möbius ladders embedded in S^3.

Theorem 3. *Let M_n be a Möbius ladder which is embedded in S^3 with loop K. Suppose that $h: S^3 \to S^3$ is an orientation reversing diffeomorphism with $h(M_n) = M_n$ and $h(K) = K$, and the order of h is some finite number p. If $h(K) = -K$ then $p = 2$, and if $h(K) = +K$ then $p = 4$.*

Proof. First suppose that $h(K) = -K$. Then h^2 fixes K pointwise. Hence h^2 also fixes each rung α_i pointwise. But h^2 is orientation preserving, so by Smith theory [Sm], if h^2 is not the identity map then the fixed point set of h^2 is either the empty set or a simple closed curve. Hence h^2 is the identity map.

Now suppose that $h(K) = +K$. Then h performs a cyclic permutation of the rungs α_i. By Lemma 3, the fixed point set of h is two points. Suppose $p = 2$, then $h(\alpha_i) = \alpha_i$ for all i. Hence h fixes a point of each α_i. Since $n \geqq 3$, this is a contradiction. Thus $p \neq 2$, so the map $g = h^2$ is not the identity. Since the fixed point set of g contains the fixed point set of h, by Smith Theory [Sm] the fixed point set of g must be a simple closed curve J. The order of h must be even, since h is orientation reversing. Thus $r = p/2$ is an integer, and the map $f = h^r$ has order two. Since $f(K) = +K$ we can use the same argument as above to show that f cannot be orientation reversing. So f is orientation preserving, and hence r must be even. Thus, in fact, $f = g^{p/4}$. This implies that the fixed point set of f contains the fixed point set of g. Since f is orientation preserving the fixed point set of f cannot contain more than a simple closed curve. Thus the fixed point set of f is J, so J intersects each rung α_i. Now this implies that g fixes a point of each α_i. Hence $g(\alpha_i) = \alpha_i$ for all i, and so the order of g is two. Therefore the order of h is four. □

Observe that, up to conjugacy, there is only one order 4 element of $\mathrm{Sym}(M_n)$ which preserves the orientation of K. Also, for n odd, there is only one conjugacy class of $\mathrm{Sym}(M_n)$ of order two which reverses the orientation of K. For n even, there are two conjugacy classes of $\mathrm{Sym}(M_n)$ of order two which reverse the orientation of K, one of which leaves no rung invariant and one of which leaves two rungs invariant. However, this latter symmetry actually fixes one of the invariant rungs pointwise. Hence if this symmetry were realizable by a finite order diffeomorphism $h: S^3 \to S^3$, it would contradict Lemma 3. Thus Theorem 3 completes our analysis of how elements of $\mathrm{Sym}(M_n)$ can be represented by orientation reversing elements of $\mathrm{Sym}(S^3, M_n)$ and $\mathrm{Sym}(\mathbb{R}^3, M_n)$.

3. Orientation Preserving Symmetries of Möbius Ladders

We consider how elements of $\mathrm{Sym}(M_n)$ can be represented by orientation preserving elements of $\mathrm{Sym}(S^3, M_n)$ and $\mathrm{Sym}(\mathbb{R}^3, M_n)$. All those elements of $\mathrm{Sym}(M_n)$ which are induced by rotations of K can be represented by periodic orientation preserving elements of $\mathrm{Sym}(S^3, M_n)$. In Fig. 5 we illustrate such an example with $n=3$. In order to facilitate our explanation of the diffeomorphism h, we have drawn a central axis A which is perpendicular to the plane containing the loop K. The action of h can be seen as the composition of a rotation by 120° about the axis A followed by a rotation by 120° about the loop K. An analogous example works for any n.

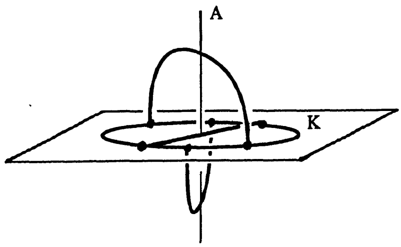


Fig. 5

In contrast, for $\mathbb{R}^3$ we have the following theorem.

Theorem 4. *Let M_n be a Möbius ladder which is embedded in $\mathbb{R}^3$ with loop K. Let $h: \mathbb{R}^3 \to \mathbb{R}^3$ be an orientation preserving finite order diffeomorphism with $h(M_n) = M_n$, and $h(K) = K$. If the order of h is even, then the order of h is two.*

Proof. Let r be the order of h and let J be the fixed point set of h. By Smith Theory [Sm], the fixed point set of an orientation preserving finite order diffeomorphism of $\mathbb{R}^3$ is a line. Thus for all $i<r$ the fixed point set of h^i is also J. We consider the cases where J intersects K and where J is disjoint from K separately. First, suppose J intersects K. Then the intersection of J and K is two points, and h reverses the orientation of K. In this case, h^2 fixes K pointwise in addition to J. Therefore h^2 is the identity.

Now suppose that J is disjoint from K. Then h preserves the orientation of K, and hence cyclically permutes the rungs α_i. By hypothesis r is assumed to be even, so $p=r/2$ is an integer. Now, h^p is a rotation of K of order two, and hence $h^p(\alpha_i) = \alpha_i$ for all i. Thus h^p fixes a point of each rung α_i. So J intersects every rung. But this implies that h fixes a point of each rung, and so h^2 fixes every rung pointwise. Thus again, h^2 fixes K pointwise in addition to J. So, in fact, $r=2$. □

For every n, up to conjugacy, $\mathrm{Sym}(M_n)$ has precisely one element of order two which respects the orientation of K. We see as follows that for every n there is an embedding of M_n in $\mathbb{R}^3$ such that this element can be realized by an orientation preserving element of $\mathrm{Sym}(\mathbb{R}^3, M_n)$ of order two. Let M_3 be the Möbius ladder illustrated in Fig. 5. Let $g: \mathbb{R}^3 \to \mathbb{R}^3$ be a rotation by 180° about the central axis which is perpendicular to the plane containing the loop K. Then $g(M_3) = M_3$, $g(K) = +K$, the fixed point set of g is the central axis, and the order of g is two. Observe that for any n we can construct an analogous example.

Now we provide an example to show that we can have any odd order orientation preserving symmetry of a Möbius ladder in $\mathbb{R}^3$. Figure 6 illustrates a Möbius ladder M_3 which is invariant under a rotation of order three about a central axis. Observe that the same rotation will work for any number of rungs which is a multiple of three. Also, for any odd number r, let K be the boundary of a band with r half twists, then for any $n > 0$, we can construct an analogous Möbius ladder M_{nr} with loop K, and M_{nr} will be invariant under a rotation of order r.

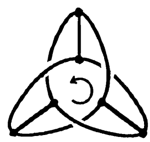


Fig. 6

Note that in an example which is constructed as in Fig. 6, the loop K will be knotted. This will not always be the case for every embedding of a Möbius ladder in $\mathbb{R}^3$ with an odd order symmetry. For example, the embedding of M_3 illustrated in Fig. 7 is invariant under a rotation of order three about a central axis. However, in Theorem 5 we will prove that, for n odd, if M_n is embedded in $\mathbb{R}^3$ so that it has an odd order symmetry, then at least one of three "special" cycles in M_n is knotted. We begin by introducing two special cycles, in addition to K, which are contained in M_n.

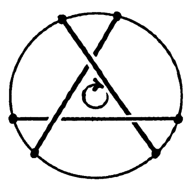


Fig. 7

Suppose that n is odd. Then K contains $2n$ edges, and its vertices occur at the endpoints of the rungs $\alpha_1, \ldots, \alpha_n$. Label the edges of K consecutively by $a_1, b_1, a_2, b_2, \ldots, a_n, b_n$. Let $R = \alpha_1 \cup \ldots \cup \alpha_n \cup a_1 \cup \ldots \cup a_n$ and let $B = \alpha_1 \cup \ldots \cup \alpha_n \cup b_1 \cup \ldots \cup b_n$. Figure 8 illustrates the R and B for the Möbius ladder in Fig. 7.

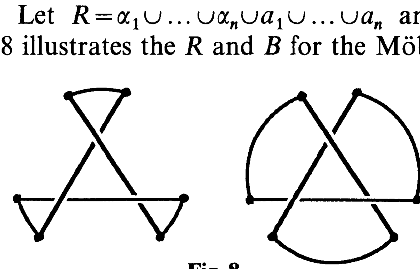


Fig. 8

By our construction it is clear that R is a collection of one or more simple closed curves. We see that R is precisely one simple closed curve as follows. Observe that each edge of K is connected by a rung to the $n+1^{\text{th}}$ subsequent edge. In particular, since n is odd, $n=2k+1$. Hence, $n+1=2(k+1)$. Now, since the edges of K are labelled alternately by a's and b's, the $n+1^{\text{th}}$ subsequent edge after a_i is a_{i+k+1} (counting the indices modulo n). Therefore for each i, a_i and a_{i+k+1} are in the same component of R. Now since n and $k+1$ are relatively prime, it follows that R has only one component. Thus, when n is odd R is one simple closed curve. Similarly for n odd, B is one simple closed curve.

Before we state our theorem concerning the curves R and B, we prove the following lemma.

Lemma 4. *Let L be an unknotted simple closed curve in S^3. Let $h: S^3 \to S^3$ be a finite order orientation preserving diffeomorphism such that $h(L)=L$, and the fixed point set of h is a simple closed curve A, which is disjoint from L. Then* $\operatorname{lk}(L, A)=1$.

Proof. Let Q denote the complement in S^3 of an open tubular neighborhood of L which is invariant ander h. Since L is unknotted Q is a solid torus. By the Equivariant Dehn's Lemma [MY] there is an embedded meridional disk D in Q with $h(D)=D$ or $h(D)\cap D=\emptyset$. Suppose $D\cap A=\emptyset$, then ∂D is trivial in $\pi_1(S^3-A)$. Hence L is trivial in $\pi_1(S^3-A)$. But this is impossible since $h(L)=L$. Thus A intersects D in $n>0$ points. Now since A is fixed pointwise by h, we must have $h(D)=D$. Thus $h|D$ is a periodic diffeomorphism of a disk with n fixed points. So $n=1$. Therefore $\operatorname{lk}(L, A)=1$. □

Theorem 5. *Let $n \geq 3$ be odd, and let M_n be a Möbius ladder which is embedded in $\mathbb{R}^3$. Suppose there is a diffeomorphism $h: \mathbb{R}^3 \to \mathbb{R}^3$ with $h(M_n)=M_n$ and $h(K)=K$. Suppose further that h is orientation preserving with odd order. Then at least one of the simple closed curves R, B, or K is knotted.*

Proof. Let p be the order of h. Since K has $2n$ edges and p is odd, p must divide n. Thus $h(R)=R$ and $h(B)=B$. Also because p is odd, no individual rung is invariant under h. The fixed point set of an orientation preserving finite order diffeomorphism of $\mathbb{R}^3$ must be an embedded line by Smith Theory [Sm]. Extend h to a map $g: S^3 \to S^3$ by mapping the point at infinity to itself. Now the fixed point set of g is a simple closed curve A which is disjoint from each of K, R, and B. So we can consider the mod 2 algebraic linking number of A with each of these cycles. Let ω_1, ω_2, and ω_3 be the mod 2 algebraic linking numbers of A with R, B, and K respectively.

Suppose that all three loops R, B, and K are unknotted. Then, by Lemma 4, $\omega_1=\omega_2=\omega_3=1$. However, by our construction of R and B, in $H_1(S^3-A, \mathbb{Z}_2)$ we have $[R]+[B]=[K]$. So, as a sum in $\mathbb{Z}_2$, $\omega_1+\omega_2=\omega_3$. This contradiction implies that one of R, B, or K must be knotted. □

Now we shall consider the diffeomorphisms which are induced by reflections of K. In contrast with the elements induced by rotations of K, all of those elements of $\operatorname{Sym}(M_n)$ which are induced by reflections of K can be realized by periodic orientation preserving elements of $\operatorname{Sym}(\mathbb{R}^3, M_n)$. In Fig. 9 we illustrate a Möbius ladder M_4 embedded in $\mathbb{R}^3$ in such a way that there is an orientation preserving diffeomorphism $g: \mathbb{R}^3 \to \mathbb{R}^3$ with $g(M_4)=M_4$ and $g(K)=-K$. The diffeomorphism

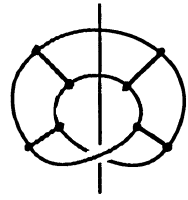


Fig. 9

$g: \mathbb{R}^3 \to \mathbb{R}^3$ is obtained by rotating by 180° about the axis A which is indicated in the Fig. 9. This diffeomorphism has order two, has fixed point set an embedded line, and $g(M_4) = M_4$ and $g(K) = -K$. For any even n we can construct an example which is analogous to this M_4. For any n which is odd we can construct a similar example but where the axis contains one of the rungs. This is illustrated for M_3 in Fig. 10. Therefore for any n, there is a Möbius ladder M_n which is embedded in $\mathbb{R}^3$ with an order two diffeomorphism $g: \mathbb{R}^3 \to \mathbb{R}^3$ with fixed point set an embedded line, where $g(M_n) = M_n$ and $g(K) = -K$.

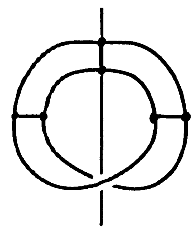


Fig. 10

4. Symmetries of Wheels

In order to try to generalize our results about chirality for Möbius ladders to a larger class of graphs, we introduce the following definitions.

Definition. A *wheel* is a graph consisting of a loop K and rungs $\alpha_1, \ldots, \alpha_n$, where K is a simple closed curve and the α_i are disjoint arcs with their endpoints at distinct points of K.

Definition. Let G be a wheel with loop K and rungs $\alpha_1, \ldots, \alpha_n$. Let B be the set of all rungs β_i with the property that the two components of $K - \beta_i$ contain equal numbers of endpoints of $\alpha_1, \ldots, \alpha_n$. If B contains $r > 0$ rungs, then we define the *Möbius hub* N_r of G to be the loop K together with all the rungs β_i in B.

Observe that the Möbius hub of a Möbius ladder is the Möbius ladder itself.

Lemma 5. *Let G be a wheel with Möbius hub N_r. If $r \geqq 3$, then N_r is a Möbius ladder.*

Proof. Let β be one of the rungs of N_r, and let C_1 and C_2 be the components of $K - \beta$, where K denotes the loop of G. Suppose that N_r is not a Möbius ladder, then without loss of generality there exists a rung γ of N_r such that γ has both endpoints in C_1. Now we can label the components D_1 and D_2 of $K - \gamma$ so that D_1 is contained in C_1 and D_2 contains C_2. By the definition of Möbius hub, C_1 and C_2 contain the same number of endpoints of the rungs $\alpha_1, \ldots, \alpha_n$ of the wheel G. But since D_1 is properly contained in C_1, there must be fewer endpoints of $\alpha_1, \ldots, \alpha_n$ in D_1 than there are in D_2. But this contradicts the fact that γ is also one of the rungs of the Möbius hub N_r. Therefore N_r must have actually been a Möbius ladder. □

Definition. Let G be a wheel with loop K, and let L_p be a subgraph of G containing the loop K and some rungs $\gamma_1, \ldots, \gamma_p$. Then L_p is said to be a *maximal Möbius subgraph* if both

1) L_p is a Möbius ladder, and

2) If α is a rung of G which is not a rung of L_p, then $L_p \cup \{\alpha\}$ is not a Möbius ladder.

Observe that since the Möbius hub of a wheel is a Möbius ladder it is contained in a (possibly non-unique) maximal Möbius subgraph. However, the Möbius hub of a wheel and a maximal Möbius subgraph of that wheel are, in general, different. Even assuming that $r \geqq 3$, the Möbius hub N_r of a wheel G is not necessarily a maximal Möbius subgraph. For example Fig. 11 illustrates a wheel where the Möbius hub N_3 is not a maximal Möbius subgraph. Here, for each β_i in N_3, the two components of $K - \beta_i$ each contain six endpoints of rungs of G. The rungs α_1 and α_2 are not in N_3 because, for $i = 1$ and $i = 2$, one component of $K - \alpha_i$ contains more endpoints than the other component. However, $N_3 \cup \alpha_1$ and $N_3 \cup \alpha_2$ are each maximal Möbius subgraphs of G. It is also easy to construct examples of wheels which have no Möbius hub yet have any number of maximal Möbius subgraphs.

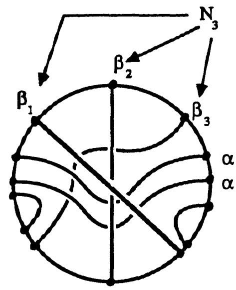


Fig. 11

Lemma 6. *Let G be a wheel with loop K. Suppose there is a homeomorphism h of the graph G such that $h(K) = K$. If G has Möbius hub N_r, then $h(N_r) = N_r$; also, if there is a p such that G contains a unique maximal Möbius subgraph L_p with p rungs, then $h(L_p) = L_p$.*

Proof. Let $A = \{\alpha_1, \ldots, \alpha_n\}$ denote the set of all rungs of G and let B denote the subset of A consisting of rungs of N_r. Since $h(K) = K$, we must have $h(A) = A$. By the definition of the Möbius hub, B is the set of all rungs β_i such that the two components of $K - \beta_i$ contain equal numbers of endpoints of $\alpha_1, \ldots, \alpha_n$. Since h is a homeomorphism of G, in fact $h(B) = B$. Thus $h(N_r) = N_r$. Also, since L_p is the unique maximal Möbius subgraph with p rungs $h(L_p) = L_p$. □

Definition. A wheel G with loop K is *intrinsically chiral* if for any embedding of G in S^3 there is no orientation reversing diffeomorphism $h: S^3 \to S^3$ with $h(G) = G$ and $h(K) = K$.

We have shown in Theorem 2 that any Möbius ladder with an odd number of rungs is intrinsically chiral. This, together with Lemma 4, easily leads us to Theorem 6.

Theorem 6. *Let G be a wheel, and let $p \geqq 3$ be an odd number. Suppose that either the Möbius hub of G has p rungs, or G contains a unique maximal Möbius subgraph with p rungs. Then G is intrinsically chiral.*

Proof. Suppose that G is embedded in S^3 in such a way that there is an orientation reversing diffeomorphism $h: S^3 \to S^3$ with $h(G)=G$ and $h(K)=K$. By Lemma 4, if G has Möbius hub N_p, then $h(N_p)=N_p$, or if L_p is the unique maximal Möbius subgraph of G with p rungs, then $h(L_p)=L_p$. In either case, since $p \geqq 3$ and p is odd this contradicts Theorem 2. Therefore, in either case, G is intrinsically chiral. □

This theorem provides us with a way to construct many examples of intrinsically chiral wheels. However, not all intrinsically chiral wheels satisfy the hypotheses of this theorem. For example, Lemma 4 can also be used to construct various other types of intrinsically chiral wheels. An example of this other type is illustrated in Fig. 12. Here, the wheel G has no Möbius hub; however, it has a unique maximal Möbius subgraph L_4 with four rungs. By Lemma 4, any homeomorphism of h of G with $h(K)=K$, would also leave L_4 setwise invariant. But $G=L_4 \cup \beta$, thus $h(\beta)=\beta$. Hence also $h(\alpha)=\alpha$. Now let M_3 denote the wheel G after the rungs α and β have been removed. Then M_3 is a Möbius ladder with an odd number of rungs, and $h(M_3)=M_3$. Again this contradicts Theorem 2, so in fact G is intrinsically chiral.

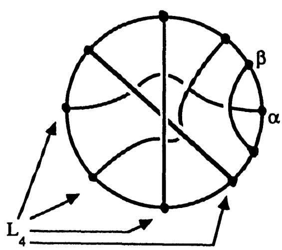


Fig. 12

References

[CG] Conway, J.H., Gordon, C.McA.: Knots and links in spatial graphs. J. Graph Theory **7**, 445–453 (1983)

[HG] Guy, R.K., Harary, F.: On the Möbius ladders. Can. Math. Bull. **10**, 493–496 (1967)

[MY] Meeks, W., Yau, S.-T.: Topology of three dimensional manifolds and the embedding problems in minimal surface theory. Ann. Math. **112**, 441–485 (1980)

[Ro] Rolfsen, D.: Knots and links. Berkeley: Publish or Perish Press 1976

[Si] Simon, J.: Topological chirality of certain molecules. Topology **25**, 229–235 (1986)

[Sm] Smith, P.A.: Transformations of finite period. II. Ann. Math. **40**, 670–711 (1939)

[Wa] Walba, D.: Stereochemical topology. In: R.B. King (ed.): Chemical applications of topology and graph theory, pp. 17–32. Amsterdam: Elsevier 1983

[WRH] Walba, D., Richards, R., Haltiwanger, R.C.: Total synthesis of the first molecular Möbius strip. J. Am. Chem. Soc. **104**, 3219–3221 (1982)

Received May 29, 1987; in revised form June 13, 1988

Math. Ann. 283, 285–299 (1989)

Mathematische Annalen

A Root System for the Lyons Group

Wolfram Neutsch[1] and Werner Meyer[2]

[1] Institut für Astrophysik der Universität Bonn, Auf dem Hügel 71, D-5300 Bonn 1, Federal Republic of Germany
[2] Max-Planck-Institut für Mathematik, Gottfried-Claren-Strasse 26, D-5300 Bonn 3, Federal Republic of Germany

0. Introduction

Sims (1973) proved the existence and uniqueness of the sporadic group Ly predicted by Lyons (1972) through the computer-aided construction of a presentation which, unfortunately, is rather cumbersome and does not lead to an insight into the structure of the group.

Meanwhile, much more information on Ly has become available.

Kantor (1981) found a Tits geometry for Ly which is "almost" a building. Meyer et al. (1985) gave the absolute minimal representation of Ly (111-dimensional over $\mathbb{F}_5$) which had been conjectured by Meyer and Neutsch (1984) and, independently, by Woldar (1987). Later, Wilson [1984, 1985] compiled the list of all maximal subgroups in Ly. His investigation uses the minimal representation explicitly, while the verification of the latter depends on Sims' presentation. For that reason, it would be of great interest to have a simpler existence and uniqueness proof.

Inspired by Kantor's results, we were led to the idea of giving a symmetric presentation of a group Γ by making use of the beautiful geometry of Ly.

Our relations are shown to be fulfilled by certain generators ("roots") of the Lyons group, and most probably they define Ly itself.

The construction is carried out in a fashion nearly identical to the methods of Chevalley theory employed to study the Tits buildings of the groups of Lie type ($G_2(5) < \mathrm{Ly}$ should be considered as a prototype). The geometric spirit of our presentation renders this possible. It is a first step towards an understanding of the Lyons group.

1. Relations in a Group of Ly Type

We say a group Λ is of Ly type if it has the following properties:

(1) Λ is simple;

(2) Λ contains an involution z with $C_\Lambda(z) \cong 2 \wedge A_{11}$.

Lyons [1972] shows that a group fulfilling (1) and (2) is of order

$$|\mathrm{Ly}| = 2^8.3^7.5^6.7.11.31.37.67 \tag{1.1}$$

and that it contains a unique conjugacy class of subgroups $\cong G_2(5)$. Let Λ_0 be one of them and B a Borel subgroup of Λ_0, i.e. a 5-Sylow normalizer.

Then B is also a Borel group in Λ.

Furthermore, let $T \cong 4^2$ be a (maximal) torus in B, N_0 and N its normalizers in Λ_0 and Λ, respectively, and $W_0 = N_0/T$ and $W = N/T$ the corresponding Weyl groups.

From the theory of Chevalley groups, cf. e.g. Carter (1972), we deduce

$$W_0 \cong D_{12} \cong S_3 \times S_2 \tag{1.2}$$

while Kantor [1981] shows

$$W \cong S_4 \times S_3 \ . \tag{1.3}$$

A proper subgroup of Λ or Λ_0 which contains a Borel group will be called parabolic.

Kantor [1981] has shown that Λ contains exactly three conjugacy classes of maximal parabolic groups. They can be associated with the points P, lines L and planes F of a Tits geometry with the Buekenhout diagram

```
                        6
o-----------o-----------o
5           5           5
P           L           F
G2(5)    5^(1+4):4S6   5^3.SL3(5)
```
(1.4)

Two objects (points, lines or planes) are called incident with each other if their intersection (as groups) is parabolic.

The apartment $A(T)$ associated with T is the set of all objects fixed by T. $A(T)$ is a subgeometry with Buekenhout diagram

```
                        6
o-----------o-----------o
1           1           1
```
(1.5)

and may be represented (Kantor (1981)) by a simplicial complex A of dimension 2:

```
              1c -------- 2a -------- 1b
             /  \        /  \        /  \
            /    \      /    \      /    \
           /      \    /      \    /      \
         4a -------- 3b -------- 4c -------- 3a
        /  \        /  \        /  \        /  \
       /    \      /    \      /    \      /    \
      /      \    /      \    /      \    /      \
    1b -------- 2c -------- 1a -------- 2b -------- 1c
      \      /    \      /    \      /    \      /
       \    /      \    /      \    /      \    /
        \  /        \  /        \  /        \  /
         3a -------- 4b -------- 3c -------- 4a
           \      /    \      /    \      /
            \    /      \    /      \    /
             \  /        \  /        \  /
              1c -------- 2a -------- 1b
```

Fig. 1

Here the 0-, 1-, 2-simplices correspond to the 12 points, 36 lines, 24 planes of the apartment, respectively.

The Weyl group $W \cong S_4 \times S_3$ acts as S_4 on the numbers $\{1,2,3,4\}$ and as S_3 on the letters $\{a,b,c\}$.

In analogy to Chevalley theory we now define the root groups associated with T as the groups X obeying the conditions:

(1) $X \cong (\mathbb{F}_5, +) \cong 5$;

(2) $T \leqq N(X)$.

It follows from the known structure of $C_{Ly}(5B)$ (Lyons (1972)) that all root groups must be generated by $5A$-elements.

Since $C_{Ly}(5A) \cong 5^{1+4} : (2 \wedge A_6)$ does not contain a Klein four group, $C_T(X) \cong 4$.

The A-normalizer of a $5A$-group is a line. Thus there is a natural bijection between the root groups and the lines in $A(T)$.

The extension of T with the commutator subgroup

$$W' \cong A_4 \times A_3 \tag{1.6}$$

of W splits, so there is a unique element k of order 3 in N which corresponds to the Weyl element (abc) and centralizes T. In fact (Lyons [1972]),

$$C_A(T) = T \times K \tag{1.7}$$

with

$$K = \langle k \rangle \tag{1.8}$$

Furthermore, there is a set of 16 complements of T in the unique $T : A_4$ fixing the letters a, b, c. These groups are evidently conjugate under T, so we may elect an arbitrary one of them and denote it by Ω.

Ω is generated by 4 elements ω_i $(1 \leqq i \leqq 4)$ which correspond to the 120°-rotations with centres in the points whose names contain the number i.

Each ω_i is uniquely determined by the choice of Ω and the corresponding Weyl permutation, namely

$$\omega_1 \to (234); \quad \omega_2 \to (143); \quad \omega_3 \to (124); \quad \omega_4 \to (132) \quad \text{in} \quad W. \tag{1.9}$$

The group

$$\langle \omega_i, k \rangle = \Omega \times K \cong A_4 \times A_3 \tag{1.10}$$

(one of just 16 complements of T in $N' = T : W' \cong 4^2 : (A_4 \times A_3)$) is represented as a regular permutation group on the root groups X_L.

This allows us to specify a set of 36 generators ("roots") for each of the 36 X_L.

We are free to take any generator for one of them, e.g. $X(1a, 2b)$. Call it $x(1a, 2b)$. Then apply $\Omega \times K$ to this root to define the remaining ones. A complete system of 36 root elements generated in this way will be called a standard (root) system.

Without restriction of generality we may assume that the Chevalley relations of $G_2(5)$ hold in the form described below (the exact exponents depend on the choice of $x(1a, 2b)$, but this clearly does not matter, because all allowed possibilities are equivalent and lead to the same group):

It will be convenient to define an orientation of the lines in A according to the rules

$$a \to b \;, \quad b \to c \;, \quad c \to a \;. \tag{1.11}$$

Now we consider a point P in A.

The 6 lines incident with P form a complete set of long roots for the stabilizer $A(P)$ of P, isomorphic to the Chevalley group $G_2(5)$, while the short roots are given by the sides of the (small) hexagon with centre P spanned by the long roots.

We denote the long and short roots by L_i and K_i ($i \in \mathbb{F}_7^\times$), respectively, in the following manner:

```
     P2      K1      P3
       o-----------o
      / \         / \
 K3  /   \       /   \  K5
    /     \     /     \
   /       \   /       \
P6 o-----------o-----------o P1
   \       /   \       /
 K2 \     /     \     / K4
     \   /       \   /
      \ /         \ /
       o-----------o
     P4      K6      P5
```

Fig. 2

where L_i points from P to P_i for $i=1,2,4$ (squares in $\mathbb{F}_7$) and from P_i to P for $i=3,5,6$ (non-squares).

The 12 roots in Fig. 2 follow each other in the same order as they do in the standard G_2 root system.

Then the nontrivial Chevalley relations are

$$[L_i, L_{2i}] = L_{3i}^4 \;, \tag{1.12}$$

$$[K_i, K_{3i}] = L_{2i}^3 \;, \tag{1.13}$$

$$[K_i, K_{2i}] = L_{2i}^3 K_{3i}^3 L_{6i}^2 \;, \tag{1.14}$$

$$[K_i, L_{4i}] = L_{2i}^4 K_{3i} L_{6i}^4 K_{2i} \;, \tag{1.15}$$

$$[L_i, K_{3i}] = K_{5i} L_{3i} K_i^4 L_{2i}^4 \tag{1.16}$$

combined with the information that for all $i \in \{1,2,4\}$ the mappings

$$K_i \Rightarrow \begin{vmatrix} 1 & 1 \\ . & 1 \end{vmatrix} \;, \quad K_{-i} \Rightarrow \begin{vmatrix} 1 & . \\ 4 & 1 \end{vmatrix} \;, \tag{1.17}$$

and

$$L_i \Rightarrow \begin{vmatrix} 1 & 1 \\ . & 1 \end{vmatrix} \;, \quad L_{-i} \Rightarrow \begin{vmatrix} 1 & . \\ 3 & 1 \end{vmatrix} \tag{1.18}$$

are isomorphisms from $\langle K_i, K_{-i} \rangle$ and $\langle L_i, L_{-i} \rangle$ onto $SL_2(5)$.

It should be noted that our relations differ slightly from those described, e.g., in Humphreys [1975]. This is due to our more symmetric choice of the roots which is more convenient in the context of the Lyons group.

For later reference, we construct an explicit $G_2(5)$ root system in the 7-dimensional minimal representation over $\mathbb{F}_5$ as the full automorphism group of the "septime" algebra with generators e_i ($i \in \mathbb{F}_7$) and the skew-symmetric product given by

$$e_{i+j} \cdot e_{i+2j} = e_{i+4j} \qquad (i \in \mathbb{F}_7\,;\; j \in \{1,2,4\})\;. \tag{1.19}$$

Up to conjugacy in $G_2(5)$, our matrices are uniquely determined:

$$L_1 = \left|\begin{matrix} 1&.&.&.&.&.&. \\ .&1&.&.&.&.&. \\ .&.&1&.&3&1&. \\ .&.&.&1&.&.&. \\ .&.&2&.&1&.&1 \\ .&.&4&.&.&1&2 \\ .&.&.&.&4&3&1 \end{matrix}\right|, \quad L_2 = \left|\begin{matrix} 1&.&.&.&.&.&. \\ .&1&.&.&2&1&. \\ .&.&1&.&.&.&. \\ .&.&.&1&4&2&. \\ .&3&.&1&1&.&. \\ .&4&.&3&.&1&. \\ .&.&.&.&.&.&1 \end{matrix}\right|, \quad L_4 = \left|\begin{matrix} 1&.&.&.&.&.&. \\ .&1&3&.&.&.&1 \\ .&2&1&1&.&.&. \\ .&.&4&1&.&.&3 \\ .&.&.&.&1&.&. \\ .&.&.&.&.&1&. \\ .&4&.&2&.&.&1 \end{matrix}\right|, \tag{1.20}$$

$$L_6 = \left|\begin{matrix} 1&.&.&.&.&.&. \\ .&1&.&.&.&.&. \\ .&.&1&.&1&3&. \\ .&.&.&1&.&.&. \\ .&.&4&.&1&.&3 \\ .&.&2&.&.&1&4 \\ .&.&.&.&2&1&1 \end{matrix}\right|, \quad L_5 = \left|\begin{matrix} 1&.&.&.&.&.&. \\ .&1&.&.&4&3&. \\ .&.&1&.&.&.&. \\ .&.&.&1&2&4&. \\ .&1&.&3&1&.&. \\ .&2&.&1&.&1&. \\ .&.&.&.&.&.&1 \end{matrix}\right|, \quad L_3 = \left|\begin{matrix} 1&.&.&.&.&.&. \\ .&1&1&.&.&.&3 \\ .&4&1&3&.&.&. \\ .&.&2&1&.&.&1 \\ .&.&.&.&1&.&. \\ .&.&.&.&.&1&. \\ .&2&.&4&.&.&1 \end{matrix}\right|, \tag{1.21}$$

$$K_1 = \left|\begin{matrix} 1&1&.&2&.&.&. \\ 4&3&.&4&.&.&. \\ .&.&1&.&1&3&. \\ 3&4&.&4&.&.&. \\ .&.&4&.&1&.&2 \\ .&.&2&.&.&1&1 \\ .&.&.&.&3&4&1 \end{matrix}\right|, \quad K_2 = \left|\begin{matrix} 1&.&1&.&.&.&2 \\ .&1&.&.&4&2&. \\ 4&.&3&.&.&.&4 \\ .&.&.&1&2&1&. \\ .&1&.&3&1&.&. \\ .&3&.&4&.&1&. \\ 3&.&4&.&.&.&4 \end{matrix}\right|, \quad K_4 = \left|\begin{matrix} 1&.&.&.&1&2&. \\ .&1&1&.&.&.&3 \\ .&4&1&2&.&.&. \\ .&.&3&1&.&.&4 \\ 4&.&.&.&3&4&. \\ 3&.&.&.&4&4&. \\ .&2&.&1&.&.&1 \end{matrix}\right|, \tag{1.22}$$

$$K_6 = \left|\begin{matrix} 1&1&.&3&.&.&. \\ 4&3&.&1&.&.&. \\ .&.&1&.&4&3&. \\ 2&1&.&4&.&.&. \\ .&.&1&.&1&.&2 \\ .&.&2&.&.&1&4 \\ .&.&.&.&3&1&1 \end{matrix}\right|, \quad K_5 = \left|\begin{matrix} 1&.&1&.&.&.&3 \\ .&1&.&.&1&2&. \\ 4&.&3&.&.&.&1 \\ .&.&.&1&2&4&. \\ .&4&.&3&1&.&. \\ .&3&.&1&.&1&. \\ 2&.&1&.&.&.&4 \end{matrix}\right|, \quad K_3 = \left|\begin{matrix} 1&.&.&.&1&3&. \\ .&1&4&.&.&.&3 \\ .&1&1&2&.&.&. \\ .&.&3&1&.&.&1 \\ 4&.&.&.&3&1&. \\ 2&.&.&.&1&4&. \\ .&2&.&4&.&.&1 \end{matrix}\right|. \tag{1.23}$$

Conjugation with $\Omega \times K$ merely permutes the roots (without exponents). Because of this fact, all standard systems are equivalent and lead to the same set of relations.

The lines in $A(T)$ form 3 parallel classes of 12 lines each. Every parallel class splits into 2 connected components called (great) circles.

A special line pair is a pair (L, L') of lines which are contained in a great circle and either have one point in common ("long" pair) or are mutual antipodes ("short" pair). The reason for this notation is that short and long special pairs form opposite pairs of short and long roots, respectively, in a certain $G_2(5)$ subgroup.

Since two opposite long root groups in $G_2(5)$ have the same centralizer in T, this must also be true for all 6 lines in a great circle.

The centralizer in the point $P \in A(T)$ of an arbitrary T-involution z contains 4 roots in each parallel class. Thus the group $\langle \Pi \rangle$ generated by a parallel class Π is a subgroup of $H = C_A(z) \cong 2 \wedge A_{11}$.

Let $\bar{H} = H/\langle z \rangle \cong A_{11}$.

As all subgroups isomorphic to 4^2 are conjugate in H we may assume without restriction that

$$\bar{T} = T/\langle z \rangle = \langle (1234)(5678), (1234)(8765) \rangle \ . \tag{1.24}$$

In H or $\bar{H}$ exactly 12 groups $\cong 5$ exist which are normalized by T or $\bar{T}$. Except for a permutation of the letters $\{1, 2, 3, 4, 5, 6, 7, 8, 9, X, E\}$ normalizing T only the following correspondence between the roots and the permutations in $\bar{H}$ is allowed by the Chevalley relations for the points:

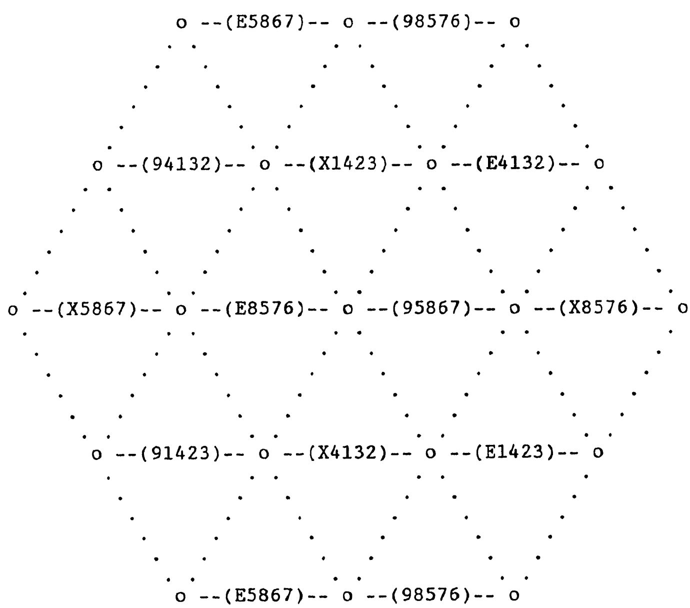


Fig. 3

It is obvious that these roots generate $2 \wedge A_{11}$.

2. Definition and Simple Geometric Properties of the Group Γ

The results of Sect. 1 lead us to the definition of a group Γ as follows:

Γ is generated by 36 elements x_L, bijectively associated with the lines L in A (Fig. 1). The defining relations of Γ are

(1) $C(P)$-relations for every point P in A;

(2) $S(\Pi)$-relations for each parallel class Π in A.

Here $C(P)$ is the set of Chevalley relations of $G_2(5)$ (cf. Sect. 1) while $S(\Pi)$ is an arbitrary system of defining relations for $2 \wedge A_{11}$. Best suited for our purpose are the Schur relations:

$$t_i^3=1 \qquad (1 \leqq i \leqq 9) \ , \tag{2.1}$$

$$(t_i \cdot t_j)^2=z \qquad (1 \leqq i,j \leqq 9;\ i \neq j) \ , \tag{2.2}$$

$$z^2=1 \tag{2.3}$$

where the generator t_i corresponds to the permutation (iXE).

We are now able to translate between the two sets of generators of $2 \wedge A_{11}$ (here $x(P,Q)$ is the root element which belongs to the line connecting the points P and Q):

$$t_9=x(3c,4a)\cdot x(4b,3c)^{-1}\cdot x(3a,4b)\cdot x(4a,3b)^{-1} \cdot x(3a,4b)\cdot x(4b,3c)\cdot x(3c,4a)^{-1} \ , \tag{2.4}$$

$$t_1=x(3a,4b)^{-1}\cdot t_9\cdot x(3a,4b) \ , \qquad t_2=x(3a,4b)^2\cdot t_9\cdot x(3a,4b)^{-2} \ , \tag{2.5}$$

$$t_3=x(3a,4b)\cdot t_9\cdot x(3a,4b)^{-1} \ , \qquad t_4=x(3a,4b)^{-2}\cdot t_9\cdot x(3a,4b)^2 \ , \tag{2.6}$$

$$t_5=x(1a,2b)^{-1}\cdot t_9\cdot x(1a,2b) \ , \qquad t_6=x(1a,2b)^2\cdot t_9\cdot x(1a,2b)^{-2} \ , \tag{2.7}$$

$$t_7=x(1a,2b)\cdot t_9\cdot x(1a,2b)^{-1} \ , \qquad t_8=x(1a,2b)^{-2}\cdot t_9\cdot x(1a,2b)^2 \ . \tag{2.8}$$

The reverse transformations are:

$$x(3a,4b)=t_9^{-1}\cdot t_1^{-1}\cdot t_4\cdot t_2^{-1}\cdot t_3\cdot t_1\cdot t_9 \ , \qquad x(1a,2b)=t_9^{-1}\cdot t_5^{-1}\cdot t_8\cdot t_6^{-1}\cdot t_7\cdot t_5\cdot t_9 \ , \tag{2.9}$$

$$x(4b,3c)=t_2\cdot t_4\cdot t_1^{-1}\cdot t_3\cdot t_2^{-1} \ , \qquad x(2b,1c)=t_6\cdot t_8\cdot t_5^{-1}\cdot t_7\cdot t_6^{-1} \ , \tag{2.10}$$

$$x(3c,4a)=t_1^{-1}\cdot t_4\cdot t_2^{-1}\cdot t_3\cdot t_1 \ , \qquad x(1c,2a)=t_5^{-1}\cdot t_8\cdot t_6^{-1}\cdot t_7\cdot t_5 \ , \tag{2.11}$$

$$x(4a,3b)=t_9^{-1}\cdot t_4^{-1}\cdot t_1\cdot t_3^{-1}\cdot t_2\cdot t_4\cdot t_9 \ , \qquad x(2a,1b)=t_9^{-1}\cdot t_8^{-1}\cdot t_5\cdot t_7^{-1}\cdot t_6\cdot t_8\cdot t_9 \ , \tag{2.12}$$

$$x(3b,4c)=t_3\cdot t_1\cdot t_4^{-1}\cdot t_2\cdot t_3^{-1} \ , \qquad x(1b,2c)=t_7\cdot t_5\cdot t_8^{-1}\cdot t_6\cdot t_7^{-1} \ , \tag{2.13}$$

$$x(4c,3a)=t_4^{-1}\cdot t_1\cdot t_3^{-1}\cdot t_2\cdot t_4 \ , \qquad x(2c,1a)=t_8^{-1}\cdot t_5\cdot t_7^{-1}\cdot t_6\cdot t_8 \ . \tag{2.14}$$

By the main result of Meyer et al. [1985] the Lyons group possesses a 111-dimensional irreducible representation over $\mathbb{F}_5$. In this we can easily identify 36 elements which generate Ly and satisfy all relations defining Γ.

Hence we have

Lemma 1. (a) *The Lyons group is a homomorphic image of* Γ;

(b) Γ *has a* 111-*dimensional nontrivial representation over* $\mathbb{F}_5$.

Let us now consider the following subconfigurations of A in Fig. 1:

```
   o-----o          o-----o-----o          o-----o-----o
  / \   / \              / \                \   / \   /
 /   \ /   \            /   \                \ / F \ /
o-----P-----o          o--L--o                o-----o
 \   / \   /            \   /                  \   /
  \ /   \ /              \ /                    \ /
   o-----o          o-----o-----o                o
```

a **b** **c**

Fig. 4a–c

The subgroups of Γ generated by the lines in Fig. 4a–c (respectively) are called $\Gamma(P)$; $\Gamma(L)$; $\Gamma(F)$.

For any set $\{\mathcal{O}, \mathcal{O}', \ldots\}$ of objects we define $\Gamma(\mathcal{O}, \mathcal{O}', \ldots)$ as the intersection of the groups $\Gamma(\mathcal{O})$, $\Gamma(\mathcal{O}'), \ldots$

With the above notation we have

Theorem 1. (a) $\Gamma(P) \cong G_2(5)$;
(b) $\Gamma(L) \cong 5^{1+4} : (2 \wedge A_6)$;
(c) $\Gamma(F) \cong 5^3 \cdot SL_3(5)$.

Theorem 2. *The lines in the following configurations* (*omitting the dotted lines*):

(a)
```
o-----o
```

(b)
```
   o
  / \
 /   \
o-----o
```

(c)
```
   o
  / \
 /   \
o-----o
 \   /
  \ /
   o
```

(d)
```
o-----o-----o
 .   . .   .
  . .   . .
   o.....o
  . .   . .
 .   . .   .
o-----o-----o
```

(e)
```
o-----o-----o
 .   . .   .
  . .   . .
   o-----o
  . .   . .
 .   . .   .
o-----o-----o
```

(f) great circle

(g) parallel class

generate groups which are isomorphic to:

(a) 5;
(b) 5^3;
(c) 5^{1+4};
(d) $2 \wedge A_6$;
(e) $5 \times (2 \wedge A_6)$;
(f) $2 \wedge A_7$;
(g) $2 \wedge A_{11}$.

Proof of Theorems 1 and 2. (1a) Due to the $C(P)$-relations, $\Gamma(P)$ is a homomorphic image of $G_2(5)$ (using a theorem of Steinberg, cf. Carter (1972), Theorem 12.1.1), so $\Gamma(P)$ is $\cong G_2(5)$ or $\cong 1$.

In the latter case a root group in $\Gamma(P)$ and hence also in $\Gamma(P')$ for a neighbouring point P' of P would be trivial, so $\Gamma(P')=1$, too. This leads to $\Gamma=1$, contradicting Lemma 1.

(2g) Because of the $S(\Pi)$-relations, $\langle \Pi \rangle$ is a homomorphic image of $2 \wedge A_{11}$, so is $\cong 2 \wedge A_{11}$, A_{11} or 1. Only the first possibility is in conformity with (1a), since a long special line pair generates $SL_2(5) \cong 2 \wedge A_5$. The remaining statements in Theorem 2 now follow immediately from (1a) and (2g).

(1b) Since (2c) and (2d) hold, we need only show that 5^{1+4} is normalized by $2 \wedge A_6$. This follows from (1a), applied to the two points incident with L.

(1c) Analogously to (1b), we conclude with the help of (2b) that the three lines incident with F generate a normal subgroup $\Gamma_0(F) \cong 5^3$ of $\Gamma(F)$.

The images in $\Gamma(F)/\Gamma_0(F)$ of the root subgroups in $\Gamma(F)$ fulfill all of the Chevalley relations for the group $SL_3(5)$ (which is defined by these relations) if we map them as follows:

```
o --a-- o --A-- o                      o       o
 \     / \     /                        \     /
  C   / F \   b                          C   b
   \ /     \ /                            \ /
    o ----- o            ===>      o --a-- o --A-- o
     \     /                              / \
      c   B                              B   c
       \ /                              /     \
        o                              o       o
```

Fig. 5

$SL_3(5)$ is simple, and according to (2c) not all of the images can be trivial; thus $\Gamma(F) \cong 5^3 \cdot SL_3(5)$. This extension does not split, since $\Gamma(F)$ contains a 5-Sylow subgroup of $G_2(5)$ and therefore elements of order 25. This establishes Theorems 1 and 2.

We now define for an arbitrary (long or short) special line pair L, L' the groups $T_{LL'}$ and $Q_{LL'}$ as follows:

Let $T_{LL'}$ be the common normalizer of the root groups L and L' and let $Q_{LL'}$ be the set normalizer of $\{L, L'\}$ in $\langle L, L' \rangle \cong SL_2(5)$.

Furthermore, for each great circle K and each parallel system Π we introduce the abbreviations

$$T_K = \langle T_{LL'} : L, L' \text{ special line pair in } K \rangle \ , \tag{2.15}$$

$$Q_K = \langle Q_{LL'} : L, L' \text{ special line pair in } K \rangle \ , \tag{2.16}$$

$$T_\Pi = \langle T_{LL'} : L, L' \text{ special line pair in } \Pi \rangle \ , \tag{2.17}$$

$$Q_\Pi = \langle Q_{LL'} : L, L' \text{ special line pair in } \Pi \rangle \ , \tag{2.18}$$

as well as

$$T = \langle T_{LL'} : L, L' \text{ special line pair} \rangle \ , \tag{2.19}$$

$$Q = \langle Q_{LL'} : L, L' \text{ special line pair} \rangle \tag{2.20}$$

and for any point P:

$$T(P) = \langle T_{LL'} : L, L' \text{ special line pair in } \Gamma(P) \rangle \ . \tag{2.21}$$

Of course, $T(P)$ is the torus in $\Gamma(P) \cong G_2(5)$ belonging to our apartment. We next show

Theorem 3. (a) *For a special line pair L, L' in the circle K we have $T_{LL'} = T_K \cong 4$;*
(b) *for every parallel class Π : $T_\Pi \cong 4 \times 2$;*
(c) *for all points P : $T(P) = T \cong 4^2$.*

Proof. (a) and (b) follow from an easy calculation in $\langle \Pi \rangle \cong 2 \wedge A_{11}$. Trivially, we have $T(P) \leqq T$. With (a) we deduce for every great circle K with an arbitrary but fixed P that $T_K \leqq T(P)$. Since $\langle T_K \rangle = T$, we get (c).

Theorem 4. (a) *For all special line pairs $L, L' : Q_{LL'} = N_{\langle L, L' \rangle}(T_{LL'}) \cong Q_8$, the quaternion group of order* 8; *the intersection of T with $Q_{LL'}$* is $T_{LL'}$;
(b) *T is a normal subgroup of Q;*
(c) *each element q of Q permutes the lines of A, inducing an automorphism of A as a simplicial complex;*
(d) *the image of this action is the full automorphism group $S_4 \times S_3$ of A.*

Proof. The first part of (a) is immediate since $\langle L, L' \rangle \cong SL_2(5)$. The second part can be verified in $\Gamma(P)$ for an appropriate point P.

In this $\Gamma(P)$ we also see that $Q_{LL'}$ normalizes $T(P) = T$, thus the same holds true for $Q = \langle Q_{LL'} \rangle$. Furthermore, each $T_{LL'}$ is contained in $Q_{LL'}$, hence in Q; so $T = \langle T_{LL'} \rangle$ is a subgroup of Q. This proves (b).

Let q be an element of $Q_{LL'}$. If q is contained in $T_{LL'} < T$, (c) holds trivially. If q is in $Q_{LL'} \backslash T_{LL'}$, q induces a permutation of the groups of order 5 which are normalized by T in each of the groups $\Gamma(P)$ and $\langle \Pi \rangle$ where P is any point with $L, L' < \Gamma(P)$ and Π the parallel system containing L and L'. But all these groups of order 5 are root groups.

From the $C(P')$-relations for appropriate points P' we find that the 16 remaining roots are also permuted. Inspection of the permutations generated by Q easily leads to (c) and (d).

3. Some Geometric Subgroups of Γ

Let Π be a parallel class and P a point in A. The group $H=\langle\Pi\rangle$ is $\cong 2\wedge A_{11}$ by Theorem 2.g. We denote the unique involution in $Z(H)$ by z.

We now prove

Theorem 5. *The intersection of H and $\Gamma(P)$ is $C_{\Gamma(P)}(z)\cong(1/2).2\wedge(S_5\times S_5)$.*

Proof. Since all pairs (Π, P) are equivalent under Q (Theorem 4.d), we may restrict ourselves to the case $P=1a$ and $\Pi=$parallel system of Fig. 3. Then H and $\Gamma(P)$ obviously contain the 4 roots $x(1a,2b)$, $x(2c,1a)$, $x(3b,4c)$, $x(4b,3c)$ which generate a group $SL_2(5)\underset{2}{\vee}SL_2(5)\cong 2\wedge(A_5\times A_5)$ of index 2 in $C_{\Gamma(P)}(z)$ $\cong(1/2).2\wedge(S_5\times S_5)$. This group is enlarged by $T-T<H$ and $T<\Gamma(P)$ because of Theorem 3.c – to the full centralizer of z in $\Gamma(P)$. As z is in the centre of H, the intersection of H and $\Gamma(P)$ is a subgroup of $C_{\Gamma(P)}(z)$; hence the proposition.

We want to consider several groups which are defined symmetrically with respect to the apartment $A(T)$.

Let

$$U_1=\Gamma(1a,1b,1c)\ ,\qquad U_2=\Gamma(2a,2b,2c)\ , \tag{3.1}$$

$$U_3=\Gamma(3a,3b,3c)\ ,\qquad U_4=\Gamma(4a,4b,4c)\ , \tag{3.2}$$

and

$$U=\langle U_1, U_2, U_3, U_4\rangle\ . \tag{3.3}$$

It will be convenient to have a systematic notation for the circles, parallel systems and corresponding $2\wedge A_{11}$-subgroups in Γ:

We denote the circle containing the points with numbers i and j by K_{ij} and the parallel system consisting of K_{ij} and K_{kl} by $\Pi_{ij.kl}$. The corresponding T-involution will be called $z_{ij.kl}$, and we set $H_{ij.kl}=\langle\Pi_{ij.kl}\rangle$.

Hence the torus elements $z_{12.34}$, $z_{13.24}$, $z_{14.23}$ are canonically associated with the double transpositions in the symmetric group S_4, while the circles K_{12}, K_{13}, K_{14}, K_{23}, K_{24}, K_{34} belong to the transpositions of S_4.

Let us now investigate the groups $U_i\,(1\leqq i\leqq 4)$ and U:

Theorem 6. (a) $U_1\cong U_2\cong U_3\cong U_4\cong U_3(3)$;
(b) $U'=U$; $U/Z(U)\cong U_4(3)\cong O_6^-(3)$; $Z(U)\leqq 4\times 3^2$.

Proof. We define

$$a=x(4b,3c)^4x(3b,4c)^2\ \Rightarrow\ (132)\ \text{ in }\ H_{12.34}\ , \tag{3.4}$$

$$b=x(4b,3c)^1x(3b,4c)^3\ \Rightarrow\ (143)\ \text{ in }\ H_{12.34}\ , \tag{3.5}$$

$$c=x(4b,2c)^1x(2b,4c)^3\ \Rightarrow\ (124)\ \text{ in }\ H_{13.24}\ , \tag{3.6}$$

$$r=x(1a,2b)^1x(2a,1b)^3\ \Rightarrow\ (568)\ \text{ in }\ H_{12.34}\ , \tag{3.7}$$

$\Gamma(1a)$, $\Gamma(1b)$, $\Gamma(1c)$ contain the $2\wedge A_5$-groups $\langle x(3b,4c),x(4b,3c)\rangle$, $\langle x(3c,4a),x(4c,3a)\rangle$, $\langle x(3a,4b),x(4a,3b)\rangle$ of $H_{12.34}$, acting on the sets $\{1,2,3,4,X\}$, $\{1,2,3,4,E\}$, $\{1,2,3,4,9\}$, respectively.

Their intersection, the $2\wedge A_4$-group on $\{1,2,3,4\}$, is thus contained in $\Gamma(1a,1b,1c)=U_1$.

Obviously, analogous results for $H_{13.24}$ and $H_{14.23}$ hold. Hence, by (3.4), (3.5), (3.6),

$$\langle a,b,c\rangle \leqq U_1 = \Gamma(1a,1b,1c) \leqq \Gamma(1a) \ . \tag{3.8}$$

In $\Gamma(1a)$ we easily verify – see (1.20), ..., (1.23) – that

$$a^3 = b^3 = c^3 = 1 \ , \tag{3.9}$$

$$aba = bab \ , \qquad aca = cac \ , \qquad bcb = cbc \ , \tag{3.10}$$

$$a^b c^{-1} a^b = c^{-1} a^b c^{-1} \ ; \qquad b^a c^{-1} b^a = c^{-1} b^a c^{-1} \ . \tag{3.11}$$

These relations form a presentation of the finite simple group $U_3(3)$, cf. Aschbacher and Hall [1973].

Since $\langle a,b,c\rangle$ is nontrivial, we deduce

$$U_3(3) \cong \langle a,b,c\rangle \leqq U_1 \leqq \Gamma(1a) \cong G_2(5) \ . \tag{3.12}$$

By inspection of the maximal subgroups of $G_2(5)$ we are left with three candidates for U_1, namely $\langle a,b,c\rangle \cong U_3(3)$, $N_{\Gamma(1a)}(\langle a,b,c\rangle) \cong G_2(2)$ and $\Gamma(1a) \cong G_2(5)$ (Conway et al. [1985]).

But, by Theorem 5, $C_{U_1}(z_{12.34})$ is the intersection of U_1 and $H_{12.34}$, hence $C_{U_1}(z_{12.34}) = \langle a,b,T\rangle \cong 4S_4$.

$G_2(2)$ and $G_2(5)$ do not contain involution centralizers of this form, so

$$U_1 = \langle a,b,c\rangle \cong U_3(3) \ . \tag{3.13}$$

Since U_1, U_2, U_3, U_4 are conjugate to each other under Q, (a) follows. Furthermore, r and c are both contained in $\Gamma(3a)$ where we immediately establish the relations

$$r^3 = 1 \ , \qquad rcr = crc \tag{3.14}$$

while in $H_{12.34} \cong 2 \wedge A_{11}$ the elements a and b evidently commute with r:

$$ra = ar \ , \qquad rb = br \ . \tag{3.15}$$

By a result of Aschbacher and Hall [1973] the relations (3.9, 3.10, 3.11, 3.14, 3.15) form a presentation of the full Schur cover of the finite simple group $U_4(3) \cong O_6^-(3)$, so with the abbreviation

$$U_0 = \langle a,b,c,r\rangle \tag{3.16}$$

we get (because $U_4(3)$ is simple and $U_0 \neq 1$)

$$U_0' = U_0 \ , \qquad U_0/Z(U_0) \cong U_4(3) \tag{3.17}$$

and $Z(U_0)$ is a factor of the Schur multiplier $12 \times 3 \cong 4 \times 3^2$ of $U_4(3)$. To complete the proof of our theorem it remains to show that $U_0 = U$. First we have $a,b,c \in U_1 \leqq U$ and $r \in U_3 \leqq U$, so $U_0 \leqq U$.

The reverse inequality amounts to $U_i \leqq U_0$ for all $i \in \{1,2,3,4\}$. Clearly this is true for $U_1 = \langle a,b,c\rangle$.

The intersection of U_1 and U_3 contains the torus T as well as c. Since $\langle c,T\rangle \cong 4S_4$ is maximal in U_3 and centralizes $z_{13.24}$, while $r \in U_3$ does not, we get

$$U_3 = \langle c,T,r\rangle \leqq U_0 \ . \tag{3.18}$$

Let now $i=2$ or 4. The intersections of U_i with U_1 and U_3 are different maximal subgroups ($\cong 4S_4$) of $U_i \cong U_3(3)$ and therefore they together generate U_i. Since they are contained in $\langle U_1, U_3\rangle \leqq U_0$, this completes the proof of the required equality

$$U=\langle a,b,c,r\rangle \tag{3.19}$$

at the same time establishing the theorem.

Having chosen a suitable unitary basis, the matrices in $SU_4(3)$ corresponding to the elements a,b,c,r are found to be

$$a=\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1-i & -1+i \\ 0 & 0 & 1+i & 1+i \end{vmatrix}, \quad b=\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1-i & 1-i \\ 0 & 0 & -1-i & 1+i \end{vmatrix}. \tag{3.20}$$

$$c=\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-i & 0 & -1+i \\ 0 & 0 & 1 & 0 \\ 0 & 1+i & 0 & 1+i \end{vmatrix}, \quad r=\begin{vmatrix} 1+i & -1-i & 0 & 0 \\ 1-i & 1-i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}. \tag{3.21}$$

where i is a square root of -1 in $\mathbb{F}_9$.

The matrices in $SO_6^-(3)$ are given by

$$a=\begin{vmatrix} 1 & . & . & . & . & . \\ . & 1 & . & . & . & . \\ . & . & 1 & 1 & 2 & 2 \\ . & . & 2 & 1 & 1 & 2 \\ . & . & 1 & 2 & 1 & 2 \\ . & . & 1 & 1 & 1 & 1 \end{vmatrix}, \quad b=\begin{vmatrix} 1 & . & . & . & . & . \\ . & 1 & . & . & . & . \\ . & . & 1 & 1 & 1 & 1 \\ . & . & 2 & 1 & 2 & 1 \\ . & . & 2 & 1 & 1 & 2 \\ . & . & 2 & 2 & 1 & 1 \end{vmatrix}. \tag{3.22}$$

$$c=\begin{vmatrix} 1 & 2 & . & . & 1 & 1 \\ 1 & 1 & . & . & 1 & 2 \\ . & . & 1 & . & . & . \\ . & . & . & 1 & . & . \\ 2 & 2 & . & . & 1 & 2 \\ 2 & 1 & . & . & 1 & 1 \end{vmatrix}, \quad r=\begin{vmatrix} 1 & . & . & . & . & . \\ . & 1 & . & . & . & . \\ . & . & 1 & 2 & 2 & 2 \\ . & . & 1 & 1 & 2 & 1 \\ . & . & 1 & 1 & 1 & 2 \\ . & . & 1 & 2 & 1 & 1 \end{vmatrix}. \tag{3.23}$$

4. The Lyons Group as a Homomorphic Image of Γ

We now define elements a,b,c,d,x in Γ by

$$a=\tau\cdot x(3b,4c)^4\cdot x(4b,3c)^2\ , \tag{4.1}$$

$$b=x(3b,4c)^2\ , \tag{4.2}$$

$$c=x(4b,2c)^3\ , \tag{4.3}$$

$$d=x(1a,2b)\cdot x(2c,1a)^3\cdot x(1a,2b)\ , \tag{4.4}$$

$$x=x(3b,4c)\cdot x(4b,3c)\cdot x(3b,4c)\cdot\tau'\cdot x(4a,3b)\cdot x(3b,4c)^3\cdot x(4a,3b) \tag{4.5}$$

where τ and τ' are the torus elements

$$\tau = x(4b,2c)\cdot x(2b,4c)^4\cdot x(4b,2c)^2\cdot x(2b,4c)^2\,, \tag{4.6}$$

$$\tau' = x(1a,4b)^4\cdot x(4c,1a)\cdot x(1a,4b)^2\cdot x(4c,1a)^2\ . \tag{4.7}$$

We verified by computer that the images $\tilde{a},\tilde{b},\tilde{c},\tilde{d},\tilde{x}$ of a,b,c,d,x in the 111-dimensional representation (cf. Lemma 1) obey all of the relations of Sims [1973], and hence they generate the Lyons group.

Furthermore,

$$\langle \tilde{a},\tilde{b},\tilde{c},\tilde{d}\rangle \cong G_2(5) \tag{4.8}$$

while

$$\langle a,b,c,d\rangle \leqq \Gamma(1a) \cong G_2(5)\ . \tag{4.9}$$

Therefore

$$\langle a,b,c,d\rangle = \Gamma(1a) \tag{4.10}$$

$x\in Q$ by Theorem 4.c permutes the 36 root groups and corresponds to the automorphism $(12)(34)\cdot(ac)$ of the apartment.

$\langle a,b,c,d,x\rangle$ contains the 12 root groups in $\Gamma(1a)$ and, e.g.,

$$X(1a,2b)^x = X(1b,2c)\ . \tag{4.11}$$

Because of the Chevalley relations these 13 root groups generate Γ. This shows the validity of

Theorem 7. (a) $\langle a,b,c,d,x\rangle = \Gamma$;
(b) *Ly is a homomorphic image of* Γ.

Remark. To prove a relation in any subgroup Δ of Γ which is isomorphic to its image $\tilde{\Delta}$ in the representation, it is sufficient to check this relation for the appropriate 111-dimensional $\mathbb{F}_5$-matrices.

In particular this holds true for the Sims relations which are expressed in elements of Δ alone.

We may apply this to the following three subgroups:

$$\Delta_x = \Gamma(1a) \cong G_2(5)\ , \tag{4.12}$$

$$\Delta_c = H_{12.34} \cong 2 \wedge A_{11}\ , \tag{4.13}$$

$$\Delta_d = \langle \Gamma(L(2c,1a)), T\rangle \cong 5^{1+4}:4S_6\ . \tag{4.14}$$

The isomorphisms $\Delta_x \cong \tilde{\Delta}_x$ and $\Delta_c \cong \tilde{\Delta}_c$ have been verified in Theorems 1.a and 2.c, respectively.

$\Delta_d \cong \tilde{\Delta}_d$ follows immediately from Theorem 2.e and the fact that all 36 root groups are normalized by T.

These arguments suffice to prove the validity of all Sims relations except three.

We believe that the remaining relations also follow from our presentation of Γ, but we have not yet been able to show this.

5. Summary

The goal of this paper is to construct a root system for the Lyons group Ly in analogy to those of the Chevalley groups.

We make ample use of geometric properties of Ly.

We are confident that similar ideas can be applied to other (all ?) sporadic groups as well, perhaps in the long run leading to an understanding of these peculiar structures.

Concerning the geometry of the Lyons group itself, more information may be gained by a careful study of the 111-dimensional minimal representation over $\mathbb{F}_5$.

Some initial results in that direction have been obtained.

We hope to present them – together with a proof of the isomorphism of the group Γ (defined in Sect. 2) with Ly – in the near future.

Acknowledgements. We are deeply indebted to Prof. Dr. Joachim Neubüser (RWTH Aachen) and his collaborators who supplied us in a most generous way with informations on several finite groups, character tables etc.

Without their friendly support and encouragement this investigation could not have been carried through.

References

Aschbacher, M., Hall, M.: Groups generated by a class of elements of order 3. J. Alg. **24**, 591–612 (1973)

Carter, R.: Simple groups of Lie type. New York: Wiley Interscience 1972

Conway, J.H., Curtis, R.T., Norton, S.P., Parker, R.A., Wilson, R.A.: Atlas of finite groups. Oxford: Clarendon Press 1985

Humphreys, J.E.: Linear algebraic groups. (Graduate Texts in Mathematics, Vol. 21). Berlin Heidelberg New York: Springer 1975

Kantor, W.: Some geometries that are almost buildings. Eur. J. Comb. **2**, 239–247 (1981)

Lyons, R.: Evidence for a new finite simple group. J. Alg. **20**, 540–569 (1972)

Meyer, W., Neutsch, W.: Über 5-Darstellungen der Lyonsgruppe. Math. Ann. **267**, 519–535 (1984)

Meyer, W., Neutsch, W., Parker, R.: The minimal 5-representation of Lyons' sporadic group. Math. Ann. **272**, 29–39 (1985)

Sims, C.C.: The existence and uniqueness of Lyons' group, in: Finite Groups '72 (Gainesville Conference), 138–141, Amsterdam: North-Holland 1973

Wilson, R.A.: The subgroup structure of the Lyons group. Math. Proc. Cambr. Philos. Soc. **95**, 403–409 (1984)

Wilson, R.A.: The maximal subgroups of the Lyons group. Math. Proc. Cambr. Philos. Soc. **97**, 433–436 (1985

Woldar, A.J.: On the maximal subgroups of Lyons' group. Comm. Alg. **15**, 1195–1203 (1987)

Received April 1, 1988; in revised form June 13, 1988

Math. Ann. 283, 301–328 (1989)

Mathematische Annalen

Dubrovin Valuation Rings and Henselization

Adrian R. Wadsworth*,**

Department of Mathematics, University of California at San Diego, La Jolla, CA 92093, USA

1. Introduction: Dubrovin Valuation Rings

Valuation theory is increasingly being recognized as a useful tool in studying the arithmetic of finite dimensional division algebras. However, there are some serious obstacles in noncommutative valuation theory that have limited its application. Notably, a valuation on the center F of a division ring D (with $[D:F]<\infty$) need not extend to a valuation on D (though when it does extend, the extension is unique). In addition, even if one starts working with division rings, one is led inevitably to consider also matrices over division rings, which arise with tensor products or scalar extensions. But there is not yet a reasonable notion of valuation for matrix rings.

N.I. Dubrovin introduced a few years ago in [D 1] and [D 2] a generalized notion of valuation ring which overcomes both of these obstacles. By basing his approach on the notion of a place in the category of simple Artinian rings, he obtained a significantly larger class of rings than the classical valuation rings on division rings. For example, the Dubrovin valuation rings restricting to a discrete valuation ring V of the center are precisely the maximal orders over V (cf. (1.15) below). Although there is no actual valuation associated with Dubrovin's rings, they nonetheless possess many of the properties of valuation rings, and have excellent extension properties. In fact, if S is a central simple F-algebra (with $[S:F]<\infty$) and V is a valuation ring of the field F, then there is always a Dubrovin valuation ring B of S with $B\cap F=V$; moreover, any two such Dubrovin valuation rings are conjugate in S, hence isomorphic (cf. Theorem A in Sect. 2). Thus, Dubrovin valuation rings appear to be very natural objects for studying the internal structure of division algebras and central simple algebras.

---

* Supported in part by the National Science Foundation

** Some of the research for this paper was carried out while the author was visiting the Wilhelms-Westfälische Universität of Münster, West Germany and the Université Catholique de Louvain of Louvain-la-Neuve, Belgium. The author would like to thank the mathematicians at both universities for their kind hospitality

While the theory of Dubrovin valuation rings is still quite new, these rings have already proved useful in working with the classical noncommutative valuation rings. For example, some of the difficult theorems in [JW] and [M1] on division algebras over Henselian fields have been generalized in [M2] to results about Dubrovin valuation rings, and these more general theorems have turned to have easier and more natural proofs. I would expect many more applications of Dubrovin valuation rings as the theory is developed further.

We will here prove some theorems on the structure of Dubrovin valuation rings, and use Henselization techniques to show that they are closely related to the classical valuation rings on division rings. Among other things, we define a value group for Dubrovin valuation rings, prove an "Ostrowski theorem" for such rings, and give various characterizations of Dubrovin valuation rings integral over their centers. Our main results are all stated in Sect. 2, with the proofs given in Sects. 3–5.

I wish to thank K. Mathiak and J. Gräter for introducing me to Dubrovin valuation rings. I thank Gräter also for showing me his results with Brungs which appear in [BG2] – these were what convinced me to begin investigating Dubrovin's rings. In addition, I would like to thank J.-P. Tignol for some enlightening conversations. The main results of this paper were announced in [W2].

For the rest of this section, we recall the basic properties of Dubrovin valuation rings proved by Dubrovin in [D1] and [D2], and give some examples of such rings. This will provide some introduction to these rings, which may be unfamiliar to many readers. The properties quoted here will be used extensively later in the paper.

A few words on notation: If B is any ring, we write

B^* for the group of units of B;

$Z(B)$ for the center of B;

J(B) for the Jacobson radical of B;

$M_k(B)$ for the $k \times k$ matrix ring over B.

If S is an algebra over a field F, then $[S:F]$ denotes the dimension of S over F. The term "field" always means a commutative field. When there is a clearly defined monomorphism f from one object into another, we will routinely identify the domain of f with its image. Thus, canonical injections become inclusions, and canonical isomorphisms are written as equalities. This will occur particularly frequently in comparing residue rings and value groups of different valuation rings.

Dubrovin's definition in [D1] of a noncommutative valuation ring is based on the idea of a place in the category of simple Artinian rings. Let S be simple Artinian. We call a subring B of S a *Dubrovin valuation ring* (*of* S) if

(i) B has an ideal I such that B/I is simple Artinian;

(ii) for each $s \in S - B$ there are b_1, $b_2 \in B$, such that $b_1 s$, $s b_2 \in B - I$.

(Dubrovin called such a ring a noncommutative valuation ring, and the terminology in [BG2] is S-valuation ring.) Dubrovin showed [D1, Sect. 1, Proposition 3] that the ideal I is actually the Jacobson radical $J(B)$. Furthermore [D1, Sect. 1, Theorem 4], B is a left and right order in S, hence a prime left and right Goldie ring. In addition, it is easy to check that $Z(B) = B \cap Z(S)$ and $Z(B)$ is a valuation ring of the field $Z(S)$. Dubrovin proved the following further significant properties of a Dubrovin valuation ring B (cf. [D1, Sect. 1, Theorems 4, 7; Sect. 2, Theorem 4] and [D2, Sect. 1, Proposition 2; Sect. 2, Theorem 1]):

(1.1) (Bézout) Every finitely generated left (resp. right) ideal of B is principal.

(1.2) (semihereditary) Every finitely generated left (resp. right) ideal of B is projective as a B-module.

(1.3) B has the "k-chain property": There is an integer $k>0$ such that for any $n>k$ and any $a_1, \dots, a_n \in B$, the left ideal $Ba_1 + \dots + Ba_n$ is generated by k of the a_i. Likewise for right ideals. (The smallest such k turns out to be the matrix size of $B/J(B)$.)

(1.4) The two-sided ideals of B are linearly ordered by inclusion, while the left ideals (resp. right ideals) are in general not linearly ordered. Likewise, the $B-B$ bimodules of S are linearly ordered.

(1.5) (Morita invariance) If $S \cong M_m(D)$, then $B \cong M_m(C)$, where C is a Dubrovin valuation ring of D. Furthermore, for any natural number k, $M_k(B)$ is a Dubrovin valuation ring of $M_k(S)$. If $e \in B$ is idempotent ($e \neq 0$), then eBe is a Dubrovin valuation ring of eSe.

(1.6) (composition of places) If T is a subring of B with $J(B) \subseteq T$, then T is a Dubrovin valuation ring of S iff $T/J(B)$ is a Dubrovin valuation ring of the simple Artinian ring $B/J(B)$. (This property is particularly useful for building examples of Dubrovin valuation rings.)

(1.7) (overrings) Let A be any overring of B, i.e., a ring with $B \subseteq A \subseteq S$. Then, A is a Dubrovin valuation ring of S, $J(A)$ is a prime ideal of B, A is the left (and right) localization of B with respect to the elements of B regular mod $J(A)$, and $B/J(A)$ is a Dubrovin valuation ring of $A/J(A)$.

(1.8) (localization) Suppose $[S:F] < \infty$, where $F = Z(S)$. Let $V = B \cap F$. For any prime ideal Q of B, set $P = Q \cap V$. Then P is a prime ideal of V, and the central localization B_P (of B as a V-algebra with respect to $V-P$) is an overring of B with $J(B_P) = Q$. Thus, (combining this with (1.7)) there is a one-to-one correspondence between prime ideals of B, prime ideals of V, and overrings of B. (Distinct prime ideals P of V yield distinct overrings of B, as $B_P \cap F = V_P$.)

Consider, by way of comparison, the valuation rings R that arise from valuations on S when S is a division ring. Such a subring R of S is characterized by the following properties (cf. [S, p. 12]):

(1.9) (i) For every $s \in S^*$, $s \in R$ or $s^{-1} \in R$.
(ii) For every $s \in S^*$, $sRs^{-1} = R$.

We will call a ring R satisfying properties (1.9) (i) and (ii) an *invariant valuation ring* of the division ring S. (The name comes from (ii) – invariance under inner automorphisms.) As is easy and well-known, for such an R the left ideals are the same as the right ideals and are linearly ordered, $J(R)$ is the unique maximal ideal of R, and the residue ring $R/J(R)$ is a division ring. Also the associated valuation on S is completely determined by R: The value group of R is

$$\Gamma_R = S^*/R^*, \tag{1.10}$$

which is made into a (totally) ordered abelian group by setting $s_1R^* \leqq s_2R^*$ iff $s_1R \supseteq s_2R$. Then the associated valuation is $v: S^* \to \Gamma_R$ given by $s \mapsto sR^*$.) Clearly, every invariant valuation ring is a Dubrovin valuation ring.

A Dubrovin valuation ring B of a simple Artinian ring S need not be invariant, but it has the following features in common with invariant valuation rings. First, B has a *residue ring*

$$\bar{B}=B/J(B), \tag{1.11}$$

which is simple Artinian, though not necessarily a division ring. Second, although there is in general no valuation associated to B, we can still define a value group, as follows: Set

$$\mathrm{st}(B)=\{s\in S^*: sBs^{-1}=B\}, \tag{1.12}$$

the stabilizer of B under the action of S^*. Indeed, $\mathrm{st}(B)$ coincides with the normalizer of B^* in S^*. Then define the *value group* Γ_B of B by

$$\Gamma_B=\mathrm{st}(B)/B^*. \tag{1.13}$$

The elements of Γ_B are in one-to-one correspondence with the fractional two-sided ideals of the form $sB=Bs$ for $s\in \mathrm{st}(B)$. Note that Γ_B is an ordered group with respect to the ordering given by $sB^*\leqq tB^*$ iff $sB\supseteqq tB$ ($s,t\in\mathrm{st}(B)$). This is a total ordering, by property (1.4). Of course, if B is an invariant valuation ring of a division ring, then Γ_B coincides with the usual value group of B.

Besides invariant valuation rings, we note the following significant examples of Dubrovin valuation rings.

(1.14) *Example.* If V is any commutative valuation ring and A is any Azumaya algebra over V, then A is a Dubrovin valuation ring. (See (3.4) below or [D2, Sect. 2, Proposition 1]).

(1.15) *Example.* Let V be a discrete (rank 1) valuation ring of a field F, and let S be a simple algebra with $Z(S)=F$ and $[S:F]<\infty$. Let B be a subring of S with $B\cap F=V$. Then B is a Dubrovin valuation ring of S iff B is a maximal order of V in S.

Proof. Dubrovin proved in [D1, Sect. 1, Theorem 4] that B is a Dubrovin valuation ring of S iff $B/J(B)$ is simple Artinian, every finitely-generated left or right ideal of B is principal, and B is a left and right order of S. If B is a maximal order of V in S, then B has all of these properties by [Re, Theorem 18.7, p. 179]; so B is a Dubrovin valuation ring of S. Conversely, suppose B is Dubrovin, and $B\cap F=V$. Since $[S:F]<\infty$, B is a p.i.-ring. Because V is Noetherian, Formanek's theorem [F, Theorem 1] shows B is a finitely-generated V-module. In addition, $BK=S$, as B is prime p.i. (cf. [Rw, Theorem 1.7.9, p. 53]). Thus, B is an order of V in S, and the overring property (1.7) shows B is a maximal order. □

One may view Dubrovin valuation rings in finite dimensional central simple algebras as a reasonable generalization to arbitrary commutative valuation base rings of the notion of maximal orders over discrete valuation rings. The results stated in the next section are known for maximal orders over discrete valuation rings; it would be interesting to see how much further the rich theory of maximal orders can be extended to Dubrovin valuation rings.

2. Statement of Theorems

In this section we state the main theorems of this paper (the lettered theorems) and some of their corollaries. The theorems and corollaries will be proved in Sects. 3–5. The similarities in approach to several of the proofs and the intertwinings of the arguments dictate that the theorems be proved together rather than sequentially.

A crucial property of Dubrovin valuation rings is that they give essentially unique extensions of valuation rings of the center:

(2.1) **Theorem.** *Let S be a simple Artinian ring and let $F=Z(S)$, with $[S:F]<\infty$. If V is any valuation ring of F, then there is a Dubrovin valuation ring B of S with $B\cap F=V$.*

Theorem A. *With S, F as in Theorem 2.1, if B and B_0 are two Dubrovin valuation rings of S with $B\cap F=B_0\cap F$, then there is a $u\in S^*$ with $uBu^{-1}=B_0$.*

Theorem 2.1 was proved by Dubrovin in [D2, Sect. 3, Theorem 2], with a more understandable proof given in [BG2, Theorem 3.8]. Theorem A was proved for $B\cap F$ of finite rank by Brungs and Gräter in [BG2, Theorem 5.4]; the proof we give below will have no such restriction.

Our main theorem, Theorem B below, describes what happens to a Dubrovin valuation ring with passage to the Henselization of the valuation on the center. To place this in context, we recall the corresponding situation for invariant valuation rings. It has long been known (cf. [S, Theorem 9, p. 53]) that if V is a Henselian valuation ring of a field F, then inside any F-division algebra D with $[D:F]<\infty$ there is a unique invariant valuation ring R extending V (i.e., $R\cap F=V$). However, if V is not Henselian, then V has at most one extension to an invariant valuation ring of D, but possibly no extension at all (cf. [W1, Corollary] or [Er2, Corollary 1]). Now, every field with valuation ring (F, V) has an essentially unique Henselization (F^h, V^h). Recently, Morandi has shown that the Henselization can be used to determine whether V extends to D:

(2.2) **Theorem.** *Let D be a division ring, let $F=Z(D)$ with $[D:F]<\infty$, and let V be a valuation ring of F. Let (F^h, V^h) be the Henselization of (F, V). Then,*

(i) *V extends to an invariant valuation ring R of D iff $D\otimes_F F^h$ is a division ring.*

(ii) *Suppose $D\otimes_F F^h$ is a division ring. Let R^h be the (unique) extension of V^h to an invariant valuation ring of $D\otimes_F F^h$. Then $R=R^h\cap D$, $\Gamma_R=\Gamma_{R^h}$, and $\bar{R}=\overline{R^h}$.*

Theorem 2.2 is proved in [M1, Theorem 2]*. When $D\otimes_F F^h$ is not a division ring there is still a Dubrovin valuation ring B of D extending V; our main theorem provides an analogue to Theorem 2.2 for B.

Before stating Theorem B we must set up some more notation. For the rest of this section fix a Dubrovin valuation ring B of a simple Artinian ring S, and let $F=Z(S)$ and $V=B\cap F$, a valuation ring of F. We always assume that $[S:F]<\infty$. Let (F^h, V^h) be the Henselization of the valued field (F, V). In addition to the

* Part (i) of this theorem is asserted in [Er2], but Ershov's proof has a gap I do not know how to fill

terminology $\bar{B}$, st(B), Γ_B introduced in (1.11)–(1.13) we write

π_B for the natural projection: $B \to \bar{B} = B/J(B)$;

t_B = matrix size of $\bar{B}$ (i.e., $\bar{B} \cong M_{t_B}(E)$, where E is a division ring);

n_B = matrix size of $S \otimes_F F^h$;

$s_B = n_B/t_B$ (which we will see is always an integer).

For any F-algebra C, $\mathrm{Aut}_F C$ denotes the group of F-automorphisms of C. But if C is a field, we also write $\mathscr{G}(C/F)$ for the Galois group $\mathrm{Aut}_F C$.

With B as above, note that for any $s \in \mathrm{st}(B)$, conjugation by s is a V-automorphism of B, so it induces a $\bar{V}$-automorphism of $\bar{B}$. That is, there is a well-defined homomorphism

$$\varphi_B : \mathrm{st}(B) \to \mathrm{Aut}_{\bar{V}} \bar{B} \quad \text{given by} \quad \varphi_B(s)(\bar{b}) = \overline{sbs^{-1}},$$

where $\bar{b} = \pi_B(b)$, for $b \in B$. Every automorphism of $\bar{B}$ maps $Z(\bar{B})$ onto $Z(\bar{B})$. Further, if $s \in B^* \cdot F^*$ then $\varphi_B(s)$ is the identity on $Z(\bar{B})$. Thus, as $\mathrm{st}(B)/(B^* \cdot F^*) = \Gamma_B/\Gamma_V$, φ_B induces a homomorphism

$$\theta_B : \Gamma_B/\Gamma_V \to \mathscr{G}(Z(\bar{B})/\bar{V}).$$

For any V-algebra A and any prime ideal P of V, we write A_P for the localization of A at P; so $A_P = A \otimes_V V_P$.

We can now give our main theorem. Let B be a Dubrovin valuation ring and let S, F, V, F^h, V^h be as described above, with $[S:F] < \infty$, and write

$$S \otimes_F F^h \cong M_{n_B}(D^h),$$

where D^h is a division ring. Since V^h is Henselian, there is a unique invariant valuation ring R of D^h with $R \cap F^h = V^h$.

Theorem B. (i) *$\bar{B} \cong M_{t_B}(\bar{R})$ where $\bar{R}(=R/J(R))$ is a division ring.*

(ii) *$\Gamma_B = \Gamma_R$.*

(iii) *Using the isomorphism $Z(\bar{B}) \cong Z(\bar{R})$ induced by* (i) *above, we have a commutative diagram:*

$$\begin{array}{ccc} \Gamma_B/\Gamma_V & \xrightarrow{\cong} & \Gamma_R/\Gamma_V \\ {\scriptstyle\theta_B}\downarrow & & {\scriptstyle\theta_R}\downarrow \\ \mathscr{G}(Z(\bar{B})/\bar{V}) & \xrightarrow{\cong} & \mathscr{G}(Z(\bar{R})/\overline{V^h}) \end{array}.$$

Corollary B. *With the notation as above, the maps θ_B and φ_B are surjective and $Z(\bar{B})$ is a normal (but not necessarily separable) extension field of $\bar{V}$ of finite degree. If $Z(\bar{B})$ is separable over $\bar{V}$, then it is actually abelian Galois over $\bar{V}$. Γ_B is an abelian group and Γ_B/Γ_V is finite.*

Theorem B allows us to extend the "Ostrowski theorem" to Dubrovin valuation rings: For B, S, F, V as above, define the *defect* of B by

$$\delta(B) = [S:F]/([\bar{B}:\bar{V}]\,|\Gamma_B:\Gamma_V|\,(n_B/t_B)^2).$$

Theorem C. *For B and R as in Theorem B, $\delta(B) = \delta(R)$. Consequently, $\delta(B) = 1$ if* $\mathrm{char}(\bar{V}) = 0$, *and $\delta(B) = p^a$ for some integer $a \geq 0$ if* $\mathrm{char}(\bar{V}) = p \neq 0$.

Note that if B is an invariant valuation ring $n_B = t_B = 1$ (cf. Theorem 2.2), so Theorem C reduces to the Ostrowski theorem for such valuation rings proved for V Henselian by Draxl [Dr, Theorem 2] and in general by Morandi [M1, Theorem 3].

In order to prove Theorems A and B and also to clarify the relation between the integer invariants n_B and t_B we must consider other Henselizations – at the localizations of V. Let $\{P_i\}_{i\in I}$ be the (linearly ordered) set of nonzero prime ideals of V. For each $i\in I$, let (F_i, V_i) be the Henselization of (F, V_{P_i}), and let $S\otimes_F F_i \cong M_{m_i}(D_i)$, where D_i is a division ring. So each $m_i \leqq [S:F]$. Note that for $i, s\in I$, if $P_i \subsetneqq P_s$ then V_{P_s} is a refinement of V_{P_i}, so we may view $F_i \subseteqq F_s$ (cf. the discussion of Henselization in the next section); hence $m_i | m_s$. Thus, if $m\in\{m_i : i\in I\}$ there is a prime ideal P_j maximal such that $m_j = m$. (For, let $I_m = \{i\in I : m_i = m\}$. Then set $P_j = \bigcup_{i\in I_m} P_i$, which is a prime ideal as the P_i are linearly ordered. Since F_j is the direct limit of the F_i, $i\in I_m$, we have $m_j = m$.) We call such a P_j a *jump prime ideal* of V with respect to S. So, P_j is a jump prime ideal iff for each $P_s \supsetneqq P_j$, $m_s > m_j$. We write $j(V, S)$ for the number of jump prime ideals, and call this the *jump rank* of V (re S). The jump rank is a convenient invariant for induction arguments, since it is always finite, even when the usual rank (= Krull dimension) of V is infinite. Since the prime ideals of our Dubrovin valuation ring B correspond to the primes of V, we say Q is a jump prime ideal of B if $Q\cap V$ is a jump prime ideal of V re S.

Now, let A be any ring with $B\subseteqq A\subseteqq S$. Then we know (cf. (1.7)) that A is a Dubrovin valuation ring of S and is a localization of B. Let $W = A\cap F$, a localization of V. We set $\tilde{V} = V/J(W)$, which is a valuation ring of the field $\bar{W} = W/J(W)$. Now, $Z(\bar{A})$ is a field extension of $\bar{W}$ of finite degree (see below). Set

(2.3) $\quad \ell_{B,A}$ = the number of extensions of $\tilde{V}$ to valuation rings of $Z(\bar{A})$.

Theorem D. *With B, S, F as above, let $Q_1 \subsetneqq Q_2 \subsetneqq \ldots \subsetneqq Q_k$ be a finite set of prime ideals of V which includes all the jump primes of V re S. Set $\ell_i = \ell_{B_{Q_i}, B_{Q_{i-1}}}$. Then,*

$$n_B = t_B \ell_2 \ell_3 \ldots \ell_k.$$

The next theorem and its corollary are crucial for the inductive proofs of Theorems A, B, and D.

Theorem E. *With B, S, F as above, let A be a ring with $B\subseteqq A\subseteqq S$, and let $W = A\cap F$. Set $\tilde{B} = B/J(A)$, which is a Dubrovin valuation ring of $\bar{A}$. Then,*

(i) *There is an exact sequence*

$$(\Gamma_{B,A}) \qquad 0\to \Gamma_{\tilde{B}} \to \Gamma_B \to \Gamma_A \to \mathscr{G}(Z(\bar{A})/\bar{W})/H \to 0,$$

where $H = \{\tau\in\mathscr{G}(Z(\bar{A})/\bar{W}) : \tau(\tilde{B}\cap Z(\bar{A})) = \tilde{B}\cap Z(\bar{A})\}$.

(ii) $t_B = t_{\tilde{B}} \geqq t_A$.

(iii) $n_B = n_{\tilde{B}}(n_A/t_A)\ell_{B,A}$, *i.e.*, $s_B = s_{\tilde{B}} s_A \ell_{B,A}$.

Corollary E. *With B, S, F, A, $\tilde{B}$ as in Theorem E, let Q be a prime ideal of B with $Q\supseteqq J(A)$. If $Q/J(A)$ is a jump prime ideal of $\tilde{B}$, then Q is a jump prime ideal of B.*

The converse to Corollary E is not true in general: One can construct examples in which Q is a jump prime ideal of B even though $Q/J(A)$ is not a jump prime ideal of $\tilde{B}$.

The invariant t_B of a Dubrovin valuation ring B is bounded above by n_B and below by the matrix size of S. The rings achieving the extreme values of t_B are particularly interesting:

Theorem F. *The following are equivalent:*

(i) $t_B = n_B$.

(ii) *B is integral over V.*

(iii) *Every principal two-sided ideal of B is principal as a left ideal and as a right ideal of B.*

(iv) *Every two-sided ideal of B is generated by elements of* $\mathrm{st}(B)$.

(v) *For all rings A, E with $B \subseteq E \subsetneq A \subseteq S$, $\ell_{E,A} = 1$.*

(vi) *$B \otimes_V V^h$ is a Dubrovin valuation ring of $S \otimes_F F^h$.*

(vii) *There is a Dubrovin valuation ring B^h of $S \otimes_F F^h$ with $B^h \cap S = B$ and $B^h \cap F^h = V^h$.*

(Yet another condition, (vi′), equivalent to those in Theorem F will be given at the beginning of Sect. 4.)

The equivalence of conditions (iii) and (iv) to the others in Theorem F is due to Morandi [M2], who has also found further interesting properties and characterizations of Dubrovin valuation rings integral over their centers. Note that condition (iv) says that the two-sided ideals of B are classified by the value group of B, just as for a commutative valuation ring. So this holds iff B is integral over $Z(B)$. Observe also that whenever V has rank 1 the conditions of Theorem F all hold, as Theorem D shows $t_B = n_B$.

Theorem G. *Suppose (in addition to the standing hypotheses) that S is a division ring. Then the following are equivalent:*

(i) $t_B = 1$.

(ii) *For each $s \in S^*$, $s \in B$ or $s^{-1} \in B$.*

(iii) *B has only finitely many different conjugates in S.*

(iv) *The set T of elements of S integral over V is a ring.* $\left(\textit{In fact, } T = \bigcap_{s \in S^*} sBs^{-1}.\right)$

When these equivalent conditions hold, the number of conjugates of B is exactly n_B.

The rings described in Theorem G, which satisfy condition (i) of (1.9) but not (ii), have been studied extensively by Mathiak, Gräter, and others (cf. [Ma]). We call them *total valuation rings.* Mathiak has defined a value group for a total valuation ring B as $S^* \Big/ \bigcap_{s \in S^*} sB^*s^{-1}$. Gräter has proved [G, Theorem 3.4] that our value group Γ_B is isomorphic to the center of Mathiak's value group. Of the conditions in Theorem G, (ii)⇒(iv) and (ii)⇒(iii) were proved by Brungs and Gräter in [BG1, Theorems 1, 3], who also showed in [BG1, Theorem 1] that the number of conjugates of B satisfying (ii) is bounded by $\sqrt{[S:F]}$. The easy equivalence (i)⇔(ii) appears in [BG2, Lemma 2.2].

The following corollary is immediate from Theorems 2.2, F, and G:

Corollary G. *Suppose (in addition to the standing hypotheses) that S is a division ring. Then the following are equivalent:*

(i) $n_B = 1$.

(ii) *B is an invariant valuation ring.*

(iii) *B is a total valuation ring and B is integral over V.*

3. Preliminaries

In this section we give preliminary results to prepare for the proofs of the theorems stated in Sect. 2. We prove some special cases of these theorems, in preparation for the main argument, which begins in Sect. 4.

The following notation will be fixed throughout this section: B is a Dubrovin valuation ring of a simple Artinian ring S, $F = Z(S)$, and $V = B \cap F$. We adopt as a standing hypothesis that $[S:F] < \infty$. Note that (F, V) is a valued field, i.e., F is the quotient field of the valuation ring V. We will write $(F, V) \subseteq (F', V')$ if (F', V') is another valued field, with $F \subseteq F'$ and $V' \cap F = V$.

We begin with some general lemmas that prepare the way for a useful result (Proposition 3.3) on extending Dubrovin valuation rings.

(3.1) **Lemma.** *Suppose (K, Y) is a valued field, T is a K-algebra (containing K), and C is a subring of T with $C \cap K = Y$. Let $N \subseteq M$ be Y-modules. Then,*

(a) *M is a flat Y-module iff M is torsion-free.*

(b) *$N \otimes_Y C \subseteq M \otimes_Y C$, $M \subseteq M \otimes_Y C$, and $(N \otimes_Y C) \cap M = N$.*

Proof. (a) This is well-known, and is actually true if Y is an invariant valuation ring of a division ring K (with $[K:Z(K)] < \infty$) – cf. [JW, Sect. 2].

(b) Since C is a torsion-free, hence flat Y-module we may identify $N \otimes_Y C$ with its image in $M \otimes_Y C$. We have the exact sequence

$$\mathrm{Tor}_1^Y(M, C/Y) \to M \otimes_Y Y \to M \otimes_Y C.$$

Since $C \cap K = Y$, C/Y is a torsion-free, so flat, Y-module; hence $\mathrm{Tor}_1^Y(M, C/Y) = 0$. This shows the map $M = M \otimes_Y Y \to M \otimes_Y C$ is injective. Now set $N_1 = (N \otimes_Y C) \cap M \supseteq N$. Then N_1/N is the kernel of

$$M/N = (M/N) \otimes_Y Y \to (M/N) \otimes_Y C = (M \otimes_Y C)/(N \otimes_Y C).$$

Since $\mathrm{Tor}_1^Y(M/N, C/Y) = 0$, we have $N_1 = N$, as desired. □

(3.2) **Lemma.** *For any Dubrovin valuation ring B,* $\mathrm{st}(B) \cap (B - J(B)) = B^*$.

Proof. The inclusion $\supseteq$ is clear. For the reverse inclusion, take $s \in \mathrm{st}(B) \cap (B - J(B))$. Then $BsB = B$ as the ideals of B are linearly ordered and $B/J(B)$ is simple. But $BsB = Bs = sB$ as $s \in \mathrm{st}(B)$. Hence, $s \in B^*$. □

Let B be our Dubrovin valuation ring of S, and let B' be a Dubrovin valuation ring of a simple Artinian ring S', where $S \subseteq S'$. We say that B' is a *compatible extension* of B if $B' \cap S = B$, $J(B) \subseteq J(B')$, and $\mathrm{st}(B) \subseteq \mathrm{st}(B')$. When this occurs, $J(B') \cap B = J(B)$, as $J(B)$ is the maximal ideal of B, and we view $\bar{B} \subseteq \bar{B}'$ via the natural inclusion. Furthermore, clearly $\mathrm{st}(B') \cap S = \mathrm{st}(B)$ and $B'^* \cap \mathrm{st}(B) = B^*$; thus, the obvious homomorphism $\Gamma_B \to \Gamma_{B'}$ is injective and order-preserving, and we view $\Gamma_B \subseteq \Gamma_{B'}$. We say that B' is an *immediate compatible extension* of B if B' is a compatible extension of B with $\bar{B}' = \bar{B}$ and $\Gamma_{B'} = \Gamma_B$.

(3.3) **Proposition.** *With B, S, F, V as above, let K be a field with $K \subseteq F$ and $[F:K] < \infty$, and let $Y = V \cap K$. Let T be a simple K algebra with $[T:K] < \infty$. Suppose T contains a ring $C \supseteq Y$ which is a free Y-module of rank $[T:K]$. Set $\bar{C} = C/J(Y)C$. Suppose $S \otimes_K T$ and $\bar{B} \otimes_{\bar{Y}} \bar{C}$ are simple rings. Then $B \otimes_Y C$ is a Dubrovin valuation ring of $S \otimes_K T$ and is a compatible extension of B. Moreover, $\overline{B \otimes_Y C} = \bar{B} \otimes_{\bar{Y}} \bar{C}$ and $\Gamma_{B \otimes_Y C} = \Gamma_B$.*

Proof. Let $B' = B \otimes_Y C$ and $S' = S \otimes_K T = S \otimes_Y C$. We have $B' \subseteq S'$. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a base of C as a free Y-module. Then $\{\alpha_1, \ldots, \alpha_n\}$ is also an F-base of T. Take any $\gamma \in S'$. Then γ has a unique representation

$$\gamma = \sum s_i \otimes \alpha_i$$

with $s_i \in S$; note that $\gamma \in B'$ iff each $s_i \in B$. Let $\varrho : B \otimes_Y C \to \bar{B} \otimes_{\bar{Y}} \bar{C}$ be the canonical epimorphism, mapping $\sum s_i \otimes \alpha_i$ to $\sum \bar{s}_i \otimes \bar{\alpha}_i$ for $s_i \in B$. Let $J = \ker(\varrho)$. Since $\bar{\alpha}_1, \ldots, \bar{\alpha}_n$ is a $\bar{V}$-base of $\bar{C}$, $\gamma \in J$ iff each $\bar{s}_i = 0$, iff each $s_i \in J(B)$. By hypothesis B'/J is simple, and is Artinian since it is a finitely-generated $\bar{B}$-module.

To verify that B' is Dubrovin, take any $\gamma = \sum s_i \otimes \alpha_i \in S' - B'$. That is, some $s_i \in S - B$. Now, as B is Dubrovin, $\sum s_i B = rB$ for some $r \in S - R$ (cf. (1.1)). Write $s_i = rb_i$ with $b_i \in B$ and $r = \sum s_i c_i$ with $c_i \in B$. We have

$$\gamma = \sum s_i \otimes \alpha_i = (r \otimes 1)(\sum b_i \otimes \alpha_i).$$

Since B is Dubrovin there is a $d \in R$, such that $dr \in B - J(B)$. Then $d \otimes 1 \in B'$ and $(d \otimes 1)\gamma = \sum drb_i \otimes \alpha_i$. Note that not all $drb_i \in J(B)$, as otherwise $dr = \sum drb_i c_i \in J(B)$, contradicting the choice of d. Hence, $(d \otimes 1)\gamma \in B' - J$. By a symmetric argument there is a $d' \otimes a \in B'$ with $\gamma(d' \otimes 1) \in B' - J$. Thus, B' is a Dubrovin valuation ring with $J(B') = J$ and $\bar{B}' \cong \mathrm{im}(\varrho) = \bar{B} \otimes_{\bar{Y}} \bar{C}$.

Note that $C \cap K = Y$ as C is integral over Y. Hence by (3.1b), $B' \cap S = B$. Since the inclusions $J(B) \subseteq J(B')$ and $\mathrm{st}(B) \subseteq \mathrm{st}(B')$ are clear, B' is a compatible extension of B.

To show $\Gamma_{B'} = \Gamma_B$ it suffices to verify $\mathrm{st}(B') = \mathrm{st}(B) \cdot B'^*$. For this, take any $\gamma \in \mathrm{st}(B')$ and write $\gamma = (r \otimes 1)(\sum b_i \otimes \alpha_i)$ just as before. Then each $b_i \in B$. Suppose every $b_i \in J(B)$. Then as $r = \sum rb_i c_i$ with $c_i \in B$, we have $0 = r(1 - \sum b_i c_i)$. Since $1 - \sum b_i c_i \in 1 + J(B) \subseteq B^*$, we have $r = 0$, so $\gamma = 0$, a contradiction. Hence, some $b_i \notin J(B)$; thus we have expressed $\gamma = (r \otimes 1)\varepsilon$, where $\varepsilon = \sum b_i \otimes \alpha_i \in B' - J(B')$, and $r \in S^*$ as $\gamma \in S'^*$. Take any $b \in B$. Then $(b \otimes 1)\varepsilon \in B'$, so B' contains

$$\gamma(b \otimes 1)\varepsilon\gamma^{-1} = (r \otimes 1)\varepsilon(b \otimes 1)(r^{-1} \otimes 1) = \sum rb_i br^{-1} \otimes \alpha_i.$$

So, $rb_i br^{-1} \in B$, for all $b \in B$ and all i. Hence,

$$B \supseteq \sum rb_i Br^{-1} = (\sum s_i B)r^{-1} = rBr^{-1}.$$

Because $rBr^{-1} \cap F = B \cap F$, this inclusion implies by (1.8) that $rBr^{-1} = B$. Hence, $r \in \mathrm{st}(B)$. Thus, $\varepsilon = (r \otimes 1)^{-1}\gamma \in \mathrm{st}(B') \cap (B' - J(B'))$; by Lemma 3.2, $\varepsilon \in B'^*$. Therefore, $\gamma = (r \otimes 1)\varepsilon \in \mathrm{st}(B) \cdot B'^*$, as desired. $\square$

The first two corollaries are known ([D2, Sect. 2, Proposition 1] and (1.5) above) except for the value group formulas.

(3.4) **Corollary.** *Let C be any Azumaya algebra over any commutative valuation ring V. Then C is a Dubrovin valuation ring of its ring of quotients and $\Gamma_C = \Gamma_V$.*

Proof. Proposition 3.3 applies with $B = V$ and $K = F$. □

(3.5) **Corollary.** *For a Dubrovin valuation ring B and any n, $M_n(B)$ is a Dubrovin valuation ring with $\overline{M_n(B)} = M_n(\bar{B})$ and $\Gamma_{M_n(B)} = \Gamma_B$.*

Proof. Apply. Proposition 3.3 with $C = M_n(V)$. □

(3.6) **Corollary.** *With the notation defined before* (3.1) *let K be a subfield of F with $[F:K] < \infty$, and let $Y = V \cap K$. Let $(K, Y) \subseteq (N, U)$ with N algebraic over K, and N and F linearly disjoint over K. Suppose,*

(i) *for each field L with $K \subseteq L \subseteq N$ and $[L:K] < \infty$, $[\overline{U \cap L} : \bar{Y}] = [L:K]$;*
(ii) *$Z(\bar{B})$ and $\bar{U}$ are linearly disjoint over $\bar{Y}$.*

Then $B \otimes_Y U$ is a Dubrovin valuation ring of $S \otimes_K N$ which is a compatible extension of B, with $\overline{B \otimes_Y U} = \bar{B} \otimes_{\bar{Y}} \bar{U}$ and $\Gamma_{B \otimes_Y U} = \Gamma_B$.

Proof. Take any field L with $K \subseteq L \subseteq N$ and $[L:K] < \infty$, and let $C = U \cap L$. Since C has maximal residue degree over Y, C is a free Y-module and the ramification index must be 1; hence $J(C) = J(Y) \cdot C$. Consequently, $C/J(Y) \cdot C$ is the field $\bar{C} = C/J(C)$. Hypothesis (ii) assures that $Z(\bar{B})$ and $\bar{C}$ are linearly disjoint over $\bar{Y}$, so $Z(\bar{B}) \otimes_{\bar{Y}} \bar{C}$ is a field and $\bar{B} \otimes_{\bar{Y}} \bar{U} \cong \bar{B} \otimes_{Z(\bar{B})} (Z(\bar{B}) \otimes_{\bar{Y}} \bar{C})$ is simple. Likewise, $S \otimes_K L$ is simple, as L and $F = Z(S)$ are linearly disjoint over K. Thus, Proposition 3.3 applies to $B \otimes_Y C$. The desired properties of $B \otimes_Y U$ follow by easy direct limit arguments, as U is the direct limit of the C's. □

For arbitrary valued fields (F, V) the Henselization of V plays much the same rôle as the completion plays for rank 1 valuation rings. We recall now the facts we need about Henselization. (For background on Henselian valuation rings, see e.g. [R1], [R2], or [E].)

A *Henselization* of a valued field (F, V) is a valued field (F^h, V^h) such that $(F, V) \subseteq (F^h, V^h)$, V^h is Henselian, and for any Henselian valued field (F', V') with $(F, V) \subseteq (F', V')$ there is a unique F-homomorphism $f: F^h \to F'$ such that $f^{-1}(V') = V^h$. Every valued field (F, V) has a Henselization which is unique up to F-isomorphism (cf. [E, p. 131]). We use the notation $(F^h, V^h) = (F, V)^H$ to indicate that (F^h, V^h) is a Henselization of (F, V). It is known [E, (17.11), (17.19)] that F^h is separable over F and V^h is an immediate extension of V, i.e., $\overline{V^h} = \bar{V}$ and $\Gamma_{V^h} = \Gamma_V$.

Now, supposee $(F^h, V^h) = (F, V)^H$, and let K be an algebraic extension field of F and Y an extension of V to K. Then (cf. [E, (17.16), (17.13)]),

(3.7) *there is a compositum $K \cdot F^h$ of K and F^h over F such that $(K \cdot F^h, Y') = (K, Y)^H$, where Y' is the unique extension of V^h to $K \cdot F^h$. Moreover, there is a 1–1 correspondence between the extensions of V to K and the equivalence classes of composites of K and F^h over F.*

(The references mentioned in [E] cover only K/F separable in (3.7). But the generalization to K/F algebraic is easy.) Two extreme cases of (3.7) will arise frequently: As F^h is separable over F, K is linearly disjoint to F^h over F iff V has a unique extension to K. On the other hand, if $[K:F]<\infty$, and V has $[K:F]$ different extensions to K, then there is an embedding $(K, Y)\subseteq(F^h, V^h)$. For, the composites of K over F are the indecomposable summands of $K\otimes_F F^h$. In this case there are $[K:F]$ summands, so each is F^h.

We will exploit heavily the relationship between Henselizations of comparable valuation rings of the same field. Let (F, V) be a valued field, let $P\neq(0)$ be a prime ideal of V, and let $W=V_P$. Let $(F_1, W_1)=(F, W)^H$ and $(F_2, V_2)=(F, V)^H$. Let $W_2=V_{2P}$, i.e., the localization of V_2 at P. This W_2 is the valuation ring of F_2 with $V_2\subseteqq W_2$ and $W_2\cap F=W$. Since V_2 is Henselian, so is W_2 (cf. [R1, p. 210, Proposition 9]). Therefore, by the universal mapping property for the Henselization, we may view $(F_1, W_1)\subseteqq(F_2, W_2)$. Let $V_1=V_2\cap F_1\subseteqq W_1$; let $\tilde{V}=V/J(W)$, a valuation ring of $\bar{W}=W/J(W)$, and likewise let $\tilde{V}_i=V_i/J(W_i)\subseteqq\bar{W}_i$, $i=1,2$. Observe that when we identify $\bar{W}_1=\bar{W}$, then $\tilde{V}_1=\tilde{V}$. Now view (F_2, W_2) as an extension of (F_1, W_1). The key fact we will use is:

(3.8) *(F_2, W_2) is the (unique) unramified extension of (F_1, W_1) such that $(\bar{W}_2, \tilde{V}_2)=(\bar{W}_1, \tilde{V}_1)^H$.*

This can be verified using the universal mapping property for Henselization (as in the proof of [M1, Theorem 2]) or by an argument using decomposition groups. That (F_2, W_2) is unramified over (F_1, W_1) means $\bar{W}_2$ is separable over $\bar{W}_1$ and that for each field K, $F_1\subseteqq K\subseteqq F_2$ with $[K:F_1]<\infty$, $[\overline{W_2\cap K}:\bar{W}_1]=[K:F_1]$.

Certain special cases of the theorems and corollaries in Sect. 2 are needed in order to prove the theorems in general. We now give those partial results:

(3.9) **Proposition.** (a) *If Theorem B holds for a Dubrovin valuation ring B, then Corollary B and Theorem C hold for B.*

(b) *Theorem E (i) holds for the ring B if Theorem B holds for the ring A and Theorem A holds for $\tilde{B}$ in $\bar{A}$.*

(c) *Theorem E (ii) holds.*

(d) *Theorem E (iii) holds for the ring B if Theorem B holds for the ring A.*

(e) *Corollary E holds for the ring B if Theorem B holds for the ring A.*

(f) *Implications* (vi)$\Rightarrow$(vii) *and* (vii)$\Rightarrow$(ii) *of Theorem F hold for any Dubrovin valuation ring B.*

Before proving Proposition 3.9, we recall a result we need of Brungs and Gräter [BG2, Lemma 4.1]. (This is a special case of Theorem A).

(3.10) **Lemma.** *Let D be a division ring with $[D:Z(D)]<\infty$, and let B be a Dubrovin valuation ring of D, and R an invariant valuation ring of D. If $B\cap Z(D)=R\cap Z(D)$, then $B=R$.*

Proof of Proposition 3.9. (a) The properties in Corollary B are all known for any invariant valuation ring of a finite-dimensional division algebra (cf. [JW, Proposition 1.7], or for the Henselian case, see [DK, p. 96] or [Er1, Proposition 1]). Since these properties hold for the R in Theorem B, the theorem assures that they also hold for B. (Note that the surjectivity of θ_B implies the surjectivity of φ_B, since the Skolem-Noether Theorem shows $\mathrm{Aut}_{Z(\bar{B})}\bar{B}\subseteqq\mathrm{im}(\varphi_B)$.)

The fact that Γ_B is abelian also follows from Dubrovin's analysis of the ideals in a Dubrovin valuation ring [D2, Sect. 2, Proposition 4]. As for Theorem C, assuming Theorem B we have

$$\begin{aligned}\delta(B) &= [S:F]/([\bar{B}:\bar{V}]\,|\Gamma_B:\Gamma_V|\,(n_B/t_B)^2)\\ &= n_B^2[D^h:F^h]/(t_B^2[\bar{R}:\overline{V^h}]\,|\Gamma_R:\Gamma_{V^h}|\,(n_B/t_B)^2)\\ &= \delta(R),\end{aligned}$$

as $n_R = t_R = 1$. The Ostrowski theorem holds for $\delta(B)$ since it holds for $\delta(R)$ by [Dr, Theorem 2].

(b) We have Dubrovin valuation rings B, A with $B \subseteq A \subseteq S$ and $\tilde{B} \subseteq \bar{A}$. The assumptions are that Theorem B holds for A and that any Dubrovin valuation ring of $\bar{A}$ contracting to $\tilde{B} \cap Z(\bar{A})$ in $Z(\bar{A})$ is conjugate to $\tilde{B}$. Note that Corollary B holds for A by (a) above. The maps φ_A, π_A, θ_A are the ones defined in Sect. 2 (but now for A). Let $W = A \cap F$, and let $Y = \tilde{B} \cap Z(\bar{A})$, a valuation ring of $Z(\bar{A})$.

Observe that $\mathrm{st}(B) \subseteq \mathrm{st}(A)$ as A is a central localization of B by (1.8). Note that if $s \in \mathrm{st}(A)$, then $s \in \mathrm{st}(B)$ iff $\varphi_A(s)(\tilde{B}) = \tilde{B}$. This holds since $B = \pi_A^{-1}(\tilde{B})$. The inclusion $\mathrm{st}(B) \hookrightarrow \mathrm{st}(A)$ induces a homomorphism $\alpha: \Gamma_B \to \Gamma_A$ with image $A^* \cdot \mathrm{st}(B)/A^*$. We claim that

$$A^* \cdot \mathrm{st}(B) = \{s \in \mathrm{st}(A) : \varphi_A(s)(Y) = Y\}.$$

For, the inclusion $\subseteq$ is clear. To see $\supseteq$ take any $s \in \mathrm{st}(A)$ such that $\varphi_A(s)(Y) = Y$. By Theorem A, $\tilde{B}$ and $\varphi_A(s)(\tilde{B})$ are conjugate in $\bar{A}$, so there is an $a \in A^*$ with $\varphi_A(a)(\tilde{B}) = \varphi_A(s)(\tilde{B})$. Then $a^{-1}s \in \mathrm{st}(B)$ as $\varphi_A(a^{-1}s)(\tilde{B}) = \tilde{B}$. Hence, $s = a(a^{-1}s) \in A^* \cdot \mathrm{st}(B)$, establishing the claim.

Now, the map $\theta_A: \Gamma_A/\Gamma_W \to \mathscr{G}(Z(\bar{A})/\bar{W})$ is surjective by Corollary B. The claim just proved shows the image of $A^* \cdot \mathrm{st}(B)$ in Γ_A/Γ_W is $\theta_A^{-1}(H)$, where $H = \{\tau \in \mathscr{G}(Z(\bar{A})/\bar{W}) : \tau(Y) = Y\}$. This yields the exactness of the diagram of Theorem E (i) at Γ_A and at $\mathscr{G}(Z(\bar{A})/\bar{W})/H$.

The kernel of α is $(\mathrm{st}(B) \cap A^*)/B^*$. Now, the restriction of π_A to A^* is a group epimorphism: $A^* \to \bar{A}^*$ with kernel $1 + J(A) \subseteq 1 + J(B) \subseteq B^*$. Clearly $\pi_A(\mathrm{st}(B) \cap A^*) = \mathrm{st}(\tilde{B})$ and $\pi_A(B^*) = \tilde{B}^*$. Hence, via π_A, $\ker(\alpha) \cong \mathrm{st}(\tilde{B})/\tilde{B}^* = \Gamma_{\tilde{B}}$, showing the diagram is exact at Γ_B and $\Gamma_{\tilde{B}}$. This proves Theorem E (i).

(c) With B, A, $\tilde{B}$ as in Theorem E, note that $t_B = t_{\tilde{B}}$ as $\bar{\tilde{B}} = \bar{B}$. Also, $t_{\tilde{B}} \geqq t_A$ because $\tilde{B}$ is $t_A \times t_A$ matrices over some Dubrovin valuation ring, by (1.5) applied to $\tilde{B}$ in $\bar{A}$.

(d) With B, A, $\tilde{B}$, W as in Theorem E, let $\tilde{V} = V/J(W) = \tilde{B} \cap \bar{W}$, a valuation ring of $\bar{W}$. Let $Z(\bar{A})_{\mathrm{sep}}$ denote the separable closure of $\bar{W}$ in $Z(\bar{A})$, which by Corollary B is abelian Galois over $\bar{W}$. (Corollary B holds for A by (a) above since we are assuming Theorem B holds for A.) Let $L' \subseteq Z(\bar{A})_{\mathrm{sep}}$ be the decomposition field of $\tilde{B} \cap Z(\bar{A})_{\mathrm{sep}}$ over $\tilde{V}$, i.e., the fixed field of

$$\{\tau \in \mathscr{G}(Z(\bar{A})_{\mathrm{sep}}/\bar{W}) : \tau(\tilde{B} \cap Z(\bar{A})_{\mathrm{sep}}) = \tilde{B} \cap Z(\bar{A})_{\mathrm{sep}}\}.$$

Because $Z(\bar{A})$ is normal over $\bar{W}$,

$$\begin{aligned}[L':\bar{W}] &= \text{the number of extensions of } \tilde{V} \text{ to } Z(\bar{A})_{\mathrm{sep}}\\ &= \text{the number of extensions of } \tilde{V} \text{ to } Z(\bar{A})\\ &= \ell_{B,A}.\end{aligned}$$

Because $Z(\bar{A})$ and L' are normal over $\bar{W}$ and $\tilde{B}\cap L'$ extends uniquely to $Z(\bar{A})$, each extension of $\tilde{V}$ to L' extends uniquely to $Z(\bar{A})$. So, there are $\ell_{B,A}$ extensions of $\tilde{V}$ to L'. Set $\ell=\ell_{B,A}$.

Let $(F_1, W_1)=(F, W)^H$, and write $S\otimes_F F_1\cong M_{n_A}(D_1)$, where D_1 is an F_1-central division ring. After identifying $\bar{W}=\bar{W_1}$, let $V_1=\pi_{W_1}^{-1}(\tilde{V})$, a valuation ring of F_1 with $V_1\cap F=V$. Let A_1 be the invariant valuation ring of D_1 such that $A_1\cap F_1=W_1$, which exists as W_1 is Henselian. By Theorem B for A, $\bar{A}\cong M_{t_A}(\bar{A_1})$. We identify $Z(\bar{A_1})$ with $Z(\bar{A})$.

Because W_1 is Henselian there is an "inertial lift" L of L' in D_1 (cf. [JW, proof of Theorem 2.9]); that is, L is a field, $F_1\subseteq L\subseteq D_1$, with L separable over F_1, such that $[L:F_1]=[L':\bar{W}]=\ell$, and, setting $W_L=A_1\cap L$, $\bar{W_L}=L'$ in $Z(\bar{A_1})$. Let $D_L=C_{D_1}(L)$, the centralizer of L in D_1, and let $A_L=A_1\cap D_L$, an invariant valuation ring of D_L. Since $\bar{W_L}$ lies in $Z(\bar{A_1})$ and is separable over $\bar{W_1}$, [JW, Lemma 1.8(a)] shows that $\bar{A_L}=\bar{A_1}$. Let $\tilde{V_L}=\tilde{B}\cap\bar{W_L}$, and let $V_L=\pi_{W_L}^{-1}(\tilde{V_L})$, a valuation ring of L with $V_L\cap F_1=V_1$.

Let $(F_2, V_2)=(F, V)^H$ and let $W_2=V_{2J(W)}$ (the localization of V_2 such that $W_2\cap F=W$). As noted before (3.8) we may view $(F_1, W_1)\subseteq(F_2, W_2)$ with $V_2\cap F_1=V_1$. Let $\tilde{V_i}=V_i/J(W_i)$, $i=1,2$. Then (3.8) says (F_2, W_2) is the unramified extension of (F_1, V_1) such that $(\bar{W_2}, \tilde{V_2})=(\bar{W_1}, \tilde{V_1})^H$. Because $\tilde{V_1}$ has $[L:F_1]$ extensions to $\bar{W_L}=L'$, V_1 has $[L:F_1]$ extensions to L. So, by (3.7), as $(F_2, V_2)=(F_1, V_1)^H$, we may view $(L, V_L)\subseteq(F_2, V_2)$.

Set $A_3=A_L\otimes_{W_L}W_2$, a subring of $S_3=D_L\otimes_L F_2$. To show A_3 is a Dubrovin valuation ring of S_3, we invoke Corollary 3.6 with $B=A_L$ (which is invariant, hence Dubrovin), $K=F=L$, $Y=W_L$, and $(N, U)=(F_2, W_2)$. Condition (i) of (3.6) holds as (F_2, W_2) is unramified over (L, W_L). Also $(\bar{W_2}, \tilde{V_2})=(\bar{W_L}, \tilde{V_L})^H$, since

$$(\bar{W_1}, \tilde{V_1})\subseteq(\bar{W_L}, \tilde{V_L})\subseteq(\bar{W_2}, \tilde{V_2})=(\bar{W_1}, \tilde{V_1})^H.$$

Because V_L extends uniquely from $\bar{W_L}$ to $Z(\bar{A_L})$ (recalling $\bar{W_L}=L'$ and $Z(\bar{A_L})=Z(\bar{A_1})=Z(\bar{A})$), by (3.7) $Z(\bar{A_L})$ and $\bar{W_2}$ are linearly disjoint over $\bar{W_L}$. Thus, by Corollary 3.6 A_3 is a Dubrovin valuation ring of S_3, with $\bar{A_3}=\bar{A_L}\otimes_{\bar{W_L}}\bar{W_2}$.

Let $S_3\cong M_k(D_3)$, where D_3 is a division ring with $Z(D_3)=Z(S_3)=F_2$. Then by the Morita property (1.5) $A_3\cong M_k(A'_3)$, where A'_3 is a Dubrovin valuation ring of D_3 with $A'_3\cap F_2=A_3\cap F_2=W_2$. Because W_2 is Henselian, there is an invariant valuation ring of D_3 extending W_2; by (3.10) that invariant ring is A'_3. Hence $\bar{A'_3}$ is a division ring, showing that k is the matrix size of $\bar{A_3}$.

To compute $n_{\tilde{B}}$ we need a Henselization of $Z(\bar{A})$ with respect to $\tilde{B}\cap Z(\bar{A})$. Since $(\tilde{B}\cap Z(\bar{A}))\cap L'=\tilde{V_L}$ and $(L', \tilde{V_L})^H=(\bar{W_2}, \tilde{V_2})$, by (3.7) the field $Z(\bar{A})\otimes_{\bar{W_L}}\bar{W_2}$ is the desired Henselization. Because $\bar{A}\cong M_{t_A}(\bar{A_1})$,

$$\begin{aligned} n_{\tilde{B}}&=t_A\cdot\text{matrix size of } \bar{A_1}\otimes_{Z(\bar{A})}(Z(\bar{A})\otimes_{\bar{W_L}}\bar{W_2})\\ &=t_A\cdot\text{matrix size of } \bar{A_3}\qquad(\text{as } \bar{A_1}=\bar{A_L})\\ &=t_A\cdot k.\end{aligned}$$

But we also have

$$\begin{aligned} S\otimes_F F_2&\cong[(S\otimes_F F_1)\otimes_{F_1}L]\otimes_L F_2\\ &\cong M_{n_A}(D_1\otimes_{F_1}L)\otimes_L F_2\\ &\cong M_{n_A\ell}(D_L\otimes_L F_2)\\ &\cong M_{n_A\ell k}(D_3).\end{aligned}$$

Hence, $n_B = n_A \ell k = n_A \ell_{B,A} n_{\tilde{B}}/t_A$, proving Theorem E (iii).

(e) With the notation of Corollary E, let $E = B_{(Q \cap V)}$. Then E is a Dubrovin valuation ring of S with $B \subseteq E \subseteq A$ (cf. (1.8)), and $\tilde{B} = B/J(A) \subseteq \tilde{E} = E/J(A)$ are Dubrovin valuation rings of $\bar{A}$ with $J(\tilde{E}) = Q/J(A)$. Since we are assuming Theorem B holds for A, we have just shown Theorem E (iii) holds. By applying this theorem for B in A and again for E in A, we obtain

$$n_B/n_E = (n_{\tilde{B}}/n_{\tilde{E}}) \cdot (\ell_{B,A}/\ell_{E,A}). \tag{3.11}$$

Observe that $\ell_{B,A} \geqq \ell_{E,A}$ since $\tilde{V} = V/J(W)$, as a refinement of $\tilde{T} = (E \cap F)/J(W)$, has at least as many extensions to $Z(\bar{A})$ as does $\tilde{T}$. In addition, $n_B \geqq n_E$ since the Henselization of F with respect to $E \cap F$ lies in the Henselization with respect to V. Likewise, $n_{\tilde{B}} \geqq n_{\tilde{E}}$. Thus, if $n_B = n_E$, then $n_{\tilde{B}} = n_{\tilde{E}}$ and $\ell_{B,A} = \ell_{E,A}$.

We prove the contrapositive of Corollary E. Suppose Q is not a jump prime ideal of B. Then there is a prime ideal Q' of B, $Q' \supsetneqq Q$, such that if $E' = B_{(Q' \cap V)}$, $n_{E'} = n_E$. We have $B \subseteq E' \subseteq E \subseteq A$. Formula (3.11) and the subsequent remarks are still valid with E' replacing B. Thus, $n_{E'} = n_E$ implies $n_{\tilde{E}'} = n_{\tilde{E}}$. Since $\tilde{E}' \supsetneqq \tilde{E}$ this implies that $J(\tilde{E})$ is not a jump prime ideal of $\tilde{B}$, as desired.

(f) For (vi) ⇒ (vii) of Theorem F, suppose $B \otimes_V V^h$ is a Dubrovin valuation ring of $S \otimes_F F^h$, where $(F^h, V^h) = (F, V)^H$. Set $B^h = B \otimes_V V^h$. Note that $S \otimes_F F^h = S \otimes_V V^h$. Then, $B^h \cap S = B$ by (3.1) (b), so $B^h \cap F = B \cap F = V$. Hence, $B^h \cap F^h$ is a ring containing V^h which contracts to V in F. Thus, $B^h \cap F^h = V^h$, yielding (vii) of Theorem F.

(vii) ⇒ (ii) Suppose there is a Dubrovin valuation ring B^h of $S \otimes_F F^h$ with $B^h \cap S = B$ and $B^h \cap F^h = V^h$. Write $S \otimes_F F^h = M_n(D^h)$, where D^h is a division ring and $n = n_B$. Let R be the invariant valuation ring of D^h with $R \cap F^h = V^h$. Then $B^h \cong M_n(R)$ by (1.5) and (3.10). Because R is integral over V^h (e.g., by [W1, Corollary]), R is locally finite, by Shirshov's theorem on integral p.i.-rings [Rw, pp. 206–207]. Hence, $B^h \cong M_n(R)$ is locally finite, so integral over V^h. For any $\beta \in B^h$, let f be the minimal (monic) polynomial of β over F^h, and let g be any monic polynomial in $V^h[X]$ with $g(\beta) = 0$. For any root γ of f in any extension field of F^h, $g(\gamma) = 0$, so γ is integral over V^h. Thus, the coefficients of f, as polynomials in the γ's are also integral over V^h. Hence, $f \in V^h[X]$ as V^h is integrally closed. Now, take any $b \in B$. The minimal polynomial of b over F is the same as its minimal polynomial (viewing $b \in S \otimes_F F^h$) over F^h. Because $B \subseteq B^h$, this polynomial has coefficients in $V^h \cap F = V$. Therefore, b is integral over V, proving (ii) of Theorem F. The proof of Proposition 3.9 is now complete. □

The next two lemmas will be used in proving that $B \otimes_V V^h$ is a Dubrovin valuation ring when B is integral over V.

(3.12) Lemma. *With B, S, F, V as defined before* (3.1), *let K be a subfield of F, $Y = V \cap K$, and let $(K, Y) \subseteq (K', Y')$. If $S \otimes_K K'$ is simple Artinian and $B \otimes_Y Y'$ is a Dubrovin valuation ring of $S \otimes_K K'$, then $B \otimes_Y Y'$ is a compatible extension of B. Furthermore, if $\bar{B} \otimes_{\bar{Y}} \bar{Y}'$ is simple, then $\overline{B \otimes_Y Y'} = \bar{B} \otimes_{\bar{Y}} \bar{Y}'$, and if also $J(Y) \cdot Y' = J(Y')$, then $J(B \otimes_Y Y') = J(B) \otimes_Y Y'$.*

Proof. Let $B' = B \otimes_Y Y'$, viewed as a subring of $S \otimes_K K' = S \otimes_Y Y'$. By (3.1), we have $B' \cap S = B$. Also, $J(B) \subseteq J(B) \otimes_Y Y' \subseteq J(B')$, as $J(B) \otimes_Y Y'$ is a proper ideal of B' by (3.1), while the linear ordering of ideals of B' (1.4) assures that $J(B')$ is the unique maximal ideal of B'. Clearly, $\mathrm{st}(B)\subseteq \mathrm{st}(B')$. So B' is a compatible extension of B. Consider the epimorphism $\gamma: B' \to \bar{B}\otimes_Y \overline{Y'} = \bar{B}\otimes_{\bar{Y}} \overline{Y'}$. If $\mathrm{im}(\gamma)$ is simple, then $\ker(\gamma)$ is the unique maximal ideal of B', which is $J(B')$, and $\overline{B'} = B'/J(B') = \mathrm{im}(\gamma)$. If further $J(Y)\cdot Y' = J(Y')$, then $\bar{B}\otimes_Y \overline{Y'} = \bar{B}\otimes_Y Y'$; so $J(B') = \ker(\gamma) = J(B)\otimes_Y Y'$, as desired. □

Here is the setup for the next lemma: Let $B\subseteq A$ be Dubrovin valuation rings of a simple Artinian ring S. Let $F = Z(S)$ (with $[S:F]<\infty$), $V = B\cap F$, $W = A\cap F$, and $\tilde{B} = B/J(A)$, a Dubrovin valuation ring of $\bar{A}$. Let K be a subfield of F with $[F:K]<\infty$ and let $U = V\cap K$, $X = W\cap K$, and $\tilde{U} = U/J(X)$. Let $(K', U') = (K, U)^H$ and $(K_1, X_1) = (K, X)^H$, and set $X' = X\cdot U'$, a valuation ring of K' with $X'\cap K = X$. As noted in (3.8) we may view $(K_1, X_1)\subseteq (K', X')$. Set $\tilde{U}' = U'/J(X')$. Note that by (3.8) $(\overline{X'}, \tilde{U}') = (\bar{X}, \tilde{U})^H$.

(3.13) **Lemma.** *In the setup just described, assume B is integral over U. If $A\otimes_X X_1$ and $\tilde{B}\otimes_{\tilde{U}} \tilde{U}'$ are both Dubrovin valuation rings, then $B\otimes_U U'$ is a Dubrovin valuation ring.*

Proof. Because B is integral over U, we must have A (a central localization of B) integral over X and $\tilde{B}$ integral over $\tilde{U}$. Recall [E, (13.3)] that the integral closure of X in F is the intersection of all the extensions of X to valuation rings of F. So, because $W\subseteq A$ is integral over X, W must be the unique extension of X to F. Hence $F\otimes_K K_1$ is a field by (3.7). This assures $S\otimes_K K_1$ is a simple Artinian ring. Let $S_1 = S\otimes_K K_1$, $A_1 = A\otimes_X X_1$, and $B_1 = B\otimes_U U_1$, where $U_1 = U'\cap K_1$. Let $P = J(X)\subseteq U$, so $A = B\cdot W = B_P$. Because $A = A_P$, $X = U_P$, and $X_1 = U_{1P}$, we have $A_1 = A\otimes_U U_1$ and likewise $J(A)\otimes_X X_1 = J(A)\otimes_U U_1$. Note also that $\bar{A}\otimes_{\bar{X}} \overline{X_1} = \bar{A}$, since $\overline{X_1} = \bar{X}$. Because we are assuming A_1 is a Dubrovin valuation ring, Lemma 3.12 (applied to A, S, F, W, ...) says $\overline{A_1} = \bar{A}$ and $J(A_1) = J(A)\otimes_X X_1 = J(A)\otimes_U U_1 \subseteq B_1$. Set $\tilde{U}_1 = U_1/J(X_1)$; so $\tilde{U}_1 = \tilde{U}$ in $A_1 = \bar{A}$. Then, as $J(X)\cdot U_1 = J(X)\cdot X_1 = J(X_1)$, we have $B_1/J(A_1) = \tilde{B}\otimes_U U_1 = \tilde{B}\otimes_{\tilde{U}} \tilde{U}_1 = \tilde{B}$, which is a Dubrovin valuation ring of $\overline{A_1}$. Hence, by (1.6) B_1 is a Dubrovin valuation ring of S_1. Set $\tilde{B}_1 = B_1/J(A_1) = \tilde{B}$.

Let $A' = A_1\otimes_{X_1} X'$ and $B' = B_1\otimes_{U_1} U' = B\otimes_U U'$. Since $(\overline{X'}, \tilde{U}') = (\overline{X_1}, \tilde{U}_1)^H$ and $\tilde{B}_1 = \tilde{B}$ is integral over $\tilde{U}_1 = \tilde{U}$, the argument used above to see S_1 is simple Artinian applies here to see that $\overline{A_1}\otimes_{\overline{X_1}} \overline{X'}$ is simple Artinian, i.e. $Z(\overline{A_1})$ and $\overline{X'}$ are linearly disjoint over $\overline{X_1}$. Then, as (K', X') is an unramified extension of (K_1, X_1), Corollary 3.6 shows A' is a Dubrovin valuation ring of S' with $\overline{A'} = \overline{A_1}\otimes_{\overline{X_1}} \overline{X'}$. Since $J(X_1)\cdot X' = J(X')$ we may invoke Lemma 3.12 again, obtaining $J(A') = J(A_1)\otimes_{\overline{X_1}} X' = J(A_1)\otimes_{U_1} U'\subseteq B'$. Furthermore, since $J(X_1)\cdot U' = J(X_1)\cdot X' = J(X')$, we have $B'/J(A') = \tilde{B}_1\otimes_{U_1} U' = \tilde{B}\otimes_{\tilde{U}} \tilde{U}'$, which by hypothesis is a Dubrovin valuation ring. Hence, by (1.6) B' is a Dubrovin valuation ring of S', as desired. □

4. The Main Argument

The basic argument for proving the theorems stated in Sect. 2 is an induction on jump rank. We give that argument in this section, while deferring the rest of the proofs to Sect. 5. Throughout this section B will be some fixed Dubrovin valuation ring of an F-central simple algebra S (with $[S:F]<\infty$), and $V=B\cap F$, a valuation ring of F. A will be some Dubrovin valuation ring with $B\subseteqq A\subseteqq S$, and $W=A\cap F$. While B will be fixed, the choice of A will vary depending on the context. We further set $\tilde{B}=B/J(A)$, which is a Dubrovin valuation ring of $\bar{A}=A/J(A)$; set $\tilde{V}=V/J(W)=\tilde{B}\cap\bar{W}$, a valuation ring of $\bar{W}$; and set $Y=\tilde{B}\cap Z(\bar{A})$, a valuation ring of the field $Z(\bar{A})$ with $Y\cap\bar{W}=\tilde{V}$. The following diagram indicates the inclusion relations among these rings:

$$\begin{array}{ccc} F\subseteqq S & \qquad & \bar{W}\subseteqq Z(\bar{A})\subseteqq\bar{A} \\ |\quad\;\; | & & |\qquad\; |\qquad\; | \\ V\subseteqq B & & \tilde{V}\subseteqq\; Y\;\subseteqq\tilde{B} \end{array}$$

In order to prove Theorem F (ii)$\Rightarrow$(vi), it is necessary for the inductive process to prove the following stronger result:

(4.1) **Proposition.** *Let B be a Dubrovin valuation ring of a simple Artinian ring S. Let K be a subfield of $Z(S)$, with $[S:K]<\infty$. Let $U=B\cap K$, a valuation ring of K, and let (K',U') be the Henselization of (K,U). Suppose B is integral over U. Then $B\otimes_U U'$ is a Dubrovin valuation ring of $S\otimes_K K'$, and $B\otimes_U U'$ is an immediate compatible extension of B.*

We now prove Theorem A, Theorem B, Theorem D, Theorem F (i)$\Rightarrow$(vi), Proposition 4.1, and Theorem G (iii)$\Rightarrow$(i). In the proof of Theorem D, we assume $Q_1, Q_2, \ldots, Q_k$ are precisely the jump prime ideals of V with respect to S.

Proof. The proof is by a primary induction on the dimension $[S:F]$ and a secondary induction on the jump rank $j(V,S)$ of V with respect to S. Note that if $[S:F]=1$, then $B=V$ and everything holds trivially. Thus, we may assume $[S:F]>1$. We break the rest of the proof four parts: I. rank$(V)=1$; II. rank$(V)>1$, $j(V,S)=1$, and V has a minimal nonzero prime ideal; III. $j(V,S)=1$ and V has no minimal nonzero prime ideal; IV. $j(V,S)>1$. The primary induction hypothesis will be invoked only in part IV. For Theorem F we will actually prove (i)$\Rightarrow$(vi′), where (vi′) reads:

(vi′) *$B\otimes_V V^h$ is a Dubrovin valuation ring of $S\otimes_F F^h$, and is an immediate compatible extension of B.*

I

Assume rank$(V)=1$. The conjugacy theorem (Theorem A) was proved for B in this case by Brungs and Gräter [BG2, Theorem 5.2]. Let $(\hat{F},\hat{V})$ be the completion of (F,V) (with respect to the topology of the ideals of V), and let $\hat{S}=S\otimes_F\hat{F}$. We identify S and $\hat{F}$ with their images in $\hat{S}$. Let $\hat{B}$ be a Dubrovin valuation ring of $\hat{S}$ with $\hat{B}\cap\hat{F}=\hat{V}$. It is shown in [D2, Sect. 3, Lemma 1] and more convincingly in [BG2, Lemma 3.4] that $\hat{B}\cap S$ is a Dubrovin valuation ring of S contracting to V in F. By Theorem A, B and $\hat{B}\cap S$ are conjugate, so we may assume $B=\hat{B}\cap S$. The proof that $\hat{B}\cap S$ is Dubrovin in [D2] or [BG2] shows that $J(\hat{B})\cap S=J(B)$ and $\bar{B}=\bar{\hat{B}}$.

Because $\operatorname{rank}(V)=1$, we can take the Henselization (F^h, V^h) of (F, V) to be: F^h is the separable closure of F in $\hat{F}$ and $V^h=\hat{V}\cap F^h$ (cf. [E, (17.18)]). Let $S^h=S\otimes_F F^h \subseteq \hat{S}$, and let $B^h=\hat{B}\cap S^h$. Since $(\hat{F}, \hat{V})$ is the completion of (F^h, V^h), the arguments quoted in the preceding paragraph show that B^h is a Dubrovin valuation ring of S^h with $B^h\cap F^h=V^h$, $J(\hat{B})\cap S^h=J(B^h)$, and $\overline{B^h}=\overline{\hat{B}}$. Hence, $B^h\cap S=B$, $J(B^h)\cap S=J(B)$, and $\bar{B}=\overline{B^h}$.

We next show $B^h=B\cdot V^h$. Clearly $B\cdot V^h\subseteq B^h$. Let $\{b_1,\dots,b_m\}\subseteq B$ be any F-base of S. Take any $\beta\in B^h$ and write $\beta=\sum b_i\gamma_i$ with $\gamma_i\in F^h$. Since F is dense in $\hat{F}$ as $\operatorname{rank}(V)=1$, F is dense in F^h. Hence, there exist $c_i\in F$ with $\gamma_i-c_i\in V^h$, $1\leq i\leq m$. Then $\sum b_ic_i=\beta-\sum b_i(\gamma_i-c_i)\in B^h\cap S=B$. Hence, $\beta=\sum b_ic_i+\sum b_i(\gamma_i-c_i)\in B\cdot V^h$. So, $B^h=B\cdot V^h$. Since $B\otimes_V V^h$ embeds in $S\otimes_V V^h$ by (3.1), this shows $B^h=B\otimes_V V^h$. Thus, B is integral over V by (3.9) (f), and B^h is a compatible extension of B.

To see $\Gamma_{B^h}=\Gamma_B$, i.e. $\operatorname{st}(B^h)=\operatorname{st}(B)\cdot B^{h*}$, take any $\delta\in\operatorname{st}(B^h)$ and write $\delta=\sum b_i\varepsilon_i$ and $\delta^{-1}=\sum b_i\gamma_i$, with the b_i as above and ε_i, $\lambda_i\in F^h$. There is an ideal $I\neq(0)$ of V^h, such that $\lambda_iI\subseteq J(V^h)$, for each i. Since F is dense in F^h we may choose $e_i\in F$ with $\varepsilon_i-e_i\in I$, each i. Let $d=\sum b_ie_i\in S$. Then,

$$\delta^{-1}(d-\delta)=\sum_i\sum_j b_ib_j\lambda_i(e_j-\varepsilon_j)\in B\cdot J(V^h)\subseteq J(B^h).$$

Hence $1+\delta^{-1}(d-\delta)\in 1+J(B^h)\subseteq B^{h*}$, so that

$$d=\delta[1+\delta^{-1}(d-\delta)]\in\operatorname{st}(B^h)\cdot B^{h*}\cap S=\operatorname{st}(B^h)\cap S=\operatorname{st}(B).$$

Thus, $\delta=d[1+\delta^{-1}(d-\delta)]^{-1}\in\operatorname{st}(B)\cdot B^{h*}$, showing $\Gamma_{B^h}=\Gamma_B$. Therefore, B^h is an immediate compatible extension of B; so (vi′) of Theorem F holds.

We can now prove Theorem B. We have $S^h\cong M_n(D^h)$ for some F^h-central division ring D^h, with $n=n_B$. Since V^h is Henselian, V^h extends to an invariant valuation ring R of D^h. Then $M_n(R)$ is a Dubrovin valuation ring of $M_n(D^h)$. By Theorem A, $B^h\cong M_n(R)$ by a V^h-isomorphism. Thus, $\bar{B}\cong\overline{B^h}\cong\overline{M_n(R)}\cong M_n(\bar{R})$, and hence $t_B=n=n_B$. Furthermore, $\Gamma_B=\Gamma_{B^h}\cong\Gamma_{M_n(R)}=\Gamma_R$, using Corollary 3.5. The middle isomorphism is the identity on Γ_{V^h}, and hence can be viewed as an equality in the divisible hull of $\Gamma_{V^h}=\Gamma_V$. We obtain the commutative diagram of Theorem B (iii) by combining the corresponding diagrams from B to B^h (as B^h is immediate over B) and from R to $M_n(R)$ to B^h. Theorem D holds as $n_B=t_B$ and there are no ℓ_i since $k=1$.

Now, let U, K and U', K' be as in Proposition 4.1, and suppose B is integral over U. Then V is integral over U, so V is the unique extension of U to F. By (3.7) $F\otimes_K K'$ is a field, and $F\otimes_K K'$ can be identified with F^h so that $V^h\cap K'=U'$. Then $S^h=S\otimes_F F^h=S\otimes_K K'=S\otimes_U U'$, and $B\otimes_U U'$ embeds in S^h by (3.1). Because $\operatorname{rank}(U)=\operatorname{rank}(V)=1$, F is dense in F'. Hence, the argument above proving $B^h=B\cdot V^h$ applies here to show $B^h=B\cdot U'=B\otimes_U U'$. This yields Proposition 4.1.

For Theorem G (iii)$\Rightarrow$(i) suppose S is a division ring and $t_B>1$. (So $S\neq F$.) Recall that in $M_k(D)$ for any division ring D (except the field with two elements) and any $k\geq 2$, the only proper subrings invariant under all inner automorphisms are central. (For this, see [H, Theorem 3], [K, Satz 3], or for $k>2$ [Ro, Corollary 1].) Therefore, as $B^h\subseteq S^h$, of matrix size t_B,

$$V\subseteq\bigcap_{s\in S^{h*}}(sB^hs^{-1}\cap S)\subseteq B^h\cap Z(S^h)\cap S=V^h\cap S=V,$$

so equality holds throughout. Each $sB^h s^{-1} \cap S$ is a Dubrovin valuation ring of S contracting to V, so a conjugate of B. There must be infinitely many such conjugates, since for any $s_1, \ldots, s_k \in S^*$ elementary localization theory shows $\left(\bigcap_{i=1}^{k} s_i B s_i^{-1}\right) \cdot F = S \neq F$. This completes part I.

Before going to part II, we prove further cases of Proposition 4.1:

(4.2) *Proposition 4.1 holds if* rank(U) *is finite or if B is an invariant valuation ring, or if $B \cong M_n(B')$, where B' is an invariant valuation ring.*

Proof of (4.2). The case rank$(U)=1$ was settled in part I above. Proposition 4.1 then holds whenever rank$(U)<\infty$ by induction on rank(U), with the induction step provided by Lemma 3.13.

Next, assume that S is a division ring and that B is an invariant valuation ring of S. Let $\{s_1, s_2, \ldots, s_\ell\}$ be a base of S as a K-vector space, and let $s_i s_j = \sum_{k=1}^{\ell} a_{ijk} s_k$ with $a_{ijk} \in K$. Let K_0 be any subfield of K such that all $a_{ijk} \in K_0$ and K_0 is finitely-generated over the prime subfield. Let S_0 be the K_0-vector space (and algebra) spanned by $\{s_1, s_2, \ldots, s_\ell\}$. Then, $K_0 \subseteq Z(S_0)$ as $K \subseteq Z(S)$, and $S_0 \otimes_{K_0} K = S$; hence, S_0 has no zero divisors, and as $[S_0 : K_0] = [S : K] < \infty$, S is a division ring. Let $B_0 = B \cap S_0$, which is an invariant valuation ring of S_0; let $U_0 = U \cap K_0 = B_0 \cap K_0$, a valuation ring of K_0. For any given $b \in B_0$, let $f \in K_0[X]$ be the minimal polynomial of b over K_0. Then, as $S = S_0 \otimes_{K_0} K$, f is also the minimal polynomial of b over K. But because $b \in B$ is integral over U, which is integrally closed, the coefficients of f lie in U. Hence $f \in (U \cap K_0)[X] = U_0[X]$; thus, B_0 is integral over U_0. Let $(K_0', U_0') = (K_0, U_0)^H$. Now, the rank of U_0 is finite (by, e.g. [B, Sect. 10, No. 3, Corollary 2]) since the transcendence degree of K_0 over the prime field is finite. Hence, by the finite rank case already proved, $B_0 \otimes_{U_0} U_0'$ is a Dubrovin valuation ring of $S_0 \otimes_{K_0} K_0'$. Because K is the direct limit of such fields K_0, and B (resp. U, U') is the direct limit of the corresponding B_0 (resp. U_0, U_0'), $B \otimes_U U'$ is the direct limit of the (compatible) Dubrovin valuation rings $B_0 \otimes_{U_0} U_0'$. It follows easily that $B \otimes_U U'$ is a Dubrovin valuation ring (in fact an invariant valuation ring, in view of Theorem 2.2).

Since we now have Proposition 4.1 for any invariant valuation ring, observe that it also follows for $B = M_n(R)$ where R is an invariant valuation ring. For, $B \otimes_U U' = M_n(R \otimes_U U')$ which is Dubrovin by (1.5) as $R \otimes_U U'$ is Dubrovin. This yields (4.2). □

The following general setup occurs repeatedly in the rest of the proof:

(4.3) *Setup.* Given B, S, F, V as usual, P will be a specified prime ideal of V, $A = B_P$, and $W = V_P = A \cap F$. Let $(F_1, W_1) = (F, W)^H$, and let $S_1 = S \otimes_F F_1$. Set $\tilde{V}_1 = \tilde{V}$ viewed in $W_1 = \bar{W}$, and let $V_1 = \pi_{W_1}^{-1}(\tilde{V}_1)$, the valuation ring of F_1 with $V_1 \subseteq W_1$ and $V_1 \cap F = V$. Set $A_1 = A \otimes_W W_1$. It will in every case be known that $n_A = t_A$ and that A_1 is a Dubrovin valuation ring of S_1 which is an immediate compatible extension of A, with $A_1 \cap F_1 = W_1$. Set $\tilde{B}_1 = \tilde{B}$ viewed in $\overline{A_1} = \bar{A}$, and let $B_1 = \pi_{A_1}^{-1}(\tilde{B}_1)$, a Dubrovin valuation ring of S_1 by (1.6), with $B_1 \cap F_1 = V_1$. Evidently, B_1 is a compatible extension of B, with $\overline{B_1} = \bar{B}$.

II

Now suppose rank$(V)>1$, but $j(V,S)=1$. Suppose further that V has a minimal prime ideal $P\neq(0)$. Use this P to form A, W, $F_1,\ldots$ as in Setup 4.3 above. The properties of A_1 specified in Setup 4.3 are known from part I. Write $S_1\cong M_n(D_1)$ where D_1 is an F_1-central division ring and $n=n_A=t_A$. Then by (1.5) $B_1\cong M_n(B_1')$, where B_1' is a Dubrovin valuation ring of D_1. Assume for convenience that a set of matrix units has been chosen in B_1 so that $B_1=M_n(B_1')$ (and $S_1=M_n(D_1)$). Set $A_1'=B_{1P}'$, which assures $A_1=B_{1P}=M_n(A_1')$.

Set $(F_2,V_2)=(F,V)^H$. (This is the (F^h,V^h) of Theorem B.) As noted in Sect. 3, we may view $(F_1,V_1)\subseteq(F_2,V_2)$. Set $D_2=D_1\otimes_{F_1}F_2$ and $S_2=S_1\otimes_{F_1}F_2=M_n(D_2)$. Since $S_2\cong S\otimes_F F_2$ the matrix size of S_2 is n_B. But $n_B=n_A=n$ because $j(V,S)=1$. Hence, D_2 is a division ring (namely, the D^h of Theorem B). We write R for the invariant valuation ring of D_2 with $R\cap F_2=V_2$. Then $R\cap D_1$ is an invariant valuation ring, and because $(F_2,V_2)=(F_1,V_1)^H$, Morandi's Theorem 2.2 says R is an immediate extension of $R\cap D_1$. Note also that by (3.10) $R\cap D_1=B_1'$, since $(R\cap D_1)\cap F_1=V_1=B_1'\cap F_1$. Set $B_2=M_n(R)$, a Dubrovin valuation ring of S_2 with $B_2\cap F_2=R\cap F_2=V_2$; B_2 is a compatible extension of $B_1=M_n(B_1')$, hence a compatible extension of B. In view of Corollary 3.5, B_2 is actually an immediate compatible extension of B_1. Thus,

$$\bar{B}=\bar{\tilde{B}}=\bar{\tilde{B}}_1=\bar{B}_1=\bar{B}_2=M_n(\bar{R}).$$

Since R is invariant, $\bar{R}$ is a division ring; hence $t_B=n=n_B$. This proves (i) of Theorem B, and also Theorem D as there are no ℓ_i, since we are assuming the Q_i are only the jump prime ideals of V.

Since B_1' is an invariant valuation ring it is integral over $B_1'\cap F_1=V_1$ (cf. [W1, Corollary]). Set $\tilde{B}_1'=B_1'/J(A_1')\subseteq\bar{A}_1'$. Then $\tilde{B}_1'\cap Z(\bar{A}_1')$ is a valuation ring integral over $\tilde{V}_1$. So $\tilde{B}_1'\cap Z(\bar{A}_1')$ is the unique extension of $\tilde{V}_1$ from $\bar{W}_1$ to $Z(\bar{A}_1')$, i.e., $\ell_{B_1',A_1'}=1$.

We can now prove Theorem A for B: Let B_0 be another valuation ring of S with $B_0\cap F=V=B\cap F$. Let $A_0=B_{0_P}$, so that $A_0\cap F=W=A\cap F$. Since rank$(W)=1$, we saw in part I that A_0 and A are conjugate. Thus, we may assume that $A_0=A$. Let $\tilde{B}_0=B_0/J(A)$, a Dubrovin valuation ring of $\bar{A}$. Since $\bar{A}=\bar{A}_1=M_n(\bar{A}_1')$, [BG2, Theorem 2.4] or [D1, Sect. 1, Theorem 7] says $\tilde{B}_0$ is conjugate to $M_n(C)$ for some Dubrovin valuation ring C of $\bar{A}_1'$. Then, as $C\cap\bar{W}_1=\tilde{B}_0\cap\bar{W}=\tilde{V}=\tilde{V}_1$ and $\ell_{B_1',A_1'}=1$, we have $C\cap Z(\bar{A}_1')=\tilde{B}_1'\cap Z(\bar{A}_1')$. So, (3.10) yields $C=\tilde{B}_1'$, since $\tilde{B}_1'$ is invariant. Thus, $\tilde{B}_0$ is conjugate to $M_n(\tilde{B}_1')=\tilde{B}$ in $\bar{A}$. Therefore, as $B=\pi_A^{-1}(\tilde{B})$ and $B_0=\pi_A^{-1}(\tilde{B}_0)$, B and B_0 are conjugate in A, proving Theorem A for B.

In the previous paragraph we saw that the conjugacy Theorem A holds for $\tilde{B}=\tilde{B}_1$ in $\bar{A}=\bar{A}_1$. Since Theorem B holds for A and for A_1 by part I, Proposition 3.11(b) shows that we have the exact sequences $(\Gamma_{B,A})$ and (Γ_{B_1,A_1}) of Theorem E(i). Because B_1 is a compatible extension of B, there is a map of complexes $(\Gamma_{B,A})\to(\Gamma_{B_1,A_1})$, to which we apply the 5-lemma to see $\Gamma_{B_1}=\Gamma_B$. Thus, $\Gamma_B=\Gamma_{B_1}=\Gamma_{B_2}=\Gamma_R$ as $B_2=M_n(R)$ (recall Corollary 3.5); this yields (ii) of Theorem B. Since the equality $\Gamma_B=\Gamma_R$ follows from identifications corresponding to inclusions $B\subseteq B_2$ and $R\subseteq B_2$, Theorem B (iii) follows at once. Note also we have now shown that B_1 is an immediate compatible extension of B. Since B_2 is immediate compatible over B_1, B_2 is also immediate compatible over B.

Because $B_2 \cap S = B$ and $B_2 \cap F_2 = V_2$, Proposition (3.9) (f) shows B is integral over V. Since $\tilde{B} = \tilde{B_1} \cong M_n(\tilde{B_1'})$ where $\tilde{B_1'}$ is an invariant valuation ring, (4.2) shows $\tilde{B} \otimes_{\bar{V}} \bar{V_2}$ is a Dubrovin valuation ring. Since we also know $A \otimes_W W_1 = A_1$ is a Dubrovin valuation ring, Lemma 3.13 shows $B \otimes_V V_2$ is a Dubrovin valuation ring. Since $B \otimes_V V_2 \subseteq B_2$ and $(B \otimes_V V_2) \cap F_2 = V_2 = B_2 \cap F_2$ it follows from (1.8) that $B \otimes_V V_2 = B_2$, which is an immediate compatible extension of B, yielding (vi′) of Theorem F. Likewise, suppose $K, U, X, K_1, X_1, K', U', \tilde{U}'$ are as defined just before Lemma 3.13, and suppose B is integral over U. Then, $A \otimes_X X_1$ is Dubrovin by part I and $\tilde{B} \otimes_{\tilde{U}} \tilde{U}'$ is Dubrovin by (4.2), so $B \otimes_U U'$ is Dubrovin by (3.13). Thus, Proposition 4.1 holds.

It remains to prove (iii)$\Rightarrow$(i) of Theorem G for this B. Suppose S is a division ring and $t_B > 1$. Now, $\tilde{B}$ is a Dubrovin valuation ring of $\bar{A}$, and the matrix size of $\bar{A}$ is $t_A = n_A = n_B = t_B > 1$. Consequently, the argument of part I (invoking [H], etc.) shows that $\tilde{B}$ has infinitely many conjugates. Each lifts via π_A^{-1} to a different conjugate of B. This completes part II.

Before going on to part III, let us verify that Theorem B and Theorem D hold if B is an Azumaya algebra over V. For this, let $B^h = B \otimes_V V^h$ where $(F^h, V^h) = (F, V)^H$, and let $S^h = S \otimes_F F^h \cong M_n(D^h)$ where $n = n_B$ and D^h is a division ring. Then, B^h is an Azumaya algebra over V^h (cf. [DI, p. 61]), so by Corollary 3.4, $\Gamma_B = \Gamma_V = \Gamma_{V^h} = \Gamma_{B^h}$. Also, $\overline{B^h} = \bar{B} \otimes_{\bar{V}} \overline{V^h} = \bar{B}$, so B^h is an immediate compatible extension of B. Now, as B^h is Dubrovin, $B^h \cong M_n(B^{h\prime})$, where $B^{h\prime}$ is a Dubrovin valuation ring of D^h with $B^{h\prime} \cap F^h = V^h$. Because V^h is Henselian there is an invariant valuation ring R of D^h with $R \cap F^h = V^h$. By (3.10) $B^{h\prime} = R$. Thus, $\bar{B} = \overline{B^h} = M_n(\bar{R})$ (so $n = t_B$, as $\bar{R}$ is a division ring). We have established (i) and (ii) of Theorem B, and (iii) holds trivially since $Z(\bar{B}) = \bar{V}$ as B is Azumaya. Further, if E, A are rings with $B \subseteq E \subseteq A \subseteq S$, then A is Azumaya since it is a central localization of B (cf. (1.8)). Hence, $Z(\bar{A}) = \overline{A \cap F}$, which assures that $\ell_{E,A} = 1$. This yields Theorem D, since $n_B = n = t_B$ and all the $\ell_i = 1$.

III

Suppose now that $\operatorname{rank}(V) > 1$ and $j(V, S) = 1$, but V has no minimal nonzero prime ideal. The correspondence between prime ideals of B and V (cf. (1.8)) assures that $(0) = \bigcap Q$ as Q ranges over the nonzero prime ideals of B. We now invoke a variation of the Azumaya algebra argument of [D2, Sect. 3] and [BG2, Sect. 3]. If $[S:F] = k^2$, then B has p.i.-degree k. Then there is a homogeneous polynomial f with integer coefficients such that f is an identity for all algebras of p.i.-degree $< k$, but $f(b_1, \ldots, b_m) \neq 0$ for some $b_1, \ldots, b_m \in B$. So, there is a prime ideal Q of B with $f(b_1, \ldots, b_m) \notin Q$. Let $P = Q \cap V$, and use this P in Setup 4.3 to form A, W, F_1, S_1, A_1, etc. Then, $J(A) = Q$. For $\bar{b}_i = \pi_A(b_i)$ we have $f(\bar{b}_1, \ldots, \bar{b}_m) = \overline{f(b_1, \ldots, b_m)} \neq 0$ in $\bar{A}$. Therefore, $\bar{A}$ has p.i.-degree $k =$ p.i.-degree of A. By the Artin-Procesi theorem [C, Theorem 9, p. 465], A is an Azumaya algebra over W. The properties of A_1 required for Setup 4.3 follow because A is an Azumaya algebra. With this choice of A nearly all the arguments of part II carry through without change. The necessary information about A and A_1 (which in part II was obtained from part I results) is contained in the observations above about Azumaya algebras.

The only argument from part II which does not carry over is the proof of the conjugacy Theorem A for B. For this, let B_0 be another valuation ring of S with $B_0 \cap F = V = B_1 \cap F$. Then, just as with B, there exists a prime ideal P_0 of V, with

B_{0P_0} Azumaya. Pick a prime ideal $P_1 \neq (0)$ of V with $P_1 \subseteq P \cap P_0$. Redefine A to be B_{P_1} and set $A_0 = B_{0P_1}$. Then A and A_0 are Azumaya algebras over V_{P_1} since they are central localizations of the Azumaya algebras B_P and B_{0P_0}. Because V_{P_1} is a valuation ring the functorial map of Brauer groups $\mathrm{Br}(V_{P_1}) \to \mathrm{Br}(F)$ is injective by [Sa, Lemma 1.2]. Hence, $[A_0] = [A]$ in $\mathrm{Br}(V_{P_1})$. Since projective V_{P_1}-modules are free this implies $M_\ell(A_0) \cong M_\ell(A)$ for some ℓ; hence, by the cancellation theorem for Azumaya algebras over semilocal rings (cf. [OS, Corollary 1]), $A_0 \cong A$ by a V_{P_1}-algebra isomorphism. This isomorphism extends by central localization to an F-automorphism of S, which by Skolem-Noether is inner. Consequently, A and A_0 are conjugate. The rest of the argument for conjugacy of B and B_0 is the same as in part II. This completes part III.

IV

Now assume $j(V, S) > 1$. We argue by induction on the jump rank (within the primary induction on $[S:F]$). Let $Q_1 \subsetneqq Q_2 \subsetneqq \dots \subsetneqq Q_k = J(V)$ be the jump prime ideals of V with respect to S. For some i, $1 \leq i < k$, let $A = B_{Q_i}$ and let $W = A \cap F = V_{Q_i}$. Specific choices of i will be made later. Using this A and W define $\tilde{B}$, $\tilde{V}$, Y, as described at the beginning of Sect. 4. The jump prime ideals of W re S are clearly $Q_1, Q_2, \dots, Q_i$, so $j(W, S) = i < j(V, S)$. Hence, by induction Theorem B holds for the ring A. Consequently, by Proposition 3.9 (e), we may invoke Corollary E to see that $j(Y, \bar{A}) \leq j(V, S) - i < j(V, S)$. In addition, Theorem C applies to A by (3.9)(a), yielding $[\bar{A} : Z(\bar{A})] \leq [\bar{A} : \bar{W}] \leq [S:F]$. Consequently, the results we want to prove about B hold by induction for $\tilde{B}$ as well as for A.

To prove the conjugacy Theorem A for B, let B_0 be another Dubrovin valuation ring of S with $B_0 \cap F = V = B \cap F$. Let $A_0 = B_{0Q_i}$, which is a Dubrovin valuation ring of S with $A_0 \cap F = W = A \cap F$. Then A and A_0 are conjugate in S, since Theorem A holds for A by induction. Hence, we may assume $A_0 = A$. Let $\tilde{B}_0 = B_0/J(A)$ and $Y_0 = \tilde{B}_0 \cap Z(\bar{A})$. Then Y and Y_0 are each valuation rings of $Z(\bar{A})$ extending $\tilde{V}$ in $\bar{W}$. Now, by (3.9)(a), Corollary B applies for A. Hence, $Z(\bar{A})$ is normal over $\bar{W}$, so there is a $\tau \in \mathscr{G}(Z(\bar{A})/\bar{W})$ with $\tau(Y_0) = Y$. Since θ_A is surjective (by Corollary B again) τ can be induced by conjugation by some element of $\mathrm{st}(A)$. After conjugating B_0 by such an element, we may assume $Y_0 = Y$. By Theorem A for $\tilde{B}$, which holds by induction, $\tilde{B}_0$ and $\tilde{B}$ are conjugate in $Z(\bar{A})$. Hence, B_0 and B are conjugate, completing the proof of Theorem A.

We can also settle Proposition 4.1. For, since B is integral over U, A is integral over X and $\tilde{B}$ integral over $\tilde{U}$. By induction, Proposition 4.1 holds for A and for $\tilde{B}$, hence it holds for B by Lemma 3.13.

We next dispose of Theorem G (iii)$\Rightarrow$(i). For this, suppose S is a division ring and $t_B > 1$. If $t_A = 1$, then $\bar{A}$ is a division ring with Dubrovin valuation ring $\tilde{B}$ and $t_{\tilde{B}} = t_B > 1$. By induction, $\tilde{B}$ has infinitely many different conjugates. Their inverse images in A yield infinitely many different conjugates of B. On the other hand, if $t_A > 1$, then by induction A has infinitely many different conjugates. For any $s \in S^*$, $(sBs^{-1})_{Q_i} = sB_{Q_i}s^{-1} = sAs^{-1}$. Thus, whenever $sAs^{-1} \neq tAt^{-1}$ we have $sBs^{-1} \neq tBt^{-1}$. So, the infinitely many conjugates of A yield infinitely many conjugates of B, as desired.

Now drop the assumption that S is a division ring. We can quickly settle Theorem D (assuming the Q_i are just the jump primes of V re S). For this choose Q_i (for the construction of A, W, etc.) to be the next to last jump prime Q_{k-1}. Then,

$j(Y, \bar{A}) \leqq k-(k-1)=1$, so that $n_{\tilde{B}}=t_{\tilde{B}}$ by parts I–III above. Of course also $t_{\tilde{B}}=t_B$. Note that the jump prime ideals of W re S are $Q_1, \ldots, Q_{k-1}$, and that the $\ell_2, \ldots, \ell_{k-1}$ of Theorem D are the same for A as for B, while the ℓ_k for B is $\ell_{B,A}$. Theorem B holds for A by induction, hence we have Theorem E(iii) for B by (3.9)(d). Thus,

$$n_B = n_{\tilde{B}}(n_A/t_A)\ell_{B,A}$$
$$= t_B(\ell_2 \ldots \ell_{k-1})\ell_k,$$

as desired.

We now prove Theorem B under the assumption that $n_B=t_B$. For this, any choice of Q_i may be made $(1 \leqq i < k)$ for defining A, W, etc. Since Theorem B for A and Theorem A for $\tilde{B}$ hold by induction, Proposition 3.9 yields Theorem E for B in A. Hence, (with $s_B = n_B/t_B$) we have

$$s_B = s_{\tilde{B}} s_A \ell_{B,A}.$$

By assumption $s_B=1$, while $s_{\tilde{B}}$ and s_A are positive integers by Theorem D (which holds by induction), as is $\ell_{B,A}$. This forces

$$n_{\tilde{B}}=t_{\tilde{B}}, \quad n_A=t_A, \quad \text{and} \quad \ell_{B,A}=1.$$

The last equation means Y is the unique extension of $\tilde{V}$ to $Z(\bar{A})$. Hence, the exact sequence $(\Gamma_{B,A})$ of Theorem E(i) is actually the short exact sequence

$$(\Gamma_{B,A}) \qquad 0 \to \Gamma_{\tilde{B}} \to \Gamma_B \to \Gamma_A \to 0.$$

Now, using Q_i for the P, define F_1, W_1, S_1, V_1, A_1, B_1, etc. as in Setup 4.3. Since $n_A=t_A$, the required properties of A_1 follow from Theorem F(i)$\Rightarrow$(vi'), which holds for A by induction. Also, since $j(W_1, S_1)=1$ as W_1 is Henselian, Theorem B holds for A_1 by induction, so by (3.9) Theorem E holds for B_1 in A_1. The exact sequence (Γ_{B_1,A_1}) is actually short exact for the same reason as for $(\Gamma_{B,A})$ (as $\tilde{V}_1=\tilde{V}$ in $Z(\bar{A}_1) = Z(\bar{A})$). Furthermore, the compatibility of B in B_1 assures there is a map of complexes $(\Gamma_{B,A}) \to (\Gamma_{B_1,A_1})$, which by the 5-lemma is an isomorphism. Hence, $\Gamma_{B_1}=\Gamma_B$ and B_1 is an immediate compatible extension of B.

Let $(F_2, V_2)=(F, V)^H=(F^h, V^h)$. As noted in Sect. 3, we may view $(F_1, V_1) \subseteqq (F_2, V_2)$. Let $S_2 = S \otimes_F F_2 = S_1 \otimes_{F_1} F_2$; let $W_2 = V_{2Q_i}$, so $W_2 \cap F_1 = W_1$; and let $\tilde{V}_2 = V_2/J(W_2) \subseteqq \bar{W}_2$. By (3.8), W_2 is an unramified extension of W_1 with $(\bar{W}_2, \tilde{V}_2) = (\bar{W}_1, \tilde{V}_1)^H$. Also, since $\tilde{V}_1=\tilde{V}$ has a unique extension to $Z(\bar{A}_1)$, $Z(\bar{A}_1)$ is linearly disjoint to $\bar{W}_2$ over $\bar{W}_1$ (cf. (3.7)). Set $A_2 = A_1 \otimes_{W_1} W_2 \subseteqq S_2$. By Corollary 3.6, A_2 is a Dubrovin valuation ring of S_2 which is compatible with A_1, and $\Gamma_{A_2}=\Gamma_{A_1}$ and $\bar{A}_2 = \bar{A}_1 \otimes_{\bar{W}_1} \bar{W}_2$. View $\bar{A}_2$ as $\bar{A}_1 \otimes_{Z(\bar{A}_1)} (Z(\bar{A}_1) \otimes_{\bar{W}_1} \bar{W}_2)$. Note that by (3.7) the unique extension, call it Y_2, of $\tilde{V}_2$ to $Z(\bar{A}_1) \otimes_{\bar{W}_1} \bar{W}_2 (=Z(\bar{A}_2))$ is the Henselization of $\tilde{B}_1 \cap Z(A_1)$. Since $\tilde{B}_1 = \tilde{B}$, $n_{\tilde{B}}=t_{\tilde{B}}$, and $j(Y, \bar{A}) < j(V, S)$, Theorem F(i)$\Rightarrow$(vi') holds by induction for $\tilde{B}_1$. Hence, if we set $\tilde{B}_2 = \tilde{B}_1 \otimes_{\tilde{V}_1} \tilde{V}_2$, $\tilde{B}_2$ is a Dubrovin valuation ring of $\bar{A}_2$ which is an immediate compatible extension of $\tilde{B}_1$, with $\tilde{B}_2 \cap Z(\bar{A}_2) = Y_2$. Then $\pi_{A_2}^{-1}(\tilde{B}_2)$ is a Dubrovin valuation ring of S_2 contracting to $B_1 \cap S = B$ in S and to V_2 in F_2. Hence, B is integral over V by (3.9)(f).

Let $B_2 = B \otimes_V V_2$, a Dubrovin valuation ring of S_2 by Proposition 4.1 (which was proved above for B); B_2 is a compatible extension of B. Since $B_2 \subseteqq \pi_{A_2}^{-1}(\tilde{B}_2)$ and both these Dubrovin valuation rings contract to V_2 in F_2, we must have $B_2 = \pi_{A_2}^{-1}(\tilde{B}_2)$ (cf. (1.8)). So, $\tilde{B}_2 = B_2/J(A_2)$. An analogous argument shows $B_1 \otimes_V V_2 = \pi_{A_2}^{-1}(\tilde{B}_2) = B_2$ which yields that B_2 is a compatible extension of B_1. Now, as W_2

and Y_2 are Henselian, $j(W_2, S_2) = j(Y_2, \overline{A_2}) = 1$, so by induction Theorem B holds for A_2 and Theorem A holds for $\tilde{B_2}$. Thus, Proposition (3.9)(b) shows that Theorem E (i) holds for B_2 in A_2. The diagram (Γ_{B_2, A_2}) is a short exact sequence, as the Henselian valuation $\tilde{V_2}$ has a unique extension to $Z(\overline{A_2})$. Moreover, from the compatibility of B_1 in B_2, there is a map of complexes $(\Gamma_{B_1, A_1}) \to (\Gamma_{B_2, A_2})$, which is an isomorphism by the 5-lemma. Hence, $\Gamma_{B_2} = \Gamma_{B_1}$. Since $\overline{B_1} = \overline{\overline{B_1}} = \overline{\overline{B_2}} = \overline{B_2}$, B_2 is an immediate compatible extension of B_1, hence of B. (This yields Theorem F (i)$\Rightarrow$(vi$'$) for B.)

Now, $S_2 \cong M_n(D_2)$, where $n = n_B$ and D_2 is an F_2-central division ring. Since V_2 is Henselian, there is an invariant valuation ring of D_2, called R in Theorem B, such that $R \cap F_2 = V_2$. Because R is the only Dubrovin valuation ring of D_2 contracting to V_2, by (1.5) and (3.10) $B_2 \cong M_n(R)$. Thus, $\bar{B} = \overline{B_2} \cong M_n(\bar{R})$ where $n = n_B = t_B$, and $\Gamma_B = \Gamma_{B_2} = \Gamma_R$ by Corollary 3.5. This yields (i) and (ii) of Theorem B. Finally, in the diagram below,

$$\begin{array}{ccccc}
\Gamma_B/\Gamma_V & \xrightarrow{=} & \Gamma_{B_2}/\Gamma_{V_2} & \xrightarrow{=} & \Gamma_R/\Gamma_{V_2} \\
\downarrow{\scriptstyle \theta_B} & & \downarrow{\scriptstyle \theta_{B_2}} & & \downarrow{\scriptstyle \theta_R} \\
\mathscr{G}(Z(\bar{B})/\bar{V}) & \xrightarrow{\cong} & \mathscr{G}(Z(\overline{B_2})/\overline{V_2}) & \xrightarrow{\cong} & \mathscr{G}(Z(\bar{R})/\overline{V_2})
\end{array}$$

the left square is commutative as B_2 is compatible with B, and the right square is commutative as $M_n(R)$ is compatible with R and $B_2 \cong M_n(R)$. The commutative outer rectangle establishes (iii) of Theorem B.

It remains to prove Theorem B in case $t_B < n_B$. We have shown Theorem D, which says

$$n_B = t_B \ell_2 \ldots \ell_k,$$

where $\ell_j = \ell_{B_{Q_j}, B_{Q_{j-1}}}$. Since $t_B < n_B$ some $\ell_j > 1$. Choose i, $1 \leqq i < k$, so that $\ell_2 = \ldots = \ell_i = 1$, but $\ell_{i+1} > 1$. (If $\ell_2 > 1$, set $i = 1$.) For this part of the argument, set $A = B_{Q_i}$ and $W = A \cap F$, with the i just selected. We have $\ell_{B,A} \geqq \ell_{A_{Q_{i+1}}, A} = \ell_{i+1} > 1$, since $\tilde{V}$ has at least as many extensions to $Z(\bar{A})$ as (the coarser valuation ring) $V_{Q_{i+1}}/J(W)$. Note also that $j(W, S) = i < j(V, S)$ and by Theorem D for A, $n_A = t_A \ell_2 \ldots \ell_i = t_A$.

Now define F_1, W_1, S_1, V_1, A_1, B_1, etc. as in Setup 4.3, using Q_i for the P. Theorem F (i)$\Rightarrow$(vi$'$), which holds for A by induction, assures that A_1 has the required properties. For the same reasons as in the $n_B = t_B$ case above, Theorem E (i) applies to B in A and to B_1 in A_1, yielding exact (but no longer short exact) sequences $(\Gamma_{B,A})$ and (Γ_{B_1, A_1}). Because B_1 is a compatible extension of B there is a map of complexes $(\Gamma_{B,A}) \to (\Gamma_{B_1, A_1})$, which is an isomorphism by the 5-lemma (Recall that $\tilde{V_1} = \tilde{V}$ and $\tilde{B_1} = \tilde{B}$ in $\overline{A_1} = \bar{A}$.) Hence, B_1 is an immediate compatible extension of B.

We have $S_1 \cong M_m(D_1)$, where $m = n_A$ and D_1 is a division ring. Then by (1.5), with appropriate choice of idempotents, $B_1 = M_m(B_1')$ and $S_1 = M_m(D_1)$ where B_1' is a Dubrovin valuation ring of D_1; further, $A_1 = M_m(A_1')$ where $A_1' = B'_{1Q_i}$. We have $\bar{B} = \overline{B_1} = M_m(\overline{B_1'})$, $\Gamma_B = \Gamma_{B_1} = \Gamma_{B_1'}$ by Corollary 3.5, and a commutative diagram (from the compatibility of B_1 with B and with B_1'):

$$\begin{array}{ccccc}
\Gamma_B/\Gamma_V & \xrightarrow{=} & \Gamma_{B_1}/\Gamma_{V_1} & \xrightarrow{=} & \Gamma_{B_1'}/\Gamma_{V_1} \\
\downarrow{\scriptstyle \theta_B} & & \downarrow{\scriptstyle \theta_{B_1}} & & \downarrow{\scriptstyle \theta_{B_1'}} \\
\mathscr{G}(Z(\bar{B})/\bar{V}) & \xrightarrow{\cong} & \mathscr{G}(Z(\overline{B_1})/\overline{V_1}) & \xrightarrow{\cong} & \mathscr{G}(Z(\overline{B_1'})/\overline{V_1})
\end{array}.$$

Since $[D_1:F_1]=[S:F]/m^2$, Theorem B holds for B_1' by the primary induction if $m>1$. In this case Theorem B follows for B by what we have just seen.

Thus, we may assume $m=1$, i.e., $S_1=D_1$. Because W_1 is Henselian, there is an invariant valuation ring of S_1 contracting to W_1; by (3.10) that ring is A_1.

Now, let $Z(\bar{A})_{\text{sep}}$ be the separable closure of $\bar{W}$ in $Z(\bar{A})$, and let $L'\subseteq Z(\bar{A})_{\text{sep}}$ be the decomposition field of $\tilde{B}\cap Z(\bar{A})_{\text{sep}}$ over $\tilde{V}$, so that $[L':\bar{W}]=\ell_{B,A}>1$. Just as in the proof of (3.9)(d), since L' is normal over $\bar{W}$ (as $Z(\bar{A})_{\text{sep}}$ is abelian Galois over $\bar{W}$), $\tilde{V}$ has $\ell_{B,A}$ different extensions from $\bar{W}$ to L', each of which extends uniquely to $Z(\bar{A})$. Identify L' with its image in $\bar{A}_1=\bar{A}$. Since W_1 is Henselian there is a field L, $F_1\subseteq L\subseteq S_1$, which is an inertial lift of L' over $\bar{W}_1=\bar{W}$. That is, L is separable over F, with $[L:F_1]=[L':\bar{W}]=\ell_{B,A}$, and $\bar{W}_L=L'$ in $\bar{A}_1$, where W_L is the valuation ring $A_1\cap L$. Let S_L be the centralizer of L in S_1, and let $A_L=A_1\cap S_L$. Then A_L is an invariant valuation ring of S_L with $\bar{A}_L=\bar{A}_1$ by [JW, Lemma 1.8(a)]. Let $B_L=B_1\cap S_L$, $\tilde{B}_L=B_L/J(A_L)$, $V_L=B_L\cap L$, and $\tilde{V}_L=V_L/J(W_L)$. Then B_L is a Dubrovin valuation ring of S_L, since $B_L=\pi_{A_L}^{-1}(\tilde{B}_1)$. Furthermore, B_1 is a compatible extension of B_L. We also have $[S_L:L]<[S_L:L]\cdot[L:F_1]^2=[S:F]$. Consequently, by the primary induction, Theorem B holds for B_L. We work back from B_L to B.

As in the proof of (3.9)(d), we may view $(L,V_L)\subseteq(F,V)^H$. So, since $(L,V_L)^H=(F,V)^H=(F^h,V^h)$ we have the same R for B as for B_L in Theorem B. Thus,

$$\bar{B}=\bar{\tilde{B}}=\bar{\tilde{B}}_L=\bar{B}_L\cong M_t(\bar{R}),$$

where $t=t_{B_L}=t_{\tilde{B}_L}=t_{\tilde{B}}=t_B$. As to the value groups, consider the diagram

$$\begin{tikzcd} 0 \arrow[r] & \Gamma_{\tilde{B}_L} \arrow[r] \arrow[d, "="'] & \Gamma_{B_L} \arrow[r] \arrow[d] & \Gamma_{A_L} \arrow[r] \arrow[d] & 0 \\ 0 \arrow[r] & \Gamma_{\tilde{B}_1} \arrow[r] & \Gamma_{B_1} \arrow[r] & \Gamma_{A_1} \arrow[r] & \mathscr{G}(Z(\bar{A}_1)/\bar{W}_1)/H \arrow[r] & 0. \end{tikzcd}$$

The rows of the diagram are the exact sequences (Γ_{B_L,A_L}) and (Γ_{B_1,A_1}) given by Theorem E (i). (This theorem applies to B_L in A_L and B_1 in A_1 by Proposition 3.9(b), since Theorem B holds for A_L and A_1 (as W_L and W_1 are Henselian) and Theorem A holds for $\tilde{B}_L=\tilde{B}_1=\tilde{B}$ by induction.) The top row is short exact because $\tilde{V}_L=\tilde{V}$ has a unique extension to $Z(\bar{A}_1)=L'$. In the bottom row

$$H=\{\tau\in\mathscr{G}(Z(\bar{A}_1)/\bar{W}_1):\tau(\tilde{B}_1\cap Z(\bar{A}_1))=\tilde{B}_1\cap Z(\bar{A}_1)\}=\mathscr{G}(Z(\bar{A}_1)/L').$$

The diagram is commutative because B_1 is a compatible extension of B_L. Furthermore, [JW, Lemma 1.8 (a)] shows that Γ_{A_L} maps onto $\theta_{A_1}^{-1}(\mathscr{G}(Z(\bar{A}_1)/L')$ in $\Gamma_{A_1}/\Gamma_{V_1}$. Thus, in the diagram above the image of $(\Gamma_{A_L}\to\Gamma_{A_1})$ contains the image of $(\Gamma_{B_1}\to\Gamma_{A_1})$. The diagram therefore shows $\Gamma_{B_1}=\Gamma_{B_L}$. Thus, $\Gamma_B=\Gamma_{B_1}=\Gamma_{B_L}=\Gamma_R$. Finally, (iii) of Theorem B follows from the commutativity of the diagram

$$\begin{tikzcd} \Gamma_B/\Gamma_V \arrow[r] \arrow[d, "\theta_B"'] & \Gamma_{B_1}/\Gamma_{V_1} \arrow[r] \arrow[d, "\theta_{B_1}"'] & \Gamma_{B_L}/\Gamma_{V_L} \arrow[r] \arrow[d, "\theta_{B_L}"'] & \Gamma_R/\Gamma_{V^h} \arrow[d, "\theta_R"'] \\ \mathscr{G}(Z(\bar{B})/\bar{V}) \arrow[r] & \mathscr{G}(Z(\bar{B}_1)/\bar{V}_1) \arrow[r] & \mathscr{G}(Z(\bar{B}_L)/\bar{V}_L) \arrow[r] & \mathscr{G}(Z(\bar{R})/\bar{V^h}) \end{tikzcd}$$

The first two squares of the diagram are commutative by the compatibility of B_1 over B and B_1 over B_L. The right square is commutative by Theorem B (iii) for B_L. This completes the proof of Theorem B, completing part IV of the main proof. □

We have now fully proved Theorem A and Theorem B. Thus, Corollary B, Theorem C, Theorem E, and Corollary E follow by Proposition. 3.9. We have proved Theorem D assuming that the Q_i are the jump prime ideals of V re S. This assumption was needed in part IV only to assure that $j(V_{Q_{k-1}}, S) < j(V, S)$, allowing us to apply Theorem B to $A = B_{Q_{k-1}}$ by induction on jump rank. Since we have now proved Theorem B in general, the argument in part IV (now by induction on k instead of jump rank) proves Theorem D as originally stated. We have also proved Proposition 4.1.

5. Proofs of Theorems F and G Completed

Proof of Theorem G. As we noted after the statement of Theorem G, (i)$\Leftrightarrow$(ii)$\Rightarrow$(iv) and (ii)$\Rightarrow$(iii) were proved by Brungs and Gräter. They also showed in [BG1, Theorem 4] that if the set of elements of S integral over V is a ring, then V extends to a total valuation ring B_1 of S. Since B is conjugate to B_1 by Theorem A, B is also a total valuation ring of S. This yields (iv)$\Rightarrow$(ii) of Theorem G, while (iii)$\Rightarrow$(i) was proved in Sect. 4 above.

It remains only to prove the formula for the number of conjugates when (i)–(iv) hold. We do this by induction on the jump rank. Let c_B be the number of conjugates of B. Note first that if $n_B = 1$, then B is an invariant valuation ring by Theorem 2.2 and Proposition 3.10; so, $c_B = 1 = n_B$. Now assume $n_B > 1 = t_B$. Then $j(V,S) > 1$ by Theorem D. Let P be the smallest jump prime ideal of V re S, and let $A = B_P$. With this P and A define W, $\tilde{V}$, $\tilde{B}$, and Y as at the beginning of Sect. 4. Since $j(W, S) = 1$, $n_A = t_A \leqq t_B = 1$ by Theorem D and Theorem E(ii). Hence $n_A = 1$, so that A is an invariant valuation ring. Therefore, all the conjugates of B lie in A, and are thus determined by their images in $\bar{A}$. That is, c_B equals the number of Dubrovin valuation rings $\tilde{B}_i$ of $\bar{A}$ with $\tilde{B}_i \cap \bar{W} = \tilde{V}$. For any such $\tilde{B}_i$, $\tilde{B}_i \cap Z(\bar{A})$ is one of the $\ell_{B,A}$ extensions of $\tilde{V}$ to $Z(\bar{A})$. If $\tilde{B}_i \cap Z(\bar{A}) = Y = \tilde{B} \cap Z(\bar{A})$, then $\tilde{B}_i$ is a conjugate of $\tilde{B}$ in $\bar{A}$. The number $c_{\tilde{B}}$ of such conjugates is $n_{\tilde{B}}$ by induction. (Note that $j(Y, \bar{A}) < j(V, S)$ by Corollary E. The hypotheses of Theorem G hold for $\tilde{B}$ as $\bar{A}$ is a division ring since A is invariant, and $t_{\tilde{B}} = t_B = 1$.) Since $Z(\bar{A})$ is normal over $\bar{W}$ by Corollary B (or [JW, Proposition 1.7]), $\mathscr{G}(Z(\bar{A})/\bar{W})$ acts transitively on the extensions of $\tilde{V}$ to $Z(\bar{A})$. Each automorphism in $\mathscr{G}(Z(\bar{A})/\bar{W})$ extends to a $\bar{W}$-automorphism of $\bar{A}$ by the surjectivity of θ_A in Corollary B. Hence, each extension of $\tilde{V}$ to $Z(\bar{A})$ is the center of $c_{\tilde{B}}$ Dubrovin valuation rings of $\bar{A}$. Therefore,

$$c_B = c_{\tilde{B}} \cdot \ell_{B,A} = n_{\tilde{B}} \cdot \ell_{B,A} = n_B,$$

as desired, where the last equality is given by Theorem E (iii). $\square$

Proof of Theorem F. We have already proved (vi) $\Rightarrow$ (vii) $\Rightarrow$ (ii) (in Proposition 3.9(f)), (ii) $\Rightarrow$ (vi) (in Proposition 4.1), and (i) $\Rightarrow$ (vi′) (in Sect. 4). Clearly (vi′) $\Rightarrow$ (vi). To complete the proof of Theorem F we now show (ii) $\Rightarrow$ (v) $\Rightarrow$ (i) and (vi′) $\Rightarrow$ (iv) $\Rightarrow$ (iii) $\Rightarrow$ (v).

(ii)$\Rightarrow$(v) Suppose B is integral over V, and take any (Dubrovin valuation) rings E, A with $B \subseteq E \subseteq A \subseteq S$. Let $T = E \cap F$, $W = A \cap F$, $\tilde{T} = T/J(W) \subseteq \bar{W} \subseteq \bar{A}$, $\tilde{E} = E/J(A) \subseteq \bar{A}$, and $P = J(T) \subseteq V$. Because B is integral over V, $E = B_P$ is integral over $T = V_P$. Hence $\tilde{E}$ is integral over $\tilde{T}$, which is a valuation ring of the field $\bar{W}$. Since the valuation ring $\tilde{E}\cap Z(\bar{A})$ is integral over $\tilde{T}$, it is the only extension of $\tilde{T}$ to $Z(\bar{A})$. Thus, $\ell_{E,A}=1$, proving (v).

(v)$\Rightarrow$(i) This is immediate from Theorem D since by hypothesis all the $\ell_i=1$.

(vi')$\Rightarrow$(iv) Let $B^h=B\otimes_V V^h$, which we are assuming is a Dubrovin valuation ring of $S\otimes_F F^h$. We have $S\otimes_F F^h\cong M_n(D^h)$ where $n=n_B$ and D^h is a division ring. Let R be the invariant valuation ring of D^h extending the Henselian valuation ring V^h of F^h. After making a convenient choice of matrix units (and copy of D^h), we may assume $B^h=M_n(R)$. We first prove (iv) for B^h. Every ideal I of B^h has the form $M_n(I')$ where I' is an ideal of R. Then $I'-\{0\}$ is a generating set of I as an ideal of B^h, and $I'-\{0\}\subseteq D^{h*}=\mathrm{st}(R)\subseteq\mathrm{st}(B^h)$.

To prove (iv) for B first note that $\Gamma_{B^h}=\Gamma_B$ since we are assuming B^h is an immediate compatible extension of B. Now, take any ideal I of B. If $\{\beta_i\}\subseteq\mathrm{st}(B^h)$ is a generating set of the ideal $I\otimes_V V^h$ of B^h, write each $\beta_i=b_iu_i$ with $b_i\in\mathrm{st}(B)$ and $u_i\in B^{h*}$. Let I_1 be the ideal of B generated by $\{b_i\}$. Then the ideal of B^h generated by I_1 is $I_1\otimes_V V^h$ but is also $I\otimes_V V^h$. Hence, by Lemma 3.1(b),

$$I=(I\otimes_V V^h)\cap V=(I_1\otimes_V V^h)\cap V=I_1,$$

which proves (iv) for B.

(iv)$\Rightarrow$(iii) Let $I=BbB$. By (iv), I is generated as an ideal by some $\{s_i\}\subseteq\mathrm{st}(B)$. Because the ideals of B are linearly ordered by inclusion (cf. (1.4)), we have $I=\bigcup_i Bs_iB$. So, there is an s_j with $b\in Bs_jB$. Then $I=Bs_jB=Bs_j=s_jB$, as $s_j\in\mathrm{st}(B)$, proving (iii).

(iii)$\Rightarrow$(v) Take any $b\in B$, $b\neq0$. We have from (iii) $BbB=Bc=dB$ for some c, $d\in B$. Since every nonzero two-sided ideal of B contains a regular element, $c\in S^*$. We have $cB\subseteq dB=Bc$, so $cBc^{-1}\subseteq B$. Since $cBc^{-1}\cap F=B\cap F=V$, the correspondence (1.8) between overrings of cBc^{-1} and localizations implies $cBc^{-1}=B$. Hence, $c\in\mathrm{st}(B)$ and $BbB=Bc=cB$. Likewise, as any nonzero $t\in S$ can be thrown into B by multiplying by a nonzero element of F, BtB can be generated by an element of $\mathrm{st}(B)$ as a cyclic left and right B-module.

Now consider rings E, A with $B\subseteq E\subseteq A\subseteq S$. For any $t\in\mathrm{st}(E)$ we have just seen there is an $s\in\mathrm{st}(B)$, such that $BtB=Bs=sB$. Then, as $tE=EtE$, we have $tE=BtE=BtBE=sBE=sE$, yielding $t=su$ with $u\in E^*$. Thus, $\mathrm{st}(E)=\mathrm{st}(B)\cdot E^*$. Likewise, $\mathrm{st}(A)=\mathrm{st}(B)\cdot A^*=\mathrm{st}(B)\cdot E^*\cdot A^*=\mathrm{st}(E)\cdot A^*$. Let $\tilde{E}=E/J(A)\subseteq\bar{A}$, $T=E\cap F$, $W=A\cap F$, and $\tilde{T}=T/J(W)=\tilde{E}\cap\bar{W}$. By Corollary B, $Z(\bar{A})$ is normal over $\bar{W}$ and θ_A maps $\mathrm{st}(A)$ onto $\mathscr{G}(Z(\bar{A})/\bar{W})$. Hence, $\mathrm{st}(A)$ acts transitively on the set of extensions of $\tilde{T}$ to $Z(\bar{A})$. But in this action every element of $\mathrm{st}(E)\cdot A^*$ sends $\tilde{E}\cap Z(\bar{A})$ to itself. Thus, $\tilde{E}\cap Z(\bar{A})$ is the only extension of $\tilde{T}$ to $Z(\bar{A})$. So, $\ell_{E,A}=1$, proving (v). This completes the proof of Theorem F. □

References

[B] Bourbaki, N.: Algèbre commutative. Che. 6, valuations. Paris: Hermann 1964

[BG1] Brungs, H.H., Gräter, J.: Valuation rings in finite dimensional division algebras. J. Algebra (to appear)

[BG2] Brungs, H.H., Gräter, J.: Extensions of valuation rings in central simple algebras. Trans. Am. Math. Soc. (to appear)

[C] Cohn, P.M.: Algebra, Vol. 2. London: Wiley 1977

[DI] Demeyer, F., Ingraham, E.: Separable algebras over commutative rings. (Lectures Notes in Math., Vol. 181). Berlin Heidelberg New York: Springer 1971

[DR] Draxl, P.: Ostrowski's theorem for Henselian valued skew fields. J. Reine Angew. Math. **354**, 213–218 (1984)

[DK] Draxl, P., Kneser, M. (eds.): SK_1 von Schiefkörpern. (Lecture Notes in Math., Vol. 778). Berlin: Springer 1980

[D1] Dubrovin, N.I.: Noncommutative valuation rings. Tr. Mosk. Mat. O.-va., **45**, 265–280 (1982). English trans: Trans. Mosc. Math. Soc. **45**, 273–287 (1984)

[D2] Dubrovin, N.I.: Noncommutative valuation rings in simple finite-dimensional algebras over a field. Mat. Sb. **123**, 496–509 (1984); English trans: Math. USSR Sb. **51**, 493–505 (1985)

[E] Endler, O.: Valuation theory. New York: Springer 1972

[Er1] Ershov, Yu.L.: Henselian valuations of division rings and the group SK_1. Mat. Sb. **117**, 60–68 (1982); English transl.: Math. USSR Sb. **45**, 63–71 (1983)

[Er2] Ershov, Yu.L.: Valued division rings, pp. 53–55, in Fifth All Union Symposium on the Theory of Rings, Algebras, and Modules, Akad. Nauk SSSR Sibirsk. Otdel., Inst. Mat., Novosibirsk, 1982 (in Russian)

[F] Formanek, E.: Noetherian PI-Rings. Commun. Algebra **1**, 79–86 (1974)

[G] Gräter, J.: Valuations on finite-dimensional division algebras and their value groups. Arch. Math. **51**, 128–140 (1988)

[H] Hua, L.K.: A note on the total matrix ring over a non-commutative field. Ann. Soc. Math. Pol. **25**, 188–198 (1952)

[JW] Jacob, B., Wadsworth, A.: Division algebras over Henselian fields. J. Algebra (to appear)

[K] Kasch, F.: Invariante Untermoduln des Endomorphismenringes eines Vektorraums. Arch. Math. **4**, 182–190 (1953)

[Ma] Mathiak, K.: Valuations of skew fields and projective Hjelmslev spaces. (Lecture Notes in Math., Vol. 1175). Berlin Heidelberg New York: Springer 1986

[M1] Morandi, P.: The Henselization of a valued division algebra. J. Algebra (to appear)

[M2] Morandi, P.: Valuation rings in division rings and central simple algebras, doctoral dissertation, Univ. of Calif. at San Diego, 1988

[OS] Ojanguren, M., Sridharan, R.: Cancellation of Azumaya algebras. J. Algebra **18**, 501–505 (1971)

[Re] Reiner, I.: Maximal orders. London: Academic Press 1975

[R1] Ribenboim, P.: Théorie des valuations. Montréal: Presses Univ. Montréal 1968

[R2] Ribenboim, P.: Equivalent forms of Hensel's lemma. Expo. Math. **3**, 3–24 (1985)

[Ro] Rosenberg, A.: The Cartan-Brauer-Hua theorem for matrix and local matrix rings. Proc. Am. Math. Soc. **7**, 891–898 (1956)

[Rw] Rowen, L.H.: Polynomial identities in ring theory. New York: Academic Press 1980

[Sa] Saltman, D.: The Brauer group and the center of generic matrices. J. Algebra **97**, 53–67 (1985)

[S] Schilling, O.F.G.: The theory of valuations. Math. Surveys, No. 4, Am. Math. Soc., Providence, R.I., 1950

[W1] Wadsworth, A.R.: Extending valuations to finite dimensional division algebras. Proc. Am. Math. Soc. **98**, 20–22 (1986)

[W2] Wadsworth, A.R.: Dubrovin valuation rings, pp. 359–374. In Perspectives in ring theory. F. van Oystaeyen, L. Le Bruyn (eds.). NATO ASI Series, Series C, Vol. 233. Dordrecht: Kluwer 1988

Received July 6, 1987

Math. Ann. 283, 329–332 (1989)

Mathematische Annalen

Immersed Hypersurfaces with Constant Weingarten Curvature

Klaus Ecker[1] and Gerhard Huisken[2]

[1] Centre for Mathematical Analysis, Australian National University, GPO Box 4, Canberra A.C.T. 2601, Australia

[2] Department of Mathematics, R.S. Phys. S. Australian National University, GPO Box 4, Canberra A.C.T. 2601, Australia

In a recent paper Korevaar [5] used the Alexandrov reflection principle to show that closed embedded hypersurfaces in $\mathbb{R}^{n+1}$, $\mathbb{H}^{n+1}$ or the upper hemisphere of $\mathbb{S}^{n+1}$ are umbilic spheres provided a certain function f of the principal curvatures $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ is constant. He only had to assume that f is positive on the positive cone $\mathscr{C}=\{\lambda|\lambda_i>0\ \forall i\}$ and that f is elliptic on the component Γ of $\{\lambda\,|\,f(\lambda)>0\}$ which contains $\mathscr{C}$. Here f is said to be elliptic if $\partial f/\partial\lambda_i>0$ for all i, $1\leqq i\leqq n$. This generalization of earlier sphere theorems (see [9] and [10] for references) cannot be extended to hypersurface immersions in view of recent counterexamples, [11].

However, assuming additional curvature conditions Walter derived in [10] global results for hypersurface immersions in a space $N^{n+1}(c)$ of constant curvature c, which have a constant higher mean curvature function H_r. Here H_r is the r-th symmetric function of the principal curvatures. It was shown that such hypersurfaces are of constant mean curvature H_1, provided they have non-negative sectional curvature *and* non-negative principal curvatures. As a consequence they have to be isoparametric with at most two distinct principal curvatures and can therefore be completely classified.

Here we show that it is not necessary to assume all principal curvatures to be non-negative. Moreover we extend Walter's result to general symmetric functions $f=f(\lambda)$. Let M^n be a smooth, connected and compact manifold without boundary. Then we have the following result.

1. Theorem. *Let $F:M^n\to N^{n+1}(c)$ be a smooth isometric hypersurface immersion of M^n into a Riemannian manifold of constant curvature c, such that the sectional curvature of M^n is non-negative. Assume that $f=f(\lambda)$ is a smooth symmetric function of $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ satisfying the following conditions*

i) *$f(\lambda)>0$ whenever $\lambda\in\mathscr{C}$.*

ii) *On the component $\Gamma(f)$ of $\{\lambda\,|\,f(\lambda)>0\}$ containing $\mathscr{C}$, f is elliptic (i.e. $\partial f/\partial\lambda_i>0\ \forall i$) and concave (i.e. $[\partial^2 f]\leqq 0$).*

If $f(\lambda)\equiv\mathrm{const}>0$ for each principal curvature vector $\lambda=\lambda(p)$, $p\in M^n$, and if $\lambda(p_0)\in\Gamma(f)$ for some point $p_0\in M^n$, then the mean curvature H_1 of M^n is constant and hence M^n is isoparametric with at most two distinct principal curvatures.

Remarks. i) We only have to show that the mean curvature H_1 is constant: Nomizu and Smyth established in [8] that then $F(M^n)$ has parallel second fundamental form. This in turn implies that F is isoparametric with at most two distinct principal curvatures by a result of Lawson, [6].

ii) The condition that $\lambda(p_0) \in \Gamma(f)$ for at least one point $p_0 \in M$ is automatically satisfied if $F(M^n)$ is contained in $\mathbb{R}^{n+1}$, $\mathbb{H}^{n+1}$ or in the upper hemisphere of $\mathbb{S}^{n+1}$. More generally it is sufficient that $F(M^n)$ lies in the domain of a strictly convex function, compare [10, Remark 5B.].

iii) It is shown in [4] that $f=(H_r)^{1/r}$ satisfies our conditions for all r, $1 \leqq r \leqq n$. Other examples including the harmonic means functions $f(\lambda)=(\lambda_1^{-1}+\lambda_2^{-1} + \dots \lambda_n^{-1})^{-1}$ can be found in [7, Chap. 2].

For the proof of Theorem 1 we need an inequality for concave symmetric functions in the plane.

2. Lemma. *Let $f=f(x,y)$ be a symmetric function on $\mathbb{R}^2$ which is concave on an open convex and symmetric subset G. Then the inequality*

$$\left(\frac{\partial f}{\partial x}(x_0,y_0)-\frac{\partial f}{\partial y}(x_0,y_0)\right)(x_0-y_0) \leqq 0$$

holds for every $(x_0,y_0) \in G$.

Proof. Consider the straight line $\gamma(t)$ orthogonal to the diagonal $\{x=y\}$ through (x_0,y_0), parametrized by

$$\gamma(t)=\tfrac{1}{2}(x_0+y_0-2t, x_0+y_0+2t).$$

Since f is concave and symmetric in x and y, $f|_\gamma$ has a maximum at $t=0$, is nonincreasing in t for $t \geqq 0$ and nondecreasing in t for $t \leqq 0$. As

$$\frac{d}{dt} f(\gamma)(t)=\frac{\partial f}{\partial y}(\gamma(t))-\frac{\partial f}{\partial x}(\gamma(t)),$$

this implies the desired inequality.

Proof of Theorem 1. Since F is a smooth immersion and f is a smooth symmetric function of the λ_i's, f as a function of the principal curvature vector $\lambda(p)$, $p \in M^n$, is a smooth function on M^n. Let ∇ denote covariant differentiation on $F(M^n)$ and let f'_{ij} be the derivative of f when considered as a function of the second fundamental form $A=\{h_{ij}\}$. Since $f(h_{ij})$ is constant on M^n by assumption, the Laplace-Beltrami operator Δ applied to f yields zero. Computing in a local orthonormal frame and summing over repeated indices we obtain

$$\begin{aligned} 0=\Delta f &= \nabla_m \nabla_m f = \nabla_m(f'_{ij}\nabla_m h_{ij}) \\ &= f'_{ij}\Delta h_{ij} + f''_{ijkl}\nabla_m h_{ij}\nabla_m h_{kl}. \end{aligned}$$

Now observe that as in [10, identity 3.16]

$$\Delta h_{kl} = \nabla_k \nabla_l H_1 + R_{ikim} h_{ml} + R_{iklm} h_{mi},$$

where R_{ijkl} is the curvature tensor on $F(M^n)$. Thus we obtain

$$\begin{aligned} 0 &= f'_{ij}\nabla_i\nabla_j H_1 + f''_{ijkl}\nabla_m h_{ij}\nabla_m h_{kl} \\ &\quad + f'_{ij}(R_{kikm}h_{mj}+R_{kijm}h_{mk}). \end{aligned}$$

Now write

$$\sigma_{ij} = R_{ijij} \quad \text{(no sum)}$$

for the sectional curvatures on M^n and rotate at a given point the coordinate system such that $\{h_{ij}\}$ is diagonal. Then the last equation reads

$$0 = \sum_i \frac{\partial f}{\partial \lambda_i} \nabla_i \nabla_i H_1 + \sum_{i,k} \frac{\partial^2 f}{\partial \lambda_i \partial \lambda_k} \nabla_m h_{ii} \nabla_m h_{kk} + \sum_{i,k} \frac{\partial f}{\partial \lambda_i} (\sigma_{ik}\lambda_i - \sigma_{ik}\lambda_k).$$

The second term on the RHS is non-positive due to the concavity of f. The last term can be written as

$$\frac{1}{2} \sum_{i,k} \left(\frac{\partial f}{\partial \lambda_k} - \frac{\partial f}{\partial \lambda_i} \right) (\lambda_k - \lambda_i) \sigma_{ik}.$$

When restricted to the variables (λ_i, λ_k), f satisfies the conditions of Lemma 2 for each pair (i, k). Since $\sigma_{ik} \geqq 0$ by assumption, the last expression is less than or equal to zero. Hence we finally conclude

$$\sum_i \frac{\partial f}{\partial \lambda_i} \nabla_i \nabla_i H_1 \geqq 0.$$

Our ellipticity assumption on f and the strong maximum principle then yield $H_1 \equiv \text{const}$, completing the proof of Theorem 1.

We may now proceed as in [10, Sect. 4] to classify all isoparametric hypersurfaces with at most two distinct principal curvatures which satisfy our curvature assumptions. The only difference appears in the case $c > 0$, where we get a much larger class of examples. The additional examples are generalized Clifford tori which arise since we are not restricted to hypersurfaces with non-negative principal curvatures. This partially generalizes the result of Cheng and Yau in [3], where the special case $f = (H_2)^{1/2}$ was considered.

Using similar notation as in [10], we define the family of hypersurfaces $\mathscr{H}_c$ in $\mathbb{R}^{n+1}$, $\mathbb{H}^{n+1}(c)$, and $\mathbb{S}^{n+1}(c)$ as follows

For $c > 0$: $\mathscr{H}_c$ is the family of all small umbilic hyperspheres and generalized Clifford tori in $\mathbb{S}^{n+1}(c)$.

For $c = 0$: $\mathscr{H}_c$ is the family of all hyperspheres and orthogonal spherical hypercylinders in $\mathbb{R}^{n+1}$.

For $c < 0$: $\mathscr{H}_c$ is the family of all geodesic distance spheres, horospheres, and geodesic hypercylinders in $\mathbb{H}^{n+1}(c)$.

Furthermore, given a complete space $N^{n+1}(c)$ of constant curvature c, let $\varrho(\mathscr{H}_c)$ be the image of the family $\mathscr{H}_c$ under the associated universal covering ϱ. Then proceeding as in [10] we obtain the following consequence of Theorem 1.

3. Corollary. *Let $F: M^n \to N^{n+1}(c)$ be an isometric immersion into a complete Riemannian manifold of constant sectional curvature c such that the sectional curvature of M^n is non-negative. Let the function $f = f(\lambda)$ be as in Theorem 1 with $\lambda(p_0) \in \Gamma(f)$ for some $p_0 \in M^n$. If $f(\lambda) \equiv \text{const} > 0$ on $F(M^n)$, then $F(M^n) \in \varrho(\mathscr{H}_c)$.*

In the special case where $c > 0$ and $N^{n+1}(c) = \mathbb{P}^{n+1}(c)$ is the $(n+1)$-dimensional real projective space, we conclude from [10, Lemma 4.7]:

4. Corollary. *Let $F: M^n \to \mathbb{P}^{n+1}(c)$ be an isometric immersion into $\mathbb{P}^{n+1}(c)$ and let $f = f(\lambda)$ be as in Theorem 1. If $F(M^n)$ has non-negative sectional curvature, $\lambda(p_0) \in \Gamma(f)$ for some point $p_0 \in M^n$ and if $f(\lambda) \equiv \mathrm{const} > 0$ on $F(M^n)$, then F is an embedding onto a distance sphere of radius $< \pi/2\sqrt{c}$ or a covering map onto a Clifford quadric in $\mathbb{P}^{n+1}(c)$.*

Finally we can extend Corollary 4.C in [10] to general functions $f(\lambda)$ without assuming a lower bound on the sectional curvature.

5. Corollary. *Let $F: M^n \to N^{n+1}(c)$ be an isometric hypersurface immersion into a Riemannian manifold $N^{n+1}(c)$ of constant sectional curvature c and let $f = f(\lambda)$ be as in Theorem 1. If $F(M^n)$ has strictly positive sectional curvature, $\lambda(p_0) \in \Gamma(f)$ for some point $p_0 \in M^n$ and if $f(\lambda) \equiv \mathrm{const} > 0$ on $F(M^n)$, then $F(M^n)$ is umbilic and has constant curvature.*

Proof. This is an immediate consequence of Corollary 3 since positive sectional curvature rules out all examples with a product structure.

References

1. Caffarelli, L., Nirenberg, L., Spruck, J.: Nonlinear second order elliptic equations IV, Starshaped compact Weingarten hypersurfaces. Current topics in partial differential equations. Y. Ohya, K. Kasahara, N. Shimakura (eds.), pp. 1–26. Tokyo: Kinokunze 1986
2. Cartan, E.: Families de surfaces isoparamétrique dans les espaces à courbure constante. Ann. Mat. Pura Appl. **17**, 177–191 (1938)
3. Cheng, S.Y., Yau, S.T.: Hypersurfaces with constant scalar curvature. Math. Ann. **247**, 81–93 (1980)
4. Garding, L.: An inequality for hyperbolic polynomials. J. Math. Mech. **8**, 957–965 (1959)
5. Korevaar, N.J.: Sphere theorems via Alexandrov for constant Weingarten curvature hypersurfaces – Appendix to a note of A. Ros. J. Differ. Geom. **27**, 221–223 (1988)
6. Lawson, H.B., Jr.: Local rigidity theorems for minimal hypersurfaces. Ann. Math. **89**, 187–197 (1969)
7. Mitrinovic, D.S.: Analytic inequalities. Grundlehren der Math. Wissenschaften. Berlin Heidelberg New York: Springer 1970
8. Nomizu, K., Smyth, B.: A formula of Simons' type and hypersurfaces with constant mean curvature. J. Differ. Geom. **3**, 367–377 (1969)
9. Ros, A.: Compact hypersurfaces with constant scalar curvature and a congruence theorem. J. Differ. Geom. **27**, 215–220 (1988)
10. Walter, R.: Compact hypersurfaces with a constant higher mean curvature function. Math. Ann. **270**, 125–145 (1985)
11. Wente, H.C.: Counterexample to a conjecture of H. Hopf. Pac. J. Math. **121**, 193–243 (1986)
12. Yau, S.T.: Submanifolds with constant mean curvature, I, II. Am. J. Math. **96**, 346–366 (1974); **97**, 76–100 (1975)

Received June 3, 1988

Math. Ann. 283, 333–351 (1989)

Mathematische Annalen

Classification of Supersingular Abelian Varieties

Ke-Zheng Li

Nankai Institute of Mathematics, Tianjin, People's Republic of China

0. Introduction

Let k be an algebraically closed field of characteristic $p>0$. An abelian variety X over k is called *supersingular* if $H^1_{\mathrm{cris}}(X/W)$ has Newton slopes all equal to $\frac{1}{2}$, or equivalently, if there is an isogeny $\varrho: E^g \to X$, where E is a supersingular elliptic curve, and $g = \dim(X)$. (This comes from a series of work by Eichler, Oort, Deligne, Shioda and Ogus. See [22, p. 113], [21, p. 35], [23, p. 586], [19, p. 59]). Oda and Oort [18] have studied the classification problem of supersingular abelian varieties in the case when $a(X)=1$ at the crystal level, where $a(X) = \dim_k \mathrm{Hom}(\alpha_p, X)$ [18, p. 595]. The other cases are relatively special but much more complicated.

In this paper we first study the classification problem of all cases at the crystal level (Sect. 1). This solves a problem left open by Oda and Oort [18]. Then, in Sect. 2, we study the fine moduli problem. Two kinds of additional structures are given to the families of supersingular abelian varieties so that they have fine moduli. The proofs are constructive.

In Sect. 3, we study what appears to be a difficult question: How can we recover a family of supersingular abelian varieties from its crystalline cohomology? This is solved by virtue of the so-called "α-sheaf".

In Sect. 4 we deal with the problem of how, for an arbitrary family of supersingular abelian varieties, we can modify it so that it has one of the additional structures of Sect. 2 (and hence is induced by a morphism to the fine moduli). For one of the structures, namely the "level structure", this is solved by Ogus.

Finally, we study the coarse moduli problem in Sect. 5. We prove that for any set of integer invariants, there is a coarse moduli space of supersingular abelian varieties having that set of integer invariants. As an example, we calculate the number of points in Oda-Oort's space which correspond to one isomorphism class of abelian varieties.

This paper contains the main part of the dissertation work directed by Professor Arthur Ogus, as well as some unpublished results of Ogus.

I wish to thank Professor Robin Hartshorne for much good advice. Also I wish to thank Professor Ken Ribet and Professor Lucien LeCam for their review of the dissertation.

1. Supersingular Dieudonné Crystals

We use the notation of [18].

Let $W = W(k)$ be the Witt ring of k, and $A = W[F, V]$ be the associative ring satisfying the following defining relations:

i) $FV = VF = p$;

ii) $Fa = a^{\sigma}F$, $Va = a^{\sigma^{-1}}V$ for all $a \in W$, where σ is the absolute Frobenius map.

Definition. A *Dieudonné crystal* is a left A-module which is free of finite rank as a W-module.

Let $A_{1,1} = A/A(F - V)$. A finite direct sum of (say g copies of) $A_{1,1}$ is called a *superspecial* Dieudonné crystal (of genus g). A Dieudonné crystal M is called *supersingular* (of genus g) if it is isomorphic to an A-submodule of a superspecial Dieudonné crystal (of genus g) of finite colength. A *trace map* [19, Sect. 6] of M is an isomorphism $\mathrm{tr}: \Lambda^{2g}M \to W[-g]$ of Dieudonné crystals, where $W[-g]$ is free of rank 1 over W generated by some element x such that $Fx = Vx = p^g x$.

Clearly the isogeny ϱ in Sect. 0 induces a supersingular Dieudonné crystal structure on $H^1_{\mathrm{cris}}(X/W)$, and it has a canonical trace map coming from $\Lambda^{2g}H^1_{\mathrm{cris}}(X/W) \simeq H^{2g}_{\mathrm{cris}}(X/W)$. If $g > 1$, $H^1_{\mathrm{cris}}(X/W)$ is superspecial iff $X \simeq E^g$ for any supersingular elliptic curve E (cf. [23, p. 586]). In this case we also say that X is *superspecial*. Ogus proved the following Torelli theorem [19, Theorem 6.2].

Theorem. *If* $g > 1$, *then the functor* $(H^1_{\mathrm{cris}}, \mathrm{tr})$ *defines a bijection*

$$\frac{\left\{\begin{matrix}\textit{supersingular abelian varieties}\\ \textit{of dimension g over k}\end{matrix}\right\}}{\textit{isomorphisms}} \xrightarrow{\sim} \frac{\left\{\begin{matrix}\textit{supersingular Dieudonné crystals}\\ \textit{of genus g together with a trace map}\end{matrix}\right\}}{\textit{isomorphisms}}$$

furthermore, for any two supersingular abelian varieties X, Y *of dimension g, there is a canonical isomorphism*

$$\mathrm{Hom}(Y, X) \otimes_{\mathbb{Z}} \mathbb{Z}_p \xrightarrow{\sim} \mathrm{Hom}_{A,\mathrm{tr}}(H^1_{\mathrm{cris}}(X/W), H^1_{\mathrm{cris}}(Y/W)).$$

The following lemmas are purely elementary, and the proofs are left to the reader.

Lemma 1.1. *Suppose that* M *is a supersingular Dieudonné crystal, and* $N \subset M$ *is an* A*-submodule such that* $N + (F, V)M = M$. *Then* $N = M$.

Let $\dot{M} = Ax_1 \oplus Ax_2 \oplus \ldots \oplus Ax_g$ be a superspecial Dieudonné crystal of genus g, where $Ax_i \simeq A_{1,1}$, i.e., the annihilator of x_i is $A(F - V)$ $(1 \leqq i \leqq g)$. To give an endomorphism of $\dot{M}$ (as an A-module) is equivalent to giving $x'_1, \ldots, x'_g \in \dot{M}$ such that $(F - V)x'_i = 0$ for all i, $1 \leqq i \leqq g$. Let $x \in \dot{M}$. Then we can uniquely write

$$x = (a_1 + b_1F)x_1 + (A_2 + b_2F)x_2 + \ldots + (a_g + b_gF)x_g,$$

where $a_1, \ldots, a_g, b_1, \ldots, b_g \in W$. Hence $(F - V)x = 0$ if and only if

$$0 = (F - V)x = \sum_{i=0}^{g} (p(b_i^{\sigma} - b_i^{\sigma^{-1}}) + (a_i^{\sigma} - a_i^{\sigma^{-1}})F)x_i,$$

i.e., $b_1^{\sigma} = b_1^{\sigma^{-1}}$, $a_1^{\sigma} = a_1^{\sigma^{-1}}$, or $a_i, b_i \in W(\mathbb{F}_{p^2})$ $(1 \leqq i \leqq g)$.

Let $H=W(\mathbb{F}_{p^2})[F]/(F^2-p, Fa-a^\sigma F, a\in W(\mathbb{F}_{p^2}))$. Then it is easy to check that H is a generalized quaternion algebra [7, Sect. 2.4] over $W(\mathbb{F}_p)$. Furthermore, there is an anti-automorphism $\phi: H\to H$ given by $\phi(a+bF)=a^\sigma-bF$. This gives a "*norm*" map $\|: H\to W(\mathbb{F}_p)$ by

$$|a+bF|=(a+bF)\phi(a+bF)=aa^\sigma-pbb^\sigma.$$

Clearly $|\alpha\beta|=|\alpha|\cdot|\beta|$ for $\alpha,\beta\in H$.

Lemma 1.2. *The map* $\|: H\to W(\mathbb{F}_p)$ *is surjective.*

Now $\tilde{M}\overset{\text{def}}{=}\ker(F-V:\dot{M}\to\dot{M})$ can be viewed as an H-module:

$$\tilde{M}=Hx_1\oplus Hx_2\oplus\ldots\oplus Hx_g.$$

Clearly $\dot{M}\simeq W\otimes_{W(\mathbb{F}_{p^2})}\tilde{M}$. Hence we say that $\tilde{M}$ is the "*skeleton*" of $\dot{M}$.

Let v be the p-adic valuation map on $K(\mathbb{F}_{p^2})$ [the quotient field of $W(\mathbb{F}_{p^2})$]. Then it is clear that $v(|a+bF|)=\min(2v(a), 2v(b)+1)$ for $a,b\in K(\mathbb{F}_{p^2})$. If

$$\alpha,\beta\in H\otimes_{W(\mathbb{F}_{p^2})}K(\mathbb{F}_{p^2}),$$

then clearly $\alpha\beta^{-1}\in H$ if and only if $v(|\alpha|)\geqq v(|\beta|)$. From this one can easily deduce that any H-submodule of a free H-module is again free, using an argument similar to that for modules over a DVR.

If $\dot{M}_1,\dot{M}_2$ are two superspecial A-submodules of $\dot{M}$, then clearly the corresponding skeletons $\tilde{M}_1,\tilde{M}_2$ are H-submodules of $\tilde{M}$. Hence

$$\dot{M}_1+\dot{M}_2\simeq(\tilde{M}_1+\tilde{M}_2)\otimes_{W(\mathbb{F}_{p^2})}W \quad\text{and}\quad \dot{M}_1\cap\dot{M}_2\simeq(M_1\cap M_2)\otimes_{W(\mathbb{F}_{p^2})}W$$

are superspecial because $\tilde{M}_1$ and $\tilde{M}_2$ are free H-modules. Furthermore, if $\dot{M}_1$ and $\dot{M}_2$ are of genus g, then $\tilde{M}_1$ and $\tilde{M}_2$ are of rank g over H, and so is $\tilde{M}_1\cap\tilde{M}_2$. Hence $\dot{M}_1\cap\dot{M}_2$ is of genus g also. Therefore we get (cf. [12, Theorem 3.1])

Lemma 1.3. *The sum and intersection of two superspecial subcrystals of* $\dot{M}$ *are again superspecial. If* M *is an arbitrary subcrystal of* $\dot{M}$ *of finite colength, then there is a smallest superspecial subcrystal* $S'(M)$ *containing* M. *Furthermore, there is a largest superspecial subcrystal (of genus* g*)* $S_0(M)$ *contained in* M.

Note that $S'(M)$ and $S_0(M)$ do not depend on $\dot{M}$. They have minimality and maximality respectively in $M\otimes_W K$, where K is the quotient field of $W(k)$.

For a $g\times g$ non-singular matrix over H, we can still define its "determinant" as in [7, Sect. 2.1]. There the "determinant" of a non-singular matrix is an element of $H-(0)/C$, where C is the commutator subgroup of H^*. In our case, $\|$ clearly factors through $H-(0)/C$. Hence we can define "*determinant*" to be an element in $W(\mathbb{F}_p)$, i.e., it is 0 if the rows are linearly dependent, and for a non-singular matrix, if the "determinant" in the sense of [7] is αC, $\alpha\in H-(0)$, then we define our determinant to be $|\alpha|$.

For a matrix $T=(a_{ij}+b_{ij}F)$ over H corresponding to an endomorphism of $\dot{M}$, we can rewrite it as a W-linear transformation with respect to the basis

$x_1, Fx_1, \ldots, x_g, Fx_g$ of $\dot{M}$ over $W(\mathbb{F}_{p^2})$ as follows

$$\tilde{T}=\begin{pmatrix} a_{11} & b_{11} & \cdots & a_{1g} & b_{1g} \\ pb_{11}^{\sigma} & a_{11}^{\sigma} & \cdots & pb_{1g}^{\sigma} & a_{1g}^{\sigma} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{g1} & b_{g1} & \cdots & a_{gg} & b_{gg} \\ pb_{g1}^{\sigma} & a_{g1}^{\sigma} & \cdots & pb_{gg}^{\sigma} & a_{gg}^{\sigma} \end{pmatrix}$$

It is not hard to check that the determinant of T is just equal to $\det(\tilde{T})$. Note that H is isomorphic to the algebra of the matrices of the form $\begin{pmatrix} a & b \\ pb^{\sigma} & a^{\sigma} \end{pmatrix}$, $a, b \in W(\mathbb{F}_{p^2})$.

A matrix over H is invertible if and only if its determinant is not in $pW(\mathbb{F}_p)$. An automorphism of $\dot{M}$ preserves a given trace map if and only if the corresponding matrix over H has determinant 1. In fact,

$$\mathrm{tr} \circ \Lambda^{2g}(T) = |T| \cdot \mathrm{tr} : \Lambda^{2g}\dot{M} \to W[-g].$$

Lemma 1.4. *Let $a_1, \ldots, a_g \in k$. Then the following are equivalent.*

i) *$a_1, \ldots, a_g$ are linearly independent over $\mathbb{F}_{p^2}$;*

(ii) $$D(a_1, \ldots, a_g) \overset{\text{def.}}{=} \begin{vmatrix} a_1 & \cdots & a_g \\ a_1^{p^2} & \cdots & a_g^{p^2} \\ a_1^{p^4} & \cdots & a_g^{p^4} \\ \vdots & \ddots & \vdots \\ a_1^{p^{2g-2}} & \cdots & a_g^{p^{2g-2}} \end{vmatrix} \neq 0.$$

In fact, $D(y_1, \ldots, y_g) = (-1)^{g(g-1)/2} \prod_{\substack{1 \leqq i \leqq g \\ \lambda_1, \ldots, \lambda_g \in \mathbb{F}_{p^2}}} (y_i + \lambda_{i+1} y_{i+1} + \ldots + \lambda_g y_g)$.

In the following, we will denote by $\bar{a}$ the image of $a \in W$ under $W \to W/pW \simeq k$.

Lemma 1.5. *Let*

$$v = (a_1 + b_1 F)x_1 + \ldots + (a_g + b_g F)x_g \in \dot{M}.$$

Then $\dot{M} = S'(Av)$ if and only if $D(\bar{a}_1, \ldots, \bar{a}_g) \neq 0$. And in this case, $F^{g-1}\dot{M} = (F, V)^{g-1}v$.

Lemma 1.6. *Let M be an arbitrary supersingular Dieudonné crystal. Then there exists $v \in M - FS'(M)$ such that $S'(Av) = S'(M)$.*

For an arbitrary M, let

$$S(M) = \{v \in M \otimes_W K \mid (F, V)^{g-1} v \subset M\}, \qquad S_0'(M) = (F, V)^{g-1} M.$$

By Lemma 1.5 and 1.6, there exists $v \in S(M)$ such that

$$(F, V)^{g-1} v = F^{g-1} S'(S(M)) \subset M.$$

Hence $S'(S(M)) \subset S(M)$, i.e., $S(M)$ is superspecial.

Corollary 1.7. *The relations* $F^{g-1}S(M)=S_0(M)$, $F^{g-1}S'(M)=S'_0(M)$ *hold.*

We will call $S(M)$ the *level structure* of M, and $S'(M)$ the *colevel structure* of M.

Remark 1.8. The dual W-module $M^\vee=\mathrm{Hom}_W(M,W)$ has a natural A-module structure. Namely, if $m\in M$, $f\in M^\vee$, then $(Ff)(m)=f(Vm)^\sigma$, $(Vf)(m)=f(Fm)^{\sigma^{-1}}$ [17, p. 83]. It is clear that $S(M)^\vee=S'_0(M^\vee)$, $S'(M)^\vee=S_0(M^\vee)$, as W-submodules of $\mathrm{Hom}_W(M,K)$.

Let $M^i=M\cap F^iS'(M)$. Then $F:M^i\to M^{i+1}$ induces a map

$$\bar{F}:\bar{M}^i=M^i/M^{i+1}\to\bar{M}^{i+1}=M^{i+1}/M^{i+2}.$$

Similarly V induces a map $\bar{V}:\bar{M}^i\to\bar{M}^{i+1}$. It is clear that $\bar{F}$ (resp. $\bar{V}$) is injective because it is induced by a bijection from $F^iS'(M)/F^{i+1}S'(M)$ to $F^{i+1}S'(M)/F^{i+2}S'(M)$.

Lemma 1.9. *For a given i,* $\mathrm{Im}\,\bar{F}\neq\mathrm{Im}\,\bar{V}:\bar{M}^i\to\bar{M}^{i+1}$ *and hence* $\dim_k\bar{M}^i<\dim_k\bar{M}^{i+1}$, *unless* $\dim_k\bar{M}^i=g$.

Corollary 1.10. *If* $M=Av$, *then* $S(M)=S'(M)$.

Conversely, if M is not generated by one vector, then $S(M)\neq S'(M)$. This is clear because $\dim_k\bar{M}^{g-2}=g$ in this case, hence $M^{g-2}\subsetneqq S'_0(M)$ is already superspecial.

We will use the following notation:

$$s'_i=\dim_k(M\cap F^{i-1}S'(M)/M\cap F^iS'(M))\qquad(1\leqq i\leqq g),$$

$$t'_i=\dim_k((F,V)M\cap F^iS'(M)/(F,V)M\cap F^{i+1}S'(M))\qquad(1\leqq i\leqq g-1),$$

$$s_i=\dim_k(M\cap F^{i-1}S(M)/M\cap F^iS(M))\qquad(1\leqq i\leqq g),$$

$$t_i=\dim_k(F^{-1}M\cap V^{-1}M\cap F^{i-1}S(M))/(F^{-1}M\cap V^{-1}M\cap F^iS(M))\qquad(1\leqq i\leqq g-1).$$

Clearly s'_i, t'_i, s_i, t_i are integer invariants satisfying the following inequalities and equalities:

i) $1\leqq s'_i\leqq s'_{i+1}$, $s'_g=g$, and $s'_i<s'_{i+1}$ unless $s'_i=g$;
ii) $s'_i\leqq t'_i\leqq s'_{i+1}$;
iii) $s_i\leqq s_{i+1}$, $s_g=g$, and $s_i<s_{i+1}$ unless $s_{i+1}=0$;
iv) $s_i\leqq t_i\leqq s_{i+1}$;
v) $s_i(M)+s'_{g-1}(M^\vee)=g$;
vi) $a(M)=\sum_{i=1}^{g}s'_i-\sum_{i=1}^{g-1}t'_i$, where $a(M)=\dim_k(M/(F,V)M)$ [18, p. 598], and so on.

A sequence of integers $(s_1,\ldots,s_g)=s$ will be called an *index* if $0\leqq s_1\leqq\ldots\leqq s_g=g$, and $s_i<s_{i+1}$ unless $s_{i+1}=0$.

Proposition 1.11. *For any supersingular Diedudonné crystal M, there is a canonical filtration* $M^i=M\cap F^i\dot{M}$ $(0\leqq i\leqq g-1)$, *where* $\dot{M}=S'(M)=Ax_1\oplus\ldots\oplus Ax_g$ *is superspecial. Then* $\bar{M}^i=M^i/M^{i+1}$ *can be viewed as a k-vector subspace of* $F^i\dot{M}/F^{i+1}\dot{M}$. *F (resp. V) induces an injective σ-linear (resp.* σ^{-1}*-linear) homomorphism* $\bar{M}^i\to\bar{M}^{i+1}$, *and there is* $v\in M-F\dot{M}$ *such that* $\dot{M}=S'(Av)$. *Conversely, given a superspecial crystal* $\dot{M}=Ax_1\oplus\ldots\oplus Ax_g$, *and a k-linear subspace* $\bar{M}^i$ *of* $F^i\dot{M}/F^{i+1}\dot{M}$ *for each i* $(0\leqq i\leqq g-1)$ *such that* $(\bar{F},\bar{V})\bar{M}^i\subseteqq\bar{M}^{i+1}$, *and such that* $\bar{M}^0$ *contains a vector*

$(\bar{a}_1, \ldots, \bar{a}_g)$ satisfying $D(\bar{a}_1, \ldots, \bar{a}_g) \neq 0$, there is a unique supersingular Dieudonné crystal $M \subseteq \dot{M}$ such that $S'(M) = \dot{M}$ and $M^i/M^{i+1} = \bar{M}^i$ as k-subspaces of $F^i\dot{M}/F^{i+1}\dot{M}$.

2. Flag Type Level Structures

We first fix some notation on group schemes. For the basic facts of group schemes, see [4–6] or [14].

Let S be a scheme of finite type over k. Let $\pi: G \to S$ be a group scheme with multiplication $m: G \times_S G \to G$, zero section $o: S \to G$ and inverse $\iota: G \to G$, all being morphisms over S. We will always assume that π is separated, hence o is a closed immersion. Let $\mathcal{M}$ be the ideal sheaf of o. Let $\omega_{G/S} = o^*\mathcal{M}$. Then canonically $\Omega^1_{G/S} \simeq \pi^*\omega_{G/S}$, and

$$\mathrm{Lie}(G/S) \simeq \mathcal{H}om_S(\omega_{G/S}, O_S) \simeq \omega_{G/S}^\vee,$$

where $\mathrm{Lie}(G/S)$ is the sheaf of (left) invariant derivations.

We say $G \to S$ is an *abelian scheme* if π is flat and proper with geometrically integral fibers. It is called *supersingular* if each of its closed fibers over S is a supersingular abelian variety.

For convenience, we will use the following notation. For any scheme $\tau: X \to S$, and any base change $\sigma: T \to S$, denote $X \times_\sigma T = X \times_S T$ in order to emphasize σ. In particular, if $\sigma = F^n$, where $F: S \to S$ is the Frobenius map, we write $X^{(p^n)} = X \times_\sigma S$.

If π is flat, finite and commutative, then we can define its *Cartier dual* $G^\vee$ (see [14, III.14]). The *relative Frobenius map* $F_{G/S}: G \to G^{(p)}$ is the unique morphism such that $\mathrm{pr}_1 \circ F_{G/S}$ is the absolute Frobenius map of G, where pr_1 is the first projection of $G^{(p)} = G \times_F S$. The *Verschiebung map* $V_{G/S}: G^{(p)} \to G$ is defined by the dual of

$$F_{G^\vee/S}: G^\vee \to (G^\vee)^{(p)} \simeq (G^{(p)})^\vee.$$

We have [6, p. 29]

Lemma 2.1. *The relations $F_{G/S} \circ V_{G/S} = p \cdot \mathrm{id}_{G^{(p)}}$, $V_{G/S} \circ F_{G/S} = p \cdot \mathrm{id}_G$ hold.*

Lemma 2.2 [14, p. 138]. *There is a canonical isomorphism of sheaves of O_S-modules*

$$\mathcal{H}om_S(G, G_{a/S}) \xrightarrow{\sim} \mathrm{Lie}(G^\vee/S).$$

The left hand side is isomorphic to the subsheaf of O_S of sections t such that $m^*(t) = t \otimes 1 + 1 \otimes t$, $o^*(t) = 0$, which will be called the *α-sheaf* of G.

Let $X \to S$ be a scheme of finite type. By a *G-action* on X we mean an S-morphism $\varrho: G \times_S X \to X$ such that

i) $\varrho \circ (m \times_S \mathrm{id}_X) = \varrho \circ (\mathrm{id}_G \times_S \varrho): G \times_S G \times_S X \to X$;

ii) $\varrho \circ (o \times_S \mathrm{id}_X) = \mathrm{id}_X: X \to X$, via $X \simeq S \times_S X$.

The action ϱ is called *free* if $(\varrho, \mathrm{pr}_2): G \times_S X \to X \times_S X$ is a closed immersion. Furthermore, ϱ is called *affine* if there exists an open affine covering $\{U_i, i \in I\}$ of S, and an open affine covering $\{V_{ij}, j \in J_i\}$ of $f^{-1}(U_i)$ for each $i \in I$, such that $\varrho(G \times_S V_{ij}) = V_{ij}$ for every pair $i \in I$, $j \in J_i$.

Lemma 2.3 [14, p. 111]. *Suppose that $G \to S$ is flat and finite and that ϱ is affine. Then ϱ has a scheme-theoretic quotient $\tau: X \to Y$ (i.e., τ is the universal morphism such that $\tau \circ \varrho = \tau \circ \mathrm{pr}_2 : G \times_S X \to Y$), where τ is finite and Y is of finite type over S. Furthermore, if ϱ is free, then τ is flat, and the quotient commutes with base change of S.*

Let $\mathscr{G}_S$ be the category of group schemes over S whose morphisms are homomorphisms. Then S is a beginning and ending object of $\mathscr{G}_S$. The kernel and cokernel are defined in the categorical sense. The kernel always exists: $\ker(f: G \to G')$ is the pullback of f and $o_G: S \to G'$. And Lemma 2.3 gives

Corollary 2.4. *Suppose H is a finite, flat, closed subgroup scheme of a commutative group scheme G of finite type over S. If the action of H on G via m is affine, then there exists a quotient group scheme G/H.*

Example. Suppose $f: G \to G'$ is an isogeny (i.e., surjective and quasi-finite) of abelian schemes. Then for any flat subgroup scheme H of $\ker(f)$, the action of H on G is affine. Indeed, by [8, p. 136], f is finite and flat, and H fixes $f^{-1}(U)$ for any open affine $U \subseteq G'$.

Definition. A finite, flat group scheme $G \to S$ is called an *α-group* over S if $F_{G/S} = 0$, $V_{G/S} = 0$.

The following fact is well known, and the proof is left to the reader.

Lemma 2.5. *If G is an α-group over S, then locally over S, G is isomorphic to $S \times \left(\alpha_p \times \overset{n}{\dots} \times \alpha_p\right)$, where $\alpha_p \simeq \operatorname{Spec} k[t]/(t^p)$ with the group scheme structure given by $m^*(t) = t \otimes 1 + 1 \otimes t$, $\iota^*(t) = -t$, $o^*(t) = 0$.*

The above n, if it is a constant, is called the *α-rank* of G. It is equal to $\operatorname{rank}_{O_S}(\omega_{G/S})$ (Note that for any flat α-group G, $\omega_{G/S}$ is flat). Furthermore, $\operatorname{rank}_{O_S}(\pi_* O_G) = p^n$. Lemma 2.5 also shows that $\omega_{G/S}$ is canonically isomorphic to the α-sheaf of G. Therefore we obtain

Corollary 2.6. *There is an anti-equivalence of categories*

$$((\text{flat coherent sheaves of } O_S\text{-modules})) \leftrightarrow ((\alpha\text{-groups over } S))$$

$$\text{the } \alpha\text{-sheaf of } G \leftarrow\!\shortmid G$$

$$\mathscr{F} \mapsto \operatorname{Spec}(\operatorname{Sym}(\mathscr{F})/\mathscr{F}^{[p]})$$

compatible with the functor $^\vee$, *where $\mathscr{F}^{[p]}$ is the ideal of $\operatorname{Sym}(\mathscr{F})$ generated by the p^{th} powers of the sections of $\mathscr{F}$. (If f is a section of $\mathscr{F}$, then $m^*(f) = f \otimes 1 + 1 \otimes f$.)*

Remark 2.7. Suppose G is a finite commutative group scheme over S such that $F_{G/S} = 0$. Then $\omega_{G/S}$ is flat iff $G \to S$ is flat. In this case, the canonical map from the α-sheaf of $G^\vee$ to $\omega_{G^\vee/S}$ is surjective.

For any supersingular abelian variety X of dimension g over k, Sect. 1 tells us that $H^1_{\text{cris}}(X/W)$ has a natural filtration, defining an index $s = (s_1, \dots, s_g)$ [and also $t = (t_1, \dots, t_{g-1})$], which is a geometric invariant. This inspires us to make the following definition. We fix a supersingular elliptic curve E because the isomorphism class of E^g is independent of the choice of E when $g > 1$, as we have seen in Sect. 1.

Definition. *A flag type level structure* of index s over S is a supersingular abelian scheme $A \to S$ together with isogenies (of supersingular abelian schemes) $\varrho_i : A_i^{(p)} \to A_{i-1} (1 \leqq i \leqq g-1)$ over S such that

i) $A_{g-1} = E^g \times S$, $A_0 = A$;

ii) $\ker \varrho_i$ is a flat α-group of α-rank s_i.

For a noetherian scheme S, let $\mathscr{C}_S =$ ((schemes over S)). We need the following lemma to construct the moduli spaces of supersingular abelian schemes with level structure.

Lemma 2.8. *Let $\mathscr{F}$ be a locally free sheaf of O_S-modules of* rank r. *Let $\mathscr{F}'$ be a coherent O_S-submodule of $\mathscr{F}$. Then the functor $P : \mathscr{C}_S \to$ ((sets)) defined by*

$$P(\tau : X \to S) = \{\textit{locally free quotient } O_S\textit{-modules } \mathscr{E} \textit{ of } \tau^* \mathscr{F} \textit{ of } \operatorname{rank} n\}$$

is represented by a relative Grassmannian $\mathbf{G}$. *Furthermore, the following hold.*

i) *Let $Q : \mathscr{C}_S \to$ (sets)) be defined by*

$$Q(\tau : X \to S) = \{\textit{locally free quotients } h : \tau^* \mathscr{F} \twoheadrightarrow \mathscr{E} \textit{ of } \operatorname{rank} n \textit{ such that } \tau^* \mathscr{F}' \subseteqq \ker h\}$$

Then Q is represented by a closed subscheme $\mathbf{G}'$ of $\mathbf{G}$.

ii) *If $\mathscr{F} = \pi_* O_G$, where $\pi : G \to S$ is a finite flat group scheme, then the functor*

$$R : \mathscr{C}_S \to (sets)),$$

$$(X \to S) \mapsto \{\textit{flat closed subgroup schemes of } X \times_S G \textit{ over } X \textit{ of } \operatorname{rank} n\}$$

is represented by a closed subscheme $\mathbf{G}_G$ of $\mathbf{G}$.

iii) *Therefore, if $\mathscr{F}'$ defines a closed subgroup scheme G' of G (not necessarily flat), then the functor*

$$R' : \mathscr{C}_S \to ((sets)),$$

$$(X \to S) \mapsto \{\textit{flat closed subgroup schemes of } X \times_S G' \textit{ over } X \textit{ of } \operatorname{rank} n\}$$

is represented by $\mathbf{G}' \cap \mathbf{G}_G (= \mathbf{G}' \times_{\mathbf{G}} \mathbf{G}_G)$.

Proof. [15, p. 32] gives the universality of the (absolute) Grassmannian, and the relative case comes from abstract nonsense. Then we need to check that Q, R are determined by algebraic conditions respectively. This is boring but without any difficulties. The last statement is just abstract nonsense. Q.E.D.

For an (arbitrary) abelian scheme A over S, we can define its *Verschiebung map* as follows. Since $F_{A/S} : A \to A^{(p)}$ is flat [8, p. 136], $\cdot p : A \to A$ is also flat, and since $\ker(F_{A/S}) \hookrightarrow \ker(\cdot p)$, there exists a unique $V : A^{(p)} \to A$ such that $V \circ F = \cdot p$. When $S = \operatorname{Spec} k$, *this coincides with the usual definition* of V, as the dual of $F_{A^\vee/S}$, where $A^\vee$ is the dual of A [14, Sect. 13].

Theorem 2.9. *Given an index s (see Sect. 1), the functor*

$$T_s : ((k\textit{-schemes})) \to ((sets)),$$

$$X \mapsto \{\textit{flag type level structures of index } s \textit{ over } X\}/\textit{isomorphisms}$$

is represented by a projective scheme S_s over k.

Proof. We use inverse induction to construct projective schemes $S_{g-1}, \ldots, S_0$ together with supersingular abelian schemes $\pi_i: B_i \to S_i$ $(0 \leqq i < g)$. Take $S_{g-1} = \mathrm{Spec}(k)$, $B_{g-1} = E^g$. Given $\pi_i: B_i \to S_i$, let

$$K_1 = \ker(F_{B_i^{(p)}/S_i}: B_i^{(p)} \to B_i^{(p^2)}), \qquad K_2 = \ker(V_{B_i/S_i}: B_i^{(p)} \to B_i),$$
$$K = K_1 \cap K_2 (= K_1 \times_{B_i^{(p)}} K_2).$$

Then K is a closed subgroup scheme of a flat group scheme K_1. Note that a flat finite closed subgroup scheme of $B_i^{(p)}$ is an α-group if and only if it is a subgroup scheme of K. Now by Lemma 2.8, the functor

$$R: \mathscr{C}_{S_i} \to ((\mathrm{sets})),$$
$$(X \to S_i) \mapsto \{\text{flat closed subgroup schemes of } X \times_{S_i} K \text{ of } \alpha\text{-rank } s_i\}$$

is represented by a relatively projective morphism $S_{i-1} \to S_i$, and a flat closed subgroup scheme G of $S_{i-1} \times_{S_i} K$ of α-rank s_i. Then let $B_{i-1} = B_i^{(p)} \times_{S_{i-1}} / G$.

Clearly $S_s = S_0$ is projective over k. We claim that S_s represents T_s.

Suppose for $S \to \mathrm{Spec}(k)$, a flag type level structure

$$\{A_i \to S,\ \varrho_i: A_i^{(p)} \to A_{i-1}\ (0 \leqq i < g)\}$$

is given. Let $\sigma_{g-1}: S \to S_{g-1} = \mathrm{Spec}(k)$ be the (unique) trivial morphism. Then $A_{g-1} \simeq B_{g-1} \times_{S_{g-1}} S$. Given $\sigma_i: S \to S_i$ such that $A_i \simeq B_i \times_{S_i} S$, by Lemma 2.8 there exists a unique $\sigma_{i-1}: S \to S_{i-1}$ such that $G \times_{S_{i-1}} S = \ker(\varrho_i: A_i^{(p)} \to A_{i-1})$ as subgroup schemes of $K \times_{S_i} S$. Hence

$$\begin{aligned} A_{i-1} &\simeq A_i^{(p)}/G \times_{S_{i-1}} S \simeq B_i \times_{\sigma_i \circ F_S} S/G \times_{S_{i-1}} S \simeq B_i \times_{F \circ \sigma_i} S/G \times_{S_{i-1}} S \\ &\simeq B_i^{(p)} \times_{\sigma_i} S/G \times_{\sigma_{i-1}} S \simeq (B_i^{(p)} \times_{S_i} S_{i-1}) \times_{\sigma_{i-1}} S/G \times_{\sigma_{i-1}} S \\ &\simeq B_{i-1} \times_{S_{i-1}} S. \end{aligned}$$

Thus $A_0 \simeq B_0 \times_{S_0} S$. To see the uniqueness of the morphism $S \to S_s$, note that for $0 \leqq i < g$, $S \to S_i$ is the composite of $S \to S_s$ and $S_s \to S_i$, and that the above argument shows the uniqueness of $S \to S_i$ inductively. Q.E.D.

Remark 2.10. For a supersingular abelian scheme $A \to S$, a "*level structure*" means an isogeny $f: E^g \times S \to A$ over S such that $\ker(f) \subset \ker(F^{g-1})$. Hence $\deg(f)$ is a power p^n, and $n \leqq g(g-1)$. It is clear that the functor

$$\hat{T}_k: \mathscr{C}_k \to ((\mathrm{sets})),$$
$$S \mapsto \{\text{level structures of degree } p^n \text{ over } S\}/\text{isomorphisms}$$

is represented by a projective scheme $\hat{S}_n$.

3. Approaches via Crystals

Let M be a crystal on S such that $pM = 0$. Then $F^*_{S/W} M$ is canonically determined by M_S, the value of M on the PD-thickening $(S, 0)$, where $F_{S/W}$ is the Frobenius map. Indeed, in this case, we need only consider PD-thickenings $U \hookrightarrow T$ such that

$pO_T=0$. Then there is a unique σ_T making the following diagram commutative:

$$\begin{tikzcd} U \arrow[r] \arrow[d, "F_U"'] & T \arrow[d, "F_T"] \arrow[dl, "\sigma_T"'] \\ U \arrow[r] & T \end{tikzcd}$$

Hence $F_T^* M_T \simeq \sigma_U^* M_U$. Furthermore, given any coherent sheaf $\mathscr{E}$ on S, one can obtain a crystal (via σ) on S which is killed by p [16, Sect. 4]. Such a crystal is called *degenerate*. Thus we get a right exact functor

$$((\text{coherent sheaves on } S)) \to ((\text{crystals on } S))$$

denoted by σ^*, by abuse of notation.

By an (F, V)-crystal on S we will mean a crystal M together with morphisms $F_M: F_{S/W}^* M \to M$ and $V_M: M \to F_{S/W}^* M$ such that $F_M \circ V_M = p \cdot \mathrm{id}_M$, $V_M \circ F_M = p \cdot \mathrm{id}_{F_{S/W} M}$.

Definition. A *crystal on S with flat type level structure* of index s consists of the following data.

i) (F, V)-crystals M_i $(0 \leqq i < g)$, such that

a) $M_{g-1} \simeq \varrho^* \dot{M}$, where $\varrho: S \to \mathrm{Spec}(k)$ is the projection and $\dot{M}$ is defined in Sect. 1; and

b) M_{i-1} is an (F, V)-subcrystal of $F_{S/W}^* M_i$ containing $\mathrm{Im}(V_{M_i})$ and $\mathrm{Im}(F_{F_{S/W}^* M_i})$;

ii) Flat coherent sheaves of O_S-modules N_i of rank s_i together with epimorphisms $h_i: (M_i/\mathrm{Im}(F_{M_i}))_S \twoheadrightarrow N_i$ and isomorphisms $n_i: F_{S/W}^* M_i / M_{i-1} \xrightarrow{\sim} \sigma^* N_i$.

For a group scheme G over S, we will denote its Dieudonné module by $\mathbf{D}(G)$. We need to quote [1, Sect. 4.3] as the following lemma.

Lemma 3.1. *Let G be a flat group scheme over S. If $F_{G/S}=0$, then canonically $\mathbf{D}(G) \simeq \sigma^*(\omega_{G/S})$. If $V_{G/S}=0$, then canonically $\mathbf{D}(G) \simeq \sigma^*(\mathrm{Lie}(G^\vee/S))$.*

Theorem 3.2. *By taking Dieudonné modules, there is a fully faithful functor $\mathbf{D}_S$ from the category of isomorphism classes of flag type level structures of index s over S to the category of isomorphism classes of crystals with flag type level structures of index s on S. If σ^* is faithful, then $\mathbf{D}_S$ is a natural equivalence.*

Proof. Look at the definition in Sect. 2. Let $M_i = R^1(\pi_i)_{\mathrm{cris}*}(O_{A_i}/W)$. Then i) in the above definition is clear. For ii), first note that $M_i/\mathrm{Im}(F_{M_i}) \simeq \mathbf{D}(\ker(F_{A_i}))$. Then using Lemma 3.1 we see that

$$(M_i/\mathrm{Im}(F_{M_i}))_S \simeq F_S^* \omega_{\ker(F_{A_i})/S} \simeq \omega_{(\ker(F_{A_i}))^{(p)}/S}$$

canonically. Also $\mathbf{D}(\ker(F_{A_i^{(p)}})) \to \mathbf{D}(\ker(\varrho_i))$ is canonically induced by

$$\omega_{\ker(F_{A_i^{(p)}})/S} \simeq \omega_{(\ker(F_{A_i}))^{(p)}/S} \to \omega_{\ker(\varrho_i)/S} \simeq N_i \overset{\text{def.}}{=} \text{the } \alpha\text{-sheaf of } \ker(\varrho_i).$$

Now assume that σ^* is faithful. We use induction to define a quasi-inverse of $\mathbf{D}_S$. Given the data in the definition of crystals with flag type level structure, suppose we have constructed $A_i \to S$. Let $K_1' = \ker(F_{A_i/S})$, $K_1 = \ker(F_{A_i^{(p)}/S}) \simeq K_1'^{(p)}$, $K_2 = \ker(V_{A_i/S})$, and $K = K_1 \cap K_2$. Then ii) gives an epimorphism $h_i: (M_i/\mathrm{Im}(F_{M_i}))_S \simeq \omega_{K_1/S} \twoheadrightarrow N_i$. Furthermore, $V_{K_1'/S}$ induces $V_{K_1'/S}^*: \omega_{K_1'/S} \to \omega_{K_1/S}$, and

$\operatorname{coker}(V^*_{K'_1/S}) \simeq \omega_{K/S}$. Since σ^* is faithful, b) implies that $h_i \circ V^*_{K'_1/S} : \omega_{K'_1/S} \to N_i$ is the zero map. Hence N_i is also a quotient of $\omega_{K/S}$. Let $\mathscr{F}_2$ be the α-sheaf of K_2, and $\mathscr{F}$ be the α-sheaf of K. By Remark 2.7, the induced morphism $\mathscr{F}_2 \to \omega_{K_2/S}$ is surjective, hence so is $\mathscr{F} \to \omega_{K/S}$. Therefore N_i is also a quotient of $\mathscr{F}_2$. This defines a closed subgroup scheme G of K. Clearly $N_i \simeq \omega_{G/S}$, so G is flat, again by Remark 2.7. Now let $A_{i-1} = A_i^{(p)}/G$.

Even when σ^* is not faithful, we can still recover a supersingular abelian scheme with flag type level structure from its crystal as above. Hence $\mathbf{D}_S$ is fully faithful. Q.E.D.

So far we have not taken account of the integer invariants t_i $(1 \leqq i < g)$ of Sect. 1. Let $t = (t_1, \ldots, t_{g-1})$. We can modify the definition in Sect. 2 by changing "index s" to "index (s, t)", and adding

iii) $\ker(F_{A_i^{(p)}/S}) \cap \ker(V_{A_i/S})$ is flat of α-rank t_i.

We can also define a functor

$$T_{s,t} : ((\text{reduced } k\text{-schemes})) \to ((\text{sets})),$$

$$S \mapsto \{\text{flag type level structures of index } (s, t) \text{ over } S\}/\text{isomorphisms}.$$

Then one can prove that $T_{s,t}$ is represented by a locally closed subscheme of S_s, just as in Theorem 2.9.

Again use induction. Suppose we have got a locally closed subscheme $S_{i,t}$ of S_i. Then it is easy to see that t_i is an upper-semicontinuous function on the set of closed points. Thus there is a greatest locally closed subset U of $S_{i,t}$ with reduced induced structure such that $K \times_{S_i} U$ is a flat α-group of α-rank t_i. Since giving a flat subgroup scheme of a flat α-group is equivalent to giving a flat quotient of its α-sheaf, we see that $S_{i-1,t}$ is a relative Grassmannian over U. In particular, if t is the smallest possible (i.e., such that $S_{s,t}$ is not empty) in the lexicographic order, then $S_{s,t}$ is a smooth open subset of S_s. By Sect. 1, one can easily see that S_s is not empty for any s. Hence we get

Proposition 3.3. *For any index s, S_s contains a non-empty smooth open subset.*

4. Supersingular Abelian Schemes

Suppose we have a supersingular abelian scheme $f: A \to S$ of dimension g. Then it is natural to ask: does it admit a flag type level structure? The answer is negative in general. (See the counter examples below.) However, it is still possible to get a flag type level structure after making some changes in S. We need the following lemma.

Lemma 4.1 (Ogus). *If S is normal, then there is an étale morphism $e: S' \to S$ such that there is an isogeny $E^g \times S' \to A \times_S S'$ (over S').*

*Proof.** Let $l \neq p$ be a prime, and consider the local systems $E_n \stackrel{\text{def.}}{=} R^1 F_* \mathbb{Z}/l^n\mathbb{Z}$ on S. According to a theorem of Grothendieck [9, Proposition 4.4.], A/S is isogeneous to a constant family (i.e., a product of an abelian variety over k with S) if and only if

* Unpublished proof of A. Ogus

E_n is constant for all $n>0$. Thus, it will suffice to prove that this condition can be achieved after some finite étale covering $S'\to S$.

The key case is when k is an algebraic closure of $\mathbb{F}_p$. In this case, we can find a field $\mathbb{F}_q\subset k$ and a descent $f_0: A_0\to S_0$ of f to varieties of finite type over $\mathbb{F}_q$. Let us choose m large enough so that $l^m\geqq 4g$ and a finite étale covering $S'_0\to S_0$ on which E_m is constant. Replace S_0 by S'_0 to simplify the notation. Replacing q by a power, we can assume that $q=p^d$, with d even, and that $p^{d/2}\equiv 1 \pmod{l^m}$.

Let σ be a geometric point of S_0 and consider the exact sequence of groups

$$1\to\pi_1(S,\sigma)\to\pi_1(S_0,\sigma)\to\mathrm{Gal}(k/\mathbb{F}_q)\to 1\,.$$

Let $E=\varprojlim E_m$. Then $E(\sigma)$ is a $\mathbb{Z}_l[\pi_1(S_0,\sigma)]$-module, and is l-torsion free. It suffices to prove that the action ϱ of $\pi_1(S,\sigma)$ on $E(\sigma)\otimes\mathbb{Q}_l$ is trivial.

For each closed point s of S_0 and each Frobenius element F_s of $\pi_1(S_0,\sigma)$, the trace $A(s)$ of $\varrho(F_s)$ is the same as the trace of $F_{A_{\bar{s}}}^{\deg(s)}$ on $H^1(A_{\bar{s}},\mathbb{Q}_l)$. As $A_{\bar{s}}$ is a supersingular abelian variety, we know

i) $A(s)$ is an integer;

ii) $|A(s)|\leqq 2gp^{d(s)/2}$ (Riemann hypothesis);

iii) $p^{d(s)/2}$ divides $A(s)$ [by supersingularity, noting that $d|d(s)$, so $d(s)$ is even];

iv) $A(s)\equiv 2g \pmod{l^m}$ (since E_m is constant).

Let $B(s)=A(s)\cdot p^{-d(s)/2}$. Then $B(s)\in\mathbb{Z}$, $|B(s)|\leqq 2g$, and $B(s)\equiv 2g \pmod{l^m}$ since $p^{d(s)/2}$ is a power of $q^{1/2}$ and $q^{1/2}\equiv 1 \pmod{l^m}$. Then since $l^m\geqq 4g$, $B(s)=2g$ and $A(s)=2gp^{d(s)/2}$ for every s.

Now consider the representation ϱ' of $\pi_1(S_0,\sigma)$ attached to a constant family of supersingular abelian varieties. The argument above shows that ϱ and ϱ' have the same trace on Frobenius elements, and hence on all elements, by Chebataroff. It follows that ϱ' and ϱ have isomorphic semi-simplifications. In particular, the semi-simplification ϱ_{ss} of ϱ is trivial when restricted to $\pi_1(S,\sigma)$. Then $\varrho|_{\pi_1(S,\sigma)}$ admits a decomposition series whose successive quotients are constant. By a (deep!) theorem of Deligne [3, p. 383, Théorème 3.4.1], $\varrho|_{\pi_1(S,\sigma)}$ is in fact semi-simple, hence trivial. Q.E.D.

Theorem 4.2. *If S is integral, then there are a blowing up $b: S_b\to S$, a normalization $n: S_n\to S_b$, and an étale covering $e: S_e\to S_n$ such that $A^{(p^{2g-3})}\times_S S_e$ admits a flag type level structure.*

Proof. Let $K_2=\ker(V_{A/S})$, $K=\ker(V_{K_2^\vee/S})$. Letting m be the rank of

$$\mathrm{Im}(V^*_{K_2^\vee/S}:\omega_{K_2^\vee/S}\to\omega_{K_2^{\vee(p)}/S}),$$

there is an open dense subset $U\subset S$ on which $\mathrm{coker}(V^*_{K_2^\vee/S})$ is flat of rank $g-m$. By Lemma 2.8, the functor

$$R':\mathscr{C}_S\to((\text{sets})),$$

$$(X\to S)\mapsto\{\text{flat closed subgroup schemes of } K\times_S X \text{ over } X \text{ of } \alpha\text{-rank } g-m\}$$

is represented by a relatively projective morphism $b': S'\to S$ and a flat closed subgroup scheme $G\hookrightarrow K\times_S S'$ over S'. But $b'|_{b'^{-1}(U)}$ should be an isomorphism. Since $G^\vee$ is an α-group and is a quotient of $K_2^{(p)}$, it gives an isogeny $A_1\to A^{(p)}\times_S S'$ with kernel $G^\vee$. Replace S' by the irreducible component (with the reduced induced structure) which maps surjectively to S. Then b' becomes a blowing up.

Take S' instead of S and A_1 instead of A, then repeat this procedure, and so on. We will finally obtain a blowing up $b: S_b \to S$ and an isogeny $i: A' \to A^{(p^{g-1})} \times_S S_b$ over S_b such that:

i) $\ker(i) \subsetneqq \ker(F^{g-1})$;

ii) $\ker(F_{A'^{(p)}/S_b}) \cap \ker(V_{A'/S_b})$ is flat of α-rank g on an open dense subset of S_b;

iii) There are isogenies $A' = A'_{g-1} \to A'_{g-2} \to \ldots \to A'_0 \simeq A^{(p^{g-1})} \times_S S_b$, such that every kernel is an α-group.

Condition ii) implies that $\ker(F_{A'^{(p)}/S_b}) = \ker(V_{A'/S_b})$, and hence A' is superspecial (i.e., all of the closed fibers are superspecial).

Now pull back A' via the normalization $n: S_n \to S_b$. Then using Lemma 4.1, we get an étale covering $e: S_e \to S_n$ and an isogeny over S_e:

$$h: E^g \times S_e \to A' \times_{S_b} S_e .$$

Let $H = \ker(h)$. First we have a decomposition $H \simeq H_1 \times_S H_2$, where H_1 is étale and H_2 is local. Since H_1 is a closed subgroup scheme of $\ker(\cdot\, t_{E^g \times S_e}) \simeq \ker(\cdot\, t_{E^g}) \times S_e$ for some integer $t (p \nmid t)$, we see that $H_1 \simeq H_0 \times S_e$, where H_0 is étale over k, and $H_1 \hookrightarrow E^g \times S_e$ is induced by a morphism $H_0 \to E^g$. Thus we may assume $H_1 = 0$ by taking E^g/H_0 (also superspecial) instead of E^g.

Suppose $H' = \ker(F_{H/S})$ has α-rank r on an open dense subset U. As above, we have a blowing up $b': S_{eb} \to S_e$ and a flat closed subgroup scheme $H_3 \subsetneqq H' \times_{S_e} S_{eb}$ of α-rank r. In fact, $H' \hookrightarrow \ker(F_{E^g}) \times S_e$ induces a closed immersion $\eta: S_{eb} \to G_{g,r} \times S_e$, where $G_{g,r}$ is the Grassmannian. If $x \in U$, then E^g/H'_x is still superspecial by Sect. 1. But there are only a finite number of subgroup schemes $\bar{H} \subset \ker(F_{E^g})$ of α-rank r such that $E^g/\bar{H}$ is superspecial. Thus $\mathrm{pr}_1 \circ \eta(b'^{-1}(U))$ is just one point. Therefore b' is an isomorphism and $H' \simeq H'_0 \times S_e$ for some $H'_0 \subset E^g$. Replacing E^g by E^g/H'_0 (also superspecial) and repeating this argument, h will finally become an isomorphism. Q.E.D.

Corollary 4.3. *If $A \to S$ has superspecial closed fibers and there is an isogeny $E^g \times S \to A$, then $A \simeq E^g \times S$.*

Remark 4.4. Letting $s_0 = (1, 2, \ldots, g)$, we note that S_{s_0} has special importance, since every supersingular abelian variety has a flag type level structure of index s_0 [18, Theorem 2.2]. By slightly modifying the proof of Theorem 4.2, one shows that under the conditions of Theorem 4.2, there are a blowing up $b: S_b \to S$, a normalization $n: S_n \to S_b$, and an étale covering $e: S_e \to S_n$ such that $A^{(p^{2g-3})} \times_S S_e$ has a flag type level structure of index s_0.

Example 1 (cf. [21, Remark 10]). This example shows the necessity of the étale covering in Theorem 4.2. Let $S = \mathrm{Spec}(k[x, y, (x-y)^{-1}]$, $G = \mathbb{Z}/2\mathbb{Z} = \{e, \sigma\}$. Define a G-action on S by letting e act as id_S, and σ act by switching x and y. Also define a G-action on $E^2 = E \times E$ by letting e act as id_{E^2}, and σ act by switching factors. Then we get a free action of G on $A = E^2 \times S$ compatible with the (free) action of G on S. By Lemma 2.3, we get quotients $A' = A/G$ and $S' = S/G$. It is easy to check that the abelian scheme structure of A over S induces an abelian scheme structure of A' over S'. Also S' is smooth. We claim that $A' \to S'$ does not admit a level structure, not even over an open dense subset of S'. Indeed, suppose we have an isogeny $f: E^2 \times S' \to A'$ over S'. Then by the above corollary, we may assume that f is an isomorphism. The projection $\varrho: S \to S'$ induces a morphism $g: A \to E^2 \times S$ over S such that $(\mathrm{id} \times \varrho) \circ g$ is the projection $A \to A'$. Here g must be an isomorphism, judging by its degree. Let x be a closed point of S. Then $g - g_x \times \mathrm{id}_S$ maps A_x to one point, hence by rigidity [14, p. 43], it factors through S, i.e., $g - g_x \times \mathrm{id}_S = 0$, or $g = g_x \times \mathrm{id}_S$. Now the projection $A \to A'$ is equal to both $g_x \times \varrho$ and $(g_x \times \varrho) \circ \sigma = (g_x \circ \sigma) \times \varrho$, so $g_x = g_x \circ \sigma$, a contradiction.

Example 2. This example shows the necessity of the blowing up in Theorem 4.2. Let $g=3$, and $f: E^2 \times \hat{S}_3 - \hat{A}$ over $\hat{S}_3$ represent $\hat{T}_3$ (see Sect. 2). Consider the morphism $\varrho: S_{s_0} \to \hat{S}_3$ induced by the level structure $E^2 \times S_{s_0} \to A_0$ over S_{s_0}. We claim that ϱ is surjective. Indeed, for the fiber over a closed point $x \in \hat{S}_3$, f induces a flag type level structure $\hat{A}_{xi} = E^2/((\ker F^{2-i}) \cap (\ker f_x))$ of index either $(1,2,3)$ or $(0,0,3)$. In the later case, $\ker(f_x) = \ker(F_{E^2})$, hence for any $G \subset \ker(F_{E^2})$ of α-rank 2, the flag type level structure $E^2 \to E^2/G \to E^2/\ker(F_{E^2})$ is compatible with the level structure f at x. Therefore $\varrho^{-1}(x)$ is of dimension $2(\simeq \mathbf{P}^2)$.

Now let us calculate S_{s_0}. We use the notation used in the proof of Theorem 2.9. First, $S_1 \simeq \mathbf{P}^2 \simeq \mathrm{Proj}\, k[x_0, x_1, x_2]$. Let $G = \ker(B_2^{(p)} \times S_1 \to B_1)$, $K = \ker(F_{B_1^{(p)}/S_1}) \cap \ker(V_{B_1/S_1})$. Then $G' = (\ker(F_{B_2^{(p)} \times S_1/S_1})/G)^{(p)}$ is a flat subgroup scheme of K of α-rank 1, and

$$\omega_{G'/S_1} \simeq (\ker(\omega_{B_2^{(p)} \times S_1/S_1} \to \omega_{G/S_1}))^{(p)} \simeq O_{S_1}(-1)^{(p)} \simeq O_{S_1}(-p).$$

Let $G'' = K/G'$. Clearly both $F_{B_1^{(p)}/S_1}$ and V_{B_1/S_1} factor through

$$j: B_1^{(p)} \to B_1^{(p)}/G' (\simeq B_2 \times S_1).$$

Let $F_{B_1^{(p)}/S_1} = F' \circ j$, $V_{B_1/S_1} = V' \circ j$. Then $G'' \simeq \ker(F') \cap \ker(V')$. Hence

$$\omega_{G''/S_1} \simeq \omega_{B_2 \times S_1/S_1}/(\mathrm{Im}(F'^*) \cap (\mathrm{Im}(V'^*)),$$

which is isomorphic to the cokernel of

$$\begin{aligned} &O_{S_1}(-1) \oplus O_{S_1}(-p^2) \to O_{S_1}^{\oplus 3} \\ &\qquad (u,v) \mapsto (x_0 u + x_0^{p^2} v, x_1 u + x_1^{p^2} v, x_2 u + x_2^{p^2} v). \end{aligned}$$

In particular, ω_{G''/S_1} is torsion free. Hence ω_{K/S_1} is also torsion free, by the exact sequence

$$0 \to \omega_{G''/S_1} \to \omega_{K/S_1} \to \omega_{G'/S_1} \to 0.$$

Since $S_0 \simeq \mathrm{Proj}_{O_{S_1}}(\omega_{K/S_1})$, we see that S_0 is irreducible.

Let S_n be the normalization of $\hat{S}_3$. We claim that for any étale covering $S_e \to S_n$, $\hat{A} \times_{\hat{S}_3} S_e$ does not admit a flag type level structure. Indeed, if it had, then we would get a surjective morphism $S_e \to S_0$ inducing $\hat{A} \times_{\hat{S}_3} S_e \simeq A_0 \times_{S_0} S_e$ over S_e. Since $S_e \to \hat{S}_3$ is finite, $\hat{A} \times_{\hat{S}_3} S_e \to S_e$ would have only a finite number of superspecial closed fibers. But we have seen that $A_0 \times_{S_0} S_e \to S_e$ has an infinite number of superspecial closed fibers, a contradiction.

Therefore the étale covering and the blowing up in Theorem 4.2 are both necessary.

5. Coarse Moduli Spaces

Let $A \to S$ be a supersingular abelian scheme. Then for any closed point $x \in S$, $G_x = \ker(F) \cap \ker(V)$ of A_x corresponds to $\bar{M}^0$ in Sect. 1, where $M = H^1_{\mathrm{cris}}(A_x/W)$. Hence the α-rank of G_x is equal to $s_1'(M)$. Therefore s_1' is an upper-semicontinuous function of S_{cl}. Similarly s_2' is an upper-semicontinuous function on the open subset of S_{cl} of all closed points whose fibers have minimal s_1', and so on.

Let M be a supersingular Dieudonné crystal of genus g over W. Then it has a canonical flag type level structure $M_i = F^{g-1-i}S(M) + M$ $(0 \leqq i < g)$ whose index s is clearly the smallest possible in the lexicographic order. This flag type level structure is called *rigid*. Similarly we can define a rigid flag type level structure for a supersingular abelian variety.

Now consider the moduli space S_s. Any $\varrho \in \mathrm{Aut}(E^g)$ gives another flag type level structure of index s over S_s, hence gives an automorphism of S_s. Therefore $\mathrm{Aut}(E^g)$ acts on S_s. The action is finite since for any $\varrho, \varrho' \in \mathrm{Aut}(E^g)$ such that $\varrho - \varrho' \in p^g \operatorname{End}(E^g)$, $\varrho - \varrho' = 0$ on the closed subgroup schemes $\ker(A_i^{(p)} \to A_{i-1})$ $(1 \leqq i < g)$, hence ϱ and ϱ' act in the same way on S_s. Therefore the action has a quotient $\tilde{S}_s$ by Lemma 2.3. Similarly, using the method of Sect. 3, we see that $\mathrm{Aut}(\dot{M}, \mathrm{tr})$ acts on S_s and the quotient space is just $\tilde{S}_s$. But $\mathrm{Aut}(\dot{M})$ also acts on S_s and we get a quotient space $\hat{S}_s$ (which was described in [18, p. 606]). We use similar notation for an index (s,t). The actions may not be transitive on a set of all closed points with isomorphic A_0 [resp.., (M_0, tr) or M_0] fibers, but they are transitive if one of the fibers is rigid, because any isomorphism $M_0 \tilde{\to} M_0'$ is induced by an isomorphism $S(M) \tilde{\to} S(M_0')$. Clearly the closed points with rigid fibers form an open subset, which we denote by an upper r. By Theorem 4.2, it is easy to show that $(\tilde{S}^r_{s,t})^{(p^{2g-3})}$ can be viewed as the "coarse moduli space" of supersingular abelian varieties of index (s,t). In other words, we have

Proposition 5.1. *If S is a normal scheme of finite type over k, and $\pi: A \to S$ is a supersingular abelian scheme whose closed fibers are all of the same index (s,t), then there is a unique morphism $f: S \to (\tilde{S}^r_{s,t})^{(p^{2g-3})}$ over k such that for any closed point x of S, $\pi^{-1}(x)$ is isomorphic to the supersingular abelian variety corresponding to $f(x)$.*

Proof. We may assume S is integral. By Theorem 4.2, there is an étale covering $e: S_e \to S$ such that $A^{(p^{2g-3})} \times_S S_e$ has a flag type level structure of index (s,t). (We don't need blowing up since the group scheme K in the proof of Theorem 4.2 is flat.) This gives a unique morphism $S_e \to S^r_{s,t}$, hence a morphism $f': S_e \to \tilde{S}^r_{s,t}$. Clearly f' maps every closed fiber of e to one point. Take another étale covering $S_e' \to S$ such that $S_e' \times_S S_e$ is a disjoint union of copies of S_e'. Then $f' \circ \mathrm{pr}_2: S_e' \times_S S_e \to \tilde{S}^r_{s,t}$ factors through S_e' set-theoretically, hence scheme-theoretically [10, Ex. II.4.2]. So f' factors through S by the following lemma (whose proof is easy and is left to the reader).

Lemma 5.2. *Suppose $X \to S$, $Y \to S$ are both flat surjective morphisms of finite type of Noetherian schemes. Let $Z = X \times_S Y$. Then S is the scheme-theoretic push-out of* $\mathrm{pr}_1: Z \to X$ *and* $\mathrm{pr}_2: Z \to Y$.

The only thing remaining to check now is the coincidence of $A_x = \pi^{-1}(x)$ with the supersingular abelian variety A_x' corresponding to $f(x)$. The above shows that

$A_x^{(p^{2g-3})} \simeq A'_x$. To get rid of (p^{2g-3}) one can consider $S^{(p^{3-2g})}$ instead of S. Then one gets $S^{(p^{3-2g})} \to \tilde{S}^r_{s,t}$, or $S \to (\tilde{S}^r_{s,t})^{(p^{2g-3})}$. Q.E.D.

It is natural to ask: What are the degrees (of the generic fibers) of the finite morphisms $S^r_{s,t} \to \tilde{S}^r_{s,t}$, $\tilde{S}^r_{s,t} \to \hat{S}^r_{s,t}$, etc.? As an example, we calculate the special case $s=s_0$ here. In this case t can only be $(1,2,\ldots,g-1)$. We need the following proposition.

Proposition 5.3. *Let $M=Av$, where v satisfies Lemma 1.5 and is general enough (i.e., $\bar{v}$ is contained in some non-empty Zariski open subset of $k^{\oplus g}$). Then an automorphism $h \in \mathrm{Aut}_A(\dot{M})$ stabilizes M if and only if there is $\lambda \in W(\mathbb{F}_{p^2})^*$ $(=W(\mathbb{F}_{p^2}) - pW(\mathbb{F}_{p^2}))$, such that $(h-\lambda I)\dot{M} \subseteq F^{g-1}\dot{M}$, and $\lambda - \lambda^\sigma \in p^{\left[\frac{g-1}{2}\right]} W(\mathbb{F}_{p^2})$.*

The normal subgroup of all automorphisms satisfying the conditions in Proposition 5.3 will be denoted by H.

Lemma 5.4. *Let $T=(a_{ij})$, $T'=(b_{ij})$ be two $g \times g$ matrices over k, let n be an integer, $0 \leqq n \leqq 2g-3$, and $y_1,\ldots,y_g$, $y'_1,\ldots,y'_g$ be indeterminates. Let*

$$R(T,T',n,y_1,\ldots,y_g,y'_1,\ldots,y'_g) = \begin{vmatrix} \sum_{j=1}^{g} a_{j1}y_j^{p^n} + b_{j1}y_j'^{p^n} & \cdots & \sum_{j=1}^{g} a_{jg}y_j^{p^n} + b_{jg}y_j'^{p^n} \\ y_1 & \cdots & y_g \\ y_1^{p^2} & \cdots & y_g^{p^2} \\ \vdots & \ddots & \vdots \\ y_1^{p^{2g-4}} & \cdots & y_g^{p^{2g-4}} \end{vmatrix}$$

Then $R(T,T',n,y_1,\ldots,y_g,y'_1,\ldots,y'_g)=0$ if and only if $T=a_{11}I$ and $T'=0$ when n is even, or $T=T'=0$ when n is odd.

Proof. Look at the Laplacian expansion of $R(T,T',n,y_1,\ldots,y_g,y'_1,\ldots,y'_g)$. If $b_{ij} \neq 0$, then there is only one term of the form $cy_i'^{p^n} \prod_{l<j} y_l^{p^{2i-2}} \prod_{l>j} y_l^{p^{2i-4}}$, $c \neq 0$, a contradiction. If $a_{ij} \neq 0$, $i \neq j$, then there is only one term of the form

$$cy_i^{p^{n+1}} \prod_{\substack{l \neq i,j \\ s_l < s_m \text{ if } l<m}} y_l^{p^{s_l}}, \quad c \neq 0,$$

also impossible. If n is odd and $a_{ii} \neq 0$, then there is only one term of the form

$$cy_i^{p^n} \prod_{i<j} y_l^{p^{2i-2}} \prod_{l>j} y_l^{p^{2i-4}}, \quad c \neq 0,$$

still impossible. Finally, if n is even and $i \neq j$, then there are exactly two terms of the form

$$cy_i^{p^n} y_j^{p^n} \prod_{\substack{l \neq i,j \\ s_l < s_m \text{ if } l<m}} y_l^{p^{s_l}},$$

namely $c = \pm a_{ii}$ and $c = \mp a_{jj}$. Hence $a_{ii} = a_{jj}$. Q.E.D.

Proof of Proposition 5.3. The proof of sufficiency is easy. We now prove necessity. Fix a basis $x_1,\ldots,x_g$ of $\dot{M}$. For every n, $0 \leqq n < g$, we want to obtain $\lambda_n \in W(\mathbb{F}_{p^2})^*$ such that $h \equiv \lambda_n I \pmod{F^n \dot{M}}$ and $\lambda_n - \lambda_n^\sigma \equiv 0 \pmod{p^{[n/2]}}$. We use induction on n.

Take $\lambda_0=1$. Suppose we have got λ_n. Then there are two possible cases.

i) n is even. In this case we may assume that $\lambda_n \in W(\mathbb{F}_p)^*$. Let $h'=h-\lambda_n I = p^{n/2}(a_{ij}+b_{ij}F)$. Then h stabilizing M implies that $h'v \in p^{n/2}\dot{M} \cap M = (F^n v, F^{n-1}Vv, \ldots, V^n v)$. Since

$$h'v = p^{n/2} \sum_{i,j=1}^{g} (a_i + b_i F)(a_{ij}+b_{ij}F)x_j$$

and

$$F^{n-l}V^l v = p^{n/2} \sum_{i=1}^{g} (a_i^{\sigma^{n-2l}} + b_i^{\sigma^{n-2l}} F)x_i,$$

modulo $F^{n+1}\dot{M}$ we get (still denoting by $\bar{a}$ the image of $a \in W$ in $W/pW \simeq k$)

$$0=\begin{vmatrix} \sum_{i=1}^{g} \bar{a}_i\bar{a}_{i1} & \cdots & \sum_{i=1}^{g} \bar{a}_i\bar{a}_{ig} \\ \bar{a}_1^{\sigma^{-n}} & \cdots & \bar{a}_g^{\sigma^{-n}} \\ \bar{a}_1^{\sigma^{-n+2}} & \cdots & \bar{a}_g^{\sigma^{-n+2}} \\ \vdots & \ddots & \vdots \\ \bar{a}_1^{\sigma^{-n+2g-4}} & \cdots & \bar{a}_g^{\sigma^{-n+2g-4}} \end{vmatrix} = R((\bar{a}_{ij}), (0), n, \bar{a}_1, \ldots, \bar{a}_g, \bar{b}_1, \ldots, \bar{b}_g)^{\sigma^{-n}}.$$

Hence by Lemma 5.4, $(\bar{a}_{ij}) = \bar{a}_{11} I$ when v is general enough. Note that $\bar{a}_{ij} \in \mathbb{F}_{p^2}$ and there are only a finite number of $g \times g$ matrix over $\mathbb{F}_{p^2}$. Therefore we can take $\lambda_{n+1} \in W(\mathbb{F}_{p^2})$ such that $h \equiv \lambda_{n+1} I \pmod{F^{n+1}\dot{M}}$ and $\lambda_{n+1} \equiv \lambda_n \pmod{p^{n/2}}$.

ii) n is odd. Then we can write

$$h \equiv \lambda_n I + p^{\frac{n-1}{2}}(b_{ij})F \left(\operatorname{mod} p^{\frac{n+1}{2}}\right),$$

$$\lambda_n^\sigma - \lambda_n \equiv p^{\frac{n-1}{2}} \mu \left(\operatorname{mod} p^{\frac{n+1}{2}}\right), \quad \mu \in W(\mathbb{F}_{p^2}).$$

Let

$$v' = hv - \lambda_n v \equiv p^{\frac{n-1}{2}} \sum_{j=1}^{g} \left(\sum_{i=1}^{g} a_i b_{ij} + \mu b_j \right) F x_j \left(\operatorname{mod} p^{\frac{n+1}{2}} \dot{M}\right).$$

Since $v' \in F^n \dot{M} \cap M$ we have

$$0=\begin{vmatrix} \sum_{i=1}^{g} \bar{a}_i\bar{b}_{i1} + \bar{\mu}\bar{b}_1 & \cdots & \sum_{i=1}^{g} \bar{a}_i\bar{b}_{ig} + \bar{\mu}\bar{b}_g \\ \bar{a}_1^{\sigma^{-n}} & \cdots & \bar{a}_g^{\sigma^{-n}} \\ \vdots & \ddots & \vdots \\ \bar{a}_1^{\sigma^{-n+2g-4}} & \cdots & \bar{a}_g^{\sigma^{-n+2g-4}} \end{vmatrix} = R((\bar{b}_{ij}^\sigma), \bar{\mu}^\sigma I, n, \bar{a}_1, \ldots, \bar{a}_g, \bar{b}_1, \ldots, \bar{b}_g)^{\sigma^{-n}}.$$

Since n is odd and v is general enough, we have $\bar{\mu}=0$, $\bar{b}_{ij}^\sigma = 0$. Therefore

$$h \equiv \lambda_n I \left(\operatorname{mod} p^{\frac{n+1}{2}}\right). \quad \text{Q.E.D.}$$

Now $S^r_{s_0}$ is smooth, and there is an open dense subset of S_{s_0} on which the action of $\mathrm{Aut}_A(\dot{M})/H$ is free and transitive for every set of closed points with isomorphic M_0 fibers. Therefore

$$\begin{aligned}&\text{the degree of } S^r_{s_0} \to \hat{S}^r_{s_0}\\ &= \#(\mathrm{Aut}_A(\dot{M})/H)\\ &= \begin{cases} p^{(g-2)(2g^2+g)+2g-3[g/2]}(p+1)(p^4-1)(p^6-1)\dots(p^{2g}-1) & \text{if} \quad g>2 \\ p^2(p^4-1) & \text{if} \quad g=2 \end{cases}.\end{aligned}$$

Now consider the degree of $\tilde{S}^r_{s_0} \to \hat{S}^r_{s_0}$. We have seen that $\mathrm{Aut}_A(\dot{M}, \mathrm{tr})$ is the normal subgroup of $\mathrm{Aut}_A(\dot{M})$ consisting of all matrices of determinant 1. Let $H' = H \cap \mathrm{Aut}(\dot{M}, \mathrm{tr})$. Then the degree is equal to the order of $(\mathrm{Aut}(\dot{M})/H)/(\mathrm{Aut}(\dot{M}, \mathrm{tr})/H')$. When $g \neq 2$, letting f be the $2g^{\text{th}}$ power map of $(W(\mathbb{F}_p)/p^{[g/2]}W(\mathbb{F}_p))^* : f(a) = a^{2g}$, this is equal to

$$\begin{aligned}\frac{\#((W(\mathbb{F}_p)/p^{[g/2]}W(\mathbb{F}_v))^*)}{\#((W(\mathbb{F}_p)/p^{[g/2]}W(\mathbb{F}_p))^{2g*})} &= \#(\ker(f))\\ &= \#\{\text{solutions of } x^{2g}=1 \text{ in } \mathbb{Z}/p^{[g/2]}\mathbb{Z}\}\,.\end{aligned}$$

If $g=2$, we also get that the degree is equal to the number of solutions of $x^2=1$ in $\mathbb{F}_p$. Summarizing, we obtain

$$\text{degree of } \tilde{S}^r_{s_0} \to \hat{S}^r_{s_0} = \begin{cases} \mathrm{G.C.D.}(2g, (p-1)p^{[g/2]-1}) & \text{if} \quad p \neq 2,\ g>2 \\ 2\mathrm{G.C.D.}(2g, p^{[g/2]-2}) & \text{if} \quad p=2,\ g>3 \\ 2 & \text{if} \quad p \neq 2,\ g=2 \\ 1 & \text{if} \quad p=2,\ g \leqq 3. \end{cases}$$

References

1. Berthelot, P., Breen, L., Messing, W.: Théorie de Dieudonné cristalline II. Lecture Notes in Mathematics, Vol. 930. Berlin Heidelberg New York: Springer 1982
2. Berthelot, P., Ogus, A.: Notes on crystalline cohomology. Mathematical Notes. Princeton: Princeton University Press 1978
3. Deligne, P.: La conjecture de Weil II. Publ. Math. IHES **52**, 137–252 (1980)
4. Demazure, M., Gabriel, P.: Groupes algebraiques. Tome I. Amsterdam: North-Holland 1970
5. Demazure, M., Grothendieck, A., et al.: SGA 3: Schemas en Groups I, II, III, Lecture Notes in Mathematics 151, 152, 153. Berlin Heidelberg New York: Springer 1970
6. Demazure, M.: Lectures on p-Divisible groups. (Lecture Notes in Mathematics, Vol. 302). Berlin Heidelberg New York: Springer 1972
7. Dieudonné, J.: La geometrie des groupes classiques. Berlin Heidelberg New York: Springer 1955
8. Grothendieck, A.: EGA III: Étude cohomologique des faisceaux coherents, Publ. Math. IHES **11** (1961), **17** (1963)
9. Grothendieck, A.: Un théorème sur les homomorphisms de schémas abelians. Invent. Math. **2**, 59–78 (1966)
10. Hartshorne, R.: Algebraic geometry. GTM 52. Berlin Heidelberg New York: Springer 1977
11. Katz, N.: Slope filtrations of F-crystals. Astérisque **63**, 113–164 (1979)
12. Manin, Yu.I.: The theory of commutative formal groups over fields of finite characteristic. Russ. Math. Surv. **18**, 1–80 (1963)

13. Matsumura, H.: Commutative algebra. New York: Benjamin 1970
14. Mumford, D.: Abelian varieties. Oxford: Oxford University Press 1970
15. Mumford, D.: Lectures on curves on an algebraic surface, Ann. Math. Stud. **59**, 1–212 (1966)
16. Nygaard, N., Ogus, A.: Tate's conjectures for K3 surfaces of finite height, Ann. Math. **122**, 461–507 (1985)
17. Oda, T.: The first De Rham cohomology group and Dieudonné modules. Ann. Sci. Ec. Norm. Sup. **2**, 63–135 (1969)
18. Oda, T., Oort, F.: Supersingular abelian varieties, Intl. Symp. on Algebraic Geometry, 595–621. Kyoto (1977)
19. Ogus, A.: Supersingular K3 crystals, Astérisque **64**, 3–86 (1979)
20. Oort, F.: Commutative group schemes. Lecture Notes in Mathematics, Vol. 15. Berlin Heidelberg New York: Springer 1966
21. Oort, F.: Which abelian surfaces are product of elliptic curves? Math. Ann. **214**, 35–47 (1975)
22. Oort, F.: Subvarieties of moduli spaces. Invent. Math. **24**, 95–119 (1974)
23. Shioda, T.: Supersingular K3 surfaces, algebraic geometry. Copenhagen 1978. (Lecture Notes in Mathematics, Vol. 732). Berlin Heidelberg New York: Springer 1978
24. Tate, J., Oort, F.: Group schemes of prime order. Ann. Sci. Ec. Norm. Sup., 4° ser. **3**, 1–21 (1970)

Received November 12, 1986; in revised form June 25, 1987 and July 1, 1988

Math. Ann. 283, 352 (1989)

Mathematische Annalen

Erratum

Compact Differentiable 4-Folds with Quaternionic Structures

Ma. Kato

Department of Mathematics, Sophia University, Kioi-cho, Chiyoda-ku, Tokyo 102, Japan

Math. Ann. **248**, 79–96 (1980)

The statement of Proposition 8 in "compact differentiable 4-folds with quaternionic structures", was incomplete and a subcase was missing. We add the following subcase to Proposition 8.

(v) $\{B_2, cv_{B_2}\} = (K \times \{c^3 I\}) \cup (cv_{B_2})(K \times \{c^3 I\}) \cup (cv_{B_2})^2 (K \times \{c^3 I\})$,

where $c \in \mathbf{R}$, $0 < c < 1$, and $v_{B_2} = \frac{1}{\sqrt{2}} \begin{pmatrix} \varrho_4^3 & \varrho_4^3 \\ \varrho_4 & -\varrho_4 \end{pmatrix}$.

Therefore, p. 95↑8, "...are involutions of S^3/H" should read as "...are of order 2 or 3 as an element of the diffeomorphism group of S^3/H". The error was in (41). In case $K = B_2$, the possibility of $(\det g)^{-1/2} g = v_{B_2}$ was missing. This error had its origin in that of Lemma 4 in my previous paper "Topology of Hopf surfaces", J. Math. Soc. Japan 27, 222–238 (1975). The correction of that paper will appear in J. Math. Soc. Japan.

Acknowledgement. I am very much grateful to M. Ue who kindly informed me about the mistake.

Received October 17, 1988

Math. Ann. 283, 353–365 (1989)

Mathematische Annalen

Scalar Curvatures on S^n

Wenxiong Chen*

Institute of Mathematics, Academia Sinica, Beijing, Peoples Republic of China and Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104-6395, USA

0. Introduction

Let (M^n, g) be a compact Riemannian manifold of dimension $n \geqq 3$, and $R(x)$ a smooth function on M^n. One would like to ask if $R(x)$ can be the scalar curvature of some metric $\tilde{g}$ that is pointwise conformal to the original metric g. This is an interesting problem in geometry. If we let $\tilde{g} = u^{4/(n-2)}g$, then it is equivalent to finding a positive solution of the differential equation

$$\begin{cases} -\Delta_g u + \dfrac{n-2}{4(n-1)} R_g u = \dfrac{n-2}{4(n-1)} R u^\tau, \\ u > 0 \end{cases} \tag{0.1}$$

where $\tau = (n+2)/(n-2)$ and R_g is the scalar curvature of the metric g.

In case R is a constant, this is the well-known Yamabe Problem and the affirmative answer was completed by Schoen [6]. Then the attention is turned to non-constant $R(x)$. Recently, Escobar and Schoen [4] obtained some good results in this respect.

It is known that the problem becomes difficult when $M^n = S^n$ with the standard metric g_0. In this situation, Kazdan and Warner [5] found an obstruction to the solvability of the corresponding equation:

$$\begin{cases} -\Delta_{g_0} u + v_n u = K(x) u^\tau \\ u > 0, \end{cases} \tag{*}$$

where $v_n = n(n-2)/4$ and $K(x) = [(n-2)/4(n-1)]R(x)$. More obstructions were found recently by Bourguignon and Ezin [2]. These show that for quite a few functions $K(x)$ Problem (*) has no solution. Then for which K can one solve (*)? This has been an interesting problem for years.

* Research supported in part by NSF Grant DMS 85-03302

In paper [4], under the assumptions:

K_0) K is positive somewhere [known to be a necessary condition for the sovability of (*)];

K_1) (Symmetry condition) $K(hx)=K(x)$ for any $h\in\Gamma$, $x\in S^n$; where Γ is some finite group of isometries on S^n;

K_2) (Flatness condition) there is $p\in S^n$, such that

$$K(p)=\max_{S^n} K \quad \text{and} \quad \nabla^j K(p)=0\,, \quad j=1,2,\dots,n-2;$$

Escobar and Schoen proved that if Γ consists of the isometries acting *without* fixed point, then Problem (*) has a solution. Their main idea is to estimate the quotient by Green's function on S^n/Γ, a manifold *without* boundary.

Now, if the isometries in Γ *have* fixed point, S^n/Γ is a manifold *with* boundary. In the following, we will deal with this situation. To overcome the boundary difficulty, we construct a symmetric (in the sense of condition $\mathscr{K}_1$)) "Green's" function on the whole S^n for estimating. Combined with the other nonlinear analysis techniques, imposing some conditions on K at certain fixed points of Γ, we prove the existence of solutions for Problem (*). This generalizes some of results in paper [4].

We have recently learnt an announcement of interesting results of Bahri and Coron [1] on the S^3 problem. They replaced the symmetry conditions by conditions on the critical points of K.

After completing our paper, we were told about the results of Vaugon [7] on this problem. However, as we will show in our Remark 2.2, our results are much stronger than his.

Throughout this paper, we write the Laplacian and the gradient on the standard S^n as Δ and ∇ respectively.

Our Main Existence Results

Let $\Gamma=\{h_1,\dots,h_m\}$ where $h_i(i=1,\dots,m)$ are isometries on S^n.

Let $F_j=\{x\in S^n|\{h_1x,\dots,h_mx\}$ has exactly j distinct components$\}$ $j=1,\dots,m$. Clearly, $S^n=\bigcup_{j=1}^{m} F_j$, and $\bigcup_{j=1}^{m-1} F_j$ is the fixed points set under the action of the isometries of Γ. We will denote F_Γ the set F_1, the common fixed points set of all the isometries in Γ.

We revise the flatness condition K_2) in [4] as $\mathscr{K}_2$). Let $D_j=\bigcup_{i=1}^{j} F_i$. For any $j=2,\dots,m$, there is $p_j\in D_j$, such that

$$K(p_j)=\max_{D_j} K$$

and

$$\nabla^i K(p_j)=0\,, \quad i=1,\dots,n-2\,.$$

It is easily seen that, if the fixed point set $\bigcup_{j=1}^{m-1} F_j$ is empty, then $\mathscr{K}_2$) is just the flatness condition K_2) in [4]. As we will show in Lemma 1.3, condition $\mathscr{K}_2$) is satisfied automatically for $n=3$.

Theorem. *Assume $K_0)$, $K_1)$, and $\mathscr{K}_2)$. Then Problem $(*)$ has a solution if one of the following conditions is satisfied:*

1) $F_\Gamma = \emptyset$.
2) $\max_{F_\Gamma} K \leqq 0$.
3) *There is j_0, $2 \leqq j_0 \leqq m$, such that*

$$j_0^{2/(n-2)} \max_{F_\Gamma} K \leqq \max_{D_{j_0}} K.$$

4) $0 < \max_{F_\Gamma} K < \frac{1}{\omega_n} \int_{S^n} K(x) dV$, *where dV and ω_n are the volume element and volume of standard S^n respectively.*
5) *There exists $x_0 \in F_\Gamma$, such that*

$$K(x_0) = \max_{F_\Gamma} K, \quad and \quad \Delta K(x_0) > 0.$$

Outline of the Proof of Theorem

Let $H_* = \left\{ u \in H^1(S^n) \,\middle|\, \int_{S^n} K(x) |u|^{\tau+1} dV > 0 \right\}$. By $K_0)$, it is easily seen that $H_* \neq \emptyset$. And obviously, H_* is an open subset in $H^1(S^n)$. Write

$$X_\Gamma = \{ u \in H^1(S^n) \mid u(hx) = u(x), \text{ a.e., for any } h \in \Gamma \}.$$

Define

$$J(u) = \frac{1}{2} \int_{S^n} [|\nabla u|^2 + v_n u^2] dV - \frac{1}{\tau+1} \int_{S^n} K(x) |u|^{\tau+1} dV.$$

By the symmetry condition $K_1)$, it is known that positive critical points of the functional J in $H_* \cap X_\Gamma$ are solutions of Problem $(*)$.

Define

$$M = \{ u \in H_* \cap X_\Gamma \mid u \neq 0, \langle J'(u), u \rangle = 0 \}$$

and

$$b = \inf_{u \in M} J(u) \tag{0.2}$$

It is not difficult to see that if $u \in M$ and $J(u) = b$, then u is a critical point of J in $H_* \cap X_\Gamma$.

Analogous to [3], one can show that $b > 0$, and there is a sequence $\{u_k\} \subset M$, such that

$$J(u_k) \to b \quad \text{and} \quad J'(u_k) \to 0, \quad \text{as} \quad k \to \infty. \tag{0.3}$$

Then $\{u_k\}$ is bounded in $H^1(S^n)$, hence exists a subsequence (still denote by $\{u_k\}$) converging weakly to some element u_0 in $H^1(S^n)$. This leads to [3]

$$J'(u_0) = 0 \quad \text{and} \quad J(u_0) \leqq b. \tag{0.4}$$

Case 1. If u_0 is not identically equal to 0. Then it is not difficult to show that $u_0 \in H_* \cap X_\Gamma$, so by (0.4), $u_0 \in M$ and $J(u_0) = b$. Replace u_0 by $|u_0|$ if necessary, we obtain a solution of Problem $(*)$.

Case 2. If $u_0=0$. It is wellknown (e.g. cf. [1]) that there exist finite points $\{x_1, \ldots, x_s\} \subset S^n$ and a constant $c_0>0$, such that for any $\varepsilon>0$,

$$u_k \to 0 \quad \text{as} \quad k\to\infty \quad \text{in} \quad H^1\left(S^n \Big\backslash \bigcup_{i=1}^{s} \mathfrak{B}_\varepsilon(x_i)\right) \tag{0.5}$$

and

$$\int_{\mathfrak{B}_\varepsilon(x_i)} |\nabla u|^2 dV \geqq c_0, \tag{0.6}$$

where $\mathfrak{B}_\varepsilon(x)$ is a geodesic ball with radius ε and centered at x on standard S^n.

Under the assumptions of our Theorem, we will show that Case 2 can never happen. Otherwise, we have the following

Lemma 1.1. *Let $x_i \in S^n$ be a point of concentration of the sequence $\{u_k\}$ as was mentioned in the Case 2, then*

$$K(x_i)>0 \quad \text{and} \quad b \geqq \frac{1}{n} S^{n/2} \sum_{i=1}^{s} (K(x_i))^{(2-n)/2},$$

where

$$S = \inf_{\varphi\in H^1(S^n)} \frac{\int_{S^n} (|\nabla\varphi|^2 + v_n \varphi^2)dV}{\left\{\int_{S^n} |\varphi|^{2n/(n-2)}dV\right\}^{(n-2)/2}}$$

is the best constant in Sobolev embedding.

While by using a symmetric Green's function estimate, we can show that

Lemma 1.2. *For every $2\leqq j \leqq m$, and for any $p_j \in F_j$,* s.t.

$$K(p_j)>0 \quad \text{and} \quad \nabla^i K(p_j)=0, \quad \text{for} \quad i=1,2,\ldots,n-2.$$

We have

$$b < \frac{j}{n} S^{n/2}(K(p_j))^{(2-n)/2}.$$

Now, by $\mathscr{K}_2$) and the above two Lemmas, we must have

$$K(x_i)>0 \quad \text{and} \quad \{x_1, \ldots, x_s\} \subset F_\Gamma. \tag{0.7}$$

Then again by Lemma 1.1,

$$b \geqq \frac{1}{n} S^{n/2} \left(\max_{F_\Gamma} K\right)^{(2-n)/2}. \tag{0.8}$$

However, if K satisfies one of the conditions in the Theorem, we are able to derive contradictions with (0.7) or (0.8). Thus prove the existence of solutions for Problem (*).

1. Lemmas and the Proofs

As in Sect. 0, let M and b be defined by (0.2), $\{u_k\}$ be a sequence in M satisfying (0.3). Suppose Case 2 happen, i.e.

$$u_k \rightharpoonup 0, \quad \text{as} \quad k\to\infty; \quad \text{in} \quad H^1(S^n).$$

Then holds the following

Lemma 1.1. $K(x_i)>0$, *and*

$$b \geqq \frac{1}{n} S^{n/2} \sum_{i=1}^{s} (K(x_i))^{(2-n)/2}, \tag{1.1}$$

where $\{x_1, \ldots, x_s\}$ *is defined by* (0.5) *and* (0.6).

Proof. 1) Since $u_k \in M$, we have

$$\int_{S^n} \{|\nabla u_k|^2 + v_n u_k^2\} dV = \int_{S^n} K(x) |u_k|^{\tau+1} dV.$$

This implies

$$J(u_k) = \frac{1}{n} \int_{S^n} K(x) |u_k|^{\tau+1} dV. \tag{1.2}$$

2) Since $J'(u_k) \to 0$, as $k \to \infty$; one has

$$-\Delta u_k + v_n u_k = K(x) |u_k|^{\tau-1} u_k + o_k(1), \tag{1.3}$$

where $o_k(1) \to 0$ as $k \to \infty$ in the dual space of $H^1(S^n)$.

Let $\eta_i \in C_0^\infty(\mathfrak{B}_\varepsilon(x_i))$; $0 \leqq \eta_i \leqq 1$, and

$$\eta_i = 1 \quad \text{for} \quad x \in \mathfrak{B}_{\varepsilon/2}(x_i).$$

Note that $u_k \rightharpoonup 0$ in $H^1(S^n)$ and consequently $u_k \to 0$ in $L^2(S^n)$, and by (0.5), one can easily verify that

$$\int_{\mathfrak{B}_\varepsilon(x_i)} |\nabla(\eta_i u_k)|^2 dV = \int_{\mathfrak{B}_\varepsilon(x_i)} K |\eta_i u_k|^{\tau+1} dV + o_k(1). \tag{1.4}$$

(1.4) and (0.6) imply, for k sufficiently large,

$$\int_{\mathfrak{B}_\varepsilon(x_i)} K |\eta_i u_k|^{\tau+1} dV \geqq c_0/2. \tag{1.5}$$

Noting that $\{u_k\}$ is bounded in $H^1(S^n)$ and consequently, bounded in $L^{\tau+1}(S^n)$, while ε is arbitrary, by the continuity of K, one can easily seen that $K(x_i)$ is positive $(i=1, \ldots, s)$.

4) Since $u_k \to 0$ in $L^2(S^n)$, by (1.4) and (1.5), we have

$$\frac{\int_{\mathfrak{B}_\varepsilon(x_i)} \{|\nabla(\eta_i u_k)|^2 + v_n(\eta_i u_k)^2\} dV}{\left\{\int_{\mathfrak{B}_\varepsilon(x_i)} K |\eta_i u_k|^{\tau+1} dV\right\}^{2/(\tau+1)}} \leqq \left(\int_{\mathfrak{B}_\varepsilon(x_i)} K |\eta_i u_k|^{\tau+1} dV\right)^{2/n} + o_k(1).$$

Then by the definition of S and η_i, as well as the continuity of K,

$$\int_{\mathfrak{B}_{\varepsilon/2}(x_i)} K |u_k|^{\tau+1} dV \geqq K(x_i)^{(2-n)/2} S^{n/2} + o_k(1) + \alpha(\varepsilon),$$

where $\alpha(\varepsilon) \to 0$ as $\varepsilon \to 0$. Now, (1.1) follows from (1.2) easily.

Lemma 1.2. *For* $2 \leqq j \leqq m$, *if there exists* $p_j \in F_j$, *s.t.*

$$K(p_j) > 0 \quad \textit{and} \quad \nabla^i K(p_j) = 0, \quad i = 1, \ldots, n-2.$$

Then

$$b < \frac{j}{n} S^{n/2}(K(p_j))^{(2-n)/2}. \tag{1.6}$$

Proof. For conciseness, we prove this for $j=2$, and the similar argument works for a general $j>2$.

1) Let q_0 be any point in F_Γ. Then obviously $-q_0 \in F_\Gamma$. Let

$$\pi : S^n \backslash \{-q_0\} \to \mathbb{R}^n$$

be the stereographic projectioon map with q_0 lying on $0 \in \mathbb{R}^n$, and $\tilde{g} = \pi^*(\delta_{ij})$ be the pullback metric on $S^n \backslash \{-q_0\}$.

For $p_2 \in F_2$, write $\{h_1 p_2, \ldots, h_m p_2\} = \{p_0, \bar{p}_0\}$. At pole p_0, the Green's function of the conformal Laplacian $\Delta_{\tilde{g}}$ under the metric $\tilde{g}$ on $S^n \backslash \{-q_0\}$ is $a|p-p_0|^{2-n}$, where $|p-p_0|$ is the distance between the 2 points under the metric $\tilde{g}$, and a is a constant.

Claim.

$$|hp - hp_0| = |p - p_0|, \quad \text{for any} \quad p \in S^n \backslash \{-q_0\}, h \in \Gamma. \tag{1.7}$$

Proof of the Claim. For any $\zeta \in S^n$, $\pi(\zeta)$ is the tangent vector at q_0 of the geodesic (on the standard S^n) linking q_0 and ζ. Since h is an isometry on S^n,

$$\angle(\pi(p), \pi(p_0)) = \angle(\pi(hp), \pi(hp_0)), \tag{1.8}$$

where $\angle(X, Y)$ stands for the angle between the 2 vectors X and Y in $\mathbb{R}^n$. Let $d(\,,\,)$ be the geodesic distance on standard S^n. Note that $q_0 \in F_\Gamma$, we have

$$d(hp_0, q_0) = d(hp_0, hq_0) = d(p_0, q_0),$$
$$d(hp, q_0) = d(p, q_0).$$

Hence,

$$|\pi(hp_0)| = |\pi(p_0)| \quad \text{and} \quad |\pi(hp)| = |\pi(p)|. \tag{1.9}$$

Now, from (1.8) and (1.9), it is easy to see that

$$|\pi(hp) - \pi(hp_0)| = |\pi(p) - \pi(p_0)|.$$

Therefore (1.7) is true. The same equality holds for $\bar{p}_0$.

Let

$$G(x) = |x - p_0|^{2-n} + |x - \bar{p}_0|^{2-n}, \qquad x \in S^n \backslash \{-q_0\}.$$

Then

$$G(hx) = G(x), \quad \text{for any} \quad h \in \Gamma,\ x \in S^n \backslash \{-q_0\}.$$

In fact, by (1.7),

$$G(hx) = |x - h^{-1}p_0|^{2-n} + |x - h^{-1}\bar{p}_0|^{2-n}$$
$$= |x - p_0|^{2-n} + |x - \bar{p}_0|^{2-n} = G(x)$$

due to the definition of $\{p_0, \bar{p}_0\}$. Where h^{-1} is the inverse of h.

2) Let

$$\lambda(x)=\left|\cos\frac{d(q_0,x)}{2}\right|^{2-n}.$$

Then

$$\lambda\in C^\infty(S^n\backslash\{-q_0\})\quad\text{and}\quad \tilde{g}=\lambda^{\tau-1}g_0.$$

Define

$$\lambda G(-q_0)=\lim_{x\to -q_0}\lambda G(x).$$

Then it is easy to see that

$$\lambda G\in X_\Gamma. \tag{1.10}$$

Observe that the functions

$$u_{\varepsilon,p}(x)=\left(\frac{\varepsilon}{\varepsilon^2+|x-p|^2}\right)^{(n-2)/2}$$

are solutions of the equation

$$-\Delta_{\tilde{g}}u_{\varepsilon,p}=n(n-2)u^\tau_{\varepsilon,p},\quad\text{for}\quad x\in S^n\backslash\{-q_0\},$$

while $-\Delta_{\tilde{g}}$ and $-\Delta+v_n$ are conformal, we have

$$-\Delta(\lambda u_{\varepsilon,p})+v_n\lambda u_{\varepsilon,p}=n(n-2)(\lambda u_{\varepsilon,p})^\tau,\quad\text{for}\quad x\in S^n. \tag{1.11}$$

Let $\mathfrak{B}_\varrho(p)$ be the ball centered at p with radius ϱ under the metric $\tilde{g}$. Choose ϱ_0 sufficiently small that

$$\mathfrak{B}_{\varrho_0}(p_0)\cap\mathfrak{B}_{\varrho_0}(\bar{p}_0)=\emptyset\quad\text{and}\quad -q_0\notin\mathfrak{B}_{2\varrho_0}(p_0)\cup\mathfrak{B}_{2\varrho_0}(\bar{p}_0).$$

Let $\Psi\in C_0^\infty(\mathfrak{B}_{2\varrho_0}(p_0)\cup\mathfrak{B}_{2\varrho_0}(\bar{p}_0))$, $\Psi(x)=1$ for $x\in\mathfrak{B}_{\varrho_0}(p_0)\cup\mathfrak{B}_{\varrho_0}(\bar{p}_0)$ and Ψ only depends on $|x-p_0|$ or on $|x-\bar{p}_0|$ in $\mathfrak{B}_{2\varrho_0}(p_0)$ or in $\mathfrak{B}_{2\varrho}(\bar{p}_0)$ respectively.

Define

$$\varphi(x)=\begin{cases} u_{\varepsilon,p_0}(x) & x\in\mathfrak{B}_{\varrho_0}(p_0)\\ u_{\varepsilon,\bar{p}_0}(x) & x\in\mathfrak{B}_{\varrho_0}(\bar{p}_0)\\ \varepsilon_0(G(x)-\Psi\alpha(x)) & x\in\{\mathfrak{B}_{2\varrho_0}(p_0)\backslash\mathfrak{B}_{\varrho_0}(p_0)\}\cup\{\mathfrak{B}_{2\varrho_0}(\bar{p}_0)\backslash\mathfrak{B}_{\varrho_0}(\bar{p}_0)\}\\ \varepsilon_0 G(x) & x\in S^n\backslash\{\mathfrak{B}_{2\varrho_0}(p_0)\cup\mathfrak{B}_{2\varrho_0}(\bar{p}_0), \end{cases}$$

where

$$\alpha(x)=\begin{cases}|x-\bar{p}_0|^{2-n}-A & x\in\mathfrak{B}_{2\varrho_0}(p_0)\\ |x-p_0|^{2-n}-A & x\in\mathfrak{B}_{2\varrho}(\bar{p}_0)\end{cases}\quad\text{with}\quad A=|p_0-\bar{p}_0|^{2-n}.$$

In order for φ to be continuous, we require (cf. [6])

$$\varepsilon_0(\varrho_0^{2-n}+A)=\left(\frac{\varepsilon}{\varepsilon^2+\varrho_0^2}\right)^{(n-2)/2}.$$

By (1.10), it is not difficult to verify that $\lambda\varphi \in X_\Gamma$, i.e. $\lambda\varphi(hx) = \lambda\varphi(x)$ for any $h \in \Gamma$. Moreover, by the assumption $K(p_0) = K(\bar{p}_0) > 0$, one can choose ε so small that

$$\int_{S^n} K|\lambda\varphi|^{\tau+1} dV > 0.$$

Now $\lambda\varphi \in H_* \cap X_\Gamma$.

We are going to use $\lambda\varphi$ to estimate the quotient

$$Q(u) = \frac{E(u)}{\left\{\int_{S^n} K|u|^{\tau+1} dV\right\}^{2/(\tau+1)}},$$

where

$$E(u) = \int_{S^n} (|\nabla u|^2 + v_n u^2) dV.$$

The following approach is somewhat standard (cf. [6, 4]).

By (1.11), we have

$$\begin{aligned} &\int_{\mathfrak{B}_{\varrho 0}(p_0)} \{|\nabla(\lambda u_{\varepsilon, p_0})|^2 + v_n(\lambda u_{\varepsilon, p_0})^2\} dV \\ &= n(n-2) \int_{\mathfrak{B}_{\varrho 0}(p_0)} (\lambda u_{\varepsilon, p_0})^{\tau+1} dV + \int_{\partial\mathfrak{B}_{\varrho 0}(p_0)} \lambda u_{\varepsilon, p_0} \frac{\partial}{\partial v}(\lambda u_{\varepsilon, p_0}) dS \\ &\leqq S\left\{\int_{\mathfrak{B}_{\varrho 0}(p_0)} (\lambda u_{\varepsilon, p_0})^{\tau+1} dV\right\}^{(n-2)/n} + \int_{\partial\mathfrak{B}_{\varrho 0}(p_0)} \lambda u_{\varepsilon, p_0} \frac{\partial}{\partial v}(\lambda u_{\varepsilon, p_0}) dS. \end{aligned} \tag{1.12}$$

The same inequality holds for $\mathfrak{B}_{\varrho 0}(\bar{p}_0)$ due to the symmetry of $\lambda\varphi$. Here we have used the fact that

$$S = n(n-2)\left\{\int_{\mathbb{R}^n} u_\varepsilon^{\tau+1} dx\right\}^{2/n}. \tag{1.13}$$

Note that

$$-\Delta(\lambda G) + v_n(\lambda G) = 0 \quad \text{for} \quad x \in S^n \setminus \{\mathfrak{B}_{\varrho 0}(p_0) \cup \mathfrak{B}_{\varrho 0}(\bar{p}_0)\},$$

and due to the symmetry of $\lambda\varphi$, we have

$$\begin{aligned} E(\lambda\varphi) \leqq\; & 2S\left\{\int_{\mathfrak{B}_{\varrho 0}(p_0)} |\lambda\varphi|^{\tau+1} dV\right\}^{(n-2)/2} \\ & + 2\int_{\partial\mathfrak{B}_{\varrho 0}(p_0)} \left\{\lambda u_{\varepsilon, p_0} \frac{\partial}{\partial v}(\lambda u_{\varepsilon, p_0}) - \varepsilon_0^2 \lambda G \frac{\partial}{\partial v}(\lambda G)\right\} dS + c\varrho_0 \varepsilon_0^2. \end{aligned}$$

Since $\lambda > 0$, and $\lambda \in C^\infty(S^n \setminus \{-q_0\})$, similar to [6], we obtain

$$E(\lambda\varphi) \leqq 2S\left\{\int_{\mathfrak{B}_{\varrho 0}(p_0)} |\lambda\varphi|^{\tau+1} dV\right\}^{(n-2)/n} - a_0 A \varepsilon_0^2 + c\varrho_0^{-n} \varepsilon_0^{\tau+1} + c\varrho_0 \varepsilon_0^2$$

with $a_0 > 0$.

Now, due to K_1) and the flatness assumption

$$\nabla^i K(p_0) = 0, \qquad i = 1, \ldots, n-2$$

analogous to [4], we arrive at

$$E(\lambda\varphi) \leqq 2^{2/n} K(p_0)^{(2-n)/n} S \left\{ \int_{\mathfrak{B}_{\varrho_0}(p_0) \cup \mathfrak{B}_{\varrho_0}(p_0)} K(x) |\lambda\varphi|^{\tau+1} dV \right\}^{(n-2)/n}$$
$$- a_0 A \varepsilon_0^2 + c(\varepsilon^{n-1} + \varrho_0^{-n} \varepsilon_0^{\tau+1} + \varrho_0 \varepsilon_0^2)$$
$$\leqq 2^{2/n} K(p_0)^{(2-n)/n} S \left\{ \int_{S^n} K(x) |\lambda\varphi|^{\tau+1} dV \right\}^{(n-2)/n}$$
$$- a_0 A \varepsilon_0^2 + c(\varrho_0^{-2n} \varepsilon_0^{\tau+1} + \varrho_0 \varepsilon_0^2).$$

Because $a_0 A > 0$, we can choose ϱ_0 small and ε_0 smaller to verify

$$E(\lambda\varphi) < 2^{2/n} K(p_0)^{(2-n)/n} S \left\{ \int_{S^n} K(x) |\lambda\varphi|^{\tau+1} dV \right\}^{(n-2)/n}.$$

That is

$$Q(\lambda\varphi) < 2^{2/n} K(p_0)^{(2-n)/n} S. \tag{1.14}$$

Choose a constant t such that $\langle J'(t\lambda\varphi), \lambda\varphi \rangle = 0$, then

$$t\lambda\varphi \in M, \quad \text{and} \quad J(t\lambda\varphi) = \frac{1}{n} Q(\lambda\varphi)^{n/2}.$$

Therefore $b \leqq J(t\lambda\varphi)$, and (1.6) ($j=2$) follows from (1.14). This completes the proof.

Lemma 1.3. *The flatness condition $\mathscr{K}_2$) is always satisfied for $n=3$.*

Proof. It suffice to show that for any $2 \leqq j \leqq m$, at maximal points of K on D_j,

$$\nabla K = 0.$$

Consider the case $j = m-1$. Suppose that $p_0 \in D_{m-1}$,

$$K(p_0) = \max_{D_{m-1}} K, \quad \text{but} \quad \nabla K(p_0) \neq 0. \tag{1.15}$$

Then by the definition of D_{m-1}, there is at least one $h \in \Gamma$, such that $hp_0 = p_0$. By the linearity of h and because of $K(hx) = K(x)$ for any $x \in S^3$, we have

$$\nabla K(x)|_{x=p_0} = \nabla K(hx)|_{x=p_0} = h \nabla K(x)|_{x=hp_0} = h \nabla K(x)|_{x=p_0}. \tag{1.16}$$

Let Π be the plane in $\mathbb{R}^4$ spanned by the vectors $\nabla K(p_0)$ and p_0, let $S^1 = \Pi \cap S^3$. Then due to $hp_0 = p_0$ and (1.16), and the linearity of h, it is easy to see that

$$hx = x \quad \forall x \in S^1$$

which implies $S^1 \subset D_{m-1}$. Hence

$$K(p_0) = \max_{x \in S^1} K.$$

This leads to $\nabla K(p_0) = 0$, a contradiction with (1.15). Therefore we must have

$$\nabla K(p_0) = 0.$$

Similarly, one can prove that for any $2 \leqq j \leqq m$, at maximal points of K on D_j, holds $\nabla K = 0$.

Remark 1.1. In fact, the conclusion of Lemma 1.3 is true on S^n for any n and for any $j = 1, 2, \ldots, m$.

2. Existence Theorems and the Proofs

Throughout this section, we assume that $K(x)$ satisfies the conditions K_0), K_1), and $\mathscr{K}_2$).

Let's still consider the sequence $\{u_k\} \subset M$, such that

$$J(u_k) \to b, \quad \text{and} \quad J'(u_k) \to 0, \quad \text{as} \quad k \to \infty.$$

As we argued in Sect. 0, in order to find a solution of Problem (*), one only needs to show that Case 2 can't happen. We argue indirectly. Suppose Case 2 occurs, i.e.

$$u_k \rightharpoonup 0 \quad \text{as} \quad k \to \infty, \quad \text{in} \quad H^1(S^n).$$

Then by Lemma 1.1, one must have

$$K(x_i) > 0 \quad \text{and} \quad b \geqq \frac{1}{n} S^{n/2} \sum_{i=1}^{s} K(x_i)^{(2-n)/2}, \tag{2.1}$$

where $\{x_1, \ldots, x_s\}$ is defined by (0.5) and (0.6).

Under the assumptions of the following theorems, we will derive contradictions with (2.1), therefore prove the existence of solutions for Problem (*).

Theorem 2.1. *If $F_\Gamma = \emptyset$ or if $\max\limits_{F_\Gamma} K \leqq 0$. Then Problem (*) has a solution.*

Proof. Since $K(x_i) > 0$, under either one of the assumption of the Theorem, we have

$$\{x_1, \ldots, x_s\} \cap F_\Gamma = \emptyset.$$

Consequently, there exists $x_i \in F_j$ with $2 \leqq j \leqq m$, $1 \leqq i \leqq s$. By the symmetry of K, there are j points in $\{x_1, \ldots, x_s\}$ belonging to D_j, hence by (2.1),

$$b \geqq \frac{j}{n} S^{n/2} \left[\max_{D_j} K\right]^{(2-n)/2}. \tag{2.2}$$

while on the other hand, by $\mathscr{K}_2$) and the definition of D_j, there exists $p \in F_k$ with $2 \leqq k \leqq j$, such that

$$K(p) = \max_{D_j} K \quad \text{and} \quad \nabla^i K(p) = 0, \qquad i = 1, \ldots, n-2.$$

Hence by Lemma 1.2,

$$b < \frac{k}{n} S^{n/2} \left[\max_{D_j} K\right]^{(2-n)/2}.$$

An obvious contradiction with (2.2). This completes the proof.

The following theorems deal with the situation that $\max\limits_{F_\Gamma} K > 0$.

Theorem 2.2. *If there is j_0, $2 \leqq j_0 \leqq m$, s.t.*

$$j_0^{2/(n-2)} \max_{F_\Gamma} K \leqq \max_{D_{j_0}} K. \tag{2.3}$$

Then Problem (*) has a solution.

Proof. First case, $\{x_1,\ldots,x_s\}\cap F_\Gamma=\emptyset$, the proof is then the same as the one in Theorem 2.1.

Second case, $\{x_1,\ldots,x_s\}\cap F_\Gamma\neq\emptyset$, then

$$b\geqq\frac{1}{n}S^{n/2}\left[\max_{F_\Gamma}K\right]^{(2-n)/2}. \tag{2.4}$$

while by $\mathscr{K}_2$) and Lemma 1.2,

$$b<\frac{j}{n}S^{n/2}\left[\max_{D_j}K\right]^{(2-n)/2}\quad\forall j=2,\ldots,m.$$

Consequently,

$$j\left[\max_{D_j}K\right]^{(2-n)/2}>\left[\max_{F_\Gamma}K\right]^{(2-n)/2}\quad\forall j=2,\ldots,m.$$

This contradicts with (2.3). The proof is completed.

Now, from the proofs of Theorem 2.1 and Theorem 2.2, we see that in order to find a solution of Problem (*), it suffice to derive a contradiction with (2.4), i.e. to verify the following inequality

$$b<\frac{1}{n}S^{n/2}\left[\max_{F_\Gamma}K\right]^{(2-n)/2}. \tag{2.5}$$

Theorem 2.3. *If*

$$0<\max_{F_\Gamma}K<\frac{1}{\omega_n}\int_{S^n}K(x)dV. \tag{2.6}$$

Then Problem (*) *has a solution.*

Proof. We are going to verify (2.5). By the assumption $0<\int_{S^n}K(x)dV$, we see that any constant $c\in H_*$. Choose a proper c such that $\langle J'(c),c\rangle=0$, then $c\in M$. Hence

$$b\leqq J(c)=\frac{1}{n}S^{n/2}\left\{\omega_n\Big/\int_{S^n}K(x)dV\right\}^{(n-2)/2} \tag{2.7}$$

because $S=\dfrac{n(n-2)}{4}\omega_n^{2/n}$. Now it is easily seen that (2.6) and (2.7) imply (2.5). This completes the proof.

Theorem 2.4. *If there exists* $x_0\in F_\Gamma$, *such that*

$$K(x_0)=\max_{F_\Gamma}K>0\quad\textit{and}\quad\Delta K(x_0)>0.$$

Then Problem (*) *admits a solution.*

Proof. Let

$$w_\varepsilon(x)=(2\varepsilon)^{(n-2)/2}\{\varepsilon^2+4+(\varepsilon^2-4)\cos d(x_0,x)\}^{(2-n)/2}.$$

Then it is not difficult to verify that

$$-\Delta w_\varepsilon+v_nw_\varepsilon=n(n-2)w_\varepsilon^\tau,\qquad\forall x\in S^n, \tag{2.8}$$

Note that

$$hx_0 = x_0 \quad \text{and} \quad d(hx_0, hx) = d(x_0, x) \quad \forall h \in \Gamma,\ x \in S^n,$$

it is easily seen that

$$w_\varepsilon(hx) = w_\varepsilon(x) \quad \forall h \in \Gamma,\ x \in S^n.$$

That is

$$w_\varepsilon \in X_\Gamma.$$

Since $K(x_0) > 0$, and

$$w_\varepsilon(x) \to \begin{cases} 0 & x \neq x_0 \\ +\infty & x = x_0 \end{cases} \quad \text{as} \quad \varepsilon \to 0.$$

We see that for sufficiently small ε, $w_\varepsilon \in H_*$. Choose a suitable constant t_ε such that $\langle J'(t_\varepsilon w_\varepsilon), w_\varepsilon \rangle = 0$, then

$$t_\varepsilon w_\varepsilon \in M \quad \text{and} \quad b \leqq J(t_\varepsilon w_\varepsilon) = \frac{1}{n}[Q(w_\varepsilon)]^{n/2}. \tag{2.9}$$

Now, let's estimate the quotient $Q(w_\varepsilon)$. Again let π be the stereographic projection from $S^n \backslash \{-x_0\}$ to $\mathbb{R}^n$, with x_0 lying on the origin of $\mathbb{R}^n$. Let $B_\varrho(0)$ be the ball of radius ϱ and centered at origin in $\mathbb{R}^n$. Obviously, $B_\varrho(0) = \pi(\mathfrak{B}_\varrho(x_0))$, where $\mathfrak{B}_\varrho(x)$ is defined in Sect. 1. Taking into account of the fact that

$$\int_D w_\varepsilon^{\tau+1} dV = \int_{\pi(D)} u_\varepsilon^{\tau+1} dx \quad \forall D \subset S^n \tag{2.10}$$

by (1.13), (2.8) and through a direct calculation, we obtain

$$\begin{aligned}\int_{S^n} \{|\nabla w_\varepsilon|^2 + v_n w_\varepsilon^2\} dV \leqq S \left\{ \int_{\mathfrak{B}_\varrho(x_0)} w_\varepsilon^{\tau+1} dV \right\}^{(n-2)/n} \\ + c \int_{\mathbb{R}^n \backslash B_\varrho(0)} u_\varepsilon^{\tau+1} dx.\end{aligned} \tag{2.11}$$

Using the second order Taylor expansion of the function $K(x)$ at point x_0, taking into account that

$$\int_{B_\varrho(0)} y_i y_j u_\varepsilon(y)^{\tau+1} dy = 0 \quad \text{for} \quad i \neq j,\ 1 \leqq i, j \leqq n$$

due to the symmetry of u_ε (here $y = (y_1, \ldots, y_n)$); we arrive at

$$\int_{\mathfrak{B}_\varrho(x_0)} w_\varepsilon^{\tau+1} dV \leqq \frac{1}{K(x_0)} \int_{\mathfrak{B}_\varrho(x_0)} [K(x) - \tfrac{1}{4} \Delta K(x_0) d^2(x, x_0)] w_\varepsilon^{\tau+1} dV \tag{2.12}$$

for ϱ sufficiently small.

By the assumption that $\Delta K(x_0) > 0$, boundedness of K and (2.10), (2.11), and (2.12), we have

$$\begin{aligned}\int_{S^n} \{|\nabla w_\varepsilon|^2 + v_n w_\varepsilon^2\} dV \leqq S[K(x_0)]^{(2-n)/n} \left\{ \int_{S^n} K(x) w_\varepsilon^{\tau+1} dV \right\}^{(n-2)/n} \\ - c_1 \int_{B_\varrho(0)} |x|^2 u_\varepsilon^{\tau+1} dx + c \int_{\mathbb{R}^n \backslash B_\varrho(0)} u_\varepsilon^{\tau+1} dx\end{aligned}$$

with the constant $c_1 > 0$. Now, by an elementary calculus, it is not difficult to verify that, for ϱ small and ε much smaller,

$$Q(w_\varepsilon) < S[K(x_0)]^{(2-n)/n}.$$

This verifies (2.5) due to (2.9), and the proof is completed.

Remark 2.1. By Theorem 2.1, if the fixed points set on S^n under the action of Γ is empty, and if K satisfies K_0), K_1), and $\mathscr{K}_2$), the Problem (*) has a solution, which implies this part of results in paper [4].

Remark 2.2. In paper [7], the sufficient conditions for Problem (*) to have a solution are, in the language of our paper

1) K satisfies K_0) and K_1)
2) for all $x \in S^n$,

$$b < \frac{j(x)}{n} S^{n/2}[K(x)]^{(2-n)/2} \tag{2.13}$$

where $j(x) = j$ for $x \in F_j$, $j = i, 2, \ldots, m$.

However, in our paper, due to Lemma 1.1, we only require condition (2.13) be satisfied on at most m points. Moreover, if K satisfies the flatness condition $\mathscr{K}_2$) (it is satisfied automatically for $n = 3$) then we only require (2.13) be satisfied at one point on S^n, that is [cf. (2.5)], if there is $x_1 \in F_\Gamma$, such that

$$K(x_1) = \max_{F_\Gamma} K \quad \text{and} \quad b < \frac{1}{n} S^{n/2}[K(x_1)]^{(2-n)/2}. \tag{2.14}$$

Furthermore, we provide some direct and verifiable conditions on K (cf. Theorems 2.1–2.4) so that (2.14) can be satisfied. Now one can see that our results are much stronger than that in paper [7].

Acknowledgement. The author would like to thank Prof. Kazdan for his helpful suggestions and discussions.

References

1. Bahri, A., Coron, J.: Vers une théorie des points critiques à l'infini. Preprint
2. Bourguignon, J.P., Ezin, J.P.: Scalar curvature functions in a conformal class of metric and conformal transformations. Trans. Am. Math. Soc. **301**, 723–736 (1987)
3. Brezis, H., Nirenberg, L.: Positive solutions of nonlinear elliptic equations involving critical Sobolev exponents. Commun. Pure Applied Math. **36**, 437–477 (1983)
4. Escobar, J., Schoen, R.: Conformal metrics with prescribed scalar curvature (to appear)
5. Kazdan, J., Warner, F.: Curvature functions for compact 2-manifolds. Ann. Math. **99**, 14–47 (1974)
6. Schoen, R.: Conformal deformations of a Riemannian metric to constant scalar curvature. J. Differ. Geom. **20**, 479–495 (1984)
7. Vaugon, M.: Transformation conforme de la courbure scalaire. Analyse non linéaire. Ann. Inst. Henri Poincaré **3**, 55–65 (1986)

Received March 9, 1987; in revised form September 9, 1987

Math. Ann. 283, 367–374 (1989)

Mathematische Annalen

Even Unimodular 8-Dimensional Quadratic Forms Over $\mathbb{Q}(\sqrt{2})$

J. S. Hsia[1,*] and D. C. Hung[2]

[1] Department of Mathematics, Ohio State University, Columbus, OH 43210, USA
[2] Department of Mathematical Sciences, SUNY at Binghamton, Binghamton, NY 13901, USA

Dedicated to Professor O. T. O'Meara on his 60th birthday

Introduction

The problem of classifying integral positive definite quadratic forms (or lattices) has been studied by many people. One frequent approach is the method of Kneser [6], which involves studying lattices adjacent to a given one. An example of this is Niemeier's complete classification of 24-dimensional even positive definite unimodular lattices over $\mathbb{Z}$ [9]. Unfortunately, this method in general requires very extensive calculations and it is difficult to verify that the enumeration is complete. By adopting approaches based on the theory of modular forms, algebraic coding, and Siegel's analytic theory of quadratic forms, it is possible to simplify the classification and indeed to give proofs of the completeness of the enumeration, see e.g. [13, 2, 3]. In this paper we consider even positive definite unimodular lattices over the ring of integers R in $\mathbb{Q}(\sqrt{2})$. By even we mean that $B(x,x) \in 2R$. There is a unique genus of such lattices in each dimension that is a multiple of 4. It is known that the 4 dimensional genus has only one class [12] which we denote by Δ_4'. Computations of the Minkowski-Siegel mass [5] suggest that further algebraic classification is feasible only for dimension 8. Several classes have been found in the 8-dimensional genus, including one with an empty root system [12, 5]. Recall that the root system of a lattice is the set of vectors of norm 2. We shall complete the enumeration. Our approach is based on a combination of Kneser's method and Siegel's theory of quadratic forms. Siegel's mass formula and his theorem for degree one Hilbert-Eisenstein series are used to verify that our enumeration is complete. A neighbor graph for the genus will be given in the last section. We note that the graph provides an alternate check of the completeness of the genus. Unless otherwise indicated, all terminology and notations will follow those of [10, 3].

* Research partially supported by N.S.F.

Enumeration

Let $F=\mathbb{Q}(\sqrt{2})$, R the ring of integers of F. Then $R=\mathbb{Z}[\varepsilon]$, where $\varepsilon=1+\sqrt{2}$ is the fundamental unit of F. Aside from the classical root systems of ADE-types, there are two new irreducible root systems over R [8], namely

$$\begin{aligned}\Delta_n(n\geqq 2)&=\{z\in I_n \mid B(z, e_1+\ldots+e_n)\equiv 0 \,(\mathrm{mod}\sqrt{2})\}\\ &=\langle\sqrt{2}e_1, e_1+e_2, \ldots, e_1+e_n\rangle\end{aligned}$$

and

$$\Delta_4'=\Delta_4+\langle(e_1+\ldots+e_4)/\sqrt{2}\rangle=\langle\sqrt{2}e_1, (e_1+\ldots+e_4)/\sqrt{2}, e_1+e_3, e_1+e_4\rangle$$

where $\{e_i\}$ is an orthonormal basis and $I_n=\langle e_1, \ldots, e_n\rangle$. We have

$$\det \Delta_4'=1$$

so Δ_4' generates an even unimodular lattice, which yields the only class in the quaternary genus. In the genus of 8-dimensional even unimodular lattices, it is known that there are at least 5 classes [12, 5], given here by their root system configurations:

$$E_8,\ 2\Delta_4',\ \Delta_8,\ 2D_4 \quad \text{and} \quad \emptyset\ .$$

An even unimodular lattice with the root system Γ will be denoted by L_Γ or just Γ when there is no confusion.

The classes E_8, $2\Delta_4'$ and L_{2D_4} can be obtained by the Kneser method as neighbors of L_{Δ_8} with respect to the prime $\sqrt{2}$. On the other hand, the existence of an empty root lattice $L_\emptyset$ was shown using a technique which is analogous to the Construction A of [7] in coding theory. Now $L_\emptyset$ is not adjacent to the classes E_8, $2\Delta_4'$, L_{Δ_8} and L_{2D_4}. To see this, let K be a neighbor of $L_\emptyset$. Then

$$\begin{aligned}K&=Rx+(L_\emptyset)_x\\ &=Rx+\{y\in L_\emptyset \mid B(y,x)\in R\}\end{aligned}$$

for some $x\notin L_\emptyset$, $\sqrt{2}x\in L_\emptyset$, $Q(x)\in 2R$. If K is equivalent to E_8, $2\Delta_4'$, L_{Δ_8} or L_{2D_4}, then K contains a binary sublattice

$$B\cong\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}$$

adapted to the base $\{u, v\}$. It is clear that $u, v\notin L_\emptyset$, hence there exist $c_1, c_2\in R-\sqrt{2}R$ such that

$$u=c_1x+y_1 \quad \text{and} \quad v=c_2x+y_2$$

for some $y_1, y_2\in(L_\emptyset)_x$. Since $|R/\sqrt{2}R|=2$, we may assume that $c_1=c_2=1$. It follows that $u-v\in(L_\emptyset)_x\subset L_\emptyset$, which is impossible because $B(u-v, u-v)=2$. This suggests that there is a "missing link" between $L_\emptyset$ and the remaining classes. This will be filled by a lattice with the root system $4\Delta_2$. Specifically, consider 4 copies of Δ_2 in the 8-dimensional space:

$$4\Delta_2=\langle\sqrt{2}e_1, e_1+e_2\rangle\perp\langle\sqrt{2}e_3, e_3+e_4\rangle\perp\langle\sqrt{2}e_5, e_5+e_6\rangle\perp\langle\sqrt{2}e_7, e_7+e_8\rangle\ .$$

Then

$$L_{4\Delta_2}=4\Delta_2+\langle(\varepsilon e_1+e_2+\varepsilon e_3+e_4)/\sqrt{2},\ (\varepsilon e_5+e_6+\varepsilon e_7+e_8)/\sqrt{2}\ ,\\ e_1+e_3+e_5+e_7,\ (\varepsilon e_1+e_2+\varepsilon e_5+e_6)/\sqrt{2}\rangle\ . \tag{1}$$

Alternately, $L_{4\Delta_2}$ can be obtained by the neighborhood method using the base lattice

$$2\Delta_4'=\langle\sqrt{2}e_1,(e_1+\ldots+e_4)/\sqrt{2},e_1+e_3,e_1+e_4\rangle\perp\\ \langle\sqrt{2}e_5,(e_5+\ldots+e_8)/\sqrt{2},e_5+e_7,e_5+e_8\rangle$$

and the vector $(\varepsilon e_1+e_2+\varepsilon e_3+e_4+\varepsilon e_5+e_6+\varepsilon e_7+e_8)/2$. Using (1) for $L_{4\Delta_2}$ and taking its neighbor which contains $(\varepsilon e_1+e_2+\varepsilon e_3+e_4+\varepsilon e_5+e_6+\varepsilon e_7+e_8)/2$, one obtains a lattice that has an empty root system. We have

Proposition 1. *There exists an 8-dimensional even unimodular lattice $L_{4\Delta_2}$, over $\mathbb{Q}(\sqrt{2})$ which has the root system $4\Delta_2$.*

Remark. It is not clear that each of the known root systems gives rise to a unique class of lattices. In particular, there may be more than one class of lattice which has an empty root system. We will show that there are no further classes in our genus via Siegel's mass formula. Alternate checks will be given using Siegel's theorem for degree one Hilbert-Eisenstein series and the neighbor graph of the genus.

Algebraic Descent

Let L be an R-lattice of $\operatorname{rank}_R(L)=m$. The algebraic descent L_0 of L is the $\mathbb{Z}$-lattice $L_0=L$ of $\operatorname{rank}_{\mathbb{Z}}(L_0)=2m$ together with the quadratic form Q_0 defined by

$$Q_0(x):=Tr_{F/\mathbb{Q}}(Q(x)/2\sqrt{2}\varepsilon)\ .$$

If (L,Q) is even unimodular over R, then (L_0,Q_0) is even unimodular over $\mathbb{Z}$. For $Q(x)=a+b\varepsilon$, $a,b\in\mathbb{Z}$ we have

$$Q_0(x)=Tr_{F/\mathbb{Q}}((a+b\varepsilon)/2\sqrt{2}\varepsilon)=a$$

so that

$$Q_0(x)=2 \text{ iff } Q(x)\in\{2,2\sqrt{2}\varepsilon,2(1+2\varepsilon)\}\ .$$

If u,v are any two roots in L, then $B(u,v)=0,\ \pm1,\ \pm\sqrt{2}$, hence $B_0(u,v)=0,\ \pm1,\ \mp1$ respectively. Let Γ be an irreducible root system over $\mathbb{Q}(\sqrt{2})$. In [5], it was shown that if Γ is "old" (i.e. the classical ADE root system) then the algebraic descent Γ_0 is 2 copies of Γ, whereas if $\Gamma=\Delta_n$ $(n\geqq2)$, Δ_4' then Γ_0 is D_{2n}, E_8 respectively. It follows that the 8-dimensional even unimodular lattices E_8 and $2\Delta_8'$ both descend to $2E_8$, while L_{Δ_8} descends to $L_{D_{16}}$. By considering also the vectors of Q-length $2\sqrt{2}\varepsilon$, it was determined that L_{2D_4} descends to $2E_8$. We shall determine the algebraic descent of

$L_{4\Delta_2}$. In this case there are 384 vectors of Q-length $2\sqrt{2}\varepsilon$ in $L_{4\Delta_2}$ given by

$$M_{ij}=\{(*\,\varepsilon e_i * e_{i+1} * \varepsilon e_j * e_{j+1})/\sqrt{2}, (*\,\varepsilon e_i * e_{i+1} * e_j * \varepsilon e_{j+1})/\sqrt{2}\,,$$
$$(*\, e_i * \varepsilon e_{i+1} * \varepsilon e_j * e_{j+1})/\sqrt{2}, (*\, e_i * \varepsilon e_{i+1} * e_j * \varepsilon e_{j+1})/\sqrt{2}\}$$

where i,j are *odd* integers satisfying $1 \leqq i < j \leqq 7$ and $*$ denotes an arbitrary sign. Each pair of i,j gives a family of 64 vectors and there are 6 such families. The vectors in each family belong to the same irreducible component upon algebraic descent. Moreover for any two families M_{ij}, $M_{i'j'}$, there exist w, w' from M_{ij}, $M_{i'j'}$, respectively such that $B_0(w,w')\neq 0$. This shows that the descent of $L_{4\Delta_2}$ must be $L_{D_{16}}$. Indeed, the root system $4\Delta_2$ descends to $4D_4$ which accounts for the remaining 96 roots in D_{16}. Thus

Proposition 2. *The algebraic descent of $L_{4\Delta_2}$ is the 16-dimensional even unimodular lattice over $\mathbb{Z}$ with the root system D_{16}.*

Remark. By a similar argument, one can show that the algebraic descent of $L_\emptyset$ is also $L_{D_{16}}$.

Mass Formula

Let L be a positive definite integral R-lattice of rank m. The Minkowski-Siegel mass of the genus of L is given by

$$M(L)=\sum_{i=1}^{h}\frac{1}{e(L_i)}$$

where $\{L_1,\dots,L_h\}$ is a set of distinct representatives of the isometry classes in the genus of L, and $e(L_i)$ is the order of the orthogonal group $O(L_i)$ of L_i. If L is even unimodular, we have the following formula from [5] (see also [11])

$$M(L)=\frac{4^{1-m}L_F(m/2,\chi_m)\prod_{i=1}^{m/2-1}\zeta_F(2i)}{(\sqrt{8})^{\frac{-m(m-1)}{2}}\prod_{i=1}^{m}\pi^i\Gamma^{-2}\left(\frac{i}{2}\right)}$$

where $\chi_m(p)=\left(\frac{-1}{p}\right)^{m/2}$, $L_F(s,\chi_m)=\prod_p(1-\chi_m(p)N_p^{-s})^{-1}$ and $\zeta_F(\cdot)$ is the Dedekind zeta function.

Let M_m denote the mass for the genus of rank m. Then

$$M_4=\frac{4^{-3}(\zeta_F(2))^2}{(\sqrt{8})^{-6}\prod_{i=1}^{4}\pi^i\Gamma^{-2}\left(\frac{i}{2}\right)}=\frac{1}{2^8\cdot 3^2}$$

and

$$M_8=\frac{4^{-7}\zeta_F^2(4)\zeta_F(2)\zeta_F(6)}{(\sqrt{8})^{-28}\prod_{i=1}^{8}\pi^8\Gamma^{-2}\left(\frac{i}{2}\right)}=\frac{11^2\cdot 19^2}{2^{18}\cdot 3^5\cdot 5^2\cdot 7}\,.$$

In order to verify that our earlier enumeration for the 8-dimensional genus is complete, it is essential to compute the order of the automorphism group of each lattice. For those lattices with nonempty root system, we follow a method in [2]. First we decompose the root lattice in L into irreducible components

$$L_1 \perp \ldots \perp L_t \ .$$

Then we let $G_2(L)$ be the factor group of $O(L)$ by the normal subgroup $S(L)$ consisting of those elements which leave invariant all the L_i. Moreover, let $G_0(L)$ be the normal subgroup of $S(L)$ consisting of those elements which, for all i, act trivially on $L_i^{\#}/L_i$. Here $L_i^{\#}$ is the dual lattice of L_i. Finally, we let $G_1(L)$ be the factor group $S(L)/G_0(L)$. If we denote $g_k(L)=|G_k(L)|$ for $0 \leqq k \leqq 2$, then $e(L)=g_0 g_1 g_2$. With the help of a computer we have the following table

L_i	$g_0(L_i)$	$g_1(L_i)$	$g_2(L_i)$	$e(L_i)$
E_8	$2^{14}3^5 5^2 7$	1	1	$2^{14}3^5 5^2 7$
$2\Delta_4'$	$(2^8 \cdot 3^2)^2$	1	2	$2^{17} \cdot 3^4$
Δ_8	$2^{15} \cdot 3^2 \cdot 5 \cdot 7$	1	1	$2^{15} \cdot 3^2 \cdot 5 \cdot 7$
$2\Delta_4$	$(2^3 \cdot 4!)^2$	6	2	$2^{14} \cdot 3^3$
$4\Delta_2$	$(2^3)^4$	2^3	4!	$2^{18} \cdot 3$
$\emptyset$	—	—	—	$2^{14} \cdot 3^2 \cdot 5 \cdot 7$

The lattice $L_\emptyset$ has a base consisting of vectors of norm 4. The automorphism group of $L_\emptyset$ is computed by considering the permutations of its 3360 norm 4 vectors. Upon summing the reciprocals of the $e(L_i)$ we obtain

$$\sum_1^6 \frac{1}{e(L)} = \frac{11^2 \cdot 19^2}{2^{18} \cdot 3^2 \cdot 5^2 \cdot 7}$$

which is exactly the mass predicted by the mass formula. Thus we have:

Theorem. *There are precisely 6 distinct classes of even unimodular lattices of rank 8 over $\mathbb{Q}(\sqrt{2})$ which are distinguished by their root systems E_8, $2\Delta_4'$, Δ_8, $2D_4$, $4\Delta_2$ and $\emptyset$.*

Remark. In the quaternary case, the order of the automorphism group of Δ_4' is computed to be $2^8 \cdot 3^2$. Hence the mass formula verifies that Δ_4' is the unique class in that genus.

Theta Series

Let H be the upper half plane. A Hilbert modular form of weight k for the Hilbert modular group $SL_2(\mathbb{Z}[\varepsilon])$ is a holomorphic function f on H^2 satisfying the condition

$$f\left(\frac{az_1+b}{cz_1+d}, \frac{\bar{a}z_2+\bar{b}}{\bar{c}z_2+\bar{d}}\right) = (cz_1+d)^k(\bar{c}z_2+\bar{d})^k f(z_1, z_2)$$

for any matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}[\varepsilon])$. Here $\bar{a}$ is the conjugation of a. Every Hilbert modular form f has a Fourier expansion of the form

$$f(z)=\sum_{v \gg 0} c_f(v) e^{2\pi i\sigma(vz/2\sqrt{2}\varepsilon)} = \sum c_f(a+b\varepsilon)\,[a,b]\ ,$$

where $[a,b]=\exp\left[2\pi i\left(\left(\frac{a+b\varepsilon}{2\varepsilon\sqrt{2}}\right)z_1+\left(\overline{\frac{a+b\varepsilon}{2\varepsilon\sqrt{2}}}\right)z_2\right)\right]$. Let L be an even unimodular lattice over $\mathbb{Q}(\sqrt{2})$ of rank m. Then the theta series of L

$$\begin{aligned}\Theta_L(z) &= \sum_{x\in L} e^{2\pi i\sigma\left(\frac{Q(x)z}{2\varepsilon\sqrt{2}}\right)} \\ &= \sum_{v\gg 0} c_L(v) e^{2\pi i\sigma\left(\frac{vz}{2\varepsilon\sqrt{2}}\right)}\end{aligned}$$

is a Hilbert modular form of weight $\frac{m}{2}$. Here $c_L(v)=\#\{x\in L \mid Q(x)=2v\}$. If $L_1=L$, $L_2,\dots,L_h$ is a complete representative system of the distinct classes in the genus $\operatorname{gen}(L)$ of L, then Siegel's theorem [11] on the average number of representations of a number by $\operatorname{gen}(L)$ is given by

$$\frac{1}{M(L)}\sum_{i=1}^{h}\frac{\theta_{L_i}(z)}{e(L_i)}=G_{\frac{m}{2}}(z)\ , \tag{2}$$

where $G_{\frac{m}{2}}(z)=1+\sum c_{\frac{m}{2}}(v)e^{2\pi i\sigma\left(\frac{vz}{2\varepsilon\sqrt{2}}\right)}$ is the Eisentein series of weight $\frac{m}{2}$. From [4] we have

$$c_{\frac{m}{2}}(v)=b_{\frac{m}{2}}\sum_{(\beta)\mid(v)}(\operatorname{sign} N\beta)^{\frac{m}{2}}|N\beta|^{\frac{m}{2}-1}$$

and

$$b_{\frac{m}{2}}=\frac{(2\pi)^m\sqrt{8}}{\left(\Gamma\left(\frac{m}{2}\right)\right)^2 8^{\frac{m}{2}}\zeta_{\mathbb{Q}(\sqrt{2})}\left(\frac{m}{2}\right)}\ .$$

For $m=4$ and 8, we compute

$$b_2=\frac{(2\pi)^4\sqrt{8}}{8^2\zeta_{\mathbb{Q}(\sqrt{2})}(2)}=48$$

$$b_4=\frac{(2\pi)^8\sqrt{8}}{(3!)^2 8^4\zeta_{\mathbb{Q}(\sqrt{2})}(4)}=\frac{480}{11}\ .$$

Applying (2) to the genus of 8-dimensional even unimodular lattices, we have

$$\frac{1}{M_8}\times\sum_{i=1}^{h}\frac{c_{L_i}(1)}{e(L_i)}=b_4=\frac{480}{11}.$$

Using the 6 lattices in the genus and $M(L)=M_8$, we have

$$\frac{1}{M_8}\times\sum_{i=1}^{6}\frac{c_{L_i}(1)}{e(L_i)}=\frac{2^{18}\cdot3^2\cdot5^2\cdot7}{11^2\cdot19^2}$$

$$\cdot\left(\frac{2^4\cdot3\cdot5}{2^{14}\cdot3^5\cdot5^2\cdot7}+\frac{2^5\cdot3}{2^{17}\cdot3^4}+\frac{2^7}{2^{15}\cdot3^2\cdot5\cdot7}+\frac{2^4\cdot3}{2^{14}\cdot3^3}+\frac{2^5}{2^{18}\cdot3}\right)=\frac{480}{11}.$$

This calculation shows that the only lattices which admit vectors of quadratic norm 2 are exactly the five classes E_8, $2\Delta_4'$, L_{Δ_8}, L_{2D_4} and $L_{4\Delta_2}$. Since the uniqueness of $L_\emptyset$ has been established previously by the mass formula, it follows that there are precisely 6 distinct classes in the 8-dimensional even unimodular genus over $\mathbb{Q}(\sqrt{2})$.

Remark. For the quarternary genus of even unimodular lattices over $\mathbb{Q}(\sqrt{2})$, (2) gives

$$\frac{1}{M_4}\times\sum_{i=1}^{h}\frac{c_{L_i}(1)}{e(L_i)}=b_2=48.$$

Since $c_{\Delta_4'}(1)=48$ this shows that Δ_4' is the only class in the genus.

Neighbor Graph

We close this paper by constructing the neighbor graph of the genus of 8-dimensional even unimodular lattices over $\mathbb{Q}(\sqrt{2})$ at the prime $p=(\sqrt{2})$. For a lattice L in the genus, the vertices of the graph $R(L,p)$ are those lattices $M\in\operatorname{gen}L$ such that $M_q=L_q$ for all primes $q\neq p$. Two vertices M and M' are joined by an edge in $R(L,p)$ if $[M:M\cap M']=[M':M\cap M']=Np$. In this case, we say that M and M' are neighbors (or adjacent) in $R(L,p)$. $R(L,p)$ is a connected graph and it contains a representative of every class in the genus of L. For each lattice in the graph, the number of neighbors is the same as the number of isotropic lines in L/pL. A result in [1, p. 21] shows that if $\dim FL=2m$, where m is the Witt index of FL at p, then this number is given by

$$\frac{[(Np)^m-1][(Np)^{m-1}+1]}{Np-1}.$$

It follows that each lattice in $R(L,p)$ has exactly 135 neighbors. We will now present the graph. Since the number of neighbors is quite large, we will not give the details and just remark that they can be obtained by the neighbor method. Moreover, we will identify isometric lattices by their root systems.

Let $N(L,K,p)$ denote the number of neighbors of L that are isometric to K. We have the following table of $N(L,K,p)$ for the lattices in the 8-dimensional genus over $\mathbb{Q}(\sqrt{2})$:

$L \backslash K$	E_8	$2\Delta_4'$	Δ_8	$2D_4$	$4\Delta_2$	$\emptyset$
E_8	0	0	135	0	0	0
$2\Delta_4'$	0	18	36	0	81	0
Δ_8	2	35	28	70	0	0
$2D_4$	0	0	3	96	36	0
$4\Delta_2$	0	6	0	64	49	16
$\emptyset$	0	0	0	0	105	30

If L and K are neighbors, then it is shown in [1, p. 48] that

$$\frac{N(L,K,p)}{N(K,L,p)}=\frac{e(L)}{e(K)}\ . \tag{3}$$

If follows from (3) that the graph obtained above agrees with our earlier computations of the orders of automorphism groups, thus giving an additional check of the completeness of our list.

References

1. Benham, J.W.: Graphs, representations, and spinor genera. Thesis, Ohio State University (1981)
2. Conway, J.H., Sloane, N.J.A.: On the enumeration of lattices of determinant one. J. Number Theory **15**, 83–94 (1982)
3. Costello, P.J., Hsia, J.S.: Even unimodular 12-dimensional quadratic forms over $\mathbb{Q}(\sqrt{5})$. Adv. Math. **64**, 241–278 (1987)
4. Gundlach, K.B.: Die Bestimmung der Funktionen zu einigen Hilbertschen Modulgruppen. J Reine Angew. Math. **220**, 109–153 (1964)
5. Hsia, J.S.: Even positive definite unimodular quadratic forms over real quadratic fields. To appear in Rocky Mountain J. Math.
6. Kneser, M.: Klassenzahlen definiter quadratischer Formen. Arch. Math. **8**, 241–250 (1957)
7. Leech, J., Sloane, N.J.A.: Sphere packings and error-correcting codes. Can. J. Math. **23**, 718–745 (1971)
8. Mimura, Y.: On 2-lattices over real quadratic integers. Math. Sem. Notes (Kobe Univ.) **7**, 327–342 (1979)
9. Niemeier, H.-V.: Definite quadratische Formen der Dimensional 24 und Diskriminant 1. J. Number Theory **5**, 142–178 (1973)
10. O'Meara, O.T.: Introduction to quadratic forms. Berlin Heidelberg New York: Springer 1971
11. Siegel, C.L.: Über die analytische Theorie der quadratischen Formen III. Ann. Math. **38**, 212–291 (1937); Gesam. Abh. **II**, 469–548 (1966)
12. Takada, I.: On the classification of definite unimodular lattices over the ring of integers in $\mathbb{Q}(\sqrt{2})$. Math. Jap. **30**, 423–433 (1985)
13. Venkov, B.B.: On the classification of integral even unimodular 24-dimensional and quadratic forms. (Russian), Trudy Mat. Inst. Steklov **148**, 65–76 (1978); (English Translation), Proc. Steklov Inst. Math. **4**, 63–74 (1980)

Received October 12, 1987

Note added in proof. A theoretical proof of the uniqueness of the empty root lattice is also possible based on a method similar to that used in [3] for the $\mathbb{Q}(\sqrt{5})$ case.

Math. Ann. 283, 375–380 (1989)

Mathematische
Annalen

On the Gauss Map of Surfaces in R^n

Sun Cun-Jin

Department of Mathematics, Suzhou University, Suzhou, Jiangsu Province, People's Republic of China

Introduction

In [1], Hoffman and Osserman gave necessary and sufficient conditions **A** (2.20, 2.21) for a map $\mathbf{\Phi}$ from a Riemann surface S_0 to $\mathbf{C}^n\backslash\{0\}$ satisfying $\mathbf{\Phi}\cdot\mathbf{\Phi}=0$ to represent the Gauss map $\bar{\mathbf{G}}:\mathrm{S}_0\to\mathbf{Q}_{n-2}\subset\mathbf{CP}^{n-1}$ of a conformal immersion $\mathbf{X}:\mathrm{S}_0\to\mathbf{R}^n$.

In this paper we introduce necessary and sufficient conditions **B** [(6, 7) of Theorem 1], which are equivalent to **A** but are more explicit. Moreover, in the proof of Theorem 1, formula (13) shows that if **H** is the mean curvature vector of $\mathbf{X}(\mathrm{S}_0)\subset\mathbf{R}^n$, $(\log|\mathbf{H}|)_z$ can be expressed by a differential expression of order 2 in $\mathbf{\Phi}$, which generalizes a result of Kenmotsu [3, Theorem 3].

Moreover, Theorem 1 yields somewhat shorter proofs of the results of Hoffman-Osserman for the cases $n=3$, 4 [1, Sect. 3; 2].

In Theorem 2, we give a constructive proof which allows us to explicitly represent $\mathbf{X}(\mathrm{S}_0)$ by his Gauss map $\mathbf{\Phi}$, and to generalize in all dimensions the representation theorem of Kenmotsu [3, Theorem 4].

The Results

Let S_0 be a Riemann surface and the map

$$\mathbf{X}:\mathrm{S}_0\to\mathbf{R}^n \tag{1}$$

be a locally conformal immersion. If $z=\xi+i\eta$ is a local parameter on S_0, and $(x_1,\ldots,x_n)$ are coordinates in $\mathbf{R}^n$, then the map defining surface **S** is given locally in the form

$$\mathbf{X}(z),\qquad \mathbf{X}=(x_1,\ldots,x_n). \tag{2}$$

The Gauss map $\mathbf{G}:\mathrm{S}_0\to\mathbf{Q}_{n-2}$ is defined by

$$\mathbf{G}(z)=\left[\frac{\partial\mathbf{X}}{\partial\bar{z}}\right], \tag{3}$$

where $\mathbf{Q}_{n-2}=\{\mathbf{Z}\in\mathbf{CP}^{n-1}|\mathbf{Z}^2=0\}$ is the complex quadric in $\mathbf{CP}^{n-1}$. We may represent $\mathbf{G}$ locally in the form $[\mathbf{\Phi}]$ such that

$$\bar{\mathbf{G}}(z)=\left[\frac{\partial \mathbf{X}}{\partial z}\right]=[\mathbf{\Phi}(z)], \tag{4}$$

where $\mathbf{\Phi}(z)=(\varphi_1,\ldots,\varphi_n)\in\mathbf{C}^n\backslash\{0\}$ satisfies

$$\mathbf{\Phi}\cdot\mathbf{\Phi}=\sum_{k=1}^{n}\varphi_k^2=0. \tag{5}$$

Theorem 1. *Let* $\mathbf{S}$ *be an oriented surface in* $\mathbf{R}^n$ *given by* (1). *Let* $\mathbf{\Phi}$ *be the Gauss map of* $\mathbf{S}$ *in the sense of* (3, 4). *If the mean curvature vector* $\mathbf{H}$ *is not zero on* $\mathbf{S}$, *then the Gauss map* $\mathbf{\Phi}$ *must satisfy*

$$\mathbf{Im}\{(\bar{\mathbf{\Phi}}_z\cdot\bar{\mathbf{\Phi}}_z)^{1/2}(\mathbf{\Phi}_{\bar{z}}-\eta\mathbf{\Phi})\}=0, \tag{6}$$

$$\mathbf{Im}\left\{\left(\frac{\mathbf{\Phi}_{\bar{z}}\cdot\mathbf{\Phi}_{z\bar{z}}}{\mathbf{\Phi}_{\bar{z}}\cdot\mathbf{\Phi}_{\bar{z}}}-\frac{\mathbf{\Phi}_z\cdot\bar{\mathbf{\Phi}}}{|\mathbf{\Phi}|^2}\right)_{\bar{z}}\right\}=0, \tag{7}$$

where η *is defined by*

$$\eta=\frac{\mathbf{\Phi}_{\bar{z}}\cdot\bar{\mathbf{\Phi}}}{|\mathbf{\Phi}|^2}. \tag{8}$$

Proof. The formula (4) means

$$\frac{\partial \mathbf{X}}{\partial z}=\psi\mathbf{\Phi} \tag{9}$$

for some function $\psi:\mathbf{S}_0\to\mathbf{C}$. We note that the surface is regular wherever ψ does not vanish. By (2.17, 2.18) in [1], we know

$$|\mathbf{\Phi}|^2\bar{\psi}\mathbf{H}=\mathbf{\Phi}_{\bar{z}}-\eta\mathbf{\Phi}. \tag{10}$$

By taking the symmetric product of (10), we get

$$|\mathbf{\Phi}|^4\bar{\psi}^2|\mathbf{H}|^2=\mathbf{\Phi}_{\bar{z}}\cdot\mathbf{\Phi}_{\bar{z}}. \tag{11}$$

From (10) and (11), we obtain

$$(\bar{\mathbf{\Phi}}_z\cdot\bar{\mathbf{\Phi}}_z)^{1/2}(\mathbf{\Phi}_{\bar{z}}-\eta\mathbf{\Phi})=\pm|\mathbf{\Phi}|^4|\mathbf{H}|\,|\psi|^2\mathbf{H}.$$

Thus (6) holds. This concludes the proof of the first part of the theorem.

By (2.19) in [1], we know

$$(\log\psi)_{\bar{z}}+\eta=0.$$

Thus we have

$$(\log\bar{\psi}^2)_z+2\bar{\eta}=0. \tag{12}$$

By (11), we know that $\mathbf{H}\neq 0$ is equivalent to $\mathbf{\Phi}_{\bar{z}}\cdot\mathbf{\Phi}_{\bar{z}}\neq 0$. From (11, 12), we have

$$\left(\log\frac{\mathbf{\Phi}_{\bar{z}}\cdot\mathbf{\Phi}_{\bar{z}}}{|\mathbf{\Phi}|^4|\mathbf{H}|^2}\right)_z+2\bar{\eta}=0.$$

Thus we get

$$(\log|\mathbf{H}|^2)_z = (\log(\mathbf{\Phi}_{\bar z}\cdot\mathbf{\Phi}_{\bar z}))_z - (\log|\mathbf{\Phi}|^4)_z + 2\bar\eta .$$

Using the above formula and noting (8), we obtain

$$(\log|\mathbf{H}|)_z = \frac{\mathbf{\Phi}_{\bar z}\cdot\mathbf{\Phi}_{z\bar z}}{\mathbf{\Phi}_{\bar z}\cdot\mathbf{\Phi}_{\bar z}} - \frac{\mathbf{\Phi}_z\cdot\bar{\mathbf{\Phi}}}{|\mathbf{\Phi}|^2}. \tag{13}$$

Since $|\mathbf{H}|$ is real, (7) holds. This completes the proof of the theorem.

Remark 1. We can prove that conditions (2.20, 2.21) in [1] (designated by **A**) are equivalent to conditions (6, 7) (designated by **B**). In fact, the proof that **B** $\Rightarrow$ **A** is obvious, and the proof that **A** $\Rightarrow$ **B** is the proof of Theorem 1.

Remark 2. We will see that Theorem 1 yields somewhat shorter proofs of the results of Hoffman-Osserman for the cases $n=3$, 4 [1, Sect. 3; 2].

When **S** is the surface in $\mathbf{R}^4$, $\mathbf{\Phi}(z)$ is determined by (3.4) in [1]. By (3.15) in [1], we obtain

$$\mathbf{\Phi}_{\bar z}\cdot\mathbf{\Phi}_{\bar z} = \mathbf{V}\cdot\mathbf{V} = (F_1\mathbf{A} - F_2\bar{\mathbf{A}})\cdot(F_1\mathbf{A} - F_2\bar{\mathbf{A}}) = -2F_1F_2|\mathbf{A}|^2, \tag{14}$$

where $\mathbf{V} = \mathbf{\Phi}_{\bar z} - \eta\mathbf{\Phi}$, F_i is defined by (3.5) in [1], **A** is defined by (3.14) in [1]. By (14), we know that $\mathbf{H} \neq 0$ is equivalent to $F_1F_2 \neq 0$. Using (3.15) in [1] and (14), we find

$$\begin{aligned}(\bar{\mathbf{\Phi}}_z\cdot\bar{\mathbf{\Phi}}_z)^{1/2}(\mathbf{\Phi}_{\bar z} - \eta\mathbf{\Phi}) &= \pm i\sqrt{2}|\mathbf{A}|(\bar F_1\bar F_2)^{1/2}(F_1\mathbf{A} - F_2\bar{\mathbf{A}})\\ &= \pm i\sqrt{2}|\mathbf{A}|(|F_1|(F_1\bar F_2)^{1/2}\mathbf{A} - |F_2|(\bar F_1F_2)^{1/2}\bar{\mathbf{A}}).\end{aligned}$$

Thus we get

$$\mathbf{Im}\{(\bar{\mathbf{\Phi}}_z\bar{\mathbf{\Phi}}_z)^{1/2}(\mathbf{\Phi}_{\bar z} - \eta\mathbf{\Phi})\} = \pm\sqrt{2}|\mathbf{A}|(|F_1| - |F_2|)\,\mathbf{Re}\{(F_1\bar F_2)^{1/2}\mathbf{A}\}. \tag{15}$$

We will show that

$$\mathbf{Re}\{(F_1\bar F_2)^{1/2}\mathbf{A}\} \neq 0. \tag{16}$$

Assume (16) fails at some point $z_0 \in \mathbf{S}_0$. Then

$$\mathbf{Re}\{(F_1(z_0)\bar F_2(z_0))^{1/2}\}\,\mathbf{Re}\{\mathbf{A}(z_0)\} - \mathbf{Im}\{(F_1(z_0)\bar F_2(z_0))^{1/2}\}\,\mathbf{Im}\{\mathbf{A}(z_0)\} = 0,$$

but

$$(F_1(z_0)\bar F_2(z_0))^{1/2} \neq 0, \qquad \text{(because of } F_1(z_0)F_2(z_0) \neq 0).$$

Thus we have

$$\mathbf{Re}\{\mathbf{A}(z_0)\} = \lambda\,\mathbf{Im}\{\mathbf{A}(z_0)\} \qquad (\text{or } \mathbf{Im}\{\mathbf{A}(z_0)\} = \mu\,\mathbf{Re}\{\mathbf{A}(z_0)\}).$$

Hence

$$\begin{gathered}\mathbf{A}(z_0) = \mathbf{Re}\{\mathbf{A}(z_0)\} + i\,\mathbf{Im}\{\mathbf{A}(z_0)\} = (\lambda + i)\,\mathbf{Im}\{\mathbf{A}(z_0)\}\\ (\text{or } \mathbf{A}(z_0) = (1 + i\mu)\,\mathbf{Re}\{\mathbf{A}(z_0)\}).\end{gathered}$$

By (3.14) in [1], we know that $\mathbf{A}^2 \equiv 0$, thus $\mathbf{Im}\{\mathbf{A}(z_0)\} = \mathbf{Re}\{\mathbf{A}(z_0)\} = 0$. This clearly contradicts (3.16) in [1]. Thus (16) holds. By (15, 16), we know that for a surface in $\mathbf{R}^4$ condition (6) becomes (3.7) in [1].

By (3.5, 3.16) in [1] and (14), we get

$$\Phi_{\bar z}\cdot\Phi_{\bar z}=-4(f_1)_{\bar z}(f_2)_{\bar z}, \tag{17}$$

which implies that $\mathbf{H}\neq 0$ is equivalent to $(f_1)_{\bar z}(f_2)_{\bar z}\neq 0$. Thus we find

$$\begin{aligned}\frac{\Phi_{\bar z}\cdot\Phi_{z\bar z}}{\Phi_{\bar z}\cdot\Phi_{\bar z}}&=\frac{1}{2}(\log(\Phi_{\bar z}\cdot\Phi_{\bar z}))_z=\frac{1}{2}\{(\log(f_1)_{\bar z})_z+(\log(f_2)_{\bar z})_z\}\\&=\frac{1}{2}\left\{\frac{(f_1)_{z\bar z}}{(f_1)_{\bar z}}+\frac{(f_2)_{z\bar z}}{(f_2)_{\bar z}}\right\}.\end{aligned} \tag{18}$$

Using (3.4, 3.12) in [1], we find

$$\frac{\Phi_z\cdot\bar\Phi}{|\Phi|^2}=\frac{\bar f_1(f_1)_z}{1+|f_1|^2}+\frac{\bar f_2(f_2)_z}{1+|f_2|^2}. \tag{19}$$

By (18, 19), we know that for surfaces in $\mathbf{R}^4$ condition (7) becomes (3.8) in [1].

When $\mathbf{S}$ is a surface in $\mathbf{R}^3$, $\Phi(z)$ is determined by (3.6) in [2]. Thus we find

$$\Phi_{\bar z}\cdot\Phi_{\bar z}=4(f_{\bar z})^2, \tag{20}$$

and

$$\frac{\Phi_z\cdot\bar\Phi}{|\Phi|^2}=\frac{2\bar f f_z}{1+|f|^2}. \tag{21}$$

By (3.9) in [2] and (20), we obtain

$$(\bar\Phi_z\cdot\bar\Phi_z)^{1/2}(\Phi_{\bar z}-\eta\Phi)=\pm 4|f_{\bar z}|^2\mathbf{N}. \tag{22}$$

Since $\mathbf{N}$ is a real vector, condition (6) is always satisfied for surfaces in $\mathbf{R}^3$. By (20), we know that $\mathbf{H}\neq 0$ is equivalent to $f_{\bar z}\neq 0$. Thus we have

$$\frac{\Phi_{\bar z}\cdot\Phi_{z\bar z}}{\Phi_{\bar z}\cdot\Phi_{\bar z}}=\frac{1}{2}(\log(\Phi_{\bar z}\cdot\Phi_{\bar z}))_z=\frac{f_{z\bar z}}{f_{\bar z}}. \tag{23}$$

By (21, 23), we know that for surfaces in $\mathbf{R}^3$ condition (7) becomes

$$\mathbf{Im}\left\{\left(\frac{f_{z\bar z}}{f_{\bar z}}-\frac{2\bar f f_z}{1+|f|^2}\right)_{\bar z}\right\}=0. \tag{24}$$

Remark 3. The Eq. (13) obtained in the proof of Theorem 1 generalizes Kenmotsu's result [3, Theorem 3].

By (18, 19), we know that for surfaces in $\mathbf{R}^4$ the Eq. (13) becomes

$$(\log|\mathbf{H}|^2)_z=\mathbf{S}(f_1)+\mathbf{S}(f_2), \tag{25}$$

where

$$\mathbf{S}(f_k)=\frac{(f_k)_{z\bar z}}{(f_k)_{\bar z}}-\frac{2\bar f_k(f_k)_z}{1+|f_k|^2},\qquad k=1,2. \tag{26}$$

By (21, 23), we know that for surfaces in $\mathbf{R}^3$ the Eq. (13) becomes

$$(\log|\mathbf{H}|)_z=\frac{f_{z\bar z}}{f_{\bar z}}-\frac{2\bar f f_z}{1+|f|^2}. \tag{27}$$

This is Kenmotsu's result [3, Theorem 3].

Theorem 2. *Let* $\mathbf{S}_0$ *be a simply connected Riemann surface, and* $\mathbf{G}:\mathbf{S}_0\to\mathbf{Q}_{n-2}$ *be a map into the complex quadric. Represent* $\mathbf{G}$ *locally by a map* $\boldsymbol{\Phi}$ *into* $\mathbf{C}^n\setminus\{0\}$ *in the sense that* $\bar{\mathbf{G}}=[\boldsymbol{\Phi}]$. *Define* η *in terms of* $\boldsymbol{\Phi}$ *by* (8). *If* $\mathbf{V}=\boldsymbol{\Phi}_{\bar z}-\eta\boldsymbol{\Phi}$ *is not zero on* $\mathbf{S}_0$, *then there exists a conformal immersion* $\mathbf{X}:\mathbf{S}_0\to\mathbf{R}^n$ *with Gauss map* $\mathbf{G}$ *if and only if* $\boldsymbol{\Phi}$ *satisfies* (6, 7).

Proof. If there exists a conformal immersion $\mathbf{X}:\mathbf{S}_0\to\mathbf{R}^n$ with Gauss map $\mathbf{G}$, we know that (6, 7) hold by Theorem 1.

Conversely, if $\boldsymbol{\Phi}(z)$ satisfies (6, 7), first we put

$$T(z)=\frac{\boldsymbol{\Phi}_{\bar z}\cdot\boldsymbol{\Phi}_{z\bar z}}{\boldsymbol{\Phi}_{\bar z}\cdot\boldsymbol{\Phi}_{\bar z}}-\frac{\boldsymbol{\Phi}_z\cdot\bar{\boldsymbol{\Phi}}}{|\boldsymbol{\Phi}|^2}, \tag{28}$$

thus we have $\mathbf{Im}\{T_{\bar z}\}=0$. Using Lemma 2.1 in [1], we have

$$h=\exp\int\mathbf{Re}\{2T(z)dz\}, \tag{29}$$

where h satisfies

$$(\log h)_z=T(z). \tag{30}$$

Second, we put

$$\psi(z)=\frac{(\bar{\boldsymbol{\Phi}}_z\cdot\bar{\boldsymbol{\Phi}}_z)^{1/2}}{|\boldsymbol{\Phi}|^2 h}. \tag{31}$$

By (28), (30), and (31), we have

$$(\log\bar\psi)_z=\frac{1}{2}(\log(\boldsymbol{\Phi}_{\bar z}\cdot\boldsymbol{\Phi}_{\bar z}))_z-(\log h)_z-(\log(\boldsymbol{\Phi}\cdot\bar{\boldsymbol{\Phi}}))_z=-\frac{\boldsymbol{\Phi}\cdot\bar{\boldsymbol{\Phi}}_z}{|\boldsymbol{\Phi}|^2}.$$

Thus we get

$$\frac{\psi_{\bar z}}{\psi}+\eta=0. \tag{32}$$

Using (31, 32) and condition (6), we obtain

$$\begin{aligned}\mathbf{Im}(\psi\boldsymbol{\Phi})_{\bar z}&=\mathbf{Im}(\psi_{\bar z}\boldsymbol{\Phi}+\psi\boldsymbol{\Phi}_{\bar z})=\mathbf{Im}\{\psi(\boldsymbol{\Phi}_{\bar z}-\eta\boldsymbol{\Phi})\}\\&=\frac{1}{|\boldsymbol{\Phi}|^2 h}\,\mathbf{Im}\{(\bar{\boldsymbol{\Phi}}_z\cdot\bar{\boldsymbol{\Phi}}_z)^{1/2}(\boldsymbol{\Phi}_{\bar z}-\eta\boldsymbol{\Phi})\}=0.\end{aligned}$$

By the proof of Theorem 2.6 in [1], we know that there is a surface $\mathbf{X}:\mathbf{S}_0\to\mathbf{R}^n$ such that

$$\frac{\partial\mathbf{X}}{\partial z}=\psi(z)\boldsymbol{\Phi}(z). \tag{33}$$

Clearly, $\mathbf{X}(\mathbf{S}_0)$ is a conformal immersed surface with Gauss map $\mathbf{G}$.

Remark 4. We may write (33) in the form

$$\mathbf{X}=\int\mathbf{Re}\{2\psi\boldsymbol{\Phi}dz\}, \tag{34}$$

or

$$\mathbf{X} = \int \mathbf{Re}\left\{\frac{2(\bar{\mathbf{\Phi}}_z \cdot \bar{\mathbf{\Phi}}_z)^{1/2}\mathbf{\Phi}}{|\mathbf{\Phi}|^2 h}\,dz\right\}, \tag{35}$$

where h is defined by (29). By (11, 31), we know that mean curvature vector $\mathbf{H}$ of $\mathbf{X}(\mathbf{S}_0)$ satisfies $|\mathbf{H}| = h$.

By Theorem 2.5 in [1], we know that if $\mathbf{X}: \mathbf{S}_0 \to \mathbf{R}^n$ is a conformal immersed surface with Gauss map $\mathbf{G}$, and the mean curvature vector $\mathbf{H}$ is not zero on $\mathbf{X}(\mathbf{S}_0)$, then $\mathbf{X}(\mathbf{S}_0)$ is determined uniquely by $\mathbf{G}$, up to translation and homothety. And $\mathbf{X}(\mathbf{S}_0)$ can be explicitly given in terms of $\mathbf{\Phi}$ by (35). Furthermore, if h is the scalar mean curvature of $\mathbf{X}(\mathbf{S}_0)$, namely, $|\mathbf{H}| = h$, then $\mathbf{X}(\mathbf{S}_0)$ can be described explicitly from h and $\mathbf{\Phi}$ by (35).

Therefore formula (35) generalizes the representation theorem of Kenmotsu [3, Theorem 4] to euclidean n-space.

References

1. Hoffman, D.A., Osserman, R.: The Gauss map of surfaces in $\mathbf{R}^n$, J. Differ. Geom. **18**, 733–754 (1983)
2. Hoffman, D.A., Osserman, R.: The Gauss map of surfaces in $\mathbf{R}^3$ and $\mathbf{R}^4$. Proc. Lond. Math. Soc. **50**, 27–56 (1985)
3. Kenmotsu, K.: Weierstrass formula for surfaces of prescribed mean curvature. Math. Ann. **245**, 89–99 (1979)
4. Kenmotsu, K.: The mean curvature vector of surfaces in $\mathbf{R}^4$. Preprint (1986)

Received June 30, 1987; in revised form February 12, 1988

Math. Ann. 283, 381–409 (1989)

Mathematische Annalen

Finite Groups and Hecke Operators

Geoffrey Mason*,**

Department of Mathematics, University of California, Division of Natural Sciences, 379 Applied Sciences, Santa Cruz, CA 95064, USA

1. Introduction

One of the residual mysteries of the classification of the finite simple group concerns the connections between the Monster group M and certain genus zero function fields associated to elliptic modular functions. These relationships were developed by Conway-Norton, Thompson et al. at the end of the last decade, and concerned something which is now called a *Thompson series*; thus it is conjectured that there is a sequence $\gamma_{-1}=1_M, \gamma_1, \gamma_2, \ldots$ of characters of the group M such that the formal power series

$$\Gamma_M = \sum \gamma_n q^n \tag{1.1}$$

has (among others) the following properties: (i) if q is interpreted as $e^{2\pi i z}$ ($z \in \mathfrak{h}$ = upper halfplane) then $\sum \gamma_n(1) q^n$ is the modular function j-744 (here 1 is of course the identity of M); (ii) for each $g \in M$ the q-expansion $\sum \gamma_n(g) q^n$ is that of a hauptfunktion of some level $N=N(g)$ divisible by the order $o(g)$ of g. More precisely if $f=f_g=\sum \gamma_n(g) q^n$ then it is conjectured that the invariance group G of f in $SL_2(R)$ contains $\Gamma_0(N)$ as a normal subgroup, that the compactified Riemann surface $(\mathfrak{h}/G)^*$ is a sphere, and that f generates the field of modular functions on $(\mathfrak{h}/G)^*$).

Despite the overwhelming evidence for the truth of this conjecture, we seem as far today from understanding it as ever. One of the difficulties which makes the conjecture so mysterious and compelling is the apparent disparity of data which must be reconciled. Almost as remarkable is that the only real progress to date has been achieved by Frenkel et al. [FLM] using methods of Kac-Moody Lie algebras, adding another ingredient to the brew.

The point of view which we adopt in the present paper is foundational. We attempt to develop the beginnings of a theory of q-expansions of the type in (1.1) for

* Research supported in part by a grant from the National Science Foundation

** Part of this research was also carried out during the author's stay at the Max Planck Inst. für Mathematik, Bonn, in 1983

an *arbitrary* finite group. To paraphrase a remark of Langlands made elsewhere but which applies equally well here, if the problem does not fall to a series of vigorous assaults then we must prepare for a long siege.

As soon as a more general point of view is adopted, it becomes natural to consider formal q-expansions (1.1) where not only are the γ_n allowed to be *generalized* (or virtual) characters of the finite group G, but also if $g \in G$ then the q-expansion $\sum \gamma_n(g) q^n$ is the Fourier expansion of a modular form of some weight k, level N and character ε and depending on g (more precisely on the conjugacy class of G determined by g). Although there is no need to so restrict ourselves, we assume throughout this paper that the following additional conditions hold: that for each $g \in G$ the q-expansion $\sum \gamma_n(g) q^n$ is a modular form on $\Gamma_0(N)$ for some N depending on g and that this q-expansion has rational integer coefficients. Thus we arrive at the

Definition. Let G be a finite group. The formal q-expansion

$$\Gamma = \sum \gamma_n q^n$$

is called a *Thompson series* (for G) if each γ_n is a rational-valued, generalized character of G and if for each $g \in G$ the q-expansion

$$\Gamma_g = \sum \gamma_n(g) q^n$$

is that of a modular form on $\Gamma_0(N)$ for some $N = N(g)$, integral weight $k(g)$ and character ε_g.

Generally, we will be concerned with Thompson series where either all $k(g) = 0$ (the original situation pertaining to the Monster), or at least each Γ_g is meromorphic. There are also important examples where each Γ_g is holomorphic, etc., but we will usually not dwell on the various situations that may arise concerning the analytic properties of the individual forms.

One can now ask an apparently naive question: can one extend (some of) the elementary theory elliptic modular forms to the context of Thompson series? In this paper we are concerned in particular with the possibility of defining Hecke operators for Thompson series. That the answer, in at least some situations, is affirmative is part of the theory for the sporadic group M_{24} as developed in [M3]. There we constructed a Thompson series for M_{24} whose corresponding L-series $\sum_{n \geq 1} \gamma_n / n^s$ has an Euler product. In this paper we show that the existence of this Thompson series is just part of a general theory of Hecke operators for Thompson series.

What is perhaps surprising is that the Hecke operators we construct are ultimately related to the theory of the (oriented) Bott cannibalistic class, regarded in this instance as a certain virtual character of a spin group. It is clear that this circumstance is just part of a much wider topological context, which however we will not pursue here.*

A brief summary of the paper is as follows: in Sect. 2 we define some important Thompson series and present certain spaces $\mathfrak{M}(\varrho)$ of Thompson series which will be our analogues of spaces of modular forms with a given weight and character, and

* Such a context seems to be provided by elliptic cohomology

on which our generalized Hecke operators will operate. In Sect. 3 we study the Bott cannabilistic class, and in Sect. 4 show how to define the Hecke operators themselves. In Sects. 5 and 6 we give some illustrations of how one can use the Hecke operators to construct (i) Thompson series which are simultaneous eigenforms and which include the M_{24} example alluded to above as a special case; (ii) Thompson series analogues of Eisenstein series; (ii) Thompson series for which the identity Γ_1 is of the form $j(q)+$constant. In an appendix we list the η-functions associated to the Conway group $\cdot O$ and which play a rôle in several places.

It remains only to record thanks to those several individuals who have contributed, in one way or another, to the results contained herein. It was Oliver Atkin who computed the q-expansions of the forms listed in the appendix and thereby provoked the author into thinking about Hecke operators, while Marvin Knopp pointed out the existence of the paper [Ba] and Michel Broué supplied a copy of his work [Br]. Finally, it is a pleasure to thank Professor F. Hirzebruch and the Max Planck Institut in Bonn for their hospitality during 1983 and for giving me a chance to think about modular forms for so many uninterrupted hours.

2. The Spaces $\mathfrak{M}(\varrho)$

In this section we establish some notation, recall some earlier results of relevance, and introduce the spaces $\mathfrak{M}(\varrho)$ of Thompson series on which our Hecke operators will act.

Our results will depend on a rational representation of the finite group G:

(2.1) V is an even-dimensional Q-vector space and

$$\varrho: G \to SL(V)$$

a representation of G by unimodular matrices.

In this situation we denote the characteristic polynomial of $\varrho(g)$ by $\chi^g(t)$. As ϱ is a rational representation one knows (cf. [M2]) that

$$\chi^g(t) = \prod_{i \geqq 1} (t^i - 1)^{k(i)} \tag{2.2}$$

for certain integers $k(i)$ which are zero unless i divides the order $o(g)$ of $\varrho(g)$.

Given (2.2), we also set

$$\begin{aligned} d(g) &= \prod i^{k(i)}, \\ k(g) &= \tfrac{1}{2}\sum k(i), \\ \Delta(g) &= (-1)^{k(g)} d(g). \end{aligned} \tag{2.3}$$

Note that $k(g)=\frac{1}{2}\dim V^g$ is an integer where V^g is the subspace of g-invariants of V, so that we can define the Dirichlet character ε_g via

$$\varepsilon_g(p) = \left(\frac{\Delta(g)}{p}\right) \tag{2.4}$$

for odd primes p.

As usual we denote by η the Dedekind eta-function

$$\eta(z) = q^{1/24} \prod_{n=1}^{\infty} (1-q^n),$$

where $q=e^{2\pi iz}$, z in the upper half-plane. If g is as in (2.2), set

(2.5)
$$\eta_g(z)=\prod_{i\geqq 1}\eta(iz)^{k(i)}$$

and define functions ω_n on G, $n=1,2,\dots$ by

(2.6)
$$\Omega_G=\sum_{n\geqq 1}\omega_n q^n,$$
$$\Omega_g=\sum_{n\geqq 1}\omega_n(g)q^n=\eta_g(z).$$

From the results of [M2], for example, we know that Ω_G is a Thompson series for G. Furthermore, one knows that $\eta_g(z)$ has weight $k(g)$ and Dirichlet character induced from ε_g. For this and a discussion of the level, see Sect. 4 of [M5] or Proposition 3.2 of [Br] for example.

An important situation with which we will be interested concerns the case where G is a suitable group of isometries of an even-dimensional even lattice $\mathbb{L}$. Thus $\mathbb{L}$ is a free abelian group with a positive-definite, G-invariant, symmetric, integral inner produce $(\,,\,)$, which is *even*, i.e., $(x,x)\in 2\mathbb{Z}$ for $x\in\mathbb{L}$. This situation was first investigated by Thompson [T], where the following Thompson series was introduced:

(2.7)
$$\Theta_G=\sum_{n\geqq 0}\alpha_n q^n,\qquad \alpha_n\in RG,$$
$$\Theta_g=\sum_{n\geqq 0}\alpha_n(g)q^n=\theta_{\mathbb{L}^g}(z),$$

where $\theta_{\mathbb{L}^g}(z)$ is the theta-function of the lattice $\mathbb{L}^g$ of g-invariants in $\mathbb{L}$. In [M5] it is shown that

(2.8) If $\mathbb{L}$ is unimodular and $g\in G$ acts on $\mathbb{L}$ with determinant 1, and on $\mathbb{L}\otimes_{\mathbb{Z}}Q$ with spinor norm 1, then Θ_g has Dirichlet character ε_g.

Now fix a representation ϱ of G as in (2.1). We define

(2.9) $\mathfrak{M}(\varrho)=$ complex vector space spanned by those Thompson series Γ_G of G such that for $g\in G$, $\Gamma_g(z)$ is a form of weight $k(g)$ and Dirichlet character induced from ε_g.

Thus we do not specify the precise level of the forms Γ_g. In any case from the preceding we have

Theorem 2. *$\Omega_G\in\mathcal{M}(\varrho)$. Moreover if $\mathbb{L}$ is a unimodular lattice and ϱ is a representation of G as isometries of $\mathbb{L}$ satisfying* (2.8) *then $\Theta_G\in\mathfrak{M}(\varrho)$.*

There are variations on $\mathfrak{M}(\varrho)$; one that we make use of is $\mathfrak{S}(\varrho)$ – those Thompson series Γ_G in $\mathfrak{M}(\varrho)$ for which each $\Gamma_g(z)$ is holomorphic and a cusp-form.

We remark finally that whereas each of the theta-series Θ_g of (2.7) are holomorphic forms, the same is not necessarily the case with the forms η_g. From Sects. 4, 5 of [Ba] we have

(2.10) If $\eta(z)$ is given by (2.5) and has level N, its order at the cusp r/s is

$$\frac{N}{24(N,s^2)}\sum_{i\geqq 1}\frac{k(i)}{i}(s,i)^2.$$

3. On the Bott Cannibalistic Class

As preparation for our construction of Hecke operators in Sect. 4, we study in this section certain generalized characters of the real orthogonal groups $O(n, R)$ and their universal covering groups $\mathrm{Spin}(n, R)$.

Recall first that the appropriate Clifford algebras and their representations afford so-called spin modules for the spin groups. Using the notation of Chap. 13 of [Hu], $\mathrm{Spin}(2r+1)$ has a (complex) spin module $\Delta(r)=\Delta$ of dimension 2^r, whilst the corresponding module for $\mathrm{Spin}(2r)$ decomposes into the sum of the two half-spin modules $\Delta^+(r)$, $\Delta^-(r)$ each of dimension 2^{r-1}. Moreover, with respect to the canonical embedding $\mathrm{Spin}\,(2r-1)\to\mathrm{Spin}(2r)$, the restriction of both $\Delta^+(r)$ and $\Delta^-(r)$ to $\mathrm{Spin}(2r-1)$ are isomorphic to the spin module $\Delta(r-1)$.

There is a rather more recondite analogue of these constructions where, roughly speaking, 2 is replaced by some odd prime p, or more generally by an odd integer k. This involves the so-called Bott cannibalistic classes. We will not be concerned in this paper with any of the topological aspects of these characteristic classes, only with some of their formal properties. We refer the reader to the appropriate sections of [Hu] or [Bo] for the general theory; the more formal theory may be found, for example, in [Td] or [AT].

Thus first let k be any positive integer. Then the Bott cannibalistic class θ_k operates on special λ-rings, in particular if E is the natural n-dimensional module for $O(n, R)$ then $\theta_k(E)$ is a certain (generalized) module for $O(n, R)$ of dimension k^n. If k is odd then there is a refinement: namely there is the so-called oriented Bott class θ_k^{or} which operates on oriented λ-rings. In particular $\theta_k^{\mathrm{or}}(E)$ exists if $\dim E$ is even and we have

(3.1) $$\theta_k^{\mathrm{or}}(E)^2=\theta_k(E), \qquad k \text{ odd}.$$

We note here that $\theta_k^{\mathrm{or}}(E)$ is a (generalized) module for $O(2n, R)$ rather than its covering group.

(3.2) **Lemma** *There is a generalized module F for $O(2n-1, R)$ such that restriction yields an isomorphism of $O(2n-1, R)$-modules*

$$\theta_k^{\mathrm{or}}(E)\cong kF.$$

Proof. Write $E=E_0\oplus T$ where $\dim T=2$. Then the fact that θ_k^{or} is an exponential map (on spaces of even dimension) yields an isomorphism

$$\theta_k^{\mathrm{or}}(E)=\theta_k^{\mathrm{or}}(E_0)\theta_k^{\mathrm{or}}(T)$$

of $O(2n-2, R)$-modules. Moreover $\theta_k^{\mathrm{or}}(T)$ is the trivial module of dimension k, so that

(3.3) $$\theta_k^{\mathrm{or}}(E)=k\theta_k^{\mathrm{or}}(E_0)$$

as $O(2n-2, R)$-modules.

Now restriction induces maps of character rings

$$R\,\mathrm{Spin}(2n)\xrightarrow{f}R\,\mathrm{Spin}(2n-1)\xrightarrow{g}R\,\mathrm{Spin}(2n-2)$$

and g is an injection (Proposition 13.13 of [Hu]). Moreover $R\,\mathrm{Spin}(2n-2)$ is a free $R\,\mathrm{Spin}(2n-1)$ module with generators 1, $\Delta^+(n-1)$ (loc. cit.). But it is clear from (3.3) that $k\theta_k^{\mathrm{or}}(E_0)$ belongs to $\operatorname{im} g$, so the freeness of $R\,\mathrm{Spin}(2n-2)$ means that

$\theta_k^{\text{or}}(E)$ is itself the restriction of a generalized Spin$(2n-1)$-module, call it F. The lemma follows immediately.

Notation

1. β_k is the generalized character of $O(n, R)$ afforded by $\theta_k(E)$. It has degree k^n.
2. ψ_2 is the character of Spin$(2r, R)$ afforded by the half-spin module $\Delta^+(r)$, while for odd k we let ψ_k be the generalized character of $O(2r-1, R)$ afforded by the module F of Lemma 3.2. We sometimes also use ψ_2 for the restriction to Spin$(2r-1, R)$; ψ_k has degree k^{r-1}.
3. β_2^{or} is the character of Spin(n, R) afforded by the spin module $\Delta(r)$, $n=2r+1$ or $2r$; for k odd, β_k^{or} is the generalized character of $O(2r, R)$ afforded by $\theta_k^{\text{or}}(E)$. So β_k^{or} has degree k^r.

By (3.1), Lemma 3.2 we get

(3.3) (a) $k\psi_k = \beta_k^{\text{or}}$ *as generalized characters of* Spin$(2r-1, R)$.

(b) $$\beta_k^{\text{or}})^2 = \begin{cases} \beta_k & \text{if } \dim E = 2r, \\ 2\beta_k & \text{if } \dim E = 2r+1 \text{ (and } k=2). \end{cases}$$

There are explicit formulae for these characters. Thus we have

$$\beta_k(g) = \prod (1 + \mu + \ldots + \mu^{k-1}), \tag{3.4}$$

where the product runs over the eigenvalues μ of g (with multiplicity). Furthermore for odd k we have

$$\beta_k^{\text{or}}(g) = \prod_{*} (\mu^{k-1/2} + \mu^{k-3/2} + \ldots + 1 + \ldots + \mu^{-(k-1)/2}). \tag{3.5}$$

Here, $*$ means that the product runs over a set U of eigenvalues of g defined as follows: pair the eigenvalues of g into 2-element sets $\{\mu, \mu^{-1}\}$, and choose U to contain exactly one element of each pair.

There is an alternate description as follows. Since g lies in a maximal torus of $O(n, R)$, $n=2r+1$ or $2r$, then there are real numbers $t_1, t_2, \ldots, t_r$, $0 \leqq t_i < 1$, such that g is conjugate to the matrix

$$\operatorname{diag}\left(\ldots, \begin{pmatrix} \cos 2\pi t_j, & -\sin 2\pi t_j \\ \sin 2\pi t_j, & \cos 2\pi t_j \end{pmatrix}, \ldots\right). \tag{3.6}$$

Then from (3.3) we deduce that for odd k,

$$\beta_k^{\text{or}}(g) = k^{1/2 \dim(E^g)} \prod_{t_j > 0} \frac{\sin k\pi t_j}{\sin \pi t_j}. \tag{3.7}$$

Here, the product runs over those t_j greater than 0, while E^g = the subspace of g-invariants of E, which has dimension equal to twice the number of t_j equal to 0.

The analogous formula for $k=2$ essentially involves replacing k by 2 and replacing the odd function $\sin x$ by the even function $\cos x$. Hence

(3.8) Let $g \in$ Spin(n, R), $n=2r+1$ or $2r$. Then

$$\beta_2^{\text{or}}(g) = 2^r \prod_{j=1}^{r} \cos 2\pi t_j.$$

Some further explanation of this is in order. We have regarded g as an element $(t_1, \ldots, t_r)$ of the abstract torus $T^r \cong (R/Z)^r$, and to interpret (3.8) one must map T^r onto a maximal torus of $\mathrm{Spin}(n, R)$; see Sect. 8 of Chap. 13 of [Hu]. For our purposes we do not need to be too careful with the various tori involved in this situation. Note that for n even, $\Delta(r) = \Delta^+(r) + \Delta^-(r)$.

The following definition will be relevant.

Definition. G is a group, ϱ a representation of G on some (possibly generalized) module, and k is an integer $\geqq 2$. Let $g \in G$ have finite order.

Call g *strongly k-singular* (with respect to ϱ) if some eigenvalue of $\varrho(g)$ is a primitive d-th root of unity for some $1 \neq d \mid k$. Call g k-regular (w.r.t. ϱ) if all eigenvalues of $\varrho(g)$ are l-th roots of unity for l coprime to k, i.e., $\varrho(g)$ has order coprime to k. Call g *weakly k-singular* (w.r.t. ϱ) if g is neither strongly k-singular nor k-regular.

We now choose a *rational* element $g \in SO(n, R)$, i.e., an element g whose characteristic polynomial with respect to the natural representation ϱ on E is given by (2.2). We then let $d(g)$, $k(g)$, and $\Delta(g)$ be as in (2.3).

At last we may state the main results of the present section.

Proposition 3.9. *The following holds for $k \geqq 2$:*

$$\beta_k(g) = \begin{cases} k^{2k(g)}, & \text{if } \; g \text{ is } k\text{-regular (with respect to } \varrho) \\ 0, & \text{if } \; g \text{ is strongly } k\text{-singular} \end{cases}$$

Moreover if k is a prime and g is weakly k-singular of order f then

$$\beta_k(g) = k^{2k(g)} \Big(\prod_{1 < p^\alpha \mid f/k} p^{e(p^\alpha k)} \Big)^{k-1}$$

where the product runs over non-trivial prime power divisors p^α of f/k and for $d \mid f$ we have set

$$e(d) = \sum_{d \mid i} k(i) \quad (cf.\ (2.2)).$$

Remarks. There is an analogous formula for $\beta_k(g)$ for composite k, but it is more complicated to state and in any case we do not need it. Note also that if g is weakly k-singular for k a prime then k necessarily divides the order f of g.

Theorem 3. *Let $k = 2$ or k be odd and let* $\dim E = 2r$. *Assume that g is not weakly k-singular and define g_0 as follows: $g_0 = g$ if k is odd; if $k = 2$ then g_0 is a pre-image of g under the natural map* $\mathrm{Spin}(2r, R) \to O(2r, R)$ *chosen arbitrarily if g is (strongly) 2-singular and chosen to have odd order if g is 2-regular. Then*

$$\beta_k^{\mathrm{or}}(g_0) = \left(\frac{(-1)^r \Delta(g)}{k} \right) k^{k(g)}.$$

Remarks. 1. If $k = 2$ and g is 2-regular then g_0 is, of course, uniquely determined.

2. If g is strongly k-singular then k divides $d(g)$ and so $\beta_k^{\mathrm{or}}(g_0) = 0$ by the theorem, as it must be to be consistent with (3.3b) and Proposition 3.9.

3. One verifies that if g has odd order then $(-1)^r \Delta(g) \equiv 1 \pmod 4$, so that the Kronecker symbol $\left(\frac{(-1)^r \Delta(g)}{k} \right)$ makes sense in the case $k = 2$, g 2-regular. Thus the

case $k=2$ of Theorem 3 can be reformulated as follows (and in this form it also holds if $\dim E=2r+1$):

(3.10) **Proposition.** *Let g and g_0 be as in Theorem 3. Then*

$$\beta_2^{\mathrm{or}}(g_0)=\left(\frac{2}{d(g)}\right)2^{[k(g)]}.$$

We begin with a proof of Proposition 3.10 which is based on Eisenstein's proof of the law of quadratic reciprocity, as exposited in Serre's text [Se] on p. 10. Thus we have the identity

(3.11) For m a positive odd integer,

$$\frac{\sin mx}{\sin x}=(-4)^{(m-1)/2}\prod_{j=1}^{(m-1)/2}\left(\sin^2 x-\sin^2\frac{2\pi j}{m}\right),$$

(3.12) For m a positive odd integer,

$$\prod_{j=1}^{(m-1)/2}\cos\frac{2\pi j}{m}=\left(\frac{2}{m}\right)2^{-(m-1)/2}.$$

Proof. Set $x=\pi/2$ in (3.11), noting that $\sin m\pi/2=(-1)^{(m-1)/2}$. We then obtain the equation

$$2^{m-1}\prod_{j=1}^{(m-1)/2}\cos^2\frac{2\pi j}{m}=1$$

and it is sufficient to verify that the *sign* involved in the statement of (3.12) is correct. Thus we are looking for the number of integers j which satisfy $m/4<j\leqq(m-1)/2$, for these are the values of j for which $\cos 2\pi j/m$ is negative. One readily verifies that the number of such j, say N, is even if, and only if, $m\equiv\pm1 \pmod 8$. The result follows.

We can now complete the proof of Proposition 3.10. We may assume that g has odd order (cf. the second remark following the statement of Theorem 3), in which case $\beta_2^{\mathrm{or}}(g_0)$ is given by (3.8) where we may take t_j to be the form j/m for m odd and $0\leqq j\leqq(m-1)/2$. In effect, then, (3.12) gives us the "contribution" of a single cycle of length m to the value of $\beta_2^{\mathrm{or}}(g_0)$, so if g has characteristic polynomial as in (2.2) then we get

$$\begin{aligned}\beta_2^{\mathrm{or}}(g_0)&=2^r\prod_{i\,\mathrm{odd}}\left(\frac{2}{i}\right)^{m(i)}2^{-(i-1)k(i)/2}\\&=\left(\frac{2}{d(g)}\right)2^f,\end{aligned}$$

with

$$\begin{aligned}f&=r-\sum(i-1)k(i)/2\\&=r-n/2+k(g).\end{aligned}$$

The proposition is now proved.

We turn next to the proof of Proposition 3.9. It is clear from (3.4) that if g is strongly k-singular, i.e., if g has an eigenvalue μ which is a primitive d-th root of unity for some $d \geqq 2$ which divides k, then $\beta_k(g)=0$; the converse is also true. So assume from now on that g is not strongly k-singular. Then (3.4) yields

$$(3.13) \qquad \beta_k(g)=k^{2k(g)} \prod_{\mu \neq 1} \frac{\mu^k-1}{\mu-1}.$$

Now if g is k-regular then μ^k ranges over the non-trivial eigenvalues of g as μ does (since g is rational), so in this case $\beta_k(g)=\beta^{2k(g)}$ as required. It remains to treat the case that g is weakly k-singular and k is a prime. We only sketch the proof of the desired formula as it will not be crucial in what follows. Let f be the order of g and recall that $k \mid f$.

In the following we let $\Phi_d(t)$ denote the d-th cyclotomic polynomial:

$$(3.14) \qquad \Phi_d(t)=\prod_i (t-\mu_i)=\sum_{a \mid d} (t^a-1)^{\mu(d/a)}$$

where $\mu(d/a)$ is the Möbius function and μ_i ranges over the primitive d-th roots of unity. Set also

$$(3.15) \qquad b(d)=\prod_i (\mu_i-1)=(-1)^{\varnothing(d)}\Phi_d(1)$$

We have

$$(3.16) \qquad \Phi_d(1)=\begin{cases} 0, & d=1 \\ p, & d=p^\alpha>1 \text{ is a prime power}. \\ 1, & \text{otherwise} \end{cases}$$

To see this, use (3.14) to see that we may take d to be square-free, say $d=p_1p_2 \ldots p_s$ with p_i distinct primes. The result is clear if $s=0$, and if $s \geqq 1$ we see that

$$\Phi_d(t)=\frac{\prod(t^u-1)}{\prod(t^v-1)}=\frac{\prod(1+t+\ldots+t^{u-1})}{\prod(1+t+\ldots+t^{v-1})}$$

where u, v jointly range over all divisors of d such that d/u resp. d/v has an even resp. odd number of prime factors. Then $\Phi_d(1)=U/V$ where U resp. V is the product of all u resp. v.

Finally, a given prime p_i is a divisor of exactly $\sum_{j \geqq 0} \binom{s-1}{2j}$ divisors of d with an odd number of prime factors, and exactly $\sum_{j \geqq 0} \binom{s-1}{2j+1}$ divisors of d with an even number of prime factors. If $s \geqq 2$ these numbers are equal, whence $\Phi_d(t)=1$ in this case. If $s=1$ the result is clear.

Now if we set

$$(3.17) \qquad a_k(d)=\prod_k \frac{\mu_i^k-1}{\mu_i-1}, \qquad d \geqq 2$$

where μ_i ranges over the primitive d-th roots of unity, the theory of cyclic groups yields

$$(3.18) \qquad a_k(d)=\begin{cases} b(d/k)^k/(b/d) & \text{if } k^2 \mid d \\ b(d/k)^{k-1}/b(d) & \text{if } k^2 \nmid d \end{cases}$$

Putting (3.15)–(3.18) together we get

$$a_k(d)=\begin{cases}0, & k=d,\\ p^{k-1}, & d=kp^\alpha, \alpha\geqq 1, p \text{ prime},\\ 1, & \text{otherwise}.\end{cases} \tag{3.19}$$

Finally, $a_k(d)$ gives the contribution to $\beta_k(g)$ in (3.13) which derives from a given primitive d-th root of unity. The multiplity of such an eigenvalue is given by the integer $e(d)$ of Proposition 3.9, which is now a consequence of (3.19).

We now turn to the principal result of this section – the proof of Theorem 3. There are several approaches to this result; after Proposition 3.9 and (3.3b) we know that

$$\beta_k^{\mathrm{or}}(g)=\varepsilon(g)k^{k(g)} \tag{3.20}$$

for some sign $\varepsilon(g)=\pm 1$ (g is assumed to be k-regular). Now since β_k^{or} restricts to a generalized character of $\langle g\rangle$, it is a triviality that (3.20) uniquely determines what the sign $\varepsilon(g)$ must be *in case g has odd order*, and of course Theorem 3 implicitly says just what it is. Equally trivially, $\varepsilon(g)$ is *not* determined for g of even order by (3.20) if one assumes only that β_k^{or} is a generalized character, so in this case it is necessary to go back to the Definition (3.7).

We will use a mixture of these approaches. Consider the following result:

(3.21) **Proposition.** *Let $k\geqq 3$ be an odd integer, $m\geqq 2$ an integer and assume $(m,k)=1$.*

Define an integer $N=N(m,k)$ as follows: N is the number of integers a satisfying (i) *$1\leqq a\leqq m/2$;* (ii) *$(a,m)=1$;* (iii) *$ak/2m-[ak/2m]>\frac{1}{2}$. Then the following holds:*

$$(-1)^N=\begin{cases}\left(\frac{k}{p}\right), & m=p^e>1,\ p \text{ an odd prime}\\ \left(\frac{k}{p}\right)\left(\frac{p}{k}\right), & m=2p^e>2,\ p \text{ an odd prime}\\ \left(\frac{2}{k}\right), & \mathrm{m}=2^e\geqq 8\\ \left(\frac{-2}{k}\right), & m=4\\ \left(\frac{-1}{k}\right), & m=2\\ 1, & \text{otherwise}\end{cases}$$

We will show that Proposition 3.21 is equivalent to Theorem 3. Then we give a direct proof of Proposition 3.21 in case m is even. If m is odd (which essentially corresponds to g having odd order) we give a proof based only on the fact that β_k^{or} is a generalized character, which will also provide a second proof of Proposition 3.10.

Now after the remarks following the statement of Theorem 3 we may restrict our attention to the case in which k is odd and g is k-regular. Because of (3.3, 3.7) and Proposition 3.9 we are reduced to determining the sign of the product

$$\prod_{t_j>0} \frac{\sin k\pi t_j}{\sin \pi t_j} \tag{3.22}$$

in the notation of (3.6, 3.7). Now the eigenvalues of g are given by $\exp(\pm 2\pi i t_j)$ for $0 \leqq t_j < 1$, so as g is rational we may take each t_j to lie in the interval $[0, \frac{1}{2}]$. With this choice, the denominator of (3.22) is positive.

We will now show that Proposition 3.21 implies Theorem 3. An eigenvalue of g corresponds to some value a/m of t_j, and for the purposes of studying (3.22) we may take $(a, m) = 1$, $1 \leqq a \leqq m/2$. Moreover, the condition $ak/2m - [ak/2m] > \frac{1}{2}$ says exactly that $\sin k\pi t_j = \sin k\pi a/m$ is negative. In effect, then, for a given m the value of $(-1)^N$ gives the contribution to the sign of (3.22) which accrues from a single primitive m-th root of unity together with its Galois conjugates.

Now assume that g has characteristic polynomial (2.2) and consider the sign contribution to (3.22) from a single cycle of length m. If $m = \prod p_i^{e_i}$ is the prime power decomposition then the relevant eigenvalues are all the m-th roots of unity, primitive or otherwise.

Assume first that m is odd. Then the only eigenvalues which may contribute are, from 3.21, the prime powers dividing m, and the total contribution is exactly

$$\prod_i \left(\frac{k}{p_i}\right)^{e_i} = \left(\frac{k}{m}\right). \tag{3.23}$$

Assume next that $m \equiv 2 (\operatorname{mod} 4)$, $m > 2$, with $p_1^{e_1} = 2$, say. Then Proposition 3.21 tells us that the contribution from eigenvalues *distinct from* -1 is exactly

$$\prod_{i \geqq 2} \left[\left(\frac{k}{p_i}\right)\left(\frac{p_i}{k}\right)\right]^{e_i} \prod_{i \geqq 2} \left(\frac{k}{p_i}\right)^{e_i},$$

which is simply

$$\left(\frac{2m}{k}\right), \quad (m > 2). \tag{3.24}$$

Similarly, the contribution from eigenvalues *distinct from* -1 in case $m \equiv 0 (\operatorname{mod} 4)$ is seen to be

$$\left(\frac{-2m}{k}\right). \tag{3.25}$$

We excluded the eigenvalue -1 from the above considerations since it is the (only!) eigenvalue distinct from 1 which coincides with its Galois conjugates. After the discussion of (3.6) and (3.7), one should not calculate the contribution of -1 to a single cycle, but to the whole expression; after we remember that only one-half of these eigenvalues contribute to (3.22), we obtain from (3.21) that

$$\left(\frac{-1}{k}\right)^{1/2 \sum_{m \text{ even}} k(m)} \tag{3.26}$$

is the *total* contribution from -1's. (Here and below, $k(n)$ is as in (2.2).) We can now at last make explicit the sign of (3.22); from (3.23–3.26) it is

$$(3.27)\quad \prod_{m\,\mathrm{odd}}\left(\frac{k}{m}\right)^{k(m)}\prod_{\substack{m\equiv 2(4)\\ m>2}}\left(\frac{2m}{k}\right)^{k(m)}\cdot\prod_{m\equiv 0(4)}\left(\frac{-2m}{k}\right)^{k(m)}\cdot\left(\frac{-1}{k}\right)^{1/2\sum_{m\,\mathrm{even}}k(m)}$$

Since $\det g=1$ then $\sum_{m\,\mathrm{even}} k(m)\equiv 0(\mathrm{mod}\,2)$. If we further let $c=\prod_{m\,\mathrm{odd}} m^{k(m)}$ then quadratic reciprocity allows us to write the first product in (3.27) as $\left(\frac{c}{k}\right)\left(\frac{-1}{k}\right)^{(c-1)/2}$. Then (3.27) can be re-written as

$$(3.28)\quad \begin{aligned}&\left(\frac{c}{k}\right)\left(\frac{-1}{k}\right)^{(c-1)/2}\prod_{m\,\mathrm{even}}\left(\frac{m}{k}\right)^{k(m)}\left(\frac{2}{k}\right)^{\sum_{m\,\mathrm{even}}k(m)}\left(\frac{-1}{k}\right)^{h}\\ &=\left(\frac{d(g)}{k}\right)\left(\frac{-1}{k}\right)^{h+(c-1)/2}\end{aligned}$$

where $h=\sum_{m\equiv 0(4)} k(m)+\frac{1}{2}\sum_{m\,\mathrm{even}} k(m)$. Comparing this with the statement of Theorem 3, we must show that

$$(3.29)\quad h+\frac{c-1}{2}\equiv k(g)+r\ (\mathrm{mod}\,2).$$

From the definition of c we easily find that $c-1\equiv 2\sum_{m\equiv 3(4)} k(m)\ (\mathrm{mod}\,4)$, so we can write (3.29) in the form

$$2\sum_{m\equiv 2(4)} k(m)+\sum_{m\,\mathrm{even}} k(m)+2\sum_{m\equiv 3(4)} k(m)\equiv\sum_{\mathrm{all}\,m} k(m)+\sum_{\mathrm{all}\,m} mk(m)\ (\mathrm{mod}\,4).$$

Finally this is easily seen to be equivalent to the assertion

$$2\Big(\sum_{m\equiv 0(4)} k(m)+\sum_{m\equiv 3(4)} k(m)\Big)\equiv 2\Big(\sum_{m\equiv 1(4)} k(m)+\sum_{m\equiv 2(4)} k(m)\Big)\ (\mathrm{mod}\,4)$$

which is true since $2k(g)=\sum_{\mathrm{all}\,m} k(m)\equiv 0(\mathrm{mod}\,2)$.

This completes the proof that Proposition 3.21 implies Theorem 3. The converse follows in a similar way, for by the above discussion we only have to determine the sign of (3.22) in case t_j runs over the rational numbers a/m for $(a,m)=1$ and $1\leqq a\leqq m/2$ (at least if $m\geqq 3$), which corresponds to evaluating $\left(\frac{(-1)^r\Delta(g)}{k}\right)$ in the case the eigenvalues of g are just the primitive m-th roots of unity. Then $k(g)=0$, $r=\varnothing(m)/2$, $d(g)=1$ unless m is a prime power, and Proposition 3.21 is readily deduced.

We will now prove Proposition 3.21 in case m is even, although some of the arguments hold for all m. Of course Proposition 3.2 bears much more than a mere superficial resemblance to the Gauss lemma; our proof reflects this. The case $m=2$ is trivial, so for convenience we may assume $m\geqq 3$. Set

$$F(i)=\{a\in N\,|\,(i)\ 1\leqq a\leqq m/2,\ (ii)\ (a,m)=1,\ (iii)\ (i-1)m/2<(ak/2m)<im/2\}$$

for $i=1,2,3,4$, with

$$f(i)=|F(i)|.$$

In the notation of Proposition 3.21 we have

$$N=f(3)+f(4). \tag{3.30}$$

Recall next that the Gauss lemma itself tells us how many residues $ak(\operatorname{mod} m)$ lie in the interval $(\frac{1}{2}m, m)$, or at least gives the parity of the number of such residues. In present terms it tells us that

$$(-1)^{f(2)+f(4)} \equiv k^{\varnothing(m)/2} \ (\operatorname{mod} m). \tag{3.31}$$

The next two results follow from the structure of the group $(Z/mZ)^*$.

$$k^{\varnothing(m)/2} \equiv \begin{cases} \left(\dfrac{k}{p}\right), & m=p^e \text{ or } 2p^e,\ p^e>1 \text{ an odd prime power,} \\ \left(\dfrac{-1}{k}\right), & m=4 \\ 1, & \text{otherwise,} \end{cases} \tag{3.32}$$

the congruence being $(\operatorname{mod} m)$.

(3.33) Suppose that m is neither a prime power nor twice a prime power. Then both $(Z/mZ)^*$ and $(Z/2mZ)^*$ have exponent dividing $\varnothing(m)/2$.

After these preparations we turn to the proof of Proposition 3.21 itself for m even. Let a range over the interval $1 \leqq a \leqq m/2$ with $(a, m)=1$, and consider which of the sets $F(i)$ contains the least positive residue of ak (mod $2m$). Let r denote a typical such residue. We have

$$k^{\varnothing(m)/2} \prod_{a\in F(1)} a \equiv \prod_{i=1}^{4} \prod_{r\in F(i)} r \equiv (-1)^{f(2)+f(4)} \prod_{r\in F(1)} r \prod_{r\in F(2)} (-r) \prod_{r\in F(3)} r \prod_{r\in F(4)} (2m-r), \tag{3.34}$$

the congruences being (mod $2m$). Next we claim

(3.35) For m even, $f(2)+f(3)$ is even if, and only if,

$$k^{\varnothing(m)/2} \equiv (-1)^{f(2)+f(4)} (\operatorname{mod} 2m).$$

Proof. The point is that if r_i, s_i, $i=2,3$, are typical residues in $F(i)$ then $(m-r_2)(m-s_2)$, $(r_3-m)(s_3-m)$ and $(m-r_2)(r_3-m)$ are congruent to $(-r_2)(-s_2)$, $(r_3)(s_3)$, $(-r_2)(r_3)$ (mod $2m$) respectively, since each r_i, s_i is odd (being coprime to m). But then $f(2)+f(3)$ is even, if and only if,

$$\prod_{r\in F(2)} (-r) \prod_{r\in F(3)} r \equiv \prod_{r\in F(2)} (m-r) \prod_{r\in F(3)} (r-m) \ (\operatorname{mod} 2m).$$

But clearly

$$\prod_{a\in F(1)} a \equiv \prod_{r\in F(1)} r \prod_{r\in F(2)} (m-r) \prod_{r\in F(3)} (r-m) \prod_{r\in F(4)} (2m-r)$$

so (3.34) yields that $f(2)+f(3)$ is even if, and only if, $k^{\varnothing(m)/2} \equiv (-1)^{f(2)+f(4)} (\bmod 2m)$ as claimed.

Now assume that m is neither a power of 2 nor twice an odd prime power. Then (3.33) yields $k^{\varnothing(m)/2} \equiv 1 (\bmod 2m)$, whence $f(2)+f(4)$ is even by (3.31). Then (3.35) yields that $f(2)+f(3)$ is even, whence also $f(3)+f(4)$ is even. So Proposition 3.21 holds for such m.

Next assume that $m=2^e \geqq 8$. Then $k^{\varnothing(m)/2} \equiv 1 (\bmod m)$ by (3.32), so $f(2)+f(4)$ is even by (3.31). Then by (3.35) we get $f(3)+f(4)$ even if, and only if $f(2)+f(3)$ is even; if, and only if, $k^{\varnothing(m)/2} \equiv 1 (\bmod 2m)$. But this hold if, and only if, $k \equiv \pm 1 (\bmod 8)$, i.e., $\left(\frac{2}{k}\right)=1$, so again Proposition 3.21 holds.

Now assume $m=2p^e$, $p^e>1$ an odd prime power. By (3.32) we get $k^{\varnothing(m)/2} \equiv \left(\frac{k}{p}\right)$ $(\bmod m)$, so (3.31) yields $(-1)^{f(2)+f(4)} = \left(\frac{k}{p}\right)$. Also, by (3.35) we see that $f(2)+f(3)$ is even if, and only if, $k^{\varnothing(m)/2} \equiv \left(\frac{k}{p}\right) (\bmod 2m)$, and since $k^{\varnothing(m)/2} \equiv \left(\frac{k}{p}\right) (\bmod m)$ then this is equivalent to $k^{\varnothing(m)/2} \equiv \left(\frac{k}{p}\right) (\bmod 4)$. Moreoever $k^{\varnothing(m)/2} \equiv \left(\frac{-1}{k}\right)^{p-1/2}$ $\equiv \left(\frac{k}{p}\right)\left(\frac{p}{k}\right) (\bmod 4)$, so we conclude that $(-1)^{f(2)+f(3)} = \left(\frac{p}{k}\right)$. Finally, we now get

$$(-1)^{f(3)+f(4)} = (-1)^{f(2)+f(3)}(-1)^{f(2)+f(4)} = \left(\frac{p}{k}\right)\left(\frac{k}{p}\right)$$

as required by Proposition 3.21, which is now established whenever m is even (we leave the case $m=4$ to the reader).

Finally, we prove Proposition 3.21 for m odd. By a previous discussion this amounts to the following: take g of odd order such that the eigenvalues of g are just the primitive m-th roots of unity. Then $k(g)=0$ and $\beta=\beta_k^{\mathrm{or}}$ is a generalized character satisfying

$$\beta(g)=\varepsilon(g), \quad \beta(1)=k^{\varnothing(m)/2} \tag{3.36}$$

and β is exponential. We must show that in the notation of Proposition 3.21, $\varepsilon(g)$ $=(-1)^N = \left(\frac{k}{p}\right)$ or 1 according as m is a prime power $p^e>1$ or not.

We proceed by induction on the order of g, the result being trivial if $g=1$. Assume that $m=p^e>1$ is a prime power. For each $1 \neq d \mid m$ we have $\beta(g^{m/d})$ $=\varepsilon(g^{m/d})^{\varnothing(m)/\varnothing(d)} = \varepsilon(g^{m/d}) = \left(\frac{k}{p}\right)$ for $d \neq m$ by induction. Using (3.36) and the integrality condition implicit in the equation

$$(\beta, 1)_{\langle g\rangle} \in Z \tag{3.37}$$

we get

$$\varnothing(m)\varepsilon(g) + \sum_{\substack{d \mid m \\ 1 \neq d \neq m}} \varnothing(d)\left(\frac{k}{p}\right) + k^{\varnothing(m)/2} \equiv 0 (\bmod m). \tag{3.38}$$

Now $k^{\varnothing(m)/2} \equiv \left(\frac{k}{p}\right)$ (mod m), so (3.38) becomes

$$\varnothing(m)\left(\varepsilon(g)-\left(\frac{k}{p}\right)\right)+\sum_{d\mid m}\varnothing(d)\left(\frac{k}{p}\right)=\varnothing(m)\left(\varepsilon(g)-\left(\frac{k}{p}\right)\right)+m\left(\frac{k}{p}\right)\equiv 0(\mathrm{mod}\, m),$$

whence $\varepsilon(g)=\left(\frac{k}{p}\right)$ is immediate.

Finally, if m is not a prime power then $k^{\varnothing(m)/2}\equiv 1(\mathrm{mod\,m})$, $\beta(g^{m/d})^{\varnothing(m)/\varnothing(d)}=1$ since $\varnothing(m)/\varnothing(d)$ is even (for $1\neq d\neq m$), so in this case (3.37) reads

$$\varnothing(m)\varepsilon(g)+\sum_{\substack{d\mid m\\ 1\neq d\neq m}}\varnothing(d)+1\equiv 0(\mathrm{mod}\, m)$$

and as $\sum_{d\mid m}\varnothing(d)=m$ then $\varepsilon(g)=1$ follows, as required.

4. Hecke Operators

After the contortions of the last section we are ready to construct our Hecke operators, which will act on the spaces $\mathfrak{M}(\varrho)$ of Sect. 2. Our approach will be naive, i.e., purely formal; we define our operators via their action on q-expansions, as in Sect. 3, Chap. VII of [L], for example.

Thus let A be any commutative ring with identity 1. We can define operators U_d and V_d on $A[[q]]$, $d\in N$, as follows:

$$U_d:\sum\gamma_n q^n\to\sum_{d\mid n}\gamma_n q^{n/d},\qquad V_d:\sum\gamma_n q^n\to\sum\gamma_n q^{dn}. \tag{4.1}$$

For each prime p choose $\psi_p\in A$ and set

$$\begin{aligned}&\psi_n=\psi_{p_1}^{e_1}\dots\psi_{p_r}^{e_r}\quad\text{if}\quad n=p_i^{e_i}\dots p_r^{e_r},\ p_i\ \text{distinct primes},\\ &\psi_1=1,\end{aligned} \tag{4.2}$$

and identify ψ_n with the operator which acts on $A[\![q]\!]$ by component-wise multiplication by ψ_n.

The Hecke operator T_n (with respec to the ψ_p) is defined by

$$T_n=\sum_{d\mid n}\psi_d V_d\circ U_{n/d}, \tag{4.3}$$

for example for a prime p we have

$$T_p=U_p+\psi_p V_p. \tag{4.4}$$

We then formally derive multiplicativity:

$$T_m T_n=T_{mn}\quad\text{if}\quad (m,n)=1, \tag{4.5}$$

and the Euler product representation:

$$\sum_{n\geq 1}\frac{T_n}{n^s}=\prod_p\left(1-\frac{T_p}{p^s}+\frac{\psi_p}{p^{2s}}\right)^{-1}. \tag{4.6}$$

Our principal examples (aside from the original Hecke operators themselves!) arise by taking for A a suitable ring of generalized characters. Let $n=2r$ be an integer divisible by 4 in the following.

Example 1. $A=R\,\mathrm{Spin}(n,R)$ with ψ_2 the half-spin character of $\mathrm{Spin}(n,R)$. This allows us to define T_2.

Example 2. $A=R\,\mathrm{Spin}(n-1,R)$ with ψ_p for p an odd prime the generalized character of $\mathrm{Spin}(n-1,R)$ defined following the proof of Lemma 3.2. We emphasize that in this section ψ_n for composite n is defined by (4.2); generally this will *not* be the same as the ψ_n of Sect. 3.

We deduce from Theorem 3 the following consequence (cf. (3.3a)):

(4.7) Let $g\in SO(n-1,R)$ be rational and assume that g and g_0 are as in Theorem 3 with $k=p$ prime. Then

$$\psi_p(g_0)=\left(\frac{\Delta(g)}{p}\right)p^{k(g)-1}=\varepsilon_g(p)p^{k(g)-1}.$$

Example 3. By restricting the ψ_p above to appropriate subgroups G we obtain operators T_p for $A=RG$.

Now let G, V satisfy hypothesis 2.1 with $n=\dim V=2r$ divisible by 4. By extending scalars we get a containment $G\leqq SO(n,R)$ and we can pull back G along $\mathrm{Spin}(n,R)\to SO(n,R)$ to a group $\tilde{G}$. We make use of the notation of Sect. 2, in particular we have the space $\mathfrak{M}(\varrho)$. With the generalized character $\psi_p\in RG$ for p an odd prime available, we can ask if T_p preserves $\mathfrak{M}(\varrho)$. If $\Gamma\in\mathfrak{M}(\varrho)$ with $\Gamma=\sum\gamma_n q^n$ then $T_p\Gamma\in\mathfrak{M}(\varrho)$ precisely when for each $g\in G$ the q-expansion

$$(U_p+\psi_p(g)V_p)\left(\sum\gamma_n(g)q^n\right) \tag{4.8}$$

is again a modular form of weight $k(g)$ and character ε_g. But after (4.2), (4.8) is just the action of the usual Hecke operator, so that (4.8) indeed has the required properties. Of course we must qualify this assertion with the remark that (4.8) only holds if g is not weakly p-singular and if $g\in SO(n-1,r)\subseteq SO(n,R)$.

When considering the case $p=2$ we have already seen that it is necessary to pass to the group $\tilde{G}:\psi_2$ will generally not induce a character of G. In favorable situations the extension

$$Z_2\to\tilde{G}\to G \tag{4.9}$$

will split, so that $\tilde{G}\cong Z_2\times G$, and we may indeed think of ψ_2 as a character of G itself. On the other hand if $g\in G$ is not weakly 2-singular and $\pm g_0$ are the pre-images of g with g_0 as in Theorem 3 then we have

$$\psi_2(-g_0)=-\varepsilon_g(2)2^{k(g)-1}.$$

Now $\varepsilon_g(2)=0$ if $o(g)$ is even and there is no problem; if $o(g)$ is odd, say N, then ε_g is defined modulo N (cf. Sect. 2) and $\Gamma_g(z)=\sum\gamma_n(g)q^n$ is a form on $\Gamma_0(N)$. Then $V_2(\Gamma_g(z))$ is a form on $\Gamma_0(2N)$ with the same weight and character ε_g, so that also $(U_2+\psi_2(-g_0)V_2)(\Gamma_g(z))$ is on $\Gamma_0(2N)$. In this case, then, one should perhaps think of T_2 as a map

$$T_2;\mathfrak{M}(\varrho)\to\tilde{\mathfrak{M}}(\varrho)$$

where $\tilde{\mathfrak{M}}(\varrho)$ is a space of Thompson series for $\tilde{G}$ satisfying the same sort of conditions as $\mathfrak{M}(\varrho)$ (cf. (2.10)). Of course the canonical embedding $RG \to R\tilde{G}$ induces a map $\mathfrak{M}(\varrho) \to \tilde{\mathfrak{M}}(\varrho)$, but in any case T_2 does not generally act as the usual Hecke operator on $\Gamma_g(z)$ if g has odd order. Nevertheless we will find it profitable to use the operator T_2 later.

We gather some of the preceding in the next result, which epitomizes the "nicest" situation. We use the

Definition. Given G as above, we say that T_p *exists* if T_p defined by (4.4) satisfies

$$T_p : \mathfrak{M}(\varrho) \to \mathfrak{M}(\varrho)$$

and if for each $g \in G$, $\Gamma \in \mathfrak{M}(\varrho)$, the operator $T_p(g) := U_p + \psi_p(g) V_p$ acts on $\Gamma_g(z)$ as the "usual" Hecke operator.

Theorem 4. *Let the notation and assumptions be as above.*

(i) *Assume that $G \leqq SO(n-1, R)$, and let p be an odd prime such that G contains no weakly p-singular elements. Then the Hecke operator T_p exists.*

(ii) *Assume that the extension* (6.9) *splits and that G contains no weakly 2-singular elements. Then T_2 exists.*

Remark that G will certainly contain no weakly p-singular elements if $|G|$ is not divisible by p, but these are by no means sufficient conditions. If G is represented on V by permutation matrices, for example, then it certainly contains no weakly p-singular elements for any prime p.

Three final comments are appropriate. First, although the weakly p-singular elements do not fit the formalism of Hecke operators, nevertheless they are not entirely without interest; we will encounter some in connection with the Leech lattice in Sect. 5. Secondly, if we are willing to deal with the operator pT_p rather than T_p itself then from (3.3) we see that

$$pT_p = pU_p + \beta_p^{\mathrm{or}} V_p,$$

in other words we can deal with Spin(n, R) and its subgroups other than having to stay inside Spin$(n-1, R)$ (we can in any case do this if $p=2$, so these comments mainly apply to odd p).

Finally, as soon as we know that T_p exists, we can use the results of Theorem 2 to produce new (and often interesting) Thompson series $T_p\Omega_G$ and $T_p\Theta_G$.

5. The Leech Lattice and Some Euler Products

Having constructed the operators T_p, at least under certain assumptions, it is natural to ask about the existence of Thompson series Γ_G which are *eigenforms*, i.e., satisfy

$$T_p \Gamma_G = \alpha(p) \Gamma_G \tag{5.1}$$

for some set of primes p. If $\Gamma_G = \sum_{n \geqq 0} \gamma_n q^n$ is holomorphic one knows by the usual formalism (cf. (4.6) and [L]) that (5.1) holds for the prime p if, and only if, the formal Dirichlet series associated Γ_G has an Euler factor for the prime p. In particular if

(5.1) holds for all p and if $\gamma_1 = 1_G$ then

(5.2) $$\sum_{n \geqq 1} \frac{\gamma_n}{n^s} = \prod_p \left(1_G - \frac{\gamma_p}{p^s} + \frac{\psi_p}{p^{2s}}\right)^{-1}$$

We will use the symbol Λ for the Leech lattice, characterized as the unique 24-dimensional even, unimodular, integral lattice with no vectors of length 2 (see Conway's article in [C] for this and other facts we use below about Λ). The group of isometries of Λ is the Conway group, hereby denoted by Co. If ϱ is the rational representation of Co afforded by Λ we will study the Thompson series Ω_{Co} of Sect. 2. By Theorem 2 we have $\Omega_{\mathrm{Co}} \in \mathfrak{M}(\varrho)$, and we will look more closely at the forms $\eta_g(z)$ defined by (2.2) for $g \in \mathrm{Co}$. In Appendix 2 the reader will find a tabulation of these forms and various facts about them; almost all of this material has been supplied by A.O.L. Atkin.

The following definition is useful: call g of *permutation-type* if the characteristic polynomial $\chi^g(t)$ of $\varrho(g)$ – (2.2) – is such that each $k(i)$ is non-negative. In the following we let $\eta_g(z)$ be as in (2.4) with $N(g)$ the corresponding level; thus $\eta_g(z)$ is a form on $\Gamma_0(N(g))$ of weight $k(g)$ and Dirichlet character ε_g.

The following results, which I still find remarkable, summarize some facts about the forms $\eta_g(z)$ for $g \in \mathrm{Co}$.

(5.3) Precisely one of the following holds:

(a) $k(g) = 0$, that is $\eta_g(z)$ has weight zero or equivalently g fixes no non-zero vectors in Λ.

(b) $k(g) > 0$ and g is of permutation-type. In this case $\eta_g(z)$ is a (holomorphic) cusp-form and is the unique normalized cusp-form of level $N(g)$ and weight $k(g)$.

(c) $k(g) > 0$ and g is not of permutation-type. In this case there are no non-zero cusp-forms of level $N(g)$ and weight $k(g)$ and $\eta_g(z)$ is a holomorphic Eisenstein series.

(5.4) Suppose that $k(g) > 0$. Then the Dirichlet series associated to $\eta_g(z)$ has an Euler-factor for the prime p, if, and only if, g is not weakly p-singular. In particular this holds for all $p \geqq 5$. Moreover no g is weakly p-singular for both $p = 2$ and 3, so in any case the coefficients of $\eta_g(z)$ are multiplicative.

The assertions of (5.3b) are considered in more detail in [K], [M3] and [KM]; more precisely these papers classify all eta-products of permutation-type which are multiplicative. There appear to be just three such eta-products which do *not* appear in Co, namely

$$18 \cdot 6, \quad 16 \cdot 8, \quad 3^2 \cdot 9^2.$$

As for (5.3c), one can check using Lemma 2.10 that even if g is not of permutation-type then $\eta_g(z)$ is holomorphic (we pointed out in Sect. 2 that this is definitely not the case in general).

Concerning the assertions of (5.4), if g is of permutative-type then the fact that $\eta_g(z)$ is an eigenform follows from (5.3b). As for the Eisenstein series, the multiplicativity of the coefficients still resides in the realms of the miraculous.

Now let

(5.5) $$\Omega_{\mathrm{Co}} = \Omega = \sum_{n=1}^{\infty} \omega_n q^n$$

be the Thompson series for Co given by (2.6). We also introduce the corresponding Dirichlet series

$$L_{\mathrm{Co}} = L = \sum_{n=1}^{\infty} \frac{\omega_n}{n^s}. \tag{5.6}$$

After (5.4) we know that if $g \in \mathrm{Co}$ satisfies $k(g) > 0$ then the "usual" Dirichlet series given by

$$L_g(s) = \sum_{n=1}^{\infty} \frac{\omega_n(g)}{n^s} \tag{5.7}$$

has an Euler factor for all but perhaps one prime and thus in any case has multiplicative coefficients. Let us denote the "p-part" of L by

$$L_p = \sum_{n=0}^{\infty} \frac{\omega_{p^n}}{p^{ns}},$$

and that of $L_g(s)$ by

$$L_{p,g}(s) = \sum_{n=0}^{\infty} \frac{\omega_{p^n}(g)}{p^{ns}}.$$

Thus

$$L_g(s) = \prod_p L_{p,g}(s)\,(k(g) > 0).$$

In order to clarify the nature of the series L_p we introduce the half-spin character ψ_2 of $\mathrm{Spin}(24, R)$ and the generalized character ψ_p of $SO(23, R)$ for odd p as in Examples 1 and 2 of the previous section, regarding ψ_2 as a character of $\mathrm{Spin}(23, R)$ by restriction when convenient. The next result follows from the preceding discussion and the results of Sect. 4:

(5.8) **Lemma.** *Suppose that $k(g) > 0$. If g is not weakly p-singular then*

$$L_{p,g}(s) = \left(1 - \frac{\omega_p(g)}{p^s} + \frac{\psi_p(g)}{p^{2s}}\right)^{-1};$$

in particular this holds if $p \geqq 5$.

The factors $L_{p,g}(s)$ for g weakly p-singular seem to be quite interesting; the data from Appendix 2 indicates the following result:

(5.9) If g is weakly p-singular and $k(g) > 0$ then

$$L_{p,g}(s) = \left(1 - \frac{\omega_p(g)}{p^s} + \sum_{n=2}^{\infty} \psi_p(g) \frac{p^{k(n-2)}}{p^{ns}}\right)^{-1}.$$

This can also be written in the form

$$L_{p,g}(s) = \left(1 - \frac{\omega_p(g)}{p^s} + \frac{\psi_p(g)}{p^s(p^s - p^k)}\right)^{-1}.$$

We take this opportunity to say something more about this result. In Sect. 3 we avoided the use of the values of the characters ψ_p on weakly p-singular elements, but so far as Co is concerned we may proceed as follows: first, the extension (4.9) splits (since Co has a trivial Schur multiplier [G]) so we may regard Co $\leqq \mathrm{Spin}(24, R)$. It is readily verified that the following holds:

(5.10) As Co-modules we have

$$\Delta^+(12) \cong \Lambda + \lambda^3(\Lambda)$$

(here we identify Λ with the complex Co-module $\Lambda \otimes C$ and $\lambda^3(\Lambda)$ is its third exterior power). On the other hand it will follow easily from the next section that the following holds:

(5.11)
$$\begin{aligned} \omega_2 &= -\Lambda \\ \omega_4 &= -(\lambda^3(\Lambda) - \Lambda \otimes \Lambda + \Lambda) \end{aligned}$$

(here we identify a module with its character). Comparing (5.10) and (5.11) we see that

(5.12) As Co-module, $\Delta^+(12)$ affords the character $\omega_2^2 - \omega_4$, i.e., $\psi_2 = \omega_2^2 - \omega_4$.

This shows that the term corresponding to $n=2$ in (5.9) is correct ($p=2$). There is only one class of weakly 3-singular elements g in Co, and the value of $\psi_3(g)$ is easily verified to be consistent with (5.9). We recall that if f is a normalized modular form, say $f = \sum_{n \geqq 1} a(n) q^n$, and if f is invariant under the Hecke operator T_2, then one has $a(2)^2 - a(4) = \varepsilon(2) 2^{k-1}$ where k is the weight and ε the Dirichlet character of f. The result (5.12) is, of course, the Co-analogue of this fact.

Finally, we state some specific results for subgroups of Co which fix a non-zero vector of Λ.

Theorem 5. *Let $G \leqq$ Co fix a non-zero vector of Λ. Then the Hecke operator T_p exists for G whenever G has no weakly p-singular elements (in particular if $p \geqq 5$). In this case we have*

$$T_p \Omega = \omega_p \Omega,$$

equivalently

$$L_p = \left(1 - \frac{\omega_p}{p^s} + \frac{\psi_p}{p^{2s}}\right)^{-1}.$$

Corollary. *Let $M_{24} \leqq$ Co be the usual subgroup permuting an orthonormal frame of Λ. Then we have*

$$T_p \Omega = \omega_p \Omega, \quad \textit{all primes } p,$$

and

$$\sum_{n=1}^{\infty} \frac{\omega_n}{n^s} = \prod_p \left(1 - \frac{\omega_p}{p^s} + \frac{\psi_p}{p^{2s}}\right)^{-1}.$$

(In these two results we have identified the characters ω_n and ψ_p with their restriction to G.)

Proof. Since G fixes a non-zero vector in Λ we get an embedding $G \leqq SO(23, R)$, and even $G \leqq \mathrm{Spin}(23, R)$ coming from Co $\leqq \mathrm{Spin}(24, R)$. Now Theorem 2, Theorem 4, the theory of Hecke operators and the results of this section yield the theorem. As M_{24} is a permutation group on a basis of $\Lambda \otimes Q$ the corollary follows from Theorem 5 and the remarks following Theorem 4.

In Table 5 of [C] the reader will find listed many vector-stabilizers within Co to which Theorem 5 applies, among them several sporadic groups such as Co_2, Co_3 (the two smallest Conway groups), Mc (McLaughlin), HS (Higman-Sims), as well as M_{24} and several classical groups.

Of course there is an analogue of the corollary for each of these groups, but since they will generally contain weakly p-singular elements we have to use (5.9) and consequently the factor L_p may be more complicated.

Finally we mention another sort of Euler product that we receive gratis from the results of Sects. 3 and 4. For simplicity we restrict our attention to the groups $\mathrm{Spin}(2r-1,R)$ and $O(2r-1,R)$ with r even, and let ψ_p for p a prime be as in Examples 1 and 2 of Sect. 4. Consider the formal expansion

$$\prod_p \left(1-\frac{\psi_p}{p^s}\right)^{-1}. \tag{5.13}$$

"Evaluating" (5.13) at a rational element g which is not weakly p-singular for any prime p yields the product (cf. (4.7))

$$\prod_p \left(1-\frac{\varepsilon_g(p)p^{k(g)-1}}{p^s}\right)^{-1}, \tag{5.14}$$

and (5.14) is just the Euler product for the Dirichlet series of the "twisted" Eisenstein series $E_{k(g),\varepsilon}(s)$ defined by

$$E_{k,\varepsilon}(s)=\sum_{n=1}^{\infty}\frac{\sigma_{k-1,\varepsilon}(n)}{n^s} \tag{5.15}$$

where

$$\sigma_{k-1,\varepsilon}(n)=\sum_{d\mid n}\varepsilon(d)d^{k-1}. \tag{5.16}$$

Thus if we define the generalized characters σ_n of $\mathrm{Spin}(2r-1,R)$ via

$$\sigma_n=\sum_{d\mid n}\psi_d$$

then we obtain the equivariant analogue of (5.14–5.16). Thus in particular we have

(5.17) **Lemma.** *Let* $G\leqq \mathrm{Spin}(2r-1,R)$ *be a finite group containing no weakly p-singular elements for any prime p. Then the formal identity*

$$\sum_{n=1}^{\infty}\frac{\sigma_n}{n^s}=\prod_p\left(1-\frac{\psi_p}{p^s}\right)^{-1}$$

reduces to (5.14–5.16) *at each element* $g\in G$.

We remark that we have been careful to avoid the Thompson series associated to the Dirichlet series $\sum\sigma_n n^{-s}$ of Lemma 5.17, for that entails studying the constant term of the Eisenstein series associated to (5.15).

6. Some Explicit Modules

Suppose that $\Gamma_G = \sum \gamma_n q^n$ is a Thompson series for the group G with V_n the (generalized) G-module affording the character γ_n. Then the graded G-module

$$\sum_n \oplus V_n \tag{6.1}$$

affords Γ_G in an obvious sense; conversely gives a graded G-module (6.1) we can form the series

$$\sum_n V_n q^n = \sum_n \gamma_n q^n \tag{6.2}$$

(identifying a module with the character it affords). In this section we will be concerned with the following issues: to describe genuine graded G-modules, i.e., each V_n is a genuine G-module, such that it affords a Thompson series Γ_G with the property that for each $g \in G$, $\Gamma_G(z) = \sum_n \gamma_n(g) q^n$ is a modular function, i.e., has weight 0.

This problem is inspired, of course, by the questions raised in [CN] concerning the Monster. Furthermore anyone familiar with the work of [FLM] will see some familiar objects below, though perhaps from a different perspective.

First let L be an even G-lattice (hypothesis 2.5) with $\mathbb{C}L$ the group algebra of L and $L_n = \{x \in L \mid (x, x) = 2n\}$. Then there is a natural G-grading attached to $\mathbb{C}L$:

$$\mathbb{C}L = \sum_{n \geqq 0} \oplus L_n \tag{6.3}$$

and of course $\mathbb{C}L$ affords the Thompson series Θ_G of (2.7).

Next, let V be some G-module, and consider the graded module

$$M = M(V) = \sum_{i \geqq 1} \oplus V_i \tag{6.4}$$

where each $V_i \cong V$. We let $S(M)$ and $\Lambda(M)$ be the symmetric and exterior algebras of M, each of which carries a natural G-grading. Specifically we have in terms of q-expansions,

$$\begin{aligned} S(M) &= \prod_{k=1}^{\infty} \sum_{r=0}^{\infty} S^r(V) q^{kr}, \\ \Lambda(M) &= \prod_{k=1}^{\infty} \sum_{r=0}^{\infty} \lambda^r(V) q^{kr}, \end{aligned} \tag{6.5}$$

where $S^r(V)$, $\lambda^r(V)$ are the r-th symmetric and exterior powers of V (cf. Sect. 4 of [M2] or Sect. 2.2 of [Br]).

We will also have cause to consider a variant of these modules, namely let

$$M_{1/2} = \sum_{n \geqq 1} \oplus V_{n-1/2}$$

be graded by half-integers, where again $V_{n-1/2} \cong V$; then define $\Lambda(M_{1/2})$ to be the exterior algebra on $M_{1/2}$ and let

$$\Lambda^o(M_{1/2})$$

be the submodule consisting elements of integral degree. In terms of q-expansions we may write

(6.6)
$$\Lambda^{e}(M_{1/2})=\left[\prod_{k=1}^{\infty}\sum_{r=0}^{\infty}\lambda^{r}(V)q^{r(k-1/2)}\right]_{Z}$$

where the subscript Z means that we take only the integral powers of q.

(6.7) **Lemma.** *Let ϱ be a rational representation of the finite group G by unimodular matrices of degree $24d$, $d\in Z$, on the space V. Then with M as in (6.4), the Thompson series Ω_G^{-1} is afforded by $S(M)$ with the grading decreased by d. In terms of q-expansions, Ω_G^{-1} arises from*

$$q^{-d}S(M)=q^{-d}\prod_{k=1}^{\infty}\sum_{r=0}^{\infty}S^{r}(V)q^{kr}.$$

Proof. See [M2], where it is also shown that Ω_G itself corresponds to

$$q^{d}\prod_{k=1}^{\infty}\sum_{r=0}^{\infty}(-1)^{r}\lambda^{r}(V)q^{kr}.$$

(Equation (5.11) follows from this.)

Theorem 6.1 *Let L be an even, unimodular G-lattice with $V=V_L$ the corresponding rational G-module. Assume that $G\leqq\Theta(V_L)$ and that* $\dim L=24d$, *$d\in Z$. Let X be the graded G-module*

$$q^{-d}\mathbb{C}L\otimes S(M).$$

Then the following hold:

(i) *X affords the Thompson series $\theta_{L,G}/\Omega_G=J_G$, say.*
(ii) *If $g\in G$ then $J_g(z)=\theta_{L^g}(z)/\eta_g(z)$ is a modular function.*
(iii) *$J_g(z)$ has level $N(g)$ dividing $24o(g)$.*

Remark. Part (iii) should be compared to part (4) of Thompson's conjecture in [T]. Note that X is a genuine G-module, moreover for any given G we can find an L satisfying the hypothesis of Theorem 6.1.

Proof. See [M5].

We note that if $g=1$ in Theorem 6.1 then $J_1(z)$ (the Poincaré series of X) is given by $\theta_L(z)/\eta(z)^{24d}$ and by (iii) is a modular function of level 1. The q-expansion is of the form $q^{-d}+$higher terms, in particular if $\dim L=24$ then $J_1(z)$ must differ from the modular function $j(z)$ by a constant. Let us set

$$J(q)=q^{-1}+196884q+\dots$$

so that $J(q)=j(z)-744$. We may ask for those groups G with the property that there is a Thompson series $\Gamma_G=\sum_{n\geqq1}\gamma_n q^n$ such that

(a) $\Gamma_g(z)=\sum\gamma_n(g)q^n$ is a modular function for each $g\in G$.
(b) $\Gamma_1(z)=J(q)+$constant.

Of course we can take each γ_n to be a sum of trivial G-modules – hardly an interesting situation – we are interested in non-trivial examples. Theorem 6.1

provides some examples whenever G acts on an even, unimodular lattice of dimension 24. We give another example which is in some sense "48-dimensional."

Theorem 6.2. *Let V be the standard 48-dimensional module for $SO(48, R)$, let M be as in (6.4), and let X be the graded* Spin$(48, R)$ *module given by*

$$q^{-1}\Lambda^o(M_{1/2}) \oplus q^2 \Delta^+(24)\Lambda(M).$$

Then the following hold:

(i) *The Poincaré series of X is $J(q)+1128$.*

(ii) *Let $G \leqq SO(48, R)$ be represented rationally on V, and assume that G contains no weakly 2-singular elements and also that the pull-back of G into* Spin$(48, R)$ *splits. Then X affords the Thompson series*

$$J_G = T_2 \Omega_G / \Omega_G$$

of G where T_2 is the Hecke operator of Theorem 4.

Remarks. A discussion of the case in which G contains weakly 2-singular elements or does not split with lifted to Spin$(48, R)$ is also possible, although we will ignore it here.

Proof. Let us compute the q-expansion of the level 1 form $T_2\eta(z)^{48}/\eta(z)^{48}$, which is

$$\begin{aligned}\frac{(U_2+2^{23}V_2)\left(q^2 \prod\limits_{n\geqq 1}(1-q^n)^{48}\right)}{q^2 \prod\limits_{n\geqq 1}(1-q^n)^{48}} &= q^{-1}U_2\left\{\frac{\prod\limits_{n\geqq 1}(1-q^n)^{48}}{\prod\limits_{n\geqq 1}(1-q^{2n-1})^{48}}\right\} \\ &\quad + 2^{23}q^2 \frac{\prod\limits_{n\geqq 1}(1-q^{2n})^{48}}{\prod\limits_{n\geqq 1}(1-q^n)^{48}} \\ &= q^{-1}U_2\left(\prod_{n\geqq 1}(1-q^{2n-1})^{48}\right) + 2^{23}q^2 \prod_{n\geqq 1}(1+q^n)^{48} \\ &= q^{-1}U_2\left(\prod_{n\geqq 1}(1+q^{2n-1})\right)^{48} + 2^{23}q^2 \prod_{n\geqq 1}(1+q^n)^{48} \\ &= q^{-1}\left[\prod_{n\geqq 1}(1+q^{n-1/2})^{48}\right]_Z + 2^{23}q^2 \prod_{n\geqq 1}(1+q^n)^{48}\end{aligned}$$

where the subscript "Z" indicates that we take only the integral powers of q.

Now with V as in the theorem, the formalism of Sect. 2 of [M2] shows that we obtain the graded Spin$(48, R)$-module whose Poincaré series is this last q-expansion by replacing $(1+q^i)^{48}$ by $\sum \lambda^r(V)q^{ri}$ and replacing 2^{23} by the half-spin module $\Delta^+(24)$. A comparison with (6.5, 6.6) shows that we obtain the module X of the theorem.

Since $T_2\eta(z)^{48}/\eta(z)^{48} = q^{-1}+1128+\ldots$ is a modular function of level 1 then (i) holds, and more generally Sect. 2 of [M2] shows that (ii) also holds as long as the "equivariant" operator T_2 is well-behaved, which is guaranteed by Theorem 4.1(ii).

The first few terms of the q-expansion afforded by X are as follows:

$$q^{-1}+\lambda^2 V + (\lambda^4 V + V^2)q + (\lambda^6 V + V\lambda^3 V + \lambda^2 V + V^2 + \Delta^+(23))q^2 + \ldots$$

Let us consider some finite subgroups G of $SO(48, R)$ to which Theorem 6.2 applies. Certainly if G acts on 48-dimensional lattice then X affords a Thompson series for the pullback $\tilde{G}$ of G into Spin$(48, R)$. For example if we take $G \cong \mathrm{Co} \times \mathrm{Co}$ (acting on the sum of two copies of the Leech lattice) we get a Thompson series for G, moreover by taking either one of the direct factors or a "diagonal" copy of Co we get distinct Thompson series.

Another interesting example is obtained by taking $G \cong A_{48}$ acting on its natural permutation module. Then the graded module X of Theorem 6.2 affords a Thompson series for the 2-fold covering group $\hat{A}_{48}$ and part (ii) of the theorem applies to any subgroup of A_{48} which splits over the center of $\hat{A}_{48}$. For example we may take $L_2(47)$ (acting on the projective line over $GF(47)$) of M_{24}.

One can check that in the $L_2(47)$ case each of the series $J_g(z) = T_2\eta_g(z)/\eta_g(z)$ coincides (up to a constant) with the Thompson series of an element of the same order in the Monster as in [CN], suggesting that perhaps $L_2(47) \leq M$ (although in fact it is not *). If we take $M_{24} \leq \hat{A}_{48}$ such that it acts with two orbits of length 24 on the 48 letter permuted by A_{48} then the Thompson series $T_2\Omega_{M_{24}}/\Omega_{M_{24}}$ appears to coincide with that of Theorem 6.1 (with $L =$ Leech lattice) up to a constant.

Appendix

Eta-Products for Co

We list information about the forms $\eta_g(z)$ arising from the action of the Conway group Co on the Leech lattice, as studied in Sect. 5. More precisely in Table 1 we give the Dirichlet series corresponding to $\eta_g(z)$ for those $g \in \mathrm{Co}$ with $k(g) > 0$, g not of permutation-type, and g not weakly p-singular for any prime p (cf. (5.4, 5.3c)); Table 2 gives similar information for g weakly p-singular, but with the first few terms of the p-part of the Dirichlet series given explicitly; Table 3 gives the coefficient of q^p (for the first few primes p) in the q-expansion of $\eta_g(z)$ for g of permutation type (cf. (5.3b)).

We use the following notation: for a Dirichlet character χ and integer $k \geq 1$ set

$$E_k(s, \chi) = \sum_{n=1}^{\infty} a(n) n^{-s}, \qquad a(n) = \sum_{d \mid n} \chi\left(\frac{n}{d}\right) d^{k-1}.$$

For an integer m we also let $E_k(s, \chi)^{(m)}$ be the Dirichlet series $E_k(s, \chi)$ with the Euler p-factors removed for each prime $p \mid m$. In the tables we index the elements $g \in \mathrm{Co}$ by $a \cdot b$, a being the order of g and b being used to differentiate between non-conjugate elements of the same order. The shape column gives the integers $k(i)$ in the characteristic polynomial (2.2), the weight and level being that of the corresponding eta-product $\eta_g(z)$. The column "char" gives the conductor of the primitive character which induces the Dirichlet character of $\eta_g(z)$. In Tables 1 and 2 we write E_k and $E_k^{(m)}$ for the corresponding Dirichlet series since χ is uniquely determined by the level and conductor.

---

* I thank Simon Norton for pointing this out to me

Table 1. "Eisensteins"

Element	Shape	Wt.	Lvl.	Char.	D. Series
2.2	$2^{16}1^{-8}$	4	2	1	E_4
3.2	$3^{9}1^{-3}$	3	3	-3	E_3
4.2	$2^{6}4^{4}1^{-4}$	3	4	-4	E_3
4.3	$4^{8}2^{-4}$	2	4	1	$E_2^{(2)}$
5.2	$5^{5}1^{-1}$	2	5	5	E_2
6.3	$2^{5}3^{4}6.1^{-4}$	3	6	-3	E_3
6.5	$2^{4}6^{4}1^{-2}3^{-2}$	2	6	1	$(1-3^{-s})^{-1}E_2^{(3)}$
6.6	$3^{3}6^{3}1^{-1}2^{-1}$	2	6	1	$(1-2^{-s})^{-1}E_2^{(2)}$
6.10	$1.6^{6}2^{-2}3^{-3}$	1	6	-3	$(1+2^{-s})^{-1}E_1^{(2)}$
8.2	$2^{3}48^{2}1^{-1}$	2	8	8	E_2
8.3	$8^{4}4^{-2}$	1	8	-4	$E_1^{(2)}$
9.2	$9^{3}3^{-1}$	1	9	-3	$E_1^{(3)}$
10.4	$2^{3}5^{2}.10.1^{-1}$	2	10	5	E_2
12.3	$2^{2}3^{2}4.12.1^{-2}$	2	12	12	E_2
12.6	$2^{3}6.12^{2}.1^{-1}3^{-1}4^{-2}$	1	12	-4	$(1+3^{-s})^{-1}E_1^{(3)}$
12.10	$4^{2}12^{2}2^{1}6^{-1}$	1	12	-3	$E_1^{(2)}$
12.14	$12^{2}3.12^{2}.4^{-2}$	2	12	12	$\prod_p\left[1-\left(\left(\frac{-4}{p}\right)p+\left(\frac{-3}{p}\right)\right)p^{-s}\right]^{-1}$
12.15	$1^{2}4.6^{2}.12.3^{-2}$	2	12	12	$\prod_p\left[\left(\left(\frac{-3}{p}\right)p+\left(\frac{-4}{p}\right)\right)p^{-s}\right]^{-1}$
14.2	$2^{2}.14^{2}.1^{-1}7^{-1}$	1	14	-7	E_1
16.2	$2^{2}.16^{2}.4^{-1}8^{-1}$	1	32	-8	$\prod_p\left[1-\left(\left(\frac{8}{p}\right)+\left(\frac{-4}{p}\right)\right)p^{-s}\right]^{-1}$
18.4	$2^{2}9.18.1^{-1}6^{-1}$	1	18	-3	$E_1^{(3)}$
20.2	$2^{2}5.20.1^{-1}4^{-1}$	1	20	-4	E_1
20.6	$1.2.10.20.4^{-1}5^{-1}$	1	20	-20	$\prod_p\left[1-\left(\left(\frac{5}{p}\right)+\left(\frac{-4}{p}\right)\right)p^{-s}\right]^{-1}$
24.2	$2.3.4.24.1^{-1}8^{-1}$	1	24	-8	E_1
24.8	$2.6.8.24.4^{-1}12^{-1}$	1	48	-3	$\prod_p\left[1-\left(\left(\frac{-4}{p}\right)+\left(\frac{12}{p}\right)\right)p^{-s}\right]^{-1}$
24.9	$1.4.6.24.3^{-1}8^{-1}$	1	24	-24	$\prod_p\left[1-\left(\left(\frac{-3}{p}\right)+\left(\frac{8}{p}\right)\right)p^{-s}\right]^{-1}$
30.4	$2.3.5.30.1^{-1}15^{-1}$	1	30	-15	E_1
30.10	$1.6.10.15.3^{-1}.5^{-1}$	1	30	-15	$(1+2^{-s})^{-1}\prod_{p\neq 2}\left[1-\left(\left(\frac{-3}{p}\right)+\left(\frac{5}{p}\right)\right)p^{-s}\right]^{-1}$

Table 2. Weakly p-singular elements

Element	Shape	Wt.	Lvl.	Char.	p'-part of D. series	p	p-part of D. series
4.8	$1^8 4^8 2^{-8}$	4	4	1	$E_4^{(2)}$	2	$1-8.2^{-s}-64.4^{-s}-512.8^{-s}-4096.16^{-s}\ldots$
6.15	$1^5 3.6^4 2^{-4}$	3	6	-3	$E_3^{(2)}$	2	$1-5.2^{-s}-11.4^{-s}-53.8^{-s}-203.16^{-s}\ldots$
8.9	$1^4 8^4 2^{-2} 4^{-2}$	2	8	1	$E_2^{(2)}$	2	$1-4.2^{-s}+0.4^{-s}+0.8^{3}+0.16^{-s}\ldots$
10.9	$1^3.5.10^2.2^{-2}$	2	10	5	$E_2^{(2)}$	2	$1-3.2^{-s}-1.4^{-s}-7.8^{-s}-9.16^{-s}\ldots$
12.5	$2.3^3.12^3.1^{-1}4^{-1}6^{-3}$	1	12	-3	$E_1^{(2)}$	2	$1+1.2^{-s}-1.4^{-s}+1.8^{-s}-1.16^{-s}\ldots$
12.16	$1^2 3^2 4^2.12^2.2^{-2} 6^{-2}$	2	12	1	$E_2^{(2)}$	2	$1-2.2^{-s}-4.4^{-s}-8.8^{-s}-16.6^{-s}\ldots$
12.17	$1^3.12^3 2^{-1} 3^{-1} 4^{-1} 6^{-1}$	1	12	-3	$E_1^{(2)}$	2	$1-3.2^{-s}+3.4^{-s}-3.8^{-s}+3.16^{-s}\ldots$
16.3	$1^2.16^2 2^{-1} 8^{-1}$	1	16	-4	$E_1^{(2)}$	2	$1-2.2^{-2}+0.4^{-s}+0.8^{-s}+0.16^{-s}\ldots$
18.7	$1.2.18^2 6^{-1} 9^{-1}$	1	18	-3	$(1+2^{-s})^{-1}E_1^{(6)}$	3	$1-2.3^{-s}-2.9^{-s}-2.27^{-s}\ldots$
18.8	$1^2.9.18.2^{-1}.3^{-1}$	1	18	-3	$E_1^{(6)}$	2	$1-2.2^{-s}+1.4^{-s}-2.8^{-s}+1.16^{-s}\ldots$
28.3	$1.4.7.28.2^{-1}14^{-1}$	1	28	-7	$E_1^{(2)}$	2	$1-1.2^{-s}-1.4^{-s}-1.8^{-s}-1.16^{-s}\ldots$
30.7	$2.3.5.30.6^{-1}10^{-1}$	1	30	-15	$\prod_{p\neq 2}\left[1-\left(\left(\frac{-3}{p}\right)+\left(\frac{5}{p}\right)\right)p^{-s}\right]^{-1}$	2	$1+0.2^{-s}-1.4^{-s}+2.8^{-s}-3.16^{-s}\ldots$

Table 3. The cusp-forms

Element	Shape	Wt.	Lvl.	Char.	2	3	5	7	11	13	17
1.1	1^{24}	12	1	1	-24	252	4830	-16744	534612	-577738	-6905934
2.3	2^{12}	6	4	1	0	-12	54	-88	540	-418	594
2.4	$1^8 2^8$	8	2	1	-8	12	-210	1016	1092	1382	14706
3.3	3^8	4	9	1	0	0	0	20	0	-70	0
3.4	$1^6 3^6$	6	3	1	-6	9	6	-40	-564	638	882
5.3	$1^4 5^4$	4	5	1	-4	2	-5	6	32	-38	26
4.4	4^6	3	16	-4	0	0	-6	0	0	10	-30
4.5	$2^4 4^4$	4	8	1	0	-4	-2	24	-44	22	50
4.7	$1^4 2^4 4^4$	5	4	-4	-4	0	-14	0	0	-238	322
6.7	$2^3 6^3$	3	12	-3	0	-3	0	2	0	-22	0
6.9	6^4	2	36	1	0	0	0	-4	0	2	0
6.11	$1^2 2^2 3^2 6^2$	4	6	1	-2	-3	6	-16	12	38	-126
7.2	$1^3 7^3$	3	7	-7	-3	0	0	-7	-6	0	0
8.4	$4^2 8^2$	2	32	1	0	0	-2	0	0	6	2
8.8	$1^2 2 4 8^2$	3	8	-8	-2	-2	0	0	14	0	2
10.5	$2^2 10^2$	2	20	1	0	-2	-1	2	0	2	-6
11.1	$1^2 11^3$	2	11	1	-2	-1	1	-2	1	4	-2
12.11	12^2	1	144	-4	0	0	0	0	0	-2	0
12.12	2.4.6.12	2	24	1	0	-1	-2	0	4	-2	2
14.3	1.2.7.14	2	14	1	-1	-2	0	1	0	-4	6
15.3	1.3.5.15	2	15	1	-1	-1	1	0	-4	-2	2
20.4	4.20	1	80	-20	0	0	-1	0	0	0	0
21.2	3.21	1	63	-7	0	0	0	-1	0	0	0
22.2	2.22	1	44	-11	0	-1	-1	0	1	0	0
23.1	1.23	1	23	-23	-1	-1	0	0	0	-1	0

References

[AT] Atiyah, M.F., Tall, D.O.: Group representations, λ-rings and the J-homomorphism. Topology 8, 253–297 (1969)

[Ba] Biagioli, A.J.: Products of transforms of the Dedekind eta function. Ph.D. Thesis, U. Wisconsin (Madison), 1982

[Bo] Bott, R.: Lectures on $K(X)$. New York: Benjamin 1969

[Br] Broué, M.: Groupes finis, séries formelles et fonctions modulaires. Sémin. Jacques Tits, Collège de France, 1982

[C] Conway, J.H.: Three lectures on exceptional groups. In: Powell-Higman, Finite Simple Groups (pp. 215–247). London: Academic Press 1971

[CN] Conway, J.H., Norton, S.P.: Monstrous moonshine. Bull. Lond. Math. Soc. **11**, 303–339 (1979)

[FLM] Frenkel, I., Lepowsky, J., Meurman, A.: A natural representation of the Fischer-Griess Monster with the modular function j as character. M.S.R.I. reprint, Berkeley, 1984

[G] Griess, R.: Schur multipliers of some sporadic simple groups. J. Alg. **32**, 445–446 (1974)

[H] Hecke, E.: Mathematische Werke, 2nd ed. Göttingen: Vandenhoeck u. Ruprecht 1970

[Hu] Husemöller, D.: Fibre bundles, 2nd Ed. (Graduale Texts in Mathematics, Vol. 20) Berlin Heidelberg New York: Springer 1975

[K] Koike, M.: On McKay's conjecture, to appear

[KM] Kisilevsky, H., McKay, J.: Multiplicative products of η-functions. Contemp. Math., **45** (1985)

[L] Lang, S.: Introduction to modular forms. New York: Springer 1976

[M1] Mason, G.: Applications of quasi-invertible characters. J. Lond. Math. Soc.

[M2] Mason, G.: Frame-shapes and rational characters of finite groups. J. Alg. **89**, 237–246 (1984)

[M3] Mason, G.: M_{24} and certain automorphic forms. Contemp. Math., **45**, 223–244 (1985)

[M4] Mason, G.: Modular forms and the theory of Thompson series. In: Proceedings of the Rutgers Theory Year, 1983–1984 (Aschbacher, M., et al., eds.), pp. 391–407. Cambridge: Cambridge University Press 1984

[M5] Mason, G.: Discriminants and the spinor norm. To appear in Proc. Lond. Math. Soc.

[O] Ogg, A.: Modular forms and Dirichlet series. New York: Benjamin 1969

[R] Rademacher, H.: Topics in analytic number theory. New York: Springer 1973

[Sch] Schoeneberg, B.: Elliptic modular functions. New York: Springer 1974

[Se] Serre, J.-P.: A course in arithmetic. Berlin Heidelberg New York: Springer 1973

[T] Thompson, J.G.: Finite groups and modular functions. Bull. Lond. Math. Soc. **11**, 374–351 (1979)

[Td] Tom Dieck, T.: Transformation groups and representation theory. S.L.N. 766, New York, 1979

[Z] Zassenhaus, H.: On the spinor norm. Arch. Math. **13**, 434–451 (1962)

Received July 8, 1986; in revised form May 5, 1988

Math. Ann. 283, 411–429 (1989)

Mathematische Annalen

Bielliptic Abelian Surfaces

Klaus Hulek[1] and Steven H. Weintraub[2]

[1] Mathematisches Institut, Universität Bayreuth, Postfach 101251, D-8580 Bayreuth, Federal Republic of Germany

[2] Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803, USA

0. Introduction

In this paper we consider polarized abelian surfaces (A, H_A) where H_A is a polarization of type $(1, p)$. By polarization we mean as usual a Riemann form on a lattice defining A or equivalently a line bundle on A modulo translation. We shall assume $p \geqq 5$ and prime. The latter assumption is not essential for our methods but helps to keep the computations short. Our aim is to prove a criterion for H_A to be very ample which can be formulated in terms of period matrices (Theorems 0.8 and 0.9). Note that all the results of this paper except for 0.1, 0.2, 0.8, 0.9, 0.10 and 4.4 (iii) are also true for $p=3$.

Our starting point is a result of Ramanan [R] which gives a geometric characterisation of such surfaces. In order to formulate Ramanan's result, recall that one can find a cyclic covering

$$\pi : A \to A/\mathbb{Z}_p = : B$$

onto a principally polarized abelian surface B such that H_A is the pullback of the principal polarization H_B on B. Since (B, H_B) is a principally polarized abelian surface it is necessarily a Jacobian and one of the two following possibilities occurs:

(i) $B = \operatorname{Jac} C$ where C is a smooth genus 2 curve

(ii) $B = E_1 \times E_2$ where E_1 and E_2 are elliptic curves and the principal polarization is given by the reducible curve $C = E_1 \times \{0\} \cup \{0\} \times E_2$.

The polarization H_A on A is then given by $D = \pi^{-1}(C)$. In case (i) D is a smooth curve of genus $p+1$. This case was considered in [R].

Proposition 0.1 (Ramanan). *H_A is very ample unless D and C admit elliptic involutions which commute with the Galois action of the covering $A \to B$, i.e. if and only if there is a cartesian diagram*

$$\begin{array}{ccc} D & \xrightarrow{2:1} & E \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle p:1} \\ C & \xrightarrow{2:1} & E' \end{array}$$

where E and E' are elliptic curves.

Proof. [R].
The remaining case (ii) was treated in [HL].

Proposition 0.2 (Hulek-Lange). *Let $\alpha=(\alpha_1,\alpha_2)\in E_1\times E_2$ be a p-torsion point which corresponds to the covering $A\to B$. Then H_A is very ample unless one of the following two cases occurs:*

(i) *$\alpha_1=0$ or $\alpha_2=0$. Then (A,H_A) splits as a polarized abelian surface.*
(ii) *There exists an isomorphism $\phi: E_1\to E_2$ with $\phi(\alpha_1)=\alpha_2$.*

Proof. [HL].

Remark 0.3. (i) Condition (ii) of Proposition 0.2 just means that the reducible curves C, resp. D admit $2:1$ covers onto elliptic curves which commute with the Galois covering π. This corresponds to Ramanan's result.
(ii) An independent proof can be given using Reider's result [Be].

As the reader may already have noticed, and as will become even clearer later, many of the arguments in this paper (or in this subject) use the cyclic covering $\pi: A\to B$. (Such a covering always exists but is not unique.) It is worthwhile to formalize this notion.

Definition 0.4. Let (A,H_A) be an abelian surface with a polarization of type $(1,p)$. A *root* of (A,H_A) is a (necessarily p-fold) cyclic covering map $\pi:(A,H_A)\to(B,H_B)$ with (B,H_B) a principally polarized abelian surface, i.e. a p-fold cover $\pi: A\to B$ with H_A the pullback of the principal polarization H_B via π. (A,H_A) together with a root will be called a *rooted* polarized abelian surface.

In order to formulate our result we consider the *Siegel upper half space* of *degree* 2:

$$\mathfrak{S}_2=\{\tau\in M(2\times 2,\mathbb{C}),\ \tau={}^t\tau,\ \operatorname{Im}\tau>0\}\ .$$

On $\mathbb{R}^4$ we fix the standard symplectic form

$$J=\begin{pmatrix}0 & \mathbb{1}_2\\ -\mathbb{1}_2 & 0\end{pmatrix}$$

and denote the corresponding symplectic group over $\mathbb{Q}$ by

$$Sp(4,\mathbb{Q})=\{X\in GL(4,\mathbb{Q}),\ XJ{}^tX=J\}\ .$$

Note that vectors in $\mathbb{R}^4$ will be considered as row vectors with the group $GL(4,\mathbb{Q})$ operating by multiplication from the right. The group $Sp(4,\mathbb{Q})$ operates on $\mathfrak{S}_2$ by

$$\tau\mapsto(A\tau+B)(C\tau+D)^{-1}$$

where $\begin{pmatrix}A & B\\ C & D\end{pmatrix}\in Sp(4,\mathbb{Q})$ and A, B, C and D are 2×2 matrices. Consider the groups

$$\Gamma'=Sp(4,\mathbb{Z})$$

$$\Gamma^0_{1,p}=\left\{X\in Sp(4,\mathbb{Q}),\ X\in\begin{pmatrix}\mathbb{Z} & \mathbb{Z} & \mathbb{Z} & p\mathbb{Z}\\ \mathbb{Z} & \mathbb{Z} & \mathbb{Z} & p\mathbb{Z}\\ \mathbb{Z} & \mathbb{Z} & \mathbb{Z} & p\mathbb{Z}\\ \frac{1}{p}\mathbb{Z} & \frac{1}{p}\mathbb{Z} & \frac{1}{p}\mathbb{Z} & \mathbb{Z}\end{pmatrix}\right\}$$

$$\Gamma_{1,p} = \left\{ X \in \Gamma^0_{1,p},\ X - \mathbb{1}_4 \in \begin{pmatrix} \mathbb{Z} & \mathbb{Z} & \mathbb{Z} & p\mathbb{Z} \\ p\mathbb{Z} & p\mathbb{Z} & p\mathbb{Z} & p^2\mathbb{Z} \\ \mathbb{Z} & \mathbb{Z} & \mathbb{Z} & p\mathbb{Z} \\ \mathbb{Z} & \mathbb{Z} & \mathbb{Z} & p\mathbb{Z} \end{pmatrix} \right\}$$

$${}_0\Gamma_{1,p} = \Gamma^0_{1,p} \cap \Gamma' \ .$$

As we shall see shortly the quotients of $\mathfrak{S}_2$ by these groups are moduli spaces. But first we need a lemma on the relation of the groups themselves.

Lemma 0.5. (i) *There is a tower of groups with indices as shown:*

$$\begin{tikzcd}
\Gamma^0_{1,p} \arrow[dr, dash, "p+1"] & & \Gamma' = Sp(4,\mathbb{Z}) \arrow[dl, dash, "(p+1)(p^2+1)"] \\
& \Gamma^0_{1,p} \cap \Gamma' = {}_0\Gamma_{1,p} \arrow[d, dash, "p(p-1)"] & \\
& \Gamma_{1,p} &
\end{tikzcd}$$

(ii) $\Gamma_{1,p} \lhd \Gamma^0_{1,p}$ *with quotient* $SL(2,\mathbb{Z}_p)$

$\Gamma_{1,p} \lhd {}_0\Gamma_{1,p}$ *with quotient* $\left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in SL(2,\mathbb{Z}_p) \right\}$

Proof. It is convenient to transform the problem so that we are working with integer matrices. To this end let F be the diagonal matrix $F = \mathrm{Diag}(1,1,1,p)$, let $\tilde{\Gamma}^0_{1,p} = F\Gamma^0_{1,p}F^{-1}$, and similarly for the other groups. Then $\tilde{\Gamma}^0_{1,p} = \tilde{S}p(4,\mathbb{Z})$, the group of integer matrices symplectic with respect to the matrix FJ^tF. This matrix gives the inner product

$$\langle (v_1, v_2, v_3, v_4), (w_1, w_2, w_3, w_4) \rangle = \det \begin{pmatrix} v_1 & v_3 \\ w_1 & w_3 \end{pmatrix} + p \cdot \det \begin{pmatrix} v_2 & v_4 \\ w_2 & w_4 \end{pmatrix} .$$

Let $X = (x_{ij})$ be an element of $\tilde{S}p(4,\mathbb{Z})$. Observe that $x_{21} \equiv x_{23} \equiv x_{41} \equiv x_{43} \equiv 0 \pmod{p}$. To see this let x_i denote row i of X. Then X in $\tilde{S}p(4,\mathbb{Z})$ implies

$$\langle x_1, x_3 \rangle = 1 \equiv \det \begin{pmatrix} x_{11} & x_{13} \\ x_{31} & x_{33} \end{pmatrix} \pmod{p}$$

and

$$\langle x_i, x_j \rangle = 0 \equiv \det \begin{pmatrix} x_{i1} & x_{i3} \\ x_{j1} & x_{j3} \end{pmatrix} \pmod{p} \quad \text{for} \quad i = 1, 3,\ j = 2, 4$$

The first of these equations shows that (x_{11}, x_{13}) and (x_{31}, x_{33}), regarded as vectors in $(\mathbb{Z}_p)^2$, are linearly independent, and then the second shows that (x_{21}, x_{23}) and (x_{41}, x_{43}), regarded as vectors in $(\mathbb{Z}_p)^2$, are both zero, as claimed.

Thus for $X \in \tilde{\Gamma}^0_{1,p}$,

$$X \equiv \begin{pmatrix} * & * & * & * \\ 0 & a & 0 & b \\ * & * & * & * \\ 0 & c & 0 & d \end{pmatrix} \pmod{p} ,$$

where an entry marked $*$ is allowed to be arbitrary. Then direct calculation shows

$$\tilde{\Gamma}_{1,p} = \left\{ X \in \tilde{\Gamma}^0_{1,p},\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{p} \right\}$$

and

$${}_0\tilde{\Gamma}_{1,p} = \{ X \in \tilde{\Gamma}^0_{1,p},\ c \equiv 0 \pmod{p} \} .$$

The map

$$\tilde{\Gamma}^0_{1,p} \to SL(2, \mathbb{Z}_p)$$

given by

$$M \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \pmod{p}$$

is easily checked to be a homomorphism, so its kernel, $\tilde{\Gamma}_{1,p}$, is a normal subgroup. Since there is an inclusion $SL(2, \mathbb{Z}) \to \tilde{\Gamma}^0_{1,p}$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} 1 & & & \\ & a & & b \\ & & 1 & \\ & c & & d \end{pmatrix}$$

it is then immediate that the quotient $\tilde{\Gamma}^0_{1,p}/\tilde{\Gamma}_{1,p}$ is isomorphic to $SL(2, \mathbb{Z}_p)$, of order $p(p^2-1)$. By inspection ${}_0\tilde{\Gamma}_{1,p}/\tilde{\Gamma}_{1,p}$ is then seen to be as claimed, and this group has order $p(p-1)$, giving the indices as claimed.

We are left with computing $[\Gamma' : {}_0\Gamma_{1,p}]$. We see

$$\Gamma' \cap \Gamma^0_{1,p} = \left\{ X \in Sp(4, \mathbb{Z}),\ X \equiv \begin{pmatrix} * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & 0 \\ * & * & * & d \end{pmatrix} \pmod{p} \right\}$$

where of course d is not congruent to zero mod p. This group obviously has the corresponding subgroup where $d \equiv 1 \pmod{p}$ as a subgroup of index $p-1$. On the other hand, $Sp(4, \mathbb{Z})$ acts transitively on the p^4-1 non-zero elements of $(\mathbb{Z}_p)^4$, and this latter subgroup is the stabilizer of one of them, hence has index p^4-1, and so $[\Gamma' : \Gamma' \cap \Gamma^0_{1,p}] = (p^4-1)/(p-1) = (p+1)(p^2+1)$.

We then obtain a diagram

$$\begin{tikzcd} \mathscr{A}^0_{1,p}=\mathfrak{S}_2/\Gamma^0_{1,p} & & \mathscr{A}'=\mathfrak{S}_2/\Gamma' \\ & {}_0\mathscr{A}_{1,p}=\mathfrak{S}_2/{}_0\Gamma_{1,p} \arrow[ul] \arrow[ur] & \\ & \mathscr{A}_{1,p}=\mathfrak{S}_2/\Gamma_{1,p} \arrow[u] & \end{tikzcd}$$

in which the arrows indicate branched covering maps. The covers $\mathscr{A}_{1,p}\to\mathscr{A}^0_{1,p}$ and $\mathscr{A}_{1,p}\to{}_0\mathscr{A}_{1,p}$ are Galois coverings with groups given by Lemma 0.5 (ii). Note, however, that the center $\{\pm\mathbb{1}_4\}$ of $Sp(4,\mathbb{Q})$ acts trivially on $\mathfrak{S}_2$. The non-trivial element of the center is in ${}_0\Gamma_{1,p}$ but not $\Gamma_{1,p}$. Hence the degree of the cover $\mathscr{A}_{1,p}\to{}_0\mathscr{A}_{1,p}$ is $p(p-1)/2$, and similarly for the other covers, and the groups which act as the effective Galois groups are the quotients of the above-mentioned groups by their centers of order 2.

Expressed differently, the action of $Sp(4,\mathbb{Q})$ on $\mathfrak{S}_2$ factors through the projective group $PSp(4,\mathbb{Q})$. We prefer to work with$Sp(4,\mathbb{Q})$, so that we may write down matrices unambiguously, but this preference accounts for many of the $\pm$ signs in the sequel.

These spaces are moduli spaces and have the following interpretations:

$\mathscr{A}'=\{(B,H_B),\ B$ is an abelian surface; H_B is a principal polarization$\}$

$\mathscr{A}^0_{1,p}=\{(A,H_A),\ A$ is an abelian surface; H_A is a polarization of type $(1,p)\}$

$\mathscr{A}_{1,p}=\{(A,H_A,\alpha),\ \alpha$ is a level-p structure$\}$

${}_0\mathscr{A}_{1,p}=\{(\mathrm{A},\mathrm{H}_A,\pi),\ \pi$ is a root$\}$.

The first three of these are standard (see [I1] or [H]); we prove the fourth.

Proposition 0.6. *${}_0\mathscr{A}_{1,p}$ is the moduli space of rooted abelian surfaces with polarization of type* $(1,p)$.

Proof. Given a point t in $\mathscr{A}_{1,p}$, we obtain a rooted abelian surface as follows: The point t corresponds to (A,H_A,α) with α an isomorphism $\alpha: L^\vee/L\to(\mathbb{Z}_p)^2$: Here L is a lattice defining A and $L^\vee$ is its dual. Let K be the subspace $\{(0,k)\}\subset(\mathbb{Z}_p)^2$. Then $\alpha^{-1}(K)$ is a subgroup of $A^{(p)}$, the group of p-torsion points of A. Set $B=A/\alpha^{-1}(K)$ with the obvious projection π and polarization H_B. Thus we have a map

$$\mathscr{A}_{1,p}\to\mathscr{M}$$

where $\mathscr{M}$ is the desired moduli space. Note that t_1 and t_2 in $\mathscr{A}_{1,p}$ will yield the same rooted abelian surface if and only if they define the same polarized abelian surface (A,H_A) and the corresponding level structures α_1 and α_2 satisfy $\alpha_1^{-1}(K)=\alpha_2^{-1}(K)$, i.e. $\alpha_2\alpha_1^{-1}(K)=K$.

Now $\alpha_2\alpha_1^{-1}\in SL(2,\mathbb{Z}_p)$, and the stabilizer of K in $SL(2,\mathbb{Z}_p)$ is exactly the group $\left\{\begin{pmatrix} a & b\\ 0 & d\end{pmatrix}\right\}$, so $\mathscr{M}$ is the quotient of $\mathscr{A}_{1,p}$ by the action of this group,

$$\mathscr{M}=\mathscr{A}_{1,p}\Big/\left\{\begin{pmatrix} a & b\\ 0 & d\end{pmatrix}\in SL(2,\mathbb{Z}_p)\right\}={}_0\mathscr{A}_{1,p}$$

by Lemma 0.5.

Remark 0.7. A point t in ${}_0\mathscr{A}_{1,p}$ corresponds to $\pi:(A,H_A)\to(B,H_B)$. The maps ${}_0\mathscr{A}_{1,p}\to\mathscr{A}^0_{1,p}$ and ${}_0\mathscr{A}_{1,p}\to\mathscr{A}'$ are the obvious forgetful maps taking this pair to (A,H_A) and (B,H_B) respectively.

The degree of the cover ${}_0\mathscr{A}_{1,p}\to\mathscr{A}^0_{1,p}$ being $(p+1)$ corresponds to the fact that $(\mathbb{Z}_p)^2$ has $(p+1)$ subspaces isomorphic to $\mathbb{Z}_p$.

For fixed p we define the following surfaces in $\mathfrak{S}_2$:

$$\mathfrak{H}_1=\left\{\tau=\begin{pmatrix}\tau_1 & \tau_2\\ \tau_2 & \tau_3\end{pmatrix}\in\mathfrak{S}_2,\quad \tau_2=0\right\}$$

$$\mathfrak{H}_2=\left\{\tau=\begin{pmatrix}\tau_1 & \tau_2\\ \tau_2 & \tau_3\end{pmatrix}\in\mathfrak{S}_2,\quad p\tau_1-2\tau_2=0\right\}.$$

These are Humbert surfaces in the sense of [F], [vG]. Their discriminant is $\Delta(\mathfrak{H}_1)=1$, resp. $\Delta(\mathfrak{H}_2)=4$. We denote their images under the natural projections from $\mathfrak{S}_2$ to $\mathscr{A}_{1,p}$, ${}_0\mathscr{A}_{1,p}$, $\mathscr{A}^0_{1,p}$, and $\mathscr{A}'$ by $\mathscr{H}_i$, ${}_0\mathscr{H}_i$, $\mathscr{H}_i^0$, and $\mathscr{H}_i'$ respectively ($i=1,2$). While $\mathscr{H}_2'$ in $\mathscr{A}'$ may appear to depend on p, this is in fact not the case; under the action of the element

$$\begin{pmatrix}\frac{1-p}{2} & 1 & & \\ -\frac{1+p}{2} & 1 & & \\ & & 1 & \frac{1+p}{2}\\ & & -1 & \frac{1-p}{2}\end{pmatrix}$$

of $\Gamma'=Sp(4,\mathbb{Z})$ the space $\mathfrak{H}_2$ is taken to

$$\mathfrak{H}_3=\left\{\tau=\begin{pmatrix}\tau_1 & \tau_2\\ \tau_2 & \tau_3\end{pmatrix}\in\mathfrak{S}_2,\quad \tau_1-\tau_3=0\right\}$$

which is evidently independent of p.

We can now formulate our main result as follows:

Theorem 0.8. *Let*

$$\tilde{\mathscr{A}}_{1,p}:=\mathscr{A}_{1,p}\backslash(\mathscr{H}_1\cup\mathscr{H}_2).$$

Then

$$\tilde{\mathscr{A}}_{1,p}=\{(A,H_A,\alpha)\in\mathscr{A}_{1,p},\ H_A \text{ is very ample}\}$$

i.e. $\tilde{\mathscr{A}}_{1,p}$ *is the moduli space of abelian surfaces with a very ample polarization of type* $(1,p)$ *and level-p structure.*

As an immediate consequence we get

Corollary 0.9. *Let*

$$\tilde{\mathscr{A}}^0_{1,p}:=\mathscr{A}^0_{1,p}\backslash(\mathscr{H}^0_1\cup\mathscr{H}^0_2).$$

Then

$$\tilde{\mathscr{A}}^0_{1,p} = \{(A, H_A) \in \mathscr{A}^0_{1,p},\ H_A \text{ is very ample}\}$$

i.e. $\tilde{\mathscr{A}}^0_{1,p}$ *is the moduli space of abelian surfaces with a very ample polarization of type* $(1,p)$.

Remark 0.10. The analogue for ${}_0\tilde{\mathscr{A}}_{1,p} = {}_0\mathscr{A}_{1,p} \setminus ({}_0\mathscr{H}_1 \cup {}_0\mathscr{H}_2)$ obviously holds as well.

Remark 0.11. We shall see that $\mathscr{H}_1^0$ parametrizes the abelian surfaces which split as polarized abelian surfaces, whereas $\mathscr{H}_2^0$ parametrizes the abelian surfaces described in Proposition 0.1 and Proposition 0.2 (ii).

1. Bielliptic Abelian Surfaces

We first consider principally polarized abelian surfaces (B, H_B) and write them in the form

$$B = \mathbb{C}^2 / L_\tau$$

where τ is an element in the Siegel space $\mathfrak{S}_2$ and the lattice L_τ is generated by the rows of the matrix $\begin{pmatrix} \tau \\ \mathbb{1}_2 \end{pmatrix}$ such that the polarization H_B with respect to this basis is given by the matrix J. An involution of the principally polarized abelian surface (B, H_B) is an involution $j: B \to B$ with $j^*(H_B) = H_B$. Every principally polarized abelian surface admits the involution $\iota: x \mapsto -x$. Now assume that (B, H_B) admits an additional involution $j \neq \iota$. Then j induces a symplectic automorphism of the lattice L_τ, i.e. it defines an element $X(j) \in Sp(4, \mathbb{Z})$. The involutions in $Sp(4, \mathbb{Z})$ are well known and it follows from [U, p. 198] that up to conjugation we are in one of the following two cases:

$$\pm X(j) = T := \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & -1 & \\ & & & 1 \end{pmatrix} \tag{1}$$

and

$$\tau \in \operatorname{Fix} T = \left\{ \tau = \begin{pmatrix} \tau_1 & \tau_2 \\ \tau_2 & \tau_3 \end{pmatrix} \in \mathfrak{S}_2, \quad \tau_2 = 0 \right\} = \mathfrak{H}_1 \ .$$

Remark 1.1. In this case (B, H_B) splits and the involution on $B = E_1 \times E_2$ is given by $j = \pm(-id_{E_1}, id_{E_2})$.

$$\pm X(j) = U := \begin{pmatrix} 0 & 1 & & \\ 1 & 0 & & \\ & & 0 & 1 \\ & & 1 & 0 \end{pmatrix} \tag{2}$$

and

$$\tau \in \operatorname{Fix} U = \left\{ \tau = \begin{pmatrix} \tau_1 & \tau_2 \\ \tau_2 & \tau_3 \end{pmatrix} \in \mathfrak{S}_2, \quad \tau_1 - \tau_3 = 0 \right\} = \mathfrak{H}_3 \ .$$

Definition 1.2. An involution of type (2) will be called an *elliptic involution.* If (B, H_B) admits an elliptic involution we call it a *bielliptic* principally polarized abelian surface.

Remark 1.3. Bolza [B] and Igusa [I2] studied genus 2 curves with non-trivial automorphisms. In this context they found the above surfaces. They are precisely the Jacobians of genus 2 curves C which admit an elliptic involution, i.e. a 2:1 cover onto an elliptic curve E. For a simple proof of this see Proposition 4.2. If C is reducible this means the $B = E \times E$ and $C = E \times \{0\} \cup \{0\} \times E$. The elliptic involution on C is then just given by the obvious map to E. These are the abelian surfaces corresponding to the intersection $\mathfrak{H}_1 \cap \mathfrak{H}_3$.

We shall now consider polarized abelian surfaces (A, H_A) where H_A is a polarization of type $(1,p)$. Here p is an odd prime. As before we can write

$$A = \mathbb{C}^2/L_\tau$$

with $\tau \in \mathfrak{S}_2$ and L_τ the lattice spanned by the rows of the matrix $\begin{pmatrix} \tau & \\ 1 & 0 \\ 0 & p \end{pmatrix}$ such that H_A with respect to this basis is given by the matrix

$$\begin{pmatrix} & & 1 & 0 \\ & & 0 & p \\ -1 & 0 & & \\ 0 & -p & & \end{pmatrix}$$

The dual lattice L_τ^v is given by

$$L_\tau^v = \{x, H_A(x,y) \in \mathbb{Z} \text{ for all } y \in L_\tau\} .$$

Definition 1.4. Involutions j_A of (A, H_A) and j_B of (B, H_B) (with $j_A \neq \iota, j_B \neq \iota$) form an *involutory pair* (j_A, j_B) if there is a root $\pi: (A, H_A) \to (B, H_B)$ of (A, H_A) making the following diagram commute:

$$\begin{array}{ccc} A & \xrightarrow{j_A} & A \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ B & \xrightarrow{j_B} & B \end{array} \tag{D}$$

Lemma 1.5: (i) *Let $j_A \neq \iota$ be an involution of (A, H_A), A an abelian surface with polarization H_A of type $(1,p)$. Then j_A is contained in an involutory pair (j_A, j_B).*

(ii) *Let $j_B \neq \iota$ be an involution of (B, H_B), B an abelian surface with principal polarization H_B. Then j_B is contained in an involutory pair (j_A, j_B).*

Proof. (i) The involution j_A defines an involution on L_τ which we shall also denote by j_A. Since j_A leaves the form H_A on L_τ invariant it maps $L_\tau^\vee$ into itself. Hence j_A induces an involution $\bar{j}_A$ on $L_\tau^\vee/L_\tau$ which is symplectic with respect to the form induced on this quotient. Choosing a sympletic basis of $L_\tau^\vee/L_\tau$ we can consider $\bar{j}_A$ as an element in $Sp(2, \mathbb{Z}_p) = SL(2, \mathbb{Z}_p)$. The only involutions in $SL(2, \mathbb{Z}_p)$ are $\pm id$. Hence $\bar{j}_A$ leaves every subgroup $\mathbb{Z}_p \subset L_\tau^\vee/L_\tau$ invariant and dividing out by such a subgroup gives the desired diagram (D).

(ii) Let $B^{(p)}$ be the group of p-torsion points of B. The involution j_B defines a decomposition $B^{(p)} = K_+ \oplus K_-$, where $K_\pm$ are the eigenspaces for ± 1. Note that

$K_+ \simeq K_- \simeq (\mathbb{Z}_p)^2$. Every $x \in K_+$ or K_- defines a covering $A \to B$. Let H_A be the pullback of H_B to A. Then j_B lifts to an involution j_A on A over j_B with $j^*(H_A) = H_A$.

Remark 1.6. Note that we have shown that if (A, H_A) admits an involution and $\pi : (A, H_A) \to (B, H_B)$ is *any* root of (A, H_A) so does (B, H_B). On the other hand we have seen that beginning with (B, H_B) admitting an involution exactly $2(p+1)$ of the $(p+1)(p^2+1)$ choices of (A, H_A) with $\pi : (A, H_A) \to (B, H_B)$ admit such an involution.

From what we have said above there are now the following two possibilities:

(1) j_B is an involution of type (1). Then (B, H_B) splits, i.e. $B = E_1 \times E_2$ and $j_B = \pm(-\mathrm{id}_{E_1}, \mathrm{id}_{E_2})$. The covering $A \to B$ corresponds to a p-torsion point $\alpha = (\alpha_1, \alpha_2) \in E_1 \times E_2$. Since j_B lifts to the involution j_A it follows that $\alpha_1 = 0$ or $\alpha_2 = 0$. Then (A, H_A) splits as a polarized abelian surface.

(2) j_B is an elliptic involution.

Remark 1.7. Whether case (1) or (2) occurs does not depend on the choice of the map $\pi : A \to B$. This follows e.g. from the fact that in case (2) the linear system $|H_A|$ is base point free [R], [HL].

Definition 1.8. If a diagram (D) exists such that j_B is an elliptic involution we shall also call j_A an *elliptic involution*. If (A, H_A) admits an elliptic involution we shall call it a *bielliptic* polarized abelian surface.

Remark 1.9. By Remark 1.3 the bielliptic polarized abelian surfaces are just the surfaces described in Proposition 0.1 and Proposition 0.2 (ii).

2. Computations

We recall the Humbert surface

$$\mathfrak{H}_1 = \mathrm{Fix}\, T = \{\tau, \tau_2 = 0\} .$$

We now want to introduce the involution

$$S = \begin{pmatrix} -1 & 0 & & \\ -p & 1 & & \\ & & -1 & -p \\ & & 0 & 1 \end{pmatrix}$$

Straightforward calculation shows

$$\mathrm{Fix}\, S = \{\tau \in \mathfrak{S}_2,\ p\tau_1 - 2\tau_2\} = \mathfrak{H}_2$$

and we have seen that $\mathfrak{H}_2$ and $\mathfrak{H}_3$ are equivalent under $Sp(4, \mathbb{Z})$ (although not under $\Gamma^0_{1,p}$).

In this section we want to compute the stabilizer subgroups

$$P_i^0 := \{g \in \Gamma^0_{1,p},\ g(\mathfrak{H}_i) = \mathfrak{H}_i\}$$

resp.

$$P_i = P_i^0 \cap \Gamma_{1,p} = \{g \in \Gamma_{1,p},\ g(\mathfrak{H}_i) = \mathfrak{H}_i\}\ (i = 1, 2) .$$

Proposition 2.1. *$P_i^0/P_i \simeq SL(2,\mathbb{Z}_p)$ for $i=1,2$.*

Proof. We first treat the case $i=1$. Let

$$P_1^{\mathbb{Q}} := \{g \in Sp(4,\mathbb{Q}),\ g(\mathfrak{H}_1)=\mathfrak{H}_1\} \ .$$

It follows e.g. from Franke [F, Lemma 3.2.6] that

$$P_1^{\mathbb{Q}} = \left\{ g \in Sp(4,\mathbb{Q}),\ g = \begin{pmatrix} a & 0 & b & 0 \\ 0 & a' & 0 & b' \\ c & 0 & d & 0 \\ 0 & c' & 0 & d' \end{pmatrix} \quad \text{or} \quad g = \begin{pmatrix} 0 & a & 0 & b \\ a' & 0 & b' & 0 \\ 0 & c & 0 & d \\ c' & 0 & d' & 0 \end{pmatrix} \right.$$

$$\left. \text{with} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in SL(2,\mathbb{Q}) \right\} .$$

An element of the form

$$\begin{pmatrix} 0 & a & 0 & b \\ a' & 0 & b' & 0 \\ 0 & c & 0 & d \\ c' & 0 & d' & 0 \end{pmatrix}$$

cannot be in $\Gamma_{1,p}^0$ since this would imply $a, c \in \mathbb{Z}$, $b, d \in p\mathbb{Z}$ a contradiction to $ad-bc=1$. Hence the inclusion

$$\phi_1 : SL(2,\mathbb{Z}) \times SL(2,\mathbb{Z}) \to \Gamma_{1,p}^0$$

$$\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \right) \mapsto \begin{pmatrix} a & 0 & b & 0 \\ 0 & a' & 0 & pb' \\ c & 0 & d & 0 \\ 0 & \frac{c'}{p} & 0 & d' \end{pmatrix}$$

gives an isomorphism

$$P_1^0 \simeq SL(2,\mathbb{Z}) \times SL(2,\mathbb{Z}) \ .$$

An element $\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \right)$ gives rise to an element in $\Gamma_{1,p}$ if and only if

$$\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in \Gamma_1(p) = \{M \in SL(2,\mathbb{Z}),\ M \equiv \mathbb{1} \bmod p\} \ .$$

Hence

$$P_1^0/P_1 \simeq SL(2,\mathbb{Z})/\Gamma_1(p) \simeq SL(2,\mathbb{Z}_p) \ .$$

We now treat the case $i=2$. Consider the matrix

$$g_0=\begin{pmatrix}1 & 0 & & \\ \frac{p}{2} & 1 & & \\ & & 1 & -\frac{p}{2} \\ & & 0 & 1\end{pmatrix}\in Sp(4,\mathbb{Q}) .$$

Then

$$g_0 T g_0^{-1}=S$$

and hence

$$P_2^{\mathbb{Q}}=\{g\in Sp(4,\mathbb{Q}),\ g(\mathfrak{H}_2)=\mathfrak{H}_2\}=g_0 P_1^{\mathbb{Q}} g_0^{-1} .$$

For an element

$$g=\begin{pmatrix}0 & a & 0 & b\\ a' & 0 & b' & 0\\ 0 & c & 0 & d\\ c' & 0 & d' & 0\end{pmatrix}\in P_1^{\mathbb{Q}}$$

we have

$$g_0 g g_0^{-1}=\begin{pmatrix}-\frac{p}{2}a & a & 0 & b\\ a'-\frac{p^2}{4}a & \frac{p}{2}a & b' & \frac{p}{2}(b+b')\\ -\frac{p}{2}(c+c') & c & -\frac{p}{2}d' & d-\frac{p^2}{4}d'\\ c' & 0 & d' & \frac{p}{2}d'\end{pmatrix}$$

Such an element cannot be in $\Gamma_{1,p}^0$ since this would imply $a,c\in\mathbb{Z}$; $b\in p\mathbb{Z}$ and $d'\in\frac{2}{p}\mathbb{Z}$ hence $d\in\frac{p}{2}\mathbb{Z}$ which contradicts $ad-bc=1$.

In order to determine P_2^0 and P_2 we now consider the inclusion

$$\phi_2: SL(2,\mathbb{Q})\times SL(2,\mathbb{Q})\to P_1^{\mathbb{Q}}$$

$$\left(\begin{pmatrix}a & b\\ c & d\end{pmatrix},\begin{pmatrix}a' & b'\\ c' & d'\end{pmatrix}\right)\mapsto\begin{pmatrix}a & 0 & 2b & 0\\ 0 & a' & 0 & \frac{p}{2}b'\\ \frac{c}{2} & 0 & d & 0\\ 0 & \frac{2}{p}c' & 0 & d'\end{pmatrix}$$

We then have

$$g_0\begin{pmatrix} a & 0 & 2b & 0 \\ 0 & a' & 0 & \frac{p}{2}b' \\ \frac{c}{2} & 0 & d & 0 \\ 0 & \frac{2}{p}c' & 0 & d' \end{pmatrix} g_0^{-1} = \begin{pmatrix} a & 0 & 2b & pb \\ \frac{p}{2}(a-a') & a' & pb & \frac{p}{2}(pb+b') \\ \frac{1}{2}(pc'+c) & -c' & d & \frac{p}{2}(d-d') \\ -c' & \frac{2}{p}c' & 0 & d' \end{pmatrix}$$

This matrix is in $\Gamma^0_{1,p}$ if and only if

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in SL(2,\mathbb{Z}) ,$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \bmod 2 .$$

Hence

$$P_2^0 \simeq \{(M,N) \in SL(2,\mathbb{Z}) \times SL(2,\mathbb{Z}),\ M \equiv N \bmod 2\} .$$

Moreover a pair (M,N) gives rise to an element in $\Gamma_{1,p}$ if and only if $N \in \Gamma_1(p)$. This shows that

$$P_2 \simeq \{(M,N) \in SL(2,\mathbb{Z}) \times \Gamma_1(p),\ M \equiv N \bmod 2\}$$

and hence

$$P_2^0/P_2 \simeq SL(2,\mathbb{Z})/\Gamma_1(p) \simeq SL(2,\mathbb{Z}_p) .$$

Now let X be a set and assume that the group G^0 operates transitively on X. Let G be a normal subgroup of G^0 of finite index. For $x_0 \in X$ we consider the stabilizer subgroups

$$P^0 = \{g \in G^0,\ g(x_0) = x_0\}$$

resp.

$$P = P^0 \cap G = \{g \in G,\ g(x_0) = x_0\} .$$

Lemma 2.2. *If* $[G^0 : G] = [P^0 : P]$ *then* G *acts transitively on* X.

Proof. Since G^0 acts transitively on X we have an identification

$$X = G^0/P^0 .$$

Since G is normal we have

$$G\backslash X = G\backslash(G^0/P^0) = (G^0/P^0)/G = G^0/P^0G .$$

On the other hand

$$[G^0 : P^0G]\,[P^0G : G] = [G^0 : G] = [P^0 : P] .$$

Since

$$[P^0G : G] \geqq [P^0 : P^0 \cap G] = [P^0 : P]$$

this implies $[G^0 : P^0G] = 1$, hence $\#(G\backslash X) = 1$, i.e. G acts transitively.

We now return to the quotient map

$$\sigma : \mathscr{A}_{1,p} \to \mathscr{A}^0_{1,p}$$

which is given by the natural action of the group

$$\Gamma^0_{1,p}/\Gamma_{1,p} \simeq SL(2, \mathbb{Z}_p) \ .$$

Proposition 2.3. *$\sigma^{-1}(\mathscr{H}_i^0) = \mathscr{H}_i$ for $i=1,2$.*

Proof. Clearly $\mathscr{H}_i \subset \sigma^{-1}(\mathscr{H}_i^0)$. What we have to see is that $\sigma^{-1}(\mathscr{H}_i^0)$ consists of only one component. In order to see this we consider for fixed i the following set of Humbert surfaces

$$X = \{g(\mathfrak{H}_i),\ g \in \Gamma^0_{1,p}\} \ .$$

We have to show that $\Gamma_{1,p}$ operates transitively on X. But this is now an immediate consequence of Proposition 2.1 and Lemma 2.2.

Let ${}_0\sigma : \mathscr{A}_{1,p} \to \mathscr{A}^0_{1,p}$ be the quotient map.

Corollary 2.4. *${}_0\sigma^{-1}(\mathscr{H}_i^0) = {}_0\mathscr{H}_i$ for $i=1,2$.*

Proof. ${}_0\mathscr{H}_i$ is the image of $\mathscr{H}_i$ under the quotient map $\mathscr{A}_{1,p} \to {}_0\mathscr{A}_{1,p}$. Since $\mathscr{H}_i$ has only one component, the same holds for ${}_0\mathscr{H}_i$ $(i=1,2)$.

Remark 2.5. If ${}_0P_i = P_i^0 \cap {}_0\Gamma_{1,p}$, then

$${}_0P_i/P_i \simeq \left\{\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in SL(2, \mathbb{Z}_p)\right\} \quad \text{for} \quad i=1,2 \ .$$

A similar computation to 2.1 shows

Proposition 2.6. *Let P_i' be the stabilizer of $\mathfrak{H}_i$ in $Sp(4, \mathbb{Z})$. Then $[P_i' : {}_0P_i] = 2(p+1)$ for $i=1,2$.*

Proof. In the formulas for ϕ_1 and ϕ_2 in the proof of 2.1 set $p=1$ to obtain new maps ψ_1 and ψ_2, and use ψ_i in place of ϕ_i. If we let $Q = \{X = (x_{ij}) \in Sp(4, \mathbb{Q}),\ x_{ij} = 0$ for $i+j$ odd$\}$, and $P_1'' = P_1' \cap Q$, then obviously $[P_1' : P_1''] = 2$, and we have

$$P_1' = P_1'' \cup h_0 P_1'' \ \text{for}\ h_0 = \begin{pmatrix} & & & 1 \\ & & 1 & \\ & -1 & & \\ -1 & & & \end{pmatrix}$$

Now

$$\begin{aligned} P_2' &= g_0(P_1^{\mathbb{Q}})g_0^{-1} \cap Sp(4, \mathbb{Z}) \\ &= g_0(Q \cup h_0 Q)g_0^{-1} \cap Sp(4, \mathbb{Z}) \\ &= (g_0 Q g_0^{-1} \cap Sp(4, \mathbb{Z})) \cup (h_1(g_0 Q g_0^{-1}) \cap Sp(4, \mathbb{Z})) \end{aligned}$$

where $h_1 = g_0 h_0 g_0^{-1}$. Defining $P_2'' = g_0 Q g_0^{-1} \cap Sp(4, \mathbb{Z})$, we see that, since $h_1 \in Sp(4, \mathbb{Z})$, we also have $[P_2' : P_2''] = 2$.

Now, following the proof of 2.1, we see that ${}_0P_i \subset P_i''$, $i=1,2$. Then the rest of the proof goes through unchanged, except at the end we find, for $i=1,2$,

$$P_i''/{}_0P_i \simeq SL(2,\mathbb{Z})/\left\{\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in SL(2,\mathbb{Z}),\ b' \equiv 0 \ (\operatorname{mod} p)\right\}$$

and this subgroup has index $p+1$ in $SL(2,\mathbb{Z})$.

Thus $[P_i' : {}_0P_i] = [P_i' : P_i''][P_i'' : {}_0P_i] = 2(p+1)$ as claimed.

3. Proof of the Main Result

In this section we want to prove Theorem 0.8. The main problem is to study the bielliptic polarized abelian surfaces. Recall from Sects. 1 and 2 that every bielliptic principally polarized abelian surface can be given by a period matrix of the form $\begin{pmatrix} \tau \\ \mathbb{1}_2 \end{pmatrix}$ where $\tau \in \mathfrak{H}_2$.

Proposition 3.1 *The bielliptic abelian surfaces with a polarization of type* $(1,p)$ *and level-p structure form an irreducible* 2*-dimensional family. They are parametrized by the surface* $\mathscr{H}_2 \subset \mathscr{A}_{1,p}$.

In view of Proposition 2.3 it will be enough to prove this result without level-p structure:

Proposition 3.1. *The bielliptic abelian surfaces with a polarization of type* $(1,p)$ *form an irreducible* 2*-dimensional family. They are parametrized by the surface* $\mathscr{H}_2^0 \subset \mathscr{A}_{1,p}^0$.

Proof. Let $\tau \in \mathfrak{H}_2$. Then the polarized abelian surface A defined by the period matrix $\begin{pmatrix} & \tau & \\ 1 & & 0 \\ 0 & & p \end{pmatrix}$ admits an elliptic involution. To see this note that since $\tau \in \mathfrak{H}_2$ it is of the form $\begin{pmatrix} 2\tau_1 & p\tau_1 \\ p\tau_1 & \tau_3 \end{pmatrix}$. The assertion then follows from the equality

$$\begin{pmatrix} -1 & 0 & & \\ -p & 1 & & \\ & & 1 & -1 \\ & & 0 & 1 \end{pmatrix} \begin{pmatrix} 2\tau_1 & p\tau_1 \\ p\tau_1 & \tau_3 \\ 1 & 0 \\ 0 & p \end{pmatrix} = \begin{pmatrix} 2\tau_1 & p\tau_1 \\ p\tau_1 & \tau_3 \\ 1 & 0 \\ 0 & p \end{pmatrix} \begin{pmatrix} -1 & -p \\ 0 & 1 \end{pmatrix}$$

which shows that $\begin{pmatrix} -1 & -p \\ 0 & 1 \end{pmatrix}$ induces an involution on A which lies over the elliptic involution defined by S on the corresponding principally polarized abelian surface.

We now have to show conversely that every bielliptic abelian surface with a polarization of type $(1,p)$ can be given by a period matrix of the form

$$\Omega_\tau = \begin{pmatrix} & \tau & \\ 1 & & 0 \\ 0 & & p \end{pmatrix}, \quad \tau \in \mathfrak{H}_2 .$$

In order to do this we first observe that ${}_0\mathscr{H}_2$ parameterizes rooted abelian surfaces with an elliptic involution forming part of an involutory pair. This follows as ${}_0\mathscr{H}_2$ is the inverse image of $\mathscr{H}_2^0$ (from 2.4), so that for every (A, H_A, π) in ${}_0\mathscr{H}_2$, $\pi:(A, H_A)\to(B, H_B)$, the surface (A, H_A) admits an elliptic involution j_A (as we have just shown) and then Lemma 1.5 yields j_B. Conversely, given (B, H_B) with an elliptic involution j_B, by Lemma 1.5 and Remark 1.6 there are generically $2(p+1)$ choices for a rooted (A, H_A, π), with $\pi:(A, H_A)\to(B, H_B)$, admitting an elliptic involution. We shall show below that $2(p+1)$ is the degree of the cover ${}_0\mathscr{H}_2\to\mathscr{H}_2'$, and so all of these (A, H_A, π) must be parameterized by points in ${}_0\mathscr{H}_2$.

Now let $t_0\in\mathscr{A}_{1,p}^0$ parameterize an elliptic surface (A, H_A) with polarization of type $(1,p)$ admitting an elliptic involution j_A. Then by Lemma 1.5 there is a point ${}_0t\in{}_0\mathscr{A}_{1,p}$ parameterizing (A, H_A, π) with an elliptic involution forming part of an involutory pair. Thus ${}_0t\in{}_0\mathscr{H}_2$, so $t_0\in\mathscr{H}_2^0$ as claimed.

It remains to compute the degree of the cover ${}_0\mathscr{H}_2\to\mathscr{H}_2'$. By [$U$, p. 198] the set of points $Z=\{z\in\mathfrak{S}_2\}$ whose isotropy group in $\Gamma'=Sp(4,\mathbb{Z})$ is precisely $\{\pm\mathrm{id}, \pm S\}$ forms a Zariski open set of $\mathfrak{H}_2$, which is obviously P_2'-invariant, so we may compute the degree of the cover from this set. By [F, Satz 3.3.6], if $g\in\Gamma'$, $g\notin P_2'$ then $g(Z)\cap Z=\emptyset$. On the other hand, $P_2'/\{\pm\mathrm{id}, \pm S\}$ acts freely on Z, and $\{\pm\mathrm{id}, \pm S\}\subset{}_0P_2$, so the degree of the cover is $[P_2'/\{\pm\mathrm{id}, \pm S\}:{}_0P_2/\{\pm\mathrm{id}, \pm S\}]$ $=[P_2':{}_0P_2]=2(p+1)$ by 2.6.

Proof of Theorem 0.8. This is now straightforward. It remains to show that the polarized abelian surfaces with level-p structure which split are parameterized by $\mathscr{H}_1$. In view of Proposition 2.3 it is enough to prove this without level-p structure. But there the statement is obvious.

4. Geometric Properties of Bielliptic Abelian Surfaces

In this section we shall study some geometric properties of bielliptic abelian surfaces.

Proposition 4.1. *Let (A, H_A) be a bielliptic abelian surface which is either principally polarized or has a polarization of type $(1,p)$. Then A is isogenous to a product; more precisely there exist elliptic curves E and F such that $A=E\times F/\mathbb{Z}_2\times\mathbb{Z}_2$ and such that j_A is induced by* $(\mathrm{id}_E, -\mathrm{id}_F)$.

Proof. We first assume the case of a principal polarization and in order to remain consistent with our previous notation we denote the abelian surface by (B, H_B) and the elliptic involution by j_B. As we know from Sect. 1 we can assume that j_B is induced by the symplectic matrix

$$\pm U=\pm\begin{pmatrix}0&1&&\\1&0&&\\&&0&1\\&&1&0\end{pmatrix}$$

and that B is given by a period matrix of the form $\begin{pmatrix} \tau \\ \mathbb{1}_2 \end{pmatrix}$ with $\tau \in \mathfrak{H}_3 = \operatorname{Fix} U$, i.e. τ is of the form

$$\tau = \begin{pmatrix} \tau_1 & \tau_2 \\ \tau_2 & \tau_1 \end{pmatrix} .$$

On $\mathbb{C}^2$ the involution is given by the linear map $\pm \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ operating on $\mathbb{C}^2$ by multiplication from the right. We shall restrict ourselves to the plus sign. Otherwise in what follows the roles of E_1 and E_2 will have to be interchanged.

Since $\tau \in \mathfrak{H}_3$ we have $\operatorname{Im}(\tau_1 \pm \tau_2) \neq 0$. We consider the elliptic curves

$$E := \mathbb{C}/\mathbb{Z}(\tau_1 + \tau_2) + \mathbb{Z} , \quad F := \mathbb{C}/\mathbb{Z}(\tau_1 - \tau_2) + \mathbb{Z}$$

The maps

$$\mathbb{C} \to \mathbb{C}^2 , \quad z \mapsto (z, z) \text{ resp.}$$

$$\mathbb{C} \to \mathbb{C}^2 , \quad z \mapsto (z, -z)$$

define embeddings of E and F into B such that

$$j_B|E = \mathrm{id}_E, \; j_B|F = -\mathrm{id}_F .$$

E and F intersect transversally in their respective 2-torsion points. From this it follows immediately that

$$B = E \times F/\mathbb{Z}_2 \times \mathbb{Z}_2$$

and that j_B is induced by $(\mathrm{id}_E, -\mathrm{id}_F)$.

Now let (A, H_A) be a bielliptic surface with a polarization of type $(1, p)$. Then there exists a diagram

$$\begin{array}{ccc} A & \xrightarrow{j_A} & A \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ B & \xrightarrow{j_B} & B \end{array}$$

where (B, H_B) is as above and where the covering π is given by a p-torsion point $x \in B$ with $j_B(x) = \pm x$. We shall assume that $j_B(x) = x$. The other case can be reduced to this by replacing j_B by $-j_B$. Then $x \in E$. Let $E' \to E$ be the corresponding covering. We then have a commutative diagram

$$\begin{array}{ccc} E' \times F & \to & E \times F \\ \downarrow & & \downarrow \\ A & \to & B \end{array}$$

such that $A = E' \times F/\mathbb{Z}_2 \times \mathbb{Z}_2$ and that j_A is induced by $(\mathrm{id}_{E'}, -\mathrm{id}_F)$.

In what follows it is useful to treat the case of principally polarized abelian surfaces first. So let (B, H_B) be a bielliptic principally polarized abelian surface with elliptic involution j_B and let E and F be as in Proposition 4.1.

Recall that the invariant e of a $\mathbb{P}_1$-bundle over a curve is defined by $-e = \min\{C_1^2, C_1 \text{ a section}\}$. If the base is an elliptic curve, then $e \geqq -1$, and there is a unique bundle with $e = -1$ [Ha, Theorem V.2.15].

Proposition 4.2. *There exists a curve C on B representing H_B which is mapped to itself by j_B. Moreover there is a commutative diagram*

$$\begin{array}{ccc} C & \xrightarrow{2:1} & \bar{C}=C/j_B \\ \cap\downarrow & & \cap\downarrow \\ B & \xrightarrow{2:1} & \bar{B}=B/j_B \end{array} .$$

$\bar{B}$ is the unique $\mathbb{P}_1$-bundle over E with invariant $e=-1$ and $\bar{C}$ is a section of $\bar{B}$.

Remark 4.3. In particular the double cover $C\to\bar{C}$ defines an elliptic involution on C.

Proof. We first claim that

$$E\cdot H_B=F\cdot H_B=2 .$$

Indeed this follows from

$$H_B((\tau_1+\tau_2,\tau_1+\tau_2),(1,1))=H_B((\tau_1-\tau_2,\tau_1-\tau_2),(1,-1))=2 .$$

In order to find the required curve C we first assume that (B, H_B) is the Jacobian of a smooth curve. By choosing a suitable translate of the theta-divisor we can assume that H_B is represented by a curve C which goes through the origin of B and which is tangent to F [R, Prop. 4.2.]. In particular C intersects E transversally at 0. Since $C\cdot E=2$ it must intersect E in some other point, say P. By construction $j_B(C)$ and C have the same tangent at 0 and both go through P. Hence the assumption $j_B(C)\neq C$ would imply $C^2=j_B(C)\cdot C\geqq 3$, a contradiction. If (B, H_B) is the Jacobian of a reducible genus 2 curve it is sufficient to choose C such that it has its singularity at the origin of B. Since j_B is an elliptic involution on B it must interchange the two components of C.

In order to prove the rest of the statement we look at the elliptic curves

$$F_x:=F+x , \quad x\in E .$$

Then $j_B|F_x=-\mathrm{id}$ for all $x\in E$. The fixed points of $j_B|F_x$ are the four points of $F_x\cap E$. In particular F_x/j_B is a projective line and $\bar{B}=B/j_B$ is a $\mathbb{P}_1$-bundle over the curve

$$E/2\text{-torsion points}\simeq E .$$

Two points on C are identified under j_B if and only if they are the points of intersection $C\cap F_x$, $x\in E$.

Hence $\bar{C}=C/j_B$ becomes a section of $\bar{B}$.

It remains to determine the invariant e of the $\mathbb{P}_1$-bundle $\bar{B}$. Since $C^2=2$ it follows that $\bar{C}^2=1$. Hence e is odd. If $e\geqq 1$ this would imply the existence of a section C_1 of $\bar{B}$ with $C_1^2<0$. But this is impossible since on B no effective divisor exists with negative self-intersection.

Finally let (A, H_A) be a bielliptic abelian surface with a polarization of type $(1,p)$ which lies over a principally polarized bielliptic abelian surface (B, H_B). By Proposition 4.1 we can assume that $B=E\times F/\mathbb{Z}_2\times\mathbb{Z}_2$ and $A=E'\times F/\mathbb{Z}_2\times\mathbb{Z}_2$ where $E'\to E$ is a p-fold covering. Let C be as in Proposition 4.2 and let D be its preimage in A. If C is smooth then D is a smooth curve of genus $p+1$. Otherwise D consists of two elliptic curves intersecting in p points.

Putting

$$\bar{A}=A/j_A\ ,\qquad \bar{B}=B/j_B$$

we have a commutative diagram

$$\begin{tikzcd}
D \arrow[rrr] \arrow[ddd] \arrow[dr, hook] & & & \bar{D} \arrow[ddd] \arrow[dl, hook'] \\
& A \arrow[r, "\pi_A"'] \arrow[d, "\pi"'] & \bar{A} \arrow[d, "\pi"] & \\
& B \arrow[r, "\pi_B"'] & \bar{B} & \\
C \arrow[rrr] \arrow[ur, hook] & & & \bar{C} \arrow[ul, hook']
\end{tikzcd}$$

where all horizontal maps are $2:1$ and all vertical maps are $p:1$. Note that $\bar{D}\simeq E'$ and $\bar{C}\simeq E$. In particular $D\to\bar{D}$ defines an elliptic involution on D. This is precisely the situation described in Proposition 0.1 resp. Proposition 0.2 (ii).

Proposition 4.4 (i) *$\bar{A}$ is the unique $\mathbb{P}_1$-bundle over E' with invariant $e=-1$.*

(ii) *There is a section $\tilde{D}$ of $\bar{A}$ with $\tilde{D}^2=1$ such that*

$$\mathcal{O}_A(H_A)=\pi_A^*\mathcal{O}_{\bar{A}}(\bar{D})=\pi_A^*\mathcal{O}_{\bar{A}}\left(\tilde{D}+\frac{p-1}{2}f\right)$$

where f is a fibre of $\bar{A}$.

(iii) *The linear system associated to H_A is base point free. The associated map $\phi_A: A\to\mathbb{P}_{p-1}$ factors through a map $\phi_{\bar{A}}:\bar{A}\to\mathbb{P}_{p-1}$ which embeds $\bar{A}$ as an elliptic scroll of degree p in $\mathbb{P}_{p-1}$.*

Proof. (i) It follows from Proposition 4.2 that $\bar{B}$ is the unique $\mathbb{P}_1$-bundle over E with invariant $e=-1$. We can write $\bar{B}=\mathbb{P}(\mathscr{E})$ for a suitable rank 2 bundle $\mathscr{E}$ over E. By the above construction $\bar{A}=\mathbb{P}(\pi^*\mathscr{E})$ where $\pi: E'\to E$ is the p-fold covering from above. By [HL, p. 213] $\bar{A}$ is the unique $\mathbb{P}_1$-bundle over E' with invariant $e=-1$.

(ii) Since $e=-1$ we can find a section $\tilde{D}$ of $\bar{A}$ with $\tilde{D}^2=1$ and $\bar{D}\sim\tilde{D}+af$. Since

$$p=\bar{D}^2=(\tilde{D}+af)^2=1+2a$$

it follows that $a=\frac{1}{2}(p-1)$.

(iii) The map $\pi_A: A\to\bar{A}$ defines an inclusion

$$\pi_A^*:\Gamma\left(\mathcal{O}_{\bar{A}}\left(\tilde{D}+\frac{p-1}{2}f\right)\right)\to\Gamma(\mathcal{O}_A(H_A))\ .$$

Both spaces have dimension p, i.e. π_A^* must be an isomorphism. The assertion then follows from the fact that $|\tilde{D}+kf|$ is very ample for $k\geqq 2$.

Acknowledgement. We would like to thank the Deutsche Forschungsgemeinschaft for the support granted under contract HU 337/2-1. The second author would like to thank the National Science Foundation for its support under grant DMS-8604165. This project is part of the Forschungsschwerpunkt "Komplexe Mannigfaltigkeiten".

References

[Be] Beauville, A.: Reider's method for linear systems on surfaces. Letter 1986

[B] Bolza, O.: On binary sextics with linear transformations into themselves. Am. J. Math. **10**, 47–70 (1888)

[F] Franke, H.-G.: Kurven in Hilbertschen Modulflächen und Humbertsche Flächen im Siegelraum. Bonn. Math. Schr. 104 (1978)

[vG] van der Geer, G.: Hilbert modular surfaces. (Ergebnisse der Math., Bd. 16). Berlin Heidelberg New York: Springer 1988

[Ha] Hartshorne, R.: Algebraic Geometry. (Graduate texts in Mathematics, Vol. 52). Berlin Heidelberg New York: Springer 1977

[H] Hulek, K.: Elliptische Kurven, abelsche Flächen und das Ikosaeder. To appear: Jahresbericht der DMY

[HL] Hulek, K., Lange, H.: Examples of abelian surfaces in $\mathbb{P}^4$. J. Reine Angew. Math. **363**, 201–216 (1985)

[I1] Igusa, J.-I.: Theta functions. (Grundlehren Bd. 194). Berlin Heidelberg New York: Springer 1972

[I2] Igusa, J.-I.: Arithmetic variety of moduli of genus two. Ann. Math. **72**, 612–649 (1972)

[R] Ramanan, S.: Ample divisors on abelian surfaces. Proc. Lond. Math. Soc. **51**, 231–245 (1985)

[U] Ueno, K.: On fibre spaces of normally polarized abelian varieties of dimension two, II. Singular fibres of the first kind. J. Fac. Sci. Tokyo **19**, 163–199 (1972)

Received May 18, 1988

Math. Ann. 283, 431–464 (1989)

Mathematische Annalen

Forms Derived from the Arithmetic-Geometric Inequality

Bruce Reznick*

Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801, USA

1. Introduction and Overview

A real homogeneous polynomial (a *form*) p in n variables is *positive semidefinite* or (*psd*) if $p(\underset{\sim}{x}) \geqq 0$ for all $\underset{\sim}{x}=(x_1, \ldots, x_n) \in \mathbb{R}^n$. It is *sos* if it is a *sum of squares* of forms: $p(\underset{\sim}{x})=\sum h_k^2(\underset{\sim}{x})$ for suitable h_k. Every sos form is psd. In 1888, Hilbert [13] proved that not every psd form is sos, but no explicit example was given for nearly eighty years. The set of psd forms $p(x_1, \ldots, x_n)$ of fixed degree comprises a closed convex cone. A psd form p is called *extremal* if it is extremal as an element of this cone: p is extremal if $p=h_1+h_2$, where h_i is a psd form, implies that $h_i=\alpha_i p$ for some $\alpha_i \geqq 0$. Every psd form is a sum of finitely many extremal forms.

In general, it is difficult to determine whether a particular psd form is sos or extremal. Many examples from the literature arise from monomial substitution into the arithmetic-geometric inequality (AGI); we shall call these agiforms. In this paper, we determine a necessary condition for an agiform to be sos. If the monomials are algebraically independent, this condition is sufficient, and we obtain an explicit representation of the agiform as a sum of squares of binomials. We also determine a necessary and sufficient condition for an agiform to be extremal. These expand the pools of known extremal forms and of psd forms which are not sos.

The relevant conditions are always geometric. Associated to each agiform is a polytope with lattice point vertices and a distinguished interior lattice point: the convex hull of the set of exponents used in the substitution, and the exponent of the resulting weighted geometric mean. We shall study the set of lattice points contained in this polytope. For sums of squares, the condition involves writing lattice points as averages of even lattice points. For extremality, the condition involves the parity (mod 2) of the lattice points contained in the polytope.

Hilbert did not carry out his construction in detail, and the first explicit example of a psd form which is not sos was found by Motzkin in 1967. The AGI

* Author supported in part by the Alfred P. Sloan Foundation and the National Science Foundation

(see [12, p. 17]) states:

$$\lambda_1 t_1 + \ldots + \lambda_m t_m - t_1^{\lambda_1} \ldots t_m^{\lambda_m} \geqq 0 \tag{1.1}$$

if $t_i \geqq 0$, $\lambda_i \geqq 0$ and $\sum \lambda_i = 1$. Equality holds in (1.1) if and only if, for some $c \geqq 0$, $\lambda_i > 0$ implies $t_i = c$. Motzkin [16, p. 217] presented the form

$$M(x, y, z) = x^4 y^2 + x^2 y^4 + z^6 - 3x^2 y^2 z^2 , \tag{1.2}$$

which is psd by (1.1), under the substitution $t_1 = x^4 y^2$, $t_2 = x^2 y^4$, $t_3 = z^6$, and $\lambda_i = \frac{1}{3}$ (and multiplication by 3); he showed that M is not sos. The polytope associated to M is the triangle with vertices $(4, 2, 0)$, $(2, 4, 0)$ and $(0, 0, 6)$ and the distinguished point is $(2, 2, 2)$.

Choi, Lam, and the author [3, 4, 17] have derived other psd forms which are not sos from monomial substitutions into the AGI. Two such examples are given in [3, p. 388]:

$$S(x, y, z) = x^4 y^2 + y^4 z^2 + z^4 x^2 - 3x^2 y^2 z^2 , \tag{1.3}$$

$$Q(x, y, z, w) = w^4 + x^2 y^2 + x^2 z^2 + y^2 z^2 - 4xyzw . \tag{1.4}$$

(Not all examples of psd forms which are not sos come from the AGI. Others have been found by Robinson [21], Lax and Lax [15], Schmüdgen [22], and Choi, Lam, and the author [4–7].)

Hurwitz [14] proved in 1891 that

$$G_{2d}(y_1, \ldots, y_{2d}) := y_1^{2d} + \ldots + y_{2d}^{2d} - 2d y_1 \ldots y_{2d} = \sum g_i^2(y_1, \ldots, y_{2d}) \tag{1.5}$$

for appropriate forms g_i. Hurwitz explicitly alluded to Hilbert's recent work as indicating the non-triviality of this representation. The form G_{2d} arises from the substitution $m = 2d$, $\lambda_i = \dfrac{1}{2d}$, $t_i = y_i^{2d}$ into (1.1). In fact, Hurwitz used the sum of squares representation (1.5) to prove the AGI. Let c_i be non-negative integers summing to $2d$. Upon setting c_i of the variables equal to x_i for $i = 1, \ldots, n$, (1.5) becomes:

$$G(\underset{\sim}{c})(\underset{\sim}{x}) := c_1 x_1^{2d} + \ldots + c_n x_n^{2d} - 2d x_1^{c_1} \ldots x_n^{c_n} = \sum g_i^2(x_1, \ldots, x_1, x_2, \ldots, x_n) .$$

Thus, $G(\underset{\sim}{c})$ is sos, and (1.1) is valid when $\lambda_i = \dfrac{1}{2d} c_i$; that is, for all rational $\underset{\sim}{\lambda}$ with $\lambda_i \geqq 0$ and $\sum \lambda_i = 1$. By continuity, (1.1) holds for real $\underset{\sim}{\lambda}$. The polytope associated to $G(\underset{\sim}{c})$ is the simplex with vertices $2d\underset{\sim}{e}_i$, where $\underset{\sim}{e}_i$ is the i-th unit vector, and the distinguished point is $\underset{\sim}{c}$. We shall repeatedly contrast the Motzkin form M and the Hurwitz form

$$H(x, y, z) = x^6 + y^6 + z^6 - 3x^2 y^2 z^2 \tag{1.6}$$

as prototypes of the not-sos and sos agiforms; $2H = G(\underset{\sim}{c})$ for $\underset{\sim}{c} = (2, 2, 2)$.

Following Choi and Lam, [4, p. 1], we let $P_{n,m}$ (resp. $\Sigma_{n,m}$) denote the convex cone of psd (resp. sos) forms in n variables with even degree m and let $\Delta_{n,m} = P_{n,m} \backslash \Sigma_{n,m}$. Hilbert proved that $\Delta_{n,m} = \phi$ if and only if $m = 2$ or $n = 2$ or $(n, m) = (3, 4)$. By identifying an n-ary m-ic form with the M-tuple of its coefficients,

$M=\binom{m+n-1}{m}$, $P_{n,m}$ can be viewed as a cone lying in $\mathbb{R}^M$. From elementary convexity theory, it follows that every psd form is a sum of M extremal forms. [If p is extremal and sos, then clearly p is a perfect square, but not every perfect square is extremal: $(x^2+y^2)^2=(x^2-y^2)^2+(2xy)^2$.] There are few constructions of extremal psd forms in the literature; one general result is found in [2, p. 287]. A product of distinct irreducible indefinite factors is called "purely indefinite". If h is purely indefinite, then h^2 is extremal; if h is purely indefinite and p is extremal, then ph^2 is also extremal. The product of two extremal forms need not be extremal (see [3, p. 402].)

Choi and Lam proved that the agiforms M, S, and Q are extremal, as well as not sos [4, pp. 8–9]. The results inspired [17], in which the author derived the set of extremal psd forms with four or fewer terms. Such a form either is a monomial square, $c(\underline{x}^{\underline{v}})^2$ $(c>0)$ or, after a dilation, arises from a special class of monomial substitutions into (1.1) with $\lambda_i=\frac{1}{3}$.

We introduce some notation. An n-tuple $\underline{u}=(u_1,\ldots,u_n)$ is a *lattice point* if $\underline{u}\in\mathbb{Z}^n$; $\underline{u}$ is an *even* lattice point if $u_j\in 2\mathbb{Z}$, or $\underline{u}=2\underline{v}$, where $\underline{v}$ is a lattice point. For a lattice point $\underline{u}$ with $u_j\geqq 0$, and $\underline{x}\in\mathbb{R}^n$, $\underline{x}^{\underline{u}}$ is the monomial $x_1^{u_1}\ldots x_n^{u_n}$. (When n is small, we name the variables $x, y, z, w, v, u, \ldots$). If $\underline{u}$ is even then $\underline{x}^{\underline{u}}=(\underline{x}^{\underline{v}})^2\geqq 0$ for all $\underline{x}\in\mathbb{R}^n$. We use the term *framework* (and a capital gothic letter) to denote a set $\mathfrak{U}=\{\underline{u}_1,\ldots,\underline{u}_m\}$ of even lattice points in $\mathbb{R}^n$ for which $u_{ij}\geqq 0$ and $\sum_{j=1}^{n} u_{ij}=2d$ for all i and some d. (This last condition ensures that each monomial $\underline{x}^{\underline{u}_i}$ has degree $2d$.) A *trellis* is a framework in which $\underline{u}_1,\ldots,\underline{u}_m$ comprise the vertices of a simplex. (The name is suggested by the horticultural trellis).

We collect some conditions under which a framework is a trellis. Suppose $\sum_{i=1}^{m} c_i\underline{u}_i=\underline{0}$. Then, by summing the coordinates on both sides, $0=\sum_{j=1}^{n}\left(\sum_{i=1}^{m} c_iu_{ij}\right)=2d\left(\sum_{i=1}^{m} c_i\right)$, hence $\sum_{i=1}^{m} c_i=0$ and so $\sum_{i=2}^{m} c_i(\underline{u}_i-\underline{u}_1)=\underline{0}$. Since a polytope is a simplex if and only if $\{\underline{u}_i-\underline{u}_1: i\geqq 2\}$, the edges emanating from $\underline{u}_1$, are linearly independent, it follows that $\mathfrak{U}$ is a trellis if and only if $\mathfrak{U}$ is a linearly independent set in $\mathbb{R}^n$; $m\leqq n$ in a trellis. The monomials $\underline{x}^{\underline{u}_i}$ are algebraically independent when $\prod_{i=1}^{m}(\underline{x}^{\underline{u}_i})^{t_i}=1$ implies $t_i=0$ for all i, hence $\mathfrak{U}$ is a trellis if and only if the $\underline{x}^{\underline{u}_i}$'s are algebraically independent. Finally, if $m=n$, then $\mathfrak{U}$ is a trellis if and only if $\det[u_{ij}]\neq 0$.

Suppose $\mathfrak{U}$ is a framework. We let $C(\mathfrak{U})=cvx(\mathfrak{U})\cap\mathbb{Z}^n$ and $E(\mathfrak{U})=cvx(\mathfrak{U})\cap(2\mathbb{Z})^n$ denote the lattice points (and the even lattice points) contained in the convex hull of $\mathfrak{U}$. We are interested in two sets of averages of sets of lattice points. If $\mathfrak{B}\subset\mathbb{Z}^n$, let

$$A(\mathfrak{B})=\{\tfrac{1}{2}(\underline{s}+\underline{t}):\ \underline{s},\ \underline{t}\in(\mathfrak{B}\cap(2\mathbb{Z})^n)\}$$

denote the set of averages of even points from $\mathfrak{B}$ and

$$\bar{A}(\mathfrak{B})=\{\tfrac{1}{2}(\underline{s}+\underline{t}):\ \underline{s}\neq\underline{t},\ \underline{s},\underline{t}\in(\mathfrak{B}\cap(2\mathbb{Z})^n)\}$$

denote the set of averages of distinct even points from $\mathfrak{B}$, so $A(\mathfrak{B})=\bar{A}(\mathfrak{B})\cup(\mathfrak{B}\cap(2\mathbb{Z})^n)$. Observe that $C(\mathfrak{U})$ contains $A(E(\mathfrak{U}))$; we show in [20] that $n\leqq 3$ implies $A(E(\mathfrak{U}))=C(\mathfrak{U})$, but this is false for $n\geqq 4$ (see Theorem 6.18).

If $\underline{w}\in\mathbb{Z}^n$ and $\underline{w}=\sum\lambda_i\underline{u}_i$ with $\lambda_i\geqq 0$ and $\sum\lambda_i=1$, then $\underline{w}\in C(\mathfrak{U})$; conversely, if $\underline{w}\in C(\mathfrak{U})$ then at least one such $\underline{\lambda}$ exists. Let

$$\Lambda(\underline{w})=\{\underline{\lambda}:\lambda_i\geqq 0,\ \sum\lambda_i=1 \text{ and } \underline{w}=\sum\lambda_i\underline{u}_i\}.$$

If $\mathfrak{U}$ is a trellis, then the linear independence of the $\underline{u}_i$'s implies that $\Lambda(\underline{w})=\{\underline{\lambda}\}$ is a singleton, and $(\lambda_1,\ldots,\lambda_m)$ are called the *barycentric coordinates* of $\underline{w}$ with respect to $\mathfrak{U}$. If $\lambda_i>0$ for all i, then $\underline{w}$ is *interior* to $\mathfrak{U}$.

Fix a framework $\mathfrak{U}$, and select $\underline{w}\in C(\mathfrak{U})$ and $\underline{\lambda}\in\Lambda(\underline{w})$. Under the substitution $t_i=\underline{x}^{\underline{u}_i}(\geqq 0)$ the AGI becomes

$$\lambda_1\underline{x}^{\underline{u}_1}+\ldots+\lambda_m\underline{x}^{\underline{u}_m}-(\underline{x}^{\underline{u}_1})^{\lambda_1}\ldots(\underline{x}^{\underline{u}_m})^{\lambda_m}\geqq 0. \tag{1.7}$$

Since the $\underline{u}_i$'s are even and $w_j=\sum\lambda_i\underline{u}_{ij}$,

$$(\underline{x}^{\underline{u}_1})^{\lambda_1}\ldots(\underline{x}^{\underline{u}_m})^{\lambda_m}=\prod_{i=1}^{m}\left(\prod_{j=1}^{n}|x_j|^{u_{ij}}\right)^{\lambda_i}=\prod_{j=1}^{n}|x_j|^{w_j}=|\underline{x}^{\underline{w}}|,$$

and since $\sum_{j=1}^{n}w_j=\sum_{j=1}^{n}\sum_{i=1}^{m}\lambda_i u_{ij}=2d$, it follows from (1.7) that

$$f(\mathfrak{U},\underline{\lambda},\underline{w})(\underline{x}):=\lambda_1\underline{x}^{\underline{u}_1}+\ldots+\lambda_m\underline{x}^{\underline{u}_m}-\underline{x}^{\underline{w}} \tag{1.8}$$

is a psd form. Any positive multiple of $f(\mathfrak{U},\underline{\lambda},\underline{w})$ is called an *agiform on* $\mathfrak{U}$. (Multiples are usually taken to clear the denominators of the coefficients.) If $f(\mathfrak{U},\underline{\lambda},\underline{w})(\underline{x})=0$ and $\lambda_i>0$ for $i\in I\subseteq\{1,\ldots,m\}$, then there exists $c\geqq 0$ so that $\underline{x}^{\underline{u}_i}=c$ when $i\in I$ and $\underline{x}^{\underline{w}}\geqq 0$; in particular, $f(\mathfrak{U},\underline{\lambda},\underline{w})(1,\ldots,1)=0$.

If $\mathfrak{U}$ is a trellis, f is called a *simplical agiform on* $\mathfrak{U}$. In this case, $\underline{\lambda}$ is redundant, so it is convenient to write $f=f(\mathfrak{U},\underline{w})$. The simplicial agiforms on a fixed trellis are indexed by the elements of $C(\mathfrak{U})$. If $\underline{w}\in A(\mathfrak{U})$, then either $\underline{w}=\underline{u}_i$ or $\underline{w}=\frac{1}{2}(\underline{u}_i+\underline{u}_j)$ and the agiform $f(\mathfrak{U},\underline{w})$ is simple. In the first case $\underline{\lambda}=\underline{e}_i$ and $f(\mathfrak{U},\underline{w})(\underline{x})=\underline{x}^{\underline{u}_i}-\underline{x}^{\underline{u}_i}=0$. In the second case, $\underline{\lambda}=\frac{1}{2}(\underline{e}_i+\underline{e}_j)$ and $f(\mathfrak{U},\underline{w})(\underline{x})=\frac{1}{2}\underline{x}^{\underline{u}_i}+\frac{1}{2}\underline{x}^{\underline{u}_j}-\underline{x}^{\underline{w}}=\frac{1}{2}(\underline{x}^{\underline{u}_i/2}-\underline{x}^{\underline{u}_j/2})^2$ is a binomial square. It turns out that every agiform is a convex combination of simplical agiforms (see Theorem 7.1).

If f is an agiform on a framework $\mathfrak{U}$ and $\mathfrak{U}$ is a subset of a framework $\mathfrak{B}$, then, by taking the additional monomials with coefficient 0, f is an agiform on $\mathfrak{B}$. This creates a possible ambiguity of notation if $\mathfrak{U}$ is a trellis and $\mathfrak{B}$ is not. Accordingly, we say that the form f is a *simplicial agiform* if there exists a trellis $\mathfrak{U}$ so that f is a simplicial agiform on $\mathfrak{U}$. We may always choose $\mathfrak{U}$ so that $\underline{w}$ is an interior point; if $\lambda_j=0$, then $\underline{x}^{\underline{u}_j}$ does not occur in f, and $\underline{u}_j$ may be deleted from $\mathfrak{U}$.

We return in detail to the prototypical agiforms, M and H. Each trellis lies in the plane $t_1+t_2+t_3=6$ and no information is lost in Fig. 1 by projecting onto the first two coordinates. The elements of each trellis are labeled, the even lattice points are large squares, and the other lattice points are smaller squares.

(1.9) *Example.* We define the Motzkin trellis:

$$\mathfrak{M}=\{(4,2,0),\ (2,4,0),\ (0,0,6)\}.$$

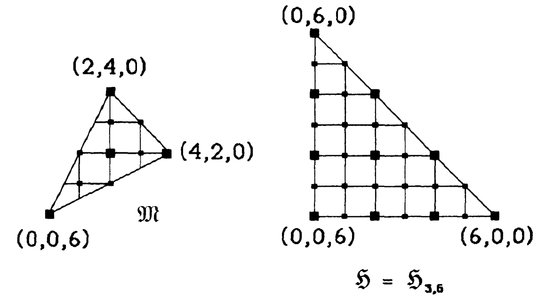


Fig. 1

[Since the points in $\mathfrak{M}$ are linearly independent, $\mathfrak{M}$ is a trellis; alternately, $cvx(\mathfrak{M})$ is a simplex, viz. a triangle.] Let $\underline{w}=(2,2,2)$ so $\Lambda(\underline{w})=\{(\frac{1}{3},\frac{1}{3},\frac{1}{3})\}$; $M=3f(\mathfrak{M},\underline{w})$ [see (1.2)] is not sos. It is easy to see from Fig. 1 that $E(\mathfrak{M})=\mathfrak{M}\cup\{\underline{w}\}$ and, since each of the ten lattice points in $cvx(\mathfrak{M})$ is an average of two even points, $C(\mathfrak{M})=A(E(\mathfrak{M}))$. Considering the distinct averages, we see that $\bar{A}(E(\mathfrak{M}))=C(\mathfrak{M})\backslash E(\mathfrak{M})$. It is particularly significant that $\underline{w}\notin\bar{A}(E(\mathfrak{M}))$; this will imply that M is not sos. By Corollary 3.4, $f(\mathfrak{M},\underline{v})$ is not sos if $\underline{v}$ is any of the four points in $C(\mathfrak{M})\backslash\bar{A}(E(\mathfrak{M}))$.

(1.10) *Example.* We define a special case of the Hurwitz trellis (see Example 1.12):

$$\mathfrak{H}=\{(6,0,0),\ (0,6,0),\ (0,0,6)\}.$$

Again, if $\underline{w}=(2,2,2)$, then $\Lambda(\underline{w})=\{(\frac{1}{3},\frac{1}{3},\frac{1}{3})\}$; and $H=3f(\mathfrak{H},\underline{w})$, as in (1.6). By Hurwitz' Theorem, H is sos; the construction of [14] gives a representation of H as a sum of nine squares which reduces to (see [12, p. 55]):

$$(1.11)\qquad 2H(x,y,z)=(x^2+y^2+z^2)\left((x^2-y^2)^2+(x^2-z^2)^2+(y^2-z^2)^2\right).$$

The representations from Hurwitz' proof are not efficient with respect to the number of squares (see [19]); we write H as a sum of five squares of binomials in (5.2). Again, $|E(\mathfrak{H})|=10$ and $C(\mathfrak{H})=A(E(\mathfrak{H}))$. In contrast to $\mathfrak{M}$, $\bar{A}(E(\mathfrak{H}))=C(\mathfrak{H})\backslash\mathfrak{H}$; there are enough even points in $C(\mathfrak{H})$ so that every non-vertex even point is an average of two distinct even points. This will imply, independently of Hurwitz, that H is sos (see Theorem 4.4).

(1.12) *Example.* The Hurwitz trellis $\mathfrak{H}_{n,2d}$ is $\{2d\underline{e}_i\}$, so $\mathfrak{H}_{3,6}=\mathfrak{H}$. Again, it is easy to see that $\mathfrak{H}_{n,2d}$ is a trellis and,

$$(1.13)\qquad C(\mathfrak{H}_{n,2d})=\{\underline{c}=(c_1,\dots,c_n):0\leqq c_i\in\mathbb{Z}\ \text{and}\ \textstyle\sum c_i=2d\}.$$

For $\underline{c}\in C(\mathfrak{H}_{n,2d})$, we have $\Lambda(\underline{c})=\left\{\frac{1}{2d}\underline{c}\right\}$ and, by Hurwitz' Theorem, $G(\underline{c})=2df(\mathfrak{H}_{n,2d},\underline{c})$ is also an sos simplicial agiform. In Example 2.4, we show that $\bar{A}(E(\mathfrak{H}_{n,2d}))=C(\mathfrak{H}_{n,2d})\backslash\mathfrak{H}_{n,2d}$; together with Theorem 4.4, this implies that $G(\underline{c})$ is sos.

Here is an overview of the rest of the paper.

In Sect. 2, we formally introduce an essential definition. If $\mathfrak{U}$ is a framework and $\mathfrak{L}$ is a set of lattice points containing $\mathfrak{U}$, then $\mathfrak{L}$ is "$\mathfrak{U}$-mediated" if every $\underline{w}$ in $\mathfrak{L}\backslash\mathfrak{U}$ is an average of two distinct even points in $\mathfrak{L}$. We give an algorithm for the construction of a maximal $\mathfrak{U}$-mediated set $\mathfrak{U}^*$ for which $A(\mathfrak{U}) \subseteqq \mathfrak{U}^* \subseteqq C(\mathfrak{U})$, and show that $\mathfrak{M}^* = A(\mathfrak{M})$ and $\mathfrak{H}^*_{n,2d} = C(\mathfrak{H}_{n,2d})$. Trellises such as these, for which $\mathfrak{U}^*$ is extreme, are called "M-trellises" and "H-trellises" respectively. We give some sufficient conditions for $\mathfrak{U}$ to be an M-trellis, which combine with previous results of the author on lattice point simplices to give a recipe for the construction of a large number of M-trellises.

In Sect. 3 we present part of the general theory of sos forms. If the agiform $f(\mathfrak{U}, \underline{\lambda}, \underline{w})$ is sos, we prove that $\underline{w} \in \mathfrak{U}^*$, by constructing a $\mathfrak{U}$-mediated set from the exponents of the monomials involved in the squares. Thus, a non-zero simplicial agiform on an M-trellis is either a binomial square or is not sos. Combined with the construction of the last section, this gives an efficient mechanism for producing large numbers of psd forms which are not sos. The essential arguments used are rephrasings and generalizations of those used by Motzkin, Choi, and Lam to show that M, S, and Q are not sos. We review some results on the Newton polytope of a psd form $p(\underline{x}) = \sum c(\underline{v})\underline{x}^{\underline{v}}: N(p) = cvx\{\underline{v} : c(\underline{v}) \neq 0\}$. [For example, $p \geqq q \geqq 0$ implies $N(p) \supseteqq N(q)$.] This is useful in studying extremality.

In Sect. 4 we show that the mediation relation $\underline{w} = \frac{1}{2}(\underline{s} + \underline{t})$, $\underline{s}$ and $\underline{t}$ even, implies an identity for the simplicial agiform $f(\mathfrak{U}, \underline{w})$ as a linear combination of $f(\mathfrak{U}, \underline{s})$, $f(\mathfrak{U}, \underline{t})$ and $(\underline{x}^{\underline{s}} - \underline{x}^{\underline{t}})^2$. If $\mathfrak{L}$ is a $\mathfrak{U}$-mediated set, we use this identity to derive a system of linear equations involving the simplicial agiforms on $\mathfrak{U}$. The solution to this system gives $f(\mathfrak{U}, \underline{w})$ as a sum of at most $|\mathfrak{L} \setminus \mathfrak{U}|$ squares of binomials. The results of sections two, three, and four combine to give one main result (Corollary 4.9):

Theorem. *The simplicial agiform* $f(\mathfrak{U}, \underline{w})$ *is* sos *if and only if* $\underline{w} \in \mathfrak{U}^*$.

It follows that every agiform on an H-trellis is sos (generalizing Hurwitz' Theorem) and an sos simplicial agiform is a sum of squares of binomials.

In Sect. 5, we apply the algorithm of Sect. 4 to write H and $M(x^k, y^k, z^k)$ (for $k = 2, 3$) explicitly as sums of squares in several inequivalent ways. In particular, we obtain H as a sum of five squares. We also discuss non-simplicial agiforms. Difficulties arise from the fact that $\Lambda(\underline{w})$ is not, in general, a singleton, so geometric information on $\underline{w}$ in $C(\mathfrak{U})$ need not translate into information about the agiform $f(\mathfrak{U}, \underline{\lambda}, \underline{w})$.

In Sect. 6, we examine four families of agiforms introduced by Motzkin, Choi, and Lam, generalizing M (twice), S and Q to more variables, and we introduce alternate generalizations of S and Q. Five of these six families of agiforms are not sos; four of them are simplicial. We compute $C(\mathfrak{U})$ and $E(\mathfrak{U})$ for suitable trellises for later use.

In Sect. 7 we give a sufficient condition for extremality. We show that every agiform is a convex combination of simplicial agiforms. An agiform f is "primitive" if it cannot be written as a non-trivial sum of other agiforms; this is weaker than extremality. We show that $f(\mathfrak{U}, \underline{\lambda}, \underline{w})$ is primitive if and only if it is simplicial and $E(\mathfrak{U}) \subseteqq (\mathfrak{U} \cup \{\underline{w}\})$. A study of the zero-sets of agiforms leads to the following equivalence relation on $\mathbb{Z}^n$: $\underline{v} \sim \underline{v}'$ if $\underline{\varepsilon}^{\underline{w}} = 1$ for $\underline{\varepsilon} \in \{-1, 1\}^n$ implies $\underline{\varepsilon}^{\underline{v}} = \underline{\varepsilon}^{\underline{v}'}$. This relation decomposes $C(\mathfrak{U})$ into equivalence classes $Z_1, \ldots, Z_r$, where $Z_1 \supseteqq \mathfrak{U} \cup \{\underline{w}\}$. We say that $\mathfrak{U}$ is "$\underline{w}$-thin" if $Z_1 = \mathfrak{U} \cup \{\underline{w}\}$ and Z_k is linearly independent for $k \geqq 2$. We show that if $f(\mathfrak{U}, \underline{\lambda}, \underline{w})$ is extremal, then it is simplicial and $\mathfrak{U}$ is $\underline{w}$-thin.

In Sect. 8, we show that if $\mathfrak{U}$ is not $\underset{\sim}{w}$-thin, then one can construct h so that $f_\alpha = f + \alpha h$ is psd for small $|\alpha|$, so $f = \frac{1}{2}(f_\alpha + f_{-\alpha})$ is not extremal. Thus, we obtain our other main result (Corollary 8.11):

Theorem. *Let f be an agiform. Then f is extremal if and only if f is simplicial and $\mathfrak{U}$ is $\underset{\sim}{w}$-thin.*

We use this to verify the extremality of M, S, and Q. The section ends with a derivation, following [17], of the simplicial agiform as the simplest extremal form which is not a monomial square. This may be viewed as an independent motivation for the study of agiforms.

In Sect. 9, we show that M, S, and Q are each generalized by a family of extremal forms; here are the next forms for each:

(1.14) $$M_4(x,y,z,w) = x^4y^2z^2 + x^2y^4z^2 + x^2y^2z^4 + w^8 - 4x^2y^2z^2w^2,$$

(1.15) $$\overline{S_4}(x,y,z,w) = x^4y^2z^2 + y^4z^2w^2 + z^4w^2x^2 + w^4x^2y^2 - 4x^2y^2z^2w^2,$$

(1.16) $$\overline{Q_6}(x,y,z,w,v,u) = u^6 + x^2y^2z^2 + y^2z^2w^2 + z^2w^2v^2 + w^2v^2x^2 + v^2x^2y^2 - 6xyzwvu.$$

We discuss S_4 and another two non-extremal primitive agiforms in some detail, and begin an analysis of "almost-agiforms."

We conclude in Sect. 10 with some open questions and areas for further research.

Finally, we should observe that our discussion centers on forms rather than inhomogeneous polynomials. This is largely a matter of taste, and we follow Choi and Lam in this decision. It is easy to change from one to the other by homogenizing a polynomial or dehomogenizing a form. The properties of being psd and sos are preserved with the obvious modifications, which we omit. It follows that the definitions and theorems of this paper have straightforward translations from forms to polynomials. The only significant changes are the deletion of the condition $\Sigma_j u_{ij} = 2d$ in the definition of framework, an accompanying adjustment in the criteria for $\mathfrak{U}$ to be a trellis, and a reduction by one in the number of variables.

Motzkin's example was $h(x,y) = x^2y^4 + x^4y^2 + 1 - 3x^2y^2$, which homogenizes to $M: M(x,y,1) = h(x,y)$. One might also homogenize each variable separately and create the biform ([5, p. 20]) $L(x,y,z,w) = x^2y^4z^2 + x^4y^2w^2 + z^4w^4 - 3x^2y^2z^2w^2$. Note that $L(x,y,1,1) = h(x,y)$ and L is a simplicial agiform on $\mathfrak{L} = \{(2,4,2,0), (4,2,0,2), (0,0,4,4)\}$.

Suppose $\mathfrak{S}$ is a polytope in $\mathbb{R}^n$ with vertices $\{\underset{\sim}{z}_1, \ldots, \underset{\sim}{z}_m\} \subset \mathbb{Z}^n_+$ and let $d = \max_i \left(\sum_{j=1}^n z_{ij} \right)$. We embed $\mathfrak{S}$ in the hyperplane $\left\{ \sum_{j=1}^{n+1} z_j = d \right\} \subset \mathbb{R}^{n+1}$ by setting $z_{i,n+1} = d - \sum_{j=1}^n z_{ij} \geqq 0$, and call the resulting simplex $\mathfrak{S}'$. Then $\mathfrak{U} = 2\mathfrak{S}'$ is a framework, and $E(\mathfrak{U}) = 2(\mathfrak{S}' \cap \mathbb{Z}^{n+1})$, so $|E(\mathfrak{U})| = |\mathfrak{S} \cap \mathbb{Z}^n|$. We return to this construction in Example 2.6.

This paper is, to a great extent, a rewriting and generalization of the author's [17], which was written very early in his study of psd and sos forms. All his subsequent papers in this subject have been written in collaboration with Man-Duen Choi and Tsit-Yuen Lam (and others). This is thus his first opportunity in many years to thank Professors Choi and Lam in print for their friendship, guidance and assistance.

2. Mediated Sets

Let $\mathfrak{U}$ be a framework. A set $\mathfrak{L} \subset \mathbb{Z}^n$ is called $\mathfrak{U}$-*mediated* if

$$\mathfrak{U} \subseteqq \mathfrak{L} \subseteqq (\bar{A}(\mathfrak{L}) \cup \mathfrak{U}). \tag{2.1}$$

That is, $\mathfrak{L}$ is $\mathfrak{U}$-mediated if it contains $\mathfrak{U}$, and every $\underline{v} \in \mathfrak{L} \setminus \mathfrak{U}$ is an average of two distinct even points in $\mathfrak{L}$; $\mathfrak{L}$ need not contain all of $\bar{A}(\mathfrak{L})$. (We will want to find the smallest $\mathfrak{U}$-mediated set containing a given lattice point.) By (2.1), every subset of $A(\mathfrak{U})$ containing $\mathfrak{U}$ is $\mathfrak{U}$-mediated (see also Theorem 2.8). We now give an algorithm for constructing the maximal $\mathfrak{U}$-mediated set, $\mathfrak{U}^*$.

(2.2) **Theorem.** *If* $\mathfrak{U}$ *is a framework, then there is a* $\mathfrak{U}$*-mediated set* $\mathfrak{U}^*$ *satisfying* $A(\mathfrak{U}) \subseteqq \mathfrak{U}^* \subseteqq C(\mathfrak{U})$ *which contains every* $\mathfrak{U}$*-mediated set.*

Proof. Define the sequence $\{\mathfrak{U}^k\}$ by $\mathfrak{U}^0 = C(\mathfrak{U})$ and $\mathfrak{U}^{k+1} = \bar{A}(\mathfrak{U}^k) \cup \mathfrak{U}$ for $k \geqq 0$. Then $\mathfrak{U}^1 = (\bar{A}(C(\mathfrak{U})) \cup \mathfrak{U}) \subseteqq C(\mathfrak{U}) = \mathfrak{U}^0$. If $\mathfrak{U}^k \subseteqq \mathfrak{U}^{k-1}$, then $\mathfrak{U}^{k+1} = (\bar{A}(\mathfrak{U}^k) \cup \mathfrak{U}) \subseteqq (\bar{A}(\mathfrak{U}^{k-1}) \cup \mathfrak{U}^k = \mathfrak{U}^k$, hence $\{\mathfrak{U}^k\}$ is a decreasing sequence of finite sets, which must stabilize. Let $\mathfrak{U}^r = \mathfrak{U}^{r+1} = \mathfrak{U}^*$; since $\mathfrak{U}^* = \mathfrak{U}^{r+1} = \bar{A}(\mathfrak{U}^r) \cup \mathfrak{U} = \bar{A}(\mathfrak{U}^*) \cup \mathfrak{U}$, $\mathfrak{U}^*$ is $\mathfrak{U}$-mediated. Further, for $k \geqq 0$, $A(\mathfrak{U}) = (\bar{A}(\mathfrak{U}) \cup \mathfrak{U}) \subseteqq (\bar{A}(\mathfrak{U}^k) \cup \mathfrak{U}) = \mathfrak{U}^{k+1}$, so $A(\mathfrak{U}) \subseteqq \mathfrak{U}^*$, and $\mathfrak{U}^* \subseteqq \mathfrak{U}^0 = C(\mathfrak{U})$.

Let $\mathfrak{L}$ be any $\mathfrak{U}$-mediated set and suppose $\underline{v} \in \mathfrak{L}$ is an extreme point of $cvx(\mathfrak{L})$. Then $\underline{v}$ cannot be an average of two distinct points in $\mathfrak{L}$; $\underline{v} \notin \bar{A}(\mathfrak{L})$. By (2.1), $\underline{v} \in \mathfrak{U}$. Thus, $cvx(\mathfrak{L}) \subseteqq cvx(\mathfrak{U})$ and $\mathfrak{L} \subseteqq C(\mathfrak{U}) = \mathfrak{U}^0$. Since $\mathfrak{L}$ is $\mathfrak{U}$-mediated, $\mathfrak{L} \subseteqq \mathfrak{U}^k$ implies $\mathfrak{L} \subseteqq (\bar{A}(\mathfrak{L}) \cup \mathfrak{U}) \subseteqq (\bar{A}(\mathfrak{U}^k) \cup \mathfrak{U}) = \mathfrak{U}^{k+1}$. It follows by induction that $\mathfrak{L} \subseteqq \mathfrak{U}^*$. □

Note that $\mathfrak{U}^* = C(\mathfrak{U})$ if and only if $C(\mathfrak{U})$ is $\mathfrak{U}$-mediated, that is, if and only if $\bar{A}(E(\mathfrak{U})) = C(\mathfrak{U}) \setminus \mathfrak{U}$.

(2.3) *Example.* (Continuing Example 1.9). Let $\mathfrak{M}$ be the Motzkin trellis. Referring to Fig. 1, we apply the algorithm of Theorem 2.2 to compute $\mathfrak{M}^*$: $\mathfrak{M}^0 = C(\mathfrak{M})$, $\mathfrak{M}^1 = \bar{A}(\mathfrak{M}^0) \cup \mathfrak{M}$. As $\underline{w} \notin \mathfrak{M}^1$, $\mathfrak{M}^1 \cap (2\mathbb{Z})^n = \mathfrak{M}$, so $\bar{A}(\mathfrak{M}^1) = \bar{A}(\mathfrak{M})$ and $\mathfrak{M}^2 = \bar{A}(\mathfrak{M}) \cup \mathfrak{M} = A(\mathfrak{M})$. Hence $A(\mathfrak{M}) \subseteqq \mathfrak{M}^* \subseteqq \mathfrak{M}^2 = A(\mathfrak{M})$, so $\mathfrak{M}^* = A(\mathfrak{M})$.

(2.4) *Example.* (Continuing Example 1.12). Let $\mathfrak{H}_{n,2d} = \{2d\underline{e}_i\}$ be the generic Hurwitz trellis. We show that $\bar{A}(E(\mathfrak{H}_{n,2d})) = C(\mathfrak{H}_{n,2d}) \setminus \mathfrak{H}_{n,2d}$, hence $\mathfrak{H}^*_{n,2d} = C(\mathfrak{H}_{n,2d})$. It suffices to write $\underline{c} \in C(\mathfrak{H}_{n,2d})$, $\underline{c} \neq 2d\underline{e}_i$, in the form $\underline{c} = \frac{1}{2}(\underline{s} + \underline{t})$, with $\underline{s} \neq \underline{t} \in E(\mathfrak{H}_{n,2d})$. Let $b_r = \sum_{i=1}^{r} c_i$, and choose k so that $b_{k-1} < d \leqq b_k$. If

$$\underline{s} = (2c_1, \ldots, 2c_{k-1}, 2d - 2b_{k-1}, 0, \ldots, 0)$$

and

$$\underline{t} = (0, \ldots, 0, 2b_k - 2d, 2c_{k+1}, \ldots, 2c_n),$$

then $\underline{c}=\frac{1}{2}(\underline{s}+\underline{t})$ with $\underline{s}, \underline{t} \in E(\mathfrak{H}_{n,2d})$. (If $\underline{s}=\underline{t}$, then $c_i=0$ for $i \neq k$; that is, $\underline{c}=2d\underline{e}_k$, a case we have excluded.)

We say that a trellis $\mathfrak{U}$ is an *M-trellis* if $\mathfrak{U}^* = A(\mathfrak{U})$; $\mathfrak{U}$ is an *H-trellis* if $\mathfrak{U}^* = C(\mathfrak{U})$. Every trellis in $\mathbb{Z}^3$ is either an H-trellis or an M-trellis (see [20]), but this is false in higher dimensions. If $A(\mathfrak{U}) = C(\mathfrak{U})$, for example if $\mathfrak{U} = \mathfrak{H}_{n,2}$, then $\mathfrak{U}$ may be both an H-trellis and an M-trellis. The argument in Example 2.3 generalizes into a useful criterion.

(2.5) **Theorem.** *If $\mathfrak{U}$ is a trellis, and either $E(\mathfrak{U}) = \mathfrak{U}$ or $E(\mathfrak{U}) = \mathfrak{U} \cup \{\underline{w}\}$ (and $\underline{w} \notin A(\mathfrak{U})$), then $\mathfrak{U}$ is an M-trellis.*

Proof. First, suppose $\mathfrak{U}^k \cap (2\mathbb{Z})^n = \mathfrak{U}$ for some k. Then $\bar{A}(\mathfrak{U}^k) = \bar{A}(\mathfrak{U})$ and $A(\mathfrak{U}) \subseteq \mathfrak{U}^* \subseteq \mathfrak{U}^{k+1} = (\bar{A}(\mathfrak{U}^k) \cup \mathfrak{U}) = (\bar{A}(\mathfrak{U}) \cup \mathfrak{U}) = A(\mathfrak{U})$, so $\mathfrak{U}^* = A(\mathfrak{U})$. It thus suffices to show that the hypothesis imply $\mathfrak{U}^k \cap (2\mathbb{Z})^n = \mathfrak{U}$ for $k=0$ or 1. If $E(\mathfrak{U}) = \mathfrak{U}$, then $\mathfrak{U}^0 \cap (2\mathbb{Z})^n = C(\mathfrak{U}) \cap (2\mathbb{Z})^n = E(\mathfrak{U}) = \mathfrak{U}$. If $E(\mathfrak{U}) = \mathfrak{U} \cup \{\underline{w}\}$, then $\mathfrak{U}^0 \cap (2\mathbb{Z})^n = \mathfrak{U} \cup \{\underline{w}\}$ and

$$\mathfrak{U}^1 = \mathfrak{U} \cup \bar{A}(\mathfrak{U} \cup \{\underline{w}\}) = A(\mathfrak{U}) \cup \{\tfrac{1}{2}(\underline{w} + \underline{u}_i)\}.$$

Since $\underline{w} \notin A(\mathfrak{U})$ by hypothesis, $\underline{w} \notin \mathfrak{U}^1$, so $\mathfrak{U}^1 \cap (2\mathbb{Z})^n = \mathfrak{U}$. □

If $\mathfrak{U}$ is a trellis, $E(\mathfrak{U}) = \mathfrak{U} \cup \{\underline{w}\}$ and $\underline{w} \in A(\mathfrak{U})$, then $\mathfrak{U}$ is not an M-trellis. It does not seem useful to elaborate on the conditions under which $\mathfrak{U}^k \cap (2\mathbb{Z})^n = \mathfrak{U}$ for $k \geqq 2$. The number of steps in the computation of $\mathfrak{U}^*$ is bounded above by $|E(\mathfrak{U}) \setminus \mathfrak{U}|$, since $\mathfrak{U}^k \cap (2\mathbb{Z})^n = \mathfrak{U}^{k+1} \cap (2\mathbb{Z})^n$ implies $\mathfrak{U}^{k+1} = \mathfrak{U}^{k+2}$ $(= \mathfrak{U}^*)$. This bound is achieved by the trellis

$$\mathfrak{U}_p = \{(0,0,2p), (2,2p-2,0), (4,2,2p-6)\}$$

for $p \geqq 3$, which is discussed more fully in [20]. (Note that $\mathfrak{U}_3 = \mathfrak{M}$.) It is not hard to see that

$$E(\mathfrak{U}_p) = \mathfrak{U}_p \cup \{(2,2,2p-4), (2,4,2p-6), \ldots, (2,2p-4,2)\}$$

and that $E(\mathfrak{U}_p^1) \setminus E(\mathfrak{U}_p) = (2,2,2p-4)$, $E(\mathfrak{U}_p^2) \setminus E(\mathfrak{U}_p^1) = (2,4,2p-6)$, etc.

We now show how to transform certain lattice point simplices into M-trellises via homogenization.

(2.6) *Example.* A *k-point n-simplex* (see [18]) is a simplex in $\mathbb{R}^n$ with vertex-set $\mathfrak{S} = \{\underline{z}_0, \ldots, \underline{z}_n\}$, such that $cvx(\mathfrak{S}) \cap \mathbb{Z}^n = \mathfrak{S} \cup \{\underline{v}_1, \ldots, \underline{v}_k\}$ and $\underline{v}_i = \sum \lambda_{ij} \underline{z}_j$ with $\sum \lambda_{ij} = 1$ and $\lambda_{ij} > 0$ [so the $\underline{v}_i$'s are strictly interior to $cvx(\mathfrak{S})$.] Since $\underline{\lambda}_i$ is the unique solution to a linear system with integer coefficients, $\lambda_{ij} \in \mathbb{Q}$. The term "0-point" is equivalent to "fundamental"; fundamental triangles are familiar, and fundamental tetrahedra have been extensively studied.

If $\mathfrak{S}$ is a 1-point n-simplex and $\underline{v}_1 = \sum \lambda_i \underline{z}_i$, write $\lambda_i = a_i/D$. Then $\sum \{k\lambda_i\} > 1$ for $2 \leqq k \leqq D-1$ by [18, p. 228], where $\{x\} = x - [x]$ denotes the fractional part of x. Conversely [p. 230], suppose $n+1$ positive rationals $\lambda_i = a_i/D$ are given, so that $\sum \lambda_i = 1$, $\sum \{k\lambda_i\} > 1$ for $2 \leqq k \leqq D-1$ and a_j and D are relatively prime for some j. (For example, $a_i = 1$, $D = n+1$.) Then there exists a canonical 1-point n-simplex $\mathfrak{S}$ with interior point $\underline{v} = \sum \lambda_i \underline{z}_i$. The only such set of rationals satisfying these conditions for $n=2$ is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$; when $n=3$, there are seven such sets, up to permutation.

Let $\mathfrak{S}=\{\underline{z}_i\}$ be a 0-point or 1-point simplex in $\mathbb{R}^n$, translated so that the components are all non-negative. As described in the introduction, $\mathfrak{S}$ may be homogenized so that we obtain the framework $2\mathfrak{S}'=\mathfrak{U}$ in $\mathbb{R}^{n+1}$. Since $\mathfrak{S}$ is a simplex, $\mathfrak{U}$ is a trellis; either $E(\mathfrak{U})=\mathfrak{U}$ or $E(\mathfrak{U})=\mathfrak{U}\cup\{2\underline{v}\}$, depending on whether $\mathfrak{S}$ is 0-point or 1-point. In the 1-point case, since $n\geq 2$ and $\underline{v}$ is strictly interior to $\mathfrak{S}$, it follows that $2\underline{v}\notin A(\mathfrak{U})$. In either case, it follows from Theorem 2.5 that $\mathfrak{U}$ is an M-trellis.

As a concrete illustration of this construction, take $n=3$ and $\underline{\lambda}=(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$. Then [p. 230] $\mathfrak{S}$ is the tetrahedron with vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, and $(-1,-1,4)$, and $\underline{z}=(0,0,1)$ has barycentric coordinates equal to $\underline{\lambda}$. After translation by $(1,1,0)$, we have $d=4$ and

$$\mathfrak{U}=2\mathfrak{S}'=\{(2,2,0,4),\ (4,2,0,2),\ (2,4,0,2),\ (0,0,8,0)\}.$$

Upon permuting the third and fourth coordinates, $\mathfrak{U}=\mathfrak{M}_4$ (see Theorem 6.9). This construction, applied to $\underline{\lambda}=\underline{\varphi}_n=\left(\frac{1}{n},\dots,\frac{1}{n}\right)$, leads to $\mathfrak{M}_{n+1}$.

We prove the following result in [20], which complements the previous discussion by giving a general way to construct H-trellises.

(2.7) **Proposition.** If $\mathfrak{U}=\{\underline{u}_1,\dots,\underline{u}_m\}$ is a trellis and $k\geq\max(2,m-2)$, then $k\mathfrak{U}$ is an H-trellis.

The final result is really a remark used to prove Corollary 4.11.

(2.8) **Theorem.** *If $\mathfrak{U}$ is a trellis and $\underline{w}\in\mathfrak{U}^*$, then there is a $\mathfrak{U}$-mediated set $\mathfrak{F}$ which contains $\underline{w}$ and has at most $|E(\mathfrak{U})\cup\{\underline{w}\}|$ elements.*

Proof. Let $\mathfrak{F}$ consist of the even points in $\mathfrak{U}^*$ plus $\underline{w}$ (if it isn't even.) Then $\bar{A}(\mathfrak{F})=\bar{A}(\mathfrak{U}^*)$, so $\mathfrak{F}$ is $\mathfrak{U}$-mediated. □

When $\mathfrak{U}=\mathfrak{H}_{n,2d}$, we can do better than this. Given $\underline{w}$, use the construction in Example 2.4 to write $\underline{w}=\frac{1}{2}(\underline{s}+\underline{t})$. Since $\underline{s}$ and $\underline{t}$ are even and lie in $\mathfrak{H}_{n,2d}$, we may repeat the argument, and write $\underline{s}$ and $\underline{t}$ as averages, etc. Unless w_1 or w_n is greater than d, $\underline{s}$ and $\underline{t}$ each have fewer than n positive components. It can be shown that this process leads to a $\mathfrak{U}$-mediated set containing $\underline{w}$ which has considerably fewer than $|E(\mathfrak{U})|$ elements. (This is the algorithm of [19], at least for $n\geq 3$.)

The concept of an H-trellis is related to Handelman's property of two-convexity (see [10, 11]), which arises in the study of the integral closure of certain commutative algebras associated to lattice point polytopes. A lattice point polytope $\mathfrak{S}$ in $\mathbb{R}^n$ is called *two-convex* if every lattice point in $2\mathfrak{S}$ is a sum of two lattice points in $\mathfrak{S}$ [10, p. 150]. After homogenization, $\mathfrak{U}=2\mathfrak{S}'$ is a framework in $\mathbb{R}^{n+1}$, and $\mathfrak{S}$ is two-convex if and only if $C(\mathfrak{U})=A(E(\mathfrak{U}))$. If $\mathfrak{S}$ is a simplex, then two-convexity is a weaker condition than being an H-trellis ($C(\mathfrak{U})=\bar{A}(E(\mathfrak{U}))\cup\mathfrak{U}$.) Handelman [11, p. 33] has proved that, among other properties, $k\mathfrak{U}$ is two-convex if $k\geq n$. This part of his result is implied by Proposition 2.7.

3. Sos Agiforms: The Necessary Condition

We begin this section with some general remarks about sos forms. Suppose $p=\sum_{k=1}^{r} h_k^2$, where $p(\underline{x})=\sum_{\underline{u}} a(\underline{u})\underline{x}^{\underline{u}}$ and $h_k(\underline{x})=\sum_{\underline{v}} b_k(\underline{v})\underline{x}^{\underline{v}}$. Then

$$\sum_{\underline{u}} a(\underline{u})\underline{x}^{\underline{u}}=\sum_k\Big(\sum_{\underline{v}} b_k(\underline{v})\underline{x}^{\underline{v}}\Big)^2. \tag{3.1}$$

Let $\underline{B}(\underline{v})$ denote the r-tuple whose k-th component is $b_k(\underline{v})$ and let $G(\underline{v},\underline{v}')=\underline{B}(\underline{v})\cdot\underline{B}(\underline{v}')=\sum_k b_k(\underline{v})b_k(\underline{v}')$. By comparing the coefficient of $\underline{x}^{\underline{u}}$ on both sides of (3.1), we obtain the equation:

$$a(\underline{u})=\sum_{\underline{v}+\underline{v}'=\underline{u}} G(\underline{v},\underline{v}')=\sum_{\underline{v}} G(\underline{v},\underline{u}-\underline{v}). \tag{3.2}$$

These observations were made, in effect, by Motzkin [16] and generalized by Choi and Lam [3,4] into the "term-inspection" method. Refinements on the term-inspection method were made in [17], and results needed in our discussion are presented in Theorem 3.6. The general theory, to be developed in [7], proceeds from the fact that $[\underline{B}(\underline{v})\cdot\underline{B}(\underline{v}')]$ is a psd matrix.

(3.3) **Theorem.** *If $\mathfrak{U}$ is a framework and $f=f(\mathfrak{U},\underline{\lambda},\underline{w})$ is an* sos *agiform, then $\underline{w}\in\mathfrak{U}^*$.*

Proof. Using the notation of (3.1), suppose $f=\sum h_k^2$ and let

$$\mathfrak{N}=\{\underline{v}\colon b_k(\underline{v})\neq 0 \text{ for some } k\}.$$

Let $\mathfrak{L}=2\mathfrak{N}\cup\mathfrak{U}\cup\{\underline{w}\}$. We show that $\mathfrak{L}$ is $\mathfrak{U}$-mediated (and so $\underline{w}\in\mathfrak{L}\subseteq\mathfrak{U}^*$) by writing each $\underline{u}\in\mathfrak{L}\backslash\mathfrak{U}$ as a sum of two distinct points in $\mathfrak{N}$; this implies that $\underline{u}$ is an average of two distinct even points in $2\mathfrak{N}\subseteq\mathfrak{L}$.

If $G(\underline{v},\underline{v}')<0$, then $b_k(\underline{v})b_k(\underline{v}')<0$ for some k, hence $\underline{v}\neq\underline{v}'$ and $\underline{v}$ and $\underline{v}'$ belong to $\mathfrak{N}$. It thus suffices to show that, for $\underline{u}\in\mathfrak{L}\backslash\mathfrak{U}$, there exists $\underline{v}$ with $G(\underline{v},\underline{u}-\underline{v})<0$. Note that $a(\underline{u}_i)=\lambda_i$, $a(\underline{w})=-1$, and $a(\underline{u})=0$ otherwise. By (3.2), we have $-1=a(\underline{w})=\sum G(\underline{v},\underline{w}-\underline{v})$, so $G(\underline{v}_0,\underline{w}-\underline{v}_0)<0$ for some $\underline{v}_0$. If $\underline{u}\neq\underline{w}$, then $\underline{u}\in\mathfrak{L}\backslash(\mathfrak{U}\cup\{\underline{w}\})$ so $a(\underline{u})=0=\sum G(\underline{v},\underline{u}-\underline{v})$. But $\underline{u}\in 2\mathfrak{N}$, so $G(\tfrac{1}{2}\underline{u},\tfrac{1}{2}\underline{u})>0$ and there must exist $\underline{v}$ with $G(\underline{v},\underline{u}-\underline{v})<0$ to make the sum vanish. □

(3.4) **Corollary.** *If $\mathfrak{U}$ is an M-trellis and $\underline{w}\notin A(\mathfrak{U})$, then the simplicial agiform $f(\mathfrak{U},\underline{w})$ is not* sos. *Any non-zero* sos *agiform on a M-trellis is a perfect square.*

Proof. If $\mathfrak{U}$ is an M-trellis and $f(\mathfrak{U},\underline{w})$ is sos, then $\underline{w}\in\mathfrak{U}^*=A(\mathfrak{U})$, so either $\underline{w}=\underline{u}_i$ (and $f=0$) or $\underline{w}=\tfrac{1}{2}(\underline{u}_i+\underline{u}_j)$ [and $f=\tfrac{1}{2}(\underline{x}^{\underline{u}_i/2}-\underline{x}^{\underline{u}_j/2})^2$]. □

Using the construction of Example 2.6 and Corollary 3.4, one can produce many psd forms which are not sos. The argument used by Motzkin, Choi, and Lam to prove that M, S, and Q are not sos was basically this: if $\underline{w}\notin\mathfrak{U}^1$, then $f(\mathfrak{U},\underline{w})$ is not sos. Since $\mathfrak{U}^1\supseteq\mathfrak{U}^*$, this is a weaker version of Theorem 3.3.

The term-inspection method can be more fully developed. Suppose $p(\underline{x})=\sum a(\underline{u})\underline{x}^{\underline{u}}$ is a (not necessarily psd) form; let $N(p)=cvx\{\underline{v}\colon a(\underline{v})\neq 0\}$ denote the *Newton polytope* or *cage* of p. If f is a simplicial agiform, then $N(f)$ is the simplex whose vertices are $\{\underline{u}_i\colon\lambda_i>0\}$. We let "$\alpha\cdot\leq s$" denote the set $\{\underline{y}\in\mathbb{R}^n\colon\underline{\alpha}\cdot\underline{y}\leq s\}$ and say that it is a *supporting half-space* for $N(p)$ if $\underset{\sim}{\alpha}\cdot\underset{\sim}{z}\leqq s$ for all $\underset{\sim}{z}\in N(p)$ and $\underset{\sim}{\alpha}\cdot\underset{\sim}{z}_0=s$ for some $\underset{\sim}{z}_0\in N(p)$. [By the definition of $N(p)$, this implies that $\underset{\sim}{\alpha}\cdot\underset{\sim}{v}=s$ for some $\underset{\sim}{v}$ with $a(\underset{\sim}{v})\neq 0$.] If $\underset{\sim}{\alpha}\cdot\leqq s$ is a supporting half-space, then $\underset{\sim}{\alpha}\cdot=s$ is called a *supporting hyperplane*. The supporting hyperplanes of p have another interpretation. For p as above, with $a(\underset{\sim}{u})\neq 0$, $\underset{\sim}{\alpha}\in\mathbb{R}^n$ and $t>0$, let

$$p(\underset{\sim}{x},\underset{\sim}{\alpha})(t)=p(t^{\alpha_1}x_1,\ldots,t^{\alpha_n}x_n)=\sum(a(\underset{\sim}{u})\underset{\sim}{x}^{u})t^{\sum\alpha_iu_i}=\sum(a(\underset{\sim}{u})\underset{\sim}{x}^{u})t^{\underset{\sim}{\alpha}\cdot\underset{\sim}{u}},$$

and let $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)=\lim_{t\to\infty} t^{-s}p(\underset{\sim}{x},\underset{\sim}{\alpha})(t)$. If $s>\max\underset{\sim}{\alpha}\cdot\underset{\sim}{u}$, then $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)=0$. If $s=\max\underset{\sim}{\alpha}\cdot\underset{\sim}{u}$, then $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)=\{\sum a(\underset{\sim}{u})\underset{\sim}{x}^{u}:\underset{\sim}{\alpha}\cdot\underset{\sim}{u}=s\}<\infty$, and since the summation is over a non-empty set, it is not identically zero. If $0<L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)$ and $r<s$, then $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},r)=\infty$. Thus, if $s<\max\underset{\sim}{\alpha}\cdot\underset{\sim}{u}$, then $L(p,\underset{\sim}{y},\underset{\sim}{\alpha},s)=\infty$ for some $\underset{\sim}{y}$. We have thus proved the following lemma.

(3.5) **Lemma.** *The half-space $\underset{\sim}{\alpha}\cdot\leqq s$ $(=s(\underset{\sim}{\alpha},p))$ is a supporting half-space of $N(p)$ if and only if $|L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)|<\infty$ for all $\underset{\sim}{x}\in\mathbb{R}^n$ and $L(p,\underset{\sim}{y},\underset{\sim}{\alpha},s)\neq 0$ for some $\underset{\sim}{y}$.*

(3.6) **Theorem.** *Let $p(\underset{\sim}{x})=\sum a(\underset{\sim}{v})\underset{\sim}{x}^{v}$ be a* psd *form.*

(i) *If $p\geqq q\geqq 0$, then $N(p)\supseteq N(q)$.*
(ii) *If $p=\sum h_k^2$, then $N(h_j)\subseteq\frac{1}{2}N(p)$ for each h_j.*
(iii) *If $\underset{\sim}{v}_0$ is an extreme point of $N(p)$, then $\underset{\sim}{v}_0$ is an even lattice point and $a(\underset{\sim}{v}_0)>0$.*
(iv) *If F is a face of the polytope $N(p)$, then $p^{(F)}(\underset{\sim}{x})=\sum_{\underset{\sim}{v}\in F} a(\underset{\sim}{v})\underset{\sim}{x}^{v}$ is* psd.

Proof. (i) Since $p(\underset{\sim}{x},\underset{\sim}{\alpha})(t)\geqq q(\underset{\sim}{x},\underset{\sim}{\alpha})(t)$, it follows from multiplication by t^{-s} that $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)\geqq L(q,\underset{\sim}{x},\underset{\sim}{\alpha},s)$. By Lemma 3.5, it follows that $s(\underset{\sim}{\alpha},p)\geqq s(\underset{\sim}{\alpha},q)$. Thus every supporting half-space of $N(p)$ contains $N(q)$. But $N(p)$, as a convex body, is the intersection of its supporting half-spaces, so $N(p)\supseteq N(q)$.

(ii) If $p=\sum h_k^2$, then $p\geqq h_j^2$, so $N(p)\supseteq N(h_j^2)$. Since $L(h^2,\underset{\sim}{x},\underset{\sim}{\alpha},2s)=(L(h,\underset{\sim}{x},\underset{\sim}{\alpha},s))^2$, Lemma 3.5 implies that $N(h_j^2)=2N(h_j)$.

(iii) If $\underset{\sim}{v}_0$ is an extreme point of $N(p)$, then $a(\underset{\sim}{v}_0)\neq 0$ by the definition of $N(p)$, and there exists a supporting hyperplane for $N(p)$ containing only $\underset{\sim}{v}_0$. That is, there exists $\underset{\sim}{\alpha}$ so that $\{\underset{\sim}{v}:\underset{\sim}{\alpha}\cdot\underset{\sim}{v}=s(\underset{\sim}{\alpha},p)\}=\{\underset{\sim}{v}_0\}$ and $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)=a(\underset{\sim}{v}_0)\underset{\sim}{x}^{v_0}$. Since p is psd, $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)\geqq 0$, so $a(\underset{\sim}{v}_0)\underset{\sim}{x}^{v_0}$ is a psd form. Since $a(\underset{\sim}{v}_0)\underset{\sim}{\varepsilon}^{v_0}\geqq 0$ for $\underset{\sim}{\varepsilon}\in\{-1,1\}^n$, it follows that $a(\underset{\sim}{v}_0)\geqq 0$ and v_{0i} is even.

(iv) The argument of (iii) may be applied, choosing $\underset{\sim}{\alpha}\cdot\leqq s$ to be a supporting hyperplane containing F. Once again, $L(p,\underset{\sim}{x},\underset{\sim}{\alpha},s)=p^{(F)}(\underset{\sim}{x})\geqq 0$. □

Both (iii) and (iv) are generalizations of the fact that a non-negative polynomial has even degree and positive leading coefficient. Parts (i), (ii), and (iii) are contained in [17, pp. 365–366], where they are proved as above, but less carefully; parts (i) and (iv) were proved by Handelman [9, p. 53]. In [11, p. 69], Handelman proved that $N(fg)=N(f)+N(g)$, with the usual convex set-sum; this implies the argument used in (ii) that $N(h^2)=2N(h)$.

4. Sos Agiforms: The Sufficient Condition

In this section, we prove the converse of Theorem 3.3 when $\mathfrak{U}$ is a trellis: if $\underset{\sim}{w}\in\mathfrak{U}^*$, then the simplicial agiform $f(\mathfrak{U},\underset{\sim}{w})$ is sos. We need two lemmas: a useful identity and a technical matrix result.

(4.1) Lemma. *Let $\mathfrak{U}$ be a trellis and suppose $\underline{w}=\frac{1}{2}(\underline{s}+\underline{t})$, where $\underline{s}$, $\underline{t}\in E(\mathfrak{U})$. Then*

$$2f(\mathfrak{U},\underline{w})(\underline{x})=f(\mathfrak{U},\underline{s})(\underline{x})+f(\mathfrak{U},\underline{t})(\underline{x})+(\underline{x}^{s/2}-\underline{x}^{t/2})^2. \tag{4.2}$$

Proof. Write $\underline{w}$, $\underline{s}$, and $\underline{t}$ in terms of the $\underline{u}_i$'s: $\underline{w}=\sum\lambda_i\underline{u}_i$, $\underline{s}=\sum\sigma_i\underline{u}_i$ and $\underline{t}=\sum\tau_i\underline{u}_i$. By the uniqueness of barycentric coordinates, $\lambda_i=\frac{1}{2}(\sigma_i+\tau_i)$. It follows that the right-hand side of (4.2),

$$(\textstyle\sum\sigma_i\underline{x}^{u_i}-\underline{x}^s)+(\sum\tau_i\underline{x}^{u_i}-\underline{x}^t)+(\underline{x}^s-2\underline{x}^w+\underline{x}^t),$$

equals the left-hand side. □

If $\underline{s}=\underline{t}=\underline{w}$, (4.2) is vacuous. If $\underline{s}=\underline{u}_k\in\mathfrak{U}$, then $f(\mathfrak{U},\underline{u}_k)(\underline{x})=\underline{x}^{u_k}-\underline{x}^{u_k}$ vanishes identically. Thus one or both of the simplicial agiforms on the right-hand side of (4.2) may disappear.

When $\mathfrak{L}$ is a $\mathfrak{U}$-mediated set, each $\underline{w}\in\mathfrak{L}\setminus\mathfrak{U}$ can be written as $\frac{1}{2}(\underline{s}+\underline{t})$, and so each $f(\mathfrak{U},\underline{w})$ satisfies a non-trivial equation of shape (4.2). This leads to an inhomogeneous linear system involving the simplicial agiforms. The next lemma will be applied to the matrix of that system.

(4.3) Lemma. *Let $A=[a_{ij}]$ be a finite matrix such that $a_{ii}=2$ and $a_{ij}\in\{0,-1\}$ if $i\neq j$. Suppose each row of A has at most two -1's and there is no principal submatrix of A in which each row has exactly two -1's. Then A is invertible, and the entries of A^{-1} are non-negative.*

Proof. Write $A=2(I-P)$ and note that the entries of P are 0 and $\frac{1}{2}$. We shall show that all the eigenvalues of P have modulus <1, so $P^r\to 0$. In this case, $A(I+P+\ldots+P^{r-1})=2(I-P^r)$, and $A^{-1}=\frac{1}{2}(I+P+P^2+\ldots)$ exists and has non-negative entries.

Let α be an eigenvalue of P and let $\underline{z}$ be a non-zero column vector with $A\underline{z}=\alpha\underline{z}$; let $\zeta=\max|z_k|$, $I=\{k:|z_k|=\zeta\}$, $T(i)=\{j:p_{ij}=\frac{1}{2}\}$, and $N(i)=|T(i)|$ [so $0\leqq N(i)\leqq 2$]. Then for $i\in I$,

$$|\alpha|\zeta=|\alpha z_i|=\left|\sum_{j\in T(i)}\frac{1}{2}z_j\right|\leqq\frac{1}{2}N(i)\zeta\leqq\zeta.$$

Since $\zeta>0$, $|\alpha|\leqq 1$. If $|\alpha|=1$, then $N(i)=2$ and for $j\in T(i)$, $|z_j|=\zeta$, so $T(i)\subseteq I$. That is, the principal submatrix of P with rows and columns taken from I has two $\frac{1}{2}$'s in each row. The corresponding principal submatrix in A has two -1's in each row, which violates the hypothesis. It follows that $|\alpha|<1$ for every eigenvalue α, completing the proof. □

(4.4) Theorem. *If $\mathfrak{U}$ is a trellis, $\mathfrak{L}$ is $\mathfrak{U}$-mediated and $\underline{w}\in\mathfrak{L}$, then the simplicial agiform $f=f(\mathfrak{U},\underline{w})$ is* sos. *To be specific, f is a sum of $|\mathfrak{L}\setminus\mathfrak{U}|$ squares of the form $c(\underline{x}^s-\underline{x}^t)^2$, where $2\underline{s}$, $2\underline{t}\in\mathfrak{L}$ and $c\geqq 0$.*

Proof. Index the points of $\mathfrak{L}\setminus\mathfrak{U}$ as $\underline{w}_1,\ldots,\underline{w}_T$, with $\underline{w}=\underline{w}_1$. Since $\mathfrak{L}$ is $\mathfrak{U}$-mediated, at least one of the following three statements is true for each $\underline{w}_i$ and suitable distinct $\underline{u}_r$, $\underline{u}_s$, and $\underline{w}_j$, $\underline{w}_k\in\mathfrak{L}\cap(2\mathbb{Z})^n$:

(4.5) (i) $\qquad \underline{w}_i=\frac{1}{2}(\underline{u}_r+\underline{u}_s)$,

(4.5) (ii) $\qquad \underline{w}_i=\frac{1}{2}(\underline{u}_r+\underline{w}_k)$,

(4.5) (iii) $\qquad \underline{w}_i=\frac{1}{2}(\underline{w}_j+\underline{w}_k) \quad (j,k\neq i)$.

For the purposes of this proof, pick exactly one correct decomposition for each $\underline{w}_i$. By Lemma 4.1, the relationships in (4.5) have, respectively, the following implications [recall that $f(\mathfrak{U}, \underline{u}_i)=0$]:

(4.6) (i) $\quad 2f(\mathfrak{U}, \underline{w}_i)(\underline{x})=(\underline{x}^{\underline{u}_r/2}-\underline{x}^{\underline{u}_s/2})^2$,

(4.6) (ii) $\quad 2f(\mathfrak{U}, \underline{w}_i)(\underline{x})=f(\mathfrak{U}, \underline{w}_k)(\underline{x})+(\underline{x}^{\underline{u}_r/2}-\underline{x}^{\underline{w}_k/2})^2$,

(4.6) (iii) $\quad 2f(\mathfrak{U}, \underline{w}_i)(\underline{x})=f(\mathfrak{U}, \underline{w}_j)(\underline{x})+f(\mathfrak{U}, \underline{w}_k)(\underline{x})+(\underline{x}^{\underline{w}_j/2}-\underline{x}^{\underline{w}_k/2})^2$.

Define the $T\times T$ matrix $A=[a_{ij}]$ as follows. For all i, $a_{ii}=2$. If $\underline{w}_i$ satisfies (4.5) (i), then the other entries in the i-th row are 0. If $\underline{w}_i$ satisfies (4.5) (ii), then $a_{ik}=-1$ and the other entries in the i-th row are 0. If $\underline{w}_i$ satisfies (4.5) (iii), then $a_{ij}=a_{ik}=-1$ and the other entries in the i-th row are 0. Let h_i denote the binomial square appearing on the right-hand side of the appropriate expression for $2f(\mathfrak{U}, \underline{w}_i)$ in (4.6), so $h_i(\underline{x})=(\underline{x}^{\underline{s}_i}-\underline{x}^{\underline{t}_i})^2$ and $2\underline{s}_i$, $2\underline{t}_i\in\mathfrak{L}$. Let $\underline{H}$ and $\underline{F}$ denote the column vectors whose i-th components are $h_i(\underline{x})$ and $f(\mathfrak{U}, \underline{w}_i)$, respectively. By the construction of A, (4.6) can be put into matrix form:

(4.7)
$$A\underline{F}=\underline{H}.$$

Lemma 4.3 was tailored to fit the matrix A; only one hypothesis is not obviously satisfied. Suppose A has a principal submatrix (with rows and columns from I) with two -1's in each row and let $\mathfrak{L}'=\{\underline{w}_i : i\in I\}$. Then every $\underline{w}_i\in\mathfrak{L}'$ satisfies (4.5) (iii) with j and k in I, so every point in $\mathfrak{L}'$ is an average of two other points in $\mathfrak{L}'$. This is clearly impossible for a finite set in $\mathbb{R}^n$; consider a point at maximum distance from the origin. It follows that A has no such principal submatrix, and Lemma 4.3 applies.

From (4.7),

(4.8)
$$\underline{F}=A^{-1}\underline{H},$$

and A^{-1} is a matrix with non-negative entries. The i-th component of $\underline{F}$ can thus be read off from (4.8) as a non-negative linear combination of the h_i's. Thus $f=f(\mathfrak{U}, \underline{w}_1)$ is a sum of $T=|\mathfrak{L}\backslash\mathfrak{U}|$ binomial squares. □

(4.9) **Corollary.** *A simplicial agiform $f(\mathfrak{U}, \underline{w})$ is* sos *if and only if $\underline{w}\in\mathfrak{U}^*$.*

Proof. Combine Theorems 3.3 and 4.4. □

(4.10) **Corollary.** *Every simplicial agiform on an H-trellis is* sos.

Example 2.4 and Corollary 4.10 generalize Hurwitz' Theorem.

(4.11) **Corollary.** *An* sos *simplicial agiform $f(\mathfrak{U}, \underline{w})$ is a sum of $|E(\mathfrak{U})\cup\{\underline{w}\}|-|\mathfrak{U}|$ squares.*

Proof. Choose $\mathfrak{L}$ using Theorem 2.8. □

It is shown in [19, pp. 110, 111] that the Hurwitzian agiform $G(\underline{c})$ is a sum of $3n-4$ squares [where $\underline{c}=(c_1, \dots, c_n)$ and $\sum c_i=2d$]; this is much smaller than the bound in Corollary 4.11. Proposition 2.7 and Theorem 4.4 can be combined.

(4.12) **Proposition.** *If $f(\mathfrak{U}, \underline{w})$ is simplicial on $\mathfrak{U}=\{\underline{u}_1, \ldots, \underline{u}_m\}$ and $k \geqq \max(2, m-2)$, then the following form is* sos*:*

$$f(\mathfrak{U}, \underline{w})(x_1^k, \ldots, x_n^k)=\lambda_1 \underline{x}^{k\underline{u}_1}+\ldots+\lambda_m \underline{x}^{k\underline{u}_m}-\underline{x}^{k\underline{w}}=f(k\mathfrak{U}, k\underline{w})(\underline{x}).$$

Finally, one can define trellis isomorphism: $(\mathfrak{U}, \mathfrak{L}) \sim (\mathfrak{V}, \mathfrak{K})$ if, after relabeling, the relevant lattice points satisfy the same averages. Clearly, if $(\mathfrak{V}, \mathfrak{K})$ is the image of $(\mathfrak{U}, \mathfrak{L})$ under a linear map, then averages are preserved. It can be proved that this is the only possible instance of trellis isomorphism: the matrix A contains enough information to give the barycentric coordinates of $\mathfrak{L}$ with respect to $\mathfrak{U}$. Suppose $p(\underline{x})=\sum h_k^2(\underline{x})$, $h_k(\underline{x})=\sum b_k(\underline{v})\underline{x}^{\underline{v}}$ and $[t_{ij}]$ is a matrix with rational entries so that $v_i'=\sum t_{ij} v_j$ is integral for every $\underline{v}$ occurring in any h_k. Under the formal substitution $x_i \rightarrow \underline{x}^{\underline{t}_i}$, we have $\underline{v} \rightarrow \underline{v}'$, and if p' and h_k' are the resulting forms, then

$$p'(\underline{x})=\sum_k h_k'^2(\underline{x})=\sum_k\Big(\sum_{\underline{v}} b_k(\underline{v}) \underline{x}^{\underline{v}'}\Big)^2.$$

In particular, if $\underline{t}_i=r\underline{e}_i$, then the substitution means replacing the variable x_i by x_i^r. It is not necessary for the t_{ij}'s to be integers; we can invert the previous example and replace x_i by $x_i^{1/r}$ in the form p'.

5. Examples of Sos Agiforms

In this section we implement the algorithm of Theorem 4.4, and write the forms $H(x, y, z)$ and $M(x^k, y^k, z^k)$ for $k=2,3$ as sums of squares. We briefly report some complications for non-simplicial agiforms.

(5.1) *Example* (Continuing Examples 1.10 and 2.4). We wish to write H as a sum of squares. Let $\mathfrak{U}=\mathfrak{H}$ and let $\mathfrak{L}$ consist of the following eight points: $\underline{u}_1=(6,0,0)$, $\underline{u}_2=(0,6,0)$, $\underline{u}_3=(0,0,6)$, $\underline{w}_1=(2,2,2)$, $\underline{w}_2=(2,4,0)$, $\underline{w}_3=(2,0,4)$, $\underline{w}_4=(4,2,0)$, and $\underline{w}_5=(4,0,2)$. We check that $\mathfrak{L}$ is $\mathfrak{U}$-mediated: $\underline{w}_1=\frac{1}{2}(\underline{w}_2+\underline{w}_3)$, $\underline{w}_2=\frac{1}{2}(\underline{w}_4+\underline{u}_2)$, $\underline{w}_3=\frac{1}{2}(\underline{w}_5+\underline{u}_3)$, $\underline{w}_4=\frac{1}{2}(\underline{w}_2+\underline{u}_1)$, and $\underline{w}_5=\frac{1}{2}(\underline{w}_3+\underline{u}_1)$. [By Theorem 2.8, we could have selected $E(\mathfrak{H})$ as the $\mathfrak{U}$-mediated set containing $(2,2,2)$; the advantage of $\mathfrak{L}$ is that it has fewer elements.] Using the terminology of the proof of Theorem 4.4, we have:

$$A=\begin{bmatrix} 2 & -1 & -1 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 \\ 0 & 0 & 2 & 0 & -1 \\ 0 & -1 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 2 \end{bmatrix},$$

and hence

$$6A^{-1}=\begin{bmatrix} 3 & 2 & 2 & 1 & 1 \\ 0 & 4 & 0 & 2 & 0 \\ 0 & 0 & 4 & 0 & 2 \\ 0 & 2 & 0 & 4 & 0 \\ 0 & 0 & 2 & 0 & 4 \end{bmatrix}.$$

The 5-tuple of binomial squares is

$$((xy^2-xz^2)^2,\ (y^3-x^2y)^2,\ (z^3-x^2z)^2,\ (x^3-xy^2)^2,\ (x^3-xz^2)^2)^T.$$

Since $2H(\underset{\sim}{x})=6f(\mathfrak{H},\underset{\sim}{w}_1)(\underset{\sim}{x})$, we can read off $2H$ as a sum of five binomial squares from the first row of $6A^{-1}$, compare with (1.12) and [19, p. 111]:

$$\begin{aligned}(5.2)\qquad 2H(x,y,z)&=2x^6+2y^6+2z^6-6x^2y^2z^2\\&=3(xy^2-xz^2)^2+2(y^3-x^2y)^2+2(z^3-x^2z)^2\\&\quad+(x^3-xy^2)^2+(x^3-xz^2)^2.\end{aligned}$$

The variable x is distinguished in (5.2), and there are two similar expressions in which y and z are distinguished. These correspond to images of $\mathfrak{L}$ under a permutation of coordinates. It turns out that H is a sum of four squares [19, p. 112], but one is the square of a trinomial; H is not a sum of three squares of forms (see [7]).

(5.3) *Example* (Continuing Example 1.9). We turn to $M(x^2,y^2,z^2)$. By Proposition 2.7, $\mathfrak{U}=2\mathfrak{M}$ is an H-trellis; we choose a smaller set. It is easy to check that $\mathfrak{L}=2\mathfrak{M}\cup\{\underset{\sim}{w}_1,\underset{\sim}{w}_2,\underset{\sim}{w}_3\}$ is $2\mathfrak{M}$-mediated, where $\underset{\sim}{w}_1=(4,4,4)$, $\underset{\sim}{w}_2=(6,6,0)$, $\underset{\sim}{w}_3=(2,2,8)$. As before, there is only one way to write the $\underset{\sim}{w}_i$'s as averages from $\mathfrak{L}$. After some simplification, we obtain the following representation:

$$\begin{aligned}M(x^2,y^2,z^2)&=x^8y^4+x^4y^8+z^{12}-3x^4y^4z^4\\&=2(x^3y^3-xyz^4)^2+(x^4y^2-x^2y^4)^2+(z^6-x^2y^2z^2)^2.\end{aligned}$$

For $M(x^3,y^3,z^3)$, we use a shortcut: $M(x^3,y^3,z^3)=H(x^2y,xy^2,z^3)$. Thus we can take $x\to x^2y$, $y\to xy^2$, $z\to z^3$ in (5.2), and obtain

$$\begin{aligned}(5.4)\qquad 2M(x^3,y^3,z^3)&=2x^{12}y^6+2x^6y^{12}+2z^{18}-6x^6y^6z^6\\&=3(x^4y^5-x^2yz^6)^2+2(x^3y^6-x^5y^4)^2\\&\quad+2(z^9-x^4y^2z^3)^2+(x^6y^3-x^4y^5)^2+(x^6y^3-x^2yz^6)^2.\end{aligned}$$

We could, of course, invert the process and derive (5.2) from (5.4).

Here is a more leisurely representation of $M(x^3,y^3,z^3)$ as a sum of binomial squares; we have suppressed the implicit $3\mathfrak{M}$-mediated set.

$$\begin{aligned}10M(x^3,y^3,z^3)&=10x^{12}y^6+10x^6y^{12}+10z^{18}-30x^6y^6z^6\\&=15(x^2y^3z^4-x^4y^3z^2)^2+12(xy^2z^6-x^3y^4z^2)^2\\&\quad+9(x^2y^3z^4-x^6y^3)^2+8(z^9-x^2y^4z^3)^2\\&\quad+6(x^3y^2z^4-x^3y^6)^2+4(xy^2z^6-x^3y^6)^2\\&\quad+3(x^2yz^6-x^4y^3z^2)^2+2(z^9-x^4y^2z^3)^2\\&\quad+(x^6y^3-x^2yz^6)^2.\end{aligned}$$

(In this case, $|\mathfrak{L}\backslash\mathfrak{U}|=9$; it is easier in practice to solve for M as a linear combination of nine specified binomial squares than it is to invert the full 9×9 matrix.)

The major obstacle in generalizing Theorem 4.4 to non-simplicial agiforms is Lemma 4.1. Suppose $\mathfrak{U}$ is not a trellis, and $\underline{w}=\frac{1}{2}(\underline{s}+\underline{t})$ with $\underline{s},\underline{t}\in E(\mathfrak{U})$. For a particular $\underline{\lambda}\in\Lambda(\underline{w})$, there may not exist $\underline{\sigma}\in\Lambda(\underline{s})$ and $\underline{\tau}\in\Lambda(\underline{t})$ with $\underline{\lambda}=\frac{1}{2}(\underline{\sigma}+\underline{\tau})$. In this case, no identity of the shape (4.2) is applicable to $f(\mathfrak{U},\underline{\lambda},\underline{w})$. For example, if

$$\mathfrak{T}=\mathfrak{H}\cup\mathfrak{M}=\{(6,0,0),\ (0,6,0),\ (0,0,6),\ (4,2,0),\ (2,4,0)\},$$

and $\underline{w}=(2,2,2)$, then M and H are both agiforms on $\mathfrak{T}$ with the same $\underline{w}$, but one is sos and the other is not. The Motzkin form corresponds to $\underline{\lambda}=(0,0,\frac{1}{3},\frac{1}{3},\frac{1}{3})\in\Lambda(\underline{w})$. There are several ways to write $\underline{w}=\frac{1}{2}(\underline{s}+\underline{t})$, but if $\underline{\lambda}=\frac{1}{2}(\underline{\sigma}+\underline{\tau})$ with $\underline{\sigma}\in\Lambda(\underline{s})$ and $\underline{\tau}\in\Lambda(\underline{t})$, then $\sigma_i=\tau_i=0$ for $i\leqq 2$, because $\sigma_i,\tau_i\geqq 0$. Thus $\sum\sigma_i\underline{u}_i$ and $\sum\tau_i\underline{u}_i$ belong to $E(\mathfrak{M})$; since $\underline{w}\notin\bar{\Lambda}(E(\mathfrak{M}))$, this is a contradiction. Of course, the failure of a lemma to generalize does not mean that the theorem also fails to generalize, and the points $(6,0,0)$ and $(0,6,0)$ in this case are basically irrelevant to the form $f(\mathfrak{U},\underline{\lambda},\underline{w})$.

Here is a less trivial example.

(5.5) *Example.* Let $\mathfrak{B}=\{(6,0,0),\ (4,2,0),\ (2,4,0),\ (0,0,6)\}$. It is not hard to see that

$$E(\mathfrak{B})=\mathfrak{B}\cup\{(2,2,2),\ (2,0,4),\ (4,0,2)\}.$$

For $\underline{w}=(2,2,2)$, a routine computation gives:

$$\Lambda(\underline{w})=\{\underline{\lambda}_t=(\tfrac{1}{3}t,\tfrac{1}{3}(1-2t),\tfrac{1}{3}(1+t),\tfrac{1}{3}):0\leqq t\leqq\tfrac{1}{2}\}.$$

We define the agiform $M_t(x,y,z)=3f(\mathfrak{B},\underline{\lambda}_t,\underline{w})$ for $0\leqq t\leqq\frac{1}{2}$:

$$\begin{aligned}(5.6)\qquad M_t(x,y,z)&=tx^6+(1-2t)x^4y^2+(1+t)x^2y^4+z^6-3x^2y^2z^2\\&=M(x,y,z)+t(x^3-xy^2)^2.\end{aligned}$$

Note that M_t is a simplicial agiform when $t=0$ and $t=\frac{1}{2}$, and is a convex combination of M_0 and $M_{1/2}$ (see also Theorem 7.1). The identity

$$M_{1/8}(x,y,z)=(x^2z-z^3)^2+\tfrac{1}{8}(3xy^2+x^3-4xz^2)^2$$

shows that $M_{1/8}$ is sos. A form is called *sbs* (see [6]) if it is a sum of squares of binomials. By Theorem 4.4, every sos simplicial agiform is also sbs. The following identity shows that $M_{1/2}$ is sbs:

$$M_{1/2}(x,y,z)=\tfrac{3}{2}(xy^2-xz^2)^2+\tfrac{1}{2}(x^3-xz^2)^2+(x^2z-z^3)^2.$$

We shall prove in [7] that M_t is sos if and only if $t\geqq 1/8$ and M_t is sbs if and only if $t=1/2$, so that the two conditions are not equivalent for non-simplicial agiforms.

6. Six Families of Agiforms

In this section we consider six families of agiforms, two generalizations for each of M, S, and Q. One family is due to Motzkin [16, p. 217], three are due to Choi and Lam [4, p. 5] and the other two are new. The subscript of a form always indicates the number of variables.

Motzkin defined a family of psd forms $\{M_n\}$ which are not sos:

$$(6.1)\qquad M_n(\underline{x})=x_1^2\dots x_{n-1}^2\left(\sum_{i=1}^{n-1}x_i^2\right)+x_n^{2n}-nx_1^2\dots x_n^2,$$

so $M_3 = M$ [cf. (1.14)]. [Motzkin actually defined $M_n(x_1, \ldots, x_{n-1}, 1)$; here as elsewhere in this section, we have renamed and renormalized the families.] Choi and Lam gave an alternative generalization of M:

$$\bar{M}_n(\underline{x}) = (n-2)x_n^{2n} + \sum_{i \neq j} x_i^{2n-2}x_j^2 - n(n-2)x_1^2 \ldots x_n^2, \tag{6.2}$$

where the sum is taken over all pairs (i,j) with $1 \leqq i \neq j \leqq n-1$.

They also defined the family $\{S_n\}$:

$$S_n(\underline{x}) = \sum_{i=1}^{n} x_i^{2n-2}x_{i+1}^2 - nx_1^2 \ldots x_n^2, \tag{6.3}$$

where $x_{n+1} = x_1$ and $S_3 = S$. We consider an alternative family:

$$\overline{S_n}(\underline{x}) = x_1^2x_2^2 \ldots x_n^2 \left(\sum_{i=1}^{n} x_i^{-2}x_{i+1}^2 - n \right), \tag{6.4}$$

where, again, $x_{n+1} = x_1$ [cf. (1.15)].

Finally, Choi and Lam generalized Q for $n = 2m$:

$$Q_{2m}(\underline{x}) = (2m-2)!\, x_{2m}^{2m} + m!\,(m-1)! \sum_{i_1 < \ldots < i_m} x_{i_1}^2 \ldots x_{i_m}^2 - 2m(2m-2)!\, x_1 \ldots x_{2m}, \tag{6.5}$$

where the sum is taken over all m-subsets of $\{1, \ldots, 2m-1\}$ and $Q_4 = 2Q$. We also consider an alternative family $\{\bar{Q}_{2m}\}$:

$$\bar{Q}_{2m}(\underline{x}) = x_{2m}^{2m} + \sum_{i=1}^{2m-1} x_i^2x_{i+1}^2 \ldots x_{i+(m-1)}^2 - 2mx_1 \ldots x_{2m}. \tag{6.6}$$

The summands are products of m consecutive elements of $\{x_1^2, x_2^2, \ldots, x_{2m-1}^2\}$, viewed cyclically, and $\overline{Q_4} = Q$ [c.f. (1.16)].

We shall show that M_n, S_n, $\bar{S}_n$, and $\bar{Q}_{2m}$ are simplicial agiforms on M-trellises, while $\bar{M}_n$ $(n \geqq 4)$ and Q_{2m} $(m \geqq 3)$ are non-simplicial agiforms. We use Theorem 3.3 to verify the assertions in [16] and [4] that M_n, S_n, and Q_{2m} are not sos; $\bar{S}_n$ and $\bar{Q}_{2m}$ are also not sos. We write $\bar{M}_4$ as a sum of squares, contradicting [4, p. 5]. We compute $E(\mathfrak{U})$ for the appropriate $\mathfrak{U}$; this is used in Theorem 9.1 to show that M_n, $\overline{S_n}$, and $\overline{Q_{2m}}$ are extremal.

Our arguments are simplified by a technical lemma and a proposition on the eigenvalues of a circulant matrix. For notational convenience, we let $\underline{\varphi}_k$ denote the k-tuple $\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)$.

(6.7) **Lemma.** *Suppose $\underline{\sigma} = (\sigma_1, \ldots, \sigma_n)$ with $\sigma_i \geqq 0$ and $\sum \sigma_i = 1$ and suppose that $2(\sigma_i - \sigma_j)$ is an integer for all i, j. Then $\underline{\sigma} = \underline{e}_j$, $\underline{\varphi}_n$, $\frac{1}{2}(\underline{e}_j + \underline{e}_k)$* or *$\frac{1}{2}(\underline{e}_j + \underline{\varphi}_n)$.*

Proof. Let $s = \min \sigma_i$, so $\underline{\sigma} = (s, \ldots, s) + \underline{\alpha}$, where each α_i is a multiple of $1/2$, $s \geqq 0$, and $ns + \sum \alpha_i = 1$. If $\sum \alpha_i = 1$, then $s = 0$ and $\underline{\sigma} = \underline{e}_j$ (if $\alpha_j = 1$) or $\underline{\sigma} = \frac{1}{2}(\underline{e}_j + \underline{e}_k)$ (if $\alpha_j = \alpha_k = 1/2$). If $\sum \alpha_i = 0$, then $\underline{\alpha} = 0$, $s = 1/n$ and $\underline{\sigma} = \underline{\varphi}_n$. If $\sum \alpha_i = 1/2$, then $\alpha_j = 1/2$, $s = 1/2n$ and $\underline{\sigma} = \frac{1}{2}(\underline{e}_j + \underline{\varphi}_n)$. $\square$

A *circulant* matrix is a square matrix whose rows are successive cyclic permutations of the first. To be specific, $C=[c_{ij}]=\mathrm{circ}(a_1,\dots,a_n)$ is the $n\times n$ matrix with $c_{ij}=a_{j-i+1}$, with the index k in a_k reduced mod n. Circulant matrices are studied in great detail in [8], from which we cite the following result (pp. 66–73).

(6.8) **Proposition.** *The eigenvalues of the matrix* $\mathrm{circ}(a_1,\dots,a_n)$ *are* $\lambda_k=\sum_{i=1}^{n} a_i\varepsilon^{k(i-1)}$, *for* $0\leqq k\leqq n-1$, *where* ε *is a primitive n-th root of unity. The eigenvalues of* $\mathrm{circ}(a_1+d,\dots,a_n+d)$ *are thus* λ_0+nd, $\lambda_1,\dots,\lambda_{n-1}$.

If $\mathfrak{U}=\{\underline{u}_1,\dots,\underline{u}_n\}\subset\mathbb{R}^n$ is a framework, then $\mathfrak{U}$ is a trellis if and only if $\det[u_{ij}]\neq 0$. For many examples in this section, the matrix $[u_{ij}]$ is circulant (or has a large circulant block), and we shall use Proposition 6.8 to show that the circulant block has non-zero determinant.

(6.9) **Theorem.** *For* $n\geqq 3$, *let*

(6.10)
$$\mathfrak{M}_n=\{(4,2,\dots,2,0),\,(2,4,\dots,2,0),\dots,(2,2,\dots,4,0),\,(0,0,\dots,0,2n)\}=\{\underline{u}_1,\dots,\underline{u}_n\}.$$

Then $\mathfrak{M}_n$ *is a trellis and* M_n *is a simplicial agiform on* $\mathfrak{M}_n$ *which is not* sos. *Further,* $E(\mathfrak{M}_n)=\mathfrak{M}_n\cup\{\underline{w}\}$, *where* $\underline{w}=(2,\dots,2,2)$, *and* $C(\mathfrak{M}_n)=A(E(\mathfrak{M}_n))$.

Proof. Note that $[u_{ij}]$ is a block matrix, consisting of the $n-1\times n-1$ block $\mathrm{circ}(4,2,\dots,2)$ and the 1×1 block $[2n]$. By applying Proposition 6.8 with $d=2$ to $\mathrm{circ}(2,0,\dots,0)=2I$, we find that $\lambda_0=2n$ and $\lambda_k=2$ for $k\neq 0$; thus $\mathrm{circ}(4,2,\dots,2)$ is non-singular and so $\mathfrak{U}$ is a trellis. It is easy to check via (6.1) that $\underline{w}=\sum\tau_i\underline{u}_i$, where $\underline{\tau}=\varphi_n$, so $M_n=nf(\mathfrak{M}_n,\underline{w})$. We shall show that $E(\mathfrak{M}_n)=\mathfrak{M}_n\cup\{\underline{w}\}$; as $\underline{w}\notin A(E(\mathfrak{M}_n))$ for $n\geqq 3$ ($v_n\in\{0,2n\}$ for $\underline{v}\in E(\mathfrak{M}_n)\setminus\{\underline{w}\}$), it follows from Corollary 3.4 that M_n is not sos.

Suppose $\underline{u}=(u_1,\dots,u_n)=\sum\lambda_i\underline{u}_i\in C(\mathfrak{M}_n)$. Then, from (6.10),

$$u_j=2\lambda_1+\dots+4\lambda_j+\dots+2\lambda_{n-1},\quad 1\leqq j\leqq n-1,\quad\text{and}\quad u_n=2n\lambda_n.$$

Since $\sum\lambda_i=1$, $u_j=2+2\lambda_j-2\lambda_n$ for $1\leqq j\leqq n-1$, so $2(\lambda_j-\lambda_n)$ is an integer; hence $2(\lambda_i-\lambda_j)$ is also an integer for all i,j. By Lemma 6.7, either $\underline{\lambda}=\underline{e}_j$ (and $\underline{u}=\underline{u}_j$), $\underline{\lambda}=\frac{1}{2}(\underline{e}_j+\underline{e}_k)$ [and $\underline{u}=\frac{1}{2}(\underline{u}_j+\underline{u}_k)$], $\underline{\lambda}=\varphi_n$ (and $\underline{u}=\underline{w}$), or $\underline{\lambda}=\frac{1}{2}(\underline{e}_j+\varphi_n)$ [and $\underline{u}=\frac{1}{2}(\underline{u}_j+\underline{w})$.] Since $\underline{w}$ and the $\underline{u}_i$'s are even, each average is a lattice point. Each of the distinct averages has at least one odd component, hence $E(\mathfrak{M}_n)=\mathfrak{M}_n\cup\{\underline{w}\}$ and $C(\mathfrak{M}_n)=A(E(\mathfrak{M}_n))$. □

(6.11) *Example.* Let $\bar{\mathfrak{M}}_n$ be the framework implicit in the definition of the family $\{\bar{M}_n\}$; $\bar{\mathfrak{M}}_n$ has $1+(n-1)(n-2)$ elements and so cannot be a trellis for $n\geqq 4$. When $n=3$, $\bar{M}_3=M$, which we know to be a non-sos simplicial agiform. It turns out that $\bar{M}_n=n(n-2)f(\bar{\mathfrak{M}}_n,\underline{\lambda},\underline{w})$ for $\underline{w}=(2,\dots,2)$ and suitable $\underline{\lambda}$. We have found representations for $\bar{M}_4$, $\bar{M}_5$, and $\bar{M}_6$ as a sum of squares and conjecture that $\bar{M}_n$ is sos for $n\geqq 4$. Here is $\bar{M}_4$:

$$\begin{aligned}&\bar{M}_4(x,y,z,w)\\&=2w^8+x^6y^2+x^6z^2+y^6x^2+y^6z^2+z^6x^2+z^6y^2-8x^2y^2z^2w^2\\&=2(w^4-x^2y^2)^2+4(xyz^2-xyw^2)^2+2(x^2y^2-x^2z^2)^2\\&\quad+(x^3y-xy^3)^2+(y^3z-yz^3)^2+(z^3x-zx^3)^2.\end{aligned}$$

(6.12) Theorem. *For $n \geqq 3$, let*

(6.13)
$$\mathfrak{S}_n = \{(2n-2, 2, 0, \ldots, 0), (0, 2n-2, 2, \ldots, 0), \ldots, (2, 0, 0, \ldots, 2n-2)\} = \{\underline{u}_1, \ldots, \underline{u}_n\}.$$

Then $\mathfrak{S}_n$ is a trellis and S_n is a simplicial agiform on $\mathfrak{S}_n$ which is not sos. *Further, $E(\mathfrak{S}_n) = \mathfrak{S}_n \cup \{\underline{w}\}$, where $\underline{w} = (2, \ldots, 2)$.*

Sketch of Proof. After showing that $[u_{ij}]$ is circulant and $\mathfrak{S}_n$ is a trellis, one shows that $S_n = nf(\mathfrak{S}_n, \underline{w})$ is a simplicial agiform with $\underline{\tau} = \varphi_n$. Again, $E(\mathfrak{S}_n) = \mathfrak{S}_n \cup \{\underline{w}\}$ will imply that $\underline{w}$ does not belong to $A(E(\mathfrak{S}_n))$, and so S_n is not sos.

If $\underline{u} = \sum \lambda_i \underline{u}_i \in E(\mathfrak{S}_n)$, then for all j, $u_j = 2\lambda_{j-1} + (2n-2)\lambda_j$ from (6.13), with $\lambda_0 = \lambda_n$. Since $\sum u_j = 2n$, if $u_j > 0$ for all j, then $u_j \geqq 2$, hence $\underline{u} = \underline{w}$. Otherwise $u_k = 0$ for some k, so $\lambda_{k-1} = \lambda_k = 0$ and $u_{k-1} = 2\lambda_{k-2}$. Thus λ_{k-2} is 0 or 1. If $\lambda_{k-2} = 1$, then $\underline{\lambda} = \underline{e}_{k-2}$ and $\underline{u} = \underline{u}_{k-2}$. If $\lambda_{k-2} = 0$, then $u_{k-1} = 0$, and the argument may be repeated. Since some u_j is positive, we eventually get $\underline{u} = \underline{u}_i \in \mathfrak{S}_n$ for some i. Thus $E(\mathfrak{S}_n) = \mathfrak{S}_n \cup \{\underline{w}\}$. The set $C(\mathfrak{S}_n)$ is larger than $A(E(\mathfrak{S}_n))$ for $n \geqq 4$ (see Theorem 9.3). □

(6.14) Theorem. *For $n \geqq 3$, let*

(6.15)
$$\bar{\mathfrak{S}}_n = \{(4, 2, 2, 2, \ldots, 0), (0, 4, 2, 2, \ldots, 2), (2, 0, 4, 2, \ldots, 2), \ldots\} = \{\underline{u}_1, \ldots, \underline{u}_n\},$$

so that $[u_{ij}]$ is circulant. Then $\bar{\mathfrak{S}}_n$ is a trellis and $\bar{S}_n$ is a simplicial agiform on $\bar{\mathfrak{S}}_n$ which is not sos. *Further, $E(\bar{\mathfrak{S}}_n) = \bar{\mathfrak{S}}_n \cup \{\underline{w}\}$, where $\underline{w} = (2, \ldots, 2)$, and $C(\bar{\mathfrak{S}}_n) = A(E(\bar{\mathfrak{S}}_n))$.*

The analysis of $\{\bar{\mathfrak{S}}_n\}$ is quite similar to that of $\{\mathfrak{M}_n\}$ and is omitted.

(6.16) Theorem. *For $m \geqq 2$, let $\mathfrak{Q}_{2m}$ denote the framework:*

(6.17)
$$\mathfrak{Q}_{2m} = \{(0, \ldots, 0, 2m)\} \cup \{(2a_1, \ldots, 2a_{2m-1}, 0) : a_j \in \{0, 1\}, \textstyle\sum a_j = m\} = \{\underline{u}_i\}.$$

Then $E(\mathfrak{Q}_{2m}) = \mathfrak{Q}_{2m}$ and Q_{2m} is an agiform on $\mathfrak{Q}_{2m}$ which is not sos.

Proof. As $\mathfrak{Q}_{2m}$ has $1 + \binom{2m-1}{m}$ elements, it is not a trellis for $2m \geqq 6$; it is easy to see that $\mathfrak{Q}_4$ is a trellis. Let $\underline{\lambda} = \left(\frac{1}{2m}, \gamma, \ldots, \gamma\right)$, where $\gamma = \frac{1}{2}\binom{2m-2}{m-1}^{-1}$ and let $\underline{w} = \sum \lambda_i \underline{u}_i \in C(\mathfrak{Q}_{2m})$. It is easy to check that $\sum \lambda_i = 1$ and $\underline{w} = (1, \ldots, 1, 1)$, so $Q_{2m} = 2m(2m-2)!\, f(\mathfrak{Q}_{2m}, \underline{\lambda}, \underline{w})$.

Since $\mathfrak{Q}_{2m}$ lies in the slab $0 \leqq x_j \leqq 2$ for $j = 1, \ldots, 2m-1$, if $\underline{u} \in E(\mathfrak{Q}_{2m})$, then $u_j = 0$ or 2 for $j < 2m$. If $u_k = 2$ for some k, then $\underline{u}$ is a convex combination of those $\underline{u}_i$'s for which $u_{i,k} = 2$, so $u_{2m} = 0$. Since $\sum u_j = 2m$, this implies that $\underline{u} \in \mathfrak{Q}_{2m}$. Otherwise, $u_j = 0$ for $1 \leqq j \leqq 2m-1$, so $\underline{u} = \underline{u}_1 \in \mathfrak{Q}_{2m}$. By looking at w_{2m}, we see that $\underline{w} \notin A(\mathfrak{Q}_{2m})$, so Q_{2m} is not sos. □

(6.18) Theorem. *For $m \geqq 2$ let $\bar{\mathfrak{Q}}_{2m} = \{\underline{u}_i\}$ be defined so that $[u_{ij}]$ has a $2m-1 \times 2m-1$ circulant block,* circ$(2, \ldots, 2, 0, \ldots, 0)$, *with m 2's and $m-1$ 0's, and the 1×1 block $2m$. Then $\bar{\mathfrak{Q}}_{2m}$ is a trellis, $E(\bar{\mathfrak{Q}}_{2m}) = \bar{\mathfrak{Q}}_{2m}$ and $C(\bar{\mathfrak{Q}}_{2m}) = A(\bar{\mathfrak{Q}}_{2m}) \cup \{\underline{w}\}$, where $\underline{w} = (1, \ldots, 1)$, and $\bar{Q}_{2m}$ is a non-*sos *simplicial agiform on $\bar{\mathfrak{Q}}_{2m}$.*

Proof. In this case, $\lambda_k = 2\sum_{i=1}^{m} \varepsilon^{k(i-1)}$, so $\lambda_0 = 2m$; as ε^m is a primitive $(2m-1)$-st root of unity, $\lambda_k = 2(1-\varepsilon^{km})/(1-\varepsilon^k) \neq 0$ for $k \geqq 1$. Thus $\bar{\mathfrak{Q}}_{2m}$ is a trellis; $\bar{Q}_{2m} = 2mf(\bar{\mathfrak{Q}}_{2m}, \underline{w})$. It is possible to argue as in Theorem 6.16 that $E(\bar{\mathfrak{Q}}_{2m}) = \bar{\mathfrak{Q}}_{2m}$; as $\underline{w} \notin A(E(\bar{\mathfrak{Q}}_{2m}))$, $\bar{Q}_{2m}$ is not sos. This may be proved more directly. For any permutation σ of $\{1, \ldots, 2m-1\}$, let

$$\sigma \underline{x} = (x_{\sigma(1)}, \ldots, x_{\sigma(2m-1)}, x_{2m}).$$

It is easily seen that $Q_{2m}(\underline{x})$ is the average (over σ) of $\bar{Q}_{2m}(\sigma\underline{x})$, up to a multiple. If $\bar{Q}_{2m}$ were sos, then Q_{2m} would also be sos, a contradiction.

Suppose $\underline{v} = \sum \lambda_i \underline{u}_i \in C(\bar{\mathfrak{Q}}_{2m})$, then

$$\text{(6.19)} \qquad v_i = 2(\lambda_i + \lambda_{i-1} + \ldots \lambda_{i-(m-1)}), \qquad 1 \leqq i \leqq m-1; \qquad v_{2m} = 2m\lambda_{2m},$$

where the indices are understood to cycle so that $\lambda_0 = \lambda_{2m-1}$, $\lambda_{-1} = \lambda_{2m-2}$, etc. With the same cyclic understanding, (6.19) implies that:

$$v_i + v_{i+(m-1)} = 4\lambda_i + 2\sum_{\substack{j=1 \\ j \neq i}}^{2m-1} \lambda_j = 2 + 2\lambda_i - 2\lambda_{2m} \quad \text{for} \quad 1 \leqq i \leqq 2m-1.$$

As before, $2\lambda_i - 2\lambda_{2m}$ is an integer for $1 \leqq i \leqq 2m-1$, and we may apply Lemma 6.7. In this case, $\frac{1}{2}(\underline{w} + \underline{u}_k)$ is not a lattice point because its $2m$-th component is not an integer. Thus $C(\bar{\mathfrak{Q}}_{2m}) = A(\bar{\mathfrak{Q}}_{2m}) \cup \{\underline{w}\}$ and it follows easily that $E(\bar{\mathfrak{Q}}_{2m}) = \bar{\mathfrak{Q}}_{2m}$. □

7. Extremal Agiforms: The Sufficient Condition

Recall that a psd form p is extremal if $p = h_1 + h_2$ with h_i psd implies that $h_i = \alpha_i p$, or equivalently, if $p(\underline{x}) \geqq h(\underline{x}) \geqq 0$ for all $\underline{x}$ (written $p \geqq h \geqq 0$) implies that $h = \alpha p$. Choi and Lam proved that the agiforms M, S, and Q are extremal, and it was shown in [17] that any extremal form with four or fewer terms is, up to a scaling of variables, either a monomial square or a simplicial agiform $f = f(\mathfrak{U}, \underline{w})$, where $|\mathfrak{U}| = 3$, $E(\mathfrak{U}) = \mathfrak{U} \cup \{\underline{w}\}$ and $\underline{w} = \frac{1}{3}(\underline{u}_1 + \underline{u}_2 + \underline{u}_3)$. In the next two sections, we generalize the methods of [17] to agiforms in more than three variables. In this section, we give a weaker version of extremality and discuss the zeros of agiforms. This gives a sufficient condition, which we later prove is necessary.

An agiform f is *primitive* if $f = \sum g_i$, where each g_i is an agiform, implies that $g_i = \alpha_i f$. (Keep in mind that agiforms do not comprise a cone.) An extremal agiform is primitive; it was shown in [17] that the converse is true when $n \leqq 3$, it is false when $n = 4$. It turns out that this is a geometric condition: f is primitive if and only if it is a simplicial agiform on a trellis $\mathfrak{U}$ with $E(\mathfrak{U}) \subsetneqq (\mathfrak{U} \cup \{\underline{w}\})$. (Necessity follows from Lemma 4.1.)

We begin with two decomposition theorems of independent interest.

(7.1) **Theorem.** *Every agiform is a convex combination of simplicial agiforms.*

Proof. Fix a framework $\mathfrak{U} = \{\underline{u}_1, \ldots, \underline{u}_m\}$, and let $f = f(\mathfrak{U}, \underline{\lambda}, \underline{w})$ be an agiform on $\mathfrak{U}$. The set $A(\underline{w})$ is the intersection of an affine subspace of $\mathbb{R}^m$ (the solutions to $\sum \lambda_i \underline{u}_i = \underline{w}$) and a simplex ($\lambda_i \geqq 0$, $\sum \lambda_i = 1$), and so is a closed convex polytope in $\mathbb{R}^m$. Hence $\underline{\lambda}$ is a convex combination of extreme points, say $\underline{\lambda} = \sum \beta_j \underline{\sigma}_j$, where $\underline{\sigma}_j$ is extremal, $\beta_j > 0$ and $\sum \beta_j = 1$. Thus

$$f(\mathfrak{U}, \underset{\sim}{\lambda}, \underset{\sim}{w})(\underset{\sim}{x}) = \sum_i \lambda_i \underset{\sim}{x}^{u_i} - \underset{\sim}{x}^{w} = \sum_i \Big(\sum_j \beta_j \sigma_{ji} \underset{\sim}{x}^{u_i}\Big) - \underset{\sim}{x}^{w}$$

$$= \sum_j \beta_j \Big(\sum_i \sigma_{ji} \underset{\sim}{x}^{u_i} - \underset{\sim}{x}^{w}\Big) = \sum_j \beta_j f(\mathfrak{U}, \underset{\sim}{\sigma}_j, \underset{\sim}{w})(\underset{\sim}{x}).$$

It follows that f is not extremal unless $\underset{\sim}{\lambda}$ is an extreme point in $\Lambda(\underset{\sim}{w})$. We now show that if $\underset{\sim}{\sigma}$ is an extreme point, then $f(\mathfrak{U}, \underset{\sim}{\sigma}, \underset{\sim}{w})$ is simplicial. After reindexing, we assume that $\sigma_i > 0$ precisely for $1 \leqq i \leqq r$; we now show that $\mathfrak{U}' = \{\underset{\sim}{u}_i : 1 \leqq i \leqq r\}$ is a linearly independent set, so $f = f(\mathfrak{U}', \underset{\sim}{w})$ is a simplicial agiform. Suppose $\sum_{i=1}^{r} \alpha_i \underset{\sim}{u}_i = \underset{\sim}{0}$; then $\sum_{j=1}^{n} \Big(\sum_{i=1}^{r} \alpha_i u_{ij}\Big) = 2d \Big(\sum_{i=1}^{r} \alpha_i\Big) = 0$. Let $\alpha_j = 0$ for $j > r$ and let $\underset{\sim}{\alpha} = (\alpha_1, \ldots, \alpha_m)$. Then $\sum_{i=1}^{m} (\sigma_i + c\alpha_i)\underset{\sim}{u}_i = \underset{\sim}{w}$ and $\sum_{i=1}^{m} (\sigma_i + c\alpha_i) = 1$ for all c. For $|c|$ sufficiently small, $\sigma_i + c\alpha_i \geqq 0$ for all i, so $\underset{\sim}{\sigma} + c\underset{\sim}{\alpha} \in \Lambda(\underset{\sim}{w})$. This contradicts the extremality of $\underset{\sim}{\sigma}$ unless $\underset{\sim}{\alpha} = \underset{\sim}{0}$. $\square$

(7.2) **Theorem.** *If $f = f(\mathfrak{U}, \underset{\sim}{w})$ is a simplicial agiform, where $\underset{\sim}{w}$ is interior to $\mathfrak{U}$, and $\underset{\sim}{s} \in E(\mathfrak{U})$, $\underset{\sim}{s} \notin (\mathfrak{U} \cup \{\underset{\sim}{w}\})$, then $f(\mathfrak{U}, \underset{\sim}{w}) = f(\mathfrak{B}, \underset{\sim}{w}) + \beta f(\mathfrak{U}, \underset{\sim}{s})$, where $\beta > 0$ and $\mathfrak{B}$ is a trellis contained in $E(\mathfrak{U})$.*

Proof. Suppose $\underset{\sim}{s} = \sum \alpha_i \underset{\sim}{u}_i \in E(\mathfrak{U})$, where $\underset{\sim}{s} \notin \mathfrak{U} \cup \{\underset{\sim}{w}\}$ and $\underset{\sim}{\alpha} \in \Lambda(\underset{\sim}{s})$. Since $\underset{\sim}{\lambda}$ is interior, $\lambda_i > 0$; after reindexing, we may assume that $\alpha_1/\lambda_1 \geqq \alpha_j/\lambda_j$ for $2 \leqq j \leqq m$, so $\alpha_1 > 0$. Let $\mathfrak{B} = \{\underset{\sim}{s}, \underset{\sim}{u}_2, \ldots, \underset{\sim}{u}_m\} (\neq \mathfrak{U})$. Define $\underset{\sim}{\beta}$ by $\beta_1 = \lambda_1/\alpha_1$ and $\beta_j = \lambda_j - \alpha_j \beta_1$ for $2 \leqq j \leqq m$. Then $\beta_i > 0$ and a routine computation, which we omit, shows that the barycentric coordinates of $\underset{\sim}{w}$ with respect to $\mathfrak{B}$ are given by $(\beta_1, \ldots, \beta_m)$. It follows that

$$f(\mathfrak{B}, \underset{\sim}{\beta}, \underset{\sim}{w})(\underset{\sim}{x}) + \beta_1 f(\mathfrak{U}, \underset{\sim}{\alpha}, \underset{\sim}{s})(\underset{\sim}{x})$$

$$= \beta_1 \underset{\sim}{x}^{s} + \sum_{j=2}^{m} (\lambda_j - \beta_1 \alpha_j)\underset{\sim}{x}^{u_j} - \underset{\sim}{x}^{w} + \beta_1\Big(\alpha_1 \underset{\sim}{x}^{u_1} + \sum_{j=2}^{m} \alpha_j \underset{\sim}{x}^{u_j} - \underset{\sim}{x}^{s}\Big)$$

$$= \sum_{j=1}^{m} \lambda_j \underset{\sim}{x}^{u_j} - \underset{\sim}{x}^{w} = f(\mathfrak{U}, \underset{\sim}{\lambda}, \underset{\sim}{w})(\underset{\sim}{x}). \quad \square$$

Geometrically speaking, the point $\underset{\sim}{s}$ subdivides the simplex $cvx(\mathfrak{U})$ into subsimplices, and $\underset{\sim}{w}$ lies in the one which maximizes α_j/λ_j.

(7.3) **Theorem.** *The agiform $f = f(\mathfrak{U}, \underset{\sim}{\lambda}, \underset{\sim}{w})$ is primitive if and only if it is a simplicial agiform on a trellis $\mathfrak{U}$ for which $E(\mathfrak{U}) \subseteq (\mathfrak{U} \cup \{\underset{\sim}{w}\})$.*

Proof. First suppose f is primitive; by Theorem 7.1, f is a simplicial agiform. Choose $\mathfrak{U}$ so that $f = f(\mathfrak{U}, \underset{\sim}{w})$ and write $\underset{\sim}{w} = \sum \lambda_i \underset{\sim}{u}_i$ with $\lambda_i > 0$. If $\underset{\sim}{s} \in E(\mathfrak{U})$, $\underset{\sim}{s} \notin (\mathfrak{U} \cup \{\underset{\sim}{w}\})$; by Theorem 7.2, $f(\mathfrak{U}, \underset{\sim}{w})$ is a convex combination of $f(\mathfrak{U}', \underset{\sim}{w})$ and $f(\mathfrak{U}, \underset{\sim}{s})$. Since the term $\underset{\sim}{x}^{s}$ has coefficient -1 in $f(\mathfrak{U}, \underset{\sim}{s})$ and does not appear in $f(\mathfrak{U}, \underset{\sim}{w})$, $f(\mathfrak{U}, \underset{\sim}{s}) \neq \alpha f(\mathfrak{U}, \underset{\sim}{w})$; this contradicts the primitivity of f.

Suppose f is simplicial on $\mathfrak{U}$ with $E(\mathfrak{U}) \subseteq (\mathfrak{U} \cup \{\underset{\sim}{w}\})$ and $f(\mathfrak{U}, \underset{\sim}{w}) = \sum g_i$, where $g_i \neq \alpha_i f$ is an agiform. Since the coefficient of $\underset{\sim}{x}^{w}$ is -1 in f, it must be negative in some g_i, so $g_i = \alpha_i f(\mathfrak{B}, \underset{\sim}{w})$ for some $\mathfrak{B}$. But $f \geqq g_i$, so by Theorem 3.6(i), $N(f) = \mathfrak{U} \supsetneq \mathfrak{B} = N(g_i)$. This is a contradiction, since $\underset{\sim}{w}$ is not contained in the convex hull of a proper subset of $\mathfrak{U}$. $\square$

We remark that, if $f(\mathfrak{U}, \underline{w})$ is primitive, then either f is a binomial square or $\mathfrak{U}$ is an M-trellis by Theorem 2.5 and f is not sos. Since extremality implies primitivity, we have already shown that, if $f(\mathfrak{U}, \underline{\lambda}, \underline{w})$ is extremal, then f is simplicial and $E(\mathfrak{U}) \subseteqq \mathfrak{U} \cup \{\underline{w}\}$.

Let p be a (not necessarily psd) form. The *zero-set of* p, $\{\underline{z} : p(\underline{z}) = 0\}$, is denoted by $\mathfrak{Z}(p)$. Since p is homogeneous, $\underline{z} \in \mathfrak{Z}(p)$ if and only if $r\underline{z} \in \mathfrak{Z}(p)$ for all real r. It will be convenient to consider $\underline{z}$ and $-\underline{z}$ separately, so we do not define $\mathfrak{Z}(p)$ projectively. Let D_i denote the operator $\partial/\partial x_i$. If $h = h(x_1, \dots, x_n)$ is a form and T is a subset of $\mathbb{R}^n$, then h is *second-order at* T if $D_i h(\underline{y}) = 0$ for all $\underline{y} \in T$ and $i = 1, \dots, n$. If p is psd and $p(\underline{z}) = 0$, then $D_i p(\underline{z}) = 0$ for all i, so p is second-order at $\mathfrak{Z}(p)$. If p is an *m-ic* form which is second-order at T, then $\sum x_i D_i p(\underline{x}) = mp(\underline{x})$ implies that $T \subseteqq \mathfrak{Z}(p)$.

(7.4) Lemma. *Suppose $z_i \neq 0$ for all i and all $\underline{z} \in T$ and $h(\underline{x}) = \sum_{k=1}^{s} c(\underline{v}_k)\underline{x}^{\underline{v}_k}$. Then h is second-order at T if and only if, for all $\underline{z} \in T$ and $i = 1, \dots, n$,*

$$\sum_{k=1}^{s} \{v_{ki}\underline{z}^{\underline{v}_k}\} c(\underline{v}_k) = 0 . \tag{7.5}$$

Proof. Since $x_i D_i(\underline{x}^{\underline{v}}) = v_i \underline{x}^{\underline{v}}$, the left-hand side of (7.5) is simply $z_i D_i h(\underline{z})$; $z_i D_i h(\underline{z}) = 0$ if and only if $D_i h(\underline{z}) = 0$ because $z_i \neq 0$. □

The relevance of second-order sets to extremal forms is shown by the next lemma, which follows from this discussion and Theorem 3.6.

(7.6) Lemma. *If p and q are forms and $p \geqq q \geqq 0$, then $N(p) \supseteqq N(q)$, $\mathfrak{Z}(p) \subseteqq \mathfrak{Z}(q)$ and q is second-order at $\mathfrak{Z}(p)$.*

For $\underline{v} \in \mathbb{Z}^n$, let $\mathfrak{O}(\underline{v}) = \{j : v_j \text{ is odd}\}$. Let $G_n = \{-1, 1\}^n$; G_n forms a group under component-wise multiplication: $\underline{\varepsilon} \cdot \underline{\varepsilon}' = (\varepsilon_1\varepsilon'_1, \dots, \varepsilon_n\varepsilon'_n)$. For a set $I \subseteqq \{1, \dots, n\}$, let $G_n(I)$ denote the subgroup $\left\{\underline{\varepsilon} \in G_n : \prod_{i \in I} \varepsilon_i = 1\right\}$, so $G_n(\phi) = G_n$ and $|G_n(I)| = 2^{n-1}$ for $I \neq \phi$. If $k \in I \setminus J$, then $\Big(1, \dots, \underset{k\text{-th}}{-1}, \dots, 1\Big) \in G_n(J) \setminus G_n(I)$; $I \neq J$ implies $G_n(I) \neq G_n(J)$. As $\underline{\varepsilon}^{\underline{v}} = \{\prod \varepsilon_i : v_i \text{ is odd}\}$, $G_n(\mathfrak{O}(\underline{v})) = \{\underline{\varepsilon} : \underline{\varepsilon}^{\underline{v}} = 1\}$.

For $I \subseteqq \{1, \dots, n\}$, usually $I = \mathfrak{O}(\underline{w})$, we define I-congruence on $\mathbb{Z}^n$ by: $\underline{v} \equiv_I \underline{v}'$ if $\underline{\varepsilon}^{\underline{v}} = \underline{\varepsilon}^{\underline{v}'}$ for all $\underline{\varepsilon} \in G_n(I)$. [Equivalently, $\underline{v}$ and $\underline{v}'$ are I-congruent when $\underline{\varepsilon}^{\underline{w}} = 1$ implies $\underline{\varepsilon}^{\underline{v}} = \underline{\varepsilon}^{\underline{v}'}$, or when $\mathfrak{O}(\underline{v} - \underline{v}')$ is either ϕ or I.] We write $\equiv$ for $\equiv_\phi$; if $\underline{v} \equiv \underline{v}'$, then $\underline{v} \equiv_I \underline{v}'$ for all I. We shall be interested in the decomposition of $C(\mathfrak{U}) = Z_1 \cup \dots \cup Z_r$ into I-equivalence classes. If $\mathfrak{U}$ is a trellis and $I = \mathfrak{O}(\underline{w})$, then $\underline{w} \equiv_I \underline{0}$; hence $\mathfrak{U} \cup \{\underline{w}\}$ always lies within one I-class. We shall always index the classes so that $(\mathfrak{U} \cup \{\underline{w}\}) \subseteqq Z_1$.

(7.7) Lemma. *If $f = f(\mathfrak{U}, \underline{w})$ is a simplicial agiform, then $G_n(\mathfrak{O}(\underline{w})) \subset \mathfrak{Z}(f)$.*

Proof. By (1.8), $f(\mathfrak{U}, \underline{w})(\underline{z}) = 0$ if $\underline{z}^{\underline{u}_1} = \dots = \underline{z}^{\underline{u}_m} = \underline{z}^{\underline{w}} = 1$. For any $\underline{\varepsilon} \in G_n$, $\underline{\varepsilon}^{\underline{u}_i} = 1$; as noted earlier, $G_n(\mathfrak{O}(\underline{w})) = \{\underline{\varepsilon} \in G_n : \underline{\varepsilon}^{\underline{w}} = 1\}$. □

Simplicial agiforms have other zeros, but these are not useful to us. For example, suppose $\underline{z}^{\underline{u}_1} = \dots = \underline{z}^{\underline{u}_m} = 0$ [as when $\mathfrak{U} = \mathfrak{M}$ and $\underline{z} = (1, 0, 0)$.] Then

$\underline{z}^{\underline{v}} = \prod(\underline{z}^{\underline{u}_i})^{\sigma_i} = 0$ for $\underline{v} = \sum \sigma_i \underline{u}_i \in C(\mathfrak{U})$ with $\sigma_i \geqq 0$, since at least one σ_i is positive. Thus, if $N(q) \subseteq C(\mathfrak{U})$, then $\underline{z} \in \mathfrak{Z}(q)$.

(7.8) **Theorem.** *The form* $h(\underline{x}) = \sum_{j=1}^{s} c(\underline{v}_j)\underline{x}^{\underline{v}_j}$ *is second-order at* $G_n(I)$ *if and only if the vector equation*

(7.9)
$$\sum_{\underline{v}_j \in Z_k} c(\underline{v}_j)\underline{v}_j = \underline{0}$$

holds for each I-class Z_k *of the set* $\{\underline{v}_1, \ldots, \underline{v}_s\}$.

We require two technical lemmas (c.f. [2, pp. 295–296].)

(7.10) **Lemma.** *Suppose* $\underline{w}$ *is a lattice point and* $I \subseteq \{1, \ldots, n\}$. *Then*

(7.11)
$$\sum_{\underline{\varepsilon} \in G_n(I)} \underline{\varepsilon}^{\underline{w}} = \begin{cases} |G_n(I)|, & \text{if } \quad \mathfrak{O}(\underline{w}) = \phi \text{ or } I \\ 0, & \text{otherwise}. \end{cases}$$

Proof. Consider the subgroups $G_n(I)$ and $G_n(\mathfrak{O}(\underline{w})) = \{\underline{w} : \underline{\varepsilon}^{\underline{w}} = 1\}$. If $\mathfrak{O}(\underline{w}) = \phi$ or I, then $\underline{\varepsilon}^{\underline{w}}$ is always 1 on the left-hand side of (7.11), and the sum is $|G_n(I)|$. Otherwise, $G_n(I) \cap G_n(\mathfrak{O}(\underline{w}))$ is a proper subgroup of $G_n(I)$ which must have index two. Thus exactly half the summands in (7.11) are 1 and the other half are -1. [A slicker proof is this: $\chi_{\underline{w}}(\underline{\varepsilon}) = \underline{\varepsilon}^{\underline{w}}$ is a group character on $G = G_n(I)$, and either $\chi = 1_G$ or $\sum_{g \in G} \chi(g) = 0$.] □

(7.12) **Lemma.** *Suppose A is a set of lattice points and* $I \subseteq \{1, \ldots, n\}$. *Then the following two systems of linear equations are equivalent:*

(7.13)
$$\sum_{\underline{v} \in A} b(\underline{v})\underline{\varepsilon}^{\underline{v}} = 0 \quad \textit{for all } \underline{\varepsilon} \in G_n(I),$$

(7.14)
$$\sum_{\underline{v} \in Z} b(\underline{v}) = 0 \quad \textit{for every I-congruence class } Z \textit{ of } A.$$

Proof. Let $A = Z_1 \cup \ldots \cup Z_r$ be a decomposition of A into I-congruence classes, and choose $\underline{v}_k \in Z_k$. Then, multiplying (7.13) by $\underline{\varepsilon}^{\underline{v}_k}$ $(= \underline{\varepsilon}^{-\underline{v}_k})$ and summing over $\underline{\varepsilon} \in G_n(I)$ gives:

(7.15)
$$0 = \sum_{\underline{\varepsilon} \in G_n(I)} \left(\sum_{\underline{v} \in A} b(\underline{v})\underline{\varepsilon}^{\underline{v} - \underline{v}_k} \right) = \sum_{\underline{v} \in A} b(\underline{v}) \left(\sum_{\underline{\varepsilon} \in G_n(I)} \underline{\varepsilon}^{\underline{v} - \underline{v}_k} \right).$$

By Lemma 7.10, the inner sum is 0 in (7.15), unless $\mathfrak{O}(\underline{v} - \underline{v}_k) = \phi$ or I, in which case $\underline{v} \equiv_I \underline{v}_k$, and the sum is $|G_n(I)|$. Thus, (7.15) reduces to

$$0 = |G_n(I)| \cdot \sum_{\underline{v} \in Z_k} b(\underline{v}),$$

which implies (7.14). Conversely, suppose (7.14) holds for each Z_k. Fix $\underline{\varepsilon} \in G_n(I)$ and sum over the Z_i's, noting that $\underline{v} \in Z_k$ implies $\underline{\varepsilon}^{\underline{v}} = \underline{\varepsilon}^{\underline{v}_k}$. Then

$$\sum_{\underline{v} \in A} b(\underline{v})\underline{\varepsilon}^{\underline{v}} = \sum_{k=1}^{r} \left(\sum_{\underline{v} \in Z_k} b(\underline{v})\underline{\varepsilon}^{\underline{v}} \right) = \sum_{k=1}^{r} \underline{\varepsilon}^{\underline{v}_k} \left(\sum_{\underline{v} \in Z_k} b(\underline{v}) \right) = 0;$$

that is, (7.14) implies (7.13). □

Proof of Theorem 7.8. Lemmas 7.6 and 7.12 combine to show that h is second-order at $G_n(I)$ if and only if, for all k,

$$\sum_{\underline{v}_j \in Z_k} c(\underline{v}_j) v_{ji} = 0, \quad \text{for } i = 1, \ldots, n. \tag{7.16}$$

But (7.16) is just the i-th component of the vector equation (7.9). □

Since $\underline{1} \in G_n(I)$, it follows that $\sum_{\underline{v}_j \in Z_k} c(\underline{v}_j) = 0$ for all classes Z_k.

Theorem 7.8 suggests the following definition. Suppose $\mathfrak{U}$ is a trellis, $\underline{w} \in C(\mathfrak{U})$ is an interior point, $I = \mathfrak{O}(\underline{w})$ and $C(\mathfrak{U})$ is partitioned into I-congruence classes $\{Z_k\}$, where $Z_1 \supseteq \mathfrak{U} \cup \{\underline{w}\}$. Then $\mathfrak{U}$ is *$\underline{w}$-thin* if

$$Z_1 = \mathfrak{U} \cup \{\underline{w}\}, \tag{7.17}$$

$$Z_k \text{ is a linearly independent set if } k \geqq 2. \tag{7.18}$$

If $\underline{w}$ is even, then $Z_1 = E(\mathfrak{U})$, so (7.17) is equivalent to primitivity. One can show using Theorem 2.5 that, if $\mathfrak{U}$ is $\underline{w}$-thin, then either $\mathfrak{U}$ is an M-trellis or $|\mathfrak{U}| = 2$ and $|E(\mathfrak{U})| = 3$. The same trellis may be $\underline{w}$-thin and not $\underline{w}'$-thin. Since $|E(\mathfrak{U}_p)| = p - 2$, the M-trellis $\mathfrak{U}_p$ cannot be $\underline{w}$-thin if $p \geqq 4$. Since $C(\mathfrak{U})$ is contained in $\{\sum x_i = m\}$, (7.18) is satisfied if $|Z_k| \leqq 2$.

(7.19) **Theorem.** *Suppose $f = f(\mathfrak{U}, \underline{w})$ is a simplicial agiform on a trellis $\mathfrak{U}$, where $\underline{w}$ is interior to $\mathfrak{U}$. If $\mathfrak{U}$ is $\underline{w}$-thin, then f is extremal.*

Proof. Write $f(\underline{x}) = \sum_{i=1}^{s} a(\underline{v}_i) \underline{x}^{\underline{v}_i}$, where $C(\mathfrak{U}) = \{\underline{v}_1, \ldots, \underline{v}_s\}$; $a(\underline{u}_i) = \lambda_i$ (>0 as $\underline{w}$ is interior to $\mathfrak{U}$), $a(\underline{w}) = -1$, and $a(\underline{v}) = 0$ otherwise. Suppose $f \geqq h \geqq 0$. As $N(f) \cap \mathbb{Z}^n = C(\mathfrak{U})$, Lemma 7.8 implies that $h(\underline{x}) = \sum_{i=1}^{s} c(\underline{v}_i) \underline{x}^{\underline{v}_i}$ is second-order at $\mathfrak{z}(f)$, which contains $G_n(I)$ by Lemma 7.7. By Theorem 7.8, the $c(\underline{v}_i)$'s satisfy (7.9), where the Z_k's are the I-congruence classes of $C(\mathfrak{U})$. Since $\mathfrak{U}$ is $\underline{w}$-thin, (7.18) implies that $c(\underline{v}) = 0$ if $\underline{v} \in Z_k$, $k \geqq 2$ and for Z_1, (7.9) and (7.17) imply:

$$\underline{0} = \sum_{i=1}^{m} c(\underline{u}_i) \underline{u}_i + c(\underline{w}) \underline{w}. \tag{7.20}$$

As $\mathfrak{U}$ is a trellis, the $\underline{u}_i$'s are linearly independent, and since $\underline{w} = \sum \lambda_i \underline{u}_i$, the complete solution to (7.20) is: $c(\underline{u}_i) = \alpha \lambda_i$, $c(\underline{w}) = -\alpha$ for some α. Thus $c(\underline{v}_i) = \alpha a(\underline{v}_i)$ for all $\underline{v}_i$; that is, $h = \alpha f$. Therefore, f is extremal. □

Parts of this argument were used by Choi and Lam to establish the extremality of M, S, and Q (see, for example, the remark after the proof of Theorem 3.8 in [4, p. 11].) These methods are not applicable to all extremal forms. For example, $p(x, y) = (x - y)^4$ is an extremal form in $P_{2,4}$, but if $q(x, y) = (x - y)^2 b(x, y)$, where b is any psd quadratic form, then $N(q) \subseteq N(p)$ and q is second-order at $\mathfrak{z}(p) = \{(r, r)\}$. Informally, our method is effective when p has only second-order zeros, except at the unit vectors, where $N(p)$ provides more information. [For example, the zeros of M are second-order at G_3, but sixth-order at $(0, 1, 0)$ in the direction of $(0, 0, 1)$, etc.]

We now verify that M, S, and Q are extremal. If $\mathfrak{U} = \mathfrak{M}$ and $\underline{w} = (2, 2, 2)$, then $\mathfrak{O}(\underline{w}) = \phi$ and the I-classes are congruence classes mod 2. Referring back to Fig. 1, we see that the I-classes for $C(\mathfrak{M})$ are:

$$Z_1=\mathfrak{M}\cup\{\underline{w}\}=\{(4,2,0),(2,4,0),(0,0,6),(2,2,2)\},$$
$$Z_2=\{(3,3,0),(1,1,4)\},\quad Z_3=\{(2,1,3),(2,3,1)\}\quad\text{and}$$
$$Z_4=\{(1,2,3),(3,2,1)\}.$$

These satisfy (7.17) and (7.18) so $\mathfrak{M}$ is $\underline{w}$-thin and M is extremal. If $\underline{w}'=(1,1,4)$, then $\mathfrak{O}(\underline{w})=\{1,2\}$ and $\underline{v}\equiv_{I'}\underline{v}'$ if v_1-v_1' and v_2-v_2' are either both even or both odd. Thus, the I'-classes for $\mathfrak{M}$ are $Z_1'=Z_1\cup Z_2$ and $Z_2'=Z_3\cup Z_4$, which fail (7.17) and (7.18), so $\mathfrak{M}$ is not $\underline{w}'$-thin. Note that

$$6f(\mathfrak{U},\underline{w}')=x^4y^2+x^2y^4+4z^6-6xyz^4=M(x,y,z)+3(z^3-xyz)^2$$

is not extremal. This is no accident.

If $\mathfrak{S}=\mathfrak{S}_3=\bar{\mathfrak{S}}_3$ and $\underline{w}=(2,2,2)$, then the I-classes for $C(\mathfrak{S})$ are:

$$Z_1=\mathfrak{S}\cup\{\underline{w}\}=\{(4,2,0),(0,4,2),(2,0,4),(2,2,2)\},$$
$$Z_2=\{(1,3,2),(3,1,2)\},\quad Z_3=\{(2,1,3),(2,3,1)\},\quad\text{and}$$
$$Z_4=\{(1,2,3),(3,2,1)\},$$

so $\mathfrak{S}$ is $\underline{w}$-thin and $S=3f(\mathfrak{S},\underline{w})$ is extremal.

If $\mathfrak{Q}=\mathfrak{Q}_4$ and $\underline{w}=(1,1,1,1)$, then $\mathfrak{O}(\underline{w})=\{1,2,3,4\}$ and $\underline{v}\equiv_I\underline{v}'$ if v_i-v_i' is either always even or always odd. Then the I-classes of $C(\mathfrak{Q})$ are:

$$Z_1=\mathfrak{Q}\cup\{\underline{w}\}=\{(4,0,0,0),(0,2,2,0),(0,2,0,2),(0,0,2,2),(1,1,1,1)\},$$
$$Z_2=\{(2,1,1,0),(0,1,1,2)\},\quad Z_3=\{(2,1,0,1),(0,1,2,1)\}\quad\text{and}$$
$$Z_4=\{(2,0,1,1),(0,2,1,1)\}.$$

Again, (7.17) and (7.18) are satisfied, so $Q=4f(\mathfrak{Q},\underline{w})$ is extremal.

8. Extremal Agiforms: The Necessary Condition

In this section we prove the converse to Theorem 7.19. We begin by showing that, if $\mathfrak{U}$ is not $\underline{w}$-thin, then there exists a form $h\neq\alpha f$ with $N(h)\subseteq N(f)$ which is second-order at $G_n(I)$. We study $f_\alpha=f+\alpha h$ on the orthants of $\mathbb{R}^n$ (so the signs of the variables are absorbed into the coefficients) by reversing the original substitution into the AGI. Although fractional exponents occur, the variables are non-negative. Finally, we show that f_α is psd for $|\alpha|$ small, so $f=\frac{1}{2}(f_\alpha+f_{-\alpha})$ is not extremal.

(8.1) **Lemma.** *If $\mathfrak{U}$ is not $\underline{w}$-thin, then there exists a non-zero $h\neq\alpha f$, involving monomials from a single I-class of $C(\mathfrak{U})$, which is second-order at $G_n(I)$.*

Proof. If (7.18) is violated, there is a non-zero solution to (7.9) for some $k\geqq 2$; if (7.17) is violated, there is a solution to (7.9) for $k=1$ which is not a multiple of $\{a(\underline{v}_i)\}$. In either case, by Theorem 7.8, the resulting form $h(\underline{x})=\sum c(\underline{v})\underline{x}^{\underline{v}}$ is second-order at $G_n(I)$ and $h\neq\alpha f$. The monomials in h come from the chosen Z_k. □

(8.2) **Theorem.** *If $\underline{w}$ is interior to $\mathfrak{U}$ and $\mathfrak{U}$ is not $\underline{w}$-thin, then $f(\mathfrak{U},\underline{w})$ is not extremal.*

Proof. Construct h by Lemma 8.1 and let $f_\alpha(\underline{x})=f(\underline{x})+\alpha h(\underline{x})$; f_α is not a multiple of f if $\alpha \neq 0$. For $\underline{z}\in\mathbb{R}^n$, $f_\alpha(\underline{z})\geqq 0$ if $f(\underline{z})=h(\underline{z})=0$ or if $f(\underline{z})>0$ and $|h(\underline{z})/f(\underline{z})|\leqq|\alpha|^{-1}$. Write $G_n=\{\underline{\varepsilon}_1,\ldots,\underline{\varepsilon}_{2^n}\}$, and let

$$R_k=\{\underline{x}=(\varepsilon_{k1}z_1,\ldots,\varepsilon_{kn}z_n):z_i\geqq 0\},$$

so $\mathbb{R}^n=\bigcup R_k$. We shall prove there exists $\delta_k>0$ so that $f_\alpha(\underline{x})\geqq 0$ for $\underline{x}\in R_k$ if $|\alpha|<\delta_k$. For $\gamma=\min\delta_k$, $f\geqq\frac{1}{2}f_\gamma\geqq 0$, so f is not extremal.

Fix k and write $\underline{\varepsilon}_k=\underline{\varepsilon}$, so $\underline{x}^{\underline{v}}=\underline{\varepsilon}^{\underline{v}}\underline{z}^{\underline{v}}$ with $z_i\geqq 0$. Then,

$$f(\underline{x})=\lambda_1\underline{z}^{\underline{u}_1}+\ldots+\lambda_n\underline{z}^{\underline{u}_n}-\underline{\varepsilon}^{\underline{w}}\underline{z}^{\underline{w}}, \tag{8.3}$$

$$h(\underline{x})=\sum c(\underline{v}_j)\underline{\varepsilon}^{\underline{v}_j}\underline{z}^{\underline{v}_j}, \tag{8.4}$$

where, as in subsequent equations, the change of variables is implicit in the formulas. If $\underline{\varepsilon}\in G_n(I)$, then $\underline{\varepsilon}^{\underline{w}}=1$ in (8.3) and the $\underline{\varepsilon}^{\underline{v}_j}$'s are all equal in (8.4) (since the $\underline{v}_j$'s all lie in the same I-class); thus $h(\underline{x})=\pm\sum c(\underline{v}_j)\underline{z}^{\underline{v}_j}$. If $\underline{\varepsilon}\notin G_n(I)$, then $\underline{\varepsilon}^{\underline{w}}=-1$, and we can say nothing about the $\underline{\varepsilon}^{\underline{v}_j}$'s.

We reverse the substitution which turned (1.1) into f. Let $\underline{z}^{\underline{u}_i}=t_i\geqq 0$. If $\underline{v}=\sum\sigma_i\underline{u}_i$, then $\underline{z}^{\underline{v}}=\prod t_i^{\sigma_i}$; write $\underline{v}_j=\sum\sigma_{ji}\underline{u}_i$, where $\sigma_{ji}\geqq 0$ and $\sum\sigma_{ji}=1$. Then (8.3) and (8.4) become:

$$f(\underline{x})=\sum\lambda_i t_i-\underline{\varepsilon}^{\underline{w}}\left(\prod t_i^{\lambda_i}\right), \tag{8.5}$$

$$h(\underline{x})=\sum(\underline{\varepsilon}^{\underline{v}_j})c(\underline{v}_j)\left(\prod t_i^{\sigma_{ji}}\right). \tag{8.6}$$

These equations use the fact that $N(h)\subseteq C(\mathfrak{U})$ in a critical but implicit way: since $\sigma_{ji}\geqq 0$, the function $\prod t_i^{\sigma_{ji}}$ is continuous for $\underline{t}\in\mathbb{R}^m_+$. If $\underline{v}=\sum\sigma_i\underline{v}_i$ where $\sigma_k<0$ for some k, then $\prod t_i^{\sigma_i}$ is unbounded near $t_k=0$.

Using (8.5) and (8.6), let $F(\alpha,\underline{t})=f_\alpha(\underline{x})$; $f_\alpha(\underline{x})\geqq 0$ for $\underline{x}\in R_k$ if $F(\alpha,\underline{t})\geqq 0$ for $\underline{t}\in\mathbb{R}^m_+$; $F(\alpha,\underline{t})$ is homogeneous in $\underline{t}$ of degree 1, but is not a form because of fractional exponents. By its homogeneity, $f_\alpha(\underline{x})\geqq 0$ for $\underline{x}\in R_k$ if $F(\alpha,\underline{t})\geqq 0$ on the compact set

$$K=\{\underline{t}=(t_1,\ldots,t_m):t_i\geqq 0 \text{ and } \textstyle\sum t_i=m\}.$$

If $\underline{\varepsilon}_k=\underline{\varepsilon}\notin G_n(I)$, then $\underline{\varepsilon}^{\underline{w}}=-1$ and, since $\lambda_i>0$, $f(\underline{t})$ is strictly positive for $\underline{t}\in K$ from (8.5). Since f and h are continuous and $f>0$, $\varphi=h/f$ is a continuous function on K and so is bounded; $|\varphi(\underline{x})|\leqq M$ for $\underline{t}\in K$. Then, as previously noted, $f_\alpha(\underline{x})\geqq 0$ for $|\alpha|<1/M$ and $\underline{x}\in R_k$.

If $\underline{\varepsilon}_k=\underline{\varepsilon}\in G_n(I)$, then

$$F(\alpha,\underline{t})=\left\{\sum\lambda_i t_i-\prod t_i^{\lambda_i}\right\}\pm\alpha\left\{\sum c(\underline{v}_j)\left(\prod t_i^{\sigma_{ji}}\right)\right\}. \tag{8.7}$$

There is equality in the AGI on K only at $\underline{1}=(1,\ldots,1)$: $f(\underline{1})=0$ and $f(\underline{t})>0$ for $\underline{t}\in K$, $\underline{t}\neq\underline{1}$. (As $\underline{w}$ is interior to $\mathfrak{U}$, $\lambda_i>0$.) Since $\varphi=h/f$ is continuous on $K\setminus\{\underline{1}\}$, we need to show that φ is bounded near $\underline{1}$. Parameterize K by

$$\underline{t}=(1+s_1,\ldots,1+s_m)\in K, \quad \sum s_i=0, \tag{8.8}$$

and substitute into (8.7) to get the Taylor series for f and h at $\underline{1}$. First,

$$f(\underline{x})=\sum\lambda_i(1+s_i)-\prod\left(1+\lambda_i s_i+\tfrac{1}{2}\lambda_i(\lambda_i-1)s_i^2+\text{higher-order terms}\right).$$

The constant and first-order terms cancel, leaving:

(8.9) $$f(\underset{\sim}{x})=\frac{1}{2}\sum\lambda_i(1-\lambda_i)s_i^2-\sum_{i<j}\lambda_i\lambda_js_is_j+\text{higher-order terms}$$
$$=\sum_{i<j}\frac{1}{2}\lambda_i\lambda_j(s_i-s_j)^2+\text{higher-order terms}.$$

Since $\lambda_i>0$, the leading term is a psd quadratic form which vanishes only at $(c,\ldots,c)$. Since $s_m=-\sum_{i=1}^{m-1}s_i$, it is a strictly definite quadratic form in $s_1,\ldots,s_{m-1}$, and so is bounded below by $\beta\sum_{i=1}^{m-1}s_i^2$ for some $\beta>0$.

Since $h(\underset{\sim}{x})=\sum c(\underset{\sim}{v}_j)\underset{\sim}{x}^{\underset{\sim}{v}_j}$ is second-order at $\underset{\sim}{1}$, $\sum c(\underset{\sim}{v}_j)=0$; as $\underset{\sim}{v}_j=\sum\sigma_{ji}\underset{\sim}{u}_i$, (7.9) implies that

$$\underset{\sim}{0}=\sum_i\Big(\sum_j\sigma_{ji}c(\underset{\sim}{v}_j)\Big)\underset{\sim}{u}_i.$$

But the $\underset{\sim}{u}_i$'s are linearly independent, so $\sum_j\sigma_{ji}c(\underset{\sim}{v}_j)=0$ for each i. Thus,

$$h(\underset{\sim}{x})=\sum_j c(\underset{\sim}{v}_j)\Big(\prod_i(1+s_i)^{\sigma_{ij}}\Big)$$
$$=\sum_j c(\underset{\sim}{v}_j)+\sum_i\Big(\sum_j\sigma_{ji}c(\underset{\sim}{v}_j)\Big)s_i+\text{higher-order terms}.$$

It follows that leading term of h is second-order, say,

(8.10) $$h(\underset{\sim}{x})=\sum_{i,j}\gamma_{ij}s_is_j+\text{higher-order terms}$$

for some γ_{ij}. Taking $s_m=-\sum_{i=1}^{m-1}s_i$, we see that the second-order term for h is bounded in absolute value by $\kappa\sum_{i=1}^{m-1}s_i^2$. Thus, (8.9) and (8.10) together imply that $\varphi=h/f$ is bounded near $\underset{\sim}{t}=\underset{\sim}{1}$. Since φ is continuous on $K\backslash\underset{\sim}{1}$, it is bounded on K. As before, there exists $\delta_k>0$ so that $f_\alpha(\underset{\sim}{x})\geqq 0$ for all $\underset{\sim}{x}\in R_k$ and $|\alpha|\leqq\delta_k$. We conclude that f is not extremal. □

(8.11) **Corollary.** *The agiform f is extremal if and only if $f=f(\mathfrak{U},\underset{\sim}{w})$ is simplicial, $\underset{\sim}{w}$ is interior to $\mathfrak{U}$ and $\mathfrak{U}$ is $\underset{\sim}{w}$-thin.*

Finally, we derive simplicial agiforms from "first principles"; this argument was given in [17] for $n=3$. Let $p(\underset{\sim}{x})=\sum_{i=1}^{r}a(\underset{\sim}{u}_i)\underset{\sim}{x}^{\underset{\sim}{u}_i}$ $(a(\underset{\sim}{u}_i)\neq 0)$ be a form in n variables. We say that p has k *effective variables* if the $r\times n$ matrix $[u_{ij}]$ has rank k, or, equivalently, if $cvx(\{\underset{\sim}{0},\underset{\sim}{u}_1,\ldots,\underset{\sim}{u}_r\})$ lies in a k-dimensional subspace of $\mathbb{R}^n$. Since the $\underset{\sim}{u}_i$'s are contained in the hyperplane $\underset{\sim}{1}\cdot=2d$, it follows that $N(p)=cvx(\{\underset{\sim}{u}_i\})$ is, geometrically, $(k-1)$-dimensional.

(8.12) **Theorem.** *Suppose $p(\underset{\sim}{x})=\sum a(\underset{\sim}{u}_i)\underset{\sim}{x}^{\underset{\sim}{u}_i}$ is a* psd *form in n variables with k effective variables and r terms, and $r\leqq k+1$. Then after a dilation of variables, p is a sum of monomial squares and (possibly) a simplicial agiform.*

Proof. As $N(p)$ is $(k-1)$-dimensional, it has at least k extreme points. By Theorem 3.6, if $\underline{v}$ is extreme, then it is even and $a(\underline{v})>0$, so $a(\underline{v})\underline{x}^{\underline{v}}$ is a monomial square. If p is not a sum of monomial squares, then $r=k+1$ and one of the $\underline{u}_i$'s is not an extreme point. Since $N(p)$ is $(k-1)$-dimensional and has k extreme points, it is a simplex. After relabeling, we denote the extreme points by $\underline{u}_1,\ldots,\underline{u}_k$, and the non-extreme point by $\underline{w}=\sum\lambda_i\underline{u}_i$, with $\lambda_i\geqq 0$ and $\sum\lambda_i=1$. If $\underline{w}$ is even and $a(\underline{w})>0$, then p is still a sum of monomial squares. If $\underline{w}$ is not even (say w_j is odd) and $a(\underline{w})>0$, take $x_j\to -x_j$ so that $\underline{x}^{u_i}\to\underline{x}^{u_i}$ and $\underline{x}^{w}\to -\underline{x}^{w}$. We may thus assume that $p(\underline{x})=\sum a(\underline{u}_i)\underline{x}^{u_i}-a(\underline{w})\underline{x}^{w}$, where $a(\underline{u}_i)\geqq 0$ and $a(\underline{w})>0$. Suppose $\lambda_i>0$ for $i=1,\ldots,l$ and $\lambda_i=0$ for $i\geqq l+1$, and write $p(\underline{x})=p^*(\underline{x})+\sum_{i=l+1}^{r} a(\underline{u}_i)\underline{x}^{u_i}$. In the notation of Theorem 3.6 (iv), $p^*=p^{(F)}$ is psd, where F is the facial simplex with vertices $\{\underline{u}_1,\ldots,\underline{u}_l\}$. Since a sum of monomial squares is still a sum of monomial squares after a dilation, it suffices to show that p^* decomposes in the desired way. For notational convenience, we assume that $l=k$.

By hypothesis, the $k\times n$ matrix $[u_{ij}]$ has full rank, thus, the inhomogeneous linear system of equations

$$\Sigma_j u_{ij}t_j=\log\lambda_i-\log a(\underline{u}_i),\qquad i=1,\ldots,l$$

has at least one solution: $\underline{t}=(s_1,\ldots,s_n)$ and $e^{\Sigma u_{ij}s_j}=\lambda_i(a(\underline{u}_i))^{-1}$. Under the scaling $x_j\to e^{s_j}x_j$, we have $\underline{x}^{u_i}\to(e^{\Sigma u_{ij}s_j})\underline{x}^{u_i}=\lambda_i(a(\underline{u}_i))^{-1}\underline{x}^{u_i}$, and $p^*(\underline{x})\to q(\underline{x})=\sum\lambda_i\underline{x}^{u_i}-c\underline{x}^{w}$, where $c=(e^{\Sigma w_j s_j})a(\underline{w})>0$. Since p^* is psd, so is q and $0\leqq q(1,\ldots,1)=1-c$. Thus $0<c\leqq 1$, and

$$q(\underline{x})=cf(\mathfrak{U},\underline{w})(\underline{x})+(1-c)\left(\sum\lambda_i\underline{x}^{u_i}\right)$$

as asserted. □

9. Examples of Extremal Forms

In Sect. 6, each of the forms M, S, and Q was generalized by two families of agiforms. Since $\{M_n\}$ $(n>3)$ and $\{Q_{2m}\}$ $(m>2)$ are not simplicial, they cannot be extremal. In this section we use Theorem 7.19 to show that the families $\{M_n\}$, $\{\bar{S}_n\}$, and $\{\bar{Q}_{2m}\}$ consist of extremal forms, but S_n is not extremal for $n\geqq 4$. (We have already seen that M, S, and Q are extremal.) We also return to a primitive, but not extremal, agiform, first mentioned in [17].

(9.1) **Theorem.** *M_n, $\bar{S}_n$, and $\bar{Q}_{2m}$ are extremal for $n\geqq 3$ and $2m\geqq 4$.*

Proof. For all three cases, let $\underline{u}_{i,j}=\frac{1}{2}(\underline{u}_i+\underline{u}_j)$, $i<j$ and $\underline{w}_i=\frac{1}{2}(\underline{u}_i+\underline{w})$; $\underline{w}_i\in C(\mathfrak{U})$ if and only if $\underline{w}$ is even. If $\underline{v}$ and $\underline{v}'$ are in $C(\mathfrak{U})$ and $v_i\equiv_I v_i'$ for $i\leqq n-1$, then $\sum v_i=\sum v_i'$ implies $\underline{v}\equiv\underline{v}'$.

Let $\mathfrak{U}=\mathfrak{M}_n$. Since $\underline{w}=(2,\ldots,2)$ and $E(\mathfrak{M}_n)=\mathfrak{M}_n\cup\{\underline{w}\}$ by Theorem 6.9, (7.17) is satisfied, and $\equiv_I$ is congruence mod 2. For $i<j\leqq n-1$, $\underline{u}_{i,j}=(a_1,\ldots,a_{n-1},0)$ with 3's in the i-th and j-th place and 2's elsewhere. For $i\leqq n-1$, $\underline{u}_{i,n}=(1,\ldots,2,\ldots,1,n)$ and $\underline{w}_i=(2,\ldots,3,\ldots,2,1)$ with the i-th place distinguished, and $\underline{w}_n=(1,\ldots,1,n+1)$. It is easy to see that no two of these are congruent, so each Z_k is a singleton for $k\geqq 2$, confirming (7.18).

A similar argument, whose details we omit, works for $\mathfrak{U}=\mathfrak{S}_n$.

For $\mathfrak{U}=\bar{\mathfrak{Q}}_{2m}$, $m\geqq 3$, $\underline{w}=(1,\dots,1)$ and $\underline{v}\equiv\underline{v}'$ provided v_i-v_i' is always even or always odd. Let $\underline{e}_i$ denote the i-th unit vector. Then, for $i<j\leqq 2m-1$, $\underline{u}_{ij}=\underline{e}_i+\dots+\underline{e}_{i+m-1}+\underline{e}_j+\dots+\underline{e}_{j+m-1}$, with the subscripts reduced $\bmod 2m-1$; $\underline{u}_{i,2m}=\underline{e}_i+\dots+\underline{e}_{i+m-1}+(0,\dots,0,m)$, and, by Theorem 6.18, $C(\bar{Q}_{2m})$ consists of these vectors and $\bar{Q}_{2m}\cup\{\underline{w}\}$. For suitable k, the first $2m-1$ components of the $\underline{u}_{ij}$'s have the following pattern mod 2, taken cyclically: $1^k0^{m-k}1^k0^{m-1-k}$. It follows that $Z_1=\bar{Q}_{2m}\cup\{\underline{w}\}$, so (7.17) is satisfied, and no two other elements of $C(\bar{Q}_{2m})$ are I-congruent, so (7.18) is satisfied and Q_{2m} is extremal. $\square$

It can be shown that two conditions, satisfied above, are sufficient, but not necessary, for $\mathfrak{U}$ to be $\underline{w}$-thin:

(9.2) (i) $$E(\mathfrak{U})=(\mathfrak{U}\cup\{\underline{w}\})\quad\text{and}\quad C(\mathfrak{U})=A(\mathfrak{U}\cup\{\underline{w}\}),$$

(9.2) (ii) $$E(\mathfrak{U})=\mathfrak{U}\quad\text{and}\quad C(\mathfrak{U})=A(\mathfrak{U})\cup\{\underline{w}\}.$$

(9.3) **Theorem.** *For $n\geqq 4$, S_n is primitive but not extremal.*

Proof. The primitivity of S_n follows from Theorems 6.12 and 7.3. We shall show that $(\alpha,\beta,\gamma,\delta,2,\dots,2)\in C(\mathfrak{S}_n)$ for each permutation $(\alpha,\beta,\gamma,\delta)$ of $(1,1,3,3)$. These six linearly dependent points are congruent mod 2, so they belong to the same Z_k, $\mathfrak{S}_n$ is not $\underline{w}$-thin and S_n is not extremal.

Suppose $\underline{v}=(v_1,\dots,v_n)\in C(\mathfrak{S}_n)$ with $v_i\geqq 0$ and $\sum v_i=2n$. By the proof of Theorem 6.12, the barycentric coordinates of $\underline{v}$ satisfy the equations $v_j=2\lambda_{j-1}+(2n-2)\lambda_j$ for $1\leqq j\leqq n$, with the indices viewed cyclically. It is easy to invert this system. Let $\varrho=-(n-1)^{-1}$; then

(9.4) $$v_j+\varrho v_{j-1}+\varrho^2 v_{j-2}+\dots+\varrho^{n-1}v_{j+1}=2\{(n-1)+\varrho^{n-1}\}\lambda_j.$$

Let $t_j=v_j+\varrho v_{j-1}$. If n is even, (9.4) implies that λ_j is a non-negative linear combination of $t_j, t_{j-2},\dots,t_{j+2}$. If n is odd, there is the "leftover" term $\varrho^{n-1}v_{j+1}$. In any event, if $t_j\geqq 0$ for all j (and $v_j\geqq 0$ if n is odd), then $\lambda_j\geqq 0$ and $\underline{v}\in C(\mathfrak{S}_n)$. Since $(n-1)t_j=(n-1)v_j-v_{j-1}$, any $\underline{v}$ with $\sum v_j=2n$ and $v_j\in\{1,2,3\}$ belongs to $C(\mathfrak{S}_n)$ when $n\geqq 4$. $\square$

(9.5) *Example.* We decompose S_4 into psd forms by carrying out in detail the proof of Theorem 8.2. The lattice points $(3,1,3,1)$, $(3,1,1,3)$, $(1,3,3,1)$, and $(1,3,1,3)$ are linearly dependent members of the same Z_k for $C(\mathfrak{S}_4)$, and $h(x,y,z,w)=xyzw(x^2-y^2)(z^2-w^2)$ is second-order at G_4. Let

(9.6) $$\begin{aligned}f_\alpha(x,y,z,w)&=x^6y^2+y^6z^2+z^6w^2+w^6x^2-4x^2y^2z^2w^2\\&\quad+\alpha xyzw(x^2-y^2)(z^2-w^2).\end{aligned}$$

Since $f_\alpha(x,y,z,-w)=f_{-\alpha}(x,y,z,w)$, f_α is psd if and only if $f_{-\alpha}$ is psd. We show that f_2 is psd. The next two identities may be routinely verified:

(9.7) $$f_2(x,y,z,w)=(x^3y+z^3w)^2+(y^3z+w^3x)^2-2xyzw(yz+xw)^2,$$

(9.8) $$\begin{aligned}10f_2(x,y,z,w)&=(7x^6y^2+y^6z^2+3z^6w^2+9w^6x^2-20x^3yzw^3)\\&\quad+(3x^6y^2+9y^6z^2+7z^6w^2+w^6x^2-20xy^3z^3w)\\&\quad+20xyzw(xz-yw)^2.\end{aligned}$$

If $xyzw \leqq 0$, then (9.7) shows that $f_2(x,y,z,w) \geqq 0$. If $xyzw \geqq 0$, then (9.8) gives $10f_2$ as a sum of two simplicial agiforms and a non-negative term, so $f_2(x,y,z,w) \geqq 0$. Thus f_2 is psd and $S_4 = \frac{1}{2}(f_2 + f_{-2})$ is not extremal. We have been unable to prove that f_2 is extremal, and suspect that it is not.

(9.9) *Example.* A version of this example was announced in [17, p. 373]. Let

$$D(x,y,z,w) = 2x^4y^2 + 2x^2y^4 + z^4w^2 + z^2w^4 - 6x^2y^2zw, \tag{9.10}$$

$D = 6f(\mathfrak{D}, \underline{w})$, where $\underline{w} = (2,2,1,1)$ and

$$\mathfrak{D} = \{(4,2,0,0), (2,4,0,0), (0,0,4,2), (0,0,2,4)\}.$$

It is easy to check that $C(\mathfrak{D}) = A(\mathfrak{D}) \cup \{\underline{w}, \underline{w}'\}$, where $\underline{w}' = (1,1,2,2)$, so the I-classes of $C(\mathfrak{D})$ with respect to $\underline{w}$ are:

$$\begin{aligned} Z_1 &= \{(4,2,0,0), (2,4,0,0), (0,0,4,2), (0,0,2,4), (0,0,3,3), (2,2,1,1)\}, \\ Z_2 &= \{(1,1,2,2), (3,3,0,0)\}, \qquad Z_3 = \{(1,2,1,2), (1,2,2,1)\}, \\ Z_4 &= \{(2,1,1,2), (2,1,2,1)\}. \end{aligned} \tag{9.11}$$

We see that Z_k is independent for $k \geqq 2$, but $\mathfrak{D}$ is not $\underline{w}$-thin, because Z_1 contains $(0,0,3,3)$. Let

$$E(x,y,z,w) = 4x^4y^2 + 4x^2y^4 + z^4w^2 + 2z^3w^3 + z^2w^4 - 12x^2y^2zw; \tag{9.12}$$

$2D(\underline{x}) = E(\underline{x}) + (z^2w - zw^2)^2$; E is not an agiform because $(0,0,3,3)$ is not even. We shall show that E is psd; since $D \geqq \frac{1}{2}E$, this verifies that D is not extremal. Observe that (9.12) arises from the substitution $t_1 = x^4y^2$, $t_2 = x^2y^4$, $t_3 = z^4w^2$, $t_4 = z^3w^3$, $t_5 = z^2w^4$, and $\underline{\lambda} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{12}, \frac{1}{6}, \frac{1}{12})$ into (1.1). If $zw \geqq 0$, then $t_3 \geqq 0$, so $E \geqq 0$. If $zw \leqq 0$, then

$$E(x,y,z,w) = 4x^4y^2 + 4x^2y^4 + (z^2w + zw^2)^2 + (-12zw)x^2y^2$$

shows that $E \geqq 0$. Thus E is psd, as asserted.

Finally, we show that E is extremal. Note that $N(E) = C(\mathfrak{D})$ and $\mathfrak{z}(E)$ contains $G_4(\{3,4\})$ and $(0,0,1,-1)$. If $E \geqq q \geqq 0$, then q is second-order at $\mathfrak{z}(E)$ and $N(q) \subseteq C(\mathfrak{D})$. Write $q(\underline{x}) = \sum c(\underline{v}_i)\underline{x}^{\underline{v}_i}$ and apply Theorem 7.8 to $G_4(\{3,4\})$. Since Z_2, Z_3, and Z_4 are linearly independent, $c(\underline{v}_i) = 0$ unless $\underline{v}_i \in Z_1$. After relabeling, we have

$$q(x,y,z,w) = c_1x^4y^2 + c_2x^2y^4 + c_3z^4w^2 + c_4z^3w^3 + c_5z^2w^4 + c_6x^2y^2zw,$$

$$\begin{aligned} 4c_1 + 2c_2 + 2c_6 &= 2c_1 + 4c_2 + 2c_6 = 4c_3 + 3c_4 + 2c_5 + c_6 \\ &= 2c_3 + 3c_4 + 4c_5 + c_6 = 0. \end{aligned} \tag{9.13}$$

One parametric solution of the system in (9.13) is: $c_1 = c_2 = 4\alpha$, $c_3 = c_5 = \alpha + \beta$, $c_4 = 2\alpha - 2\beta$, $c_6 = -12\alpha$. Hence,

$$q(x,y,z,w) = \alpha E(x,y,z,w) + \beta(z^2w - zw^2)^2. \tag{9.14}$$

Taking $\alpha = \beta = \frac{1}{2}$ in (9.14) returns D; this is not surprising, as we have only used $\mathfrak{z}(D)$ so far. Finally, $q(0,0,1,-1) = 0$ implies $\beta = 0$, so $q = \alpha E$, and we conclude that E is extremal.

This example suggests that some classification may be possible for "almost agiforms" such as E.

10. Further Questions

We conclude with some open questions about agiforms. Of course, these are dwarfed in importance by the questions which motivated this paper. Given a psd form p, how can one tell whether it is sos? Given a psd form p, how can one tell whether it is extremal?

We begin with mediated sets. Given a framework $\mathfrak{U}$, is there an algorithm for computing $\mathfrak{U}^*$ which is more efficient than Theorem 2.2? Does the set of $\mathfrak{U}$-mediated sets for fixed $\mathfrak{U}$ have interesting properties? Given $\underline{w} \in \mathfrak{U}^*$, is there an easy way to compute a "small" $\mathfrak{U}$-mediated set containing $\underline{w}$? [One algorithm is to list all averages of distinct points in $E(\mathfrak{U})$, then build a $\mathfrak{U}$-mediated set containing $\underline{w}$ by starting with the average $\underline{w} = \frac{1}{2}(\underline{s} + \underline{t})$, then finding averages for $\underline{s}$ and $\underline{t}$ (unless they are in $\mathfrak{U}$) etc.; the branching looks horrendous.] What is the worst-case bound on size? If $\mathfrak{U} \subset \mathbb{R}^n$ lies in the hyperplane $\underline{1} \cdot = 2d$ and $\underline{w} \in \mathfrak{U}^*$, then by Theorem 2.8, $\underline{w}$ is contained a $\mathfrak{U}$-mediated set with at most $E(\mathfrak{H}_{n,2d}) + 1 = \binom{d+n-1}{d} + 1$ elements. One expects to do better. Handelman's publications contain many results and open questions related to the material in Proposition 2.7, as will [20].

After the generality of Theorem 3.3, it is painful that Theorem 4.4 is limited to simplicial agiforms; but the examples in Sect. 5 cast doubt on a general "yes-or-no" lattice point criterion for deciding whether $f(\mathfrak{U}, \underline{\lambda}, \underline{w})$ is sos in the non-simplicial case. For fixed $\underline{w}$, can anything be said about $\{\underline{\lambda} \in \Lambda(\underline{w}) : f(\mathfrak{U}, \underline{\lambda}, \underline{w}) \text{ is sos}\}$? What is the "probability" that a simplicial agiform is sos? (Is this question meaningful?) It is proved in [20] that every trellis for $m = 3$ is either an H-trellis, or the image of some $\mathfrak{U}_p$ with $p \geqq 3$, so "almost every" ternary simplicial agiform is sos.

Let R be a commutative ring, and suppose x is a sum of squares in R; x has *length* k if there exist $y_i \in R$ so that $x = y_1^2 + \ldots + y_k^2$, but x is not a sum of $k-1$ squares in R. As noted after Corollary 4.11, the Hurwitz agiform $G(\underline{c})$ has length $\leqq 3n-4$, when $R = \mathbb{R}[x_1, \ldots, x_n]$, see also [19]. The achievement of these small numbers relies on the fact that a psd binary form is a sum of two squares, which need not be binomials. By Theorem 4.4, the estimate on the size of a $\mathfrak{U}$-mediated set containing $\underline{w}$ gives an upper bound on the length. How good is this estimate in general? That is, how many fewer squares arise when we allow more terms? Can these "bigger" squares be incorporated into our algorithmic structure?

If p is a psd form, let $Y(p) = \{k \in \mathbb{Z} : p(x_1^k, \ldots, x_n^k) \text{ is sos}\}$. In [20] we study $Y(f)$ for simplicial agiforms; $Y(M_n) = \{k : k \geqq n/2\}$. If $\underline{w}$ is even and $f = f(\mathfrak{U}, \underline{w})$ is simplicial, then $Y(f) = \{k : k \geqq k_0\}$ for some k_0; for any agiform, $k \in Y(f)$ implies $k+2 \in Y(f)$. What can be said about $L(f, k)$, the length of $f(x_1^k, \ldots, x_n^k)$, beyond the obvious fact that $L(f, rk) \leqq L(f, k)$? The "Horn form" (see also [3, p. 396]) is an even quartic form $H(x_1, \ldots, x_5)$; we prove that $Y(H) = \phi$. Is $Y(p) = \phi$ possible for a psd form p in three or four variables?

In Sect. 6, circulant matrices make it easy to compute $C(\mathfrak{U})$ and $E(\mathfrak{U})$. Can we find other interesting trellises (and simplicial agiforms) using circulant matrices?

The condition of $\underline{w}$-thinness seems rather mysterious; does it have a more natural reformulation? Theorem 8.12 suggests the question of presenting the psd forms in k effective variables with $k+2$ terms and the determination of the resulting extremal forms. The form E, from Example 9.9, with $k=4$, might play the role of M for the "almost-agiforms".

There are several open algebraic questions about agiforms of a different nature. A celebrated theorem of Cassels, Ellison, and Pfister [1] states that $h(x, y) = M(x, y, 1)$ has length four when $R = \mathbb{R}(x, y)$. (By Artin's solution to Hilbert's Seventeenth Problem, every psd form is a sum of squares of *rational* functions.) What can be said about the length of the dehomogenizations of other simplicial agiforms in three variables? What about the extremal ones? These questions are probably very hard; [1] uses elliptic curves. It is known that the maximal length of an element in $\mathbb{R}(x_1, \ldots, x_n)$ is at most 2^n. Can this bound be achieved by the dehomogenization of another agiform?

Another question involves irreducibility. Binomial squares are reducible. It follows from the identity

$$a^3+b^3+c^3-3abc=(a+b+c)(a^2+b^2+c^2-ab-ac-bc)$$

[c.f. (1.11)], that $f(\mathfrak{U}, \underline{w})$ is reducible when $\mathfrak{U} = \{3\underline{u}_1, 3\underline{u}_2, 3\underline{u}_3\}$ and $\underline{w} = \underline{u}_1 + \underline{u}_2 + \underline{u}_3$. Are these the only reducible simplicial agiforms?

What can be said about monomial substitutions into agiforms? The identities $z^2M(x, y, z) = Q(xy, xz, yz, z^2)$ and $x^4z^2S(x, y, z) = M(x^2, yz, xz)$, and others, are used in [4] to simplify the proofs of extremality. Is there an efficient way to find formulas such as these?

References

1. Cassels, J.W.S., Ellison, W.J., Pfister, A.: On sums of squares and on elliptic curves over function fields. J. Number Theory **3**, 125–149 (1971), (MR 45 #1863)
2. Choi, M.D., Knebusch, M., Lam, T.Y., Reznick, B.: Transversal zeros and positive semidefinite forms. In: Colliot-Thélène, J.-L., Coste, M., Mahé, L., Roy, M.-F. (eds.) Géométrie algébrique réelle et formes quadratiques. Proceedings, Rennes 1981. (Lecture Notes Mathematics Vol. 959, pp. 273–298) Berlin Heidelberg New York: Springer 1982, (MR 84b.10027)
3. Choi, M.D., Lam, T.Y.: An old question of Hilbert. In: Orzech, G. (ed.) Proceedings Queens Univ. Quadratic Forms Conference 1976, Queens Pap. Pure Appl. Math. **46**, 385–405, (1977) (MR 58 #16503)
4. Choi, M.D., Lam, T.Y.: Extremal positive semidefinite forms. Math. Ann. **231**, 1–18 (1977), (MR 58 #16512)
5. Choi, M.D., Lam, T.Y., Reznick, B.: Real zeros of positive semidefinite forms, I. Math. Z. **171**, 1–26 (1980), (MR 81d.10012)
6. Choi, M.D., Lam, T.Y., Reznick, B.: Even symmetric sextics. Math. Z. **195**, 559–580 (1987)
7. Choi, M.D., Lam, T.Y., Reznick, B.: A combinatorial theory for sums of squares. In preparation
8. Davis, P.J.: Circulant matrices. New York: Wiley 1979 (MR 81a:15053)
9. Handelman, D.: Positive polynomials and product type actions of compact groups. Mem. Am. Math. Soc. **54** (1985), No. 320, (MR 86h:46091)
10. Handelman, D.: Integral body-building in $\mathbb{R}^3$. J. Geom. **27**, 140–152 (1986), (MR 87k:52032)
11. Handelman, D.: Positive polynomials, convex integral polytopes and a random walk problem. Lecture Notes in Mathematics Vol. 1282, Berlin Heidelberg New York: Springer 1987

12. Hardy, G.H., Littlewood, J.E., Pólya, G.: Inequalities. Cambridge: Cambridge University Press 1967
13. Hilbert, D.: Über die Darstellung definiter Formen als Summe von Formenquadraten, Math. Ann. **32**, 342–350 (1888) (=Ges. Abh., **2**, 154–161)
14. Hurwitz, A.: Über den Vergleich des arithmetischen und des geometrischen Mittels. J. Reine Angew. Math. **108**, 266–268 (1891) (=Werke, **2**, 505–507)
15. Lax, A., Lax, P.: On sums of squares. Linear Algebra Appl. **20**, 71–75 (1978), (MR 57 #3074)
16. Motzkin, T.S.: The arithmetic-geometric inequality. In: Shisha, O. (ed.) Inequalities, pp. 205–224, New York: Academic Press 1967, (=Selected Papers, pp. 203–222), (MR 36 #6569)
17. Reznick, B.: Extremal psd forms with few terms. Duke Math. J. **45**, 363–374 (1978), (MR 58 #511)
18. Reznick, B.: Lattice point simplices. Discr. Math. **60**, 219–242 (1986), (MR 87i.52022)
19. Reznick, B.: A quantitative version of Hurwitz' theorem on the arithmetic-geometric inequality. J. Reine Angew. Math. **377**, 108–112 (1987)
20. Reznick, B.: Midpoint polytopes and the map $x_i \to x_i^k$. In preparation
21. Robinson, R.M.: Some definite polynomials which are not sums of squares of real polynomials. In: Selected questions of algebra and logic. Izdat. "Nauka" Sibirsk. Otdel. Novosibirsk, pp. 264–282, 1973 [Abstract in Notices A.M.S. **16**, 554 (1969)], (MR 49 #2647)
22. Schmüdgen, K.: An example of a positive polynomial which is not a sum of squares of polynomials. A positive, but not strongly positive functional. Math. Nachr. **88**, 385–390 (1979), (MR 81b.12024)

Received October 1, 1987; in revised form May 30, 1988

Math. Ann. 283, 465–478 (1989)

Mathematische Annalen

On Murasugi's and Traczyk's Criteria for Periodic Links

Józef H. Przytycki

Department of Mathematics, Warsaw University, PL-00901 Warszawa, Poland and Department of Mathematics, University of Toronto, Toronto, M5S 1A1, Canada

1. Introduction

A link L in S^3 is called n-periodic if there is a Z_n action on S^3 with a circle as a fixed point set, which maps L onto itself, and such that L is disjoint from the fixed point set. Furthermore if L is an oriented link, we assume that each generator of Z_n preserves the orientation of L or changes it to the opposite one.

The skein polynomial (also called FLYPMOTH, generalized Jones, HOMFLY, Jones-Conway, twisted Alexander and two-variable Jones) of oriented links in S^3 can be defined uniquely by the conditions $P_{T_1}(a,z)=1$ and $aP_{L_+}(a,z)+a^{-1}P_{L_-}(a,z)=zP_{L_0}(a,z)$ where T_1 is the trivial knot and L_+, L_-, and L_0 are diagrams of oriented links which are identical except near one crossing where they look like in Fig. 1.1.

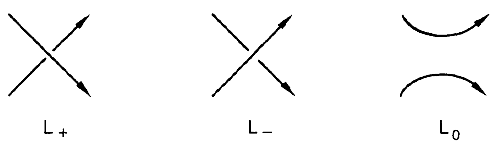


Fig. 1.1

Lemma 1.1. *Let $\mathscr{R}$ be the subring of the ring $Z[a^{\mp 1}, z^{\mp 1}]$ generated by $a^{\mp 1}$, $\frac{a+a^{-1}}{z}$ and z, then for any oriented link L, $P_L(a,z) \in \mathscr{R}$.*

Observe that z is not invertible in $\mathscr{R}$.

Proof. It is true for the trivial link of n components, T_n. Namely $P_{T_n}(a,z) = \left(\frac{a+a^{-1}}{z}\right)^{n-1} \in \mathscr{R}$. Furthermore if $P_{L_-}(a,z)$ [respectively $P_{L_+}(a,z)$] and $P_{L_0}(a,z)$ are elements of $\mathscr{R}$ then $P_{L_+}(a,z)$ [respectively $P_{L_-}(a,z)$] is an element of $\mathscr{R}$. Therefore Lemma 1.1 holds by the standard induction.

Now we can formulate our criterion for n-periodic links. It has especially simple form for a prime period (see Sect. 2 for the general statement).

Theorem 1.2. *Let L be an r-periodic oriented link and r a prime number, then the skein polynomial $P_L(a,z)$ satisfies:*

$$P_L(a,z) \equiv P_L(a^{-1},z) \bmod (r, z^r),$$

where (r, z^r) is the ideal in $\mathcal{R}$ generated by r and z^r.

The Jones polynomial of oriented links, $V_L(t)$ can be obtained from the skein polynomial $P_L(a,z)$ by substituting $a = it^{-1}$, $z = i\left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)$. Therefore

Corollary 1.3 [Tr1]. *Let L be an r-periodic oriented link and r a prime number. Then the Jones polynomial $V_L(t)$ satisfies:*

$$V_L(t) \equiv V_L(t^{-1}) \bmod (r, t^r - 1),$$

where $(r, t^r - 1)$ is the ideal in $Z[t^{\mp 1/2}]$ generated by r and $t^r - 1$.

The Kauffman polynomial of regular isotopy of non-oriented diagrams of links, $\Lambda_D(a,z)$, can be uniquely defined by the following properties:

(i) $\Lambda_O(a,z) = 1$

(ii) $\Lambda_{\rightharpoonup\!\circ}(a,z) = a\Lambda_{)}(a,z)$, $\Lambda_{\rightharpoondown\!\circ}(a,z) = a^{-1}\Lambda_{)}(a,z)$,

(iii) $\Lambda_{\times}(a,z) + \Lambda_{\times}(a,z) = z(\Lambda_{\asymp}(a,z) + \Lambda_{)(}(a,z))$

where the symbols $\times$, $\times$, $\asymp$, and)(stand for diagrams which look like that in a neighborhood of the crossing and are identical elsewhere.

The Kauffman polynomial of (ambient) isotopy of oriented links, $F_L(a,z)$ is defined by

$$F_L(a,z) = a^{-w(D)}\Lambda_D(a,z),$$

where D is any diagram of L and $w(D)$ is the planar writhe (or twist) of D defined by taking the algebraic sum of the crossings, counting $\times$ and $\times$ as $+1$ and -1 respectively.

For each L, $F_L(a,z) \in \mathcal{R}$, as in the case of the skein polynomial.

Theorem 1.4. *Let L be an r-periodic, oriented link and r a prime number. Then*

$$F_L(a,z) \equiv F_L(a^{-1},z) \bmod (r, z^r),$$

where (r, z^r) is the ideal in $\mathcal{R}$ generated by r and z^r.

To use practically Theorems 1.2 and 1.4 we need the following fact.

Lemma 1.5. *Let $w(a,z) \in \mathcal{R}$ and $w(a,z) = \sum_i v_i(a) z^i$ where $v_i(a) \in Z[a^{\mp 1}]$. Then $w(a,z) \in (r, z^r)$ if and only if $v_i(a)$ is an element of the ideal $(r, (a + a^{-1})^{r-i})$ in $Z[a^{\mp 1}]$ for each $i \leqq r$.*

Example 1.6. Consider the knot 11_{388} (in the Perko [Pe] notation); Fig. 1.2. The skein polynomial of this knot, $P_{11_{388}}(a,z) = (3 + 5a^2 + 4a^4 + a^6) + (-4 - 10a^2 - 5a^4)z^2 + (1 + 6a^2 + a^4)z^4 - a^2 z^6$ (see [LM] or [P1]). Consider the difference $w(a,z) = P_{11_{388}}(a,z) - P_{11_{388}}(a^{-1},z)$. $v_0(a)$ for $w(a,z)$ (in the notation of Lemma 1.5) is equal to $5(a^2 - a^{-2}) + 4(a^4 - a^{-4}) + a^6 - a^{-6}$. Therefore for a prime number $r \geqq 7$, $v_0(a) \notin (r, (a + a^{-1})^r) = (r, a^{2r} + 1)$. Therefore by Theorem 1.2 and Lemma 1.5 the knot 11_{388} is not r-periodic for $r \geqq 7$.

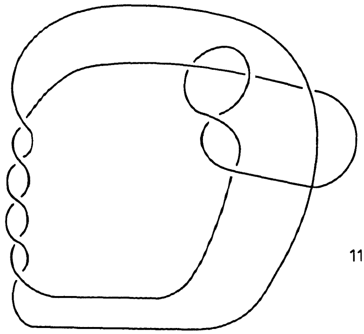


Fig. 1.2

Notice that the knot 11_{388} has the symmetric Jones polynomial $V_{11_{388}}(t) = V_{11_{388}}(t^{-1}) = t^{-2} - t^{-1} + 1 - t + t^2$ and therefore the Traczyk criterion (Corollary 1.3) cannot be applied.

Example 1.7. Consider the knot 10_{48} (in the Rolfsen notation [Ro]); Fig. 1.3. The skein polynomial of 10_{48} is symmetric ($P_{10_{48}}(a,z) = P_{10_{48}}(a^{-1},z)$; [Th, P1]) so Theorem 1.2 cannot be used to analyse periods of 10_{48}. We can use however the Kauffman polynomial to show that 10_{48} is not r-periodic for $r \geqq 7$. One can compute (see [Th] or [P1]) that $w(a,z) = F_{10_{48}}(a,z) - F_{10_{48}}(a^{-1},z) = z(a^5 + 3a^3 + 2a - 2a^{-1} - 3a^{-3} - a^{-5}) + z^2(\ldots)$. $v_1(a)$ for $w(a,z)$ (in the notation of Lemma 1.5) is equal to $a^5 + 3a^3 + 2a - 2a^{-1} - 3a^{-3} - a^{-5}$ so for $r \geqq 7$, $v_1(a) \notin (r, (a + a^{-1})^{r-1})$. Therefore by Theorem 1.4 and Lemma 1.5 the knot 10_{48} is not r-periodic for $r \geqq 7$.

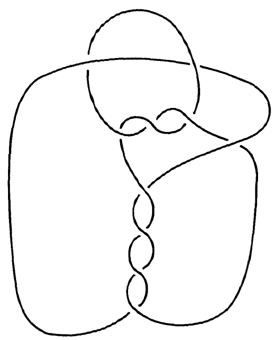


Fig. 1.3

2. Criterion for n-Periodic Links Using the Skein Polynomial

Theorem 2.1. *Let L be an n-periodic oriented link, then the skein polynomial $P_L(a,z)$ satisfies:*

$$P_L(a,z) \equiv P_L(a^{-1},z) \bmod J_n,$$

where J_n is the ideal in $\mathscr{R}$ generated by elements of type $kz^{n/k}$ where k is any divisor of n (e.g. nz and z^n are in J_n).

Corollary 2.2. *Let L be an n-periodic oriented link, then*
(i) *(Theorem 1.2) If n is a prime number then*

$$P_L(a, z) \equiv P_L(a^{-1}, z) \operatorname{mod}(n, z^n)$$

if $n = r^q$ *is a power of a prime number, then*
(ii) $P_L(a, z) \equiv P_L(a^{-1}, z) \operatorname{mod}(r^q, z^r)$
(iii) $P_L(a, z) \equiv P_L(a^{-1}, z) \operatorname{mod}(r, z^{r^q})$.

Proof of Corollary 2.2. It follows from Theorem 2.1 because the ideals (n, z^n), (r^q, z^r), and (r, z^{r^q}) are bigger then the ideal J_n.

Proof of Theorem 2.1. By the positive solution of the Smith conjecture [Sm, Thur], the fixed point set of our Z_n-action on S^3 is an unknotted circle and the action is conjugated to an orthogonal one. Therefore if we write S^3 as $R^3 \cup \infty$ we can assume that a fixed point set is a vertical axis with ∞ and our Z_n-action is generated by rotation ϕ given by the formula: $\phi(z, t) = (e^{2\pi i/n} z, t)$ where $R^3 = \{z, t : z$ complex and t real numbers$\}$. Each n-periodic link can be represented by ϕ-invariant diagram (also denoted by L); that is $\phi(L) = L$ or $-L$ where $-L$ denote the link (diagram) obtained from L by reversing its orientation; compare Fig. 2.1.

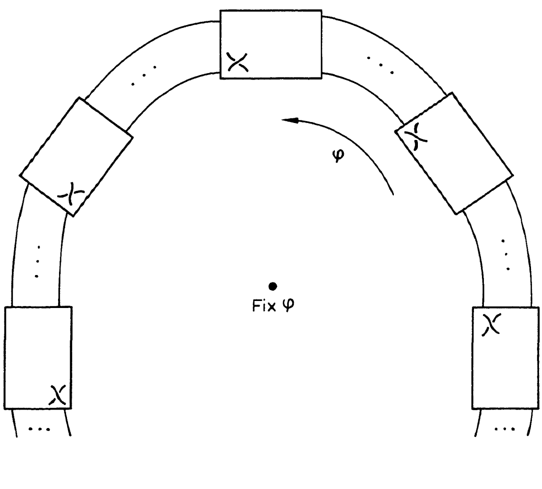


Fig. 2.1

Let $L_{\mathrm{sym}(\times)}$, $L_{\mathrm{sym}(\times)}$, and $L_{\mathrm{sym}(\asymp)}$ denote three ϕ-invariant diagrams which are identical except near the Z_n orbit of a single crossing where all n crossings of the orbit are positive in $L_{\mathrm{sym}(\times)}$, negative in $L_{\mathrm{sym}(\times)}$ and smoothed in $L_{\mathrm{sym}(\asymp)}$. We have the following fact.

Lemma 2.3.

$$a^n P_{L_{\mathrm{sym}(\times)}}(a, z) + (-1)^{n+1} a^{-n} P_{L_{\mathrm{sym}(\times)}}(a, z) \equiv z^n P_{L_{\mathrm{sym}(\asymp)}}(a, z) \operatorname{mod} \hat{J}_n,$$

where $\hat{J}_n$ *is the ideal in* $\mathcal{R}$ *generated by elements of type* $kz^{n/k}$ *where k is any divisor of n but* 1.

Proof (based on the idea of Murasugi [Mu 2]). Let p be a crossing of $L_{\mathrm{sym}(\times)}$ such that $L_{\mathrm{sym}(\times)}$ and $L_{\mathrm{sym}(\asymp)}$ differ only in crossings $p, \phi(p), \phi^2(p), \ldots, \phi^{n-1}(p)$. Let us build the (part of) binary resolving tree for $L_{\mathrm{sym}(\times)}$, which uses crossings $p, \phi(p), \ldots, \phi^{n-1}(p)$; compare Fig. 2.2. Now we analyse what value is introduced to $P_{L_{\mathrm{sym}(\times)}}(a,z)$ by leaves of our binary tree. The leaf $L_{\mathrm{sym}(\times)}$ gives $(-1)^n a^{-2n} P_{L_{\mathrm{sym}(\times)}}(a,z)$ to $P_{L_{\mathrm{sym}(\times)}}(a,z)$ and the leaf $L_{\mathrm{sym}(\asymp)}$ gives $a^{-n}z^n P_{L_{\mathrm{sym}(\asymp)}}(a,z)$. Observe that Z_n acts on leaves of the binary tree and the only fixed points of the action are $L_{\mathrm{sym}(\times)}$ and $L_{\mathrm{sym}(\asymp)}$. Let L^l be a leaf of the binary tree which is not a fixed point set and let k be the order of its orbit. Of course k divides n and $k > 1$. Our leaf has been obtained from $L_{\mathrm{sym}(\times)}$ by applying smoothing at least $\frac{n}{k}$ times so it introduced to $P_{L_{\mathrm{sym}(\times)}}(a,z)$ the value $z^{n/k}P_{L^l}(a,z)(\cdot)$. Furthermore each element of the orbit of L^l is isotopic to L^l so the orbit of L^l introduced to $P_{L_{\mathrm{sym}(\times)}}(a,z)$ the value $kz^{n/k}P_{L^l}(a,z)(\cdot)$ which is an element of the ideal $\hat{J}_n$. Therefore

$$P_{L_{\mathrm{sym}(\times)}}(a,z) \equiv (-1)^n a^{-2n} P_{L_{\mathrm{sym}(\times)}}(a,z) + a^{-r}z^r P_{L_{\mathrm{sym}(\asymp)}}(a,z) \bmod \hat{J}_n .$$

This completes the proof of Lemma 2.3. □

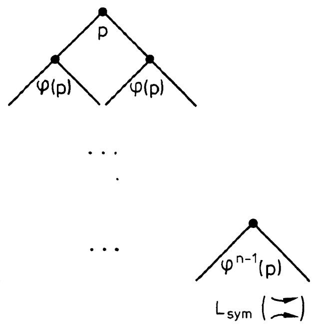


Fig. 2.2

Now we can complete the proof of Theorem 2.1. From Lemma 2.3 it follows immediately that

$$P_{L_{\mathrm{sym}(\times)}}(a,z) \equiv (-a^2)^{-n} P_{L_{\mathrm{sym}(\times)}}(a,z) \bmod J_n$$

(because $z^n \in J_n$). On the other hand one can go from L to its mirror image $\bar{L}$ using changes $L_{\mathrm{sym}(\times)} \leftrightarrow L_{\mathrm{sym}(\times)}$ therefore $P_L(a,z) \equiv (-a^2)^{-nj} P_{\bar{L}}(a,z) \bmod J_n$ where nj is equal to the writhe number of the diagram L. Now, from the well known equality $P_{\bar{L}}(a,z) = P_L(a^{-1},z)$, it follows that $P_L(a,z) \equiv (-a^2)^{-nj} P_L(a^{-1},z) \bmod J_n$. To complete the proof of Theorem 2.1 we need the following fact.

Lemma 2.4. *$a^{2n} + (-1)^{n+1}$ is in the ideal J_n.*

Proof. J_n is generated by elements $kz^{n/k}$ so the elements $kz^{n/k}\left(\frac{a+a^{-1}}{z}\right)^{n/k} = k(a+a^{-1})^{n/k}$ are also in J_n so $k(a^2+1)^{n/k} \in J_n$. To prove Lemma 2.4 it is sufficient to show that $a^{2n} + (-1)^{n+1}$ is in the ideal generated by elements $k(a^2+1)^{n/k}$ in $Z[a]$. This reduces, after substituting $y = a^2+1$, to showing that $(y-1)^n + (-1)^{n+1}$ is in the ideal generated by the elements $ky^{n/k}$ in $Z[y]$ but it follows from the well known fact that for any natural number i, $\binom{n}{i}$ is a multiplicity of $\frac{n}{gcf(n,i)}$ so $\binom{n}{i} y^i$ is a multiplicity of $\frac{n}{gcf(n,i)} y^{gcf(n,i)}$ which is in the ideal. It completes our proof of Lemma 2.4 and Theorem 2.1. $\square$

We can generalize Theorem 2.1 (or rather show its limits) by considering the following operations on link diagrams:

Definition 2.5 [P2]. (a) A t_k move is an elementary operation on an oriented link diagram L resulting in the diagram $t_k(L)$ as shown on Fig. 2.3. Two oriented links L and L' are said to be t_k equivalent if one can go from L to L' using $t_k^{\mp 1}$ moves (and isotopy).

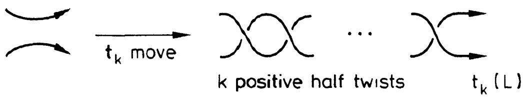


Fig. 2.3

(b) A $\bar{t}_k$ move is an elementary operation on an oriented link diagram L resulting in the diagram $\bar{t}_k(L)$, which is naturally oriented for k even, as shown on Fig. 2.4. Two oriented links L and L' are said to be $\bar{t}_k$ equivalent (k even) if one can go from L to L' using $\bar{t}_k^{\mp 1}$ moves (and isotopy).

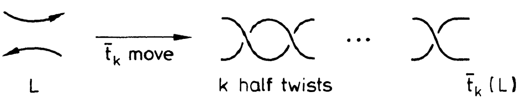


Fig. 2.4

(c) Two unoriented links L and L' are called k-equivalent (or t_k, $\bar{t}_k$ equivalent) if one can go from L to L' using $t_k^{\mp 1}$ or $\bar{t}_k^{\mp 1}$ moves and ignoring orientation. For k even k-equivalence is also well defined for oriented links.

Theorem 2.6. *For every oriented link L*

(a) $P_{t_{2n}(L)}(a,z) \equiv P_L(a,z) \bmod J_n$

(b) $P_{\bar{t}_{2n}(L)}(a,z) \equiv P_L(a,z) \bmod J_n$.

Proof. It follows essentially from Theorems 1.1 and 1.7 of [P2] but we can give a short proof independent from [P2]. We need, first, the lemma corresponding to Lemma 2.3.

Lemma 2.7. (a) $a^n P_{t_n(L)}(a,z) + (-1)^{n+1} a^{-n} P_{t_n^{-1}(L)}(a,z) \equiv z^n P_L(a,z) \bmod \hat{J}_n$

(b) $a^{-n} P_{\bar{t}_{2n}(L)}(a,z) + (-1)^{n+1} a^n P_L(a,z) \equiv z^n \left(\frac{a+a^{-1}}{z}\right)^{n-2} P_{L_\infty}(a,z) \bmod \hat{J}_n$

where $\bar{t}_{2n}(L)$, L, and L_∞ are shown on Fig. 2.5.

Fig. 2.5

Proof. (a) Consider n crossings, say $p_1, \ldots, p_n$, of the diagram $t_n(L)$ in which $t_n(L)$ differ from $t_n^{-1}(L)$ (see Fig. 2.6). Let us build the (part of) binary resolving tree for $t_n(L)$ using the crossings $p_1, p_2, \ldots, p_n$ (compare there proof of Lemma 2.3). L is the leaf of the tree obtained from $t_n(L)$ by performing n smoothings, so it introduces to

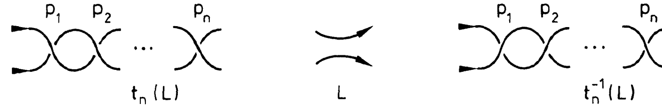


Fig. 2.6

$P_{t_n(L)}(a,z)$ the value $a^{-n}z^nP_L(a,z)$. $t_n^{-1}(L)$ is the leaf obtained from $t_n(L)$ by performing n $\mp$ changes, so it introduces to $P_{t_n(L)}(a,z)$ the value $(-1)^n a^{-2n}P_{t_n^{-1}(L)}(a,z)$. Now consider leaves obtained from $t_n(L)$ by k smoothings and $(n-k)$ $\mp$ changes. The crucial observation is that these leaves are all isotopic (say to $L^{l(k)}$) so there is $\binom{n}{k}$ such leaves and they introduce (together) to $P_{t_n(L)}(a,z)$ the value $\binom{n}{k} z^k a^{-k}(-a^2)^{n-k}P_{L^{l(k)}}(a,z)$. For $0<k<n$, $\binom{n}{k} z^k$ is in the ideal $\hat{J}_n$ (compare Lemma 2.4) so $P_{t_n(L)}(a,z) \equiv a^{-n}z^nP_L(a,z)+(-1)^n a^{-2n}P_{t_n^{-1}(L)}(a,z) \mod \hat{J}_n$ and Lemma 2.7 (a) follows. A proof of Lemma 2.7 (b) is similar to that of (a) and we omit it. □

Theorem 2.6 follows now from Lemmas 2.7 and 2.4. □

Corollary 2.8. *Let L be an oriented link which is t_{2n}, $\bar{t}_{2n}$ equivalent to an n-periodic link, then*

$$P_L(a,z) \equiv P_L(a^{-1},z) \bmod J_n. \quad \square$$

Now we can prove the practical criterion for $w(a,z) \in \mathscr{R}$ to be in the ideal J_n; it generalize Lemma 1.5.

Lemma 2.9. *Let $w(a,z) \in \mathscr{R}$ and $w(a,z) = \sum_i v_i(a)z^i$ where $v_i(a) \in Z[a^{\mp 1}]$. Then $w(a,z) \in J_n$ if an only if for any i, $v_i(a)$ is an element of the ideal $J_n^i(a)$ in $Z[a^{\mp 1}]$ where $J_n^i(a)$ is generated by elements $k(a+a^{-1})^{\max(0,n/k-i)}$ where k is any divisor of n.*

Proof. Let $\mathscr{R}'$ denote the ring $\mathscr{R}$ treated as $Z[a^{\mp 1}]$-module and consider the ideal J_n' in $\mathscr{R}'$ generated by $z^iJ_n^i(a)$ [i.e. by elements $z^ik(a+a^{-1})^{\max(0,n/k-i)}$]. This is chosen so that $v_i(a) \in J_n^i(a)$ for any i iff $w(a,z) \in J_n'$. Now Lemma 2.9 says that $J_n' = J_n$. First of all $J_n' \subset J_n$ because $z^ik(a+a^{-1})^{\max(0,n/k-i)} = kz^{n/k}\left(\frac{a+a^{-1}}{z}\right)^{\max(0,n/k-i)}$ $\times z^{\max(0,i-n/k)} \in J_n$. On the other hand J_n' is an ideal in $\mathscr{R}$, not only in $\mathscr{R}'$, because $J_n^i(a) \subset J_n^{i+1}(a)$ for any i. Therefore $J_n' \supset J_n$ because $kz^{n/k} \in z^{n/k}J_n^{n/k}(a) \subset J_n'$, and therefore $J_n' = J_n$. □

The following corollary is the slight generalization of the Traczyk result [Tr 1].

Corollary 2.10. (a) *Let L be an oriented link t_{2n}, $\bar{t}_{2n}$ equivalent to an n-periodic link. Then the Jones polynomial $V_L(t)$ satisfies:*

(i) $V_L(t) \equiv V_L(t^{-1}) \bmod J_n(t)$ *where $J_n(t)$ is the ideal in $Z\left[t^{\frac{\mp 1}{2}}\right]$ generated by elements of type $k(t-1)^{n/k}$ where k is any divisor of n. In particular*

(ii) *(see Corollary 1.3) If n is a prime number then $V_L(t) \equiv V_L(t^{-1}) \bmod (n, t^n - 1)$. If $n = r^q$ is a power of a prime number, then*

(iii) $V_L(t) \equiv V_L(t^{-1}) \bmod (r, t^{r^q} - 1)$ *and*

(iv) $V_L(t) \equiv V_L(t^{-1}) \bmod (r^q, (t-1)^r)$

(b) *consider the polynomial invariant of (global) isotopy of unoriented links defined by*

$$\hat{V}_L(t)=(t^{1/2})^{-3lk(L')}\,V_{L'}(t),$$

where L' is an oriented link which reduces to L after ignoring its orientation, and $lk(L')$ is the global linking number of L'. Then for any nonoriented link L, t_{2n}, $\bar{t}_{2n}$ equivalent to an n periodic link:

(i) *If n is odd then $\hat{V}_L(t)\equiv t^{nlk_2L}\hat{V}_L(t^{-1}) \bmod J_n(t)$ where $lk_2(L)=0$ or 1 is a global linking number of $L \bmod 2$. It does not depend on the orientation of L and for any orientation L' of L $lk_2(L)\equiv lk(L') \bmod 2$.*

In particular if $n=r^q$ is a power of odd prime then

(ii) $\hat{V}_L(t)\equiv \hat{V}_L(t^{-1}) \bmod (r, t^{r^q}-1)$ *and*

(iii) $\hat{V}_L(t)\equiv \hat{V}_L(t^{-1}) \bmod (r^q, (t-1)^r)$.

Proof. $V_L(t)=P_L(a,z)$ for $a=it^{-1}$, $z=i\left(\sqrt{t}-\dfrac{1}{\sqrt{t}}\right)$ therefore (a) is an easy corollary of Theorem 2.1 and Corollaries 2.2 and 2.8. To prove (b) of Corollary 2.10 observe that we can always orient L so that ϕ preserves the orientation (orient first L/Z_n and then lift the orientation to L). For such oriented L, denoted L', we have by part (a):

$$V_{L'}(t)\equiv V_{L'}(t^{-1}) \bmod J_n(t)$$

and therefore

$$(t^{1/2})^{3lkL'}\hat{V}_L(t)\equiv (t^{1/2})^{-3lkL'}\hat{V}_L(t^{-1}) \bmod J_n(t)$$

and so

$$\hat{V}_L(t)=t^{-3lkL'}\hat{V}_L(t^{-1}) \bmod J_n(t).$$

Now if n is an odd number then lkL' is a multiplicity of n. Furthermore $lk(L')+nlk_2(L)$ is a multiplicity of $2n$ and therefore $-3lkL'\equiv nlk_2(L) \bmod 2n$, and so the part b(i) of Corollary 2.10 follows [$t^{2n}-1$ is in $J_n(t)$; compare Lemma 2.4]. b(ii) and (iii) follow similarly as in part (a) or in Corollary 2.2.

3. Criterion for n-Periodic Links Using the Kauffman Polynomial

Theorem 3.1. *Let L be an oriented link which is t_{2n}, $\bar{t}_{2n}$ equivalent to an n-periodic link, then the Kauffman polynomial of L satisfies:*

$$F_L(a,z)\equiv F_L(a^{-1},z) \bmod J_n.$$

Our proof of Theorem 3.1 is very similar to that of Theorem 2.1 and Corollary 2.8 so we will only sketch it here.

We start by considering nonoriented diagrams of n periodic link and its invariant of regular isotopy $\Lambda(a,z)$. We use the notation analogous to that of the proof of Theorem 2.1. In particular we consider nonoriented n-periodic diagrams

$$L_{\mathrm{sym}(\times)},\ L_{\mathrm{sym}(\times)},\ L_{\mathrm{sym}(\asymp)} \quad \text{and} \quad L_{\mathrm{sym}()()}.$$

Lemma 3.2. (a) $\Lambda_{L_{\mathrm{sym}(\times)}}+(-1)^{n+1}\Lambda_{L_{\mathrm{sym}(\times)}}\equiv z^n(\Lambda_{\mathrm{sym}(\asymp)}+\Lambda_{\mathrm{sym}()()}) \bmod \hat{J}_n$

(b) $\Lambda_{L_{\mathrm{sym}(\times)}}\equiv(-1)^n\Lambda_{\mathrm{sym}(\times)} \bmod J_n$.

Proof. Let p be a crossing of $L_{\mathrm{sym}(\times)}$ such that $L_{\mathrm{sym}(\times)}$ and $L_{\mathrm{sym}(\times)}$ differ only in crossings $p, \phi(p), \ldots, \phi^{n-1}(p)$. Let us consider the (part of) trinary resolving tree for $L_{\mathrm{sym}(\times)}$ which uses crossings $p, \phi(p), \ldots, \phi^{n-1}(p)$. Z_n acts on leaves of the trinary tree and the only fixed points of the action are $L_{\mathrm{sym}(\times)}$, $L_{\mathrm{sym}(\asymp)}$, and $L_{\mathrm{sym})(}$. They introduce to $\Lambda_{L_{\mathrm{sym}(\times)}}$ the values $(-1)^n \Lambda_{L_{\mathrm{sym}(\times)}}$, $z^n \Lambda_{L_{\mathrm{sym}(\asymp)}}$, and $z^n \Lambda_{L_{\mathrm{sym})(}}$ respectively. All other orbits of the Z_n action on the leaves introduce to $\Lambda_{L_{\mathrm{sym}(\times)}}$ a value which is in the ideal $\hat{J}_n$ (see the proof of Lemma 2.3) so Lemma 3.2(a) follows. Part (b) follows immediately from (a). □

Now we will show that the equality from Theorem 3.1 holds for any n-periodic oriented link. It follows from Lemma 3.2(b) that for n-periodic oriented link diagram

$$a^{w(L_{\mathrm{sym}(\times)})} F_{L_{\mathrm{sym}(\times)}} = (-1)^n a^{w(L_{\mathrm{sym}(\times)})} F_{L_{\mathrm{sym}(\times)}} \bmod J_n .$$

Because $w(L_{\mathrm{sym}(\times)}) - w(L_{\mathrm{sym}(\times)}) = \mp 2n$ therefore by Lemma 2.4

$$F_{L_{\mathrm{sym}(\times)}} \equiv F_{L_{\mathrm{sym}(\times)}} \bmod J_n .$$

We can go from L to its mirror image $\bar{L}$ using changes $L_{\mathrm{sym}(\times)} \leftrightarrow L_{\mathrm{sym}(\times)}$ and because $F_{\bar{L}}(a,z) = F_L(a^{-1}, z)$ we have

$$F_L(a,z) \equiv F_L(a^{-1}, z) \bmod J_n .$$

To complete the proof of Theorem 3.1 we need the following Lemma.

Lemma 3.3. (a) *Let L_n and L_{-n} be a nonoriented link diagrams which are the same except the part shown on Fig. 3.1. Then*

$$\Lambda_{L_n} \equiv (-1)^n \Lambda_{L_{-n}} \bmod J_n .$$

(b) *If L and L' are oriented $t_{2n}, \bar{t}_{2n}$ equivalent links then $F_L(a,z) \equiv F_{L'}(a,z) \bmod J_n$.*

Fig. 3.1 L_n n positive half twists n negative half twists L_{-n}

Proof. (a) Consider the (part of) trinary resolving tree for L_n using crossings $p_1, p_2, \ldots, p_n$. Let $\psi(p_i)$ be defined $\psi(p_1) = p_2, \psi(p_2) = p_3, \ldots, \psi(p_n) = p_1$. Now consider the Z_n action on leaves of the tree with the generator g defined as follows: Let L^l be any leaf of the tree, then $g(L^l)$ is the leaf related to L^l as follows. If an operation ($\times$ to $\times$ or $\asymp$ or)() was performed on p_i in L^l, the same operation is performed on $\psi(p_i)$ in $g(L^l)$. The elements in an orbit of the Z_n action are all isotopic so we can complete our proof of Lemma 3.3(a) in the same manner as that of Lemma 3.2. To proof (b) of Lemma 3.3, observe that $w(L_n) - w(L_{-n}) = \mp 2n$ for any orientation of L_n (and corresponding orientation of L_{-n}) so by Lemma 2.4, $F_{L_n}(a,z) \equiv F_{L_{-n}}(a,z) \bmod J_n$. Hence Lemma 3.3(b) and Theorem 3.1 follows. □

Corollary 3.4. *Consider an invariant of (global) isotopy of nonoriented links, $\hat{F}_L(a,z)$, defined by the formula $\hat{F}_L(a,z) = a^{2lk(L')} F_{L'}(a,z)$ where L' is any orientation on nonoriented link L. Then for any nonoriented link L which is $t_{2n}, \bar{t}_{2n}$ equivalent to a Z_n periodic link*

$$\hat{F}_L(a,z) \equiv \hat{F}_L(a^{-1}, z) \bmod J_n .$$

Proof. Let L' denote L oriented in such a way that Z_n preserves the orientation [compare the proof of Corollary 2.10 (b)]. Then by Theorem 3.1 $F_{L'}(a,z) \equiv F_{L'}(a^{-1},z) \bmod J_n$. Therefore $a^{-2lk(L')}\hat{F}_L(a,z) \equiv a^{2lk(L')}\hat{F}_L(a^{-1},z) \bmod J_n$, so $\hat{F}_L(a,z) \equiv a^{4lk(L')}\hat{F}_L(a^{-1},z) \bmod J_n$. Finally by Lemma 2.4 $a^{2n} \equiv (-1)^n \bmod J_n$ so $a^{4lk(L')} \equiv 1 \bmod J_n$ and Corollary 3.4 follows. □

We can consider the simplified version, $S_L(a)$, of the Kauffman polynomial by substituting $z=a+a^{-1}$ in $F_L(a,z)$. $S_L(a)$ is interesting on its own and resemble somehow the Alexander polynomial [if L is a split link then $S_L(a)=0$; if L is a knot then $S_L(a)=1+(a+a^{-1})(\cdot)$]. Our criterion for n periodic links also simplifies when one applies $S_L(a)$. Namely, Theorem 3.1 reduces to

Corollary 3.5. *Let L be an oriented link which is t_{2n}, $\bar{t}_{2n}$ equivalent to an n-periodic link, then*

$$S_L(a) \equiv S_L(a^{-1}) \bmod J_n(a),$$

where $J_n(a)$ is the ideal in $Z[a^{\mp 1}]$ generated by elements of type $k(a+a^{-1})^{n/k}$ where k is any divisor of n. In particular for $n=r^q$ (a power of a prime number)

$$S_L(a) \equiv S_L(a^{-1}) \bmod (r, a^{2r^q}+1)$$

and

$$S_L(a) \equiv S_L(a^{-1}) \bmod (r^q, (a+a^{-1})^r). \quad \square$$

4. Examples and Further Speculations

Example 4.1. Consider the right handed trefoil knot, 3_1 (Fig. 4.1). The skein polynomial $P_{3_1}(a,z)=(-2a^{-2}-a^{-4})+a^{-2}z^2$. Then $w(a,z)=P_{3_1}(a,z)-P_{3_1}(a^{-1},z)$ $=(a+a^{-1})^3(a-a^{-1})-z^2(a+a^{-1})(a-a^{-1})$. Therefore, $w(a,z) \in J_n$ iff $n=2,3,4,6$ or 12. Hence by Theorem 2.1 the right handed trefoil knot is not n periodic for $n \neq 2,3,4,6$ or 12. If one consider the Kauffman polynomial one cannot do any better so our criteria cannot exclude periods 4, 6, and 12 (compare Lemma 4.5).

Example 4.2. Consider the knot 10_{137} [Ro], Fig 4.1. It is 10_{124} in [Th]. The Kauffman polynomial

$$\begin{aligned} F_{10_{137}}(a,z) = & -a^{-2}-1-2a^2-2a^4-a^6+z(-a^{-1}-3a^1-5a^3-3a^5) \\ & +z^2(a^{-2}+4+7a^2+8a^4+4a^6) \\ & +z^3(2a^{-1}+9a+15a^3+8a^5) \\ & +z^4(\cdot);\ [\mathrm{Th}]. \end{aligned}$$

Consider the difference $w(a,z)=F_{10_{137}}(a,z)-F_{10_{137}}(a^{-1},z)$. One can check that $w(a,z) \in J_n$ iff $n=2,3,4,6$ or 12 so 10_{137} is not n-periodic for $n \neq 2,3,4,6$, and 12. Let us check it in more detail for $n=5$. Consider $v_0(a)$ for $w(a,z)$ (in the notation of Lemma 1.5). Modulo the ideal $(5, a^5+a^{-5})$ one gets:

$$\begin{aligned} v_0(a) & \equiv -a^{-2}-1-2a^2-2a^4-a^6+a^2+1+2a^{-2}+2a^{-4}+a^{-6} \\ & \equiv a^{-2}-a^2+2a^{-4}-2a^4+a^{-4}-a^4 \equiv a^{-2}-a^2+3a^{-4}-3a^4 \\ & \not\equiv 0 \bmod (5, a^5+a^{-5}) \quad \text{so} \quad w(a,z) \notin J_5. \end{aligned}$$

The case $n=5$ is interesting because in [BZ] the knot 10_{137} has, by mistake, associated with it period 5. I would like to thank Murasugi for pointing out the possibility of mistakes in [BZ].

Example 4.3. Consider the right handed Hopf link H_1 (Fig. 4.1). The Kauffman polynomial

$$F_{H_1}(a,z)=(-a^{-3}-a^{-1})z^{-1}+a^{-2}+(a^{-3}+a^{-1})z.$$

Then $w(a,z)=F_{H_1}(a,z)-F_{H_1}(a^{-1},z)$ is in J_n iff $n=2$. Therefore the only period of H_1 is equal to 2.

Example 4.4. Consider the oriented link 6_1^3 as on Fig. 4.1. The skein polynomial $P_{6_1^3}(a,z)=(-a^{-4}-2a^{-2}-1)z^{-2}+(a^{-4}+3a^{-2}+2)+(-a^{-4}-3a^{-2}-1)z^2+a^{-2}z^4$. Then $w(a,z)=P_{6_1^3}(a,z)-P_{6_1^3}(a^{-1},z)=(a+a^{-1})\ (a-a^{-1})\left(\dfrac{a+a^{-1}}{z}\right)^2+(a+a^{-1})^2(\cdot)+z^2(\cdot)$. Therefore $w(a,z)\in J_n$ iff $n=2$. Hence 6_1^3 is not n periodic for $n>2$.

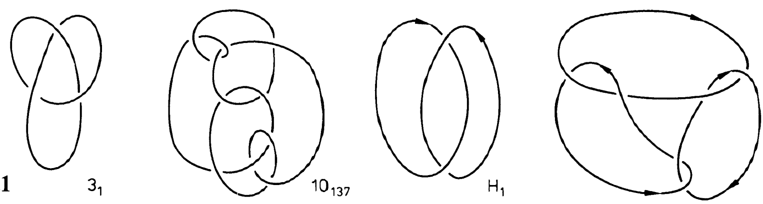


Fig. 4.1 3_1 10_{137} H_1 6_1^3

Examples 1.6, 1.7, 4.1, and 4.2 suggest that our criteria cannot exclude periods 2 and 3 for knots. For links of more than one components one can exclude period 3 as shown in Examples 4.3 and 4.4 (notice that the global linking number is equal to ∓ 1 in these examples). In fact the following holds

Lemma 4.5. *For any link L*

(a) $P_L(a,z)-P_L(a^{-1},z)\in J_2$

(b) $F_L(a,z)-F_L(a^{-1},z)\in J_2$.

For any knot K

(c) $P_K(a,z)-P_K(a^{-1},z)\in J_3$

(d) $F_K(a,z)-F_K(a^{-1},z)\in J_3$

(e) $P_K(a,z)-P_K(a^{-1},z)\in J_4$

(f) $F_K(a,z)-F_K(a^{-1},z)\in J_4$.

Therefore our criteria do not work for period 2 *and for knots and periods* 3 *and* 4.

Proof. Let us write $P_L(a,z)$ or $F_L(a,z)$ as

$$\sum_{i\leqq 0} u_i(a)\left(\frac{a+a^{-1}}{z}\right)^i+v_1(a)z+\sum_{i\geqq 2} v_i(a)z^i.$$

Then $u_i(a)$ is in $Z[a^{\mp 2}]$ so $u_i(a)-u_i(a^{-1})$ is in the ideal $((a+a^{-1})(a-a^{-1}))$ of $Z[a^{\mp 1}]$. $v_1(a)-v_1(a^{-1})$ is in the ideal $(a-a^{-1})$ of $Z[a^{\mp 1}]$ and $\sum_{i\geqq 2} v_i(a)z^i$ is in the ideal (z^2) of $\mathscr{R}$. Combining this one gets that $P_L(a,z)-P_L(a^{-1},z)$ and $F_L(a,z)-F_L(a^{-1},z)\in J_2=(2z,z^2)$.

To prove (c) of Lemma 4.5 we use the folklore fact (compare [LM], [L] or [P1]) that for any knot K, $P_K(a,z)-1$ is a multiple of $(a+a^{-1})^2-z^2$. Because $P_K(a,z)=\sum_{i\geqq 0} v_{2i}(a)z^{2i}$ where $v_{2i}(a)\in Z[a^{\mp 2}]$ so $v_0(a)-v_0(a^{-1})$ is a multiple of $(a+a^{-1})^3$ and $v_2(a)-v_2(a^{-1})$ is a multiple of $(a+a^{-1})$ so $P_K(a,z)-P_K(a^{-1},z)\in J_3=(3z,z^3)$; compare Lemma 2.9. Now consider $F_K(a,z)=\sum_{i\geqq 0} w_i(a)z^i$. $w_0(a)-w_0(a^{-1})$ is a multiple of $(a+a^{-1})^3$ (the first coefficients of $F_K(a,z)$ and $P_K(a,z)$ are the same; compare [L] or [P1]). $w_2(a)-w_2(a^{-1})$ is a multiple of $(a+a^{-1})^2$ because $w_2(a)\in Z[a^{\mp 2}]$. Finally $w_1(a)-w_1(a^{-1})$ is a multiple of $(a+a^{-1})^2$. To prove this it is enough to show that $w_1(a)$ is a multiple of $(a+a^{-1})$. By [LM2], $F_K(a,z)-1$ is a multiple of $(a+a^{-1}-z)$. Therefore $F_K(a,z)-1/a+a^{-1}-z=\sum_{i\geqq 0} s_i(a)z^i$; furthermore the first coefficients of $F_K(a,z)-1/a+a^{-1}-z$ and $P_K(a,z)-1/a+a^{-1}-z$ are the same, so $s_0(a)$ is multiple of $(a+a^{-1})$ [we use the fact that $P_K(a,z)-1$ is a multiple of $(a+a^{-1})^2-z^2$]. Finally because $w_1(a)=s_0(a)+(a+a^{-1})s_1(a)$, hence $w_1(a)$ is a multiple of $(a+a^{-1})$. By Lemma 2.9, $F_K(a,z)-F_K(a^{-1},z)\in J_3$ so Lemma 4.5(d) is proven. 4.5(e) and (f) can be proven similarly. One should only additionally observe that if a polynomial $w(a)\in Z[a^{\mp 2}]$ then $w(a)-w(a^{-1})$ is a multiple of
$$(a+a^{-1})(a-a^{-1})=(a+a^{-1})(a+a^{-1}-2a^{-1}).$$
It completes our proof of Lemma 4.5.

Our examples show that even if our link is p periodic our criteria can detect the lack of p^2 periodicity. On the other hand if our criteria do not exclude n and m-periodicity (n and m co-prime numbers) then they cannot exclude nm-periodicity. It is the case because if $w\in J_n$ and $w\in J_m$ then $w\in J_{mn}$. (Lemma 2.9 allows us easily to prove that $J_n\cap J_m=J_{nm}$.) We conclude, combining the above with Lemma 4.5 that our criteria cannot exclude a periodicity of knots for $n=2$, 3, 4, 6, and 12 (compare Examples 4.1 and 4.2).

If a link L is n-periodic then $L_*=L/Z_n$ is a link in the 3-sphere S^3/Z_n. The idea of Murasugi [Mu 2] is to compare polynomial invariant of L and L_* and to show that for some ideal I_n a polynomial invariant W satisfies:

4.6 $$W_L\equiv W_{L_*}^n \bmod I_n.$$

For the Jones polynomial $V_L(t)$, Murasugi has shown [Mu 2] that if L is an r-periodic oriented link and r a prime number then $V_L(t)\equiv V_{L_*}^r(t) \bmod I_r(t)$ where $I_r(t)$ is the ideal in $Z[t^{\mp 1/2}]$ generated by r and $\left(-\frac{t+1}{\sqrt{t}}\right)^{r-1}-1$. The similar result can be obtained for skein and Kaufmann polynomials. In the case of skein polynomial one gets
$$P_L(a,z)\equiv P_{L_*}^r(a,z) \bmod I_r,$$
where I_r is the ideal in $Z[a^{\mp 1},z^{\mp 1}]$ generated by r, $\left(\frac{a+a^{-1}}{z}\right)^{r-1}-1$ and polynomials $P_{T_{k,r}}(a,z)$ where $T_{k,r}$ is the torus link of type k,r. To make 4.6 useful one should analyse how big is the ideal I_r. The idea of a proof of 4.6 is based on the observation that, to the binary resolving tree of L_* corresponds r-periodic binary resolving tree for L (one uses triplet $L_{\mathrm{sym}(\times)}$, $L_{\mathrm{sym}(\times)}$, $L_{\mathrm{sym}(\asymp)}$). The second tree

suffices to compute $P_L(a,z) \bmod r$ assuming the values at leaves are known (compare the proof of Theorem 2.1). Finally notice that the above method can be used to prove another Murasugi result [Mu 1] that for an r-periodic knot the Alexander polynomial satisfies:

$$\Delta_L(t) \equiv \Delta^r_{L_*}(t)(1+t+\ldots+t^{\lambda-1})^{r-1} \bmod r$$

for some integer λ.

The method of our paper can be also applied for Z_n invariant links when the fixed point set of the action is a part of the link. Then we have to use Hoste-Kidwell polynomial [HK] or its simplified version introduced in [HP]. We can also analyse Z_n invariants links when Z_n acts freely on S^3. We will describe it in the sequel paper (compare [P 4]).

Finally we hope that the analysis of n-periodic (or symmetric) links in any 3-manifolds can lead to some unified theory of skein modules of coverings (compare [P 3]).

References

[APR] Anstee, R.P., Przytycki, J.H., Rolfsen, D.: Knot Polynomials and generalized mutation. Topology Appl. (to appear)

[BZ] Burde, G., Zieschang, H.: Knots. De Gruyter studies in Math. 5. Berlin New York: De Gruyter 1985

[FYHLMO] Freyd, P., Yetter, D., Hoste, J., Lickorish, W.B.R., Millet, K., Ocneanu, A.: A new polynomial invariant of knots and links. Bull. Am. Math. Soc. **12**, 239–249 (1985)

[H] Hillman, J.A.: Symmetries of knots and links, and invariants of abelian coverings. Parts I and II. Kobe J. Math. **3**, 7–24 and 149–165 (1986)

[HK] Hoste, J., Kidwell, M.: Dichromatic link invariants. Trans. Am. Math. Soc. (to appear)

[HP] Hoste, J., Przytycki, J.H.: An invariant of dichromatic links. Proc. Am. Math. Soc. (to appear)

[J] Jones, V.F.R.: Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**(2), 335–388 (1987)

[Ka] Kauffman, L.H.: An invariant of regular isotopy. Preprint 1985

[L] Lickorish, W.B.R.: The panorama of polynomials for knots, links, and skeins. Proceedings of the Santa Cruz conference on Artin's braid groups (1986) (to appear)

[LM] Lickorish, W.B.R., Millett, K.C.: A polynomial invariant of oriented links. Topology **26**, 107–141 (1987)

[LM 2] Lickorish, W.B.R., Millet, K.C.: An evaluation of the F-polynomial of a link. Preprint 1987

[Mu 1] Murasugi, K.: On periodic knots. Comment. Math. Helv. **46**, 162–174 (1971)

[Mu 2] Murasugi, K.: Jones polynomials of periodic links. Pac. J. Math. **131**, 319–329 (1988)

[Pe] Perko, K.A.: Invariants of 11-crossing knots. Publ. Math. Orsay (1980)

[P 1] Przytycki, J.H.: Survey on recent invariants in classical knot theory. Warsaw University preprint 1986

[P 2] Przytycki, J.H.: t_k Moves on links. Proceedings of the Santa Cruz conference on Artin's braid groups (1986), to appear

[P 3] Przytycki, J.H.: Skein modules of 3-manifold. Preprint 1987

[P 4] Przytycki, J.H.: On Murasugi's and Traczyk's criteria for periodic links. An extended version of the short talk given at Honolulu (1987)

[PT] Przytycki, J.H., Traczyk, P.: Invariants of links of Conway type. Kobe J. Math. **4**, 115–139 (1987)

[Ro] Rolfsen, D.: Knots and links. Math. Lect. Series 7. Berkeley: Publish or Perish 1976

[Sm] The Smith conjecture (J.W. Morgan, H. Bass). New York Academic Press 1984

[Th] Thistlethwaite, M.B.: Knots to 13-crossings. Math. Comp. (to appear)

[Thur] Thurston, W.: The geometry and topology of 3-manifolds. Mimeographed notes, 1977–1979

[Tr 1] Traczyk, P.: 10_{101} has no period 7: a criterion for periodic links. Proc. Am. Math. Soc. (to appear)

[Tr 2] Traczyk, P.: A criterion for knots of period 3. Preprint, 1988

Received May 31, 1988

Math. Ann. 283, 479–489 (1989)

Mathematische Annalen

The Quadratic Schur Subgroup Over Local and Global Fields

C. Riehm*

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada, L8S 4K1

Let K be a field of characteristic 0 and let A be a central simple K-algebra with an involution I. The restriction ω of I to K is an involution of K, and we call I an ω-involution. If ω is the identity, I is said to be an involution of the first kind, otherwise an involution of the second kind.

Suppose now that the dimension of A is n^2. If I is of the first kind, the dimension of the subspace of elements fixed by I is one of $\frac{1}{2}n(n\pm 1)$ – see [Sch, 7.5, Chap. 8]. In this case we define the *type* of I to be $+1$ if the $+$sign prevails, otherwise -1. An involution of type 1 is sometimes called an *orthogonal* involution, and one of type -1 a *symplectic* involution. The *quadratic Brauer class* $[A, I]$ is then defined to be $([A]$, type $I)$ where $[A]\in \mathrm{Br}(K)$ is the Brauer class of A. If (B, J) is another central simple algebra B with involution J of the first kind, then $I\otimes J$ is an involution of the first kind on the central simple algebra $A\otimes B$, and it is easy to check that type $I\otimes J=$(type I) (type J). It follows that the set of quadratic Brauer classes is a multiplicatively closed subset of $B(K)\times\{\pm 1\}$, and therefore also a subgroup since $B(K)$ is a torsion group; we call it the *quadratic Brauer group* $B(K, \mathrm{id})$. It follows from a theorem of A.A. Albert that

$$B(K,\mathrm{id})={}_2B(K)\times\{\pm 1\},$$

where ${}_2B(K)$ denotes the subgroup of $B(K)$ of exponent 2 – see Sect. 1.

Suppose now that $\omega\neq\mathrm{id}$. In this case one defines $[A, I]$ to be simply the Brauer class $[A]$, and the quadratic Brauer group $B(K,\omega)$ is the set of Brauer classes that arise in this way – it is again a group. It is sometimes convenient in this case to say that I is of type 0, and formally write $[A, I]=([A], 0)$ instead; one also sometimes refers to such an I as a *unitary* involution. Let K_0 be the subfield of K fixed by ω. Another theorem of Albert says that

$$B(K,\omega)=\ker \mathrm{cor}_{K/K_0},$$

where cor_{K/K_0}: $B(K)\to B(K_0)$ is the corestriction map (Theorem 8).

The quadratic Brauer group is defined more generally for a commutative ring K in [HTW]. This definition is related to the one used here in Sect. 1.

---

* Support by N.S.E.R.C. grant #A8778 gratefully acknowledged

Recall that the Schur subgroup $S(K)$ of $B(K)$ consists of the Brauer classes which are represented by a central simple direct summand of the group algebra KG for some finite group G. The quadratic Schur subgroup $S(K,\omega)$ is defined in an analogous manner: an element c in $B(K,\omega)$ is in $S(K,\omega)$ if and only if there is a finite group G and a central simple direct summand A of KG with the following property: A is stable under the canonical ω-involution Ω of KG (which inverts the elements of G and is ω-linear) and $[A,I]=c$ where I is the restriction of Ω to A.

Our principal goal is the determination of $S(K,\omega)$ in the case of K a local or global field.

Let K_c be the largest subcyclotomic extension of $\mathbb{Q}$ contained in K, and let $\tilde{K}$ be the subfield of K_c fixed by the composition of ω and complex conjugation ($\tilde{K}$ is the maximal real subfield of K_c in the case $\omega=\mathrm{id}$). If L/k is any extension of fields, denote by $L\otimes S(k)$ the subgroup of $B(L)$ of classes obtained from those in $S(k)$ by extension of scalars; it is easy to see that $L\otimes S(k)\subseteqq S(L)$.

Lemma 1. *If K is any field of characteristic 0, the image of the forgetful map $S(K,\omega)\to S(K)$ is $K\otimes S(\tilde{K})$, and $S(K,\omega)=K\otimes S(\tilde{K},\omega)$.*

Theorem 2. *Let K be an algebraic number field.*

(i) *If $\omega\neq\mathrm{id}$, $S(K,\omega)=K\otimes S(\tilde{K})$.*

(ii) *If $\omega=\mathrm{id}$ and K is totally imaginary, then*

$$S(K,\mathrm{id})=(K\otimes S(\tilde{K}))\times\{\pm 1\}.$$

(iii) *If $\omega=\mathrm{id}$ and K is not totally imaginary, then $S(K,\mathrm{id})$ consists of the quadratic Brauer classes*

$$(\beta,\varepsilon)\in(K\otimes S(\tilde{K}))\times\{\pm 1\} \tag{1}$$

with $\varepsilon=1$ resp. -1 iff β is split resp. non-split at all real primes.

Remarks. 1. Obviously $K\otimes S(\tilde{K})$ depends on a knowledge of $S(\tilde{K})$, and on the local degrees in $K/\tilde{K}$. In Chap. 7 of Yamada's book [Y], $S(\tilde{K})$ is determined in many cases.

2. We note that (i) actually holds for K an arbitrary field of characteristic 0.

3. (iii) can be given more generally: if $\omega=\mathrm{id}$ and K is any formally real field, then $S(K,\mathrm{id})$ consists of the classes (1) with $\varepsilon=1$ iff β splits in all real closures of K (see Theorem 9). This theorem also contains a simple variant of the Benard-Schacher theorem on the "uniform distribution of invariants" [Y, Theorem 6.1] for formally real fields.

Theorem 3. *Let K be a local field, i.e. a finite extension of $\mathbb{Q}_p$.*

(i) *If $\omega\neq\mathrm{id}$, $S(K,\omega)=1$.*

(ii) *$S(K,\mathrm{id})={}_2S(K)\times\{\pm 1\}$ if K is an odd degree extension of an abelian extension of $\mathbb{Q}_p$, otherwise $S(K,\mathrm{id})=\{\pm 1\}$.*

The proofs are given in Sects. 2 and 3.

1. The Quadratic Brauer Group

We first recall the definition of $B(K,\omega)$ as formulated in [HTW] and, in the case of a field K of characteristic 0, indicate its relationship to the definition given in the introduction.

An *anti-structure* over a commutative ring K is a triple $\mathbf{A}=(A,I,\lambda)$ where A is an algebra (associative with 1) over K, I is an antiautomorphism of A, and λ is a unit of A satisfying $\lambda\lambda^I=1$ and

$$a^{I^2}=\lambda a\lambda^{-1} \quad \text{for all } a\in A.$$

$\mathbf{A}$ is called an ω-antistructure if the restriction of I to K is ω.

We recall that a Morita equivalence between two rings A and B is a quadruple consisting of two bimodules $M={}_BM_A$ and $N={}_AN_B$, and two bimodule isomorphisms $M\otimes_A N\to B$ and $N\otimes_B M\to A$ whose associated pairings $M\times N\to B$ and $N\times M\to A$ (both denoted by $\langle\ ,\ \rangle$), satisfy

$$\langle m,n\rangle m'=m\langle n,m'\rangle \quad\text{and}\quad \langle n,m\rangle n'=n\langle m,n'\rangle$$

for all m,m' in M and n,n' in N. A particular Morita equivalence, called a *derived* Morita equivalence, is obtained as follows: let M be a progenerator for A (i.e. a finitely generated projective module such that A is a direct summand of some direct product $M\times M\times\ldots\times M$), set $N=\mathrm{Hom}_A(M,A)$ and $B=\mathrm{End}_A M$; then the bimodule isomorphisms are given by the canonical maps

$$M\otimes_A \mathrm{Hom}_A(M,A)\to \mathrm{End}_A M\,,\quad\text{and}\quad \mathrm{Hom}_A(M,A)\otimes_B M\to A\,.$$

Suppose now that $\mathbf{A}=(A,I,\lambda)$ and $\mathbf{B}=(B,J,\mu)$ are antistructures and that we have a Morita equivalence between the rings A and B, effected by the modules M and N. Make N into a $B-A$ bimodule by twisting by I and J: $bna:=a^I nb^J$. Suppose that $h:M\to N$ is a bimodule isomorphism satisfying

$$\langle h(m\lambda),m'\rangle^I=\langle h(m'),\mu m\rangle \tag{2}$$

for all m,m' in M. Then we say that the two antistructures are *quadratic Morita equivalent* (cf. [HTW, FM, H]). The quadratic Brauer group as defined in [HTW] is the set of quadratic Morita classes of ω-antistructures on Azumaya algebras, and is a group under tensor product. We shall denote it by $B(K,\omega)'$ in order to distinguish it from the group $B(K,\omega)$ defined in the introduction. We note that there is also a forgetful homomorphism of $B(K,\omega)'$ into $B(K)$ given by $[A,I,\lambda]\mapsto[A]$.

There is also a notion of *derived* quadratic Morita equivalence: Suppose that we have a Morita equivalence between the rings A and B, effected by M and N, and let $\mathbf{A}=(A,I,\lambda)$ be an antistructure. Make N into a right A-module via I, and suppose that $h:M\to N$ is an isomorphism of A-modules. Then

(i) *there is a unique antiautomorphism J on B such that h is also a B-isomorphism when N is made into a left B-module via J,*

and

(ii) *there is a unique unit μ in B such that (B,J,μ) is an antistructure and such that h effects a quadratic Morita equivalence between it and* $\mathbf{A}$.

The *scaling* ${}^u\mathbf{A}$ of an antistructure $\mathbf{A}$ by a unit u of A is the antistructure (A,I',λ') where

$$a^{I'}=u^{-1}a^I u \quad\text{and}\quad \lambda'=u^{-1}u^I\lambda\,.$$

Lemma 4. *Let* **A** *be an antistructure, let $M = A$ as right A-module, and identify both $B = \operatorname{End}_A M$ and $N = \operatorname{Hom}_A(M, A)$ with A via left multiplication. Let $h: M \to N$ be any isomorphism where N is a right A-module via I. Then h is of the form $h(m) = um$ for some unit u in A, and the derived antistructure is* ${}^u\mathbf{A}$.

This is an easy calculation. We note that it follows from this that scaling an antistructure does not change the quadratic Morita equivalence class – one need only define the map h as above using the scaling unit u.

We assume from now on that K is a field of characteristic 0.

Theorem 5. *Two antistructures on the same simple algebra are quadratic Morita equivalent if and only if they are mutual scalings.*

Proof. Let **A** and **B** be the antistructures. As mentioned above, the sufficiency follows from the lemma. So assume that they are quadratic Morita equivalent, say via the bimodules M and N and the isomorphism $h: M \to N$. Since **A** and **B** have the same underlying ring A, $A \cong \operatorname{End}_A M$ and so $M \cong A$; we shall therefore identify M with A. Thus we can also identify N with A, acting via left multiplication. Then

$$h(a) = h(1)a^J = a^I h(1)$$

and so $a^J = u^{-1} a^I u$ with $u = h(1)$ (which is a unit since h is an isomorphism). Similarly (2) with $m = m' = 1$ is $h(\lambda)^I = h(1)\mu$, which implies that $\mu = u^{-1} u^I \lambda$. □

Theorem 6. *There is an isomorphism $\pi: B(K, \omega) \to B(K, \omega)'$ which takes $[A, I] \mapsto [A, I, 1]$.*

Proof. Let $[A, I] \in B(K, \omega)$. If k is a positive integer, there is a canonical "extension" $\tilde{I}$ of I to the $k \times k$ matrices $\tilde{A} = A(k \times k)$, given by "conjugate transpose", $(a_{ij})^{\tilde{I}} = {}^t(a_{ij}^I)$. It is easy to check that $\tilde{I}$ has the same type as I, so $[\tilde{A}, \tilde{I}] = [A, I]$. Now suppose that $[B, J] = [A, I]$. Since $[A] = [B]$, we can choose k and another integer ℓ so that $\tilde{A} \cong \tilde{B} = B(\ell \times \ell)$; we shall assume that $\tilde{A}$ and $\tilde{B}$ are equal. By the Skolem-Noether theorem, there is a unit u of $\tilde{A}$ such that $\tilde{J} = (\operatorname{inn} u) \circ \tilde{I}$ where $\operatorname{inn} u$ is the inner automorphism $a \mapsto u^{-1}au$. Then (see [Sch, Sect. 7, Chap. 8]) $u^{\tilde{I}} = u$ if $\tilde{I}$ is of the first kind (since then $\tilde{J}$ is also of the first kind and has the same type as $\tilde{I}$), and this can also be assumed if $\tilde{I}$ is of the second kind. Thus $(\tilde{B}, \tilde{J}, 1)$ is the scaling ${}^u(\tilde{A}, \tilde{I}, 1)$ and so $[\tilde{B}, \tilde{J}, 1] = [\tilde{A}, \tilde{I}, 1]$ by Lemma 4. The fact that π is well-defined now follows from:

Lemma 7. $[\tilde{A}, \tilde{I}, \tilde{\lambda}] = [A, I, \lambda]$ *for any antiautomorphism I, where $\tilde{\lambda} = \lambda E_k$ (E_k the identity matrix of degree k).*

Proof. Take $M = A(k \times 1)$. In the corresponding derived Morita equivalence, we may take $B = \tilde{A}$ operating by left multiplication on M, and $N = A(1 \times k)$ operating on M by both left and right multiplication. We let $h: M_A \to N_A$ be the map $h(m_i) = {}^t(m_i^I)$. It is straightforward to check that the resulting derived Morita equivalence yields the desired result. □

We now return to the proof of Theorem 6. It is clear that π is a homomorphism.

If $\omega \neq \mathrm{id}$, it is injective since

$$\begin{tikzcd} B(K,\omega) \arrow[rr, "\pi"] \arrow[dr, "?"'] & & B(K,\omega)' \arrow[dl, "?"] \\ & B(K) & \end{tikzcd}$$

is commutative and the forgetful map on $B(K,\omega)$ is simply the identity map. If $\omega = \mathrm{id}$, the kernel of π is certainly contained in the subgroup $([K], \pm 1)$ of order 2. Suppose $[M(m,K), I, 1] = [M(n,K), J, 1]$; By Lemma 7 we can assume that $m = n$, and so $(M(n,K), J, 1) = {}^{u}(M(n,K), I, 1)$ for some unit $u \in M(n,K)$ by Theorem 5. Then $u^{-1}u^{I}1 = 1$ so $u^{I} = u$. It is easy to see that a is fixed by I iff $u^{-1}a$ is fixed by $J = (\mathrm{inn}\, u) \circ I$. Thus type I = type J, which implies that π is injective.

To show that π is surjective, suppose that $[A, I, \lambda] \in B(K,\omega)'$. Since $\lambda\lambda^{I} = 1$, A supports an ω-involution J by [Sch, 8.2, Chap. 8], and so by Theorem 5 and the Skolem-Noether theorem, we may assume that I itself is an involution. Then $\lambda \in K^*$. If $\omega \neq \mathrm{id}$, another scaling (using Hilbert's Theorem 90) shows that we can take $\lambda = 1$, so $[A, I, \lambda] = \pi[A, I]$. Suppose $\omega = \mathrm{id}$. Then $\lambda = \pm 1$; we are finished if $\lambda = 1$ so suppose $\lambda = -1$. By Wedderburn's theorem we can assume $A = M(n, D)$ for some division algebra D. One can show, using the results in ibid, that there is an ω-involution J on D and that I differs by an inner automorphism from $(d_{ij}) \mapsto {}^{t}(d_{ij}^{J})$. Thus there is a unit u in A such that $u^{I} = -u$, and scaling by it yields $[A, I, -1] = [A, (\mathrm{inn}\, u) \circ I, 1]$ which is obviously in the image of π. □

Theorem 8.

$$B(K,\omega) = \begin{cases} {}_2B(K) \oplus \{\pm 1\} & \text{if} \quad \omega = \mathrm{id}, \\ \ker \mathrm{cor}_{K/K_0} & \text{if} \quad \omega \neq \mathrm{id}. \end{cases}$$

Proof. By a theorem of Albert, [Sch, 8.4, Chap. 8], a central simple algebra has a K-involution iff its Brauer class has order 1 or 2. The expression for $B(K, \mathrm{id})$ follows from this and the fact that $M(2,K)$, for example, has involutions of both types, namely transpose, and transpose followed by the inner automorphism with respect to $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Similarly the expression when $\omega \neq \mathrm{id}$ is another theorem of Albert – see 9.5, ibid.

2. The Schur Subgroup Over a Number Field

We begin by proving Lemma 1.

Suppose that $[A, I] \in S(K,\omega)$, say that A is the direct summand of the group algebra KG with I induced on A by the canonical ω-involution Ω of KG. There is a unique absolutely irreducible character χ of G which corresponds to A. The center of A is $K = K(\chi)$, so the values of χ lie in K. Since $\mathbb{Q}(\chi)$ is a cyclotomic extension of $\mathbb{Q}$, this means that the values of χ lie in K_c. Now consider the formula for the idempotent

$$e_\chi = \frac{n}{g} \sum_{s \in G} \chi(s^{-1})s,$$

where $n=\chi(1)$ and g is the order of G. Now Ω permutes the primitive central idempotents of KG, and since A is stable under it, it fixes e_χ. Therefore

$$e_\chi=\frac{n}{g}\sum_{s\in G}\chi(s^{-1})^{\omega}s^{-1}=\frac{n}{g}\sum_{s\in G}\chi(s^{-1})^{*\omega}s,$$

where * is complex conjugation. On comparing the expressions for e_χ, we see that the values of χ are fixed by $*\omega$, and so $\tilde{K}(\chi)=\tilde{K}$. This means that the direct summand $\tilde{A}$ of $\tilde{K}G$ which belongs to χ has center $\tilde{K}$, and so since $K\otimes\tilde{K}G=KG$, it follows at once that $K\otimes\tilde{A}=A$ since $K\otimes\tilde{A}$ is simple. Therefore im $S(K,\omega)\subseteqq K\otimes S(\tilde{K})$.

We now show the reverse inclusion. Let $\tilde{\Omega}$ be the canonical ω-involution of the group algebra $\tilde{K}G$. Because $*\omega=\mathrm{id}$ on $\tilde{K}$, ω acts on $\tilde{K}$ via complex conjugation. This implies that $\tilde{\Omega}$ leaves invariant all simple factors of $\tilde{K}G$. (Indeed the proof is almost identical to Theorem 13.3, Chap. 8, [Sch]: If T is the algebra trace of $\tilde{K}G$, then it is easy to see that $T(xy^{\tilde{\Omega}})$ is a positive definite hermitian form on $\tilde{K}G$ (with G as an orthogonal basis). Thus $T(xx^{\tilde{\Omega}})>0$, which implies that every simple factor is $\tilde{\Omega}$-invariant.) Thus if $\tilde{A}$ is a central simple factor of $\tilde{K}G$, it is clear that $K\otimes\tilde{A}$ is a central simple factor of KG and is invariant under the canonical ω-involution of KG, and so Lemma 1 is proved. □

Theorem 9. *Let K be a formally real field. If $\beta\in S(K)$ is split in at least one real closure of K, then it is split in all real closures of K. $S(K,\mathrm{id})$ consists of all*

$$(\beta,\varepsilon)\in S(K)\times\{\pm 1\}$$

with $\varepsilon=1$ iff β is split at all real closures.

Proof. As in the previous proof, any simple component of a group algebra KG is stable under the canonical K-involution (of KG). Suppose for the moment that K is real closed. It is easy to check that Frobenius' theorem on simple algebras over $\mathbb{R}$ [Sch, Theorem 6.4, Chap. 8] and the Frobenius-Schur theory of representations over $\mathbb{R}$ [S, 13.2] hold more generally for real closed fields. Therefore if $(\beta,\varepsilon)\in S(K,\mathrm{id})$, ε must be 1 if β is split and must be -1 if β is non-split (in which case β is the class of the unique non-commutative central division algebra over K, the quaternion algebra $(-1,-1)$). Now suppose again that K is merely formally real, and that $(\beta,\varepsilon)\in S(K,\mathrm{id})$. If $\hat{K}$ is a real closure of K, then $(\hat{K}\otimes\beta,\varepsilon)\in S(\hat{K},\mathrm{id})$ and so ε is 1 if β splits in $\hat{K}$ and is -1 otherwise. Since this holds for any real closure, the first statement of the theorem follows, and the second is a consequence of this and Lemma 1 and the fact that $K\otimes S(\tilde{K})=S(K)$. □

Lemma 10. *Let k be a finite extension of $\mathbb{Q}$, and let K/k be a finite extension of even degree. Then there exists a finite prime $\mathfrak{p}$ of k and a prime $\mathfrak{P}$ of K lying over $\mathfrak{p}$ with the property that the local extension $K_{\mathfrak{P}}/k_{\mathfrak{p}}$ also has even degree.*

Proof. Let L be the normal closure of K/k, let $\mathscr{G}$ be the Galois group of L/k and $\mathscr{H}$ that of L/K. Choose $\tau\in\mathscr{G}-\mathscr{H}$ such that $\tau^2\in\mathscr{H}$, for example by considering a 2-Sylow subgroup of $\mathscr{H}$ contained in a 2-Sylow subgroup of $\mathscr{G}$. By the Tchebotarev density theorem [CF, p. 227], there is a prime $\mathfrak{P}'$ of L which is unramified over k and whose Frobenius automorphism is τ. Let $\mathfrak{P}$ and $\mathfrak{p}$ be resp. the primes of K and k lying below $\mathfrak{P}'$. Then the decomposition group of $\mathfrak{P}'/\mathfrak{p}$ is

$$\mathscr{Z}(\mathfrak{P}'/\mathfrak{p})=\langle\tau\rangle=\mathrm{Gal}(L_{\mathfrak{P}'}/k_{\mathfrak{p}}).$$

Now $\tau \notin \mathrm{Gal}(L_{\mathfrak{P}'}/K_{\mathfrak{P}})$ since the latter group is a subgroup of $\mathscr{H}$ but

$$\tau^2 \in \mathscr{G}(L/K) \cap \mathscr{Z}(\mathfrak{P}'/\mathfrak{p}) = \mathscr{Z}(\mathfrak{P}'/\mathfrak{P}) = \mathrm{Gal}(L_{\mathfrak{P}'}/K_{\mathfrak{P}}).$$

It follows at once that $K_{\mathfrak{P}}/k_{\mathfrak{p}}$ has even degree. □

We can now prove Theorem 2. Parts (i) and (iii) follow from Lemma 1 and Theorem 9 respectively. Therefore we assume that K is totally imaginary and that $\omega = \mathrm{id}$. We must show that $([\mathrm{K}], -1)$ is in $S(K, \mathrm{id})$.

Since $K/\mathbb{Q}$ has even degree, there is a finite prime $\mathfrak{P}$ in K such that $K_{\mathfrak{P}}/\mathbb{Q}_p$ also has even degree by Lemma 10. Let Q be the quaternion algebra over $\mathbb{Q}$ with non-trivial Hasse invariants at p and ∞, and invariant 0 at the other primes. By a theorem of M. Benard and K.L. Fields [Y, Theorem 7.2], $S(\mathbb{Q})$ consists of the quaternion algebras and so $[Q] \in S(\mathbb{Q})$. Since $\mathbb{Q}$ is formally real, every simple component of a rational group algebra $\mathbb{Q}G$ is invariant under the canonical involution, so $([Q], \varepsilon) \in S(\mathbb{Q}, \mathrm{id})$ for a suitable choice of $\varepsilon = \pm 1$. By Theorem 9, $\varepsilon = -1$ since $\mathbb{R} \otimes Q$ is non-split. Thus $([K \otimes Q], -1) \in S(K, \mathrm{id})$. But $K \otimes Q$ is split at $\mathfrak{P}$, and hence at all other primes lying over p by the "uniform distribution theorem" of Benard and Schacher [Y, Theorem 6.1]. Since K is totally imaginary, this means that $K \otimes Q$ is split, so $([K], -1) \in S(K, \mathrm{id})$ as desired. □

3. The Schur Subgroup Over a $\mathfrak{p}$-Adic Field

We now assume that K is a finite extension of $\mathbb{Q}_p$ for some p, and we shall prove Theorem 3 in several stages.

Case 1: $\omega \neq \mathrm{id}$. For *any* finite extension K/K_0 of local fields, the corestriction map cor_{K/K_0} is injective [CL, Proposition 1, Chap. XI, and Theorem 1, Chap. XIII], and so $S(K, \omega) = B(K, \omega) = 1$ by Theorem 8.

From now on we assume that $\omega = \mathrm{id}$. We first show that the kernel of the forgetful map on $S(K, \mathrm{id})$ is ± 1. Let Q be a rational quaternion algebra which is split at p but not split at ∞. Then because $S(\mathbb{Q}) = {}_2B(\mathbb{Q})$ and every direct summand of a rational group algebra $\mathbb{Q}G$ is stable under the canonical $\mathbb{Q}$-involution, $([Q], \varepsilon) \in S(\mathbb{Q}, \mathrm{id})$ for some choice of $\varepsilon = \pm 1$. By the usual argument of extending to $\mathbb{R}$, we see that $\varepsilon = -1$. Now extend to K to show that $([K], -1) \in S(K, \mathrm{id})$, as desired.

We can assume for the rest of the proof that ${}_2S(K) = \pm 1$ since ${}_2S(K)$ is either trivial or ± 1 (recall that $B(K) = \mathbb{Q}/\mathbb{Z}$ – cf. [CL, Proposition 6, Chap. XIII]). We shall have to construct "quadratic Schur algebras", that is central simple algebras which are direct summands of group algebras (over K) and which are stable under the canonical K-involution of the group algebra – or what is the same, which are the images of KG under an irreducible K-representation of the finite group G and which admit an involution of the first kind which inverts the images of the elements of G. This is done by the use of a crossed-product algebra $A = (K(\zeta)/K, z)$ (see [MO, Sect. 29]) using a cocycle $z \in Z^2(\mathrm{Gal}(K(\zeta)/K), \mu(K(\zeta))$, where ζ is a suitable root of unity and $\mu(K(\zeta))$ is the group of roots of unity of $K(\zeta)$. Thus A has a distinguished basis $\{u_\sigma : \sigma \in \mathrm{Gal}(K(\zeta)/K)\}$ over $K(\zeta)$ with multiplication defined by

$$(a_\sigma u_\sigma)(b_\tau u_\tau) = a_\sigma b_\tau^\sigma z(\sigma, \tau) u_{\sigma\tau}$$

for any a_σ and b_τ in $K(\zeta)$.

Lemma 11. *Suppose that the values of z are actually ± 1 and that there is $\iota \in \mathrm{Gal}(K(\zeta)/K)$ such that $\zeta^{\iota} = \zeta^{-1}$. Then $A = (K(\zeta)/K, z)$ is a quadratic Schur algebra.*

Proof. We can assume that z is normalized, i.e. $z(\sigma, \tau) = 1$ if either σ or τ is the identity. A is easily seen to be a "Schur algebra" for the group $G = \bigcup_{\sigma} \langle \pm \zeta \rangle u_\sigma$ since G spans A over K (the representation space is of course any simple A-module). We must show that there is a K-involution on A which inverts the elements of G. Consider the K-linear map I on A which, for each σ, takes $a_\sigma u_\sigma$ to $a_\sigma^{\iota\sigma^{-1}} u_\sigma^{-1}$ ($a_\sigma \in K(\zeta)$). A straightforward calculation shows that I has the desired properties. □

Lemma 12. *If p_1 is an odd prime such that $K(\zeta_{p_1})/K$ is a Galois extension of even degree, then there is an automorphism ι of $K(\zeta_{p_1})/K$ which inverts ζ_{p_1}.*

Proof. Note that $\mathbb{Q}(\zeta_{p_1})/\mathbb{Q}$ has even degree and is cyclic, so the unique element of order 2 in its Galois group is complex conjugation, which inverts ζ_{p_1}. The restriction of the Galois group of $K(\zeta_{p_1})/K$ to $\mathbb{Q}(\zeta_{p_1})$ is injective and so the image contains complex conjugation, whence the lemma. □

A standard technique for constructing crossed-product algebras is to use a *cyclic* extension $K(\zeta)/K$; in this case one can assume that the algebra has the form

$$A = \sum K(\zeta) u_\sigma^i, \qquad 0 \leqq i < (K(\zeta):K) = n,$$

where $u_\sigma^n = a \in K^*$; in particular the u_σ^i form a distinguished basis. A is often denoted by $(K(\zeta)/K, \sigma, a)$ or simply $(K(\zeta)/K, a)$. Furthermore A is split iff a is a norm in the extension $K(\zeta)/K$. See [MO, 30.4], for example. Less well known is the fact that there is a similar construction, due to Yamada, for any finite cyclotomic extension – see [Y, Chap. 2]. We shall use his construction in the bicyclic case only:

Yamada's Lemma. *Suppose that ζ is a root of unity and that $K(\zeta)/K$ has Galois group the direct product of two cyclic groups $\langle \varrho \rangle$ and $\langle \sigma \rangle$ of finite orders r and s resp. Let a, b, and c be roots of unity in $K(\zeta)$ satisfying*

$$a^{\varrho - 1} = b^{\sigma - 1} = 1, \qquad a^{\sigma - 1} = N_\varrho c, \qquad b^{\varrho - 1} = N_\sigma c^{-1},$$

where, for example, $N_\varrho c = c^{1 + \varrho + \varrho^2 + \dots + \varrho^{r-1}}$. Then there is a crossed-product algebra ("bicyclic algebra") $A = (K(\zeta)/K,\ a,\ b,\ c)$ which has a distinguished basis $u_\varrho^i u_\sigma^j$, $0 \leqq i < r$, $0 \leqq j < s$, with the property that

$$u_\varrho^r = a, \qquad u_\sigma^s = b, \qquad u_\sigma u_\varrho = c u_\varrho u_\sigma.$$

If λ and μ are roots of unity in $K(\zeta)^*$, and if $v_\varrho := \lambda u_\varrho$ and $v_\sigma := \mu u_\sigma$, then the elements $v_\varrho^i v_\sigma^j$ form a distinguished basis for the bicyclic algebra $A = (K(\zeta)/K, a', b', c')$ where

$$a' = (N_\varrho \lambda) a, \qquad b' = (N_\sigma \mu) b, \qquad c' = \lambda^{\sigma - 1} (\mu^{\varrho - 1})^{-1} c. \quad \square$$

We now return to the proof of Theorem 3.

Case 2: $\omega = \mathrm{id}$, $K/\mathbb{Q}_p$ abelian, p odd, and $\mu(K)_2 = \pm 1$. By a theorem of Janusz, [J], $S(K)$ is generated by the class of a cyclic algebra $(K(\zeta_p)/K, \zeta)$ where ζ generates the group of roots of unity in K with order prime to p. Since the Brauer class of a cyclic algebra $(L/K, a)$ is multiplicative in a, it is easy to see that the class of $(K(\zeta_p)/K, -1)$

is the non-trivial element of ${}_2S(K)$. By the Yamada-Fontaine theorem [Y, 4.4′, 4.5], $S(K)$ is a cyclic group of order $(p-1)/e_0$ where e_0 is the tame ramification index of $K/\mathbb{Q}_p$. Since $\mathbb{Q}_p(\zeta_p)/\mathbb{Q}_p$ has tame index $p-1$, $S(K(\zeta_p))$ must be trivial. Thus $K(\zeta_p)/K$ has even degree since the scalar extension map $B(K)\to B(K(\zeta_p))$ is multiplication by $(K(\zeta_p):K)$ when viewed as a map $\mathbb{Q}/\mathbb{Z}\to\mathbb{Q}/\mathbb{Z}$. This case then follows from Lemmas 11 and 12.

Case 3: $\omega=\mathrm{id}$, $K/\mathbb{Q}_p$ abelian, p odd, $|\mu(K)_2|>2$, and $K(\zeta_p)/K$ a ramified quadratic extension. Suppose $\mu(K)_2$ has order 2^h $(h\geqq 2)$. Let ζ be a primitive 2^{h+1}-root of unity; then $K(\zeta)$ is an unramified quadratic extension of K and so is disjoint from $K(\zeta_p)$. Thus $K(\zeta,\zeta_p)=K(\zeta\zeta_p)$ has Galois group over K generated by elements ϱ and σ of order 2 where the fixed field of ϱ is $K(\zeta)$ and that of σ is $K(\zeta_p)$. By Yamada's lemma, there is a bicyclic algebra $A=(K(\zeta\zeta_p)/K, 1, 1, -1)$. Suppose that the residue class field of K has q elements (so q is a power of p). The gcd of $q+1$ and $q-1$ is 2, so $q+1$ is not a power of 2 since the fact that K contains the fourth roots of unity implies that $q-1$ is divisible by 4. It follows that there is an odd prime p_1 which divides q^2-1 but not $q-1$. It is easy to see then that $K(\zeta)=K(\zeta_{p_1})$ and $K(\zeta,\zeta_p)=K(\zeta_{p_1p})$.

It follows from Lemma 12 that $\varrho\sigma$ inverts ζ_{p_1p} and so A is a quadratic Schur algebra by Lemma 11. The proof for case 3 will be finished by showing that A is non-split.

We define a new distinguished basis of A by taking $v_\varrho:=\zeta u_\varrho$ (with ζ a 2^{h+1}-root of unity as before) and $v_\sigma=u_\sigma$. Since $N_\varrho\zeta=\zeta^2$ and $\zeta^{\sigma-1}=-1$, $A=(K(\zeta_{p_1p})/K, \zeta^2, 1, 1)$ by Yamada's Lemma. In particular v_ϱ and v_σ commute and one sees easily that

$$A\cong(K(\zeta_p)/K,\zeta^2)\otimes(K(\zeta_{p_1})/K,1).$$

The second factor is of course split, and so this case will follow if we show that ζ^2 is not a norm in $K(\zeta_p)/K$. Suppose that ζ^2 *is* a norm, say $\zeta^2=N\alpha$ with $\alpha\in K(\zeta_p)$. Certainly α must be a unit; we can write $\alpha=\zeta'\beta$ where ζ' is a $(q-1)^{st}$ root of unity and β is a 1-unit, i.e. is $\equiv 1 \bmod \mathfrak{P}$ where $\mathfrak{P}$ is the maximal ideal of the ring of integers of $K(\zeta_p)$. Now $N\beta$ is also a 1-unit, and is also a root of unity since both of $N\alpha$ and $N\zeta'$ are, and so must be a p-power root of unity. Since ζ^2 is a 2-power root of unity, we can therefore assume that $\beta=1$, and that ζ' is also a 2-power root of unity. Since $K(\zeta_p)/K$ is totally ramified, $\mu(K(\zeta_p))_2=\mu(K)_2$ and so ζ' is a power of ζ^2. This is impossible since $N\zeta'=\zeta'^2=\zeta^2$. This finishes the proof of case 3.

Case 4: $\omega=\mathrm{id}$, $K/\mathbb{Q}_p$ abelian, p odd, and $|\mu(K)_2|>2$. Let $K(\zeta_p)_u$ be the maximal unramified subextension of $K(\zeta_p)/K$. Since $S(K)_2\neq 1$, $K(\zeta_p)/K(\zeta_p)_u$ must be tamely ramified of even degree; let L be the unique intermediate field of which $K(\zeta_p)$ is a quadratic extension. We can therefore apply Case 3 (and its proof) to find an odd prime $p_1\neq p$ and a cocycle $z\in Z^2(\mathrm{Gal}(K(\zeta_{p_1p})/L), \pm 1)$ such that the corresponding crossed-product algebra is non-split and such that there is an $\iota\in\mathrm{Gal}(K(\zeta_{p_1p})/L)$ which inverts ζ_{p_1p}. Let $z'\in Z^2(\mathrm{Gal}(K(\zeta_{p_1p})/K, \pm 1)$ be the corestriction of z. As mentioned earlier, cor is injective on the Brauer group over a local field, and so the crossed-product algebra corresponding to z' is non-split. Therefore this case follows from Lemma 11.

Case 5: $\omega=\mathrm{id}$, $K/\mathbb{Q}_p$ abelian, and $p=2$. It is known in this case that the non-trivial Brauer class in $S(K)$ $(={}_2S(K))$ is represented by a bicyclic algebra of the following form (see [R], for example): Let h be the smallest integer $\geqq 2$ with the property that there is an odd integer m such that $L := \mathbb{Q}_2(\zeta_{2^h}, \zeta_m)$ contains K; we can assume that the residue class degree f of L/K is $\equiv 0 \bmod 2^h$. The Galois group $\mathcal{G}$ of L/K is the bicyclic group $\langle\sigma_1\rangle \times \langle\sigma_2\rangle$ where

(i) σ_1 is of order 2, inverts $\zeta = \zeta_{2^h}$, and has fixed field $K(\zeta_m)$.

(ii) σ_2 is of order f and has fixed field $K(\zeta_4)$.

Then the bicyclic algebra is $A := (L/K, 1, 1, \zeta)$.

As was indicated in [R], one can replace m by any odd multiple m' of m. We shall choose m' in such a way that $\mathbb{Q}_2(\zeta_{2^h}, \zeta_{m'}) = \mathbb{Q}_2(\zeta_{2^h}, \zeta_n)$ for some odd positive integer n which is relatively prime to the order of $\mu(K)$. The following lemma is useful in this regard:

Lemma. *If q is a power of 2, $\mathbb{F}_{q^2}$ is generated over $\mathbb{F}_2$ by a primitive $(q+1)^{st}$ root of 1.*

Proof. Let $\mathbb{F}$ be a proper subfield of $\mathbb{F}_{q^2}$, say with q' elements. Then $q^2 = q'^r$ for some $r \geqq 2$, so $q' \leqq q$. Therefore $q'-1$ is not divisible by $q+1$, whence the lemma. □

Now let q be the number of elements in the residue class field of $\mathbb{Q}_2(\zeta_{2^h}, \zeta_m)$; we can assume $m = q+1$. Clearly $q+1$ is relatively prime to $q-1$, hence a fortiori to $|\mu(K)|$ as well. By the lemma we can choose $m' = q^2 - 1$ and $n = q+1$. Suppose that m has been replaced by m'. Let $f' = \frac{1}{2}f$ and $K' =$ the fixed field of $\sigma_2^{f'}$. If p_1 is any prime divisor of $q+1$, $K'(\zeta_{p_1}) = \mathbb{Q}_2(\zeta_{2^h}, \zeta_m)$ since p_1 does not divide $q-1$. Therefore it follows from Lemma 12 that $\sigma_2^{f'}$ inverts a $(q+1)^{st}$ root of unity ζ'; moreover $2^h | f$ implies that $\sigma_2^{f'}\zeta = \zeta$.

We shall now show that A is a quadratic Schur algebra. Let $\{u_1^i u_2^j\}$ be the distinguished basis of A over L. Let $v_1 = u_1^{-1}$ and $v_2 = u_2^{-1}$. Under the multiplication $\cdot$ of the opposite algebra A^0, $v_1 \cdot v_2 \cdot v_1^{-1} \cdot v_2^{-1} = \zeta^{-1}$, $v_1^2 = 1 = v_2^f$, and $v_i \cdot \lambda \cdot v_i^{-1} = \sigma_i(\lambda)$ for $i = 1, 2$. Thus $A^0 = (L/K, 1, 1, \zeta^{-1})$. Let $J : A \to A^0$ be the additive map defined by $J(\lambda u_1^i u_2^j) = \sigma_1 \sigma_2^{f'}(\lambda) \cdot v_1^i \cdot v_2^j$. Then J is an isomorphism of K-algebras, i.e. J is a K-involution, and it inverts the elements of the group $\langle\zeta, \zeta', u_1, u_2\rangle$. This finishes the proof of Case 5.

Case 6. $K/\mathbb{Q}_p$ an arbitrary finite extension.

Lemma. *Let K_1 be the maximal abelian extension of $\mathbb{Q}_p$ contained in K. Then*

$$\mathrm{im}(S(K, \mathrm{id}) \to S(K)) = K \otimes \mathrm{im}(S(K_1, \mathrm{id}) \to S(K_1)).$$

Proof. The product of the restriction maps

$$\mathrm{Gal}(K_c K_1/\mathbb{Q}_p) \to \mathrm{Gal}(K_c/\mathbb{Q}) \times \mathrm{Gal}(K_1/\mathbb{Q}_p)$$

is clearly injective, so the subextension $K_c K_1/\mathbb{Q}_p$ of $K/\mathbb{Q}_p$ is abelian. Therefore $K_c K_1 = K_1$, i.e. $K_c \subseteqq K_1$, and so $K_{1c} = K_c$ and $\tilde{K}_1 = \tilde{K}$. It follows from Lemma 1 that

$$\mathrm{im}(S(K, \mathrm{id}) \to S(K)) = K \otimes S(\tilde{K}) = K \otimes (\mathrm{im}(S(K_1, \mathrm{id}) \to S(K_1)). \quad \square$$

By the earlier cases we know that $\mathrm{im}(S(K_1, \mathrm{id}) \to S(K_1)) = {}_2S(K_1)$ and so we get

$$\mathrm{im}(S(K, \mathrm{id}) \to S(K)) = K \otimes {}_2S(K_1).$$

Moreover $S(K) = K \otimes S(K_1)$ by [Y, Proposition 4.6], so that ${}_2S(K) \neq 1$ implies that ${}_2S(K_1) \neq 1$ since $S(K_1)$ is a finite group. Theorem 3 follows from this and the fact that the scalar extension map $B(K_1) \to B(K)$, when these two groups are identified with $\mathbb{Q}/\mathbb{Z}$, is multiplication by the degree of K/K_1. $\square$

Acknowledgement. I am grateful to H. Kisilevsky and J. Labute for providing the proof of Lemma 10.

References

[CF] Cassels, J.W.S., Frohlich, A.: Algebraic number theory. Washington: Thompson 1967

[CL] Serre, J.-P.: Corps locaux. Act. Sci. et Ind. 1296. Paris: Hermann 1962

[FM] Frohlich, A., McEvett, A.M.: Forms over rings with involution. J. Algebra **12**, 79–104 (1969)

[H] Hahn, A.: A hermitian Morita theorem for algebras with anti-structure. J. Algebra **93**, 215–235 (1985)

[HTW] Hambleton, I., Taylor, L., Williams, E.B.: An introduction to the maps between surgery obstruction groups. In: Algebraic topology, Aarhus 1982 (Lecture Notes in Math., Vol. 1051, pp. 49–127) Berlin Heidelberg New York: Springer 1984

[MO] Reiner, I.: Maximal Orders, L.M.S. Monographs 5. London: Academic Press 1975

[R] Riehm, C.: The Schur subgroup of the Brauer group of a local field. L'Enseignement Mathématique (to appear)

[Sch] Scharlau, W.: Quadratic and hermitian forms. Grundl. Math. Wiss. 270. Berlin Heidelberg New York: Springer 1985

[S] Serre, J.-P.: Linear representations of finite groups, Graduate Texts in Mathematics 42. Berlin Heidelberg New York: Springer 1977

[Y] Yamada, T.: The Schur subgroup of the Brauer group. Lecture Notes in Math. 397. Berlin Heidelberg New York: Springer 1974

Received June 1, 1988

Math. Ann. 283, 491–497 (1989)

Mathematische Annalen

L'ensemble des algèbres de Lie algébriques n'est pas Zariski-dense dans la variété des algèbres de Lie de dimension $M \geqq 9$

R. Carles

Université de Poitiers, Mathematiques, 40, Avenue du Recteur Pineau, F-86022 Poitiers Cedex, France

Introduction

Soient L_m la variété des lois d'algèbres de Lie de dimension m sur le corps $\mathbb{C}$ des nombres complexes, R_m la sous-variété des lois résolubles et N_m celle des lois nilpotentes sur lesquelles opère le groupe $GL_m(\mathbb{C})$ par l'action canonique habituelle. Les notions topologiques sont relatives à la topologie de Zariski et $\bar{X}$ désigne l'adhérence de X. Dans [2] et [3] avait été mis en évidence le rôle important joué par les algèbres de Lie algébriques dans la description de L_m et de R_m et qui peut se formuler ainsi:

a) toute algèbre de Lie rigide (c.a.d. dont l'orbite est ouverte) est algébrique;

b) l'ensemble des algèbres de Lie algébriques (respectivement algébriques résolubles) est dense dans L_m (respectivement R_m) pour $m \leqq 7$.

L'assertion (b) avait été prouvée [2] pour le cas des algèbres de Lie décomposables au prix d'une démonstration essentiellement technique cas par cas sur chaque type de la classification de Vergne [8]. Il en découlait une construction aisée des composantes de L_m et de R_m pour $m \leqq 7$. Le cas algébrique étudié dans ce travail vérifie (b) sans difficulté.

Dans cet article est démontré principalement que l'assertion (b) est fausse pour $m \geqq 9$, résultat déjà conjecturé dans [2]. Pour cela on construit une nouvelle famille de composantes irréductibles de L_m et de R_m si $m \geqq 9$ pour lesquelles un ouvert non vide ne contient aucune algèbre de Lie algébrique (ni décomposable). Ces composantes sont engendrées par des déformations liées à l'existence du centre de certaines algèbres de Lie décomposables. La question reste ouverte pour $m = 8$.

1. Ensembles d'algèbres de Lie décomposables, algébriques

Si $\mathfrak{g}$ est une algèbre de Lie de dimension finie, on désigne par $\Delta(\mathfrak{g})$ l'algèbre de Lie de ses dérivations, $Z(\mathfrak{g})$ son centre, $\mathfrak{r}(\mathfrak{g}) = \mathfrak{r}$ son radical et $\mathfrak{n}(\mathfrak{g}) = \mathfrak{n}$ son plus grand idéal nilpotent. Un tore sur $\mathfrak{g}$ est un espace vectoriel constitué de dérivations semi-simples de $\mathfrak{g}$ qui commutent. L'algèbre $\mathfrak{g}$ est dite décomposable (ou presque algébrique) si elle admet une décomposition (dite normale) $\mathfrak{u} \oplus \mathfrak{s} \oplus \mathfrak{n}$ de radical

$\mathfrak{u}\oplus\mathfrak{n}$, avec $\mathfrak{s}$ une sous-algèbre de Levi et $\mathfrak{u}$ une sous-algèbre abélienne qui commute avec $\mathfrak{s}$ telle que $\mathrm{ad}_{\mathfrak{g}}\mathfrak{u}$ soit un tore sur $\mathfrak{g}$. Les lois décomposables $\mathfrak{g}$ d'idéal $\mathfrak{s}\oplus\mathfrak{n}(\mathfrak{g})$ fixé s'obtiennent à isomorphisme près comme le produit semi-direct $\tau\oplus\mathfrak{s}\oplus\mathfrak{n}$ où τ est le sous-espace d'un tore extérieur maximal T_e sur $\mathfrak{s}\oplus\mathfrak{n}$ (un tore est dit extérieur s'il ne contient aucune dérivation intérieure non nulle [2]). Si sa dimension r est fixée, le tore τ varie dans la Grassmannienne $\mathrm{Gr}_r(T_e)$; l'ensemble des lois obtenues a pour orbite sous l'action canonique de $GL_m(\mathbb{C})$ un ensemble irréductible et constructible (c.a.d. réunion finie de localement fermés) noté $L_m^0(\mathfrak{s}\oplus\mathfrak{n})$ et étudié dans [2]. Si $\mathfrak{s}$ est nul, $L_m^0(\mathfrak{n})$ est contenu dans R_m et l'on désigne encore cet espace par $R_m^0(\mathfrak{n})$.

Une algèbre de Lie algébrique est une algèbre de Lie décomposable dont le tore $\mathrm{ad}_{\mathfrak{g}}\mathfrak{u}$ est algébrique (c.a.d. une algèbre de Lie linéaire algébrique au sens de Chevalley). Un tore algébrique est caractérisé par un ensemble de poids α qui s'écrivent $\sum_{i=1}^{r} c_i\alpha_i$ sur une base de poids α_i $(1\leqq i\leqq r)$ avec $c_i\in\mathbb{Q}$ $(1\leqq i\leqq r)$. L'algèbre de Lie des dérivations de $\mathfrak{s}\oplus\mathfrak{n}$ est algébrique, chaque tore maximal est algébrique et l'on peut choisir T_e algébrique: les tores algébriques constituent alors une partie dense de $\mathrm{Gr}_r(T_e)$. Ces tores donnent une famille d'algèbres de Lie algébriques dense dans $L_m^0(\mathfrak{s}\oplus\mathfrak{n})$. Une généralisation de ces résultats conduit à la proposition suivante:

(1.1) **Proposition.** *L'ensemble L_m^0 (respectivement R_m^0) des algèbres de Lie (respectivement résolubles) décomposables est constructible dans L_m (respectivement R_m) pour tout m. L'ensemble des algèbres de Lie (respectivement résolubles) algébriques est dense dans L_m^0 (respectivement R_m^0) pour tout m.*

Preuve. Les lois d'algèbres de Lie dont le radical est nilpotent forment une partie constructible X_n de L_n à cause de la semi-continuité des applications qui à $\mathfrak{g}$ associent les dimensions de $\mathfrak{n}(\mathfrak{g})$ et de $\mathfrak{r}(\mathfrak{g})$. Les classes de conjugaison des tores maximaux sur ces algèbres sous l'action adjointe du groupe $GL_n(\mathbb{C})$ sont en nombre fini d'après [1, 6] et soient T_i $(0\leqq i\leqq k)$ des représentants de ces classes. Soit $X_n(T)$ l'ensemble des lois qui admettent à isomorphisme près T comme tore maximal. Chaque $X_n(T_i)$ est l'orbite de l'espace constructible $X_n^{T_i}$ des lois T_i-invariantes privé des orbites des $X_n^{T_j}$ $(j\neq i)$ pour lesquelles T_j contient T_i (à conjugaison près): chaque $X_n(T_i)$ $(0\leqq i\leqq k)$ est constructible, non irréductible en général et l'on a $X_n=\coprod_{i=0}^{k} X_n(T_i)$. Par exemple $X_n(T_0)$ pour $T_0=(0)$ est l'ensemble des algèbres de Lie caractéristiquement nilpotentes de dimension n: il est constructible non vide pour tout $n\geqq 7$ [4].

Les lois décomposables $\mathfrak{g}$ de dimension m dont la partie $\mathfrak{s}\oplus\mathfrak{n}$ appartient à $X_n(T)$ sont (à isomorphisme près) le produit semi-direct $\tau\oplus\mathfrak{s}\oplus\mathfrak{n}$ où τ varie dans la grassmannienne $\mathrm{Gr}_{m-n}(T_e)$ (on peut pendre un même tore extérieur maximum algébrique $T_e\subset T$ pour toutes ces algèbres). Cet espace noté $L_m^0(X_n(T))$ est constructible, irréductible si $X_n(T)$ l'est et l'on peut calculer sa dimension: même preuve que la proposition 1.3.(1) de [2]. L'espace L_m^0 est réunion finie des $L_m^0(X_n(T))$ où n varie entre $\frac{m}{2}$ et m et T sur un nombre fini de classes, il est donc encore constructible.

Les lois algébriques s'obtiennent de la même façon en prenant les tores algébriques dans chaque $\mathrm{Gr}_{m-n}(T_e)$; on utilise le fait qu'ils constituent une partie dense.

Pour construire R_m^0 on remplace X_n par $X_n \cap N_n$. CQFD.

(1.2) *Remarque*. En prenant l'adhérence de certains des espaces $L_m^0(X)$ et $R_m^0(X)$ où X est une composante irréductible d'un $X_n(T)$ on obtient les composantes de L_m ou de R_m qui contiennent un ouvert non vide de lois décomposables. Ces espaces sont la généralisation naturelle des $L_m^0(\mathfrak{s}\oplus\mathfrak{n})$ étudiés dans [2].

On déduit l'assertion (b) de [2] et (1.1):

(1.3) **Corollaire.** *Les lois algébriques constituent une partie dense dans L_m et dans R_m pour tout $m \leqq 7$.*

(1.4) *Remarque*. L'ensemble des lois algébriques de L_m (ou R_m) n'est pas constructible pour $m \geqq 3$. Il contient le sous-ensemble constructible formé des lois dont le radical est nilpotent, l'ouvert des algèbres de Lie rigides (condition a), l'ouvert des algèbres de Lie $\mathfrak{g}$ telles que $\dim \Delta(\mathfrak{g}) \leqq \dim \mathfrak{g}$ [3].

2. Déformations d'une algèbre de Lie liées à l'existence d'un centre

Le deuxième groupe de cohomologie adjointe d'une algèbre de Lie décomposable $\mathfrak{g}=\mathfrak{s}\oplus\mathfrak{u}\oplus\mathfrak{n}$ peut se calculer à l'aide d'un théorème de factorisation de Hochschild et Serre [7, 2]:

$$H^2(\mathfrak{g},\mathfrak{g}) \simeq H^2(\mathfrak{s}\oplus\mathfrak{u},\mathbb{C})\otimes Z(\mathfrak{g}) + Z(\mathfrak{s}\oplus\mathfrak{u})\otimes(H^1(\mathfrak{n},\mathfrak{n})^{\mathfrak{s}+\mathfrak{u}}/Z(\mathfrak{s}+\mathfrak{u})) + H^2(\mathfrak{n},\mathfrak{n})^{\mathfrak{s}+\mathfrak{u}}$$

Les deux derniers termes sont constitués de classes dont certaines s'intègrent en déformations déjà rencontrées. Le premier terme s'identifie à $C^2(\mathfrak{u}, Z(\mathfrak{g}))$ et se plonge dans l'espace des 2-cocycles $Z^2(\mathfrak{g},\mathfrak{g})$; il s'interprète comme l'ensemble des déformations infinitésimales suivantes:

si $\varphi \in C^2(\mathfrak{u}, Z(\mathfrak{g}))$, soit $\mathfrak{g}_{t\varphi}$ la loi définie par les crochets

$$[x,y]_t = t\varphi(x,y) \quad \text{si} \quad (x,y)\in\mathfrak{u}^2,$$

$$[x,y]_t = [x,y] \quad \text{si} \quad (x,y)\in(\mathfrak{s}+\mathfrak{n})\times\mathfrak{g}.$$

On vérifie que $\mathfrak{g}_{t\varphi}$ est une déformation non triviale si $\varphi \neq 0$. En particulier l'algèbre de Lie $\mathfrak{g}_\varphi$ n'est pas décomposable si $\varphi \neq 0$.

On en déduit aisément le lemme suivant:

(2.1) **Lemme.** *L'orbite $\Sigma(\mathfrak{g})$ sous l'action de $GL_m(\mathbb{C})$ de l'ensemble des $\mathfrak{g}_\varphi$ avec $\varphi\in C^2(\mathfrak{u}, Z(\mathfrak{g}))$ est une partie irréductible, constructible de dimension égale à*

$$m^2 - \dim \Delta(\mathfrak{g}) + \frac{r(r-1)}{2}\dim Z(\mathfrak{g}) \qquad (r=\dim\mathfrak{u}).$$

Le sous-ensemble des algèbres de Lie décomposables de $\Sigma(\mathfrak{g})$ est l'orbite de $\mathfrak{g}$.

L'espace $Z^2(\mathfrak{g}_\varphi, \mathfrak{g}_\varphi)$ est le tangent de Zariski au point $\mathfrak{g}_\varphi$ de L_m. Si sa dimension pour $\varphi\in C^2(\mathfrak{u}, Z(\mathfrak{g}))$ est égale à celle de $\Sigma(\mathfrak{g})$, alors l'adhérence de $\Sigma(\mathfrak{g})$ est une composante irréductible et le point $\mathfrak{g}_\varphi$ est simple pour le schéma défini par les relations de Jacobi [2]. Cette dimension pourra être calculée à l'aide du lemme suivant:

(2.2) **Lemme.** *Si* $R = ad_{\mathfrak{g}}(\mathfrak{s} \oplus \mathfrak{u})$ *on a:*

$$\dim Z^2(\mathfrak{g}_\varphi, \mathfrak{g}_\varphi) = \dim Z^2(\mathfrak{g}_\varphi, \mathfrak{g}_\varphi)^R - \dim \mathfrak{gl}(\mathfrak{g})^R + \dim \mathfrak{g}^R + m^2 - m$$

Preuve. On utilise le fait que les différents R-modules qui interviennent sont semi-simples; de plus l'algèbre R annule le groupe $H(\mathfrak{g}_\varphi, \mathfrak{g}_\varphi)$ puisqu'elle est constituée de répliques de $\mathrm{ad}\,\mathfrak{g}_\varphi$. CQFD.

Considérons l'algèbre de Lie $\mathfrak{g} = \mathfrak{a} \times \mathbb{C}$ avec $\mathfrak{a} = \mathfrak{u} \oplus \mathfrak{n}$ où le tore $T = \mathrm{ad}_{\mathfrak{n}}\mathfrak{u}$ est maximal sur $\mathfrak{n}$ et vérifie:

(2.3) $\dim T \geqq 4$;

(2.4) Les poids α qui définissent l'action de T sur $\mathfrak{n}$ sont de multiplicité 1 et non nuls;

(2.5) $H^2(\mathfrak{n}, \mathfrak{n})^T = 0$.

Ces hypothèses entraînent $H^i(\mathfrak{a}, \mathfrak{a}) = 0$ pour $0 \leqq i \leqq 2$ ainsi que la proposition suivante qui donne une nouvelle famille de composantes:

(2.6) **Proposition.** *L'adhérence de* $\Sigma(\mathfrak{g})$ *pour* $\mathfrak{g} = \mathfrak{a} \times \mathbb{C}$ *où* $\mathfrak{a} = \mathfrak{u} \oplus \mathfrak{n}$ *vérifie les hypothèses* (2.3, 2.4, 2.5) *est une composante irréductible de* L_m *et de* R_m *de dimension* $m^2 - m + r\dfrac{(r-3)}{2}$ $(r = \dim \mathfrak{u})$ *qui contient un ouvert non vide constitué de lois non décomposables (donc non algébriques) ainsi que des points simples pour le schéma défini par les relations de Jacobi. L'espace* $\Sigma(\mathfrak{g})$ *est constitué d'une famille continue à* $\dfrac{r(r-1)}{2} - 1$ *paramètres d'orbites de dimension* $m^2 - m - r + 1$ *(pour* $\varphi \neq 0$*) et d'une orbite de dimension* $m^2 - m - r$ *(pour* $\varphi = 0$*).*

Preuve. Soit une base t_i $(1 \leqq i \leqq r)$ de $\mathfrak{u}$. On a $Z(\mathfrak{g}) = \mathbb{C}$ et l'on choisit φ de sorte que la sous-algèbre de Cartan $\mathfrak{h} = \mathfrak{u} \times \mathbb{C}$ de $\mathfrak{g}_\varphi$ soit une algèbre de Heisenberg si r est pair ou le produit direct par $\mathbb{C}$ d'une algèbre de Heisenberg si r est impair. Il suffit de prendre φ tel que $\varphi(t_i, t_j)$, $(i<j)$ soit nul si $j \neq \left[\frac{r}{2}\right] + i$ et égal à 1 si $j = \left[\frac{r}{2}\right] + i$ où [ ] désigne la partie entière. L'espace $Z^2(\mathfrak{g}_\varphi, \mathfrak{g}_\varphi)^R$ est l'ensemble des 2-cochaînes f qui envoient $\mathfrak{h}^2$ dans $\mathfrak{h}$, $\mathfrak{h} \times \mathfrak{n}_\alpha$ dans $\mathfrak{n}_\alpha$ $\left(\text{si } \mathfrak{n} = \bigoplus_\alpha \mathfrak{n}_\alpha \text{ est la décomposition en sous-espaces de poids pour } T\right)$, $\mathfrak{n}_\alpha \times \mathfrak{n}_\beta$ dans $\mathfrak{n}_{\alpha+\beta}$ et qui vérifient les conditions de cocycle si [ ] désigne la loi de $\mathfrak{g}_\varphi$:

$$(2.7) \qquad (t, t', t'') \in \mathfrak{u}^3, \qquad \sum_{\mathrm{circ}(tt't'')} f(\varphi(t, t'), t'') = \sum_{\mathrm{circ}(tt't'')} [t, f(t', t'')];$$

$$(2.8) \qquad (t, t', x_0) \in \mathfrak{u}^2 \times \mathbb{C}, \qquad [t, f(t', x_0)] = [t', f(t, x_0)];$$

$$(2.9) \qquad (t, t', x) \in \mathfrak{u}^2 \times \mathfrak{n}_\alpha, \qquad f(\varphi(t, t'), x) = [x, f(t, t')];$$

$$(2.10) \qquad (t, x_0, x) \in \mathfrak{u} \times \mathbb{C} \times \mathfrak{n}_\alpha, \qquad [x, f(t, x_0)] = 0;$$

$$(2.11) \qquad (t, x, y) \in \mathfrak{u} \times \mathfrak{n}_\alpha \times \mathfrak{n}_\beta, \qquad f([x, y], t) = [x, f(y, t)] + [y, f(t, x)];$$

$$(2.12) \qquad (x_0, x, y) \in \mathbb{C} \times \mathfrak{n}_\alpha \times \mathfrak{n}_\beta, \qquad f([x, y], x_0) = [x, f(y, x_0)] + [y, f(x_0, x)];$$

$$(2.13) \qquad f|\mathfrak{n} \times \mathfrak{n} \in Z^2(\mathfrak{n}, \mathfrak{n})^T;$$

La condition (2.10) donne $f(\mathfrak{u}\times\mathbb{C})\subset\mathbb{C}$ et la condition (2.8) est automatiquement vérifiée. La condition (2.9) donne pour $i<j$ et $j\neq\left[\frac{r}{2}\right]+i$: $[x, f(t_i, t_j)]=0$, soit $f(t_i, t_j)\in\mathbb{C}$; si l'on pose

$$v=f\left(t_i, t_{\left[\frac{r}{2}\right]+1}\right)$$

elle donne $f(1,x)=[x, v]$ pour tout $x\in\mathfrak{n}$ et

$$f\left(t_i, t_{\left[\frac{r}{2}\right]+i}\right)=v$$

modulo $\mathbb{C}$, d'où:

(2.14) $\quad (t,t')\in\mathfrak{u}^2, \quad f(t,t')=\varphi(t,t')v+\lambda(t,t')\cdot 1 \quad$ avec $\quad \lambda\in C^2(\mathfrak{u},\mathbb{C})$.

Les vecteurs $w=\varphi(t,t')t''+\varphi(t',t'')t+\varphi(t'',t)t'$ pour $(t,t',t'')\in\mathfrak{u}^3$ engendrent $\mathfrak{u}$ pour $r\geqq 4$ (2.3); si l'on tient compte de (2.14) l'égalité (2.7) s'écrit $f(1,w)=[w,v]$ pour tout $w\in\mathfrak{u}$, d'où avec ce qui précède:

(2.15) $\quad f(1,x)=[x,v] \quad$ pour tout $x\in\mathfrak{g}_\varphi$,

et (2.12) devient inutile. La condition (2.11) exprime que pour chaque $t\in\mathfrak{u}$ l'application $x\to f(t,x)$ appartient à $Z^1(\mathfrak{n},\mathfrak{n})^T$ d'où:

(2.16) $\quad f|\mathfrak{u}\times\mathfrak{n}\in\mathfrak{u}^*\otimes Z^1(\mathfrak{n},\mathfrak{n})^T$.

Les conditions de cocycle se résument à (2.13, 2.14, 2.15, 2.16) où le vecteur v décrit $\mathfrak{h}$ et n'intervient dans (2.15) que par l'intermédiaire de $\mathrm{ad}_{\mathfrak{g}_\varphi} v$, c'est à dire modulo le centre $\mathbb{C}$ de $\mathfrak{g}_\varphi$. On en déduit la dimension de $Z^2(\mathfrak{g}_\varphi, \mathfrak{g}_\varphi)$ par le lemme (2.2) et ce qui précède en tenant compte de (2.4) et (2.5): on trouve $m^2-m+\frac{r(r-3)}{2}$. C'est aussi la dimension de $\Sigma(\mathfrak{g})$ obtenue par le lemme (2.1) en remarquant que le théorème de Künneth donne pour un produit direct $\mathfrak{g}=\mathfrak{g}_1\times\mathfrak{g}_2$:

$$H^1(\mathfrak{g},\mathfrak{g})=H^1(\mathfrak{g}_1,\mathfrak{g}_1)+H^1(\mathfrak{g}_2,\mathfrak{g}_2)+Z(\mathfrak{g}_1)\otimes(\mathfrak{g}_2/[\mathfrak{g}_2,\mathfrak{g}_2])+(\mathfrak{g}_1/[\mathfrak{g}_1,\mathfrak{g}_1])\otimes Z(\mathfrak{g}_2).$$

Si $\mathfrak{g}_1=\mathfrak{a}$ et $\mathfrak{g}_2=\mathbb{C}$ les hypothèses (2.4) et (2.5) entraînent les égalités $H^1(\mathfrak{g},\mathfrak{g})=\mathbb{C}\oplus(\mathfrak{a}/[\mathfrak{a},\mathfrak{a}])$ et $Z(\mathfrak{g})=\mathbb{C}$ d'où $\dim\Delta(\mathfrak{g})=m+r$ et la dimension cherchée. CQFD

(2.17) **Corollaire.** *Si* $\mathfrak{a}$ *vérifie les conditions* (2.3, 2.4, 2.5), *alors la composante* $\Sigma(\mathfrak{g})$ *avec* $\mathfrak{g}=\mathfrak{a}\times\mathbb{C}$ *est égale à l'adhérence de l'ensemble* $E_c^1(\mathfrak{a})$ *des lois d'algèbres de Lie qui sont (à isomorphisme près) extensions centrales de noyau* $\mathbb{C}$ *de* $\mathfrak{a}$.

Preuve. L'espace $E_c^1(\mathfrak{a})$ contient $\Sigma(\mathfrak{g})$ et d'après [4] il est constructible, irréductible et de dimension égale à $m(m-1)-\dim\Delta(\mathfrak{a})+\dim Z^2(\mathfrak{a},\mathbb{C})-v$ avec ici $v=0$ où $Z^2(\mathfrak{a},\mathbb{C})$ désigne l'espace des 2-cocycles à valeur triviale de $\mathfrak{a}$. Les dimensions des 2 sous-espaces calculées par les 2 formules sont les mêmes d'où l'égalité de leur adhérence. CQFD

(2.18) *Remarques.* (a) Les algèbres $\mathfrak{a}=\mathfrak{u}\oplus\mathfrak{n}$ qui vérifient les conditions (2.3, 2.4, 2.5) sont nombreuses: on peut prendre pour $\mathfrak{n}$ tout produit direct de plus de 4 facteurs nilpotents de dimension $\leqq 6$ (sauf $\mathfrak{n}_{6,5}$ et $\mathfrak{n}_{6,12}$ dans la nomemclature de [8]) de type $\mathfrak{n}_n^{(l)}$ ou des recollements centraux [5].

(b) Si $\mathfrak{a}_1$ et $\mathfrak{a}_2$ sont non isomorphes, les composantes $\bar{E}_c^1(\mathfrak{a}_i)$ $(i=1,2)$ sont distinctes. On en construit ainsi un grand nombre d'ordre $\Gamma(\sqrt{m})$ où Γ est la fonction gamma d'Euler en utilisant (a) et [5].

(c) Les deux composantes de N_7 mises en évidence par Vergne [8] sont de la forme de celles du corollaire (2.17): soient $\bar{E}_c^1(\mathfrak{n}_{6,19})$ et $\bar{E}_c^1(\mathfrak{n}_{6,15})$ [4] pour les algèbres de Lie nilpotentes $\mathfrak{n}_{6,19}$ et $\mathfrak{n}_{6,15}$ dans la nomenclature de [8].

3. Conclusions

Soient $R_m(\mathfrak{n})$ le sous-espaces de R_m des algèbres de Lie de plus grand idéal nilpotent isomorphe à $\mathfrak{n}, \mathfrak{o}_n$ l'algèbre de Lie abélienne $\mathbb{C}^n, \mathfrak{n}_3$ l'algèbre de Heisenberg $[x_1, x_2]=x_3$ et $\mathfrak{r}_2 \simeq T_1 \oplus \mathfrak{o}_1$ l'algèbre de Lie $[x_1, x_2]=x_2$. Les séries $\mathfrak{o}_n$ $(n \geqq 4)$, $\mathfrak{n}_3 \times \mathfrak{o}_{n-3}$ $(n \geqq 5)$ admettent des tores maximaux de dimension n et $n-1$ qui vérifient (2.3, 2.4, 2.5); (2.6, 2.17) donnent:

(3.1) $\bar{E}_c^1(\mathfrak{r}_2^n)$ $(n \geqq 4)$ est une composante irréductible de L_{2n+1} et de R_{2n+1} $\left(\text{de dimension } \frac{9}{2}n^2 + \frac{n}{2}\right)$ contenu dans $\bar{R}_{2n+1}(\mathfrak{o}_{n+1})$ et qui contient un ouvert non vide d'algèbres de Lie non décomposables;

(3.2) $\bar{E}_c^1(\mathfrak{r}_2^{n-3} \times (T_2 \oplus \mathfrak{n}_3))$ $(n \geqq 5)$ est une composante irréductible de L_{2n} et de R_{2n} (de dimension $\frac{9}{2}n^2 - \frac{9}{2}n + 2$) contenue dans $\bar{R}_{2n}(\mathfrak{n}_3 \times \mathfrak{o}_{n-2})$ et qui contient un ouvert non vide d'algèbres de Lie non décomposables.

En améliorant (3.1) on a le résultat suivant:

(3.3) **Proposition.** *L'espace $R_m(\mathfrak{o}_n)$ pour $n > m - n \geqq 4$ contient un ouvert non vide de L_m constitué d'algèbres de Lie non décomposables et un ouvert constitué de lois décomposables dont l'adhérence est la composante $\bar{R}_m^0(\mathfrak{o}_n)$ de L_m.*

Le corollaire (1.3) avec (3.1, 3.2) ou (3.3) permet d'énoncer le principal résultat de cette étude:

(3.4) **Proposition.** *L'ensemble des algèbres de Lie algébriques (respectivement décomposables) est: dense dans L_m et dans R_m pour $m \leqq 7$; non dense dans L_m et dans R_m pour $m \geqq 9$.*

Ceci met en évidence une difficulté supplémentaire pour construire les variétés L_m et R_m si $m \geqq 9$ et peut-être $m=8$ du fait que les lois non décomposables ne sont plus dans l'adhérence d'ouverts constitués de lois décomposables bien choisies. Il reste à trouver une méthode générale pour construire ces nouvelles composantes. Ceci s'ajoute à la difficulté de classer les lois nilpotentes à partir de la dimension 7 et l'existence de familles paramétrées. Notons cependant que différentes tables d'algèbres de Lie nilpotentes en dimension 7 sont parues récemment: E. N. Safiullina, Sur la classification des algèbres de Lie nilpotentes de dimension 7 (en russe) Kazan 1976; Ancochea Bermudez, Goze, CRAS 302 (1986); Magnin, L., Sur les algèbres de Lie nilpotentes de dimension $\leqq 7$. JGP.3 (1986); M. Romdhani, Classification of real and complex Nilpotent Lie algebras of Dimension 7, à paraître dans Linear and Multilinear Algebra; Seeley, C., Nilpotent Lie algebras of Dimension 7 over the Complex Numbers, (1988). Ceci permet d'envisager l'étude de L_8, intéressante à cause de phénomènes nouveaux qui y apparaissent.

Bibliographie

1. Bratzlavsky, F.: Sur les algèbres admettant un tore d'automorphismes donné. J. Algebra **30**, 305–316 (1974)
2. Carles, R., Diakité, Y.: Sur les variétés d'algèbres de Lie de dimension $\leqq 7$. J. Algebra **91**, 53–63 (1984)
3. Carles, R.: Sur la structure des algèbres de Lie rigides. Ann. Inst. Fourier, Grenoble **34**, 65–82 (1984)
4. Carles, R.: Sur les algèbre de Lie caractéristiquement nilpotentes. Prépublication n° 5, Poitiers (1984)
5. Carles, R.: Sur certaines classes d'algèbres de Lie rigides. Math. Ann. **272**, 477–488 (1985)
6. Favre, G.: Système de poids sur une algèbre de Lie nilpotente. Manuscr. Math. **9**, 53–90 (1973)
7. Hochschild, G., Serre, J.P.: Cohomology of Lie algebras. Ann. Math. **57**, 591–603 (1953)
8. Vergne, M.: Thèse 3^e cycle, Paris (1966)

Reçu le 23 juin 1988

Math. Ann. 283, 499–509 (1989)

Mathematische Annalen

On Essential Singularities of Meromorphic Mappings

Gerd-E. Dethloff

Mathematisches Institut der Universität, Bunsenstrasse 3–5, D-3400 Göttingen, Federal Republic of Germany

1. Introduction

The notion of meromorphic mapping which is used in this article is the one of Stein [6, 7]. It is equivalent to the notion "SR-meromorph" of Stoll [10], and the idea to this notion was already given by Remmert [5]. In [10], Stoll also introduced a notion of essential singularity of a meromorphic mapping: Let X^*, Y be normal complex spaces, $A \subset X^*$ a thin subset, $X = X^* \setminus A$ and $f: X \to Y$ a meromorphic mapping. Then a point $P \in A$ is called a "SR-Singularität" of f if there doesn't exist a neighbourhood U of P in X^* and a meromorphic extension $g: U \to Y$ of $f: (U \cap X) \to Y$.

But defining the notion of essential singularity like "SR-Singularität" has at least two disadvantages:

Firstly a "SR-Singularität" only is a very weak form of a singularity, e.g. the holomorphic function

$$z^{-1}: (\mathbb{C} \setminus \{0\}) \to \mathbb{C}$$

has a "SR-Singularität" in the zero point.

Secondly the assumption that A has to be thin in X^* is very restrictive, e.g. for the holomorphic function

$$\sqrt{z}: (\mathbb{C} \setminus \mathbb{R}_0^+) \to \mathbb{C}$$

$\mathbb{R}_0^+$ is not thin in $\mathbb{C}$.

Hence in this article the notion of essential singularity of meromorphic mappings is defined differently:

Firstly it is allowed to replace Y by a "bigger" normal complex space Z (that shall mean that Y is an open subspace of Z) before extending f into the point P. Further examples, some of which are given in the Sects. 3–5 of this paper, let it seem to be sensible to define three "versions" of an essential singularity: The first version is the one discribed above. For the second (resp. the third) version it also is allowed to replace Y by a normal complex space Y' which is a bit "smaller" than Y and then to replace Y' by a "bigger" space Z before extending f into P. Here "smaller" shall mean that there exists a closed nowhere dense subset M of Y and a holomorphic mapping $h: Y \to Y'$ so that $h(Y \setminus M)$ is an open subspace of Y' and $h: (Y \setminus M) \to h(Y \setminus M)$ is biholomorphic (resp. local biholomorphic).

Secondly it is allowed that $A \neq X^*$ is an arbitrary closed subset of X^* and P is any point of ∂X (where ∂X denotes the border of X in respect of X^*). Now the situation around P can be very complicated (e.g. it can happen that for any connected neighbourhood U of P, $U \cap X$ has infinitely many connected components), so it is not clear at once what an extension of f into P shall be. For $g: U \to Z$ being an extension of the mapping $f:(U \cap X) \to Z$ one should demand at least that there is an open subset $O \subset U \cap X$ with $P \in \partial O$ so that the equality $f = g$ holds on O. Using the identity-lemma for meromorphic mappings [6, p. 830] it can easily be shown that this is equivalent to demand that there exists a sequence $(G_\nu)_{\nu \in \mathbb{N}}$ of connected components G_ν of $U \cap X$ with $P \in \partial \left(\bigcup_{\nu \geqq \nu_0} G_\nu \right)$ for all $\nu_0 \in \mathbb{N}$ so that the equality $f = g$ holds on $\bigcup_{\nu \geqq 1} G_\nu$. Since in this paper it is intended to define the notion of essential singularity as strong as possible this already is the right concept of extension, but it should at least be added that other concepts of extension are possible and, more generally, there are also other possibilities to define the notion of essential singularity than the one given in this article.

In Sect. 2 we first define the three notions "essential singularity of the i-th kind, $i = 1, 2, 3$" (Definition 2.2). It is an immediate consequence of this definition that if f in P has an essential singularity of the i-th kind, it there also has an essential singularity of the $(i-1)$-th kind, $i = 2, 3$. Then a proposition is proved that shows that if A is nowhere dense in X^* and for every neighbourhood U of P there exists a subneighbourhood W so that $W \cap X$ is connected (these assumptions hold e.g. if A is thin in X^*), then Definition 2.2 gives the "right" notions.

In Sects. 3–5 there are given a lot of examples of meromorphic mappings with essential singularities which especially show that there exist mappings f and points P, in which f has an essential singularity of the i-th kind, but no essential singularity of the $(i+1)$-th kind, $i = 1, 2$. Beyond that there are given two theorems which prove:

(a) If X^* is connected with $\dim X^* = n$, $A \neq X^*$ is an arbitrary closed subset of X^* and m any number equal to or greater than n and 2, there exists a pure m-dimensional normal complex space Y_m and a meromorphic mapping $f_m: X \to Y_m$ which has essential singularities of the first kind in all points of ∂X, but no single essential singularity of the second kind.

(b) Let X^* and A be as in (a) and X^* be 1-dimensional. Then there exists a meromorphic mapping which has essential singularities of the third kind in every point of ∂X.

(c) Let Y be 1-dimensional. Then every essential singularity of the first kind is one of the second kind.

Notice that (a) also is interesting in connection with extension problems for meromorphic mappings as they were examined by Stein in [7, 8] (cf. also Theorem 6.3 of this paper), because for thin A (a) especially yields that the correspondence given by $\overline{G_{f_m}} \subset X^* \times Y_m$, where G_{f_m} is the graph of $f_m: X \to Y_m$, doesn't yield a meromorphic mapping in any point $P \in A$. This result even holds if Y_m is enlarged to a "bigger" space Z as it was described above.

In Sect. 6 there are collected some propositions which can be helpful when trying to prove that a meromorphic mapping in a given point has **no** essential singularity. Among them there is an important theorem of Stein which deals with the case that A is thin in X^* and $\dim X^* - \dim A > \dim Y$ and a further theorem of Stoll which can be applied if Y is projective-algebraic.

2. Definition of Essential Singularities of Meromorphic Mappings

First we introduce some notations which will be kept up during this paper. Let X^* and Y be normal complex spaces, where X^* is connected and has a countable basis of topology. Let $A \neq X^*$ be a non-empty closed subset of X^*, $X := X^* \backslash A$ and $P \in \partial X$, where ∂X denotes the border of X in respect of X^*. Let further $f: X \to Y$ be a meromorphic mapping, $G_f \subset X \times Y$ its graph, $\check{f}: G_f \to X$ and $\hat{f}: G_f \to Y$ its canonical projections and $S_f \subset X$ the set of its (non-essential) singularities.

Definition 2.1 *(c-sequence)*. Let U be a connected neighbourhood of P in X^*. For every $\nu \in \mathbb{N}$ let G_ν be (not necessary different) connected components of $U \cap X$ with

$$P \in \partial\Big(\bigcup_{\nu \geqq \nu_0} G_\nu\Big) \quad \text{for all } \nu_0 \in \mathbb{N}.$$

Then $\mathscr{G} := (G_\nu)_{\nu \in \mathbb{N}}$ is called *c-sequence (in resp. of P and U)*. For $\mathscr{G}$ we set $|\mathscr{G}| := \bigcup_{\nu \geqq 1} G_\nu \subset U \cap X$.

Definition 2.2 *(essential singularities of meromorphic mappings)*. P is said to be an *essential singularity of the i-th kind (*ess *i*-sing*) of f (in resp. of X^*), $i = 1, 2, 3$,* if for every
connected neighbourhood U of P in X^*,
c-sequence $\mathscr{G}$ in resp. of P and U,
normal complex space Z,
$h \in \mathscr{H}_i(Y, Z)$ (see below),
there doesn't exist a meromorphic extension $g: U \to Z$ of the mapping $h \circ f: |\mathscr{G}| \to Z$.
The sets $\mathscr{H}_i(Y, Z)$ are defined as subsets of the set of *holomorphic mappings* from Y to Z as follows:
$\mathscr{H}_1(Y, Z)$ consists of all $h: Y \to Z$ for which there exists an open subset $Z_0 \subset Z$, so that $h: Y \to Z_0$ is biholomorphic.
$\mathscr{H}_2(Y, Z)$ consists of all $h: Y \to Z$ for which there exists an open subset $Z_0 \subset Z$ and a closed and nowhere dense subset $M \subset Y$, so that $h: (Y \backslash M) \to Z_0$ is biholomorphic.
$\mathscr{H}_3(Y, Z)$ consists of all $h: Y \to Z$ for which there exists a closed and nowhere dense subset M of Y, so that $h: (Y \backslash M) \to Z$ is locally biholomorphic.

It can be easily proved that Definition 2.2 defines a local property of f. Definition 2.2 becomes simpler if A has additional properties:

Proposition 2.3. *Let A be nowhere dense in X^* and assume that for every neighbourhood V of P in X^* there exists a subneighbourhood W for which $W \cap X$ is connected.*
Then P is an ess *i*-sing *of f in resp. of X^* if and only if for every*
connected neighbourhood U of P in X^*,
normal complex space Z,
$h \in \mathscr{H}_i(Y, Z)$ (cf. Definition 2.2),
there doesn't exist a meromorphic extension $g: U \to Z$ of $h \circ f: (U \cap X) \to Z$.

The *proof* is straightforward.

3. Some Relations Between Ess 1-Sing and Ess 2-Sing

Theorem 3.1. (a) *Let X^* be n-dimensional, $m \in \mathbb{N}$ with $m \geqq \max(n, 2)$. Then there exists a pure m-dimensional normal complex space Y_m and a meromorphic mapping $f_m: X \to Y_m$, so that every point of ∂X is an* ess 1-sing, *but no point of ∂X is an* ess 2-sing.

(b) *Let Y be 1-dimensional. Then if P is an* ess 1-sing *of f, it also is an* ess 2-sing *of f.*

Remark. See Remark 2 of Theorem 6.3 for a supplementation to this theorem.

Before we start with the proof of Theorem 3.1, we prove two lemmas:

Lemma 3.2. *Let M_1, $M_2 \subset X^*$ be open subsets, M_2 be connected and M_3 be a connected component of $M_1 \cap M_2$. If $M_1 \cap M_2 \neq \emptyset \neq (X^* \backslash M_1) \cap M_2$, then*

$$M_2 \cap \partial M_1 \cap \partial M_3 \neq \emptyset .$$

For the *Proof of Lemma 3.2*, we refer, if necessary, to [2, p. 39].

Lemma 3.3. *Let S be the singular locus of X^*. Then there exists a sequence $F = (x_\mu)_{\mu \in \mathbb{N}}$ with the following properties:*

(a) *$x_\mu \in X \backslash S$ for all $\mu \in \mathbb{N}$; $x_{\mu_1} \neq x_{\mu_2}$ for $\mu_1 \neq \mu_2$.*

(b) *We set $|F| := \{x_\mu : x_\mu \in F\}$. Then $|F|$ is a discrete subset of X.*

(c) *For all $x \in \partial X$, for all connected neighbourhoods U of x in X^* and for all connected components G of $U \cap X$ the set $G \cap |F|$ contains infinitely many points: $\#(G \cap |F|) = \infty$.*

Proof of Lemma 3.3. Let $\mathscr{B} := \{B_\nu, \nu \in \mathbb{N}\}$ be a countable basis of the topology of X^* consisting of connected sets and $\mathscr{L}_A$ be the set of all connected components of $B_\nu \cap X$ of those $B_\nu \in \mathscr{B}$ with $B_\nu \cap A \neq \emptyset$. The set $\mathscr{L}_A$ is countable, so we can enumerate its elements:

$$\mathscr{L}_A = \{G_\mu, \mu \in \mathbb{N}\} .$$

Since X^* has a countable basis of topology, we can introduce a metric $\delta(\cdot, \cdot)$ on X^*. Now define

$$A_\nu := \left\{ x \in X^* : \delta(x, A) < \frac{1}{\nu} \right\} \quad \text{for every } \nu \in \mathbb{N} .$$

Then $A_\nu \cap G_\nu \neq \emptyset$: Since A_ν is a neighbourhood of A, it is enough to show $\partial G_\nu \cap A \neq \emptyset$. This directly follows if we apply Lemma 3.2.

Now we can construct the sequence F:

Choose x_1 from $(A_1 \cap G_1) \backslash S$. If $x_1, \ldots, x_{\mu-1}$ are already constructed, choose x_μ from $(A_\mu \cap G_\mu) \backslash (S \cup \{x_1, \ldots, x_{\mu-1}\})$.

It follows directly from this construction that the properties (a) and (b) are fulfilled. From Lemma 3.2, applied with $M_1 = X$, $M_2 = U$, $M_3 = G$, it follows that there exists a $x_0 \in U \cap \partial X \cap \partial G$.

Assume now $\#(|F| \cap G) < \infty$. Then $V := U \backslash (G \cap |F|)$ is a neighbourhood of x_0, so there is an open set $B \in \mathscr{B}$ with $x_0 \in B \subset V$. Since $B \cap G \neq \emptyset$ there exists a connected component G_ν of $B \cap X$ in $\mathscr{L}_A$ with $G_\nu \cap (G \backslash |F|) \neq \emptyset$. Hence $G_\nu \subset G \backslash |F|$, but this contradicts $|F| \cap G_\nu \supset \{x_\nu\}$. $\square$

Proof of Part (a) of the Theorem in the Case $n \geqq 2$. First we construct a sequence F like in Lemma 3.3. Then we construct Y from X by blowing up X simultaneously in every point of $|F|$ by Hopf's σ-process [4]. If $\pi: Y \to X$ is the canonical projection, we have:

$$f := \pi^{-1}: X \to Y; \quad x \to \pi^{-1}(x) \text{ is a meromorphic mapping,} \tag{1}$$

$$S_f = |F|, \tag{2}$$

$$\pi: Y \to X \text{ is a holomorphic mapping with } \pi \circ f = \mathrm{id}|_X, \tag{3}$$

$$Y \setminus f(X \setminus |F|) \text{ is a 1-codimensional analytic set in } Y. \tag{4}$$

We now define $Y_m := Y \times \mathbb{C}^{m-n}$ and $f_m: X \to Y_m$; $x \to (f(x), 0, \ldots, 0)$ and prove the assertion of part (a) of the theorem for $n \geqq 2$:

Take a point $P \in \partial X$. It is easy to see that f_m has no ess 2-sing in P: Define U any connected neighbourhood of P, $\mathscr{G}$ any c-sequence in resp. of P and U, $Z := X^* \times \mathbb{C}^{m-n}$, $h := \pi \times \mathrm{id}^{(m-n)}: Y_m \to Z$; $(y, z_1, \ldots, z_{m-n}) \to (\pi(y), z_1, \ldots, z_{m-n})$, $g: X^* \to X^* \times \mathbb{C}^{m-n}$; $x \to (x, 0, \ldots, 0)$. As a consequence of (3) and (4) we have $h \in \mathscr{H}_2(Y_m, Z)$, and $h \circ f_m = g$ on $|\mathscr{G}|$ is obvious.

To prove that f_m has an ess 1-sing in P, we assume the contrary. Under this assumption there exist U, $\mathscr{G}$, Z, h and g like in Definition 2.2. We especially have $h \circ f_m = g$ on $|\mathscr{G}|$, hence (cf. [6, p. 837])

$$S_g \cap |\mathscr{G}| = |\mathscr{G}| \cap |F|. \tag{5}$$

S_g is an analytic subset of U, for which there exists a unique decomposition in irreducible analytic sets $(S_g)_\iota$, $\iota \in I$, so that $\{(S_g)_\iota, \iota \in I\}$ is a local finite covering of S_g, hence there exists a connected subneighbourhood $V \subset U$ of P with

$$V \cap (S_g)_\iota \neq \emptyset \text{ only for a finite number of } \iota \in I. \tag{6}$$

We have $V \cap |\mathscr{G}| \cap |F| = V \cap |\mathscr{G}| \cap S_g$ [cf. (5)]. From this we conclude with Lemma 3.3b and (6):

$$\#(V \cap |\mathscr{G}| \cap |F|) < \infty. \tag{7}$$

Because of the properties of $\mathscr{G}$ there exists a $\nu_0 \in \mathbb{N}$ with $G_{\nu_0} \cap V \neq \emptyset$, hence there is a connected component G of $V \cap X$ with $G \subset G_{\nu_0}$. An application of Lemma 3.3c to P, V and G yields $\#(|F| \cap G) = \infty$, hence $\#(V \cap |\mathscr{G}| \cap |F|) = \infty$, which contradicts (7). □

To prove part (a) of the theorem in the case $n = 1$, we need another lemma:

Lemma 3.4. *Let* X^* *be 1-dimensional. Then there exists a sequence* $F = (x_\mu)_{\mu \in \mathbb{N}}$ *in* X *with the properties* (a), (b), *and* (c) *of Lemma 3.3 and a holomorphic function* $f: X \to \mathbb{C}$ *with*

$$\{x \in X : f(x) = 0\} = |F|.$$

Proof. Choose the sequence F like in Lemma 3.3. It is sufficient to construct such a function on every connected component of X. X^* is a Riemann surface (because the singular locus is at least 2-codimensional, hence empty), so the connected components of X are Riemann surfaces, too. On every connected component of X, $|F|$ yields a Cousin-II-distribution. Since such connected components are not compact (this is a simple consequence of the properties of F), this distribution has a solution, and this solution has the desired properties. □

Proof of Part (a) of the Theorem in the Case $n=1$: Let $f: X \to \mathbb{C}$ be the function defined in the previous lemma and define $Y_m := \mathbb{C}^m$ and $f_m: X \to Y_m$; $x \to (f(x), 0, \ldots, 0)$.

Take any point $P \in \partial X$. Again, it is easy to see that P is no ess 2-sing of f_m: Let U be any connected neighbourhood of P, $\mathscr{G}$ any c-sequence in resp. of P and U, $Z := \mathbb{C}^m$, $h: Y_m \to Z$; $(z_1, z_2, \ldots, z_m) \to (z_1 z_2, z_2, \ldots, z_m)$ and $g = 0: U \to Z$.

To prove that f_m has an ess 1-sing in P, we again assume the contrary. Under this assumption there exist U, $\mathscr{G}$, Z, h, and g like in Definition 2.2. First we show:

$$P \in \partial(|F| \cap |\mathscr{G}|), \qquad |F| \cap |\mathscr{G}| \neq \emptyset. \tag{8}$$

It suffices to show that for every connected subneighbourhood $U' \subset U$ of P the nonequality $|F| \cap U' \cap |\mathscr{G}| \neq \emptyset$ holds. There exists a G_v with $G_v \cap U' \neq \emptyset$ and hence a connected component G of $U' \cap X$ with $G \subset G_v$. An application of Lemma 3.3 yields $\#(|F| \cap G) = \infty$ and hence $|F| \cap |\mathscr{G}| \cap U' \neq \emptyset$.

Let $z_0 := h(0)$. Then $g(x) = z_0$ for all $x \in |F| \cap |\mathscr{G}|$, and with $S_g = \emptyset$ (cf. [10, p. 224]) and (8) we can conclude:

$$\{x \in U : g(x) = z_0\} \supset (|F| \cap |\mathscr{G}|) \cup \{P\}. \tag{9}$$

There exists a neighbourhood W of z_0 in Z which is mapped biholomorphically on a closed analytic subspace of a domain in a $\mathbb{C}^r$. Hence g yields r holomorphic functions which, because of (9), are constant on $((|F| \cap |\mathscr{G}|) \cup \{P\}) \cap g^{-1}(W)$. Therefore if $V \subset g^{-1}(W)$ is a connected neighbourhood of P, we have with (8) and the identity-lemma on Riemann surfaces:

$$g(x) \equiv z_0 \quad \text{for all} \quad x \in V. \tag{10}$$

Since we have $h \circ f_m = g$ on $V \cap |\mathscr{G}| (\neq \emptyset)$ and h is injective, we get from (10) and (8):

$$f_m(x) \equiv 0 \quad \text{for all} \quad x \in V \cap |\mathscr{G}|.$$

But this is impossible, since $\{x \in X : f_m(x) = 0\} = |F|$ and $|F|$ only is a discrete subset of the open set $V \cap |\mathscr{G}|$. □

Before we start with the proof of part (b) of the theorem, we prove a topological lemma:

Lemma 3.5. *Let S, T be topological spaces which locally admit a metric, $C \subset S$ a closed and nowhere dense subset. Let $f: S \to T$ be continuous and $f: (S \backslash C) \to T$ be injective. Let $O_1, O_2 \subset S$ be open sets with $O_1 \cap O_2 = \emptyset$ for which $f(O_i)$ are open subsets of T and $f: O_i \to f(O_i)$ are topological maps.*

Then $f(O_1) \cap f(O_2) = \emptyset$.

Proof. Assume $W := f(O_1) \cap f(O_2) \neq \emptyset$. Since $f: O_i \to f(O_i)$ are topological maps and T locally admits a metric $W' := (W \cap [f(O_1 \cap C) \cup f(O_2 \cap C)])$ is closed and nowhere dense in W. Hence there exists $w_0 \in W \backslash W'$ and $w_1 \in O_1 \backslash C$, $w_2 \in O_2 \backslash C$ with $f(w_1) = f(w_2) = w_0$, but this is impossible because $f: (S \backslash C) \to T$ was injective. □

Proof of Part (b) of the Theorem. Let Z be a normal complex space. It suffices to show $\mathscr{H}_2(Y, Z) \subset \mathscr{H}_1(Y, Z)$. Let $h \in \mathscr{H}_2(Y, Z)$ and Y_1 be a connected component of Y. Then $h(Y_1) \subset Z$ is an open subset and $h: Y_1 \to h(Y_1)$ is biholomorphically: For $\dim Y_1 = 0$ this is a direct consequence of $h \in \mathscr{H}_2(Y, Z)$, for $\dim Y_1 = 1$ we will prove that below. The previous lemma now shows that $h: Y \to Z$ is injective, hence $h \in \mathscr{H}_1(Y, Z)$.

We still have to show that if $h \in \mathscr{H}_2(Y, Z)$ with a Riemann surface Y then $h(Y)$ is an open subset of Z and $h: Y \to h(Y)$ is biholomorphic. It is enough to show that $h: Y \to Z$ is locally biholomorphic, since then an application of Lemma 3.5 completes the proof.

Since $h(Y)$ is connected, we may assume that Z is a Riemann surface. If we introduce local charts in Y and Z in an appropriate way (cf. [1, p. 164]), we reduce our assertion to the following one:

Let $\varepsilon \in \mathbb{R}^+$, $U_\varepsilon(0) := \{z \in \mathbb{C} : |z| < \varepsilon\}$, $N \subset U_\varepsilon(0)$ a closed and nowhere dense subset, $f(z) := z^p$ with $p \in \mathbb{N}$ so that $f: (U_\varepsilon(0) \backslash N) \to \mathbb{C}$ is injective. Then $p = 1$.

Assume $p \geqq 2$. Then the two points $z_1 = \frac{\varepsilon}{2}$, $z_2 = \frac{\varepsilon}{2} e^{\frac{2\pi i}{p}}$ are different, so there are neighbourhoods O_1 (resp. O_2) of z_1 (resp. z_2) with $O_i \subset U_\varepsilon(0)$ and $O_1 \cap O_2 = \emptyset$, for which the mappings $f: O_i \to f(O_i)$ are biholomorphic. Then the Lemma 3.5 yields $f(O_1) \cap f(O_2) = \emptyset$, but this is wrong since $f(z_1) = f(z_2)$. □

4. Some Relations Between Ess 2-Sing and Ess 3-Sing

First, we introduce some special notations for this section:

$$\begin{aligned} G &:= \{z = r \cdot e^{2\pi i \alpha} : r \in \mathbb{R}^+, \alpha \in \mathbb{R}, 0 < \alpha < 1\}. \\ H &:= \{z = r \cdot e^{2\pi i \alpha} : r \in \mathbb{R}^+, \alpha \in \mathbb{R}, 0 < \alpha < \tfrac{1}{2}\}. \\ \tilde{f} &: G \to H; \quad r \cdot e^{2\pi i \alpha} \to \sqrt{r} \cdot e^{\pi i \alpha}. \\ X^* &= Y := \mathbb{C}^n, \quad X := G \times \mathbb{C}^{n-1}, \\ A &:= \mathbb{C}^n \backslash X, \quad S := \{z \in A : z_1 = 0\} \end{aligned} \tag{11}$$

and, for

$$\varepsilon \in \mathbb{R}, x = \{x_1, \ldots, x_n\} \in \mathbb{C}^n: \quad U_\varepsilon(x) := \{z \in \mathbb{C}^n : |z - x| < \varepsilon\},$$
$$U_\varepsilon(x_1) := \{z \in \mathbb{C} : |z - x_1| < \varepsilon\}.$$

We define

$$f: X \to Y; \quad (z_1, z_2, \ldots, z_n) \to (\tilde{f}(z_1), z_2, \ldots, z_n).$$

Proposition 4.1. (a) *P is no* ess 1-sing *of f for all $P \in A \backslash S$.*

(b) *P is an* ess 2-sing *of f for all $P \in S$.*

(c) *P is no* ess 3-sing *of f for all $P \in A$.*

Proof. (a) is obvious, since, if $P = (p_1, \ldots, p_n) \in A \backslash S$ and $\varepsilon \in \mathbb{R}^+$ with $\varepsilon < p_1$ we can extend $\tilde{f}$ holomorphically from $U_\varepsilon(p_1) \cap \{\operatorname{Im} z_1 > 0\}$ to $U_\varepsilon(p_1)$.

(c) is easy: Choose $Z = \mathbb{C}^n$, $h: \mathbb{C}^n \to \mathbb{C}^n$; $(z_1, z_2, \ldots, z_n) \to (z_1^2, z_2, \ldots, z_n)$. To prove (b), let $P \in S$ be arbitrary. Assume that P is no ess 2-sing of f. Then there exist U, Z, h, M, and g like in Proposition 2.3. First we want to prove:

There are points $Q_1 = (q_1, q_2, \ldots, q_n)$, $Q_2 = (-q_1, q_2, \ldots, q_n)$ in $\mathbb{C}^n$ with $q_1 \in \mathbb{R}^+$ and $\delta \in \mathbb{R}^+$ with $\delta < q_1$ in such a way, that for every two points $R_1 = (r_1, r_2, \ldots, r_n)$, $R_2 = (-r_1, r_2, \ldots, r_n)$ in $\mathbb{C}^n$ with $r_1 \in \mathbb{R}^+$ and $R_1 \in U_\delta(Q_1)$ the equality $h(R_1) = h(R_2)$ holds. (12)

Since S_g is a 2-codimensional analytic subset of U there exist a point $P' = (p'_1, \ldots, p'_n) \in A \cap U$ with $p'_1 > 0$ and an $\eta \in \mathbb{R}^+$ with $\eta < p_1$, so that we have:

$$U_\eta(P') \subset U, \quad U_\eta(P') \cap S_g = \emptyset. \tag{13}$$

Define $Q_1 := (+\sqrt{p'_1}, p'_2, \ldots, p'_n)$, $Q_2 := (-\sqrt{p'_1}, p'_2, \ldots, p'_n)$ and $\delta \in \mathbb{R}^+$ so small, that, if q denotes the mapping $\mathbb{C}^n \to \mathbb{C}^n$; $(z_1, z_2, \ldots, z_n) \to (z_1^2, z_2, \ldots, z_n)$, we have $\delta < \sqrt{p'_1}$ and $q(U_\delta(Q_1)) \subset U_\eta(P')$.

Let R_1, R_2 be like in (12) and $R := q(R_1)$. Let $(z_\nu^{(1)})_{\nu \in \mathbb{N}}, (z_\nu^{(2)})_{\nu \in \mathbb{N}}$ be sequences with $z_\nu^{(1)} \in X \cap \{\operatorname{Im} z_1 > 0\}$, $z_\nu^{(2)} \in X \cap \{\operatorname{Im} z_1 < 0\}$ and $z_\nu^{(1)} \to R \leftarrow z_\nu^{(2)}$ for $\nu \to \infty$. From (11), we conclude $f(z_\nu^{(1)}) \to R_1$, $f(z_\nu^{(2)}) \to R_2$ and hence, because $R \notin S_g$ [cf. (13)],

$$g(R) = \lim_{\nu \to \infty} g(z_\nu^{(i)}) = \lim_{\nu \to \infty} h \circ f(z_\nu^{(i)}) = h(R_i) \quad \text{for} \quad i = 1, 2$$

which proves (12).

Now define $s: \mathbb{C}^n \to \mathbb{C}^n$; $(z_1, z_2, \ldots, z_n) \to (-z_1, z_2, \ldots, z_n)$. The set $\{z \in U_\delta(Q_1) : h \circ s = h\}$ is an analytic subset of $U_\delta(Q_1)$, which contains the set $U_\delta(Q_1) \cap A$ [cf. (12)], hence $h \circ s = h$ on $U_\delta(Q_1)$.

Choose z' from the set $U_\delta(Q_1) \setminus [(U_\delta(Q_1) \cap M) \cup s(U_\delta(Q_2) \cap M)]$. Then $s(z') \in U_\delta(Q_2) \setminus M$, especially $z', s(z') \in \mathbb{C}^n \setminus M$ and $z' \neq s(z')$, but $h(z') = h(s(z'))$, what is impossible, because h is injective on $\mathbb{C}^n \setminus M$. □

5. Some Examples for Ess 3-Sing

We again use the notations introduced in Sect. 2.

Theorem 5.1. *Let X^* be a Riemann surface. Then there is a holomorphic function $f: X \to \mathbb{C}$ so that every point $P \in \partial X$ is an* ess 3-sing *of f.*

Proof. First we apply Lemma 3.4 and get a sequence $F = (x_\mu)_{\mu \in \mathbb{N}}$ with the properties (a), (b), and (c) of Lemma 3.3 and a holomorphic function $f: X \to \mathbb{C}$ with

$$\{x \in X : f(x) = 0\} = |F|. \tag{14}$$

Let us assume that there exists a point $P \in \partial X$ which is no ess 3-sing of f. Then there exist U, $\mathscr{G}$, Z, h, M, and g like in Definition 2.2. Since $\mathbb{C}$ is connected, we may assume that Z is connected, too, and hence a Riemann surface. Now we can prove (8, 9) literally as it was done in Sect. 3. From (8, 9) we can conclude with the identity-lemma for holomorphic mappings between Riemann surfaces (where $z_0 = h(0)$):

$$g(x) \equiv z_0 \quad \text{for all} \quad x \in U. \tag{15}$$

Since $|F|$ is a discrete subset of X and (14, 8) there exists a connected component G_ν of $|\mathscr{G}|$ where f is not constant, hence locally biholomorphic outside a discrete subset of G_ν. Since h is locally biholomorphic outside M, too, there exists an open subset V of $|\mathscr{G}|$ where $h \circ f$ is locally biholomorphic. This contradicts (15). □

Proposition 5.2. *Let H_n be the n-dimensional Hopf-manifold and $\pi: (\mathbb{C}^n \setminus \{0\}) \to H_n$ the canonical projection.*

Then the zero point of $\mathbb{C}^n$ is an ess 3-sing *of π.*

Proof. Define $X^*:=\mathbb{C}^n$, $A:=\{0\}$, $Y:=H_n$ and $f:=\pi$. Let $p\in\mathbb{R}^+$ be the smallest number so that for an arbitrary $z\in X$ we have $f(p\cdot z)=f(z)$, and, for this p and any $r\in\mathbb{R}^+$, define $F_r:=\{z\in X: r<|z|<p\cdot r\}$ (cf. [2, p. 146]).

Assume that the zero point is no ess 3-sing of f. Then there exist U, Z, h, and g like in Proposition 2.3. There exist open subsets $U_0\subset Y$ and $W_0\subset Z$, so that $h: U_0\to W_0$ is biholomorphic, especially W_0 is n-dimensional. We further may assume that $U_0\subset f(F_r)$ for suitable chosen $r\in\mathbb{R}^+$. For all $k\in\mathbb{N}_0$ define $V_k:=(f|_{F_{rp^{-k}}})^{-1}(U_0)$. Then all mappings $f: V_k\to f(V_k)=U_0$ are biholomorphic.

Now let $w_0\in W_0$ be arbitrary. Then there exists a point $u_0\in U_0$ and for all $k\in\mathbb{N}_0$ a point $v_k\in V_k$ with $f(v_k)=u_0$, $h(u_0)=w_0$, hence $(v_k, w_0)\in G_g$. Since $v_k\to 0$ if $k\to\infty$ and G_g is closed in $U\times Z$ we have $(0, w_0)\in G_g$, and, because $w_0\in W_0$ was arbitrary:

$$\{0\}\times W_0\subset G_g, \qquad \dim W_0=n. \tag{16}$$

Since G_g is an irreduzile n-dimensional analytic set, we therefore get the contradiction $G_g=G_g\cap(\{0\}\times Z)$. □

6. When do Ess *i*-Sing not Exist?

Proposition 6.1 (Product-Spaces). *Let $Y_1,\ldots,Y_t$ be normal complex spaces, $Y=Y_1\times\ldots\times Y_t$ and pr_j, $j=1,\ldots,t$, the canonical projections from Y to Y_t.*

(a) *If there exists a connected neighbourhood U of P in X^*, a c-sequence $\mathscr{G}$ in resp. of P and U and for every $j\in\{1,\ldots,t\}$ a normal complex space Z_j and a holomorphic mapping $h_j\in\mathscr{H}_i(Y_j,Z_j)$ such that $h_j\circ pr_j\circ f:|\mathscr{G}|\to Z_j$ can be extended to a meromorphic map $g_j: U\to Z_j$, then P is no* ess *i*-sing *of f.*

(b) *Let A be nowhere dense in X^* and assume that for every neighbourhood V of P in X^* there exists a subneighbourhood W such that $W\cap X$ is connected. Then if P is no* ess *i*-sing *of any mapping $pr_j\circ f: X\to Y_j$, $j=1,\ldots,t$, P also is no* ess *i*-sing *of f.*

Remark. There exist meromorphic mappings, for which some $pr_j\circ f$ may have ess *i*-sing in P, but f hasn't: The mapping f_m constructed in the proof of part (a) of Theorem 3.1 in the case $n=1$ has no ess 2-sing, but $pr_1\circ f_m$ has ess 3-sing, as we showed in the proof of Theorem 5.1.

Proof of Proposition 6.1. (a) Define $Z:=Z_1\times\ldots\times Z_t$, $h:=h_1\times\ldots\times h_t: Y\to Z$. It is easily proved that $h\in\mathscr{H}_i(Y,Z)$. Now we have to construct g: Let $G_{g^*}:=\{(x, z_1,\ldots,z_t): x\in U,\ z_i\in g_i(x),\ i=1,\ldots,t\}\subset U\times Z$. Then there exists a meromorphic map $g: U\to Z$ with $G_g\subset G_{g^*}$, (cf. [6, p. 839]). There further exists a closed an thin subset M^* of U such that $G_{g^*}\cap[(U\backslash M^*)\times Z]$ gives a holomorphic map, hence $G_g\cap[(U\backslash M^*)\times Z]=G_{g^*}\cap[(U\backslash M^*)\times Z]$. From the last equality it follows $g=h\circ f$ on $(|\mathscr{G}|\backslash M^*)$, hence an application of the identity-lemma for meromorphic maps (cf. [6, p. 830]) yields $g=h\circ f$ on $|\mathscr{G}|$. So g is an extension of $h\circ f$ from $|\mathscr{G}|$ to U.

(b) The proof is straightforward if we firstly apply Proposition 2.3, then the special assumption on the structure of A and at last part (a). □

Proposition 6.2 (Closed Complex Subspaces). *Let A be nowhere dense in X, U a connected neighbourhood of P in X^*, Z a normal complex space and $h\in\mathscr{H}_i(Y,Z)$. Let Z be a closed complex subspace of a normal complex space Z_0.*

Then, if $g: U \to Z_0$ *is a meromorphic extension of* $h \circ f: U \cap X \to Z$, *we have* $g(U) \subset Z$ *and* $g: U \to Z$ *is a meromorphic mapping; especially P is no* ess *i*-sing *of f.*

Proof. $\check{g}^{-1}(A \cup S_g)$ is a closed, nowhere dense subset of G_g (cf. [6, p. 823], [3, p. 167]). Since $\check{g}^{-1}(U \cap X) \subset U \times Z$ we therefore have $G_g \subset U \times Z$. Since G_g is irreducible in $U \times Z_0$ and $\check{g}$ is a proper map these properties hold in $U \times Z$, too. □

The next theorem is due to Stein [8, 9]:

Theorem 6.3. *Define* $\operatorname{fmd} f := \operatorname{Min}_{z \in G_j} \dim \hat{f}^{-1}(\hat{f}(z))$.

Let $A \subset X^*$ *be analytic and* $\operatorname{fmd} f > \dim A$.

Then the topological closure $\overline{G_f}$ *of* G_f *in* $X^* \times Y$ *is a meromorphic extension of f from X to* X^**; especially f has no* ess 1-sing *in any point* $P \in A$.

Remark 1. A simple dimension-theoretic calculation shows that $\operatorname{fmd} f \geqq \dim G_f - \dim Y = \dim X^* - \dim Y$. So Theorem 6.3 is especially true if $\operatorname{codim} A > \dim Y$.

Remark 2. Theorem 6.3 shows that the inequality "$m \geqq \max(n, 2)$" in part (a) of Theorem 3.1 can't be improved: If $m < n$, we take A an isolated point P of X^*. Then Theorem 6.3 tells that $f_m: X \to Y_m$ has no ess 1-sing in P. The fact "$m \geqq 2$" already follows from part (b) of Theorem 3.1.

The next theorem actually only is an application of Stoll's theorem (4.3) in his paper [10]:

Theorem 6.4. *Let A be thin of codimension 2 in* X^* *and Y be a projective-algebraic space.*

Then there exists a meromorphic extension $g: X^* \to Y$ *of* $f: X \to Y$*; especially f has no* ess 1-sing *in any point* $P \in A$.

Proof. We may assume that $Y = \mathbb{P}^r$ for a suitable $r \in \mathbb{N}$, since then the assertion for arbitrary Y follows from Proposition 6.2. An application of Stoll's theorem (4.3) yields:

There exists a $\mu \in \{0, \ldots, r\}$ such that $X \setminus S_f \not\subset (f|_{X \setminus S_j})^{-1}(E_\mu)$ and for all $P \in X \setminus (S_f \cup (f|_{X \setminus S_j})^{-1}(E_\mu))$ the equation

$$f(P) = (f_0(P) : f_1(P) : \ldots : f_{\mu-1}(P) : 1 : f_{\mu+1}(P) : \ldots : f_r(P))$$

with meromorphic functions $f_i: X \to \mathbb{C}$ holds, where

$$E_\mu = \{(w_0 : w_1 : \ldots : w_r) \in \mathbb{P}^r : w_\mu = 0\}.$$

Now Levi's extension-theorem (cf. [3, p. 185]) tells us that we can extend the f_i to meromorphic functions $f_i^*: X^* \to \mathbb{C}$. Let $P(f_i^*)$ be the polar sets of f_i^* and $R = \bigcup_{i=0,\ldots,n;\, i \neq \mu} P(f_i^*)$. Then

$$f^*: (X^* \setminus R) \to \mathbb{P}^r;\ P \to (f_0^*(P) : f_1^*(P) : \ldots : f_{\mu-1}^*(P) : 1 : f_{\mu+1}^*(P) : \ldots : f_r^*(P))$$

is a holomorphic mapping on which we can apply Theorem (4.3) of Stoll a second time, but this time the other way around. We get that $\overline{G_{f^*}} \subset X^* \times \mathbb{P}^r$ yields a meromorphic mapping $g: X^* \to \mathbb{P}^r$. As is easily seen with the identity-lemma for meromorphic mappings, g is an extension of f. □

The following proposition shows that in Theorem 6.4 the assumption that A is thin of codimension 2 in X^* cannot be weakened. It also gives the connection between essential singularities like they are defined in Definition 2.2 and isolated singularities like they occur in the function theory of **one** complex variable:

Proposition 6.5. *Let $B \subset \mathbb{C}$ be a domain with $0 \in B$ and $f: B \setminus \{0\} \to \mathbb{C}$ be a holomorphic function. Then the zero point is an* ess 1-sing *of f if and only if it is an isolated essential singularity in the sense of function theory of one complex variable* [1]. *In this case it even is an* ess 3-sing *of f.*

Proof. Since the zero point is no ess 1-sing of f if it is a removable singularity or a pole we only have to show:

If the zero point is an isolated essential singularity then it is an ess 3-sing of f.

Assume that it is no ess 3-sing. Then there exist U, Z, h, M, and g like in Proposition 2.3. Take $y^{(1)}, y^{(2)} \in \mathbb{C}$ with $h(y^{(1)}) \neq h(y^{(2)})$. Now with the theorem of Casorati-Weierstraß there exist two sequences $(x_\nu^{(1)})_{\nu \in \mathbb{N}}, (x_\nu^{(2)})_{\nu \in \mathbb{N}}$ in $U \setminus \{0\}$ with

$$x_\nu^{(1)} \to 0 \leftarrow x_\nu^{(2)}, \qquad f(x_\nu^{(1)}) \to y^{(1)}, \qquad f(x_\nu^{(2)}) \to y^{(2)} \quad \text{for} \quad \nu \to \infty .$$

Now we have

$$g(0) = \lim_{\nu \to \infty} g(x_\nu^{(1)}) = \lim_{\nu \to \infty} h(f(x_\nu^{(1)}) = h\left(\lim_{\nu \to \infty} f(x_\nu^{(1)})\right) = h(y^{(1)})$$

$$\neq h(y^{(2)}) = h\left(\lim_{\nu \to \infty} f(x_\nu^{(2)})\right) = \lim_{\nu \to \infty} h(f(x_\nu^{(2)})) = \lim_{\nu \to \infty} g(x_\nu^{(2)}) = g(0). \ \square$$

References

1. Fischer, W., Lieb, I.: Funktionentheorie. Braunschweig: Vieweg 1985
2. Grauert, H., Fritzsche, K.: Einführung in die Funktionentheorie mehrerer Veränderlicher. Berlin Heidelberg New York: Springer 1974
3. Grauert, H., Remmert, R.: Coherent analytic sheaves. Berlin Heidelberg New York Tokyo: Springer 1984
4. Hopf, H.: Schlichte Abbildungen und lokale Modifikationen 4-dimensionaler komplexer Mannigfaltigkeiten. Comment. Math. Helv. **29**, 132–156 (1955)
5. Remmert, R.: Holomorphe und meromorphe Abbildungen komplexer Räume. Math. Ann. **133**, 328–370 (1957)
6. Stein, K.: Maximale holomorphe und meromorphe Abbildungen. II. Am. J. Math. **86**, 823–868 (1964)
7. Stein, K.: Meromorphic mappings. Enseign. Math., II. Ser. **14**, 29–46 (1968)
8. Stein, K.: Fortsetzung holomorpher Korrespondenzen. Invent. Math. **6**, 78–90 (1968)
9. Stein, K.: Topics on holomorphic correspondences. Rocky Mt. J. Math. **2**, 443–463 (1972)
10. Stoll, W.: Über meromorphe Abbildungen komplexer Räume. I. Math. Ann. **136**, 201–239 (1958)

Received July 1, 1988

Math. Ann. 283, 511–518 (1989)

Mathematische Annalen

A Characterization of Sun-Reflexivity*

B. de Pagter

Department of Mathematics, Delft University of Technology, Julianalaan 132, 2628 BL Delft, The Netherlands

1. Introduction

The duality theory for strongly continuous semigroups of bounded linear operators (i.e., C_0-semigroups) in a Banach space was initiated by Phillips in [8]. One of the difficulties in dealing with adjoint semigroups is that the adjoint semigroup $\{T^*(t)\}_{t\geqq 0}$ of a C_0-semigroup $\{T(t)\}_{t\geqq 0}$ in a Banach space X, need not be strongly continuous in X^*. However, restricting $\{T^*(t)\}_{t\geqq 0}$ to the closed subspace $X^\odot$ of X^* on which $\{T^*(t)\}_{t\geqq 0}$ is strongly continuous, we obtain a C_0-semigroup $\{T^\odot(t)\}_{t\geqq 0}$ in $X^\odot$. Now we can repeat this construction with the semigroup $\{T^\odot(t)\}_{t\geqq 0}$, and we get a C_0-semigroup $\{T^{\odot\odot}(t)\}_{t\geqq 0}$ in the Banach space $X^{\odot\odot}$ (we refer the reader to Sect. 2 for a more detailed exposition of this construction). Then it may happen that the space $X^{\odot\odot}$ coincides with X and $\{T^{\odot\odot}(t)\}_{t\geqq 0}=\{T(t)\}_{t\geqq 0}$. If this occurs we say that X is $\odot$-reflexive ("sun-reflexive") with respect to $\{T(t)\}_{t\geqq 0}$.

In recent years this duality theory for C_0-semigroups has found various applications, in particular to differential equations (e.g. second order elliptic boundary value problems, evolution equations), and $\odot$-reflexivity plays an important role in these applications (e.g. [1, 2]). It was already shown by Phillips (see also [7, Theorem 14.6.1]) that a Banach space X is $\odot$-reflexive with respect to a C_0-semigroup $\{T(t)\}_{t\geqq 0}$ if and only if the resolvent operator $R(\lambda, A)$ is $\sigma(X, X^\odot)$-compact for all λ in the resolvent set of A, where A denotes the infinitesimal generator of $\{T(t)\}_{t\geqq 0}$. This implies in particular that X is $\odot$-reflexive whenever $R(\lambda, A)$ is a weakly compact operator.

The purpose of the present paper is to show that $\odot$-reflexivity is in fact equivalent to weak compactness of $R(\lambda, A)$. Moreover, it will be shown that in certain Banach spaces (e.g. in L^1-spaces) $\odot$-reflexivity in equivalent to compactness of $R(\lambda, A)$.

* Work on this paper was supported by the Netherlands Organization for Scientific Research (NWO)

2. Preliminaries

Let $(X, \|\cdot\|)$ be a Banach space and suppose that $\{T(t)\}_{t\geqq 0}$ is a C_0-semigroup of bounded linear operators in X. The infinitesimal generator of $\{T(t)\}_{t\geqq 0}$ is denoted by A, with domain $\operatorname{dom}(A)$. We refer the reader for the basic properties of C_0-semigroups and their generators to one of the books [3, 5, 7]. For a complex number λ in the resolvent set $\varrho(A)$ of A we denote by $R(\lambda, A)$ the bounded linear operator $(\lambda I - A)^{-1}$. For convenience of the reader and to establish notation we briefly recall the duality theory for such semigroups. For details and proofs see e.g. [7, Chap. 14].

The adjoint semigroup $\{T^*(t)\}_{t\geqq 0}$ in X^* is clearly weak*-continuous, but in general it is not a C_0-semigroup in X^*. The adjoint A^* of A is a closed and weak*-densely defined operator in X^*, which is the weak*-generator of $\{T^*(t)\}_{t\geqq 0}$, i.e.,

$$\operatorname{dom}(A^*) = \left\{ x^* \in X^* : w^* - \lim_{t \downarrow 0} \frac{T^*(t)x^* - x^*}{t} \text{ exists in } X^* \right\}$$

and $\langle x, A^*x\rangle = \lim_{t\downarrow 0} t^{-1}\langle x, T^*(t)x^* - x^*\rangle$ for all $x \in X$ and $x^* \in \operatorname{dom}(A^*)$. Recall that $\varrho(A^*) = \varrho(A)$ and $R(\lambda, A^*) = R(\lambda, A)^*$ for all $\lambda \in \varrho(A)$.

Now let $X^\odot$ be the largest subspace of X^* on which $\{T^*(t)\}_{t\geqq 0}$ is strongly continuous, i.e.,

$$X^\odot = \{x^* \in X^* : \|T^*(t)x^* - x^*\| \to 0 \quad \text{as} \quad t \downarrow 0\}.$$

Clearly $X^\odot$ is a closed subspace of X^* and $T^*(t)(X^\odot) \subseteq X^\odot$ for all $t \geqq 0$. Define $T^\odot(t) = T^*(t)|_{X^\odot} : X^\odot \to X^\odot$ for all $t \geqq 0$. Then $\{T^\odot(t)\}_{t\geqq 0}$ is a C_0-semigroup. We collect some important facts concerning this "sun-dual" in the following proposition.

Proposition 2.1 (see [7, Chap. 14]). (i) $X^\odot = \overline{\operatorname{dom}(A^*)}$, *the norm closure of* $\operatorname{dom}(A^*)$ *in* X^*; *equivalently,* $X^\odot = \overline{R(\lambda, A)^*(X^*)}$ *for all* $\lambda \in \varrho(A)$.

(ii) *If we put* $\|x\|_1 = \sup\{|\langle x, x^*\rangle| : x^* \in X^\odot, \|x^*\| \leqq 1\}$ *for all* $x \in X$, *then* $\|\cdot\|_1$ *is a norm in* X *which is equivalent to the original norm in* X; *in fact,* $\|x\|_1 \leqq \|x\| \leqq M_0 \|x\|_1$ *for all* $x \in X$, *where* $M_0 = \liminf_{\lambda\to\infty} \|\lambda R(\lambda, A)\|$.

(iii) *The infinitesimal generator* $A^\odot$ *of* $\{T^\odot(t)\}_{t\geqq 0}$ *is given by* $\operatorname{dom}(A^\odot) = \{x^* \in \operatorname{dom}(A^*) : A^*x^* \in X^\odot\}$ *and* $A^\odot x^* = A^* x^*$ *for all* $x^* \in \operatorname{dom}(A^\odot)$ *(i.e.,* $A^\odot$ *is the part of* A^* *in* $X^\odot$*).*

(iv) $\varrho(A^\odot) = \varrho(A^*) = \varrho(A)$ *and* $R(\lambda, A^\odot) = R(\lambda, A)^*|_{X^\odot}$ *for all* $\lambda \in \varrho(A)$.

Note that it follows from Proposition 2.1 (i) that $\{T^*(t)\}_{t\geqq 0}$ is a C_0-semigroup in X^* if and only if $\operatorname{dom}(A^*)$ is norm dense in X^*, which is in particular the case if X is reflexive or, of course, if $\{T(t)\}_{t\geqq 0}$ is a uniformly continuous semigroup.

Now we can repeat the above procedure with X and $\{T(t)\}_{t\geqq 0}$ replaced by $X^\odot$ and $\{T^\odot(t)\}_{t\geqq 0}$ respectively. Thus $\{T^{\odot *}(t)\}_{t\geqq 0}$ is a weak*-continuous semigroup in $X^{\odot *}$ with weak*-generator $A^{\odot *}$. Furthermore, the domain of strong continuity of $\{T^{\odot *}(t)\}_{t\geqq 0}$ in $X^{\odot *}$ is $X^{\odot\odot} = \overline{\operatorname{dom}(A^{\odot *})}$ and $\{T^{\odot\odot}(t)\}_{t\geqq 0}$ is a C_0-semigroup in $X^{\odot\odot}$ with generator $A^{\odot\odot}$. Note in particular that $\varrho(A^{\odot\odot}) = \varrho(A^{\odot *}) = \varrho(A^\odot) = \varrho(A^*) = \varrho(A)$ and $R(\lambda, A^{\odot\odot}) = R(\lambda, A^\odot)^*|_{X^{\odot\odot}}$ for all $\lambda \in \varrho(A)$.

Let $j: X \to X^{**}$ be the canonical embedding of X into its bidual, and let $r_{\odot}: X^{**} \to X^{\odot *}$ be the restriction mapping, i.e., $r_{\odot}(x^{**}) = x^{**}|_{X^{\odot}}$ for all $x^{**} \in X^{**}$. Now $j_{\odot} = r_{\odot} \circ j$ is a mapping from X into $X^{\odot *}$ such that $\langle j_{\odot}(x), x^{\odot} \rangle = \langle x, x^{\odot} \rangle$ for all $x \in X$ and $x^{\odot} \in X^{\odot}$. Since

$$\|j_{\odot}(x)\| = \sup\{|\langle x, x^{\odot}\rangle| : x^{\odot} \in X^{\odot}, \|x^{\odot}\| \leqq 1\} = \|x\|_1$$

for all $x \in X$, it follows from Proposition 2.1 (ii) above that $j_{\odot}$ is a linear norm isomorphism from X onto $j_{\odot}(X)$. In particular, $j_{\odot}(X)$ is a closed subspace of $X^{\odot *}$. In general, however, $j_{\odot}$ is not an isometric isomorphism. Furthermore, it follows from

$$\langle T^{\odot *} j_{\odot} x, x^{\odot} \rangle = \langle x, T^{\odot}(t) x^{\odot} \rangle = \langle T(t)x, x^{\odot} \rangle$$

for all $x^{\odot} \in X^{\odot}$, that $T^{\odot *} j_{\odot} x = j_{\odot} T(t)x$ for all $x \in X$ and all $t \geqq 0$. Now it is clear that $j_{\odot}(X) \subseteqq X^{\odot\odot}$.

If we identify, for a moment, X with its image $j_{\odot}(X)$, then the C_0-semigroup $\{T^{\odot\odot}(t)\}_{t \geqq 0}$ is an extension of $\{T(t)\}_{t \geqq 0}$, $\operatorname{dom}(A^{\odot\odot}) \cap X = \operatorname{dom} A$, $A^{\odot\odot}x = Ax$ for all $x \in \operatorname{dom} A$ and $R(\lambda, A^{\odot\odot})_{|X} = R(\lambda, A)$ for all $\lambda \in \varrho(A)$.

Now we recall the following definition.

Definition 2.2. *The space X is called $\odot$-reflexive ("sun-reflexive") with respect to $\{T(t)\}_{t \geqq 0}$ if $j_{\odot}(X) = X^{\odot\odot}$.*

Since $X^{\odot\odot} = \overline{\operatorname{dom} A^{\odot *}}$, it is clear that X is $\odot$-reflexive with respect to $\{T(t)\}_{t \geqq 0}$ if and only if $R(\lambda, A^{\odot})^*(X^{\odot *}) \subseteqq j_{\odot}(X)$ for all $\lambda \in \varrho(A)$.

Next we will present two simple examples to illustrate the above concepts.

Examples 2.3. (i) Let $X = C_0(\mathbb{R})$, the space of all complex continuous functions f on $\mathbb{R}$ such that $\lim_{|x| \to \infty} f(x) = 0$, with the sup-norm. For $t \geqq 0$ and $f \in C_0(\mathbb{R})$ define $T(t)f(x) = f(x+t)$ for all $x \in \mathbb{R}$. Clearly $\{T(t)\}_{t \geqq 0}$ is a C_0-semigroup in $C_0(\mathbb{R})$ and the generator A is given by

$$\operatorname{dom}(A) = \{f \in C_0(\mathbb{R}) : f \text{ is differentiable and } f' \in C_0(\mathbb{R})\},$$

$Af = f'$ for all $f \in \operatorname{dom}(A)$. The dual space of $C_0(\mathbb{R})$ can be identified with the space $M_b(\mathbb{R})$ of all bounded (complex) Borel measures on $\mathbb{R}$. The adjoint of A is given by

$$\operatorname{dom}(A^*) = \{\mu \in M_b(\mathbb{R}) : D\mu \in M_b(\mathbb{R})\}, A^*\mu = -D\mu \text{ for all } \mu \in \operatorname{dom}(A^*),$$

(where $D\mu$ denotes the distributional derivative of μ). A Borel measure μ with $D\mu \in M_b(\mathbb{R})$ is absolutely continuous with respect to Lebesgue measure m. As usual, we identify the subspace of $M_b(\mathbb{R})$ consisting of all measures which are absolutely continuous with respect to m with the space $L^1(\mathbb{R}, m)$, via the Radon-Nikodym derivative. Then $\operatorname{dom}(A^*)$ consists of all functions in $L^1(\mathbb{R}, m)$ which are of bounded variation. It is now easy to see that $X^{\odot} = \overline{\operatorname{dom}(A^*)} = L^1(\mathbb{R}, m)$ and that $T^{\odot}(t)g(x) = g(x-t)$ for all $t \geqq 0$. Furthermore,

$$\operatorname{dom}(A^{\odot}) = \{g \in L^1(\mathbb{R}, m) : g \text{ is absolutely continuous and } g' \in L^1(\mathbb{R}, m)\}$$

and $A^{\odot}g = -g'$ for all $g \in \operatorname{dom}(A^{\odot})$.

As usual, the dual of $L^1(\mathbb{R}, m)$ is identified with $L^\infty(\mathbb{R}, m)$. For $f \in L^\infty(\mathbb{R}, m)$ we then have $T^{\odot *}(t)f(x) = f(x+t)$ for all $t \geqq 0$, and the weak*-generator of

$\{T^{\odot *}(t)\}_{t \geqq 0}$ is given by

$$\operatorname{dom}(A^{\odot *}) = \{f \in L^\infty(\mathbb{R}, m) : f \text{ is absolutely continuous and } f' \in L^\infty(\mathbb{R}, m)\}$$

and $A^{\odot *} f = f'$ for all $f \in \operatorname{dom}(A^{\odot *})$. Moreover, $X^{\odot\odot} = BUC(\mathbb{R})$, the space of all bounded uniformly continuous functions on $\mathbb{R}$. The generator of $\{T^{\odot\odot}(t)\}_{t \geqq 0}$ is given by

$$\operatorname{dom}(A^{\odot\odot}) = \{f \in BUC(\mathbb{R}) : f \text{ is differentiable and } f' \in BUC(\mathbb{R})\},$$

$A^{\odot\odot} f = f'$ for all $f \in \operatorname{dom}(A^{\odot\odot})$. We see that the space $C_0(\mathbb{R})$ is not $\odot$-reflexive with respect to $\{T(t)\}_{t \geqq 0}$. Note that the spectrum of A is $\varrho(A) = \{ia : a \in \mathbb{R}\}$ and that

$$R(\lambda, A) f(x) = \int_x^\infty f(\xi) e^{\lambda(x-\xi)} d\xi$$

for all $f \in C_0(\mathbb{R})$ and $\operatorname{Re} \lambda > 0$. It is easy to verify that $R(\lambda, A)$ is not weakly compact (see also the comments at the end of this paper).

(ii) Let X be the Banach space $C(S^1)$ of all continuous functions on the unit circle S^1. For $t \geqq 0$ and $f \in C(S^1)$ we define $T(t) f(\theta) = f(\theta + t)$ (as usual we write $f(\theta)$ for $f(e^{i\theta})$). The dual of $C(S^1)$ can be identified with the space $M(S^1)$ of all Borel measures on S^1. As in (i) we find that $X^\odot = L^1(S^1, m)$ (where m denotes normalized Lebesgue measure on S^1) and $T^\odot(t) g(\theta) = g(\theta - t)$ for all $t \geqq 0$ and all $g \in L^1(S^1, m)$. The adjoint semigroup $\{T^{\odot *}(t)\}_{t \geqq 0}$ in $L^\infty(S^1, m)$ satisfies $T^{\odot *}(t) f(\theta) = f(\theta + t)$ for all $f \in L^\infty(S^1, m)$. It is now clear that $X^{\odot\odot} = C(S^1)$, and hence $C(S^1)$ is $\odot$-reflexive with respect to the C_0-semigroup $\{T(t)\}_{t \geqq 0}$. We note already that in this situation the resolvent operator $R(\lambda, A)$ of the generator A is in fact compact for all $\lambda \in \varrho(A)$.

3. A Characterization of $\odot$-Reflexivity

First we recall some relevant facts concerning weakly compact operators. Let $\mathscr{L}(X, Y)$ denote the Banach space of all bounded linear operators from Banach space X into Banach space Y. An operator $T \in \mathscr{L}(X, Y)$ is called weakly compact if the image $T(B_X)$ of the closed unit ball B_X is relatively weakly compact in Y, i.e., if $\overline{T(B_X)}$ is weakly compact (note that the norm and weak closure of the convex set $T(B_X)$ coincide). If $T \in \mathscr{L}(X, Y)$, then T is weakly compact if and only if $T^{**}(X^{**})$ is contained in Y (identifying Y with its canonical image in Y^{**}). Furthermore, the set of all weakly compact operators from a Banach space X into itself is a norm closed two-sided ideal in $\mathscr{L}(X)$. The proofs of these well-known results can be found in e.g. [5, Sect. VI.4].

Now let $\{T(t)\}_{t \geqq 0}$ be a C_0-semigroup in the Banach space X with generator A. It is useful to note that for any $\lambda \in \varrho(A)$ we have the following commutative diagram

$$\begin{tikzcd}
 & X^{**} \arrow[r, "R(\lambda, A)^{**}"] \arrow[dd, "r_\odot"] & X^{**} \arrow[dd, "r_\odot"] \\
X \arrow[ur, "j"] \arrow[dr, "j_\odot"'] & & \\
 & X^{\odot *} \arrow[r, "R(\lambda, A^\odot)^*"'] & X^{\odot *}
\end{tikzcd}$$

(where the mappings j, $j_{\odot}$, and $r_{\odot}$ are as introduced in the previous section). Furthermore observe that it is immediate from the resolvent equation that $R(\lambda, A)$ is (weakly) compact for all $\lambda \in \varrho(A)$ if and only if $R(\lambda, A)$ is (weakly) compact for some $\lambda \in \varrho(A)$. The proof of the following proposition is now simple.

Proposition 3.1 (cf. [7, Corollary to Theorem 14.6.1]). *If $R(\lambda, A)$ is weakly compact for $\lambda \in \varrho(A)$, then X is $\odot$-reflexive.*

Proof. Take $\lambda \in \varrho(A)$. If $R(\lambda, A)$ is weakly compact, then $R(\lambda, A)^{**}(X^{**}) \subseteqq j(X)$. Since the restriction mapping $r_{\odot}$ is surjective, it follows from the commutativity of the above diagram that $R(\lambda, A^{\odot})^{*}(X^{\odot *}) \subseteqq j_{\odot}(X)$. As observed in Sect. 2, this implies that X is $\odot$-reflexive with respect to $\{T(t)\}_{t \geqq 0}$. □

As mentioned in the introduction, it is shown in [7, Theorem 14.6.1], that $\odot$-reflexivity is equivalent to $\sigma(X, X^{\odot})$-compactness of $R(\lambda, A)$. Since weak compactness of $R(\lambda, A)$ clearly implies that $R(\lambda, A)$ is $\sigma(X, X^{\odot})$-compact, the above proposition is an immediate consequence of this result. The direct proof above is included for the reader's convenience. Our next objective is to show that $\odot$-reflexivity of X is in fact *equivalent* to weak compactness of $R(\lambda, A)$. The proof is divided into three lemmas.

Lemma 3.2. *Let Y be a closed subspace of a Banach space Z. We denote by $i: Y \to Z$ the inclusion mapping (so $i^*: Z^* \to Y^*$ is the restriction mapping). Let W be a closed subspace of Z^* and put $W_1 = i^*(W)$. Now suppose that $T: Z \to Z$ is a bounded linear operator which satisfies the following three conditions:* (i) $T(Z) \subseteqq Y$; (ii) $T^*(W) \subseteqq W$; (iii) $T_0^*(Y^*) \subseteqq W_1$, *where $T_0 = T_{|Y}: Y \to Y$.*

Then $(T^*)^2(Z^*) \subseteqq W$.

Proof. Defining $T_1: Z \to Y$ by $T_1 z = Tz$ for all $z \in Z$ we clearly get the commutative diagram

$$\begin{tikzcd}
Z \arrow[r, "T"] \arrow[dr, "T_1"] & Z \arrow[r, "T"] & Z \\
Y \arrow[u, "i"] \arrow[r, "T_0"'] & Y \arrow[u, "i"] \arrow[r, "T_0"'] & Y \arrow[u, "i"']
\end{tikzcd}$$

and so by taking adjoints we find the commutative diagram

$$\begin{tikzcd}
Z^* \arrow[r, "T^*"] \arrow[d, "i^*"'] & Z^* \arrow[r, "T^*"] \arrow[d, "i^*"] & Z^* \arrow[d, "i^*"] \\
Y^* \arrow[r, "T_0^*"'] & Y^* \arrow[ur, "T_1^*"] \arrow[r, "T_0^*"'] & Y^*
\end{tikzcd}$$

Hence, $(T^*)^2(Z^*) = T_1^* \circ T_0^* \circ i^*(Z^*) = T_1^* \circ T_0^*(Y^*) \subseteqq T_1^*(W_1) = T_1^* \circ i^*(W)$
$= T^*(W) \subseteqq W$. □

Lemma 3.3. *If the Banach space X is $\odot$-reflexive with respect to the C_0-semigroup $\{T(t)\}_{t \geqq 0}$, then $R(\lambda, A)^2$ is weakly compact for all $\lambda \in \varrho(A)$.*

Proof. Take in the above lemma $Z=X^*$, $Y=X^{\odot}$, $W=j(X)$ and $T=R(\lambda,A)^*$. Note that $W_1=j_{\odot}(X)$ and $T_0=R(\lambda,A^{\odot})$. It follows now from the $\odot$-reflexivity of X that $R(\lambda,A^{\odot})^*(X^{\odot *})\subseteq j_{\odot}(X)$, i.e., that $T_0^*(Y^*)\subseteq W_1$. Hence, we may conclude that $(T^*)^2(Z^*)\subseteq W$, and so $[R(\lambda,A)^2]^{**}(X^{**})\subseteq j(X)$, which shows that $R(\lambda,A)^2$ is weakly compact. □

Lemma 3.4. *If $\{T(t)\}_{t\geqq 0}$ is a C_0-semigroup in the Banach space X with generator A, then $\|[\lambda R(\lambda,A)]^2R(\mu,A)-R(\mu,A)\|\to 0$ as $\lambda\to\infty$ for all $\mu\in\varrho(A)$.*

Proof. Fix $\mu\in\varrho(A)$ and let $M\geqq 1$ be such that $\|\lambda R(\lambda,A)\|\leqq M$ for all $\lambda\geqq\lambda_0$, for some $\lambda_0\in\mathbb{R}_+$. For $\lambda\neq\mu$ we have

$$\begin{aligned}\lambda R(\lambda,A)R(\mu,A)-R(\mu,A)&=\frac{\lambda}{\lambda-\mu}\{R(\mu,A)-R(\lambda,A)\}-R(\mu,A)\\&=\frac{1}{\lambda-\mu}\{\mu R(\mu,A)-\lambda R(\lambda,A)\},\end{aligned}$$

and hence

$$\|\lambda R(\lambda,A)R(\mu,A)-R(\mu,A)\|\leqq\frac{1}{\lambda-|\mu|}(\|\mu R(\mu,A)\|+M)$$

for all $\lambda>\max(\lambda_0,|\mu|)$. This shows that $\|\lambda R(\lambda,A)R(\mu,A)-R(\mu,A)\|\to 0$ as $\lambda\to\infty$. Now

$$\begin{aligned}&\|[\lambda R(\lambda,A)]^2R(\mu,A)-R(\mu,A)\|\\&\quad\leqq\|\lambda R(\lambda,A)[\lambda R(\lambda,A)R(\mu,A)-R(\mu,A)]\|+\|\lambda R(\lambda,A)R(\mu,A)-R(\mu,A)\|\\&\quad\leqq(M+1)\|\lambda R(\lambda,A)R(\mu,A)-R(\mu,A)\|\end{aligned}$$

for all $\lambda\geqq\lambda_0$, by which the lemma is proved. □

We now formulate the main result of the paper.

Theorem 3.5. *Given a C_0-semigroup $\{T(t)\}_{t\geqq 0}$ in the Banach space X with generator A, the following two statements are equivalent.*

(i) *X is $\odot$-reflexive with respect to $\{T(t)\}_{t\geqq 0}$.*

(ii) *$R(\lambda,A)$ is weakly compact for $\lambda\in\varrho(A)$.*

Proof. Assume that X is $\odot$-reflexive. It follows from Lemma 3.3 that $R(\lambda,A)^2$ is weakly compact for all $\lambda\in\varrho(A)$. Since the set of weakly compact operators is a closed two-sided ideal in $\mathscr{L}(X)$, Lemma 3.4 now implies that $R(\lambda,A)$ is weakly compact for all $\lambda\in\varrho(A)$. The converse implication is Proposition 3.1. □

In certain Banach spaces the result of the above theorem can be strengthened. For this purpose, recall that a Banach space X has the *Dunford-Pettis property* if every weakly compact operator from X into any Banach space Y maps weakly compact subsets of X onto compact subsets of Y (see e.g. Sect. II.9 in the book [9]). Clearly, if X has the Dunford-Pettis property and $T\in\mathscr{L}(X)$ is weakly compact, then T^2 is a compact operator.

Corollary 3.6. *Suppose that X is a Banach space with the Dunford-Pettis property, and let $\{T(t)\}_{t\geqq 0}$ be a C_0-semigroup in X with generator A. The following statements are equivalent.*

(i) *X is $\odot$-reflexive with respect to $\{T(t)\}_{t\geqq 0}$.*

(ii) *$R(\lambda, A)$ is compact for $\lambda \in \varrho(A)$.*

Proof. Only (i) $\Rightarrow$ (ii) needs proof. Suppose that X is $\odot$-reflexive. It follows from Theorem 3.5 that $R(\lambda, A)$ is weakly compact for all $\lambda \in \varrho(A)$. Since X has the Dunford-Pettis property, this implies that $R(\lambda, A)^2$ is compact for all $\lambda \in \varrho(A)$. Now it is a simple consequence of Lemma 3.4 that $R(\lambda, A)$ is compact for all $\lambda \in \varrho(A)$. □

We mention two important examples of Banach spaces to which the result of the corollary applies.

1) For any σ-finite measure space (Ω, Σ, μ), the space $X = L^1(\Omega, \mu)$ has the Dunford-Pettis property (this is the classical result of Dunford and Pettis [4]).

2) For any locally compact Hausdorff space Ω, the Banach space $C_0(\Omega)$ of all continuous functions on Ω vanishing at infinity (with sup-norm) has the Dunford-Pettis property. In particular, the Banach space $C(\Omega)$ of all continuous functions on a compact space Ω, has the Dunford-Pettis property (these results go back to Grothendieck [6]).

We end this paper by mentioning a criterion for weak compactness of $R(\lambda, A)$. Let A be a closed and densely defined linear operator in the Banach space X with $\varrho(A) \neq \phi$. For $x \in \mathrm{dom}(A)$ define $\|x\|_A = \|x\| + \|Ax\|$. Then $(\mathrm{dom}(A), \|\cdot\|_A)$ is a Banach space with unit ball $B_A = \{x \in \mathrm{dom}(A) : \|x\| + \|Ax\| \leqq 1\}$. Given $\lambda \in \varrho(A)$ it is easy to verify that

$$(\|R(\lambda, A)\| + \|AR(\lambda, A)\|)^{-1} R(\lambda, A)(B_X) \subseteqq B_A \subseteqq \max(|\lambda|, 1) R(\lambda, A)(B_X),$$

so $R(\lambda, A)(B_X)$ is relatively weakly compact if and only if B_A is a relatively weakly compact subset of X. Furthermore, by the Eberlein-Smulian theorem (see e.g. [5, Theorem V.6.1]), a subset S of X is relatively weakly compact if and only if S is relatively sequentially weakly compact. Combining these observations we get the equivalence of the following three statements:

(1) $R(\lambda, A)$ is a weakly compact operator for $\lambda \in \varrho(A)$;

(2) the embedding of $(\mathrm{dom}(A), \|\cdot\|_A)$ into X is weakly compact;

(3) any sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathrm{dom}(A)$ with $\sup\limits_n \|x_n\| < \infty$ and $\sup\limits_n \|Ax_n\| < \infty$ has a subsequence which is weakly convergent to an element in X.

References

1. Amann, H.: Dual semigroups and second order linear elliptic boundary value problems. Isr. J. Math. **45**, 225–254 (1983)
2. Clément, Ph., Diekmann, O., Gyllenberg, M., Heijmans, H.J.M., Thieme, H.R.: Perturbation theory for dual semigroup. I. The sun-reflexive case. Math. Ann. **277**, 709–725 (1987)
3. Davies, E.B.: One-parameter semigroups. London: Academic Press 1980
4. Dunford, N., Pettis, B.J.: Linear operations on summable functions. Trans. Amer. Math. Soc. **47**, 323–392 (1940)
5. Dunford, N., Schwartz, J.T.: Linear operators, Part I: general theory. New York: Interscience 1958

6. Grothendieck, A.: Sur les applications linéaires faiblement compactes d'espaces du type $C(K)$. Canad. J. Math. **5**, 129–173 (1953)
7. Hille, E., Phillips, R.S.: Functional analysis and semigroups. Providence: Amer. Math. Soc. 1957
8. Phillips, R.S.: The adjoint semi-group. Pac. J. Math. **5**, 269–283 (1955)
9. Schaefer, H.H.: Banach lattices and positive operators. Berlin Heidelberg New York: Springer 1974

Received September 2, 1988

Math. Ann. 283, 519–528 (1989)

Mathematische Annalen

Cohomologie des groupes et corps d'invariants multiplicatifs

Jean Barge

Centre Mathématiques, Ecole Polytechnique, F 91 128, Palaiseau, France et Université de Lausanne, Département de Mathématiques, CH-1015 Lausanne-Dorigny, Suisse

Introduction

Soit Γ un groupe fini opérant linéairement et fidèlement sur un espace vectoriel complexe V, de dimension finie. Le problème de savoir si le corps des invariants, $\mathbb{C}(V)^\Gamma$ est transcendant pur sur $\mathbb{C}$ – ou même stablement transcendant pur – a reçu une réponse négative. C'est Saltman [S 1] [S 2] qui, le premier, a donné un contre exemple. L'invariant qu'il utilise est le groupe de Brauer non ramifié, $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(V)^\Gamma)$. C'est un sous-groupe du groupe de Brauer, nul pour les corps transcendants purs sur $\mathbb{C}$, et Saltman a construit un exemple de groupe fini Γ et de représentation fidèle V, tels que $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(V)^\Gamma)$ soit non nul.

Un peu plus tard, Saltman et Bogomolov indépendemment ont calculé $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(V)^\Gamma)$ pour tout Γ.

Théorème 1 [B].

$$\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(V)^\Gamma) = \mathrm{Ker}[H^2(\Gamma;\mathbb{Q}/\mathbb{Z}) \to \Pi_A H^2(A;\mathbb{Q}/\mathbb{Z})]$$

où A parcourt tous les sous groupes abéliens de Γ.

(Voir aussi [CTS]) En particulier cet invariant ne dépend que de Γ et non pas de la représentation linéaire dès qu'elle est fidèle.

Un problème voisin et étroitement lié à la question précédente est le suivant: Soient G un groupe fini et M un G – réseau – i.e. un $\mathbb{Z}$ module libre et de type fini sur lequel G agit fidèlement. On note $\mathbb{C}[M]$ l'anneau du groupe abélien M à coefficients dans $\mathbb{C}$ et $\mathbb{C}(M)$ son corps de fractions. Calculer $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G)$. La réponse a été donnée par Saltman [S 3] en 87.

Théorème 2 [S 3].

$$\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G) = \mathrm{Ker}[H^2(G;\mathbb{Q}/\mathbb{Z} \oplus M) \to \Pi_B H^2(B;\mathbb{Q}/\mathbb{Z} \oplus M)].$$

où cette fois-ci B parcourt les sous-groupes abéliens bicycliques de G. – Un groupe est abélien bicyclique si c'est un quotient de $\mathbb{Z}^2$ –.

Le lien entre les deux problèmes (action linéaire et G-réseaux) est le suivant: Supposons que le G-réseau M soit d'indice fini dans un module de permutation P de sorte que l'on ait la suite exacte $0 \to M \to P \to F \to 0$. On construit alors Γ, le produit semi-direct de G par le groupe des caractères de F, et une action linéaire fidèle de Γ sur $\mathbb{C}(P)$ de sorte que $\mathbb{C}(P)^\Gamma = \mathbb{C}(M)^G$ (1).

Un des buts de cet article est de donner une preuve directe du théorème 2, à partir du théorème 1, grâce à l'égalité (1). Les méthodes sont purement cohomologiques et, une fois admis le théorème 1, cet article est "self contained". Le point clé est la preuve du cas particulier suivant:

Théorème 3. *Soit Γ le produit semi-direct d'un groupe bicyclique B, par un B-module fini. Alors l'application*

$$H^2(\Gamma; \mathbb{Q}/\mathbb{Z}) \to \Pi_A H^2(A; \mathbb{Q}/\mathbb{Z})$$

où A parcourt les sous-groupes abéliens de Γ, est injective.

Le théorème ci-dessus signifie (par dualité) que pour un tel Γ, toute l'homologie $H_2(\Gamma; \mathbb{Z})$ est engendrée par des classes toriques et la preuve que j'en donne s'inspire de [BG]. La démonstration repose sur le fait que $\mathbb{Z}^2$ est un groupe à dualité [Bi], et elle rent évident le rôle des groupes abéliens bicycliques. D'ailleurs on aboutit facilement à caractérisation suivante:

Théorème 4. *Soit G un groupe fini. Les conditions suivantes sont équivalentes.*

1) *Pour tout G-réseau M,* $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G) = 0$.
2) *Les sous-groupes de Sylow de G sont abéliens bicycliques.*

En fait tout exemple de G-réseau M pour lequel $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G) \neq 0$ conduit "par linéarisation" à un contre exemple au problème de Noether. C'est la méthode de Saltman [S 2] qui "démarrait" avec le plus petit groupe violant 2), à savoir $(\mathbb{Z}/2)^3$.

La première partie est consacrée à la preuve du théorème 3. Dans la deuxième nous démontrons le théorème 4. En passant, nous obtenons une preuve du théorème 2 dans le cas particulier où M est d'indice fini dans un module de permutation.

I

Cette partie est consacrée à la démonstration du théorème

Théorème I.1. *Soit B un groupe abélien bicyclique et soit F un B-module fini. Soit Γ le produit semi-direct de B par F. Alors l'application*

$$H^2(\Gamma; \mathbb{Q}/\mathbb{Z}) \to \Pi_A H^2(A; \mathbb{Q}/\mathbb{Z})$$

où A parcourt les sous-groupes abéliens de Γ, est injective.

Remarquons d'abord que l'énoncé précédent est équivalent – par la formule des coefficients universels – à celui-ci.

L'application

$$\bigoplus_A H_2(A; \mathbb{Z}) \to H_2(\Gamma; \mathbb{Z})$$

est surjective. Par ailleurs puisque $H_2(A;\mathbb{Z}) \simeq \wedge_2(A)$ on peut se restreindre aux sous-groupes A qui sont abéliens bicycliques.

Définition I.2. Nous disons qu'une classe x de $H_2(\Gamma;\mathbb{Z})$ est torique, s'il existe un homomorphisme $f: \mathbb{Z}\times\mathbb{Z}\to\Gamma$ tel que $x=f_*(u)$ où u est un générateur (choisi une fois pour toutes) de $H_2(\mathbb{Z}^2;\mathbb{Z})\simeq\mathbb{Z}$.

Le théorème I.2 est donc équivalent à la:

Proposition I.3. *Soit Γ comme dans le théorème I.1. Le groupe $H_2(\Gamma;\mathbb{Z})$ est engendré par des classes toriques.*

I.4. Construction d'un groupe universel U.

Notons U le produit semi-direct du groupe $\mathbb{Z}\times\mathbb{Z}$, par l'idéal d'augmentation, I, de son algèbre de groupe et choisissons deux générateurs X_1, X_2 du groupe $\mathbb{Z}\times\mathbb{Z}$ ce qui nous permet d'identifier I à l'idéal de l'anneau des polynômes de Laurent à deux variables engendré par X_1-1 et X_2-1.

Proposition 1.5. *Si $H_2(U;\mathbb{Z})$ est engendré par des classes toriques, il en est de même pour $H_2(\Gamma;\mathbb{Z})$ où Γ est comme dans le théorème I.1.*

Preuve. Soit donc

$$0\to F\xrightarrow{i}\Gamma\to B\to 1 \tag{1}$$

une suite exacte scindée, où B est bicyclique et F fini et soit

$$0\to H_1(B;F)\xrightarrow{\phi}\frac{H_2(\Gamma;\mathbb{Z})}{i_*(H_2(F;\mathbb{Z}))}\to H_2(B;\mathbb{Z})\to 0 \tag{2}$$

la suite exacte "des termes de bas degré" dans la suite spectrale de Hochschild-Serre.

Puisque (1) est scindée, (2) aussi et en conséquence toute classe de $H_2(\Gamma;\mathbb{Z})$ est congrue, modulo les classes toriques, à un élément de l'image de ϕ (puisque B, comme F est abélien).

Soient u_1, u_2, deux générateurs de B. La suite

$$\mathbb{Z}[B]^2\xrightarrow{d}\mathbb{Z}[B]\xrightarrow{\varepsilon}\mathbb{Z}\to 0$$

où $d(a_1, a_2)=u_1\cdot a_1-a_1+u_2\cdot a_2-a_2$ est un début de résolution libre de $\mathbb{Z}$. Il en résulte que tout élément α de $H_1(B;F)$ est la classe d'une chaine $f_1, f_2\in F\times F$ qui est un cycle, c'est-à-dire telle que

$$u_1\cdot f_1-f_1+u_2\cdot f_2-f_2=0$$

Condidérons alors le diagramme commutatif ci-dessous.

$$\begin{tikzcd}
0 \arrow[r] & I \arrow[r] \arrow[d, "h"] & U \arrow[r] \arrow[d, dashed, "k"] & \mathbb{Z}\times\mathbb{Z} \arrow[r] \arrow[d, "l"] & 0 \\
0 \arrow[r] & F \arrow[r] & \Gamma \arrow[r] & B \arrow[r] & 0
\end{tikzcd}$$

où $l(X_1)=u_1$, $l(X_2)=u_2$ et où h est l'unique application $\mathbb{Z}\times\mathbb{Z}$ équivariante telle que $h(X_1-1)=f_2$ et $h(X_2-1)=-f_1$.

Le diagramme se complète par un homomorphisme (et même plusieurs) $k: U \to \Gamma$ (car il s'agit de produits semi-directs). Il induit en homologie le diagramme commutatif suivant:

$$\begin{array}{ccc} H_1(\mathbb{Z}\times\mathbb{Z}); I) & \xrightarrow{\phi} & \dfrac{H_2(U;\mathbb{Z})}{i_*(H_2(I;\mathbb{Z}))} \\ \Big\downarrow {\scriptstyle (h,l)_*} & & \Big\downarrow {\scriptstyle k_*} \\ H_1(B;F) & \xrightarrow{\phi} & \dfrac{H_2(\Gamma;\mathbb{Z})}{i_*(H_2(F;\mathbb{Z}))} \end{array} .$$

Soit alors $\beta \in H_1(\mathbb{Z}\times\mathbb{Z}; I)$ la classe de $(1-X_2, X_1-1)$ qui est bien un cycle puisque $X_1\cdot(X_2-1)-(X_2-1)=X_2(X_1-1)-(X_1-1)$.

Il est clair que $(h,l)_*(\beta)=\alpha$ et donc que $\phi(\alpha)=k_*(\phi(\beta))$. Il suffit donc de montrer que $\phi(\beta)$ est engendré par des classes toriques. $\square$

La démonstration du théorème I.1 résulte donc maintenant de

Proposition I.6. *Soit U notre groupe universel.*
Alors $H_2(U;\mathbb{Z})$ est engendré par des classes toriques.

Pour démontrer la proposition I.6 nous utiliserons le lemme ci-dessous dont la démonstration est donnée en appendice.

Lemme I.7. *Soit*

$$0 \to A \xrightarrow{i} \Gamma \to G \to 1 \tag{1}$$

une suite exacte scindée où A est abélien qui induit en homologie la suite exacte

$$0 \to H_1(G;A) \xrightarrow{\phi} \frac{H_2(\Gamma;\mathbb{Z})}{i_*(H_2(A;\mathbb{Z}))} \to H_2(G;\mathbb{Z}) \to 0$$

Soient $x \in H_2(G;\mathbb{Z})$ et s_1, s_2 deux sections de (1). On note $d(s_1,s_2)$ la classe dans $H^1(G;A)$ de la différence s_2-s_1. Alors on a la formule

$$(s_2)_*(x)-(s_1)_*(x)=\phi[d(s_1,s_2)\cap x].$$

Preuve de I.6. On applique la formule ci-cessus à l'extension scindée

$$0 \to I \to U \to \mathbb{Z}\times\mathbb{Z} \to 1$$

en choisissant pour x "la classe fondamentale" du tore notée $u \in H_2(\mathbb{Z}\times\mathbb{Z};\mathbb{Z})$: Le cap produit par u,

$$\bigcap u : H^1(\mathbb{Z}\times\mathbb{Z}; I) \to H_1(\mathbb{Z}\times\mathbb{Z}; I)$$

est alors un isomorphisme (puisque $\mathbb{Z}\times\mathbb{Z}$ est un groupe à dualité de dimension 2 [Bi]). Il en résulte que toute classe de l'image de ϕ est égale (modulo des classes toriques) à la différence $(s_2)_*(u)-(s_1)_*(u)$ – pour $d(s_1,s_2)$ convenablement choisie –. Ces deux classes, comme u, sont toriques. $\square$

Remarque I.8. Le groupe U est en fait un quotient du groupe fondamental de la surface de genre 2 puisqu'on a $[X_1, X_2-1]=[X_2, X_1-1]$ et la classe d'homologie

$\phi(\beta)$ de $H_2(U;\mathbb{Z})$ est l'image de "la" classe d'orientation de cette surface. Il résultait alors de la preuve de la proposition I.5 que toute classe de $H_2(\Gamma;\mathbb{Z})$ à défaut d'être engendrée par des classes toriques (comme nous l'avons finalement démontré) était à priori engendré par des classes de genre 2.

Remarque I.9. Pour revenir au problème initial nous avons en fait montré que toute extension centrale non triviale de Γ par $\mathbb{Q}/\mathbb{Z}$ se détecte par pull-back sur un des trois sous-groupes abéliens de Γ suivants:

Tout d'abord F, puis un relevé quelconque de B, enfin l'image dans Γ d'un relevé bien choisi, non par de B, mais de $\mathbb{Z}\times\mathbb{Z}$ dans le pull-back de l'extension initiale par l'application $l:\mathbb{Z}\times\mathbb{Z}\to B$.

II. Applications

II.1. Soient G un groupe fini et M un G-réseau qui est d'indice fini dans un G-module permutation P.

Comme mentionné dans l'introduction si F est le conoyau de M dans P on note Γ le produit semi-direct de G par le groupe des caractères $\hat{F}$ de F. Soit $V=P\underset{\mathbb{Z}}{\otimes}\mathbb{C}$. Le corps $\mathbb{C}(P)$ s'identifie au corps des fonctions rationnelles sur V et le groupe Γ agit fidèlement et linéairement sur $\mathbb{C}(V)=\mathbb{C}(P)$ de sorte que $\mathbb{C}(P)^\Gamma=\mathbb{C}(M)^G$ et on a donc

$$\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(P)^\Gamma)=\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G)\quad (1)\,[\mathrm{S}.2]$$

Proposition II.2. *Soit B un groupe bicyclique et soit M un B réseau. Alors* $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^B)=0$.

Preuve. Si M est d'indice fini dans un B-module de permutation on applique l'égalité (1), puis le thèorème I de l'introduction, puis le théorème I.1.

Le cas général (où M n'est plus nécessairement d'indice fini dans un module de permutation) résulte des deux faits suivants:

Fait 1. Soit M un G-réseau. Alors il existe un G-réseau N tel que $M\oplus N$ soit d'indice fini dans un module de permutation.

Fait 2. Si M et N sont deux G-réseaux le groupe $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G)$ s'injecte dans $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M\oplus N)^G)$.

Nous laissons la démonstration du fait 1 en exercice.

Preuve du Fait 2. On observe d'abord que puisque $\mathbb{C}(M)$ est transcendant pur sur $\mathbb{C}$, le groupe $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G)$ est en fait un sous-groupe du groupe de Brauer relatif, à savoir $H^2(G,\mathbb{C}(M)^*)$. Il reste à voir que $H^2(G,\mathbb{C}(M)^*)$ s'injecte dans $H^2(G,\mathbb{C}(M\oplus N)^*)$.

Pour cela considérons les suites d'injection suivantes:

$$\mathbb{C}(M)^*\overset{i_1}{\hookrightarrow}(\mathbb{C}(M)[N])^*\overset{i_2}{\hookrightarrow}(\mathbb{C}(M)(N))^*=(\mathbb{C}(M\oplus N))^*$$

La première injection i_1 admet une G-rétraction (induite par l'homomorphisme de N dans le groupe trivial) et donc induit une injection en cohomologie. Quand à la deuxième i_2, son conoyau s'identifie au $\oplus_I\mathbb{Z}$ où I parcourt les idéaux principaux de l'anneau factoriel $\mathbb{C}(M)[N]$. Ces idéaux sont permutés par G et donc $H^1(G, \oplus_I \mathbb{Z})=0$. Ce qui montre que $H^2(G; (\mathbb{C}(M)[N])^*)$ s'injecte dans $H^2(G; (\mathbb{C}(M \oplus N))^*)$. $\square$

Théorème II.3 [S 3]. *Soient G un groupe fini et M un G-réseau d'indice fini dans un G-module de permutation.*

Alors

$$\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G) = \mathrm{Ker}[H^2(G; \mathbb{Q}/\mathbb{Z} \oplus M) \to \Pi_B H^2(B, \mathbb{Q}/\mathbb{Z} \oplus M)]$$

où B parcourt les sous-groupes abéliens bicycliques de G.

En appliquant le théorème I, le théorème II.3 se réduit, grâce à II.1) à la proposition suivante:

Proposition II.4 *Soit* $0 \to M \to P \to F \to 0$ *une suite exacte de G-modules, où P est de permutation et F fini. Soit* Γ *le produit semi-direct de G par le groupe des caractères* $\hat{F}$, *de F. Alors les deux groupes*

$$\mathrm{Ker}[H^2(G; \mathbb{Q}/\mathbb{Z} \oplus M) \to \Pi_B H^2(B; \mathbb{Q}/\mathbb{Z} \oplus M)]$$

et

$$\mathrm{Ker}[H^2(\Gamma; \mathbb{Q}/\mathbb{Z}) \to \Pi_A H^2(A; \mathbb{Q}/\mathbb{Z})]$$

où B parcourt les sous-groupes abéliens bicycliques de G et A les sous-groupes abéliens ou abéliens bicycliques de Γ, *sont isomorphes.*

Introduisons une notation: Si G est un groupe fini et M un G-module nous notons $K^i(G; M)$ le noyau de l'application

$$H^i(G; M) \to \Pi_B H^i(B; M)$$

où B parcourt les sous-groupes abéliens bicycliques de G. Ainsi la proposition II.4 signifie-t-elle que

$$K^2(\Gamma; \mathbb{Q}/\mathbb{Z}) \simeq K^2(G; \mathbb{Q}/\mathbb{Z}) \oplus K^2(G; M).$$

Lemme II.5. *Avec les notations précédentes,* $K^2(G; M)$ *est canoniquement isomorphe à* $K^1(G; F)$.

Preuve. Il suffit de voir que $K^1(G; P) = K^2(G; P) = 0$. C'est clair pour $K^1(G; P)$ qui est contenu dans $H^1(G; P)$ qui est nul. Par ailleurs si $P \simeq \bigoplus_H \mathbb{Z}[G/H]$ on a la suite d'isomorphismes.

$$H^2(G; P) \simeq \bigoplus_H H^2(G; \mathbb{Z}[G/H]) \simeq \bigoplus_H H^2(H; \mathbb{Z}) \simeq \bigoplus_H H^1(H; \mathbb{Q}/\mathbb{Z})$$

et un homomorphisme de H dans $\mathbb{Q}/\mathbb{Z}$, nul sur les sous-groupes cycliques est trivial; d'où la nullité de $K^2(G; P)$.

Preuve de la Proposition II.4. Il s'agit donc de montrer que $K^2(\Gamma; \mathbb{Q}/\mathbb{Z})$ est isomorphe à

$$K^2(G; \mathbb{Q}/\mathbb{Z}) \oplus K^1(G; F).$$

Considérons le diagramme de suites exactes:

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
 & K^2(G;\mathbb{Q}/\mathbb{Z}) \arrow[d] & K^2(\Gamma;\mathbb{Q}/\mathbb{Z}) \arrow[d] & K^1(G;F) \arrow[d] & \\
0 \arrow[r] & H^2(G;\mathbb{Q}/\mathbb{Z}) \arrow[r, "p^*"] \arrow[d] & \operatorname{Ker}[H^2(\Gamma;\mathbb{Q}/\mathbb{Z}) \to H^2(F;\mathbb{Q}/\mathbb{Z})] \arrow[r, "\psi"] \arrow[d] & H^1(G;F) \arrow[r] \arrow[d] & 0 \\
 & \Pi_B H^2(B;\mathbb{Q}/\mathbb{Z}) & \Pi_A H^2(A;\mathbb{Q}/\mathbb{Z}) & \Pi_B H^1(B;F) &
\end{tikzcd}$$

où la suite horizontale est la suite «des termes de bas degré» dans la suite spectrale de Hochschild-Serre, et où A, B parcourent respectivement les sous-groupes abéliens bicycliques de Γ et de G.

Nous allons montrer que ce diagramme induit la suite exacte scindée

$$0 \to K^2(G;\mathbb{Q}/\mathbb{Z}) \to K^2(\Gamma;\mathbb{Q}/\mathbb{Z}) \to K^1(G;F) \to 0$$

Il a deux points non évidents et nous laissons les autres vérifications au lecteur.

(a) $$\psi(K^2(\Gamma;\mathbb{Q}/\mathbb{Z}) \subset K^1(G;F)$$

Preuve. Soit B un sous-groupe abélien bicyclique de G.

Considérons le diagramme

$$\begin{tikzcd}
0 \arrow[r] & \hat{F} \arrow[r] \arrow[d, equal] & \Gamma \arrow[r, "p"] & G \arrow[r] & 1 \\
0 \arrow[r] & \hat{F} \arrow[r] & p^{-1}(B) \arrow[r] \arrow[u, "i"'] & B \arrow[r] \arrow[u, "j"'] & 1
\end{tikzcd}$$

et soit $x \in K^2(\Gamma;\mathbb{Q}/\mathbb{Z})$.

Par fonctorialité $j^*(\psi(x)) = \psi(i^*(x))$ mais $i^*(x) \in K^2(p^{-1}(B);\mathbb{Q}/\mathbb{Z})$ qui est nul (théorème I.1).

(b) $$\psi(K^2(\Gamma;\mathbb{Q}/\mathbb{Z})) = K^1(G;F)$$

Preuve. Soit A un sous-groupe bicyclique de Γ.

Considérons le diagramme

$$\begin{tikzcd}
0 \arrow[r] & \hat{F} \arrow[r] \arrow[d, equal] & \Gamma \arrow[r, shift right, "p"'] & G \arrow[l, shift right, "s"'] \arrow[r] & 1 \\
0 \arrow[r] & \hat{F} \arrow[r] & p^{-1}(p(A)) \arrow[r, shift right] \arrow[u, hook, "i"'] & p(A) \arrow[l, shift right, "s"'] \arrow[r] \arrow[u, hook, "j"'] & 1
\end{tikzcd}$$

où s est une section quelconque. Soit $y \in K^1(G, F)$. Je pretends que l'unique

$$x \in \operatorname{Ker}[H^2(\Gamma;\mathbb{Q}/\mathbb{Z}) \to H^2(\hat{F};\mathbb{Q}/\mathbb{Z})]$$

vérifiant $\psi(x) = y$ et $s^*(x) = 0$ est en fait dans $K^2(\Gamma;\mathbb{Q}/\mathbb{Z})$.

En effet

$$\psi(i^*(x)) = j^*(y) = 0$$

et

$$s^*(i^*(x)) = j^*(s^*(x)) = 0\,.$$

Ce qui montre que $i^*(x) \in H^2(p^{-1}(P(A)); \mathbb{Q}/\mathbb{Z})$ est nul. Comme A est un sous-groupe bicyclique arbitraire, x appartient donc à $K^2(\Gamma; \mathbb{Q}/\mathbb{Z})$. □

Remarque II.6. Le théorème II.3 se généralise immédiatement au cas où le G-réseau M n'est plus d'indice fini dans un G-module de permutation grâce au fait suivant qui est «entre les lignes» dans [S 2].

Si M et N sont deux G-réseaux, tout élément $\alpha \in \mathrm{Br}(\mathbb{C}(M)^G)$ appartient au sous groupe $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G)$ si et seulement si son image par l'application canonique appartient à $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M \oplus N)^G)$.

Nous sommes maintenant en mesure de prouver le

Théorème II.7. *Soit G un groupe fini. Les conditions suivantes sont équivalentes.*
i) *Pour tout G-réseau M on a $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G) = 0$.*
ii) *Les sous-groupes de Sylow de G sont abéliens bicycliques.*

Preuve. ii)$\to$i) Comme nous l'avons observé en II.2 le groupe $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G)$ est un sous-groupe du groupe $H^2(G; \mathbb{C}(M)^*)$.

Il en résulte que l'application: $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G) \to \Pi_S \mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^S)$ où S parcourt les sous-groupes de Sylow de G est injective. Ce dernier groupe est nul par la proposition II.2.

i)$\to$ii) Soient donc G un groupe fini et S un p sous-groupe de Sylow de G non abélien bicyclique.

Nous commençons par reproduire ici la construction de Saltman [S 2] pour exhiber un S réseau N tel que $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(N)^S)$ soit non nul.

Soit N un S réseau tel que $H^2(S; N) \simeq \dfrac{\mathbb{Z}}{\# S}$ (par exemple le noyau de l'application de $\mathbb{Z}[S \times S] \to \mathbb{Z}[S]$) dans la résolution standard de $\mathbb{Z}$.)

Notons p^n le cardinal de S et $y = p^{n-1}e$ où e est un générateur du groupe cyclique $H^2(S; N)$. Quitte à ajouter à N un S module auxilliaire on peut supposer que N est d'indice fini dams un module de permutation. Soit alors $M = N \otimes_S \mathbb{Z}[G]$. C'est un G-réseau lui aussi d'indice fini dans un module de permutation et on peut donc appliquer le théorème II.3 pour calculer $\mathrm{Br}^{\mathrm{nr}}(\mathbb{C}(M)^G)$. Montrons que l'élément $y = p^{n-1}e \in H^2(S; N) = H^2(G; M)$ qui est non nul, appartient à $K^2(G; M)$. En effet il faut montrer que y devient nul quand on le restreint aux sous-groupes abéliens bicycliques B de G; et pour cela il suffit de le voir pour les sous-groupes de Sylow B_q d'un tel B qui sont maintenant des q groupes abéliens bicycliques. C'est alors clair si $q \neq p$ et pour $q = p$ le p groupe B_p est conjugué dans G d'un sous-groupe de S, strictement contenu dans S puisque S n'est pas abélien bicyclique et dont le cardinal divise donc p^{n-1}. La restriction de $y = p^{n-1}e$ à $H^2(B_p; M)$ est donc nulle puisque ce groupe de cohomologie est tué par p^{n-1}. □

III. Appendice

Lemme III.1. *Soit* $0 \to A \xrightarrow{i} \Gamma \to G \to 1$ *une suite exacte scindée* (1) *où* A *est abélien, qui induit en homologie la suite exacte*

$$0 \to H_1(G;A) \xrightarrow{\phi} \frac{H_2(\Gamma;\mathbb{Z})}{i_*(H_2(A;\mathbb{Z}))} \to H_2(G;\mathbb{Z}) \to 0$$

Soient $x \in H_2(G,\mathbb{Z})$ *et* s_1, s_2 *deux sections de* (1). *On note* $d(s_1,s_2)$ *la classe dans* $H^1(G;A)$ *de la différence* $s_2 - s_1$. *Alors on a*

$$s_{2*}(x) - s_{1*}(x) = \phi(d(s_1,s_2) \cap x) \in \frac{H_2(\Gamma;Z)}{i_*(H_2(A;Z))}$$

Par évaluation sur les classes de cohomologies le lemme III.1 est équivalent à:

Lemme III.2. *Soit*

$$0 \to A \xrightarrow{i} \Gamma \to G \to 1 \tag{1}$$

une suite exacte scindée où A *est abélien, qui induit en cohomologie la suite exacte:*

$$0 \to H^2(G;C) \to \mathrm{Ker}[H^2(\Gamma;C) \xrightarrow{i^*} H^2(A;C)] \xrightarrow{\Psi} H^1(G;\mathrm{Hom}(A;C)) \to 0$$

où C *est un groupe abélien arbitraire. Soient* $v \in \mathrm{Ker}[H^2(\Gamma;C) \to H^2(A;C)]$ *et* s_1, s_2 *deux sections de* (1). *On note* $d(s_1,s_2)$ *la classe dans* $H^1(G;A)$ *de la différence* $s_2 - s_1$. *Alors on a*

$$s_2^*(v) - s_1^*(v) = \Psi(v) \cup d(s_1,s_2) \in H^2(G;C) \tag{2}$$

Preuve. Il est clair que le membre de gauche de l'égalité (2) ne dépend que de $\Psi(v)$. Pour démontrer l'égalité (2) on peut donc supposer par exemple que $s_1^*(v) = 0$. Si $0 \to C \to E \to \Gamma \to 1$ est une extension centrale représentant v, le groupe E est alors le produit semi-direct de G par $C \times A$ puisque d'une part $s_1^*(v) = 0$ et d'autre part $i^*(v) = 0$. La loi de multiplication dans E identifié par s_1 à $C \times A \times G$ est la suivante:

$$(c,a,g) \times (c',a',g') = (c + c' + \overline{\Psi(v)}(g, g^{-1} \cdot a'),\ a + g \cdot a', gg')$$

où $\overline{\Psi(v)}$ désigne maintenant un homomorphisme croisé dont la classe est $\Psi(v)$. Soit de même $\bar{d}$ un homomorphisme croisé représentant $d(s_1,s_2)$. On doit calculer dans E pour $g_1, g_2 \in G$, $s_2(g_1g_2)s_2(g_2)^{-1}s_2(g_1)^{-1}$ et on trouve

$$(\overline{\Psi(v)}(g_1, \bar{d}(g_2)), 0, 1)$$

ce qui signifie exactement que le 2-cocycle associée à s_2, est le cup produit des deux 1-cocycles $\overline{\Psi(v)}$ et $\bar{d}$.

Remerciements. Je remercie vivement J. L. Colliot-Thèlene et J. J. Sansuc qui m'ont initié au sujet, puis fourni tous les éclaircissements nécessaires. Je remercie également E. Ghys, J. Lannes, M. Ojanguren, T. Vust, de leurs suggestions et de l'intérêt qu'ils ont montré pour ce travail, ainsi que le Fonds National Suisse qui m'a offert un séjour de 6 mois à l'Université de Lausanne.

Bibliographie

[BG] Barge, J., Ghys, E.: Surface et cohomologie bornée. Invent. Math. **92**, 509–526 (1988)

[CTS] Colliot-Thélène, J.L., Sansuc, J.J.: Groupes de Brauer de variétés quotients. En préparation

[S 1] Saltman, D.J.: Noether's problem over an algebraically closed field. Invent. Math. **77**, 71–84 (1984)

[S 2] Saltman, D.J.: Multiplicative field invariants. J. Algebra **106**, 221–238 (1987)

[S 3] Saltman, D.J.: Multiplicative field invariants and the Brauer group. Preprint Austin (1987)

[B] Bogomolov, F.A.: The Brauer group of quotient spaces by linear group actions. Izv. Akad. Nauk. Ser. Mat. **51**, 485–516 (1987)

[Bi] Bieri, R.: Gruppen mit Poincaré Dualität. Comm. Math. Helv. **47**, 373–396 (1972)

Reçu le 9 juin 1988; version révisée reçu le 12 octobre 1988

Math. Ann. 283, 529–537 (1989)

Mathematische Annalen

Über das Primzahl-Zwillingsproblem

Dieter Wolke
Mathematisches Institut, Albertstrasse 23b, D-7800 Freiburg, Bundesrepublik Deutschland

1. Für natürliches k und $x \geqq 2$ sei

$$\psi_2(x, 2k) = \sum_{2k < n \leqq x} \Lambda(n)\Lambda(n-2k)$$

(Λ = von-Mangold-Funktion).

Nach Hardy-Littlewood [2] verhält sich ψ_2 für $x \to \infty$ vermutlich wie

$$H(x, 2k) = 2 \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{\substack{p \mid k \\ p>2}} \left(\frac{p-1}{p-2}\right) \cdot (x-2k). \tag{1.1}$$

Van der Corput [1] und Lavrik [4] (s. hierzu Montgomery [5, Chap. 15]) zeigten, daß dies im mittleren Sinn richtig ist.

Für beliebige C_1, $C_2 > 0$ kann die Anzahl der $k \leqq x/2$ mit

$$E(x, 2k) = \psi_2(x, 2k) - H(x, 2k) \neq O(x(\ln x)^{-C_1}) \tag{1.2}$$

durch $O(x(\ln x)^{-C_2})$ abgeschätzt werden. Oder:

$$\sum_{1 \leqq k \leqq x/2} (E(x, 2k))^2 = O(x^2(\ln x)^{-C}) \quad \text{für jedes } C > 0. \tag{1.3}$$

Nach der Methode von Montgomery-Vaughan [6] wurde (1.2) von Jahnke [3] wie folgt verschärft: Es existiert ein $\delta > 0$, so daß die Anzahl der $k \leqq x/2$ mit $\psi(x, 2k) = 0$ durch $O(x^{1-\delta})$ abgeschätzt werden kann.

In der vorliegenden Arbeit soll das Ergebnis von van der Corput-Lavrik dahingehend verbessert werden, daß für „fast alle" $k \leqq x$ die vermutete Asymptotik für $\psi_2(y, 2k)$ in einem y-Bereich gilt, der weit über x hinausgeht.

Satz. *Seien* ε, A, $B > 0$. *Dann gilt für alle* $k \leqq x$ *mit Ausnahme von* $O(x(\ln x)^{-A})$ *Zahlen*

$$\psi_2(y, 2k) = H(y, 2k) + O(y(\ln y)^{-B})$$

im Bereich

$$2x \leqq y \leqq x^{\frac{8}{5} - \varepsilon}.$$

(Die O-Konstanten hängen nur von ε, A und B ab).

Die Aussage wird ermöglicht durch Dichte-Sätze für die Nullstellen der L-Reihen. Die Dichte-Hypothese (s. [5, Chap. 12]) bzw. die verallgemeinerte Riemann'sche Vermutung führen mit der verwendeten Methode zu dem Intervall $[2x, x^{2-\varepsilon}]$ für y.

2. Für $D_1 > 0$ sei

$$2x \leqq N < N' \leqq 2N \leqq x^{\frac{8}{5}-\varepsilon}, \qquad Q_1 = (\ln x)^{D_1}, \qquad Q = N^{\frac{5}{8}+\frac{\varepsilon}{2}}. \tag{2.1}$$

Für hinreichend großes x ist $Q_1 < \dfrac{N}{Q} < Q$.

Im Hinblick auf die Auswertung der Summe

$$S(\alpha) = S(\alpha, N, N') = \sum_{N < n \leqq N'} \Lambda(n) e(\alpha n) \tag{2.2}$$

benötigt man folgendes

Lemma 1. *Für $q \in \mathbb{N}$ und $\beta \in \mathbb{R}$ sei*

$$L(q, \beta) = \sum_{\chi(q)} \sum_{|\varrho(\chi)| \leqq x^2} \left| \int_N^{N'} t^{\varrho - 1} e(t\beta) dt \right|$$

(ϱ durchläuft die nichttrivialen Nullstellen von $L(s, \chi)$).

(1) *Für* $1 \leqq q \leqq Q_1$, $|\beta| \leqq \dfrac{1}{qQ}$ *und beliebiges $D_2 > 0$ gilt (mit nur von D_2 abhängiger $\ll$-Konstante)*

$$L(q, \beta) \ll N^{1/2} \min(N^{1/2}, |\beta|^{-1/2}) (\ln x)^{-D_2}.$$

(2) *Für* $Q_1 < q \leqq \dfrac{N}{Q}$, $|\beta| \leqq \dfrac{1}{qQ}$ *gilt*

$$L(q, \beta) \ll N^{1/2} \min(N^{1/2}, |\beta|^{-1/2}) (\ln x)^{16}.$$

Beweis. Nach Titchmarsh [9, Lemmas 4.3 und 4.5.] ist für $\varrho = \xi + i\eta$ im Fall $\dfrac{1}{N} \leqq |\beta| \leqq \dfrac{1}{qQ}$

$$\int_N^{N'} t^{\varrho - 1} e(t\beta) dt \ll \begin{cases} N^{\xi - 1} |\beta|^{-1}, & \text{falls} \quad |\varrho| \leqq N|\beta|, \\ N^{\xi - 1/2} |\beta|^{-1/2}, & \text{falls} \quad N|\beta| < |\varrho| \leqq 8\pi N|\beta|, \\ N^{\xi} (|\varrho| + 1)^{-1}, & \text{falls} \quad 8\pi N|\beta| < |\varrho| \leqq x^2. \end{cases} \tag{2.3}$$

Durch Entwickeln von $e(t\beta)$ sieht man im Fall $|\beta| \leqq N^{-1}$

$$\int_N^{N'} t^{\varrho - 1} e(t\beta) dt \ll \frac{N^{\xi}}{1 + |\varrho|} \tag{2.4}$$

Es soll im Folgenden der Fall $\beta \geqq N^{-1}$ behandelt werden.

Sei wie üblich für $\sigma \geqq \frac{1}{2}$ und $T \geqq 1$

$N(\sigma, T, \chi)$ die Anzahl der Nullstellen $\varrho = \xi + i\eta$ von $L(s,\chi)$ mit $\xi \geqq \sigma$ und $|\eta| \leqq T$. Nach Montgomery [5, Chap. 12] gilt

$$\sum_{\chi \bmod q} N(\sigma, T, \chi) \ll \begin{cases} (qT)^{\frac{3(1-\sigma)}{2-\sigma}} \ln^9(qT), & \frac{1}{2} \leqq \sigma \leqq \frac{4}{5}, \\ (qT)^{\frac{2(1-\sigma)}{\sigma}} \ln^{14}(qT), & \frac{4}{5} \leqq \sigma < 1. \end{cases} \tag{2.5}$$

(Diese Ungleichungen sind zwischenzeitlich mehrfach verschärft worden. Eine Übersicht findet man in Richert [8, Chap. 6]. In Bezug auf unseren Satz entstehen jedoch nur geringe Verbesserungen). Für $q \leqq Q_1$, $\chi \bmod q$, $|\tau| \leqq x^2$ und

$$\sigma \geqq \sigma_0 = 1 - (\ln x)^{-4/5} \tag{2.6}$$

gilt $L(\sigma + iT, \chi) \neq 0$ (s. [7, VIII, Satz 6.2]). Für $Q_1 < q \leqq NQ^{-1}$ with $\sigma_0 = 1$ gesetzt.

Im Bereich $\frac{1}{2} \leqq \sigma \leqq \frac{4}{5}$ ist $T^{\frac{3(1-\sigma)}{2-\sigma}}$ in T wachsend, $T^{\frac{3(1-\sigma)}{2-\sigma}-1}$ fallend. Entsprechendes ist richtig für $T^{\frac{2(1-\sigma)}{\sigma}}$. (2.3, 2.5, 2.6) führen daher im Fall $N^{-1} \leqq \beta \leqq (qQ)^{-1}$ zu der Abschätzung

$$\begin{aligned} L(q,\beta) \ll \ln^{16} x \cdot \Bigg\{ & qN\beta^{1/2} + \beta^{-1/2} N^{1/2} \int_{1/2}^{4/5} ((qN\beta)^{\frac{3}{2-\sigma}} N^{-1})^{1-\sigma} d\sigma \\ & + \beta^{-1/2} N^{1/2} \int_{4/5}^{\sigma_0} ((qN\beta)^{2/\sigma} N^{-1})^{1-\sigma} d\sigma \Bigg\}. \end{aligned}$$

Im ersten Integral ist die Funktion

$$f_1(\sigma) = (1-\sigma)\left(\frac{3}{2-\sigma}\ln(qN\beta) - \ln N\right)$$

wegen $Q > N^{13/25}$ monoton steigend. Im Intervall $[\frac{4}{5}, \sigma_0]$ nimmt die Funktion

$$f_2(\sigma) = (1-\sigma)\left(\frac{2}{\sigma}\ln(qN\beta) - \ln N\right)$$

ihren größten Wert an einem der Endpunkte an. Es folgt daher

$$L(q,\beta) \ll \ln^{16} x \cdot \{qN\beta^{1/2} + q^{1/2} N^{4/5} + N^{1/2} \beta^{-1/2} ((qN\beta)^{2/\sigma_0} N^{-1})^{1-\sigma_0}\}.$$

Für $q \leqq Q_1$ ist $((qN\beta)^{2/\sigma_0} N^{-1})^{1-\sigma_0}$ nach (2.6)

$$\ll (\ln x)^{-16-D_2}.$$

Damit ergibt sich Behauptung (1) für $|\beta| \geqq N^{-1}$. Im andern Fall geht man mit (2.4) ganz analog vor. Behauptung (2) folgt in gleicher Weise, wobei mehrfach $Q \geqq N^{5/8}$ ausgenützt wird.

Für $1 \leqq q \leqq Q$ und $(a,q) = 1$ sei

$$I_{q,a} = \left[\frac{a}{q} - \frac{1}{qQ}, \frac{a}{q} + \frac{1}{qQ}\right]. \tag{2.7}$$

Lemma 2. (1) *Für* $1 \leqq q \leqq Q_1$, $(a,q)=1$, $\alpha = \frac{a}{q} + \beta \in I_{q,a}$ *und beliebiges* $D_2 > 0$ *gilt*

$$S(\alpha) = \sum_{N < n \leqq N'} \Lambda(n) e(\alpha n)$$
$$= \frac{\mu(q)}{\varphi(q)} \sum_{N < n \leqq N'} e(\beta n) + O((\ln x)^{1-D_2} \cdot \min(N, N^{1/2} |\beta|^{-1/2})).$$

(2) *Für* $Q_1 < q \leqq \frac{N}{q}$, $(a,q)=1$ *und* $\alpha \in I_{q,a}$ *gilt*

$$S(\alpha) \ll q^{-1/2} \min(N, N^{1/2} |\beta|^{-1/2}) \ln^{17} x.$$

(3) *Für* $\frac{N}{Q} < q \leqq Q$, $(a,q)=1$ *und* $\alpha \in I_{q,a}$ *gilt*

$$S(\alpha) \ll (N q^{-1/2} + N^{4/5} + N^{1/2} q^{1/2}) \ln^4 x.$$

Beweis zu (1). Nach [7, VII. Satz 4.4] besteht für $N < y \leqq N'$ und $\chi \bmod q$ die explizite Formel

$$\psi(y, \chi) = \sum_{n \leqq y} \Lambda(n) \chi(n)$$
$$= \varepsilon_\chi y - \sum_{|\varrho(\chi)| \leqq x^2} \frac{y^\varrho}{\varrho} + R(y),$$

wobei

$$\varepsilon_\chi = 1, \quad \text{falls} \quad \chi = \chi_0; \qquad \varepsilon_\chi = 0, \quad \text{falls} \quad \chi \neq \chi_0$$

und

$$R(y) \ll \min\left(\ln^2 x, \frac{y}{x^2 \|y\|} \ln^2 x\right) \left(\|y\| = \min_{a \in \mathbf{Z}} |y - a|\right). \tag{2.8}$$

Die im zitierten Satz genannte Einschränkung auf primitive Charaktere kann auf Kosten eines Fehlers $O(\ln^2 x)$ fallengelassen werden. Die dortigen Terme $v_0 \ln x$, d_0 und $B(y, \chi)$ sind ebenfalls $O(\ln^2 x)$. (Im Fall, daß zu q ein Charakter χ_1 mit einer Ausnahme-Nullstelle gehört, kann d_0 nur durch $O(q^\varepsilon)$ abgeschätzt werden. Da zu q nur ein solches χ_1 gehört, bleibt die folgende Formel für $S\left(\frac{a}{q} + \beta\right)$ auch in diesem Fall richtig.)

In üblicher Weise ergibt sich

$$S\left(\frac{a}{q} + \beta\right) = \frac{1}{\varphi(q)} \sum_{\chi(q)} \chi(a) \tau(\bar{\chi}) \sum_{N < n \leqq N'} \Lambda(n) \chi(n) \cdot e(n\beta) + O(\ln^2 x)$$
$$= \frac{\mu(q)}{\varphi(q)} \int_N^{N'} e(y\beta) dy - \frac{1}{\varphi(q)} \sum_{\chi(q)} \chi(a) \tau(\bar{\chi}) \sum_{|\varrho(\chi)| \leqq x^2} \int_N^{N'} y^\varrho e(y\beta) dy$$
$$+ O\left(\frac{1}{\varphi(q)} \sum_{\chi(q)} |\tau(\bar{\chi})| \left(|R(N)| + |R(N')| + \int_N^{N'} |R(y)| |\beta| dy\right) + \ln^2 x\right). \tag{2.9}$$

Das Integral im ersten Term ist

$$=\sum_{N<n\leqq N'} e(n\beta)+O(1+N|\beta|).$$

Der Fehler-Term erweist sich mit (2.8) und der Ungleichung $\tau(\chi)\ll q^{1/2}$ als

$$O(q^{1/2}\ln^2 x \max(1,\mathrm{N}|\beta|).$$

Damit wird (2.9) zu

$$S\left(\frac{a}{q}+\beta\right)=\frac{\mu(q)}{\varphi(q)}\sum_{N<n\leqq N'} e(n\beta)+O\left(\frac{q^{1/2}}{\varphi(q)}L(q,\beta)\right)$$
$$+O(q^{1/2}\ln^2 x\max(1,N|\beta|)). \tag{2.10}$$

Der Fehler kann mit Lemma 1 wie in (1) behauptet abgeschätzt werden. Ähnlich geht man bei (2) vor.

(3) findet man in [10, Theorem 3.1]. In etwas schwächerer Form ergibt es sich wie (1) mit (2.10) und einem Analogon zu Lemma 1.

Lemma 3. *Für $Q_1<q\leqq Q$ und $(a,q)=1$ ist*

$$\int_{I_{q,a}}|S(\alpha)|^2 d\alpha\ll\begin{cases} Nq^{-1}(\ln x)^{35}, & \text{falls}\quad Q_1<q\leqq \dfrac{N}{Q},\\ \left(\dfrac{N^2}{q^2Q}+\dfrac{N^{8/5}}{qQ}\right)\ln^8 x, & \text{falls}\quad \dfrac{N}{Q}<q\leqq Q.\end{cases}$$

Dies ergibt sich unmittelbar aus den Ungleichungen (2) und (3) von Lemma 2.

3. Die Einteilung des Einheitsintervalls in „major" und „minor arcs" geschieht wie folgt.

$$\left.\begin{aligned}\mathbf{M}&=\bigcup_{q\leqq Q_1}\ \bigcup_{1\leqq a\leqq q,\,(a,q)=1} I_{q,a}\\ \mathbf{m}&=[Q^{-1},1+Q^{-1}]\backslash\mathbf{M}.\end{aligned}\right\} \tag{3.1}$$

Es werde zuerst der Fall

$$x(\ln x)^{D_3}<N<N'\leqq 2N\leqq x^{\frac{8}{5}-\varepsilon} \tag{3.2}$$

betrachtet.

Für $k\leqq x$ läßt sich der Beitrag von $\mathbf{M}$ zu

$$\psi_2(N',2k)-\psi_2(N,2k)=\int_{Q^{-1}}^{1+Q^{-1}}|S(\alpha,N,N')|^2 e(-2k\alpha)d\alpha+O(x) \tag{3.3}$$

mit Lemma 2 leicht berechnen. Man sieht für $q \leqq Q_1$

$$\sum_{\substack{a=1\\(a,q)=1}}^{q} \int_{I_{q,a}} |S(\alpha)|^2 e(-2k\alpha)d\alpha$$

$$= \sum_{\substack{a=1\\(a,q)=1}}^{q} \frac{\mu^2(q)}{\varphi^2(q)} \int_{-1/2}^{1/2} \Big|\sum_{N<n\leqq N'} e(n\beta)\Big|^2 e\left(-2k\left(\frac{a}{q}+\beta\right)\right) d\beta$$

$$+O\left(\varphi(q)\int_0^{(qQ)^{-1}} (\ln x)^{2-2D_2} \min(N^2, N\beta^{-1})d\beta\right)$$

$$+O\left(\frac{1}{\varphi(q)}\int_{(qQ)^{-1}}^{1/2} \beta^{-2}d\beta\right) = \frac{\mu^2(q)}{\varphi^2(q)} c_q(-2k)(N'-N)$$

$$+O\left(\frac{x}{\varphi(q)}\right)+\varphi(q)(\ln x)^{2-2D_2}\cdot N+\frac{q}{\varphi(q)}Q,$$

also

$$\int_{\mathbf{M}} |S(\alpha,N,N')|^2 e(-2k\alpha)d\alpha = \sum_{q\leqq Q_1} \frac{\mu^2(q)}{\varphi^2(q)} c_q(-2k)(N'-N)$$

$$+O(N((\ln x)^{-D_3}+(\ln x)^{2+2(D_1-D_2)})) \tag{3.4}$$

$\left(N,\ N' \text{ gemäß (3.2)},\ c_q(h) = \sum_{1\leqq a\leqq q,\,(a,q)=1} e\left(\frac{a}{q}h\right)\right).$

Es soll

$$A(N,N',D_3) = \left|\left\{k\leqq x, \left|\int_{\mathbf{M}} |S(\alpha,N,N')|^2 e(-2k\alpha)d\alpha\right| > N(\ln x)^{-D_3}\right\}\right| \tag{3.5}$$

abgeschätzt werden. Man erhält

$$A(\,) \leqq N^{-2}(\ln x)^{2D_3}\cdot \int_0^1 \Big|\sum_{k\leqq x} e(k\gamma)\int_{\mathbf{m}} |S(\alpha)|^2 e(-2\alpha)d\alpha\Big|^2 d\gamma$$

$$\ll N^{-2}(\ln x)^{2D_3} \int_{\mathbf{m}} |S(\alpha_1)|^2 \int_{\mathbf{m}} |S(\alpha_2)|^2 \cdot \min\left(x, \frac{1}{\|\alpha_1-\alpha_2\|}\right) d\alpha_2 d\alpha_1. \tag{3.6}$$

Lemma 4. *Gleichmäßig in* $\alpha_1 \in \mathbf{m}$ *gilt*

$$\int_{\mathbf{m}} |S(\alpha_2)|^2 \min\left(x, \frac{1}{\|\alpha_1-\alpha_2\|}\right) d\alpha_2 \ll Nx(\ln x)^{37-D_1}.$$

Beweis. Zu festgehaltenem $\alpha_1 \in \mathbf{m}$ sei

$$\mathbf{m}_0 = \mathbf{m}_0(\alpha_1) = [\alpha_1 - x^{-1}, \alpha_1 + x^{-1}] \cap \mathbf{m},$$

$$\mathbf{m}_l = \mathbf{m}_l(\alpha_1) = \{\alpha \in \mathbf{m}, 2^{l-1}x^{-1} \leqq |\alpha-\alpha_1| \leqq 2^l x^{-1}\} \qquad \left(1 \leqq l \leqq \frac{\ln x}{\ln 2}+1\right).$$

Nach dem Dirichlet'schen Approximationssatz und (3.1) wird $\mathbf{m} \subseteq \bigcup_l \mathbf{m}_l$ überdeckt durch die Intervalle $I_{q,a}$ mit $Q_1 < q \leqq Q$, $(a,q)=1$, $1 \leqq a \leqq q$.

Es werde bei festgehaltenem l zuerst $J_{2,l}$, der Beitrag der $I_{q,a}$ mit $\frac{N}{Q} < q \leqq Q$, $I_{q,a} \cap \mathbf{m}_l \neq \emptyset$, zum obigen Integral abgeschätzt. Zu jedem $q \in \left(\frac{N}{Q}, Q\right]$ gibt es höchstens

$$\ll \frac{2^l}{x} q + 1 \ll \begin{cases} 1, & \text{falls} \quad NQ^{-1} < q \leqq x2^{-1}, \\ 2^l q x^{-1}, & \text{falls} \quad x2^{-1} < q \leqq Q \end{cases}$$

solche Intervalle. Die erste Alternative tritt nur für $2^l > xQN^{-1}$ ein. In diesem Fall ergibt sich mit Lemma 3

$$\begin{aligned} J_{2,l} &\ll \frac{x}{2^l} \ln^8 x \cdot \left(\sum_{N/Q < q \leqq x/2^l} \left(\frac{N^2}{q^2 Q} + \frac{N^{8/5}}{qQ} \right) + \sum_{x/2^l < q \leqq Q} \frac{2^l q}{x} \left(\frac{N^2}{q^2 Q} + \frac{N^{8/5}}{qQ} \right) \right) \\ &\ll \ln^9 x \cdot \left(\frac{x}{2^l} N + \frac{x}{2^l} \frac{N^{8/5}}{Q} + \frac{N^2}{Q} + N^{8/5} \right) \\ &\ll \ln^9 x \cdot \left(\frac{N^2}{Q} + \frac{N^{13/5}}{Q^2} + N^{8/5} \right) \\ &\ll Nx(\ln x)^{36-D_1}, \end{aligned} \tag{3.7}$$

nach (2.1). Im Fall $2^l \leqq xQN^{-1}$ folgt dieselbe Ungleichung.

Bezeichne analog $J_{1,l}$ den Beitrag der $I_{q,a}$ mit $Q_1 < q \leqq \frac{N}{Q}$ zu dem Integral über $\mathbf{m}_l$. Die q mit $q \geqq x2^{-l}$ liefern ähnlich wie oben mit Lemma 3

$$\ll \frac{x}{2^l} \sum_{Q_1 < q \leqq N/Q} \frac{2^l}{x} q \cdot \frac{N}{q} \ln^{35} x \ll Nx(\ln x)^{36-D_1}.$$

Es bleiben im Fall $2^l \leqq xQ_1^{-1}$ nur noch die q mit $Q_1 < q \leqq x2^{-l}$. Das Intervall $(Q_1, \min(NQ^{-1}, x2^{-l})]$ werde in $\ll \ln x$ Teile $(K, K']$ mit $K' \leqq 2K$ eingeteilt. Wegen $\left|\frac{a}{q} - \frac{a'}{q'}\right| \geqq \frac{1}{4K^2}$ für $q, q' \in (K, K']$, $(a,q)=(a',q')=1$, $\frac{a}{q} \neq \frac{a'}{q'}$ wird $\mathbf{m}_l$ von höchstens $\ll \frac{2^l}{x} K^2 + 1$ Intervallen $I_{q,a}$ mit $q \in (K, K']$ geschnitten. Diese tragen zu $J_{1,l}$ nur

$$\begin{aligned} &\ll \frac{x}{2^l} \left(\frac{2^l}{x} K^2 + 1 \right) \frac{N}{K} \ln^{35} x \ll Nx \ln^{35} x \left(\frac{K}{x} + K2^{-l} \right) \\ &\ll Nx(\ln x)^{35-D_1} \end{aligned}$$

bei. Dies führt zur Ungleichung (3.7) für $J_{1,l}$. Summation über die l ergibt die Behauptung.

Setzt man Lemma 4 in (3.6) ein, so zeigt sich

$$A(N, N', D_3) \ll x(\ln x)^{37+2D_3-D_1}. \tag{3.8}$$

Zusammenfassung von (3.3–3.5, 3.8) ergibt für N, N' gemäß (3.2) (nachdem D_2 hinreichend groß gewählt wurde):

Für alle $k \leqq x$ bis auf höchstens $O(x(\ln x)^{37+2D_3-D_1})$ Ausnahmen gilt

$$\psi_2(N', 2k) - \psi_2(N, 2k) = \sum_{q \leqq (\ln x)^{D_1}} \frac{\mu^2(q)}{\varphi^2(q)} c_q(-2k)(N'-N) + O(N(\ln x)^{-D_3}). \tag{3.9}$$

Im Fall $2x \leqq N \leqq x(\ln x)^{D_3}$ argumentiert man ganz analog mit

$$\psi_2(N', 2k) - \psi_2(N, 2k) = \int_{Q^{-1}}^{1+Q^{-1}} \sum_{N < n_1 \leqq N'} \Lambda(n_1) e(n_1\alpha)$$
$$\times \sum_{N-2k \leqq n_2 < N'-2k} \Lambda(n_2) e(-n_2\alpha) e(-2k\alpha) d\alpha .$$

Somit ist (3.9) richtig für

$$2x \leqq N < N' \leqq 2N \leqq x^{\frac{8}{5}-\varepsilon} .$$

4. Das Intervall $[2x, x^{\frac{8}{5}-\varepsilon}]$ werde aufgeteilt in $\ll (\ln x)^{D_4+1}$ aneinander anschließende Intervalle $[N_\nu, N_{\nu+1}]$ mit

$$N_{\nu+1} < N_\nu(1 + (\ln x)^{-D_4}) .$$

Für alle $k \leqq x$ bis auf höchstens $O(x(\ln x)^{38+2D_3+D_4-D_1})$ Ausnahmen und alle ν gilt daher

$$\psi_2(N_\nu, 2k) - \psi_2(2x, 2k) = \sum_{q \leqq (\ln x)^{D_1}} \frac{\mu^2(q)}{\varphi^2(q)} c_q(-2k)(N_\nu - 2x)$$
$$+ O(N_\nu(\ln x)^{D_4-D_3+1}) . \tag{4.1}$$

Für alle Nicht-Ausnahmen k ist wegen

$$\sum_{N_\nu < n \leqq N_{\nu+1}} \Lambda(n)\Lambda(n-2k) \ll N_\nu$$

(4.1) an allen $y \in [2x, x^{\frac{8}{5}-\varepsilon}]$ mit einem Fehler

$$O(y((\ln x)^{D_4-D_3+1} + (\ln x)^{-D_4}))$$

erfüllt.

Mit der Ungleichung

$$|c_q(-2k)| \leqq \frac{\varphi(q)}{\varphi\left(\dfrac{q}{(q, 2k)}\right)}$$

erhält man durch einfache elementare Rechnung, daß die Anzahl der $k \leqq x$ mit

$$\left| \sum_{q > (\ln x)^{D_1}} \frac{\mu^2(q)}{\varphi^2(q)} c_q(-2k) \right| > (\ln x)^{-D_5}$$

durch $O(x(\ln x)^{2D_5+1-2D_1})$ abgeschätzt werden kann.

Verwertet man noch das Ergebnis (1.2) von von der Corput-Lavrik, so folgt nach passender Wahl der D_v die Behauptung des Satzes für alle y im angegebenen Intervall.

Literatur

1. Corput, J.G. van der: Sur l'hypothèse de Goldbach pour presque tous les nombres pairs. Acta Arith. **2**, 266–290 (1937)
2. Hardy, G.H., Littlewood, J.E.: Some problems of "Partitio Numerorum". III: On the expression of a number as a sum of primes. Acta Math. **44**, 1–70 (1923)
3. Jahnke, Th.: Über die Anzahl der Primzahlpaare mit gegebenem Abstand. Dissertation Univ. Freiburg. 1978
4. Lavrik, A.F.: On the twin prime hypothesis of the theory of primes by the method of I.M. Vinogradov. Dokl. Akad. Nauk SSSR **132**, 1013–1015 (1960) = Sov. Math. Dokl. **1**, 700–702 (1960)
5. Montgomery, H.L.: Topics in multiplicative number theory. (Lectures Notes Mathematics 227). Berlin-Heidelberg-New York: Springer 1971
6. Montgomery, H.L., Vaughan, R.C.: The exceptional set in Goldbach's problem. Acta Arith. **27**, 353–370 (1975)
7. Prachar, K.: Primzahlverteilung. Berlin Göttingen Heidelberg: Springer 1957
8. Richert, H.-E.: Lectures on Sieve methods. Bombay: Tata Institute 1976
9. Titchmarsh, E.C.: The theory of the Riemann zeta-function. Oxford: Clarendon 1951
10. Vaughan, R.C.: The Hardy-Littlewood method. Cambridge: University Press 1981

Received July 9, 1987

Math. Ann. 283, 539–549 (1989)

Mathematische Annalen

The Rate of Convergence of a Harmonic Map at a Singular Point

Robert Gulliver[1] and Brian White[2]

[1] School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA
[2] Department of Mathematics, Stanford University, Stanford, CA 94305, USA

0. Introduction

It has been apparent since the 1968 example of De Giorgi that weak solutions of elliptic systems in $m \geqq 3$ independent variables may well have points of discontinuity [dG]. In the geometrically interesting case of harmonic mappings, the example $f(x)=x/|x|$ of a discontinuous (weakly) harmonic mapping, from Euclidean $\mathbb{R}^m$ to the standard sphere of dimension $m-1$, was given by Hildebrandt and Widman [HW]. Thus, the regularity theory for harmonic mappings between Riemannian manifolds requires a clear understanding of the behaviour of the mapping near its singular set. If $f: M^m \to N^n$ is a minimizing harmonic map with a point of discontinuity $O \in M$, then its *homogeneous tangent map* is defined as follows. Let $x=(x^1, \ldots, x^m)$ be Riemannian normal coordinates at 0; for each $0<\lambda<1$, define the blowup $f_\lambda(x) := f(\lambda x)$, as introduced in [GM]. By means of a monotonicity lemma, Schoen and Uhlenbeck showed that f_λ has uniformly bounded Dirichlet integral with respect to the Euclidean metric in the domain [SU, p. 314]. It follows that each blowup sequence $f_{\lambda(i)}$, as $\lambda(i) \to 0$, has a subsequence converging locally weakly to $f_0: \mathbb{R}^m \to N$. They also show that f_0 is harmonic and homogeneous, and that the subsequence converges in the H^1 norm [SU, p. 329]. We shall adopt the global approach of Schoen and Uhlenbeck: we choose an isometric embedding of N^n into some Euclidean $\mathbb{R}^d$, and define $H^1(M, N)$ to be the subset of $H^1(M, \mathbb{R}^d)$, the space of functions having square-integrable first partial derivatives, having values in N almost everywhere.

The analysis of Schoen and Uhlenbeck left open the important question of the uniqueness of the homogeneous tangent mapping f_0. This was resolved for any smooth f_0 by Simon in [S1] (see also [S2]):

Theorem (Simon). *Let $f \in H^1(M, N)$ be a harmonic mapping which minimizes energy on some neighborhood of $O \in M$, where N is a real-analytic manifold. Let $f_0 \in C^2(\mathbb{R}^m \backslash \{0\}, N)$ be the weak limit of some blowup sequence $f_{\lambda(i)}$ as $\lambda(i) \to 0$. Then f_0 is the unique homogeneous tangent map to f at 0, and the restrictions to the sphere of radius 1 satisfy, as $\lambda \to 0$,*

$$\|f_\lambda - f_0\|_{C^2(S^{m-1})} + \|D_\varrho f_\lambda\|_{C^1(S^{m-1})} \to 0 .$$

Here D_ϱ denotes partial differentiation in spherical coordinates of $\mathbb{R}^m$ with respect to $\varrho = |x|$. Similarly, we shall write D_α for partial differentiation with respect to x^α, $1 \leqq \alpha \leqq m$.

Note that Simon's theorem gives an estimate implying only rather slow convergence of f to its homogeneous tangent mapping f_0 (compare inequality (0.6) below). In the present paper, we show that the order of convergence, in general, *depends on the dimensions* of the domain M^m and the target manifold N^n. In the lowest dimensions for which singularities may occur, namely $m = 3$ and $n = 2$, the convergence of f to f_0 is controlled by a *positive power of* $|x|$. On the other hand, whenever $m \geqq 3$ and $n \geqq 3$, we construct examples for which this convergence is *slower* than any positive power of $|x|$.

Let us describe harmonic maps in detail. For $f \in H^1(M, N)$, the energy functional is

$$E(f) = 1/2 \int_M \gamma^{\alpha\beta} \langle D_\alpha f, D_\beta f \rangle \, d\,\mathrm{vol}_M, \tag{0.1}$$

where the Riemannian metric of M is given by $ds_M^2 = \gamma_{\alpha\beta}(x) dx^\alpha dx^\beta$, $(\gamma_{\alpha\beta})$ is the inverse of the $m \times m$ matrix $(\gamma^{\alpha\beta})$, and summation over $1 \leqq \alpha, \beta \leqq m$ is assumed. The volume form is $d\,\mathrm{vol}_M = \sqrt{\gamma(x)}\, dx^1 \dots dx^m$, where $\gamma = \det(\gamma_{\alpha\beta})$. Since we have chosen an isometric embedding of N^n into Euclidean $\mathbb{R}^d$, the inner product $\langle\ ,\ \rangle$ may be understood as the standard inner product of $\mathbb{R}^d$. Via integration by parts, one sees that $f \in H^1(M, N)$ is stationary for E if and only if the weak Laplace-Beltrami operator

$$\Delta_M f := -\gamma^{-1/2} D_\alpha(\sqrt{\gamma}\, \gamma^{\alpha\beta} D_\beta f) \tag{0.2}$$

is normal to N almost everywhere. Equivalently, the vector function f satisfies

$$\Delta_M f + \gamma^{\alpha\beta}(x) B(D_\alpha f, D_\beta f) = 0 \tag{0.3}$$

weakly, where B is the second fundamental form of N in $\mathbb{R}^d$. For any vector fields U, V tangent to N, we may define $B(U, V) := (D_U V)^\perp$, where D_U is covariant differentiation in $\mathbb{R}^d$, and at the relevant point of N, a vector $W \in \mathbb{R}^d$ is given the orthogonal decomposition $W = W^\perp + W^T$ into vectors $W^\perp$ normal to N and W^T tangent to N. Note that in Eq. (0.3), the coefficients of B depend on $f(x) \in N$.

The homogeneous tangent mapping $f_0 : \mathbb{R}^m \to N$ is also harmonic, but with respect to the Euclidean metric on $\mathbb{R}^m$. If we write $\Delta f := -D_\alpha D_\alpha f$ for the standard Laplacian, then the equation satisfied weakly by f_0 is

$$L f_0 := \Delta f_0 + B(D_\alpha f_0, D_\alpha f_0) = 0 .$$

In addition to Simon's result stated above, there is an earlier method developed by Allard and Almgren in [AA] in the context of minimal varieties, which requires an additional hypothesis but yields a stronger conclusion. The analogous proof in the context of harmonic mappings has been carried out by Simon in [S2]. Given a harmonic mapping $f_0 : S^{m-1} \to N$, that is, one whose homogeneous extension to $\mathbb{R}^m$ satisfies $L f_0 = 0$, a vector field $\phi : S^{m-1} \to TN$ along f_0 is called a *harmonic-Jacobi field* if $L f_t$ vanishes to first order in t for any family of mappings $f_t : S^{m-1} \to N$ with $\partial f_t / \partial t = \phi$ at $t = 0$. Equivalently, ϕ is a solution of the linearized equation

$$\Delta \phi + 2B(D_\alpha \phi^T, D_\alpha f) + D_\phi B(D_\alpha f, D_\alpha f) = 0, \tag{0.4}$$

where $D_\phi B$ is the covariant derivative with respect to ϕ of the second fundamental form as a tensor with values in $\mathbb{R}^d$, and using the natural connection D^T of TN for its arguments. For example if $f_t: S^{m-1} \to N$, $-\varepsilon < t < \varepsilon$, is a one-parameter family of harmonic mappings, then it follows that $\phi = \partial f_t/\partial t$ is a harmonic-Jacobi field; in this case, we may say that ϕ is *integrable.*

Theorem (Almgren-Allard [AA]; cf. [S2]). *Let $f \in H^1(M, N)$ be a harmonic mapping which minimizes energy on some neighborhood of $O \in M$.*

Let $f_0 \in C^2(\mathbb{R}^m \setminus \{0\}, N)$ be the weak limit of some blowup sequence $f_{\lambda(i)}(x) = f(\lambda(i)x)$ where $\lambda(i) \to 0$. Assume that f_0 satisfies the following integrability hypothesis:

$$\begin{aligned}&\textit{There is a k-parameter family } f: (U \subset \mathbb{R}^k) \times S^{m-1} \to N \textit{ of}\\ &\textit{harmonic maps such that } f(0, \cdot) = f_0 \textit{ and such that each}\\ &\textit{harmonic-Jacobi field } \varphi \textit{ along } f_0 \textit{ is equal to}\\ &\frac{d}{dt} f(tv, \cdot)(t=0) \textit{ for some } v \in \mathbb{R}^k.\end{aligned} \tag{0.5}$$

Then f_0 is the unique homogeneous tangent map to f at 0, and

$$\|f_\lambda - f_0\|_{C^2(S^{m-1})} + \|D_\varrho f_\lambda\|_{C^1(S^{m-1})} \leqq C\lambda^\alpha \tag{0.6}$$

for some C and $\alpha > 0$ depending on N and on f_0.

Remark 0.1. As stated in [S2, pp. 272–273], this theorem requires N to be analytic. But in the proof there, analyticity is used only to conclude (0.5) from a weaker integrability hypothesis. (cf. [S2, pp. 271–272]). Thus with (0.5), analyticity is not needed. □

It might appear likely to a casual observer that the two theorems are but special cases of a stronger result, as yet undiscovered, that concludes the λ^α convergence but does not require the strong hypothesis (0.5). However, as we shall show

(Section 1) *the λ^α convergence does* not *hold in dimensions m, $n \geqq 3$ for stationary harmonic mappings in the absence of the integrability hypothesis;*

and, on the other hand:

(Section 2) *the integrability hypothesis* (0.5), *and therefore λ^α convergence,* always *hold when the domain M has dimension* 3 *and the target manifold N has dimension* 2.

It is interesting to note that this universal integrability holds precisely in the first dimensions in which regularity fails. In fact if $n = 1$, with any m, then we are dealing with weak solutions of a single uniformly elliptic equation, which are as smooth as the coefficients allow (for reference see [G, p. 53]). If the domain dimension $m = 1$, then harmonic mappings become geodesics with constant speed parametrization, and weak solutions are again smooth [M, p. 28]. In the case of a two-dimensional domain, we may refer to Morrey's result on general variational problems [M, Theorem 1.10.4(iii) and pp. 34–37], from which it may be seen that a weakly harmonic mapping is as regular as the target manifold N.

It is also interesting to note that the method of [S1] requires analyticity of N to conclude uniqueness of the limit map f_0, whereas the method used here gives uniqueness and fast convergence without assuming analyticity.

The first author would like to thank the Consiglio Nazionale delle Ricerche for its hospitality at the University of Trento. The second author would like to acknowledge the support of the Institute for Mathematics and its Applications at the University of Minnesota, and of the Alfred P. Sloan Foundation.

1. An Example of Logarithmic Convergence

We begin by presenting an example of a stationary harmonic mapping $f: M^3 \to N^3$, having a single point of discontinuity, where the domain M and the target manifold N each have dimension three. Once this example is constructed, it may be extended to form examples $f_1: M_1^m \to N_1^n$ of harmonic mappings for arbitrary dimensions $m \geqq 3$ and $n \geqq 3$, having the same rate of convergence to their homogeneous tangent mappings. In fact, we may choose $M_1 := M \times (S^1)^{m-3}$ and $N_1 := N \times R^{n-3}$ as Riemannian product manifolds and then define $f_1(x, \theta) = (f(x), 0) \in N_1$, where $(x, \theta) \in M \times (S^1)^{m-3}$, $\theta = (\theta_4, \ldots, \theta_m)$ and where $(u, t) \in N \times \mathbb{R}^{n-3}$, $t = (t_4, \ldots, t_n)$. Then the logarithmic convergence of f to its homogeneous tangent mapping f_0 will imply a similar property for f_1 (although the homogeneous tangent mapping for f_1 will have a singular set of the form $\mathbb{R}^{m-3} \subset \mathbb{R}^m$; see Remark 1.1 below).

We choose $M^3 = B_1^3 \subset \mathbb{R}^3$, the unit ball with the standard, Euclidean metric, and refer to standard coordinates $x = (x^1, x^2, x^3) = \varrho\omega$, $\omega \in S^2$, $\varrho \geqq 0$. The target manifold N^3 shall be a hypersurface of revolution in $\mathbb{R}^4$, generated by the curve $r = \Gamma_0(z)$ in the (r, z)-plane:

$$N = \{(v, z) \in \mathbb{R}^3 \times \mathbb{R} : |v| = \Gamma_0(z)\}.$$

We shall also write $v = r\omega \in \mathbb{R}^3$ where $\omega \in S^2, r \geqq 0$. For simplicity, we may replace the coordinate z by the arc-length parameter $u = u(z)$ for the generating curves: $(du/dz)^2 = 1 + (d\Gamma_0/dz)^2$, and define $\Gamma(u)$ such that $\Gamma(u(z)) := \Gamma_0(z)$. Then the coordinates $(\omega, u) \in S^2 \times \mathbb{R}$ may be used to describe the Riemannian metric of N induced from the Euclidean metric of $\mathbb{R}^4$:

$$ds_N^2 = \Gamma(u)^2\, ds_\Sigma^2(\omega) + du^2,$$

where $\omega \in S^2$, $u \in \mathbb{R}$ and ds_Σ^2 is the canonical metric of constant Gauss curvature 1 on the sphere $\Sigma = S^2$. We consider the $O(3)$-equivariant mapping $f(x) = f(\varrho\omega) = (\omega, u(\varrho)) \in N$ in terms of the coordinates $(\omega, u) \in S^2 \times \mathbb{R}$ for N, determined by a real-valued function $u = u(\varrho)$ of one real variable. Then the energy of f may be computed in terms of $u(\varrho)$:

$$\begin{aligned} E(f) &= \int_M [(du/d\varrho)^2 + 2(\Gamma(u)/\varrho)^2]\varrho^2 d\varrho\, d\mathrm{vol}_\Sigma(\omega) \\ &= 4\pi \int_0^1 [\varrho^2 (du/d\varrho)^2 + 2\Gamma(u)^2]\, d\varrho\,, \end{aligned} \tag{1.1}$$

as follows from (0.1). The Euler-Lagrange equations may be computed directly from this formula for $E(f)$, to show that a continuous mapping $f: M \to N$ of the form $f(\varrho\omega) = (\omega, u(\varrho))$ is stationary for E if and only if $u(\varrho)$ is a weak solution of

$$\frac{d}{d\varrho}\left(\varrho^2 \frac{du}{d\varrho}\right) = 2\Gamma(u(\varrho))\Gamma'(u(\varrho)). \tag{1.2}$$

The reader will observe that the ordinary differential Eq. (1.2) has a singular point at $\varrho = 0$.

Let us consider in particular the family of functions

$$u(\varrho) = \frac{1}{\sqrt{C - 2\log\varrho}}, \qquad 0 \leqq \varrho \leqq 1, \tag{1.3}$$

for various constants $C \geqq 0$. As $\varrho \to 0$, $u(\varrho)$ converges to zero more slowly than any positive power of ϱ. A direct computation yields $d(\varrho^2 du/d\varrho)/d\varrho = u^3 + 3u^5$ for each value of C. This computation and (1.2) lead us to consider the specific function

$$\Gamma(u) = \sqrt{1 + u^4/4 + u^6/2}. \tag{1.4}$$

This choice for Γ leads to $\Gamma(u) = \Gamma_0(z)$, via a hyperelliptic integral, defined for all $-\infty < z < \infty$. In other words, (1.4) corresponds to a *complete hypersurface of revolution* $N^3 \subset \mathbb{R}^4$.

The homogeneous tangent mapping $f_0 : \mathbb{R}^3 \to N$ is given by $f_0(\varrho\omega) = (\omega, 0)$, which is an isometric parameterization of the totally geodesic sphere $\Sigma_0 = \{(\omega, u) \in N : \omega \in S^2, u = 0\}$ at the narrowest point of N. Recalling the Allard-Almgren theorem stated above, it is of interest to consider whether harmonic Jacobi fields $\phi : S^2 \to TN$ along the restricted mapping $f_0 : S^2 \to N$ are integrable. In fact, *any* harmonic mapping $f: S^2 \to N$ must have its image in Σ_0. Namely, the parallel sphere $\Sigma_K := \{(\omega, u) \in N : \omega \in S^2, u = K\}$ has principal curvature vectors, as a submanifold of N, with negative or positive component in the direction of $\partial/\partial z$, when $\Gamma_0'(z)$ is positive or negative, respectively. With Γ as in (1.4), we have $u\Gamma'(u) > 0$ for $u \neq 0$. It follows that $|u|$ may not have a positive local maximum along a harmonic mapping, and in particular, $u \equiv 0$ for any harmonic mapping $g: S^2 \to N$, by compactness of S^2. On the other hand, $\Gamma''(0) = 0$ implies that $f_t(\omega) := (\omega, t) \in N$ is harmonic *to first order* at $t = 0$, which means that $\partial/\partial z = \partial f_t(\omega)/\partial t$, at $t = 0$, is a harmonic-Jacobi field along f_0. In particular, the integrability hypothesis is violated.

It is apparent that the mapping $f: B_1 \to N$ we have constructed is not continuous at 0; for this reason, we have yet to show that it is a weak solution of the Euler-Lagrange Eqs. (0.3). To show that $f \in H^1(M, N)$, it is enough to show that the energy integral (1.1) is finite. But $\varrho du/d\varrho = u^3$ for the family of functions (1.3), which implies that the integrand of (1.1) is uniformly bounded on $0 \leqq \varrho \leqq 1$. Next, recall that f is a weak solution of (0.3) provided that for all $h \in C^\infty(M, \mathbb{R}^4)$ with compact support, there holds

$$\int_M \gamma^{\alpha\beta}(\langle D_\alpha f, D_\beta h\rangle + \langle B(D_\alpha f, D_\beta f), h\rangle) d\mathrm{vol}_M = 0. \tag{1.5}$$

Write $h = h_1 + h_0$, where $h_1(x) = 0$ for $|x| \leqq \varepsilon$. Then it follows from (1.2) that (1.5) holds with h replaced by h_1, since f is smooth on $\operatorname{supp} t(h_1)$. By choosing h_0 to be h

times a standard cutoff function, we may achieve that $h_0(x)=0$ for $|x|\geqq 3\varepsilon$ and that $|Dh_0|\leqq|Dh|+|h|/\varepsilon$. Since h is uniformly bounded, we have $\|h\|_{L^2(B_{3\varepsilon})}\leqq C\varepsilon^{3/2}$, and in particular, $h_0\to 0$ in $H^1(M,\mathbb{R}^4)$ as $\varepsilon\to 0$, implying that $\int_M\langle D_\alpha f, D_\beta h_0\rangle dx\to 0$. Meanwhile, since $f\in H^1(M,N)$, we have $B(D_\alpha f, D_\beta f)\in L^1(M,\mathbb{R}^4)$ and hence

$$\left|\int_M\langle B(D_\alpha f, D_\beta f), h_0\rangle dx\right|\leqq\|h\|_{L^\infty}\int_{B_{3\varepsilon}}|B(D_\alpha f, D_\beta f)|dx\to 0\,.$$

This shows that f is a weak solution of the Euler-Lagrange equations. We have proved the following

Theorem 1. *Given any $m, n\geqq 3$, there is a real-analytic Riemannian manifold N^n, and a harmonic mapping $f: B_1^m\to N^n$ with a discontinuity at $O\in M$, such that $f(\varrho\omega)\to f_0(\omega)$ and $\varrho D_\varrho f(\varrho\omega)\to 0$, as $\varrho\to 0$, uniformly in $C^2(S^2)$, both more slowly than any positive power of ϱ.*

Remark 1.1. As observed above, for $m>3$ the example leads to a homogeneous tangent map with singularities on an $\mathbb{R}^{m-3}$. The analysis of [AA] and of [S1] is problematic in this case. However, we may modify the example to construct an *isolated* singularity, by defining $M=B_1^m\subset\mathbb{R}^m$ and $N=\{(v,z)\in\mathbb{R}^m\times\mathbb{R}: |v|=\Gamma_0(z)\}$ in close analogy to the example above. Let $f: M\to N$ be the $O(m)$-invariant mapping $f(\varrho\omega)=(\omega, u(\varrho))$ where $u(\varrho)$ belongs to the one-parameter family (1.3). Choose the hypersurface of revolution N so that

$$\Gamma(u)=\sqrt{1+\frac{m-2}{2(m-1)}u^4+\frac{1}{m-1}u^6}\,, \tag{1.6}$$

in terms of the arc-length parameter u. Then $f: M^m\to N^m$ is a stationary harmonic mapping as in Theorem 1, which moreover has an *isolated* singularity. This carries over to any dimensions $n\geqq m\geqq 3$.

Remark 1.2. It might be noted that both theorems on convergence to the homogeneous tangent map, as stated in the introduction, require f to be locally mimimizing, while our examples are only stationary (see however Remark 2.1 below). In fact, it is a rather nontrivial exercise to prove that any specific discontinuous mapping minimizes energy. Recently, Schoen and Brézis-Coron-Lieb have announced independent proofs that the mapping $f(x)=x/|x|$ from the Euclidean ball to S^2 has minimum energy. In analogy with these results, we expect that the examples with isolated singularities just constructed have minimum energy with respect to their Dirichlet boundary data.

Remark 1.3. Suppose that the function $\Gamma(u)$ is an *arbitrary* real-analytic function which assumes its positive minimum value at the unique critical point $u=0$, which is degenerate: $\Gamma''(0)=0$. Let $k+2\geqq 4$ be the order of the first nonzero derivative of Γ at 0 (k must be even). Then there are solutions of the ordinary differential Eq. (1.2) with $u(\varrho)\to 0$. Specifically, these solutions have the asymptotic behaviour $u(\varrho)(C-ka\log\varrho)^{1/k}\to 1$ as $\varrho\to 0$, for some real constants a and C. This behavior may be proved by first finding an invariant manifold of the form $\varrho du/d\varrho=\Phi(u)$, where Φ is a function of the form $\Phi(u)=au^{k+1}+O(u^{k+2})$, and where $2\Gamma(u)\Gamma'(u)$ has the same leading term.

2. The Integrability Hypothesis for $m=3$ and $n=2$

As observed in the introduction, a harmonic mapping: $M^m \to N^n$ must be smooth if the target dimension $n=1$, if the domain dimension $m=1$ or if $m=2$. On the other hand, if m and n are both $\geqq 3$, then as we have just seen, there are discontinuous harmonic mappings with only logarithmic convergence to their homogeneous tangent mappings. This leaves exactly one pair of dimensions to be investigated: $m=3$ and $n=2$. In these dimensions, the homogeneous tangent mapping is a harmonic mapping from S^2 into N^2, which may be expected to show a rigidity not apparent in higher dimensions. In fact, we have the following

Theorem 2. *Let M^3 and N^2 be Riemannian manifolds of dimension 3 and 2, respectively. Let $f: M^3 \to N^2$ be a locally minimizing harmonic map near $O \in M$. Then f converges to a unique homogeneous tangent mapping f_0 at a rate controlled by a positive power of the distance ϱ from 0, as in inequality (0.6). Moreover, if f has a discontinuity at 0, then N has the topological type of the two-dimensional sphere or projective plane.*

In order to apply the Allard-Almgren theorem stated in the introduction, we need the integrability hypothesis for an arbitrary harmonic mapping $f_0: S^2 \to N^2$. First, we shall observe that f_0 necessarily enjoys a much stronger property than harmonicity, as is widely known. We write g in place of f_0 for the three lemmas.

Lemma 1. *Let Σ^2, N^2 be two-dimensional Riemannian manifolds, $N^2 \subset \mathbb{R}^d$, and $g \in H^1(\Sigma, N)$ a harmonic mapping. If Σ has the topological type of S^2 or of $\mathbb{R}P^2$, then g is a conformal mapping. If moreover N is not of the topological type of S^2 or $\mathbb{R}P^2$, then g is constant.*

For completeness, we give a proof for Lemma 1. Let $z \in \mathbb{C}$ be a local conformal parameter for Σ: that is, $z=x+iy$ where $ds_\Sigma^2 = \lambda(z)^2(dx^2+dy^2)$, and $\lambda(z)>0$. The existence of z follows from the uniformization theorem. Then as a mapping into $\mathbb{R}^d$, g has the differential

$$dg = g_x dx + g_y dy = g_z dz + g_{\bar z} d\bar z\,,$$

where the subscripts denote partial derivatives, and the complex partial derivatives are defined by $g_z = (g_x - ig_y)/2$ and $g_{\bar z} = (g_x + ig_y)/2$, as usual. Let the Euclidean inner product $\langle\ ,\ \rangle$ of $\mathbb{R}^d$ be extended as a *complex-bilinear* form on $\mathbb{C}^d$, and similarly, let the second fundamental form B of N, as a submanifold of $\mathbb{R}^d$, be extended as a symmetric bilinear tensor on the complexified tangent bundle to N, with values in its complexified normal bundle. We may compute $\lambda^2 \Delta_\Sigma g = -4g_{z\bar z}$ and $\lambda^2 \gamma^{\alpha\beta} B(D_\alpha g, D_\beta g) = 4B(g_z, g_{\bar z})$. The Eq. (0.3) for a harmonic mapping becomes

$$g_{z\bar z} = B(g_z, g_{\bar z})\,. \tag{2.1}$$

As noted in the introduction, since Σ has dimension 2, a weak solution $g \in H^1(\Sigma, N)$ must be of class C^2.

Recall that a *conformal* mapping $g: \Sigma \to N$ is one which preserves the Riemannian metric up to a variable factor $\sigma: \Sigma \to [0, \infty)$, that is, such that $g^* ds_N^2 = \sigma(z) ds_\Sigma^2$. For a conformal parameter $z=x+iy$, this is equivalent to $|g_x|^2 = |g_y|^2 (= \sigma\lambda^2)$ and $\langle g_x, g_y\rangle = 0$. These two equations may be written in complex notation as $\langle g_z, g_z\rangle = 0$.

We shall first show that $\langle g_z, g_z\rangle$ is locally *holomorphic*, or equivalently, that $\langle g_z, g_z\rangle_{\bar z}=0$. But $\langle g_z, g_z\rangle_{\bar z}=2\langle g_z, g_{z\bar z}\rangle=2\langle g_z, B(g_z, g_{\bar z})\rangle$ by Eq. (2.1). On the other hand, the real and imaginary parts of g_z are tangent vectors to N, while $B(g_z, g_{\bar z})$ is a (real) normal vector, so this last quantity vanishes identically. This shows that $\langle g_z, g_z\rangle$ is holomorphic on the domain of definition of the parameter z.

Now if Σ has the topological type of $\mathbb{R}P^2$, then we may compose g with the covering projection from the Riemannian universal covering space of Σ. Thereby, we may assume that Σ has the topological type of S^2. It follows from the uniformization theorem that Σ is conformally equivalent to the standard $S^2=\mathbb{C}P^1$, and in particular, we may cover Σ by two conformal coordinate charts $z:\mathbb{C}\to\Sigma$ and $\zeta:\mathbb{C}\to\Sigma$ satisfying $z\zeta\equiv 1$. By the chain rule, we obtain $z^2\langle g_z, g_z\rangle=\zeta^2\langle g_\zeta, g_\zeta\rangle$. Now $\langle g_\zeta, g_\zeta\rangle$ is a holomorphic function at $\zeta=0$, which implies that $|\langle g_z, g_z\rangle|\leqq C|z|^{-4}$ for $|z|$ sufficiently large, and hence $\langle g_z, g_z\rangle\equiv 0$ by the maximum principle. This shows that g is a conformal mapping (some readers may prefer to invoke the Riemann-Roch theorem here).

Finally, if N^2 is not topologically a sphere or projective plane, then its universal cover $\tilde N$ is conformally the disk or the plane, and g lifts to a conformal mapping $\tilde g: S^2\to\tilde N$. But a nonconstant conformal mapping must be an open mapping, which would imply that $\tilde g(S^2)$ is both compact and open as a subset of $\tilde N$, a contradiction. Therefore $\tilde g$ and g must be constant mappings. This finishes the proof of Lemma 1. □

In analogy with harmonic-Jacobi fields along a harmonic mapping, we may define a vector field $\varphi:\Sigma\to TN$ along a conformal mapping $g:\Sigma^2\to N^2$ to be a *conformal-Jacobi field* (or infinitesimal conformal mapping) if it satisfies

$$\langle \varphi_z, g_z\rangle = 0. \tag{2.2}$$

This is immediately seen to be equivalent to the vanishing of the first variation of $\langle g_z, g_z\rangle$ when g is varied in the direction of φ.

Lemma 2. *Let Σ, N be two-dimensional Riemannian manifolds, $g:\Sigma\to N$ a harmonic mapping, and $\varphi:\Sigma\to TN$ a harmonic-Jacobi field along g. If Σ has the topological type of S^2 or of $\mathbb{R}P^2$, then g is a conformal mapping, and φ is a conformal-Jacobi field.*

We shall show that $\langle\varphi_z, g_z\rangle$ is a holomorphic function on the domain of any conformal parameter z. Lemma 2 will then follow as in the proof of Lemma 1.

We may rewrite (0.4) in the form

$$\varphi_{z\bar z}=B(\varphi_z^T, g_{\bar z})+B(g_z, \varphi_{\bar z}^T)+D_\varphi B(g_z, g_{\bar z}). \tag{2.3}$$

Observe that φ_z is the same as $D_{g_z}\varphi$, since φ is a vector field along g. The component of φ_z normal to N is therefore $\varphi_z^\perp=(D_{g_z}\varphi)^\perp=:B(g_z,\varphi)=B(\varphi, g_z)=(D_\varphi g_z)^\perp$. Further, since the values of B are normal vectors to N, we have $\langle g_z, B(g_z, g_{\bar z})\rangle=0$. Differentiating this last expression in the direction of φ yields

$$\langle g_z, D_\varphi B(g_z, g_{\bar z})\rangle = -\langle D_\varphi g_z, B(g_z, g_{\bar z})\rangle = -\langle \varphi_z, B(g_z, g_{\bar z})\rangle.$$

Using twice more the fact that B has values normal to N, we obtain from (2.3) that

$$\langle g_z, \varphi_{z\bar{z}}\rangle = \langle g_z, D_\varphi B(g_z, g_{\bar{z}})\rangle = -\langle \varphi_z, B(g_z, g_{\bar{z}})\rangle .$$

Recalling (2.1), we conclude that $\langle g_z, \varphi_z\rangle_{\bar{z}} = 0$, so that $\langle g_z, \varphi_z\rangle$ is holomorphic. This completes the proof of Lemma 2. □

Lemma 3. *Let $g: \Sigma \to N$ be a nonconstant (branched) conformal mapping of degree d, where Σ is topologically S^2 and N is topologically S^2 or $\mathbb{R}P^2$. Then there is a $(4d+2)$-parameter family*

$$f: (U \subset \mathbb{R}^{4d+2}) \times \Sigma \to N$$

of conformal mappings with $f(0, \cdot) = g(\cdot)$ such that every conformal-Jacobi field along g is equal to

$$\frac{d}{dt} f(tv, \cdot)(t=0)$$

for some $v \in \mathbb{R}^{4d+2}$.

To begin the proof of Lemma 3, we may assume that N is topologically S^2 (otherwise lift g to the universal cover). Recall that Σ and N are conformally equivalent to $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$; we write conformal diffeomorphisms $T_1: \hat{\mathbb{C}} \to \Sigma$ and $T_2: \hat{\mathbb{C}} \to N$. Then g is represented by a (branched) conformal mapping $w_0: \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ (That is, $T_2 \circ w_0 = g \circ T_1$). The conformality relation $\langle g_z, g_z\rangle = 0$ implies that $w_{0z} = 0$ or $w_{0\bar{z}} = 0$. We may orient Σ so that $w_{0\bar{z}} = 0$ holds. Then as is well known, w_0 must be a rational function: $w_0(z) = P(z)/Q(z)$ for some polynomials P, Q having no common factor other than constants. Since g has mapping degree d,

$$\max\{\deg P, \deg Q\} = d .$$

Suppose that $w_0(\infty) \neq \infty$, or equivalently,

$$\deg(P) \leqq \deg(Q) = d$$

(otherwise modify T_2). Note that w_0 belongs to the $(4d+2)$ parameter family

$$w_{A,B} = \frac{P+A}{Q+B},$$

where A and B are polynomials with degree $(A) \leqq d$ and degree $(B) \leqq d-1$. Each $w_{A,B}$ corresponds to a conformal map $g_{A,B}$, and $g_{0,0} = g$.

Now let φ be a conformal-Jacobi field along g. If $w_0(z) \neq \infty$, then for some real $\alpha(z)$ and $\beta(z)$

$$T_{2*}^{-1}(\varphi(T_1(z)) = \alpha(z) D_u(w_0(z)) + \beta(z) D_v(w_0(z)) ,$$

where $w = u + iv$ and $\{D_u, D_v\}$ is the coordinate basis of vectorfields on $\mathbb{C}$. Write $\gamma(z) := \alpha(z) + i\beta(z)$. The equation $\langle g_z, \varphi_z\rangle = 0$ for a conformal-Jacobi field becomes $\alpha_z u_{0z} + \beta_z v_{0z} = 0$ where $w_0(z) := u_0(z) + iv_0(z)$. This is immediately equivalent to $\gamma_{\bar{z}} = 0$ since w_0 has only isolated branch points. Thus φ is represented by the function γ, which is holomorphic except at the poles of w_0.

We need some information about these singularities. Near any pole z_1 of w_0 (so $w_0(z_1)=\infty$) we may consider the conformal parameter $\hat{w}:=1/w$. Then g is represented by the meromorphic function $\hat{w}_0$ with $\hat{w}_0(z_1)=0$. The vector field φ is represented in terms of $D_{\hat{u}}$ and $D_{\hat{v}}$ by a locally holomorphic function $\hat{\gamma}(z)$. It follows from the chain rule that $\gamma(z)=-w_0(z)^2\hat{\gamma}(z)$. Therefore, γ has at most a pole at z_1, whose order is at most twice the order of the pole of w_0. It follows that $\gamma(z)$ may be written *globally* as

$$\gamma(z)=R(z)/Q(z)^2$$

for some complex polynomial $R(z)$. Also, since $w_0(\infty)\neq\infty$, $\gamma(\infty)\neq\infty$ and thus $\deg(R)\leqq\deg(Q^2)=2d$. Now we must show that for some polynomials A and B with $\deg A\leqq d$ and $\deg B\leqq d-1$,

$$\varphi=\frac{d}{dt}g_{tA,tB}(t=0)$$

i.e.,

$$\gamma=\frac{d}{dt}(P+tA)/(Q+tB)=\frac{AQ-BP}{Q^2}$$

i.e.,

$$R=AQ-BP.$$

Now since the complex polynomials form a Euclidean domain, and since P and Q are relatively prime, there exist polynomials A_1 and B_1 such that

$$R=A_1Q-B_1P.$$

Now divide B_1 by Q to get polynomials S and B such that

$$B_1=SQ+B$$
$$\deg(B)<\deg(Q).$$

Then, letting $A=A_1-SP$, we have

$$R=AQ-BP.$$

Since $\deg(R)\leqq 2d$ and $\deg(BP)<\deg(Q^2)=2d$, it follows that $\deg(A)\leqq d$. This completes the proof of Lemma 3. □

Theorem 2 is a direct consequence of the three lemmas. In fact, since f is locally energy minimizing, it has at least one homogeneous tangent map $f_0:\mathbb{R}^3\to N$ [SU, p. 314]. If N is not S^2 or $\mathbb{R}P^2$, then f_0 is constant by Lemma 1, and therefore f is C^α (indeed smooth) in a neighborhood of 0 [SU, p. 315]. In particular, (0.6) holds.

On the other hand, if N is S^2 or $\mathbb{R}P^2$ and f_0 is not constant, then by Lemmas 1–3, it satisfies the integrability hypothesis (0.5) of Simon's version of the Allard-Almgren theorem. This completes the proof. □

Remark 2.1. Theorem 2 (as well as the theorems quoted in the introduction) also apply to mappings that are harmonic but not necessarily energy minimizing. In that case, however, one must assume that at least one blow-up sequence $f_{\lambda(i)}$ converges strongly (say in C^2 on compact subsets of $\mathbb{R}^m \setminus \{0\}$) to the homogeneous limit f_0.

References

[G] Giaquinta, M.: Multiple integrals in the calculus of variations and nonlinear elliptic systems. Ann. Math. Study 105. Princeton (1983)

[SU] Schoen, R., Uhlenbeck, K.: A regularity theory for harmonic maps. J. Differ. Geom. **17**, 307–335 (1982)

[GM] Giusti, E., Miranda, M.: Sulla regolarità delle solutioni deboli di una classe di sistemi ellittici quasilineari. Arch. Rat. Mech. Anal. **31**, 173–184 (1968)

[M] Morrey, C.B.: Multiple integrals in the calculus of variations. Berlin Heidelberg New York: Springer 1966

[AA] Allard, W., Almgren, F.: On the radial behaviour of minimal surfaces and the uniqueness of their tangent cones. Ann. Math. **113**, 215–265 (1981)

[S1] Simon, L.: Asymptotics for a class of non-linear evolution equations, with applications to geometric problems. Ann. Math. **118**, 525–571 (1983)

[S2] Simon, L.: Isolated singularities of extrema of geometric variational problems. Lecture Notes Math., Vol. 1161, pp. 206–277. Berlin Heidelberg New York: Springer 1985

[dG] De Giorgi, E.: Un esempio di estremali discontinue per un problema variazionale di tipo ellittico. Boll. U.M.I. **4**, 135–137 (1968)

[HW] Hildebrandt, S., Widman, K.-Ø.: Some regularity results for quasilinear elliptic systems of second order. Math. Z. **142**, 67–86 (1975)

Received December 2, 1986; in revised form August 5, 1987

Math. Ann. 283, 551–572 (1989)

Mathematische Annalen

The Symmetric-Square L-Function Attached to a Cuspidal Automorphic Representation of GL_3

S. J. Patterson[1] and I. I. Piatetski-Shapiro[2]

[1] Mathematisches Institut der Universität, Bunsenstrasse 3–5, D-3400 Göttingen, Federal Republic of Germany
[2] Department of Mathematics, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69973, Israel

In [13] Rankin introduced a new method into the theory of automorphic forms which he used to determine the analytic properties of $\sum_{n \geqq 1} \tau(n)^2 n^{-s}$ where τ is the Ramanujan function. We can reformulate his results as follows. Let π_Δ be the automorphic representation of $PGL_2(\mathbf{Q}_\mathbf{A})$ associated with Δ. Let $L(s, \pi, r)$ be the L-function associated with an automorphic representation π of $G_\mathbf{A}$ where G is a reductive algebraic group over the base field k, and is a finite dimensional representation of the L-group ${}^L G$. Then Rankin proved that

$$s \mapsto L(s, \pi_\Delta, \mathrm{Sym}^2)\zeta(s)$$

where Sym^2 is the symmetric square representation $SL_2(\mathbf{C})$ and ζ is the zeta function of $\mathbf{Q}$ (including the archimedean factor), has an analytic continuation as a meromorphic function into $\mathbf{C}$. It is invariant under the replacement of s by $1-s$. The poles of this function are located at $s=0,1$ and are simple.

The same method was rediscovered shortly after Rankin's work by Selberg [14] and is usually called the Rankin-Selberg method.

At the Antwerp conference in 1972 Shimura [16] described a variant of the Rankin-Selberg method which yielded the Dirichlet series $\sum_{n \geqq 1} \tau(n^2) n^{-s}$, i.e. $L(s, \pi_\Delta, \mathrm{Sym}^2)$ which he showed to be holomorphic. This was a significant sharpening of Rankin's result and had important consequences – see, for example, [4].

The method of Rankin-Selberg can be generalized very substantially; see [9, 12] for more details. In particular, if π is an automorphic representation of $GL_r(k_\mathbf{A})$ with central quasicharacter χ one has an Euler product $L(s, \pi \times \pi)$ which has an analytic continuation as a meromorphic function into the entire plane. If π is unitary then $s=1$ is a simple pole of $L(s, \pi \times \pi)$ if and only if $\pi \cong \check{\pi}$, the contragredient of π; still assuming π to be unitary the function $L(s, \pi \times \pi)$ is holomorphic except possibly at $s=0,1$, and there is a functional equation relating $L(s, \pi \times \pi)$ and $L(1-s, \check{\pi} \times \check{\pi})$.

Since $L(s, \pi \times \pi) = L(s, \pi, \mathrm{Ten}^2)$ where Ten^2 denotes the tensor square of the standard representation of $GL_r(\mathbb{C})$, and since

$$\mathrm{Ten}^2 = \mathrm{Sym}^2 \oplus \Lambda^2$$

(Sym^2 = symmetric square, Λ^2 = alternating square) gives the decomposition of Ten^2 into irreducible representations we have

$$L(s, \pi \times \pi) = L(s, \pi, \mathrm{Sym}^2) \cdot L(s, \pi, \Lambda^2).$$

In particular, when $r = 3$ we have moreover

$$L(s, \pi, \Lambda^2) = L(s, \check{\pi} \otimes \chi)$$

where $L(s, \pi)$ is the L-function of [8].

Since the analytic continuation and functional equation of $L(s, \pi \otimes \pi\chi)$ are also known we deduce the analytic continuation as a meromorphic function, and the functional equation of $L(s, \pi, \mathrm{Sym}^2)$. What does not follow from this is that $L(s, \pi, \mathrm{Sym}^2)$ should have only finitely many poles. This has, however been proved by a different method by Shahidi [15]. Indeed if π is cuspidal one might expect $L(s, \pi, \mathrm{Sym}^2)$ to be entire, but this is not the case for the following reason. Let π_1 be a cuspidal automorphic representation of $GL_2(k_{\mathbf{A}})$ with central quasicharacter χ_1; let π be the Gelbart-Jacquet [4] lift of π_1 to $GL_3(k_{\mathbf{A}})$. Then π is cuspidal and

$$L(s, \pi, \mathrm{Sym}^2) = L(s, \pi_1, \mathrm{Sym}^4) \cdot L(s, \chi_1^2)$$

which has a pole at $s = 1$ if $\chi_1^2 = 1$. In fact $L(s, \pi \times \pi)$ has a pole at $s = 1$ precisely when $\pi \cong \check{\pi}$; see [9, p. 368]. The condition $\chi_1^2 = 1$ ensures that the condition $\pi \cong \check{\pi}$ is fulfilled. Conversely, by a theorem of Flicker's [3, II, Theorem 2.9] this condition implies that π is a lift of an automorphic representation π_1 with $\chi_1^2 = 1$. Thus the only case in which $L(s, \pi, \mathrm{Sym}^2)$ can have a pole is that which we have just discussed.

It is the objective of this paper to give a integral representation of $L(s, \pi, \mathrm{Sym}^2)$ analogous to the one given by Shimura in the case of GL_2 in [15, 17] (see also [4] for an adelic representation-theoretic account of Shimura's method). Shimura's method is based on the consideration of the Rankin-Selberg convolution of an automorphic form of the representation π with a theta-function. Here we shall do the same, although the theta-function can no longer be constructed using the Weil representation but by the technique of Eisenstein series used in [10]. The integral representation is given in Proposition 3.2 and Corollary 4.2. A peculiarity of our method is that we exploit the fact that the representation to which the theta-function belongs is degenerate i.e. it has no Whittaker model.

From this integral representation we would expect to be able to deduce that $L(s, \pi, \mathrm{Sym}^2)$ has only finitely many poles. Unfortunately we do not reach this goal in general. Let $L_v(s, \pi, \mathrm{Sym}^2)$ denote the v^{th} factor of $L(s, \pi, \mathrm{Sym}^2)$. We shall be able to limit the poles of a function

$$\prod_{v \in S} M_v(s) \cdot \prod_{v \notin S} L_v(s, \pi, \mathrm{Sym}^2)$$

where M_v runs through a certain vector space. Unfortunately we have not been able to show that for a archimedean v this space has no common zeros in the halfplane $\mathrm{Re}(s) \geqq \frac{1}{2}$. The "local integrals" M_v have a more complicated structure than is usual since they involve three representations, all of which are infinite dimensional.

It also follows from our method that when the characteristic of k is neither 0 or 2 that $L(s, \pi, \mathrm{Sym}^2)$ has a pole at $s=1$ if and only if

$$\int_{G_k \backslash G_{\mathbf{A}} / Z_{\mathbf{A}}} \varphi(g)\theta_1(g)\theta_2(g)dg \neq 0$$

where φ is some automorphic form of π and θ_1, θ_2 are certain "theta functions," albeit not ones arising from a Weil representation. In view of the discussion above this integral is not identically zero when π is a lift of an automorphic representation of $GL_2(k_{\mathbf{A}})$ for which χ_1^2 is the fourth power of a Größencharakter. We presume that the same holds when the characteristic of k is 0 but we have not been able to overcome the technical difficulties (see Proposition 5.3 and the remarks following it).

For our construction we shall have to make use of the theory of Eisenstein series. In Sect 2 we shall recall those facts which we shall need. In Sect. 3 we shall construct the Rankin-Selberg integral which is the central subject of this paper. In Sect. 4 we shall evaluate the local integral at a generic place. In Sect. 5 we prove the appropriate local functional equation. Finally in Sect. 6 we shall summarize our results.

We should note that Shahidi has given in [15] a quite different approach to the investigation of L-functions of the type considered here. He derives these results from the theory of Eisenstein series, and he proves results which are much more general than ours, and in this special case, more precise [15, Corollary 6.7]. Nevertheless the method described here is interesting in itself and may well have applications which the method of [15] cannot have.

In this paper k will denote a fixed $\mathbf{A}$-field of characteristic $\neq 2$. We shall denote for an algebraic group G defined over k the group of k-points of G by G_k, *of* k_v-points by G_v where v is a place of k, and of $k_{\mathbf{A}}$-points by $G_{\mathbf{A}}$ where $k_{\mathbf{A}}$ denotes the adele ring of k. Let $\Sigma(k)$ denote the set of places of k; let $\Sigma_\infty(k)$ (resp. $\Sigma_f(k)$) be the subset of archimedean (resp. non-archimedean) places.

We shall work with GL_2, GL_3 and subgroups of these. Let P (resp. Q) be the standard $(2,1)$ (resp. $(1,2)$) parabolic subgroup of GL_3. Let N denote the upper triangular unipotent subgroup of GL_3 or GL_2 (it will be clear from the context which is meant). Let $M(P)$, $M(Q)$ be the standard Levi factor of P, Q; let $N(P) = M(P) \cap N$, $N(Q) = M(Q) \cap N$. Let $U(P)$, U(Q) be the unipotent radical of P, Q. Let H denote the diagonal subgroups of GL_2 or GL_3; again it will be clear from the context which is intended. Let M be the normalizer of H, $W = M/H$. Let Z be the centre of GL_2 or GL_3. Let $B = H \cdot N$.

Over a local field F let μ (resp. $\mu(P)$, $\mu(Q)$) denote the square root of the modulus of the adjoint action of $H(F)$ (resp. $M(P)(F)$, $M(Q)(F)$) on the Lie algebra over F of N (resp. $U(P)$, $U(Q)$). This also yields positive quasi-cahracters $\mu_{\mathbf{A}}$ of $H_{\mathbf{A}}$, $\mu_{P,\mathbf{A}}$ of $M(P)_{\mathbf{A}}$ and $P_{\mathbf{A}}$, $\mu_{Q,\mathbf{A}}$ of $M(Q)_{\mathbf{A}}$ and $Q_{\mathbf{A}}$.

Let $\mu_2 = \{\pm 1\}$. Let us denote by $\tilde{G}$ the 2-fold metaplectic over of G when this is defined; if $G_1 \subset G$ is such that there exists a natural splitting of this covering then we denote by G_1^* the corresponding isomorphic copy of G_1 in $\tilde{G}$. We shall take those facts which we need concerning metaplectic groups from [10].

We recall here briefly the description of the metaplectic covers of GL_r over local fields and rings of adeles. These covers are characterised by their restrictions to the diagonal subgroup H. Let F be a local field and let $(\cdot,\cdot)_F$ be the 2-Hilbert symbol on F. Let $\mathrm{diag}(a_1,\ldots,a_r)$ be the diagonal matrix with a_i as the ii^{th} entry. The function $c: F^\times \times F^\times \to \{\pm 1\}$ given by

$$c(\mathrm{diag}(a_1,\ldots,a_r),\mathrm{diag}(a'_1,\ldots,a'_r)) = \prod_{i<j} (a_i, a'_j)_F$$

defines a 2-cocycle on H which extends to a 2-cocycle of $GL_r(F)$. The centre of the covering group is the lift of Z_F where

$$Z_F = \{zI \mid z \in F^\times\} \quad (r\text{ odd})$$

and

$$Z_F = \{zI \mid z \in F^{\times 2}\} \quad (r\text{ even}).$$

In the case of GL_3 we can identify $M(P)$ (resp. $M(Q)$) with $GL_2(F).Z_F$ by embedding GL_2 in the upper (resp. lower) 2×2 diagonal block. As the lift of Z_F is the centre of $GL_3(F)$ the lifts of the two factors commute. This carries over to the adelic case. The covering groups are therefore given by 2-cocycles and are so endowed with a section which we shall denote by s.

Finally we shall write π (or a similar letter) for a class of representations. If π is automorphic we shall write π_v for the local component at v. We shall often choose a representation space V of π; on this we shall write the action of G on V as left-multiplication; i.e. $G\times V\to V$; $(g,v)\mapsto gv$. We shall take an automorphic representation to be irreducible unless the contrary is stated.

2. Eisenstein Series

We shall be concerned here with Eisenstein series associated with representations of the two-fold covers of $P_{\mathbb{A}}$ and $Q_{\mathbb{A}}$ induced to that of $GL_3(k_{\mathbb{A}})$. Let $\tilde{V}$ be an automorphic representation of $\widetilde{GL}_2(k_{\mathbb{A}})$ and let $L: \tilde{V}\to\mathbf{C}$ be a $GL_2(k)^*$-invariant linear form. Let

$$\Lambda(f) = \int_{N_k^*\backslash N_{\mathbb{A}}^*} L(nf)\,dn, \qquad f\in V$$

be the corresponding $N_{\mathbb{A}}^*$-invariant linear form. This is identically zero if $(\tilde{V},L)$ is cuspidal, otherwise not. Let χ be a quasicharacter of $\tilde{Z}_{\mathbb{A}}$ trivial on Z_k^* and such that $\chi|\mu_2$ is non-trivial. We shall also assume that μ_2 acts non-trivially on $\tilde{V}$; we say then that $\tilde{V}$ and χ are "genuine." Considering $\widetilde{GL}_2(k_{\mathbb{A}})\tilde{Z}_{\mathbb{A}}$ as $\tilde{M}(P)_{\mathbb{A}}$ or $\tilde{M}(Q)_{\mathbb{A}}$ as above we construct two new representations, which we denote by $\tilde{V}_P$ and $\tilde{V}_Q$, of $\tilde{M}(P)_{\mathbb{A}}$ and $\tilde{M}(Q)_{\mathbb{A}}$. The vector space is the original $\tilde{V}$, the action of $\widetilde{GL}_2(k_{\mathbb{A}})$ is the original one, the action of $\tilde{Z}_{\mathbb{A}}$ is by χ. The linear forms L, Λ yield linear forms on $\tilde{V}$ yield linear forms L_P, Λ_P (resp. L_Q, Λ_Q) on $\tilde{V}_P$ (resp. $\tilde{V}_Q$). Note that L_P (resp. L_Q) is $M(P)_k^*$- (resp. $M(Q)_k^*$)-invariant.

Let Ω_P (resp. Ω_Q) be the group of quasicharacters of $M(P)_{\mathbb{A}}$ (resp. $M(Q)_{\mathbb{A}}$) which are trivial on $M(P)_k Z_{\mathbb{A}}$) (resp. $M(Q)_k Z_{\mathbb{A}}$). Recall that Ω_P and Ω_Q have the structure of complex manifolds. For $\omega\in\Omega_P$ (resp. $\omega\in\Omega_Q$) the real number $\sigma(\omega)$ by $|\omega(x)| = \mu_P(x)^{\sigma(\omega)}$ (resp. $|\omega(x)| = \mu_Q(x)^{\sigma(\omega)}$). This is well-defined. For $\omega\in\Omega_P$ (resp. Ω_Q) we define $\tilde{\mathbf{V}}_P(\omega)=\tilde{V}_P\otimes(\omega\mu_P)$, $\tilde{\mathbf{V}}_Q(\omega)=\tilde{\mathbf{V}}_Q\otimes(\omega\mu_Q)$. We regard $\tilde{\mathbf{V}}_P(\omega)$ and $\tilde{\mathbf{V}}_Q(\omega)$ as the fibres of holomorphic vector bundles $\tilde{\mathbf{V}}_P$ and $\tilde{\mathbf{V}}_Q$ over Ω_P and Ω_Q respectively. We regard $\tilde{V}_P(\omega)$ and $\tilde{V}_Q(\omega)$ as representation spaces of $\tilde{P}_{\mathbf{A}}$ and $\tilde{Q}_{\mathbf{A}}$ on which $U(P)^*_{\mathbf{A}}$ and $U(Q)^*_{\mathbf{A}}$ act trivially. We define $\mathbf{F}_P(\omega)$ to be the space of functions $f:\widetilde{GL}_3(k_{\mathbf{A}})\to\tilde{V}_P(\omega)$ satisfying

$$f(\gamma g)=\gamma f(g),\qquad \gamma\in\tilde{P}_A,\qquad g\in\widetilde{GL}_3(k_A),$$

and which satisfies the usual smoothness conditions (cf. [10, II.1]). This is, by right multiplication a representation space of $\widetilde{GL}_3(k_{\mathbf{A}})$. We can construct $F_Q(\omega)$, $\omega\in\Omega_Q$ analogously. Note that for $f\in F_P(\omega)$ the map

$$f\mapsto L_P(f(g))$$

is a linear form left-invariant under $M(P)^*_k U(P)^*_A$. The spaces $\mathbf{F}_P(\omega)$ and $\mathbf{F}_Q(\omega)$ are the fibres of holomorphic vector bundles $\mathbf{F}_P$ and $\mathbf{F}_Q$ over Ω_P and Ω_Q. We represent sections $f\in\mathbf{F}_P(U)$, U an open subset of Ω_P, by $g\mapsto f(g,\omega)$, $g\in GL_3(k_{\mathbf{A}})$, $\omega\in U$ and analogously with Q instead of P.

There exists $c(\tilde{V})\in\mathbf{R}$ so that over the open sets $\{\omega\in\Omega_P|\sigma(\omega)>c(\tilde{V})\}$ and $\{\omega\in\Omega_Q|\sigma(\omega)>c(\tilde{V})\}$ we have maps

$$E_P:\mathbf{F}_P\to\mathcal{O}_P\quad\text{and}\quad E_Q:\mathbf{F}_Q\to\mathcal{O}_Q$$

where $\mathcal{O}_P$ (resp. $\mathcal{O}_Q$) is the structure sheaf of Ω_P (resp. Ω_Q). The maps E_P and E_Q (Eisenstein series) are defined by

$$E_P(f,\omega)=\sum_{\gamma\in P_k^*\backslash GL_3(k)^*}L_P(f(\gamma,\omega))$$

and

$$E_Q(f,\omega)=\sum_{\gamma\in Q_k^*\backslash GL_3(k)^*}L_Q(f(\gamma,\omega)).$$

Let $\mathcal{M}_P$ (resp. $\mathcal{M}_Q$) be the sheaf of meromorphic functions on Ω_P (resp. Ω_Q). Then E_P (resp. E_Q) can be continued to maps (over Ω_P (resp. Ω_Q))

$$E_P:\mathbf{F}_P\to\mathcal{M}_P,\qquad E_Q:\mathbf{F}_Q\to\mathcal{M}_Q.$$

This is one of the central results of the theory of Eisenstein series (see [11, p. 276]). One can also give a functional equation for the E_P, E_Q and describe the singularities. This we shall now discuss.

We fix an additive character e_0 of $k_{\mathbf{A}}$ which is non-trivial but trivial on k. We shall assume that measures on $k_{\mathbf{A}}$ (or k_v) are self-dual with respect to e_0 (resp. $e_{0,v}$). In particular the measure of $k_{\mathbf{A}}/k$ is 1.

We shall assume first that $\tilde{V}$ is cuspidal – we shall later assume that it is exceptional, that is, that $\tilde{V}$ is a cuspidal Weil representation (see [5, 3.3]; [6, 4.4]). Let

$$w_{123}=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix},\qquad w_{132}=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$$

so that

$$w_{123}M(P)w_{123}^{-1}=M(Q), \qquad w_{132}M(Q)w_{132}^{-1}=M(P).$$

We define $\Omega_P \to \Omega_Q$; $\omega \mapsto \tilde{\omega}$ where

$$\tilde{\omega}(x)=\omega(w_{132}xw_{132}^{-1})$$

and analogously $\Omega_Q \to \Omega_P$; $\omega \mapsto \tilde{\omega}$. Note that $\tilde{\tilde{\omega}}=\omega$. Also $\sigma(\omega)+\sigma(\tilde{\omega})=0$. There are intertwining operators defined if $\sigma(\omega)>c(\tilde{V})$

$$I_{PQ}:\tilde{\mathbf{V}}_Q(\omega)\to\tilde{\mathbf{V}}_P(\omega);\, f\mapsto\left(g\mapsto \int_{U(Q)^*_{\mathbb{A}}} f(w_{132}ng,\omega)dn\right)$$

$$I_{QP}:\tilde{\mathbf{V}}_P(\omega)\to\tilde{\mathbf{V}}_Q(\omega);\, f\mapsto\left(g\mapsto \int_{U(P)^*_{\mathbb{A}}} f(w_{123}ng,\omega)dn\right)$$

and one has the following standard evaluations of the constant terms:

$$\int_{U(P)^*_k\backslash U(P)^*_{\mathbb{A}}} E_P(nf,\omega)dn=L_P(f(I,\omega)),$$

$$\int_{U(Q)^*_k\backslash U(Q)^*_{\mathbb{A}}} E_P(nf,\omega)dn=L_Q((I_{QP}f)(I,\omega)),$$

$$\int_{U(Q)^*_k\backslash U(Q)^*_{\mathbb{A}}} E_Q(nf,\omega)dn=L_Q(f(I,\omega))$$

and

$$\int_{U(P)^*_k\backslash U(P)^*_{\mathbb{A}}} E_Q(nf,\omega)dn=L_P((I_{PQ}f)(I,\omega)).$$

The study of Eisenstein series reduces to a large extent to the study of the I_{PQ} and I_{QP}. If $V\cong\hat{\otimes} V_v$ then the operators I_{PQ} and I_{QP} are determined by their local analogues. In particular this allows us to regularize the I_{PQ} and I_{QP}, as we shall now describe.

Although one can do this without restricting $\tilde{V}$ further we shall assume that $\tilde{V}$ is exceptional as well as being cuspidal. This means that there exists a Größencharakter α of $k^*_{\mathbb{A}}$ so that:

i) if $\alpha_v(-1)=1$ then $\tilde{V}_v\cong V_0(\alpha^*_v)$ with

$$\alpha^*(s(h^2))=\mu(h)\alpha_v(\det(h)) \qquad (h\in H_v)$$

where $V_0(\alpha^*)$ has the meaning of [10, I.2]. This means that $V_0(\alpha^*)$ is the irreducible quotient of a principal series representation (of $\bar{\varrho}(\alpha_v^{1/2}|\,|_v^{1/4},\ \alpha_v^{-1/2}|\,|_v^{1/4})$ in the notation of [5, Sect. 2]). The covariants of $V_0(\alpha^*)$ with respect to the lift of the upper triangular unipotent group is an irreducible $\tilde{H}_v$-module on which $s\{h^2\,|\,h\in H_v\}$ acts by $\alpha^*\mu^{-1}$. This suffices to identify $V_0(\alpha^*)$ with the r_α of [5, Sect. 2].

ii) there exist places v such that $\alpha_v(-1)=-1$; at such places $\tilde{V}_v$ is cuspidal if v is non-archimedean and square-integrale if v is archimedean. Here V_v is again r_α (see [5, Proposition 3.3.3]).

The construction of the global V is given in [5, Sect. 8].

$$W_{\Pi,v}(\omega_v)=\{f:\widetilde{GL}_3(k_v)\to\tilde{V}_v\,|\,f(\gamma g)=\omega_v(\gamma)\mu_{\Pi,v}(\gamma)\cdot(\gamma f(g)),$$
$$\gamma\in\tilde{\Pi}_v,\ g\in\widetilde{GL}_3(k_v),\text{ and } f \text{ locally constant}\},$$

where $\Pi = P$ or Q. Then we have maps

$$I_{QP}: W_{P,v}(\omega_v) \to W_{Q,v}(\omega_v)$$

defined as the regularization of

$$(I_{QP}f)(g) = \int_{U(Q)_v^*} f(w_{132}ng)dn$$

and

$$I_{PQ}: W_{Q,v}(\omega_v) \to W_{P,v}(\omega_v)$$

defined as the regularization of

$$(I_{PQ}f)(g) = \int_{U(P)_v^*} f(w_{123}ng)dn.$$

These are the local factors of the global intertwining operators defined above. If $v_v^0 \in W_{\Pi,v}(\omega_v)$ is such that

$$v_v^0(gk) = v_v^0(g) \quad \text{for} \quad k \in K_v^*$$

where this is meaningful, and normalized by

$$v_v^0(I) = v_v^0$$

where v_v^0 is the standard K_v^*-invariant vector of $\tilde{V}_v$ (see [10, I.2]). This defines the family of vectors with respect to which the tensor product of the $W_{\Pi,v}(\omega_v)$ can be taken. One has then at an unramified place that ([10, Proposition I.2.4])

$$I_{QP}v_v^0 = \frac{L(\omega_v^6\alpha_v^3\chi_v^{-2}|\,|_v^{-1/2})}{L(\omega_v^6\alpha_v^3\chi_v^{-2}|\,|_v^{3/2})}\, v_v^0$$

and

$$I_{PQ}v_v^0 = \frac{L(\omega_v^6\alpha_v^{-3}\chi_v^2|\,|_v^{-1/2})}{L(\omega_v^6\alpha_v^{-3}\chi_v^2|\,|_v^{3/2})}\, v_v^0.$$

where

$$\chi_v^2(x) = \chi_v(\mathbf{s}(xI))$$

and

$$\omega_v(x) = \omega_v\begin{pmatrix}1 & 0 & 0\\ 0 & x & 0\\ 0 & 0 & 1\end{pmatrix}, \quad \left(\text{resp. } \omega_v(x) = \omega_v\begin{pmatrix}1 & 0 & 0\\ 0 & x & 0\\ 0 & 0 & 1\end{pmatrix}^{-1}\right).$$

It follows from this and [10, Proposition I.2.3] by the usual techniques of the theory of Eisenstein series

$$L(\omega^6\alpha^3\chi^{-2}\|\,\|_{\mathbf{A}}^{3/2})E_P(f,\omega)$$

and

$$L(\omega^6\alpha^{-3}\chi^2\|\,\|_{\mathbf{A}}^{3/2})E_Q(f,\omega)$$

are holomorphic in ω. Moreover one has the functional equation

$$\begin{aligned}&L(\omega^6\alpha^{-3}\chi^2\|\,\|_{\mathbf{A}}^{3/2})E_Q(f,\omega)\\ &\quad = \varepsilon(\omega^6\alpha^{-3}\chi^2\|\,\|_{\mathbf{A}}^{-1/2})L(\omega^{-6}\alpha^3\chi^{-2}\|\,\|_{\mathbf{A}}^{3/2})E_P\left(\frac{L(\omega^6\alpha^{-3}\chi^2\|\,\|_{\mathbf{A}}^{3/2})}{L(\omega^6\alpha^{-3}\chi^2\|\,\|_{\mathbf{A}}^{-1/2})}\, I_{PQ}f, \tilde{\omega}\right)\end{aligned}$$

and an analogous one with P and Q interchanged. Here L and ε are the usual Tate functions; see [11, pp. 110, 111]. Note that if $f=\otimes f_v$, $f_v=v_v^0$ almost everywhere, then

$$\frac{L(\omega^6\alpha^{-3}\chi^2\| \|_{\mathbf{A}}^{3/2})}{L(\omega^6\alpha^{-3}\chi^2\| \|_{\mathbf{A}}^{1/2})} I_{PQ}f = \bigotimes_v \left\{\frac{L(\omega_v^6\alpha_v^{-3}\chi_v^2| |_v^{3/2})}{L(\omega_v^6\alpha_v^{-3}\chi_v^2| |_v^{-1/2})} I_{PQ}f_v\right\}$$

and almost all factors here are equal to v_v^0.

This suffices for the discussion of the case where $\tilde{V}$ is cuspidal. The case where $\tilde{V}$ is an exceptional non-cuspidal representation is similar although a little more complicated. In this case we see that

$$L(\omega^6\alpha^3\chi^{-2}\| \|_{\mathbf{A}}^{3/2})E_P(f,\omega)$$

and

$$L(\omega^6\alpha^{-3}\chi^2\| \|_{\mathbf{A}}^{3/2})E_Q(f,\omega)$$

have at most simple poles where

$$\omega^6\alpha^3\chi^{-2}=\| \|_{\mathbf{A}}^{\pm 3/2} \quad \text{resp. } \omega^6\alpha^{-3}\chi^2=\| \|_{\mathbf{A}}^{\pm 3/2}.$$

One has the same functional equation as before. These assertions follow from the general results on Eisenstein series [10, II.1]; [11, p. 278].

3. The Rankin-Selberg Integral

In this section we shall prove the central global result needed for our investigation. It is a formula of Rankin-Selberg type which we now formulate.

Let W be an irreducible cuspidal representation of $GL_3(k_{\mathbf{A}})$ and let $L_W: W\to\mathbf{C}$ be a non-trivial $GL_3(k)$-invariant linear form. Let θ be an irreducible automorphic representation of $\widetilde{GL}_3(k_{\mathbf{A}})$ with genuine central quasicharacter $\tilde{\chi}$; this is to be an exceptional representation of the type of [10, Theorem II.2.1]. Let $L_\theta:\theta\to\mathbf{C}$ be a non-trivial $GL_3(k)^*$-invariant linear form. We recall that it follows from [10, Theorem I.3.5] that if e is a non-degenerate character of $N_{\mathbf{A}}^*$ trivial on N_k^* then

$$\int_{N_k^*\backslash N_{\mathbf{A}}^*} L_\theta(nv)\bar{e}(n)dn=0, \qquad v\in\theta.$$

Let χ be the central quasicharacter of W. Let V be an irreducible exceptional automorphic representation of $\widetilde{GL}_2(k_{\mathbf{A}})$ (possibly cuspidal) and extend it to $\tilde{P}_{\mathbf{A}}$ and $\tilde{Q}_{\mathbf{A}}$ with central quasicharacter $(\chi\tilde{\chi})^{-1}$. Let $L: V\to\mathbf{C}$ be a non trivial $GL_2(k)^*$-invariant linear form. We form the corresponding series $E_P(f,\omega)$ and $E_Q(f,\omega)$ as in Sect. 2. We shall now prove:

Proposition 3.1. *With the notations above one has for ω with $\sigma(\omega)$ large enough, $w\in W$, $t\in\theta$, $f\in\mathbf{F}_P(\omega)$, $f'\in\mathbf{F}_Q(\omega)$ the integrals*

$$\int_{GL_3(k)\backslash GL_3(k_{\mathbf{A}})/Z_{\mathbf{A}}} L_W(gw)L_\theta(gt)E_P(gf,\omega)dg$$

and

$$\int_{GL_3(k)\backslash GL_3(k_{\mathbf{A}})/Z_{\mathbf{A}}} L_W(gw)L_\theta(gt)E_Q(gf',\omega)dg$$

converge absolutely. Define for a non-degenerate character e of $N_{\mathbf{A}}$ trivial on N_k the following functionals

$$\Lambda_W(\omega)=\int_{N_k\backslash N_{\mathbf{A}}} L_W(nw)\bar{e}(n)dn\,,\qquad w\in W$$

$$\Lambda_\theta^{01}(t)=\int_{U(P)_k\backslash U(P)_{\mathbf{A}}} L_\theta(nt)e(n)dn\,,\qquad t\in\theta$$

$$\Lambda_\theta^{10}(t)=\int_{U(Q)_k\backslash U(Q)_{\mathbf{A}}} L_\theta(nt)e(n)dn\,,\qquad t\in\theta$$

$$\Lambda_P(f(g,\omega))=\int_{(N\cap M(P))_k\backslash(N\cap M(P))_{\mathbf{A}}} L(nf(g,\omega))e(n)dn\,,\qquad (f\in\mathbf{F}_P(\omega))$$

$$\Lambda_Q(f(g,\omega))=\int_{(N\cap M(Q))_k\backslash(N\cap M(Q))_{\mathbf{A}}} L(nf(g,\omega))dn\,,\qquad f\in\mathbf{F}_Q(\omega)\,.$$

Then Λ_θ^{01} (resp. Λ_θ^{10}) is $U(Q)_{\mathbf{A}}$- (resp. $U(P)_{\mathbf{A}}$-) invariant. One has

$$\int_{GL_3(k)\backslash GL_3(k_{\mathbf{A}})/Z_{\mathbf{A}}} L_W(gW)L_\theta(gt)E_P(gf,\omega)dg$$
$$=\int_{N_{\mathbf{A}}Z_{\mathbf{A}}\backslash GL_3(k_{\mathbf{A}})} \Lambda_W(gw)\Lambda_\theta^{01}(gt)\Lambda_P(f(g,\omega))dg$$

and

$$\int_{GL_3(k)\backslash GL_3(k_{\mathbf{A}})/Z_{\mathbf{A}}} L_W(gw)L_\theta(gt)E_Q(gf',\omega)dg$$
$$=\int_{N_{\mathbf{A}}Z_{\mathbf{A}}\backslash GL_3(k_{\mathbf{A}})} \Lambda_W(gw)\Lambda_\theta^{10}(gt)\Lambda_Q(f'(g,\omega))dg\,.$$

Proof. We shall deal only with the integral involving P; the one involving Q can be treated analogously. We shall first consider the integrals formally leacing aside questions of convergence.

The usual Rankin-Selberg transformation shows that

$$\int_{GL_3(k)\backslash GL_3(k_{\mathbf{A}})/Z_{\mathbf{A}}} L_W(gw)L_\theta(gt)E_P(gf,\omega)dg$$

is equal to

$$\int_{P_k\backslash GL_3(k_{\mathbf{A}})/Z} L_W(gw)L_\theta(gt)L(f(g,\omega))dg\,.$$

Now we have the Fourier expansion

$$L_W(gw)=\sum_{p\in N_kZ_k\backslash P_k} \Lambda_W(pgw)$$

since W is cuspidal. Using this the integral becomes

$$\int_{N_kZ_{\mathbf{A}}\backslash GL_3(k_{\mathbf{A}})} \Lambda_W(gw)L_\theta(gt)L(f(g,\omega))dg$$
$$=\int_{N_{\mathbf{A}}Z_{\mathbf{A}}\backslash GL_3(k_{\mathbf{A}})}\int_{N_k\backslash N_{\mathbf{A}}} \Lambda_W(gw)L_\theta(ngt)L(f(ng,\omega))e(n)dndg\,.$$

The inner integral can be written as

$$\int_{(N\cap M(P))_k\backslash(N\cap M(P))_{\mathbf{A}}}\int_{U(P)_k\backslash U(P)_{\mathbf{A}}} \Lambda_W(gw)L_\theta(n_1n_2gt)L(f(n_2g,\omega))e(n_1)e(n_2)dn_1dn_2\,.$$

Since θ is exceptional the $(N\cap M(P))_k$-invariant function

$$n_2 \mapsto \int_{U(P)_k\backslash U(P)_\mathbf{A}} L_\theta(n_1 n_2 gt)e(n_1)dn$$

is constant (otherwise there would be a non-trivial non-degenerate Fourier coefficient); thus our integral becomes

$$\Lambda_W(gw) \int_{U(P)_k\backslash U(P)_\mathbf{A}} L_\theta(n_1 gt)e(n_1)dn_1 \int_{(N\cap M(P))_k\backslash(N\cap M(P))_\mathbf{A}} L(n_2 f(g,\omega))e(n_2)dn_2$$

which is

$$\Lambda_W(gw)\Lambda_\theta^{01}(gt)\Lambda_P(f(g,\omega)).$$

This yields the equality asserted in the proposition. The convergence of the latter integral follows from [7, Sect. 2] and the intermediate ones follow from this.

From Proposition 3.1 we deduce now:

Proposition 3.2. *The functions*

$$Z_P(\omega; w, t, f) = L(\omega^6\alpha^3\chi^2\tilde{\chi}^2\|\ \|_\mathbf{A}^{3/2}) \int_{GL_3(k)\backslash GL_3(k_\mathbf{A})/Z_\mathbf{A}} L_W(gw)L_\theta(gt)E_P(gf,\omega)dg$$

and

$$Z_Q(\omega; w, t, f') = L(\omega^6\alpha^{-3}\chi^{-2}\tilde{\chi}^{-2}\|\ \|_\mathbf{A}^{3/2}) \int_{GL_3(k)\backslash GL_3(k_\mathbf{A})/Z_\mathbf{A}} L_W(gw)L_\theta(gt)E_Q(gf',\omega)dg$$

where α is derived from V as in Sect. 2 have analytic continuations as meromorphic functions to Ω_P and Ω_Q respectively. If V is cuspidal then Z_P and Z_Q are holomorphic; if V is not cuspidal then Z_P has at most simple poles where

$$\omega^6\alpha^3\chi^2\tilde{\chi}^2 = \|\ \|_\mathbf{A}^{\pm 3/2};$$

likewise Z_Q has at most simple poles where

$$\omega^6\alpha^{-3}\chi^{-2}\tilde{\chi}^{-2} = \|\ \|_\mathbf{A}^{\pm 3/2}.$$

One has the functional equation:

$$Z_Q(\omega; w, t, f') = \varepsilon(\omega^6\alpha^{-3}\chi^{-2}\tilde{\chi}^{-2}\|\ \|_\mathbf{A}^{-1/2})Z_P(\omega; w, t, \tilde{I}_{PQ}f')$$

where

$$\tilde{I}_{PQ} = \frac{L(\omega^6\alpha^{-3}\chi^{-2}\tilde{\chi}^{-2}\|\ \|_\mathbf{A}^{3/2})}{L(\omega^6\alpha^{-3}\chi^{-2}\tilde{\chi}^{-2}\|\ \|_\mathbf{A}^{1/2})} I_{PQ}.$$

This follows from Proposition 3.1 and the results recalled in Sect. 2.

We shall next derive alternative expressions for Z_P and Z_Q as Euler products. For this we need some preparations.

Recall that if we represent W as $\otimes W_v$ then each W_v has a unique Whittaker model; thus we can represent $\Lambda_w(\otimes w_v)$ as $\prod_v \Lambda_{w,v}(w_v)$ where $\Lambda_{w,v}$ is a Whittaker functional of W_v and the product is over all places of k.

Consider next the function $P_\mathbf{A} \to \mathbf{C}$

$$g \mapsto \int_{U(P)_k\backslash U(P)_\mathbf{A}} L_\theta(ngt)dn.$$

This is P_k^*-invariant and is an automorphic form belonging to a automorphic representation θ of $P_{\mathbf{A}}$. As θ is given as the residue of Eisenstein series [10, Theorem II.2.1] it is immediate that θ_P is too. It follows from studying the constant term that θ_P is irreducible. By [10, Theorem II.2.5] each local factor has a unique Whittaker model and the global Whittaker functional is non-trivial as the representation is genuine. Thus if $\theta = \otimes\theta_v$ and $t = \otimes t_v$ it follows that

$$\int_{(N\cap M(P))_k\backslash(N\cap M(P))_{\mathbf{A}}} \int_{U(P)_k\backslash U(P)_{\mathbf{A}}} L_\theta(ungt)du\, e(n)dn$$

can be factorized as $\prod_v \Lambda_{\theta,v}^{1\,0}(g_v t_v)$. The functional $\Lambda_{\theta,v}^{1\,0}$ is determined up to a scalar multiple by its transformation property under N_v.

As V is exceptional we can represent V as $\otimes V_v$. One has again that

$$\int_{(N\cap M(P))_k\backslash(N\cap M(P))_{\mathbf{A}}} L(nf)e(n)dn$$

with

$$f = \otimes f_v$$

can be expressed as $\prod_v \Lambda_v(f_v)$. Note that $f \in \mathbf{F}_P(\omega)$ can, as usual, be represented as a finite sum of such primitive elements and that at almost all places the factor f_v is $v_v^0(\omega)$. One demands that $\Lambda_v(v_v^0(\omega)) = 1$ for almost all v.

One now has:

Proposition 3.3. *With the notations above*

$$Z_P(\omega;\, \otimes w_v, \otimes t_v, \otimes f_v)$$

is equal to

$$\prod_v \left(L_v(\omega^6\alpha^3\chi^2\tilde{\chi}^2 \|\ \|_{\mathbf{A}}^{3/2}) \int_{N_v Z_v\backslash GL_3(k_v)} \Lambda_{w,v}(gw_v)\Lambda_{\theta,v}^{0\,1}(gt_v)\cdot \Lambda_{P,v}(f_v(g,\omega))dg\right)$$

where $L_v(\phi)$ is the v^{th} factor of $L(\phi)$, and $\Lambda_{P,v}$ is the extension of Λ_v to $\mathbf{F}_{P,v}(\omega_v)$. Likewise

$$Z_Q(\omega;\, \otimes w_v, \otimes t_v, \otimes f_v)$$

is equal to

$$\prod_v \left(L_v(\omega^6\alpha^{-3}\chi^{-2}\tilde{\chi}^{-2} \|\ \|_{\mathbf{A}}^{1/2}) \int_{N_v Z_v\backslash GL_3(k_v)} \Lambda_{w,v}(gw_v)\cdot \Lambda_{\theta,v}^{1\,0}(gt_v)\Lambda_{Q,v}(f_v(g,\omega))dg\right)$$

where $\Lambda_{Q,v}$ is the extension of Λ_v to $\mathbf{F}_{Q,v}(\omega_v)$.

This is immediate from Proposition 3.2.

4. The Generic Case

In Sect. 3 we have shown how the functionals $Z_P(\omega; w, t, f)$ and $Z_Q(\omega; w, t, f)$ can be represented as Euler products. In this section we shall develop the local theory at a generic place. We shall consider a non-archimedean local field F with odd residual characteristic. Let $V(\omega)$ be a principal series representation of $GL_3(F)$ with unramified quasicharacter ω. Let w_0 be a K-invariant vector of $V(\omega)$ where K is the standard maximal compact subgroup of $GL_3(F)$. Let $e: N(F) \to \mathbf{C}$ be an unramified non-degenerate character. Let $\Lambda: V(\omega) \to \mathbf{C}$ be the Whittaker functional for e such that $\Lambda(w_0) = 1$.

Next let $V_0(\tilde{\omega})$ be an exceptional representation of $\widetilde{GL}_3(F)$ in the sense of [10, II.1]. Suppose that $\tilde{\omega}$ is unramified so that $V_0(\tilde{\omega})$ has a K^*-invariant t_0. Let e_P (resp. e_Q) be the character of $N(F)$ so that

$$e_\Pi | N(F) \cap M(\Pi)(F) = 1, \qquad e_\Pi | U(\Pi)(F) = e | U(\Pi)(F)$$

where $\Pi = P$ or Q. Let $\tilde{\Lambda}_P : V_0(\tilde{\omega}) \to \mathbf{C}$ be the unique linear form so that

$$\tilde{\Lambda}_P(nv) = \bar{e}_P(n) \tilde{\Lambda}_P(v) \qquad v \in V_0(\tilde{\omega}), n \in N(F)^*$$

and

$$\tilde{\Lambda}_P(t_0) = 1 .$$

That is the first condition defines a one-dimensional space of functionals follows from [10, Theorem I.3.6] since $\tilde{\Lambda}_P$ factors through the Jacquet map. That the normalization is possible then follows from [10, Theorem I.4.2]. Henceforth we shall write N for $N(F)$ etc.

Analogously to $\tilde{\Lambda}_P$ we can define Λ_Q. Next let $V_0(\Omega)$ be an exceptional representation of $GL_2(F)$ where Ω is again unramified. Let $v_0 \neq 0$ be a K^*-invariant vector in $V_0(\omega)$, where K now refers to the standard maximal compact subgroup of $GL_2(F)$. We define the character χ of the centre of $GL_3(F)$ to be $(\omega | \tilde{Z}(F))^{-1} \cdot (\tilde{\omega} | \tilde{Z}(F))^{-1}$. Then we can regard $V_0(\Omega)$ as a $\tilde{P}$ or a $\tilde{Q}$ representation with central quasicharacter χ. Define the function $f_0^P : \widetilde{GL}_3(F) \to V_0(\Omega)$ by

$$f_0^P(gug') = g f_0^P(g') \mu_P(g) \qquad g' \in \widetilde{GL}_3(F), g \in \tilde{P}, u \in U(P)^* ,$$

and

$$f_0^P(k) = v_0 .$$

We can define f_0^Q analogously.

There is a unique Whittaker functional on $V_0(\Omega)$ with respect to $e | (N \cap N(Q))$, taking the value 1 on v_0. Denote this by Λ_P (resp. Λ_Q).

Our main result is then:

Proposition 4.1. *If* $\Omega\left(\mathbf{s}\begin{pmatrix} \pi^2 & 0 \\ 0 & 1 \end{pmatrix}\right)$ *is small enough* (π *being a uniformizer of* F) *then*

$$\int_{NZ \backslash GL_3(F)} \Lambda(g w_0) \tilde{\Lambda}_P(g t_0) \Lambda_P(f_0^P(g)) dg$$

converges. We suppose the measure on $NZ \backslash GL_3(F)$ *to be right-invariant measure giving the open subset* $NZ \backslash NZK$ *measure* 1. *Let*

$$\omega_1 = \omega\begin{pmatrix} \pi & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \omega_2 = \omega\begin{pmatrix} 1 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \omega_3 = \omega\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \pi \end{pmatrix}$$

be the Satake parameters of $V(\omega)$. *Let*

$$X = \tilde{\omega}\left(\mathbf{s}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \pi^2 \end{bmatrix}\right) \cdot \Omega\left(\mathbf{s}\begin{pmatrix} \pi^2 & 0 \\ 0 & 1 \end{pmatrix}\right) |\pi| .$$

Then the integral is equal to

$$(1-(\omega_1\omega_2\omega_3)^2X^3)/(1-\omega_1^2X)(1-\omega_2^2X)(1-\omega_3^2X)(1-\omega_1\omega_2X)$$
$$\times(1-\omega_2\omega_3X)(1-\omega_3\omega_1X).$$

One can give a very similar evaluation of the corresponding integral with Q instead of P but we shall not formulate this explicitly.

Proof. The convergence of the integral will become clear during the proof. By splitting the integral into a sum over right K-cosets we see that it is equal to

$$\sum_{\eta\in H/Z(H\cap K)} \Lambda(\eta\omega_0)\tilde{\Lambda}_P(\eta t_0)\Lambda_P(f_0^P(\eta))\mu(\eta)^{-2}$$

where H is the diagonal subgroup of $GL_3(F)$. There are explicit formulae for each of the terms here, as we shall now explain; the proposition will then follow from carrying out the summation. We shall take η to be of the form

$$\begin{pmatrix} \pi^{f_1} & 0 & 0 \\ 0 & \pi^{f_2} & 0 \\ 0 & 0 & \pi^{f_3} \end{pmatrix}.$$

As the summand is zero unless $f_1 \geqq f_2 \geqq f_3$ we shall assume this henceforth. By Shintani's theorem, [18], we have

$$\Lambda(\eta\omega_0)\mu(\eta)^{-1} = \det\begin{pmatrix} \omega_1^{f_1+2} & \omega_1^{f_2+1} & \omega_1^{f_3} \\ \omega_2^{f_1+2} & \omega_2^{f_2+1} & \omega_2^{f_3} \\ \omega_3^{f_1+2} & \omega_3^{f_2} & \omega_3^{f_3} \end{pmatrix} \Big/ (\omega_1-\omega_2)(\omega_2-\omega_3)(\omega_1-\omega_3).$$

Next $\tilde{\Lambda}_P$ factors through the Jacquet map for $\tilde{Q}$ and so we see that

$$\begin{aligned} \mu(\eta)^{-1}\Lambda_P(\eta t_0) &= \tilde{\omega}\left(\mathbf{s}\begin{bmatrix} \pi^{f_3} & 0 & 0 \\ 0 & \pi^{f_2} & 0 \\ 0 & 0 & \pi^{f_1} \end{bmatrix}\right) && \text{if } f_2 \equiv f_3 \pmod 2 \\ &= 0 && \text{otherwise}. \end{aligned}$$

This follows from [10, Theorem I.4.2].

The same result shows that

$$\begin{aligned} \Lambda_P(f_0^P(\eta))\mu(\eta)^{-1} &= \chi(\mathbf{s}(\pi^{f_3}I))\cdot\Omega\left(\mathbf{s}\begin{pmatrix} \pi^{f_2-f_3} & 0 \\ 0 & \pi^{f_1-f_3} \end{pmatrix}\right) && \text{if } f_1 \equiv f_2 \pmod 2 \\ &= 0 && \text{otherwise}. \end{aligned}$$

Next observe that as ω is exceptional

$$\tilde{\omega}\left(\mathbf{s}\begin{bmatrix} \pi^{f_3} & 0 & 0 \\ 0 & \pi^{f_2} & 0 \\ 0 & 0 & \pi^{f_1} \end{bmatrix}\right) = |\pi|^{f_3+f_2/2}\tilde{\omega}\left(\mathbf{s}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \pi^2 \end{bmatrix}\right)^{(f_1+f_2+f_3)/2}$$

and as Ω is exceptional

$$\Omega\left(\mathbf{s}\begin{pmatrix}\pi^{f_2-f_3} & 0\\ 0 & \pi^{f_1-f_3}\end{pmatrix}\right)=|\pi|^{-(f_1-f_2)/2}\,\Omega\left(\mathbf{s}\begin{pmatrix}\pi^2 & 0\\ 0 & 1\end{pmatrix}\right)^{(f_1+f_2-2f_3)/2}$$

Let us take $f_2=0$ to fix the representative modulo Z.

The integral then becomes

$$\sum_{\substack{f_1\geqq f_2\geqq 0\\ f_1,f_2\equiv 0(\mathrm{mod.}\,2)}} X^{(f_1+f_2)/2}\det\begin{pmatrix}\omega_1^{f_1+2} & \omega_1^{f_2+1} & 1\\ \omega_2^{f_1+2} & \omega_2^{f_2+1} & 1\\ \omega_3^{f_1+2} & \omega_3^{f_2+1} & 1\end{pmatrix}\cdot\Delta^{-1}$$

where $\Delta=(\omega_1-\omega_2)(\omega_2-\omega_3)(\omega_1-\omega_3)$. This we can write formally as

$$\sum_{\substack{f_1\geqq f_2\geqq 0\\ f_1,f_2\equiv 0(\mathrm{mod.}\,2)}} det\begin{pmatrix}(X^{1/2}\omega_1)^{f_1}\cdot\omega_1^2 & (X^{1/2}\omega_1)^{f_2}\cdot\omega_1 & 1\\ (X^{1/2}\omega_2)^{f_1}\cdot\omega_2^2 & (X^{1/2}\omega_2)^{f_2}\cdot\omega_2 & 1\\ (X^{1/2}\omega_3)^{f_1}\cdot\omega_3^2 & (X^{1/2}\omega_3)^{f_3}\cdot\omega_3 & 1\end{pmatrix}\Delta^{-1}$$

The summation over f_1 can be carried out; it yields

$$\sum_{\substack{f_2\geqq 0\\ f_2\equiv 0(\mathrm{mod.}\,2)}} \det\begin{pmatrix}\omega_1^2(X^{1/2}\omega_1)^{f_2}/(1-X\omega_1^2) & \omega_1(X^{1/2}\omega_1)^{f_2} & 1\\ \omega_2^2(X^{1/2}\omega_2)^{f_2}/(1-X\omega_2^2) & \omega_2(X^{1/2}\omega_2)^{f_2} & 1\\ \omega_3^2(X^{1/2}\omega_3)^{f_2}/(1-X\omega_3^2) & \omega_3(X^{1/2}\omega_3)^{f_2} & 1\end{pmatrix}\Delta^{-1}.$$

On multiplying this out and evaluating the sum over f_2 we see this is equal to Δ^{-1} times the alternating sum over all permutations of ω_1, ω_2, ω_3 of

$$\omega_1^2\omega_2/(1-X\omega_1^2)(1-X^2\omega_1^2\omega_2^2).$$

It is now an exercise involving the invariant theory of the symmetric group of order 6 to simplify this sum. After a rather long calculation one finds the result quoted in the statement of the proposition.

This can now be applied to the Z_P and Z_Q. We find:

Corollary 4.2. *Let, in the notations of Sect. 3, v be a place so that*

i) *v is non-archimedean with odd residual characteristic,*

ii) *the character e is unramified at v,*

iii) *ω_v is unramified.*

then the v^{th} favtor of $Z_P(\omega;\otimes w_u,\otimes t_u,\otimes f_u)$ with w_v, t_v, f_v the standard unramified vectors is equal to

$$\prod_{1\leqq i\leqq j\leqq 3}(1-\omega_{i,v}\omega_{j,v}\omega_v(\pi_v)^2\alpha_v(\pi_v)\tilde{\omega}_{2,v}^2(\pi_v)\,|\pi_v|_v^{1/2})^{-1}$$

where ω_2^2 is the Größencharacter

$$\tilde{\omega}_2^2(x)=\tilde{\omega}\left(\mathbf{s}\begin{bmatrix}1 & 0 & 0\\ 0 & x^2 & 0\\ 0 & 0 & 1\end{bmatrix}\right)$$

and $\omega_{1,v}$, $\omega_{2,v}$, $\omega_{3,v}$ *are the Satake parameters of* W_v. *Also the* v^{th} *factor of* $Z_Q(\omega; \otimes w_u, \otimes t_u, \otimes f_u)$ *under the same assumptions is*

$$\prod_{1 \leqq i \leqq j < 3} (1 - (\omega_{i,v}\omega_{j,v})^{-1} \cdot \omega_v(\pi_v)^2 \cdot \alpha_v(\pi_v)^{-1} \tilde{\omega}_{2,v}^2(\pi_v)^{-1} |\pi_v|_v^{1/2})^{-1}$$

Proof. The formulae here are direct consequences of Proposition 4.1 when one compares the corresponding notations. The one point which one needs to verify is that

$$\Omega_v \begin{pmatrix} x^2 & 0 \\ 0 & 1 \end{pmatrix} = |\pi_v|_v^{1/2} \omega_v(x^2) \alpha_v(x).$$

In this we are comparing an exceptional representation as constructed in [10, I.2] with a Weil representation. The equation above follows on comparing the Whittaker models, i.e. [10, Theorem I.4.2] with [5, Proposition 2.3.3].

Before leaving this result we remark that

$$\tilde{\chi}^2(x) = \tilde{\omega}_2^2(x)^3 \varepsilon(-1, x)_{\mathbf{A}}$$

where $(\,,)_{\mathbf{A}}$ is the adelic Hilbert symbol of order 2 and $\varepsilon: \mu_2(k) \to \mathbf{C}^\times$ is the unique injective homomorphism.

5. The Local Functional Equation

The results we have already proved allow us to prove the local functional equation for our situation. Let F be a local field of characteristic $\neq 2$. Let W be a irreducible admissible representation of $GL_3(F)$. Let $e: N \to \mathbf{C}$ be a non-degenerate character and let $\Lambda: W \to \mathbf{C}$ be a non-trivial Whittaker functional. Let θ be an exceptional representation of $\widetilde{GL}_3(F)$ as constructed in [10, I.2]. Let $\Lambda^{01}: \theta \to \mathbf{C}$ be a linear form so that $\Lambda^{01}(nt) = \bar{e}_P(n)\Lambda^{01}(t)\,(n \in N)$ where $e_P: N \to \mathbf{C}$ is that character for which $e_P | M(P) \cap N = 1$ and $e_P | U(P) = e | U(P)$. Likewise we choose Λ^{10} so that $\Lambda^{10}(nt) = \bar{e}_Q(n)\Lambda^{10}(t)$ and e_Q is defined as e_P but with Q replacing P throughout.

Let $\tilde{V}$ be an exceptional, possibly cuspidal, representation of $GL_2(F)$. Then let χ (resp. $\tilde{\chi}$) be the central quasicharacters of W (resp. θ). Let us extend $\tilde{V}$ to $\tilde{M}(P)$ (resp. $\tilde{M}(Q)$) by requiring that $\tilde{Z}$ acts through $(\chi\tilde{\chi})^{-1}$. Let Ω_P (resp. Ω_Q) be the complex manifold of quasicharacters of $M(P)$ (resp. $M(Q)$) trivial on Z. Let for $\omega \in \Omega_P$ (resp. $\omega \in \Omega_Q$) $\mathbf{F}_P(\omega)$ (resp. $\mathbf{F}_Q(\omega)$) be the space of locally constant functions $f(\gamma u g) = \omega(\gamma)\mu_P(\gamma) f(g)$, $\gamma \in \tilde{M}(P)$, $u \in U(P)$, $g \in \widetilde{GL}_3(F)$ (resp. $f(\gamma u g) = \omega(\gamma)\mu_Q(\gamma) f(g)$, $\gamma \in \tilde{M}(Q)$, $u \in U(Q)$, $g \in \widetilde{GL}_3(F)$). Then $\mathbf{F}_P$ and $\mathbf{F}_Q$ are holomorphic vector bundles over Ω_P and Ω_Q respectively.

Let us realize $\tilde{V}$ as r_α as in [5, Sect. 1]. Then we can construct intertwining operators $I_{PQ}: F_Q(\omega) \to F_P(\omega)$ and $I_{QP}: F_P(\omega) \to F_Q(\omega)$ where ω has the meaning ascribed to it in Sect. 3. Let us also write for $\omega \in \Omega_P$ (resp. $\omega \in \Omega_Q$)

$$\omega(x) = \omega \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \left(\text{resp. } \omega(x) = \omega \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \right) \right).$$

Then $(I_{QP}f)(g)$ is the regularized value of

$$\frac{L(\omega^6\alpha^{-3}(\chi\tilde{\chi})^{-2}|\,|_F^{3/2})}{L(\omega^6\alpha^{-3}(\chi\tilde{\chi})^{-2}|\,|_F^{-1/2})}\int_{U(Q)^*} f(w_{132}ng)dn.$$

Let $\tilde{\Lambda}_P: \tilde{V}\to\mathbf{C}$ be a non-trivial Whittaker functional for $\tilde{V}$ considered as a $\tilde{M}(P)$-representation with respect to $\bar{e}|M(P)$; we choose $\tilde{\Lambda}_Q$ analogously. Let for $w\in W$, $t\in\theta$, $f\in\Gamma_m(\mathbf{F}_P)$, $f'\in\Gamma_m(\mathbf{F}_Q)$

$$Z_P(\omega;\, w, t, f) = L(\omega^6\alpha^3\chi^2\tilde{\chi}^2|\,|_F^{3/2})\int_{NZ\backslash GL_3(F)}\Lambda(gw)\Lambda^{01}(gt)\tilde{\Lambda}_P(f(g,\omega))dg$$

and

$$Z_Q(\omega;\, w, t, f') = L(\omega^6\alpha^{-3}\chi^{-2}\tilde{\chi}^{-2}|\,|_F^{3/2})\int_{NZ\backslash GL_3(F)}\Lambda(gw)\Lambda^{10}(gt)\tilde{\Lambda}_Q(f'(g,\omega))dg.$$

Here if F is archimedean $\Gamma_m(\mathbf{F}_P)$ (resp. $\Gamma_m(\mathbf{F}_Q)$) denotes the space of meromorphic sections of $\mathbf{F}_P$ (resp. $\mathbf{F}_Q$). If F is non-archimedean then $\Gamma_m(\mathbf{F}_P)$ (resp. $\Gamma_m(\mathbf{F}_Q)$) denotes the space of rational sections of $\mathbf{F}_P$ (resp. $\mathbf{F}_Q$), where "rational" has the usual meaning in this context. Likewise we let $\Gamma(\mathcal{O},\Omega_P))$ (resp. $\Gamma(\mathcal{O},\Omega_Q)$) be the ring of holomorphic functions (if F is nonarchimedean) on Ω_P (resp. Ω_Q). Let $\Gamma_m(\mathcal{O},\Omega_P)$ and $\Gamma_m(\mathcal{O},\Omega_Q)$ denote the corresponding rings of meromorphic or "rational" functions.

Let us call an admissible irreducible representation W of $GL_3(F)$ *relevant* if there exists a global field k, a place v of k with $k_v\cong F$ and an irreducible cuspidal representation π of $GL_3(k_{\mathbf{A}})$ with $\pi_v\cong W$. Only relevant representations play a role in any global applications which may arise. Note that a square-integrable representation is relevant by [2].

Theorem 5.1. *Let F be a local field and let W be a relevant representation of $GL_3(F)$. Let θ, $\tilde{V}$ be as above. Then one has:*

i) *The integrals defining $Z_P(\omega;\, w,t,f)$ and $Z_Q(\omega;\, w,t,f')$ converge if $\sigma(\omega)$ is large enough and represent elements of $\Gamma_m(\mathcal{O},\Omega_P)$ and $\Gamma_m(\mathcal{O},\Omega_Q)$ respectively.*

ii) *There exists $\gamma(W,\theta,\tilde{V})\in\Gamma_m(\mathcal{O},\Omega_P)$ so that*

$$Z_P(\omega;\, w,t,f) = \gamma(\omega;\, W,\theta,\tilde{V})Z_Q(\tilde{\omega};\, w,t,I_{QP}f).$$

iii) *There exists $\phi\in\Gamma(\mathcal{O},\Omega_P)^\times$ so that*

$$\tilde{I}_{PQ}\tilde{I}_{QP} = \phi\cdot Id.$$

Remarks. 1. Note that this does not allow us to compute the $\gamma(\omega;\, W,\theta,V)$; the theorem merely asserts their existence. The proof of the theorem will suggest the correct form of the $\gamma(\omega;\, W,\theta,V)$. Indeed it may be possible to actually prove this using an asymptotic analysis of Z_P and Z_Q (cf. [8, Sect. 5]) as ω becomes highly ramified. This we do not undertake here.

2. It is unfortunate that we have no local proof of Theorem 5.1(ii) even in the case of W a principal series representation. It would also be very desirable to have independent proofs in the archimedean case.

3. The results of Sect. 4 allow us to find $\gamma(W,\theta,V)$ and L_P, L_Q in the "generic" case.

Proof. (i). That the integrals converge if $\sigma(\omega)$ is large enough follows from the analysis of Sect. 2. That they represent "rational" functions when F is non-archimedean follows by a standard argument – cf. [8, Sect. 1]. An archimedean version of this, as in [8, Sect. 9] shows that the Z_P and Z_Q have analytic continuations as meromorphic functions. We shall not give these arguments in detail as they are now standard and fairly lengthy.

ii) Let us define for an admissible representation π of $GL_3(F)$ and a quasicharacter α of $F^\times$

$$L(\alpha, \pi, \mathrm{Sym}^2) = L(0, (\pi \otimes \alpha) \times \pi)/L(0, \check{\pi} \otimes \chi\alpha)$$

where χ is the central quasicharacter of π. Here $L(s, \pi_1 \times \pi_2)$ is as in [7, 9] and $L(s, \pi)$ is as in [8]. Likewise we define

$$\varepsilon(\alpha, \pi, \mathrm{Sym}^2, e) = \varepsilon(0, (\pi \otimes \alpha) \times \pi, e)/\varepsilon(0, \check{\pi} \otimes \chi\alpha, e)$$

where e is an additive character of F.

Note that if π is an automorphic representation of $GL_3(k_{\mathbb{A}})$ and if α is a Größencharakter then

$$L(\alpha, \pi, \mathrm{Sym}^2) = \prod L(\alpha_v, \pi_v, \mathrm{Sym}^2)$$

and

$$\varepsilon(\alpha, \pi, \mathrm{Sym}^2, e) = \prod \varepsilon(\alpha_v, \pi_v, \mathrm{Sym}^2, e_v)$$

exist, the first if $\sigma(\alpha)$ is large enough. The function $L(\alpha, \pi, \mathrm{Sym}^2)$ has an analytic continuation to the space of all Größencharaktere as a meromorphic function and satisfies the functional equation

$$L(\alpha, \pi, \mathrm{Sym}^2) = \varepsilon(\alpha, \pi, \mathrm{Sym}^2, e) L(\| \ \| \alpha^{-1}, \pi^v, \mathrm{Sym}^2)$$

– [8, 13.6], [9, 2.7]. Moreover in the product defining ε all but finitely many factors are 1 and $\varepsilon(\alpha, \pi, \mathrm{Sym}^2, e)$ is a monomial function.

Next we note that in the situation of Propositions 3.2 and 3.3 we have that if v is a place so that the conditions of Corollary 4.2 can be satisfied then

$$Z_P(\omega_v; w_v, t_v, f_v) = L(\omega_v^2, \alpha_v, \omega_{2,v}^2 | \ |_v^{1/2}, W_v, \mathrm{Sym}^2)$$

and

$$Z_Q(\omega_v; w_v, t_v, f_v') = L(\omega_v^2 \alpha_v^{-1} \omega_{2,v}^{-2} | \ |_v^{1/2}, \check{W}_v, \mathrm{Sym}^2)$$

where w_v, t_v, f_v, f_v' take their standard values.

Let S be a finite set of places of k so chosen that the conditions of Corollary 4.2 may be satisfied outside S. We choose then for $v \notin S$ the standard vectors as arguments. Thus we see that

$$\begin{aligned} &Z_P(\omega; \otimes w_v, \otimes t_v, \otimes f_v)/L(\omega^2 \alpha \tilde{\omega}_2^2 \| \ \|_{\mathbb{A}}^{1/2}, W, \mathrm{Sym}^2) \\ &= \prod_{v \in S} Z_P(\omega_v; w_v, t_v, f_v)/L(\omega_v^2 \alpha_v \tilde{\omega}_{2,v}^2 | \ |_v^{1/2}, \check{W}_v, \mathrm{Sym}^2). \end{aligned}$$

The right-hand sides of these expressions are finite products and therefore are convergent everywhere that the factors are finite. The left-hand sides of these are related by the functional equation. We thus obtain the following:

$$\prod_{v\in S}\frac{Z_P(\omega_v;\,w_v,t_v,f_v)}{L(\omega_v^2\alpha_v\tilde\omega_{2,v}^2|\,|_v^{1/2},W,\mathrm{Sym}^2)}$$
$$=\frac{\varepsilon(\omega^6\alpha^3\chi^2\tilde\chi^2\|\,\|_{\mathbf{A}}^{-1/2},e)}{\varepsilon(\omega^2\alpha\tilde\omega_3^2\|\,\|_{\mathbf{A}}^{1/2},W,\mathrm{Sym}^2,e)}\prod_{v\in S}\frac{Z_Q(\tilde\omega_v;\,w_v,t_v,\tilde I_{QP}f)}{L(\omega_v^{-2}\alpha_v^{-1}\tilde\omega_{2,v}^{-2}|\,|_v^{1/2},\check W_v,\mathrm{Sym}^2)}.$$

Consequently the trilinear functionals

$$Z_P(\omega_v;\,*,*,*)\quad\text{and}\quad Z_Q(\omega_v;\,*,*,I_{QP}*)$$

are proportional. We note here that for no W with a Whittaker model can one have that $Z_P(*,*,*,*)$ is identically zero by a density argument. This then shows that a function γ as asserted in (ii) exists. The argument used here also suggests strongly that $\gamma(\omega;\,W,\theta,\tilde V)$ is of the form

$$(\text{monomial})\cdot L(\omega_v^2\alpha_v\tilde\omega_{2,v}^2|\,|_v^{1/2},\mathrm{Sym}^2,W_v)/L(\omega_v^{-2}\alpha_v^{-1}\tilde\omega_{2,v}^{-2}|\,|_v^{1/2},\mathrm{Sym}^2,\check W_v).$$

Note that it is immediate that if F is non-archimedean then $\gamma(W,\theta,\tilde V)$ is "rational".

This reasoning is valid as long as W occurs in some cuspidal automorphic representation, that is, if W is relevant.

We shall now prove (iii). In the case that $\tilde V$ is a non-cuspidal exceptional representation this follows from [10, I.2] and the realization of $V_0(\omega)$ as a subrepresentation of $V({}^{w_0}\omega)$, [10, Theorem I.2.9]. We let k be a global field and let α be a Größencharakter. Let S be the set of places where $\alpha_v(-1)=-1$. Let $V^0(\alpha_v)$ be the representation denoted by r_{α_v} in [5]. Then we consider instead of the local field F the ring $k_S=\prod_{v\in S}k_v$. The same constructions can be made in this case. From the global theory (Sect. 2) and the results recalled above for the places outside S we see that the analogue of (iii) holds for

$$\tilde V=\bigotimes_{v\in S}V^0(\alpha_v)\quad(\text{tensor product over }\mathbf{C}[\mu_2(k)]).$$

Next note that if α_1 and α_2 satisfy $\alpha_1(-1)=-1$, $\alpha_2(-1)=-1$ then there exists β so that $\alpha_1=\alpha_2\beta^2$ and $V^0(\alpha_1)=V^0(\alpha_2)\otimes(\beta\circ\det)$ in the case of a local field. Thus we have only to prove (iii) over F for one quasicharacter α with $\alpha(-1)=-1$. The validity of (iii) is thus a property of F.

Consider first the case $k=\mathbf{Q}$ and p,q two rational primes. Then $\alpha(x)=(pq,x)_{2,\mathbf{A}}$ is as above with $S=\{p,q\}$ if $p,q>0$, $p,q\equiv 1\pmod 4$. Note that $\tilde I_{PQ,v}\tilde I_{QP,v}=\phi_v\mathrm{Id}$ where ϕ_v is "rational in q_v^{-s}". We have just seen that $\prod_{v\in S}\phi_v$ is monomial. As $p\neq q$ it follows in this case that ϕ_p and ϕ_q are monomial. This proves (iii) for $\mathbf{Q}_p$, $p\equiv 1\pmod 4$. Next let $\alpha(x)=(-p,x)_{2,\mathbf{A}}$, $p\equiv 1\pmod 4$. In this case $S=\{\infty,p\}$ and we deduce the validity of (iii) for $F=\mathbf{R}$. Starting from $k=\mathbf{Q}(\sqrt{-1})$ we deduce (iii) also for $F=\mathbf{C}$. To prove it in general using the same method we have only to show that given F there exist a global field k, a place v of k with $k_v\cong F$, a Größencharakter α of $K_{\mathbf{A}}^\times$ so that $\alpha_v(-1)=-1$ with the property that if $S=\{w\,|\,\alpha_w(-1)=-1\}$ and $w\in S$, $w\neq v$ then the residual characteristic of w is different to that of v. This follows as we have shown that

$$\prod_{\substack{w \in S \\ w \text{ non-arch.}}} \phi_w \quad \text{is monomial.}$$

To prove this fact let T be the set of places consisting of the archimedean places and those with the same residual characteristic as v (but $v \notin T$). Let $S = \{v\} \cup T$. Then let $U(S)$ be the group of S-units in $k_S^\times$. Let $x \in k_S^\times$ be that element so that $x_v = -1$, $x_w = 1$ ($w \in T$). Then we form the character β of $U(S) \cup U(S)x$ by demanding that $\beta | U(S) = 1$, $\beta(x) = -1$. This can be extended to a character β of $k_S^\times$ and hence to a Größencharakter β of $k_{\mathbb{A}}^\times$ unramified outside S. This β will serve as the sought for α.

In the case of F of finite characteristic we have to modify this slightly. Let v be a place of a global field k and let ϕ_v be as above. Find $m_0 > 1$ so that $\phi_v(\omega \,|\, {}^s_v)$ is not a rational function of q_v^{-ms} for $m \geqq m_0$. Let w be a place of k with a residue field of q_v^m elements, $m \geqq m_0$. Then, as above, we can find a Größencharakter α of $k_{\mathbb{A}}^\times$ unramified outside $\{v, w\}$ so that $\alpha_v(-1) = -1$, $\alpha_w(-1) = -1$. Hence $\phi_v \cdot \phi_w$ is monomial. As ϕ_w is "rational in q_v^{-ms}" it follows that both ϕ_v and ϕ_w are themselves monomial. This completes the proof of (iii).

As we have remarked above we shall only need the assertion of the theorem when W is relevant. We shall now extend this to cover the class of principal series representations.

Corollary 5.2. *Assertion* (ii) *of Theorem 5.1 is valid if W is a principal series representation.*

Proof. For a representation ψ of H let $W(\psi)$ be the corresponding principal series representation. The W form a holomorphic vector bundle over the set of all possible ψ with $\psi | Z = \chi$. Let T be the set of all ψ for which the assertion is false. If $\psi_1 \in T$ then there would exist w_1, $w_2 \in W(\psi_1)$ so that

$$\frac{Z_P(\omega; w_1, t, f)}{Z_Q(\omega; w_1, t, I_{QP} f)} \neq \frac{Z_P(\omega; w_2, t, f)}{Z_Q(\omega; w_2, t, I_{QP} f)}.$$

The set of ψ can be considered as the set of functions on an irreducible algebraic variety. The subset of relevant ψ is Zariski dense – see [1, pp. 69–72]. On the other hand the remark above shows that T is an open subset and by Theorem 5.1(ii) it contains no relevant ψ. It now follows that T is empty, and this is the assertion of the theorem.

We are indebted to the referee for pointing out this proof which yields a sharper statement than our original discussion.

Theorem 5.1 asserts the existence of the trilinear functionals Z_P and Z_Q but says nothing about the possible poles or zeros of these functions. In order to be able to derive information about $L(\omega, \pi, \mathrm{Sym}^2)$ from Proposition 3.2 we need a "non-vanishing" about Z_P. Proposition 5.3 below gives such a statement which suffices for out purposes although it is rather unsatisfactory from an aesthetic point of view.

Proposition 5.3. *Let F be a non-archimedean field of characteristic $\neq 2$; we retain the notations of Theorem 5.1. Let $\Omega_P^* \subset \Omega_P$ be a connected component of Ω_P. Then there exist $w \in W$, $t \in \theta$, $\tilde{v} \in \tilde{V}$ and a compact open subgroup K_1 of K^* so that if $w \in \Omega_P^*$*

we define $f(\omega)\in F_P(\omega)$ *by*

(i) $f(I,\omega)=\tilde{v}$
(ii) f *is* K_1*-invariant*
(iii) f *is supported on* $\tilde{P}K_1$

then

$$Z_P(\omega;\, w, t, f)=L(\omega^6\alpha^3\chi^2\tilde{\chi}^2|\,|_F^{3/2}).$$

Proof. We have to show that

$$\int \Lambda(gw)\Lambda^{01}(gt)\Lambda_P(f(g,\omega))dg=1.$$

The proof of this is now fairly standard, see, e.g. [8, Sect. 4], so we sketch only the main points. Since W has a Whittaker model it follows from the Gelfand-Kazhdan theorem that if $\phi: P(F)\to \mathbf{C}$, $\phi(ng)=e(n)\phi(g)$, $n\in N(F)$ and $\phi(zg)=\chi(z)\phi(g)$ $z\in Z(F)$, and if ϕ has compact support modulo $N(F)/Z(F)$ then there exists $w\in W$ so that

$$\Lambda(gw)=\phi(w), \qquad g\in P(F).$$

We choose a sufficiently small compact open subgroup K_2 so that $e|N(F)\cap K_2=1$, χ, $\tilde{\chi}|Z(F)\cap K_2=1$ and let ϕ be that function supported on $N(F)Z(F)K_2$ satisfying the conditions above and $\phi|K_2=1$. We then find $t\in\theta$ so that $\Lambda^{01}(t)\neq 0$, and $\tilde{v}\in\tilde{V}$ so that $\Lambda_P(\tilde{v})\neq 0$. We choose K_1 so that $K_1\subset K$ and w, t are K_1-invariant and $f(\omega)$ as in the statement of the proposition is well-defined. It is then obvious that

$$\int \Lambda(gw)\Lambda^{01}(gt)\Lambda_P(f(g,\omega))dg$$

is equal to $\Lambda^{01}(t)\cdot\Lambda_P(f(I,\omega))\cdot\mathrm{meas}(K_1)$. Since $\Lambda_P(f(I,\omega))$ does not depend on ω the assertion follows directly. This proves the proposition.

6. Global Results

We can now apply the results of the previous section to the study of the global L-function. For the convenience of the reader we recall the definitions. Let k be a global field and let π be an irreducible cuspidal automorphic representation of $GL_3(k_{\mathbf{A}})$. Let χ be the central quasicharacter of π. Let α be a Größencharakter of k, i.e. a quasicharacter of $k_{\mathbf{A}}^\times$ trivial on $k^\times$. Let

$$L(\alpha,\pi,\mathrm{Sym}^2)=L(\mathcal{O},(\pi\otimes\alpha)\times\pi)/L(\mathcal{O},\check{\pi}\otimes\chi\alpha).$$

This is an Euler product of analogously defined local factors; it converges if $\sigma(\alpha)$ is large enough. Likewise we define the monomial function

$$\varepsilon(\alpha,\pi,\mathrm{Sym}^2)=\varepsilon(\mathcal{O},(\pi\otimes\alpha)\times\pi)/\varepsilon(\mathcal{O},\check{\pi}\otimes(\chi\alpha))$$

which, given a choice of additive character, can be expressed as a product of monomials over the set of places of k. All but finitely many are equal to 1. Then, as we have already pointed out it is known that as a function of $\alpha L(\alpha,\pi,\mathrm{Sym}^2)$ has an analytic continuation as a meromorphic function to the complex manifold of all Größencharaktere. Moreover one has the functional equation

$$L(\alpha,\pi,\mathrm{Sym}^2)=\varepsilon(\alpha,\pi,\mathrm{Sym}^2)L(\|\;\|_{\mathbf{A}}\alpha^{-1},\check{\pi},\mathrm{Sym}^2).$$

We now have:

Theorem 6.1. *Suppose k is of characteristic* >2. *Then* $L(\alpha, \pi, \mathrm{Sym}^2)$ *has a pole at* α_0 *only if*

$$\pi \otimes \alpha_0 \cong \check{\pi} \quad \text{or} \quad \pi \otimes (\alpha_0 \| \, \|_{\mathbf{A}}^{-1}) \cong \check{\pi} .$$

The corresponding pole is simple.

Remarks. 1. α_0 satisfies $\alpha_0^3 = \chi^{-2}$ or $\alpha_0^3 = \chi^{-2} \| \, \|_{\mathbf{A}}^3$. If one such α_0 exists satisfying $\alpha_0^3 = \chi^{-2}$ then $\pi \otimes (\alpha_0 \chi)$ has trivial central character.

2. The restriction that the characteristic be $\neq 2$ is natural as the symmetric square behaves quite differently in this characteristic.

3. That the case when the characteristic is 0 is not covered is a consequence that no analogue of Proposition 5.3 is available in the archimedean cases. In fact, as we shall see it would suffice to be able to show that when $W, \theta, \tilde{V}$ are all unitary then there is no ω_1 satisfying $\sigma(\omega_1) \geqq \frac{1}{2}$ for which

$$Z_P(\omega_1; w, t, f) = 0$$

for all w, t, f. Although this is very probably true we do not have a proof.

Proof. By renormalizing the representations $W, \theta, \tilde{V}$ we may assume that they are unitary. Here W will be a representation which realizes the class π.

The fact that $L(0, (\pi \otimes \alpha) \times \pi)$ has a pole at the stated points follows from [9, p. 368]. Since π is cuspidal $L(0, \pi \otimes \chi\alpha)$ is finite and non-zero for such α. Hence these poles exist, and there are no further poles of $L(\alpha, \pi, \mathrm{Sym}^2)$ when $\sigma(\alpha) \geqq 1$ or $\sigma(\alpha) \leqq 0$. Thus it suffices to show that there are no poles α satisfying

$$\tfrac{1}{2} \leqq \sigma(\alpha) \leqq 1 .$$

If there are none then there are none in the region $0 < \sigma(\alpha) \leqq \frac{1}{2}$ by the functional equation and so we would have proved our assertion.

We have next, with the notations above for

$$\omega = \otimes \omega_v , \qquad t = \otimes t_v , \qquad f = \otimes f_v$$

$$Z_P(\omega; w, t, f) = L(\omega^2 \alpha \tilde{\omega}_2^2 \| \, \|_{\mathbf{A}}^{1/2}, \pi, \mathrm{Sym}^2)$$
$$\times \prod_v Z_{P,v}(\omega_v; w_v, t_v, f_v) / L(\omega_v^2 \alpha_v \tilde{\omega}_{2,v}^2 | \, |_v^{1/2}, \pi_v, \mathrm{Sym}^2) .$$

As we have already seen the second factor on the right-hand side is such that almost all factors are 1. The left-hand side is holomorphic if $0 \leqq \sigma(\omega) < \frac{1}{4}$ by Proposition 3.2. It suffices therefore to verify that each factor

$$Z_{P,v}(\omega_v; w_v, t_v, f_v) / L(\omega_v^2 \alpha_v \tilde{\omega}_{2,v}^2 | \, |_v^{1/2}, \pi_v, \mathrm{Sym}^2) ,$$

is, for a suitable choice of w_v, t_v, f_v, non-zero for any such ω.

That this is so for $Z_{P,v}(\omega_v; w_v, t_v f_v)$ follows immediately from Proposition 5.3. Write $\beta_v = \omega_v^2 \alpha_v \omega_{v,v}^2 | \, |_v^{1/2}$. Then the factor

$$1/L(\beta_v, \pi_v, \mathrm{Sym}^2) = L(\mathcal{O}, \pi_v \otimes \chi_v \beta_v) / L(\mathcal{O}, (\pi_v \otimes \beta_v) \times \pi_v) .$$

Thus it remains to show that if $\sigma(\beta_v)$ satisfies $\frac{1}{2} \leqq \sigma(\beta) < 1$ then $L(0, (\pi_v \otimes \beta) \times \pi_v)^{-1} \neq 0$ when π_v is unitary, irreducible and generic. If π_v is tempered then this follows from [9, Proposition 8.4]. If π_v is not tempered then it is a member of the complementary series (see [8, Sects. 6.1–6.4]) and the computation of [9, Proposition 9.4] (note the misprint!) shows that $L(0, (\pi_v \otimes \beta) \times \pi_v) \neq 0$ in this case as well. The fact that this argument does not hold when $\frac{1}{2}$ is replaced by any smaller number is the reason why we use the functional equation to make this deduction.

This now completes the proof of the theorem.

References

1. Bernstein, J.-N., Deligne, P., Kazhdan, D., Vigneras, M.-F.: Représentations des groupes réductifs sur un corps local. Paris: Hermann 1984
2. Clozel, L.: On limit multiplicities of discrete series representations in spaces of automorphic forms. Invent. Math. **83**, 265–284 (1984)
3. Flicker, Y.Z.: The symmetric square. (Series of manuscripts)
4. Gelbart, S., Jacquet, H.: A relation between automorphic representations of $GL(2)$ and $GL(3)$. Ann. Sci. Ec. Norm. Super., IV. Ser. **11**, 471–542 (1978)
5. Gelbart, S., Piatetski-Shapiro, I.I.: Distinguished representations and modular forms of half-integral weight. Invent. Math. **59**, 145–188 (1980)
6. Gelbart, S., Piatetski-Shapiro, I.I.: On Shimura's correspondence for modular forms of half-integral weight. In: Automorphic forms, representation theory and arithmetic. Bombay: Tata Institute 1979
7. Jacquet, H., Shalika, J.: On Euler products and the classification of automorphic representations. Am. J. Math. **103**, 499–558, 777–815 (1981)
8. Jacquet, H., Piatetski-Shapiro, I.I., Shalika, J.: Automorphic forms on $GL(3)$. Ann. Math. **109**, 169–257 (1979)
9. Jacquet, H., Piatetski-Shapiro, I.I., Shalika, J.: Rankin-Selberg convolutions. Am. J. Math. **105**, 367–464 (1983)
10. Kazhdan, D., Patterson, S.J.: Metaplectic forms. Publ. Math. IHES **59**, 35–142 (1984); **61**, 149 (1985)
11. Langlands, R.P.: On the functional equations satisfied by Eisenstein series. (Lecture Notes Mathematics, Vol. 544). Berlin Heidelberg New York: Springer 1976
12. Piatetski-Shapiro, I.I.: Euler subgroups. In: Lie groups and their representations (Gelfand, I.M., ed.). Bristol: Hilger 1975
13. Rankin, R.A.: Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions, I, II, III. Proc. Cambridge Phil Soc. **35**, 351–356, 357–372 (1939); **36**, 150–151 (1940)
14. Selberg, A.: Bemerkungen über eine Dirichletsche Reihe, die mit der Theorie der Modulformen nahe verbunden ist. Arch. Math. Naturw. B XL III 47–50 (1940)
15. Shahidi, F.: On the Ramanujan conjecture and finiteness of poles for certain L-functions. Ann. Math. **127**, 547–584 (1988)
16. Shimura, G.: Modular forms of half-integral weight. In: Modular forms in one variable, I. (Lecture Notes Mathematics, Vol. 320). Berlin Heidelberg New York: Springer 1973
17. Shimura, G.: On the holomorphy of certain Dirichlet series. Proc. Lond. Math. Soc. **31**, 79–98 (1975)
18. Shintani, T.: On an explicit formula for p-adic "class-I Whittaker functions". Proc. Jap. Acad., Ser. A **52**, 180–182 (1976)

Received December 21, 1987

Math. Ann. 283, 573–581 (1989)

Mathematische
Annalen

Additivity of Certain Functionals and the Construction of Invariant Integrals

Zoltán Sasvári

Sektion Mathematik der Technischen Universität,
Mommsenstrasse 13, DDR-8027 Dresden, German Democratic Republic

1. Introduction

Throughout the paper X denotes a locally compact Hausdorff space and $C_{00}(X)$ ($C_{00}^{+}(X)$) denotes the set of all complex-valued (nonnegative) continuous functions on X having compact support. The support of a function f is denoted by $\operatorname{supp}(f)$. Our main result is the following:

Theorem 1. *Let I be a mapping from $C_{00}^{+}(X)$ into $[0, \infty)$ and suppose that for all $f, g \in C_{00}^{+}(X)$ the following relations hold:*

(i) $I(f) \leqq I(g)$ *whenever* $f \leqq g$ *(I is monotone);*

(ii) $I(pf) = pI(f)$ *if* $p \in R$ *and* $p \geqq 0$ *(I is homogeneous);*

(iii) $I(f+g) \leqq I(f) + I(g)$ *(I is subadditive);*

(iv) $I(f+g) = I(f) + I(g)$ *whenever* $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$.

Then I is additive, i.e., $I(f+g) = I(f) + I(g)$ for all $f, g \in C_{00}^{+}(X)$.

We shall prove Theorem 1 in Sect. 3. This theorem enables us to give a short proof of a general result concerning the existence of invariant integrals in $C_{00}^{+}(X)$ (Theorem 2). For the proof of the existence we use Weil's method which is usually applied for construction of the Haar integral on locally compact groups [2, 3].

Theorem 2 is the analogue of a result of Banach [1, p. 239]. Banach considered an equivalence relation $\cong$ between subsets of X and showed that if $\cong$ satisfies certain conditions then there exists a regular measure μ on X such that $\mu(A) = \mu(B)$ whenever $A \cong B$. Our concern will be with the existence of an integral I on $C_{00}^{+}(X)$ for which $I(f) = I(g)$ whenever $f \cong g$, where $\cong$ is a certain equivalence relation in $C_{00}^{+}(X)$. The Haar integral on locally compact groups will be obtained as a special case.

We remark that Theorem 1 remains true in a more general setting (see Remark in Sect. 3). An application of this Remark will be given in Sect. 4.

2. Existence of Invariant Integrals

A mapping $I: C_{00}^{+}(X) \rightarrow [0, \infty)$ will be called an *integral* on $C_{00}^{+}(X)$ if I is additive, homogeneous and not identically zero. It is well known that for any such I there exists a nonnegative regular measure μ on X for which the equality $I(f)=\int\limits_X f\,d\mu$ holds for all $f\in C_{00}^+(X)$ [3].

We shall use the notation $\|f\|=\operatorname{supp}\{f(x):x\in X\}$, $f\in C_{00}^+(X)$. Let $\cong$ be an *equivalence relation* in $C_{00}^+(X)$, i.e., $f\cong f$; $f\cong g$ if and only if $g\cong f$; $f\cong g$ and $g\cong h$ imply $f\cong h$. An equivalence relation $\cong$ is called a *congruence relation* if the following conditions (2.1–2.3) are satisfied:

$$f\cong g \quad \text{implies} \quad \|f\|=\|g\|; \tag{2.1}$$

if $f, g\in C_{00}^+(X)$ and $g\neq 0$ then there exist a positive integer n, real numbers $c_i>0$, functions $g_i\in C_{00}^+(X)$ such that $g_i\cong g$ $(i=1,\dots,n)$ and $f\leqq \sum_1^n c_i g_i$; (2.2)

if $f\cong g$ and $f\leqq \sum_1^n c_i f_i$ $(c_i>0)$ then there exist $g_i\in C_{00}^+(X)$ such that $g_i\cong f_i$ and $g\leqq \sum_1^n c_i g_i$. (2.3)

Theorem 2. *Let $\cong$ be a congruence relation in $C_{00}^+(X)$ and let $\{f_\alpha:\alpha\in\Gamma\}\subset C_{00}^+(X)$ be a net of nonzero functions satisfying the following condition:*

if f, $g\in C_{00}^+(X)$ and $\operatorname{supp}(f)\cap\operatorname{supp}(g)=\emptyset$ *then there exists $\beta(f,g)\in\Gamma$ such that for any $\alpha\geqq\beta$ the relations $h\cong f_\alpha$,* $\operatorname{supp}(h)\cap\operatorname{supp}(f)\neq\emptyset$ *imply* $\operatorname{supp}(h)\cap\operatorname{supp}(g)=\emptyset$. (2.4)

Then there exists an integral I on $C_{00}^+(X)$ for which $I(f)=I(g)$ if $f\cong g$ and $I(f)>0$ whenever $f\neq 0$.

Proof. Choose a fixed nonzero function $f_0\in C_{00}^+(X)$. For every $f\in C_{00}^+(X)$ we set

$$J_\alpha(f):=\inf\left\{\sum_1^n c_i : f\leqq \sum_1^n c_i f_i,\ f_i\cong f_\alpha,\ c_i>0,\ i=1,\dots,n\right\}.$$

If $f\neq 0$, then $J_\alpha(f)>0$. Indeed, if $f\leqq \sum_1^n c_i f_i$ and $f_i\cong f_\alpha$ then by (2.1) we have $\|f\|\leqq\left(\sum_1^n c_i\right)\|f_\alpha\|$ and hence $J_\alpha(f)\geqq\dfrac{\|f\|}{\|f_\alpha\|}$. Let

$$I_\alpha(f):=\frac{J_\alpha(f)}{J_\alpha(f_0)},\qquad f\in C_{00}^+(X).$$

It is immediate that I_α is monotone, homogeneous and subadditive and that $I_\alpha(f_0)=1$. Moreover, we have:

$$\text{if}\quad f\cong g\quad\text{then}\quad I_\alpha(f)=I_\alpha(g); \tag{2.5}$$

$$I_\alpha(f+g)=I_\alpha(f)+I_\alpha(g)\quad\text{whenever}\quad \operatorname{supp}(f)\cap\operatorname{supp}(g)=\emptyset\quad\text{and}\quad \alpha\geqq\beta(f,g); \tag{2.6}$$

if $f\neq 0$ then there exist positive numbers $c(f)$, $C(f)$ for which $c(f)\leqq I_\alpha(f)\leqq C(f)$, $\alpha\in\Gamma$. (2.7)

Property (2.5) follows from (2.3) while (2.6) is a consequence of (2.4). In order to show (2.7) choose $f_i, f_j' \in C_{00}^{+}(X)$, $c_i, c_j' > 0$ $(i=1,\dots,n;\ j=1,\dots,n')$ so that $f_i \cong f_0$, $f_j' \cong f$ and

$$f_0 \leqq \sum_1^{n'} c_j' f_j', \quad f \leqq \sum_1^{n} c_i f_i .$$

Using the properties of I_α we get

$$I_\alpha(f_0) = 1 \leqq \left(\sum_1^{n'} c_j'\right) I_\alpha(f) \quad \text{and} \quad I_\alpha(f) \leqq \left(\sum_1^{n} c_i\right),$$

from which (2.7) follows.

The set of all mappings $J: C_{00}^{+}(X) \to [0, \infty)$ with $J(0)=0$ and $c(f) \leqq J(f) \leqq C(f)$ for nonzero f is compact in the topology of pointwise convergence (by Tikhonov's theorem). Consequently there exists a subnet of $\{I_\alpha\}$ converging to a mapping I. It is clear that I is monotone, subadditive, homogeneous and $I(f)=I(g)$ whenever $f \cong g$. Moreover, (2.6) implies that $I(f+g)=I(f)+I(g)$ if $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$. It follows from Theorem 1 that I is additive. This completes the proof.

Examples. Let $x_0 \in X$ and let $\{W_\alpha : \alpha \in \Gamma\}$ be a neighbourhood basis at x_0. We introduce a partial ordering $\geqq$ in Γ by setting $\alpha \geqq \beta$ if $W_\alpha \subset W_\beta$. For each $\alpha \in \Gamma$ we choose a nonzero function f_α such that $\operatorname{supp}(f_\alpha) \subset W_\alpha$. Suppose that H is a group of homeomorphisms of X onto itself satisfying the following conditions:

$$\bigcup_{h \in H} hV = X \quad \text{for every nonvoid open set} \quad V \subset X; \tag{2.8}$$

for each pair of compact disjoint sets F_1 and F_2, there exists $\beta(F_1, F_2) \in \Gamma$ such that $hW_\alpha \cap F_1 \neq \emptyset$ implies $hW_\alpha \cap F_2 = \emptyset$ for all $h \in H$ and $\alpha \geqq \beta(F_1, F_2)$. (2.9)

Setting $f \cong g$ if $f(x) = g(h(x))$ for some $h \in H$, we see that $\cong$ is a congruence relation in $C_{00}^{+}(X)$ such that the net $\{f_\alpha\}$ satisfies condition (2.4). [Relation (2.2) follows from the fact that, in view of (2.8), the set $\operatorname{supp}(f)$ can be covered by a finite number of sets of the form hV $(h \in H)$ where $V = \left\{x \in X : g(x) > \frac{\|g\|}{2}\right\}$.]

Note that (2.8) is always satisfied if H is transitive, i.e., for any $x, y \in X$ there exists $h \in H$ with $h(x) = y$.

Plainly condition (2.9) is satisfied in the following cases:

a) X is metrizable, $x_0 \in X$ is arbitrary, and H is the group of all isometric homeomorphisms of X onto itself; or more generally

b) X is a uniformly locally compact uniform space, $x_0 \in X$ is arbitrary, and H is a uniformly equicontinuous group of uniformisms of X.

Thus, as a special case, we obtain a theorem of Segal [4, p. 187] (see also the remarks of Goetz in [1, p. 352]).

There are classical cases where conditions (2.8, 2.9) are satisfied. The symbol X denotes a locally compact group and x_0 is always the identity of X. We list these examples:

c) H is the group of left translations $x \to yx$ $(x, y \in X)$;

d) X is compact or commutative and H is generated by the mappings $x \to x^{-1}$, $x \to yx$ and $x \to xy$ $(x, y \in X)$;

e) G is a compact subgroup of X and H is generated by the mappings $x \to yx$ $(x, y \in X)$ and $x \to xg$ $(x \in X,\ g \in G)$;

f) X admits a left invariant metric ϱ and H is the group of all invertible isometric mappings $h: X \to X$. If ϱ is also right (inverse) invariant then the corresponding (Haar) integral will be right (inverse, respectively) invariant as well.

3. Proof of Theorem 1

To illustrate our method, we consider the case where X is discrete. Then any function $f \in C_{00}^{+}(X)$ can be written as $f = \sum_{x \in X} f(x)\delta_x$ (finite sum) where $\delta_x(x) = 1$ and $\delta_x(y) = 0$ for $y \in X$, $y \neq x$. Using (ii) and (iv) in Theorem 1 we get

$$\begin{aligned} I(f+g) &= I\Big(\sum_{x \in X} (f(x)+g(x))\delta_x\Big) = \sum_{x \in X} (f(x)+g(x)) I(\delta_x) \\ &= \sum_{x \in X} f(x) I(\delta_x) + \sum_{x \in X} g(x) I(\delta_x) \\ &= I\Big(\sum_{x \in X} f(x)\delta_x\Big) + I\Big(\sum_{x \in X} g(x)\delta_x\Big) \\ &= I(f) + I(g)\,. \end{aligned}$$

In the general case, we will construct continuous functions which will play the role of the functions δ_x. We will write

$$C_{00}^{1}(X) = \{h \in C_{00}^{+}(X) : h(X) \subset [0,1]\}\,.$$

We write $C_{00}^{1}([0,1])$ as C^1.

The proof. Let $f \in C_{00}^{1}(X)$ and denote by B_f the set of all $t \in [0,1]$ with the following property: for every $\varepsilon > 0$, there exists $\delta > 0$ such that $I(fh) < \varepsilon$ whenever $h \in C_{00}^{1}(X)$ and $\operatorname{supp}(h) \subset f^{-1}([t-\delta, t+\delta])$.

We show first that the set $D_f := [0,1] \backslash B_f$ is denumerable, so that B_f is everywhere dense in $[0,1]$. Assume that D_f is nondenumerable. Then there exist a positive integer m, mutually distinct numbers $t_i \in D_f$ and $\varepsilon_i > \frac{1}{m}$ $(i = 1, 2, \ldots)$ such that for every $\delta > 0$ we can find functions $h_i \in C_{00}^{1}(X)$ with $\operatorname{supp}(h_i) \subset f^{-1}([t_i - \delta, t_i + \delta])$ and $I(fh_i) \geqq \varepsilon_i > \frac{1}{m}$. Let N be a positive integer with $N > mI(f)$ and let $\delta = \delta(N)$ be a positive number for which

$$[t_i - \delta, t_i + \delta] \cap [t_j - \delta, t_j + \delta] = \emptyset \qquad (i \neq j;\ i, j = 1, \ldots, N)\,.$$

Choose $h_i \in C_{00}^{1}(X)$ so that $\operatorname{supp}(h_i) \subset f^{-1}([t_i - \delta,\ t_i + \delta])$ and $I(fh_i) \geqq \varepsilon_i > \frac{1}{m}$. Then $\operatorname{supp}(h_i) \cap \operatorname{supp}(h_j) = \emptyset$ $(i \neq j)$ and hence

$$I(f) \geqq I\left(\sum_{1}^{N} fh_i\right) = \sum_{1}^{N} I(fh_i) \geqq \frac{N}{m},$$

contradicting the choice of N. Thus D_f is denumerable. Now suppose that $f \in C^1_{00}(X)$ and $\|f\| < 1$. We will show that $I(f)$ can be approximated by sums of the form $\sum_1^n t_i I(h_i)$ where $h_i \in C^1_{00}(X)$ and $\operatorname{supp}(h_i) \cap \operatorname{supp}(h_j) = \emptyset$ $(i \neq j)$.

Let $\varepsilon > 0$ be arbitrary and choose $t_1, \ldots, t_n \in B_f$ so that

$$0 =: t_0 < t_1 \ldots < t_n < t_{n+1} := 1 \quad \text{and} \quad \max_{i=0,\ldots,n} (t_{i+1} - t_i) < \varepsilon .$$

Note that trivially $1 \in B_f$ since $\|f\| < 1$. As $t_i \in B_f$, there is a $\delta > 0$ for which $\delta < \frac{1}{2} \min_{i=0,\ldots,n} (t_{i+1} - t_i)$ and

$$I(f\tilde{h}_i) < \frac{\varepsilon}{n+1} \tag{3.1}$$

for every $\tilde{h}_i \in C^1_{00}(X)$ with $\operatorname{supp}(\tilde{h}_i) \subset f^{-1}([t_i - \delta,\ t_i + \delta])$, $i = 1, \ldots, n+1$. Choose $g_i \in C^1$ such that

$$g_i([t_i + \delta, t_{i+1} - \delta]) = 1 \quad \text{and} \quad g_i([0,1] \backslash [t_i, t_{i+1}]) = 0 ,$$

$i = 0, \ldots, n$, and let $\tilde{g}_i \in C^1$ be the functions which are uniquely determined by the relations

$$\operatorname{supp}(\tilde{g}_i) \subset [t_i - \delta, t_i + \delta] , \quad i = 0, \ldots, n+1 , \quad \text{and} \quad \sum_0^n g_i + \sum_0^{n+1} \tilde{g}_i = 1 .$$

We set $h_i := g_i(f)$, $i = 0, \ldots, n$; $\tilde{h}_i := \tilde{g}_i(f)$, $i = 1, \ldots, n+1$, and $\tilde{h}_0 := \chi \tilde{g}_0(f)$, where χ is a function in $C^1_{00}(X)$ for which $\chi(\operatorname{supp}(f)) = 1$. Write $\tilde{h} = \sum_0^{n+1} \tilde{h}_i$. Then we have $\tilde{h}$, h_i, $\tilde{h}_j \in C^1_{00}(X)$ $(i = 0, \ldots, n;\ j = 0, \ldots, n+1)$. These functions have the following properties:

$$\operatorname{supp}(h_i) \cap \operatorname{supp}(h_j) = \emptyset \qquad (i \neq j; i, j = 0, \ldots, n);$$

$$\operatorname{supp}(\tilde{h}_i) \subset f^{-1}([t_i - \delta, t_i + \delta]); \qquad f = f\left(\sum_1^n h_i + \tilde{h}\right);$$

$$\sum_1^n t_i h_i \leqq f < \chi ; \qquad f\tilde{h}_0 \leqq \varepsilon\chi ; \qquad \sum_0^n h_i \leqq \chi ;$$

and

$$0 \leqq f h_i - t_i h_i \leqq \varepsilon h_i \qquad (i = 0, \ldots, n) .$$

The last inequality implies that $I(f h_i) \leqq (t_i + \varepsilon) I(h_i)$, $i = 0, \ldots, n$. From this and properties of the functions h_i and $\tilde{h}_i$, we get

$$\begin{aligned}
\sum_1^n t_i I(h_i) &= I\left(\sum_1^n t_i h_i\right) \leqq I(f) = I\left(f\left(\sum_0^n h_i + \tilde{h}\right)\right) \\
&\leqq \sum_0^n I(f h_i) + I(f\tilde{h}) \leqq \sum_0^n (t_i + \varepsilon) I(h_i) + I(f\tilde{h}) \\
&= \sum_1^n t_i I(h_i) + \varepsilon I\left(\sum_0^n h_i\right) + I(f\tilde{h}) \\
&\leqq \sum_1^n t_i I(h_i) + \varepsilon I(\chi) + I(f\tilde{h}) .
\end{aligned} \tag{3.2}$$

In view of (3.1) we have

$$I(f\tilde{h}) \leqq \sum_{1}^{n+1} I(f\tilde{h}_i) + I(f\tilde{h}_0) < (n+1)\frac{\varepsilon}{n+1} + \varepsilon I(\chi) = \varepsilon(1 + I(\chi)).$$

Putting this in (3.2) we obtain

$$0 \leqq I(f) - \sum_{1}^{n} t_i I(h_i) \leqq (1 + 2I(\chi)).$$

We will now prove that I is additive. Since I is homogeneous, it suffices to show that $I(f+g) = I(f) + I(g)$ for all $f, g \in C_{00}^1(X)$ with $\|f\| < 1$ and $\|g\| < 1$.

Suppose first that $cf \leqq g \leqq Cf$, where c and C are positive numbers, and choose $\chi \in C_{00}^1(X)$ for which $\chi(\mathrm{supp}(f+g)) = 1$. Let $\varepsilon > 0$ be arbitrary. It follows from the facts just proved that there exist functions h_i, h'_j, $\tilde{h}$, $\tilde{h}' \in C_{00}^1(X)$ and numbers $t_i, t'_j \in [0,1]$ $(i = 0, \dots, n;\ j = 0, \dots, m)$ such that

$$0 \leqq I(f) - \sum_{1}^{n} t_i I(h_i) \leqq \varepsilon, \quad \sum_{0}^{n} h_i + \tilde{h} = 1 \quad \text{on} \quad \mathrm{supp}(f+g),$$

$$I(f\tilde{h}) \leqq \varepsilon, \quad \sum_{1}^{n} t_i h_i \leqq f,$$

$$\mathrm{supp}(h_i) \cap \mathrm{supp}(h_j) = \emptyset \quad (i \neq j;\ i, j = 0, \dots, n),$$

and g, h'_j, $\tilde{h}'$, and t'_j satisfy the same relations with m instead of n. We have

$$\begin{aligned} I(f) + I(g) &\leqq \sum_{1}^{n} t_i I(h_i) + \sum_{1}^{m} t'_j I(h'_j) + 2\varepsilon \\ &= \sum_{1}^{n} t_i I\left(h_i\left(\sum_{1}^{m} h'_j + \tilde{h}'\right)\right) + \sum_{1}^{m} t'_j I\left(h'_j\left(\sum_{1}^{n} h_i + \tilde{h}\right)\right) + 2\varepsilon \\ &\leqq \sum_{1}^{n} t_i I\left(h_i\left(\sum_{1}^{m} h'_j\right)\right) + \sum_{1}^{m} t'_j I\left(h'_j\left(\sum_{1}^{n} h_i\right)\right) \\ &\quad + I\left(\sum_{1}^{n} t_i h_i \tilde{h}'\right) + I\left(\sum_{1}^{m} t'_j h'_j \tilde{h}\right) + 2\varepsilon. \end{aligned}$$

The sum of the first two term on the right is equal to

$$S_1 := I\left(\sum_{i=1}^{n} \sum_{j=1}^{m} (t_i + t'_j) h_i h'_j\right).$$

Since

$$\sum_{i=1}^{n} \sum_{j=1}^{m} (t_i + t'_j) h_i h'_j \leqq f + g,$$

we get $S_1 \leqq I(f+g)$. Let S_2 denote the sum of the third and fourth terms. From the inequalities $\sum_{1}^{n} t_i h_i \leqq f$ and $\sum_{1}^{m} t'_j h'_j \leqq g$, we obtain

$$S_2 \leqq I(f\tilde{h}') + I(g\tilde{h}) \leqq I\left(\frac{g}{c}\tilde{h}'\right) + I(Cf\tilde{h}) \leqq \left(\frac{1}{c} + C\right)\varepsilon,$$

and so

$$I(f)+I(g)\leqq I(f+g)+\left(2+\frac{1}{c}+C\right)\varepsilon.$$

This being true for all positive ε, it follows that $I(f+g)=I(f)+I(g)$.

Let now f, $g\in C_{00}^{+}(X)$ be arbitrary. For $\varepsilon>0$, we set $F:=f+\varepsilon(f+g)$ and $G:=g+\varepsilon(f+g)$. There exist positive numbers c and C such that $cF\leqq G\leqq CF$ and hence the relaiton

$$\begin{aligned}I(f+g)+2\varepsilon I(f+g)&=I(f+g+2\varepsilon(f+g))=I(F+G)\\&=I(F)+I(G)\geqq I(f)+I(g)\end{aligned}$$

holds for every $\varepsilon>0$. That is, we have $I(f+g)=I(f)+I(g)$. The proof is complete.

Remark. Theorem 1 is true for an arbitrary topological space X and for an arbitrary family $\mathscr{F}$ of bounded nonnegative continuous functions on X having the following properties:

(ii) fg, $af+bg\in\mathscr{F}$ whenever f, $g\in\mathscr{F}$ and a, $b\geqq 0$;

(ii) if $f\in\mathscr{F}$, $\|f\|\leqq 1$ and $h\in C^1$ then there exists $\chi\in\mathscr{F}$ such that $\chi\geqq 1$ on $\operatorname{supp}(f)$ and $\chi h(f)\in\mathscr{F}$.

Actually, it is unnecessary to require (ii) for every $h\in C^1$. It suffices for example to require (ii) for infinitely differentiable $h\in C^1$. Consequently Theorem 1 holds for the family of compactly supported nonnegative infinitely differentiable functions on R^n.

4. Additivity of the Upper Integral

Let I be an integral on $C_{00}^{+}(X)$ and denote by $\mathscr{M}^{+}$ the set of upper semicontinuous functions $f:X\to[0,\infty]$. The characteristic function of a set $A\subset X$ will be denoted by χ_A. Define the nonnegative functional $\bar{I}$ on $\mathscr{M}^{+}$ by

$$\bar{I}(f):=\sup\{I(g):g\in C_{00}^{+}(X),\ g\in f\}.$$

The functional $\bar{I}$ is monotone, homogeneous and additive on $\mathscr{M}^{+}$ [3, (11.12) and (11.14)]. Let now $h:X\to[0,\infty]$ be arbitrary and define $\bar{\bar{I}}(h)$ be setting

$$\bar{\bar{I}}(h):=\inf\{\bar{I}(g):g\in\mathscr{M}^{+},g\geqq h\}.$$

This functional is not additive in general but it is monotone, homogeneous and subadditive [3, (11.17)]. Moreover, $\bar{\bar{I}}\left(\lim\limits_n h_n\right)=\lim\limits_n\bar{\bar{I}}(h_n)$ whenever $\{h_n\}$ is an increasing sequence of nonnegative functions [3, (11.18)]. The set function $\iota(A):=\bar{\bar{I}}(\chi_A)$, $A\subset X$, is an outer measure, which is regular on the σ-algebra of ι-measurable sets [3, (11.34)].

Let $\mathscr{F}_m$ denote the set of ι-measurable functions $f:X\to[0,\infty]$. An important property of $\mathscr{F}_m$ is that $\bar{\bar{I}}$ is additive on $\mathscr{F}_m$. This can be proved for example by showing first additivity for step functions and then approximating measurable functions by step functions. To illustrate the Remark in Sect. 3, we give another proof.

Theorem 3. *$\bar{\bar{I}}$ is additive on $\mathscr{F}_m$.*

Proof. Let $\mathscr{F}$ denote the set of all bounded functions $f \in \mathscr{F}_m$ with $\bar{\bar{I}}(f) < \infty$ and $\imath(\operatorname{supp}(f)) < \infty$. It is easy to see that $\mathscr{F}$ satisfies conditions (i) and (ii) of the Remark if X is replaced by X_d, the discrete version of X. Consequently, the additivity of $\bar{\bar{I}}$ on $\mathscr{F}$ will follow from

$$\bar{\bar{I}}(f_1+f_2)=\bar{\bar{I}}(f_1)+\bar{\bar{I}}(f_2), \quad \text{whenever} \quad A_1 \cap A_2 = \emptyset, \tag{4.1}$$

where $A_i = \{x \in X : f_i(x) > 0\}$, $f_i \in \mathscr{F}$ $(i=1,2)$.

To prove (4.1), let $\varepsilon > 0$ be arbitrary and choose compact sets $F_i \subset A_i$ such that $\|f_i\| \imath(A_i \setminus F_i) < \varepsilon$ $(i=1,2)$. Then we have

$$\bar{\bar{I}}(f_i) = \bar{\bar{I}}((\chi_{F_i} + \chi_{(A_i \setminus F_i)}) f_i) \leqq \bar{\bar{I}}(\chi_{F_i} f_i) + \varepsilon .$$

Since $F_1 \cap F_2 = \emptyset$, there exist disjoint open sets U_1 and U_2, for which $U_1 \supset F_1$ and $U_2 \supset F_2$. Choose a function $h \in \mathscr{M}^+$ such that $h \geqq f_1 + f_2$ and $\bar{\bar{I}}(f_1+f_2) > \bar{I}(h) - \varepsilon$. We then have $\chi_{U_i} h \in \mathscr{M}^+$ and $\chi_{U_i} h \geqq \chi_{F_i} f_i$. Using the properties of $\bar{I}$ and $\bar{\bar{I}}$, we get

$$\begin{aligned}\bar{\bar{I}}(f_1+f_2) &> \bar{I}(h) - \varepsilon \geqq \bar{I}((\chi_{U_1} + \chi_{U_2})h) - \varepsilon = \bar{I}(\chi_{U_1} h) + \bar{I}(\chi_{U_2} h) - \varepsilon \\ &\geqq \bar{\bar{I}}(\chi_{F_1} f_1) + \bar{\bar{I}}(\chi_{F_2} f_2) - \varepsilon > \bar{\bar{I}}(f_1) + \bar{\bar{I}}(f_2) - 3\varepsilon ,\end{aligned}$$

and hence $\bar{\bar{I}}(f_1+f_2) \geqq \bar{\bar{I}}(f_1) + \bar{\bar{I}}(f_2)$. Relation (4.1) follows now from the subadditivity of $\bar{\bar{I}}$.

Finally, consider arbitrary $f_1, f_2 \in \mathscr{F}_m$. If $\bar{\bar{I}}(f_1) = \infty$ or $I(f_2) = \infty$ then $\bar{\bar{I}}(f_1+f_2) = \bar{\bar{I}}(f_1) + \bar{\bar{I}}(f_2)$. If $\bar{\bar{I}}(f_1) < \infty$ then $\imath(\{x \in X : f_i(x) = \infty\}) = 0$ and hence we may suppose that $f_i(x) < \infty$ for all $x \in X$ $(i=1,2)$. For each nonnegative integer n, we set $A_n^{(i)} := \left\{x \in X : \frac{1}{n} \leqq f_i(x) \leqq n\right\}$ and $f_n^{(i)} := \chi_{A_n^{(i)}} f_i$ $(i=1,2)$. Then we have $f_n^{(i)}(x) \to f_i(x)$ $(x \in X)$. Using the additivity of $\bar{\bar{I}}$ on $\mathscr{F}$ we obtain

$$\begin{aligned}\bar{\bar{I}}(f_1+f_2) &= \bar{\bar{I}}\left(\lim_n (f_n^{(1)} + f_n^{(2)})\right) = \lim_n \bar{\bar{I}}(f_n^{(1)} + f_n^{(2)}) \\ &= \lim_n \bar{\bar{I}}(f_n^{(1)}) + \lim_n \bar{\bar{I}}(f_n^{(2)}) = \bar{\bar{I}}(f_1) + \bar{\bar{I}}(f_2) .\end{aligned}$$

This completes the proof.

Addendum. We would like to thank the managing editor for bringing to our attention the following results of J. R. Baxter and R. V. Chacon, related to our Theorem 1.

Let M be a metric space and denote by $C^r(M)$ the set of continuous real valued functions on M. Let $\Phi : C^r(M) \to R$ be a functional such that:

(i) $\lim_{\|f\| \to 0} \Phi(f) = 0$;
(ii) $\Phi(f+g) = \Phi(f) + \Phi(g)$; if $fg = 0$;
(iii) $\Phi(f+\alpha) = \Phi(f) + \Phi(\alpha)$ for all $f \in C^r(M)$, $\alpha \in R$.

It has been shown in [J. R. Baxter, R. V. Chacon: Functionals on continuous functions, Pac. J. Math. 51, 355–362 (1974)] that if M has dimension no greater than one, Φ must be linear. In view of Theorem 1 it is quite surprising that if $M = [0,1] \times [0,1]$ then there exist nonlinear functionals on $C^r(M)$ which are bounded, continuous, monotone, and satisfy conditions (ii) and (iii) [J. R. Baxter, R. V. Chacon: Nonlinear functionals on $C([0,1] \times [0,1])$, Pac. J. Math. 48, 347–353 (1973)].

References

1. Banach, S.: Oeuvres. I. Warszawa: Éditions Scientifiques de Pologne 1967
2. Bourbaki, N.: Éléments de mathématique VI, Première partie. Integration. Paris: Hermann 1959
3. Hewitt, E., Ross, K.A.: Abstract harmonic analysis. I. Berlin Heidelberg New York: Springer 1963
4. Segal, I.E., Kunze, R.A.: Integrals and operators. Berlin Heidelberg New York: Springer 1978

Received February 15, 1988

Math. Ann. 283, 583–630 (1989)

Mathematische Annalen

Nonlinear Elliptic Equations with Singular Boundary Conditions and Stochastic Control with State Constraints

1. The Model Problem

J. M. Lasry and P. L. Lions

Ceremade, Université Paris-Dauphine, Place de Lattre de Tassigny,
F-75775 Paris Cedex 16, France

Contents

I. Introduction

I.1. General Introduction

One of the primary goals of this paper is to study various models of stochastic control problems involving constraints on the state of the system (state constraints). And by following the dynamic programming approach this is equivalent to study some nonlinear second-order elliptic equations. Then, the state constraints lead to highly singular boundary conditions. A typical example would be: let Ω be a bounded, smooth domain in $\mathbb{R}^N$, we look for a solution $u \in C^2(\Omega)$ of

$$-\Delta u + |\nabla u|^p + \lambda u = f \quad \text{in} \quad \Omega \tag{1}$$

where $p>1$, $\lambda>0$, f is a given smooth function in Ω, and the boundary condition is given by

$$u(x) \to +\infty \quad \text{as} \quad \operatorname{dist}(x, \partial\Omega) \to 0\,, \tag{2}$$

(in fact, this boundary condition will correspond to the case $1<p\leqq 2$).

We will show in this paper how such nonlinear, boundary value problems (1)–(2) can be solved and we will thus obtain existence, uniqueness and comparison results leading also to a complete solution of the stochastic control problem we are considering. It turns out that many cases have to be investigated and the results differ somewhat from one case to the other: typical behaviours involve the cases $1<p\leqq 2$, $p>2$, f blowing up near $\partial\Omega$, λ going to 0. Let us also mention that the methods introduced below allow us to treat more general nonlinear second-order elliptic equations, like more general quasilinear elliptic equations, Hamilton-Jacobi-Bellman equations, semilinear elliptic equations, first-order Hamilton-Jacobi equations, Monge-Ampère equations: in all those examples, singular boundary conditions may be encountered (and are even of a fundamental use) and we refer to Brézis [6], Crandall and Lions [7], Urbas [30], Simon [26–28], where such boundary conditions appear. And, our methods enable us to treat these equations with boundary conditions like (2).

I.2. Description of the Stochastic Control Problems

The basic model problem we are considering is a stochastic control problem where the *state of the controlled system is a diffusion process* and a typical example is the solution of the following *stochastic differential equation*

$$dX_t = a_t dt + dB_t, \qquad X_0 = x \in \mathbb{R}^N, \tag{3}$$

where B_t is a standard Brownian motion [in some probability space $(\Omega, F, F_t, P)\ldots$] and where a_t is the control process i.e. a progressively measurable stochastic process that we may choose as we wish (taking possibly into account restrictions on the controls such that, for instance, a_t takes its values into a given set $A\ldots$). A very important, particular class of controls is given by the so-called feedback controls i.e., given a function $a(\cdot)$, one looks for a solution of

$$dX_t = a(X_t)dt + dB_t, \qquad X_0 = x. \tag{4}$$

This general class of problems occur in many contexts: however, depending on the particular examples of applications, it is possible to classify those problems in the following categories. For some problems, the state process X_t may take its value in $\mathbb{R}^N$ without any restriction while in other problems the state X_t should remain in some given region $\bar{\Omega}$. In the latter case, the model is to be complemented with boundary prescriptions in case the process reaches or crosses the boundary $\partial\Omega$. Let us immediately mention that if Ω is bounded, and a_t or $a(\cdot)$ are bounded then *for all* $t>0$ $P(X_t \in \partial\Omega)>0$. The most usual models in stochastic control theory yield the following boundary prescriptions: in the case of the so-called *exit problems*, one considers the first exit time of X_t from $\bar{\Omega}$ (or the first hitting time of X_t on $\partial\Omega$) and the process is stopped at this time. The other standard model consists in a boundary mechanism which prevents the state process from escaping from $\bar{\Omega}$; the simplest of which is the *reflecting boundary condition.*

Now, at least for deterministic problems, it is well-known that another way to enforce state constraints (i.e. $X_t \in \bar{\Omega}$) is simply to restrict our attention to controls such that X_t remains in $\bar{\Omega}$ or in Ω. In the case of (nondegenerate) stochastic models like (3) or (4), this possibility does not seem to have been explored. And this is precisely the type of problems we have in mind. In view of a remark made above, it is clear enough that in order to constrain a Brownian motion in a bounded domain Ω we need to use unbounded drifts a_t or $a(\cdot)$: in other words, we will have to choose feedbacks controls which, roughly speaking, push back the state process inside Ω when it gets near $\partial\Omega$ and with an intensity which blows up at the boundary. To be more specific, we will consider the class $\mathscr{A}$ of feedback controls i.e. of, say, continuous functions on Ω, $a(\cdot)$ such that the solution X_t of (4) stays in Ω with probability 1 for all $t \geqq 0$, (and for all initial points $x \in \Omega$).

Once, admissible control processes and thus state processes have been specified we may now describe a typical example of the optimal control problems we want to study. For each $a \in \mathscr{A}$, we will consider for example the *following cost function*

$$J(x,a) = E \int_0^\infty \left\{ f(X_t) + \frac{1}{q} |a_t|^q \right\} e^{-\lambda t} dt, \qquad \forall x \in \Omega, \tag{5}$$

where $q > 1$, f is a given function on Ω say bounded from below and $\lambda > 0$ is a given parameter – the so-called *discount factor*, and where we denote by $a_t = a(X_t)$. Let us emphasize that this particular form of the *running cost* $g(x,a) = f(x) + \dfrac{1}{q} |a|^q$ is by no means essential for the analysis which follows: it just provides a simple but general enough model problem. Let us also mention that this choice of cost functions corresponds to the so-called *infinite horizon* problems and that other cases are considered in this paper.

Finally, we wish to minimize J i.e. we want to determine the *value function* (or Bellman function):

$$u(x) = \inf_{a \in \mathscr{A}} J(x,a), \qquad \forall x \in \Omega \tag{6}$$

and optimal (feedback) controls $\underset{\sim}{a}$ such that $u(x) = J(x, \underset{\sim}{a})$.

I.3. Description of the Associated Boundary Value Problem

We want now, in this section, to follow the heuristic dynamic programming approach to such optimal stochastic control problems: the dynamic programming argument (which can be viewed as a modern, extended version of Hamilton-Jacobi-Carathéodory theories for problems in the calculus of variations), leads to a nonlinear partial differential equation. More precisely, the dynamic programming principle, due to R. Bellman, indicates that the value function u given by (6) should satisfy the following second-order, quasilinear, elliptic equation

$$-\frac{1}{2} \Delta u + \frac{1}{p} |\nabla u|^p + \lambda u = f \quad \text{in} \quad \Omega, \tag{7}$$

where p is the conjugate exponent of q i.e. $p = \dfrac{q}{q-1}$. In fact, such a claim, even if we forget the heuristic aspect of Bellman's derivation of (7) is by no means obvious here, in view of the restriction to feedback controls and of the state constraints. But nevertheless (7) is to be expected for the value function u. This equation is a very particular case of the so-called Hamilton-Jacobi-Bellman equations. And at least for problems like exit problems or the ones corresponding to reflecting boundary conditions (as described in the preceding section), a rigorous derivation of the

Hamilton-Jacobi-Bellman equation and the analysis of such nonlinear p.d.e. are now available: see Fleming and Rishel [8]; Bensoussan and Lions [2, 3]; Krylov [11, 12]; Lions [16–18]; Lions and Trudinger [24, 25] and the bibliography therein.

Let us add to this general description that the exit problems lead to Dirichlet type boundary conditions like

$$u=\varphi \quad \text{on} \quad \Omega, \tag{8}$$

where φ is the exit cost i.e. the price to be paid for hitting the boundary at a point x of $\partial\Omega$. On the other hand, reflecting type boundary conditions lead to Neumann (or oblique derivative) type boundary conditions like for instance

$$\frac{\partial u}{\partial n}=\psi \quad \text{on} \quad \partial\Omega, \tag{9}$$

where n is the unit outward normal to $\partial\Omega$ and ψ is the reflection cost i.e. the price to be paid for reflecting on the boundary $\partial\Omega$ at the point x.

Finally, let us mention that another aspect of Bellman's dynamic programming argument is a rule for finding an optimal feedback control which in the case of (7) reduces to the choice

$$a(x)=-|\nabla u|^{p-2}\nabla u(x) \quad \text{for} \quad x\in\Omega. \tag{10}$$

Now, we go back to the state-constraints problem described in the preceding section and we ask ourselves the following question: what is the boundary condition (or any other characterization at $\partial\Omega$) we may expect for the value function u given by (6)? From the above considerations it is tempting to say that to discourage hitting the boundary we should impose an infinite exit cost or reflection cost i.e.

$$u(x)\to+\infty \quad \text{as} \quad \operatorname{dist}(x,\partial\Omega\to 0 \tag{11}$$

or

$$\frac{\partial u}{\partial n}(x)\to+\infty \quad \text{as} \quad \operatorname{dist}(x,\partial\Omega)\to 0 \tag{12}$$

[where $n(x)$ is defined near $\partial\Omega$ by $-\nabla(\operatorname{dist}(x,\partial\Omega))$]. More sophisticated formulations, which are also very natural from the control viewpoint, are: u is the maximum solution (or even subsolution) of (7); or: u is the upperenvelope of bounded solutions of (7).... Finally, for readers experienced with viscosity solutions, a possible form of the boundary condition could be

$$u-\varphi \quad \text{achieves its minimum over} \quad \Omega \tag{13}$$

for all $\varphi\in C^2(\bar{\Omega})$ [or $C^{1,1}(\bar{\Omega})$, or $C^1(\bar{\Omega})$, or even $C^{0,1}(\bar{\Omega})$]: this "viscosity formulation" will be explained below in Sect. IVV, see also Lions [17, 29] for the deterministic case.

It turns out, and the precise results are given in the next section, that if the latter formulations are always true, the choice between the boundary conditions (11) or (12) requires some careful analysis and will in fact depend on the behaviour of f

near $\partial\Omega$ and on q. This can easily be "justified" by a vague economical argument: if f blows up fast enough near $\partial\Omega$ of if q is large (remember that a has to blow up near $\partial\Omega$) then the cost functions will blow up at $\partial\Omega$ and so will u. Then, we should expect (11). On the other hand, if f, say, is bounded and if q is near 1 then it does not cost much to drive the state off $\partial\Omega$ and we may expect now u to be bounded on Ω. On the other hand, recalling Bellman's rule (10) for the optimal control and the fact that a cannot remain bounded if we want X_t to stay in Ω, we should expect that some condition like (12) holds. Of course, the reason for which we insist on conditions like (11) or (12) compared to a "maximum solution" characterization is because of the specific information contained in those formulations (which could turn out to be crucial for numerical purposes). Finally, note that $p=1$ is excluded in the p.d.e. results (see next Sect. I.4). This corresponds to the fact that it is impossible to force state-constraints with bounded controls. All these heuristic considerations will find their mathematical counter parts in the results presented in the next section.

I.4. Short Review of the Results

In this section, we present some of the results obtained in this paper on the simple example of the model equation (1) [equivalent to (7) after an obvious scaling]. In doing so, we follow the order of the sections below. To simplify the presentation we will always assume at least that $f\in C^1(\Omega)$, is bounded *from below*. We will denote by $d(x)=\operatorname{dist}(x,\partial\Omega)$ for all $x\in\bar{\Omega}$.

We begin with the case when the running cost f is not too large, while the other term in the cost function is quite large since we will assume $1<p\leqq 2$ i.e. $q\geqq 2$.

Theorem I.1. *Assume that $1<p\leqq 2$ and that f satisfies*

$$\lim\{f(x)d(x)^q/d(x)\to 0_+\}=C_1\geqq 0. \tag{14}$$

Then, there is a unique solution $u\in C^2(\Omega)$ of (1) such that $u(x)\to+\infty$ as $d(x)\to 0_+$. In addition, any solution $v\in C^2(\Omega)$ of (1) satisfies: $u\geqq v$ on Ω. Finally, if C_0 is the unique positive root of $\left(\frac{2-p}{p-1}\right)^p C_0^p-\frac{2-p}{(p-1)^2}C_0-C_1=0$ if $p<2$, $C_0^2-C_0-C_1=0$ if $p=2$, then u satisfies

$$\left.\begin{aligned}\lim\left\{u(x)d(x)^{\frac{2-p}{p-1}}/d(x)\to 0_+\right\}&=C_0\quad\text{if}\quad p<2\\ \lim\{u(x)\,|\mathrm{Log}\,d(x)|^{-1}/d(x)\to O_+\}&=C_0\quad\text{if}\quad p=2.\end{aligned}\right\}\ \Box \tag{15}$$

We now turn to the case when both terms in the running cost are not too large: in particular we assume that $p>2$ i.e. $1<q\leqq 2$.

Theorem I.2. *Assume that $p>2$ and that f satisfies*

$$\lim\{f(x)d(x)/d(x)\to 0_+\}=0,\quad \textit{for some}\quad \beta\in(0,p). \tag{16}$$

Then, all solutions $v\in C^2(\Omega)$ of (1) bounded from below are bounded and may be extended continuously to $\bar{\Omega}$. And there exists a maximum solution $u\in C^2(\Omega)$ of (1). In *addition, u satisfies*

$$\liminf_{y\in\Omega,\, y\to x} \{u(y)-u(x)\}\,|y-x|^{-\alpha}<0\,, \quad \textit{for all} \quad x\in\partial\Omega \tag{17}$$

where $\alpha=(p-2)/(p-1)$.

Furthermore, if $\liminf\{f(x)d(x)^\gamma/d(x)+0_+\}>0$ *for some* $\gamma\in(q,\beta)$, *then* (17) *holds with* $\alpha=1-\gamma/p$. □

Also, if additional assumptions on Ω or f are made, we are able to sharpen (17) or prove (12) [or even sharper estimates then (17) and (12)...].

The next case concerns the situation when the running cost f is blowing up near the boundary very fast. We have the

Theorem I.3. *Assume that* f *satisfies*

$$\liminf\{f(x)d(x)^\beta/d(x)\to 0_+\}>0\,, \quad \textit{for some} \quad \beta\geqq\max(p,q)\,. \tag{18}$$

Then, any solution $v\in C^2(\Omega)$ *of* (1) *bounded from below converges to* $+\infty$ *as* $d(x)$ *goes to* 0. *In addition, such a solution is unique if* (18) *is replaced by*

$$\lim\{f(x)d(x)^\beta/d(x)\to 0_+\}=C_1>0\,, \quad \textit{for some} \quad \beta\geqq\max(p,q) \tag{18'}$$

and this solution, denoted by u, satisfies

$$\lim\{u(x)d(x)^\alpha/d(x)\to 0_+\}=C_0\,, \tag{19}$$

where $d(x)^\alpha$ *is replaced by* $|\mathrm{Log}\,d(x)|^{-1}$ *if* $\beta=p\geqq q$; $\alpha=\dfrac{\beta}{p}-1$ *and* $C_0=\left(\dfrac{C_1}{\alpha}\right)^{1/p}$ *if* $\beta>\max(p,q)$; $C_0=C_1^{1/p}$ *if* $\beta=p>2$; $C_0=(1+C_1)^{1/2}$ *if* $\beta=p=2$. □

Roughly speaking, the combination of Theorems I.1–I.3 cover all possible situations. One way of unifying the above results is by the use of the viscosity formulation of the various boundary conditions encountered above namely

$$u-\varphi \quad \text{achieves its minimum over } \Omega\,, \quad \text{for all} \quad \varphi\in C^2(\bar\Omega)\,. \tag{20}$$

Theorem I.4. *Assume that* $p>1$ *and* $\beta>0$ *and that either* f *is bounded or* $f(x)d(x)^\beta$ *converges to a positive constant as* $d(x)$ *converges to* 0_+. *Then, there is a unique* $u\in C^2(\Omega)$ *solution of* (1) *satisfying* (20). □

This is a nonexhaustive list of results since we will consider below many related questions like the stochastic interpretation of the above solutions, the existence of optimal controls, the ergodic problem, i.e. $\lambda\to 0_+$, the approximation of such solutions, extensions to more general data f or Hamiltonians. Finally, we will also briefly explain how the techniques we introduce allow us to treat similar boundary conditions for other types of nonlinear equations.

I.5. Organization of the Paper

As usual in stochastic control problems, various strategies are possible. One can use p.d.e. methods to derive the existence of a smooth solution of the associated HJB equation – here a second order quasilinear elliptic equation with strong nonlinearities in the gradient and singular boundary conditions. The uniqueness question may be solved directly by p.d.e. methods or by checking that any solution is the value function. Finally, one builds an optimal control using, whenever it is possible, the solution of the HJB equation. This is why some sections below deal with purely p.d.e. questions while others are concerned with the stochastic interpretation. Another distinction is made below between what we call the model problem (1) and more general equations. This artificial distinction is made only to simplify the exposition. In fact, in all sections below, we adopt a layered presentation with gradual generalizations where we just explain the required modifications of proofs.

II. Subquadratic Hamiltonians

We will be dealing here with (1) in the case when $1<p\leqq 2$.

II.1. Bounded Data

We begin with the case of bounded data i.e. we assume that $f\in L^\infty(\Omega)$.

Theorem II.1. *There is a unique solution* $u\in W^{2,r}(\Omega)$ $(\forall r<\infty)$ *of* (1) *such that* $u(x)\to+\infty$ *as* $d(x)\to 0_+$. *In addition, if* $C_0=(p-1)^{\frac{p-2}{p-1}}(2-p)^{-1}$ *when* $p<2$, $C_0=1$ *when* $p=2$, *then* (15) *holds. Finally, let* $v\in L^1_{\mathrm{loc}}(\Omega)$ *satisfy*

$$-\Delta v+p|\xi|^{p-2}\xi\cdot\nabla v+\lambda v\leqq f+(p-1)|\xi|^p \quad in \quad \mathscr{D}'(\Omega), \qquad \forall\xi\in\mathbb{R}^n \tag{21}$$

then $v\leqq u$ *a.e. in* Ω; *in other words, u is the maximum* L^1_{loc} *subsolution.* □

Corollary II.1. *Let* $f_1, f_2\in L^\infty(\Omega)$ *and let* u_1, u_2 *be the corresponding solutions of* (1) *which go to* $+\infty$ *on* $\partial\Omega$. *Then, we have*

$$\sup_\Omega\,(u_1-u_2)^+\leqq\frac{1}{\lambda}\sup_\Omega\,(f_1-f_2)^+. \quad \square$$

Proof of Corollary II.1. $u_1-\frac{1}{\lambda}\sup_\Omega\,(f_1-f_2)^+$ is a subsolution of (1) with f replaced by f_2 so by Theorem II.1 $u_1\leqq u_2+\frac{1}{\lambda}\sup_\Omega\,(f_1-f_2)^+$. □

The proof of Theorem II.1 is unfortunately a bit longer and we split it into several parts. First (step 1), we compute the explosion rate of such a solution and this trivial computation leads to families of super and subsolutions. Next (step 2), we build a minimum and a maximum "explosive" solution which have the same leading behaviour near the boundary. Then (step 3), we prove the uniqueness and (15). Finally (step 4), we prove the "maximal subsolution" property.

Step 1. It is reasonable to try to obtain the leading term in an expansion of a solution of (1) blowing up at the boundary by the following ansatz near the boundary: $u(x)\simeq C_0 d(x)^{-\alpha}$. The most explosive term in $[-\Delta u+|\nabla u|^p+\lambda u-f]$ is then

$$-C_0\alpha(\alpha+1)d^{-\alpha-2}+C_0^p\alpha^p d^{-(\alpha+1)p},$$

where we used (twice) the fact that $|\nabla d|=1$ near the boundary (in fact, as it is well-known: $|\nabla d|=1$ at each differentiability point of d, and d is smooth near the boundary if Ω is smooth). This leads to the choices

$$\alpha=\frac{2-p}{p-1}, \quad C_0=\alpha^{-1}(\alpha+1)^{1/(p-1)} \quad \text{if} \quad p<2.$$

Of course, if $p=2$ one replaces $C_0 d^{-\alpha}$ by $-C_0 \operatorname{Log} d$ and one finds $C_0=1$.

In order to use in a meaningful way the above formal consideration, we build two families of "approximations of $C_0 d^{-\alpha}$", each of which is a two-parameter family, where we first denote by d any smooth function, say $C^2(\bar{\Omega})$, on $\bar{\Omega}$ equal to $\operatorname{dist}(x,\partial\Omega)$ near the boundary, say for $\operatorname{dist}(x,\partial\Omega)\leqq\delta_0$ with $\delta_0>0$. Then, we introduce for $\varepsilon,\delta\geqq 0$

$$\left.\begin{aligned}\bar{w}_{\varepsilon,\delta}&=(C_0+\varepsilon)(d-\delta)^{-\alpha}+C_\varepsilon\\ \underline{w}_{\varepsilon,\delta}&=(C_0-\varepsilon)(d+\delta)^{-\alpha}-C_\varepsilon\end{aligned}\right\} \tag{22}$$

for some large constant C_ε to be determined. Of course, if $p=2$ then $(d\pm\delta)^{-\alpha}$ is replaced by $-\operatorname{Log}(d\pm\delta)$. Notice also that if $\underline{w}_{\varepsilon,\delta}$ is defined and smooth on Ω, $\bar{w}_{\varepsilon,\delta}$ is only defined on $\Omega_\delta=\{x\in\Omega, \operatorname{dist}(x,\partial\Omega)>\delta\}$ at least for $\delta<\delta_0$ (δ_0 to be choosen small enough; $0\leqq\delta<\delta_0$ will always be assumed in this proof). In fact, it will be handy to consider d as a smooth function on $\mathbb{R}^N$, say $C^2(\mathbb{R}^N)$, such that: $d(x)=\operatorname{dist}(x,\partial\Omega)$ if $x\in\bar{\Omega}$, $\operatorname{dist}(x,\partial\Omega)\leqq\delta_0$; $d(x)\geqq\delta_0$ if $\operatorname{dist}(x,\partial\Omega)\geqq\delta_0$, $x\in\Omega$; $d(x)=-\operatorname{dist}(x,\partial\Omega)$ if $x\notin\Omega$ and $\operatorname{dist}(x,\partial\Omega)\leqq\delta_0$; $d(x)\leqq-\delta_0$ if $\operatorname{dist}(x,\partial\Omega)\geqq\delta_0$, $x\notin\bar{\Omega}$. Observe of course that $|\nabla d|=1$ in $\{\operatorname{dist}(x,\partial\Omega)\leqq\delta_0\}$ and that $d(x)=-\delta=\operatorname{dist}(x,\partial\Omega_\delta)$ if $\operatorname{dist}(x,\partial\Omega)\leqq\delta_0$ while $d(x)+\delta=\operatorname{dist}(x,\partial\Omega^\delta)$ if $\operatorname{dist}(x,\partial\Omega)\leqq\delta_0$, where

$$\Omega^\delta=\{x\in\mathbb{R}^N, \operatorname{dist}(x,\bar{\Omega})\leqq\delta\}=\{x\in\mathbb{R}^N/d(x)\geqq-\delta\}.$$

So that, we may consider $\underline{w}_{\varepsilon,\delta}$ to be defined on Ω^δ. (Notice that such a function d exists as soon as Ω is open bounded and has a C^2-regular boundary $\partial\Omega$.)

We conclude these preliminaries with the following computations

$$\begin{aligned}&-\Delta\bar{w}_{\varepsilon,\delta}+|\nabla\bar{w}_{\varepsilon,\delta}|^p+\lambda\bar{w}_{\varepsilon,\delta}-f\\ &\quad=-\alpha(\alpha+1)(C_0+\varepsilon)(d-\delta)^{-\alpha-2}|\nabla d|^2+\alpha(C_0+\varepsilon)(d-\delta)^{-\alpha-1}\Delta d\\ &\qquad+\alpha^p(C_0+\varepsilon)^p(d-\delta)^{-p(\alpha+1)}|\nabla d|^p+\lambda(C_0+\varepsilon)(d-\delta)^{-\alpha}+\lambda C_\varepsilon-f.\end{aligned}$$

Recalling that $\alpha+2=(\alpha+1)p$ and $\alpha^p C_0^p=\alpha(\alpha+1)C_0$, we deduce easily for $\varepsilon\leqq 1$, $\delta\leqq\delta_0$

$$-\Delta\bar{w}_{\varepsilon,\delta}+|\nabla\bar{w}_{\varepsilon,\delta}|^p+\lambda\bar{w}_{\varepsilon,\delta}-f\geqq\nu\varepsilon(d-\delta)^{-\alpha-2}+\lambda C-C(1+(d-\delta)^{-\alpha-1})$$

for some $\nu>0$, $C\geqq 0$. And we can choose C_ε large enough in order to find

$$-\Delta\bar{w}_{\varepsilon,\delta}+|\nabla\bar{w}_{\varepsilon,\delta}|^p+\lambda\bar{w}_{\varepsilon,\delta}\geqq f \quad \text{in} \quad \Omega_\delta. \tag{23}$$

Similarly, one shows that C_ε can be choosen large enough to have:

$$-\Delta\underline{w}_{\varepsilon,\delta}+|\nabla\underline{w}_{\varepsilon,\delta}|^p+\lambda\underline{w}_{\varepsilon,\delta}\leqq f \quad \text{in} \quad \Omega^\delta. \tag{24}$$

Step 2. Building a minimum "explosive" solution is easy in the subquadratic case. Indeed, one solves

$$-\Delta u_R + |\nabla u_R|^p + \lambda u_R = f \quad \text{in} \quad \Omega, \qquad u_R \in W^{2,r}(\Omega) \quad (\forall r < \infty) \tag{25}$$

with boundary conditions going to infinity (as $R \to \infty$) like for instance

$$u_R = R \quad \text{on} \quad \partial\Omega \tag{26}$$

or

$$u_R = \underline{w}_{\varepsilon, 1/R} \quad \text{on} \quad \partial\Omega, \quad \text{for any fixed} \quad \varepsilon > 0. \tag{27}$$

Since $p \leqq 2$, the existence follows from standard results on subquadratic quasi-linear equations (see for example Amann and Crandall [1]). In view of the maximum principle (we have to use here the slightly more general form of maximum principle in Sobolev spaces – see for example Bony [5] and Lions [23]) we deduce in the case of (27) for example

$$\underline{w}_{\varepsilon, 1/R} \leqq u_R \leqq u_{R'} \leqq \bar{w}_{\varepsilon'} \quad \text{if} \quad 0 < R < R', \quad \forall \varepsilon' > 0$$

and where $\bar{w}_\varepsilon = \bar{w}_{\varepsilon, 0}$. The last inequality of this string comes from the maximum principle provided we observe that $u_{R'} < \bar{w}_\varepsilon$ near $\partial\Omega$ since $\bar{w}_\varepsilon$ blows up at the boundary.

Hence, u_R is bounded in L^∞_{loc}. This combined with (25) implies that u_R is bounded in $W^{2,r}_{\text{loc}}(\Omega)$ ($\forall r < \infty$): this can be deduced either from [1] using again the fact that we are dealing with a subquadratic Hamiltonian or by using the gradient estimates of the appendix (see Lions [16, 19]) which yield bounds in $W^{1,\infty}_{\text{loc}}(\Omega)$ and then using (25). Anyway, u_R converges (as $R \to \infty$) to a solution $\underline{u}$ of (1) in $W^{2,r}_{\text{loc}}(\Omega)$ ($\forall r < \infty$) which also satisfies $\underline{w}_\varepsilon \leqq \underline{u} \leqq \bar{w}_{\varepsilon'}$ $\forall \varepsilon' > 0$. Next, we claim that $\underline{u} \geqq \underline{w}_{\varepsilon'}$ for all $\varepsilon' > 0$. Indeed for any $R' > 0$, we can find R such that $\underline{w}_{\varepsilon', 1/R'} \leqq \underline{w}_{\varepsilon', 1/R}$ and letting R' go to $+\infty$, we conclude easily.

We now claim that $\underline{u}$ is the minimum "explosive" solution of (1). Indeed, let u be another solution of (1) in $W^{2,r}_{\text{loc}}(\Omega)$ ($\forall r < \infty$) such that $u \to +\infty$ as $d(x) \to 0_+$ then by maximum principle $u \geqq u_R$ in Ω and thus passing to the limit we obtain $u \geqq \underline{u}$ in Ω.

To build a maximum explosive solution, we consider the preceding minimum explosive solution $\underline{u}_\delta$ in Ω_δ and we let δ go to 0. Recall that we have

$$(C_0 - \varepsilon)(d + \delta)^{-\alpha} - C_\varepsilon \leqq \underline{u}_\delta \leqq \bar{w}_{\varepsilon, \delta}, \qquad \forall \varepsilon > 0$$

and clearly enough $\underline{u}_\delta \geqq \underline{u}_{\delta'}$ if $0 < \delta' < \delta$. Therefore, passing to the limit, exactly as above we find a solution $\bar{u}$ of (1) such that

$$\underline{w}_\varepsilon \leqq \bar{u} \leqq \bar{w}_\varepsilon \quad \text{in} \quad \Omega.$$

The fact that $\bar{u}$ is the maximum explosive solution is proved by using again the maximum principle to show (with the above notations)

$$u \leqq \underline{u}_\delta \to \bar{u} \quad \text{as} \quad \delta \text{ goes to } 0.$$

In conclusion, we found solutions $\underline{u}$, $\bar{u} \in W^{2,r}_{\text{loc}}(\Omega)$ ($\forall r < \infty$) of (1) such that

$$\underline{w}_\varepsilon \leqq \underline{u} \leqq u \leqq \bar{u} \leqq \bar{w}_\varepsilon \quad \text{in} \quad \Omega, \quad \text{for all} \quad \varepsilon > 0, \tag{28}$$

where u is any solution of (1) in $W^{2,r}_{\text{loc}}(\Omega)$ ($\forall r < \infty$) such that $u \to \infty$ as $d \to 0_+$.

Step 3: Uniqueness. It is of course enough to show that $\underline{u} \equiv \bar{u}$ in Ω. We first observe that (28) implies that $\bar{u}(x)(\underline{u}(x))^{-1}$ converges to 1 as $d(x) \to 0_+$. Therefore, if we denote by $m = \inf_{\Omega} f(x)$, we deduce that for all $\theta \in (0,1)$

$$\underline{u}(x) > \theta \bar{u}(x) + (1-\theta)m/\lambda \quad \text{in a neighbourhood of} \quad \partial\Omega .$$

In addition, $w = \theta\bar{u} + (1-\theta)m/\lambda$ satisfies in Ω

$$-\Delta w + |\nabla w|^p + \lambda w \leqq \theta f + (1-\theta)m \leqq f .$$

Therefore, we deduce easily by the maximum principle

$$w \leqq \underline{u} \quad \text{in} \quad \Omega$$

and we conclude letting θ go to 1.

Step 4. We wish to prove that the unique explosive solution $\underline{u}$ of (1) that we built above is also the maximum L^1_{loc} subsolution. Let $v \in L^1_{\text{loc}}(\Omega)$ satisfy (21). In order to avoid some rather unpleasant technicalities, we begin with the case $f \in C(\bar{\Omega})$: in that case, we smooth v by convolution i.e. we consider $v_n = v * \varrho_n$ where $\varrho \in \mathscr{D}(\mathbb{R}^N)$, $0 \leqq \varrho$, $\int_{\mathbb{R}^N} \varrho dx = 1$, $\operatorname{supp}(\varrho) \subset B_1$ and $\varrho_n = n^N \varrho(n\,\cdot)$. Then, if $\delta > \frac{1}{n}$, we find easily

$$-\Delta v_n + |\nabla v_n|^p + \lambda v_n \leqq f * \varrho_n \quad \text{in} \quad \Omega_\delta$$

and $f * \varrho_n \leqq f + \varepsilon_n$ where $\varepsilon_n \underset{n}{\to} 0$. Therefore, we deduce

$$\left(v_n - \frac{\varepsilon_n}{\lambda_n}\right) \leqq \underline{u}_\delta \quad \text{if} \quad \delta > \frac{1}{n}$$

and we conclude letting n go to $+\infty$ and then δ go to 0_+.

If $f \in L^\infty(\Omega)$, we obtain by the above proof that $v_n \leqq \underline{u}^n_\delta$ where $\underline{u}^n_\delta$ is the explosive solution in Ω_δ corresponding to $f * \varrho_n$ $\left(\text{still if } \delta > \frac{1}{n}\right)$. In addition, the proof made above also shows that $\underline{u}^n_\delta$ is bounded in $L^\infty_{\text{loc}}(\Omega_\delta)$ and thus in $W^{2,r}_{\text{loc}}(\Omega_\delta)$ $(\forall r < \infty)$ since $f * \varrho_n$ is bounded in $L^\infty(\Omega)$: in fact, one may even choose C_ε such that

$$(C_0 - \varepsilon)(d-\delta)^{-\alpha} - C_\varepsilon \leqq \underline{u}^n_\delta \leqq (C_0 + \varepsilon)(d+\delta)^{-\alpha} + C_\varepsilon$$

(with the usual modifications if $p=2$). Then, we may pass to the limit as n goes to $+\infty$ and $\underline{u}^n_\delta$ (or subsequences) converges to a solution of (1) in Ω_δ thus below $\underline{u}_\delta$ (in fact it is $\underline{u}_\delta$ because the above inequality shows it blows up at $\partial\Omega_\delta$). Therefore, $v \leqq \underline{u}_\delta$ in Ω_δ and we conclude letting δ go to 0. □

Remark II.1. One may deduce from the above arguments the "continuity" of the explosive solution with respect to Ω, p or f (for the weak $L^\infty *$ topology).

Remark II.2. By a convenient (and technical) variation of the above method one can show that it is possible to replace $f \in L^\infty(\Omega)$ by $f \in L^p_{\text{loc}}(\Omega)$ $\left(p > \frac{N}{2}\right)$, f bounded from below and f bounded near $\partial\Omega$.

II.2. General Data

We now wish to allow some data f which may not be bounded near $\partial\Omega$.

Theorem II.2. *Let $f \in L^\infty_{loc}(\Omega)$, assume that f is bounded from below and that f satisfies* (14). *Then, Theorem II.1 still holds provided one replaces C_0 by the unique positive solution of the equation $\left(\frac{2-p}{p-1}\right)^p C_0^p - \frac{2-p}{(p-1)^2} C_0 - C_1 = 0$ if $p<2$, $C_0^2 - C_0 - C_1 = 0$ if $p=2$.*

Proof. We only present the main modifications in the preceding proof. With the above new value of C_0, one builds exactly as in the proof of Theorem II.1 a maximum explosive solution $\bar{u}$ of (1) such that

$$(C_0-\varepsilon)d^{-\alpha} - C_\varepsilon \leqq \bar{u} \leqq (C_0+\varepsilon)d^{-\alpha} + C_\varepsilon \quad \text{in} \quad \Omega, \quad \forall \varepsilon > 0. \tag{29}$$

The above equation for C_0 comes into the picture when making the formal computations of Step 1 and balancing the various leading terms in $d^{-\alpha-2} = d^{-q}$. The only modification in the proof of Theorem II.1 consists in proving that there exists a minimum explosive solution $\underline{u}$ which also satisfies (29). To this end, we observe that $\underline{w}_{\varepsilon,\delta}$ is a subsolution of (1) when Ω is replaced by Ω^δ and f is replaced by

$$f_\delta = \min(f, C_2 + C_3(d+\delta)^{-q}) \quad \text{in} \quad \Omega, \quad = C_2 + C_3(d+\delta)^{-q} \quad \text{in} \quad \Omega^\delta - \Omega,$$

where C_3, C_2 are positive constants such that $C_3 > C_1$, $C_2 + C_3 d^{-q} > f$ in Ω. Obviously, $f_\delta \in L^\infty(\Omega)$. Therefore, by Theorem II.1 and its proof, there exists a unique explosive solution u_δ of (1) with f replaced by f_δ, [obtained by an increasing limit of solutions of (1) with finite boundary values] and $u_\delta \geqq \underline{w}_{\varepsilon,\delta}$. Since $f \geqq f_\delta$, any explosive solution of (1) is above u [use the maximum principle with the approximating bounded solutions of (1)] and thus in particular $\bar{u} \geqq u_\delta$. From this, we deduce easily letting δ go to 0 the existence of a minimum explosive solution of (1) $\underline{u}$ satisfying (29).

Remark II.3. The analogues of Remarks II.1–II.2 still hold: notice only that the stability with respect to f holds with respect to the weak $* L^\infty$ topology provided the data f are uniformly bounded from below and, satisfy (14) with C_1 bounded and $f(x) \leqq Cd^{-q} + C$ for some $C \geqq 0$.

Remark II.4. The proof also shows that if $\bar{w}$ is a supersolution of (1) which blows up on $\partial\Omega$ i.e.

$$-\Delta\bar{w} + |\nabla\bar{w}|^p + \lambda\bar{w} \geqq f \quad \text{in} \quad \Omega, \quad \bar{w} \to 0 \quad \text{as} \quad d(x) \to 0_+$$

then $\bar{w} \geqq u$.

Remark II.5. If we allow f to go to $-\infty$ near $\partial\Omega$ (or some points of $\partial\Omega$) then the situation is a bit more complex. Let $f \in L^\infty_{loc}(\Omega)$, if we assume (14) with $C_1 = 0$ then the above result is no longer true. In that case, there still exists a maximum explosive solution which behaves as $C_0 d^{-\alpha}$ and is the unique solution going to $+\infty$ as $C_0 d^{-\alpha}$. However, in general, there may exist other solutions going to $+\infty$

less rapidly: indeed, consider

$$f(x)=\frac{\Delta d}{d}-\frac{|\nabla d|^2}{d^2}+\frac{|\nabla d|^p}{d^p}-\lambda\,\mathrm{Log}\,d\,.$$

If $1<p<2$, f behaves like $-\frac{1}{d^2}$ near $\partial\Omega$ and thus satisfies (14). And notice that $u(x)=-\mathrm{Log}\,d(x)$ is then a solution of (1) which goes to $+\infty$ as $d(x)$ goes to 0.

If we assume (14) and $C_1>0$, then f is bounded from below and Theorem II.2 applies. Now, if we assume (14) and $C_1<0$, then there are two positive solutions C_0 of the equation stated in Theorem II.2 say $0<C_0^-<C_0^+$ and $C_0^-\to 0$, $C_0^+\to C_0$ as $C_1\to 0_-$. Again, there exists a maximum explosive solution of (1) behaving near $\partial\Omega$ as $C_0^+d^{-\alpha}$ and it is the unique such solution. But there also exists in general another explosive solution of (1) behaving near $\partial\Omega$ as $C_0^-d^{-\alpha}$: for instance, consider $f=-\Delta w+|\nabla w|^p+\lambda w$ where $w=C_0^-d^{-\alpha}$. □

II.3. Asymptotic Expansions Near the Boundary

In this section, we want to precise a bit the behaviour near the boundary of solutions which blow up at the boundary. Even if we will not present a complete asymptotic expansion near the boundary (which should include $[q]-1$ singular terms plus a bounded term where $[q]$ denotes the integer part of q), the methods we use should give it and we leave the awful computations to a courageous reader. We will only prove the

Theorem II.3. *Let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below and assume that*

$$\lim\{f(x)d(x)^{q-1}/d(x)\to 0_+\}=0\,. \tag{30}$$

We denote by u the unique solution of (1) in $W^{2,r}(\Omega)\,(\forall r<\infty)$ which goes to $+\infty$ on $\partial\Omega$. Then, if $p\in(\frac{3}{2},2]$ i.e. $q\in[2,3)$, $u-\frac{C_0}{d^\alpha}$ is bounded on Ω when $p<2$ while $u+\log d$ is bounded on Ω when $p=2$. Next, if $p\in(1,\frac{3}{2}]$ we set

$$\left.\begin{aligned} C_1(x)&=-\frac{1}{2}\frac{\alpha}{\alpha-1}C_0\Delta d(x) \quad &&\text{if}\quad p<\frac{3}{2},\\ C_1(x)&=-\frac{1}{2}C_0\Delta d(x) &&\text{if}\quad p=\frac{3}{2},\end{aligned}\right. \tag{31}$$

and we have

$$\left.\begin{aligned} &\left\{u-\frac{C_0}{d^\alpha}\right\}d^{\alpha-1}\to C_1 &&\text{as}\quad d\to 0_+\quad \text{if}\quad p<\frac{3}{2},\\ &\left\{u-\frac{C_0}{d}\right\}|\log d|^{-1}\to C_1 &&\text{as}\quad d\to 0_+\quad \text{if}\quad p=\frac{3}{2}.\end{aligned}\right\} \tag{32}$$

Proof. We begin with the case $1<p\leqq\frac{3}{2}$. In view of the results of the previous sections, it is enough build appropriate sub and supersolutions which blow up

near $\partial\Omega$. To this end, we consider

$$\left.\begin{array}{lll} w_\varepsilon^+=\dfrac{C_0}{d^\alpha}+\dfrac{(C_1+\varepsilon)}{d^{\alpha-1}}+C_\varepsilon, & w_\varepsilon^-=\dfrac{C_0}{d^\alpha}+\dfrac{(C_1-\varepsilon)}{d^{\alpha-1}}-C_\varepsilon & \text{if } p<\dfrac{3}{2}, \\ w_\varepsilon^-=\dfrac{C_0}{d}-(C_1+\varepsilon)\operatorname{Log} d+C_\varepsilon, & w_\varepsilon^-=\dfrac{C_0}{d}-(C_1-\varepsilon)\operatorname{Log} d-C_\varepsilon & \text{if } p=\dfrac{3}{2}, \end{array}\right\} \tag{33}$$

where C_ε is a positive constant to be determined. Tedious computations show that, provided C_1 is given by (31) and C_ε is large enough, w_ε^+ (resp. w_ε^-) is a supersolution of (1) [resp. subsolution of (1)]. Therefore, $w_\varepsilon^-\leqq u\leqq w_\varepsilon^+$ in Ω for all $\varepsilon>0$ and (32) is proved.

Next, if $\frac{3}{2}<p\leqq 2$, we also want to build convenient sub and supersolutions. However, in this case, the choices are not straightforward as above. Indeed, recalling that $\alpha=\dfrac{2-p}{p-1}$ we choose

$$w_\varepsilon^+=\frac{C_0}{d^\alpha}-(C_1+\varepsilon)d^{1-\alpha}+C_\varepsilon, \quad w_\varepsilon^-=\frac{C_0}{d^\alpha}-(C_1-\varepsilon)d^{1-\alpha}-C_\varepsilon, \tag{34}$$

where

$$C_1=-\frac{1}{2}\frac{\alpha}{1-\alpha}C_0\Delta d \quad \text{if} \quad p<2, \qquad C_1=-\frac{1}{2}\Delta d \quad \text{if} \quad p=2.$$

Again, one can check that w_ε^+, w_ε^- for conveniently large C_ε are sub and supersolutions of (1) and since they go to $+\infty$ at $\partial\Omega$ we deduce that $w_\varepsilon^-\leqq u\leqq w_\varepsilon^+$ in Ω and we conclude. $\square$

Remark II.6. In the various bounds on the behaviour of explosive solutions near the boundary, it may seem strange that the leading terms are not continuous with respect to p (as p goes to 2 for example). Similarly, in (34) the term $d^{1-\alpha}$ vanishes and could seem to be irrelevant. However – and this fits well with the stochastic control interpretation – these questions disappear if we look for formal expansions of the gradient obtained by differentiating these expansions for the solution: indeed, in Theorem II.1, u behaves like

$$(p-1)^{\frac{p-2}{p-1}}\frac{1}{2-p}\,d(x)^{-\frac{2-p}{p-1}}$$

so $\nabla u(x)$ should behave like $-(p-1)^{-1/(p-1)}\nabla d(x)d^{-1/(p-1)}$ and when p goes to 2 this quantity goes to $-\nabla d(x)d^{-1}$ which is precisely the gradient of $-\log d$. A similar explanation holds for (34).

III. Infinite Boundary Conditions and Blowing up Data

In this section, we consider the case of data f blowing up at the boundary fast enough to force solutions of (1) bounded from below to blow up at the boundary. This also will yield some uniqueness results. The results of this section correspond to Theorem I.3.

III.1. Forced Infinite Boundary Conditions

Theorem III.1. *Assume that $f\in L^\infty_{\mathrm{loc}}(\Omega)$ satisfies* (18). *Then, any solution u of* (1) *$W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) which is bounded from below converges to $+\infty$ as $d(x)$ goes to* 0.

Remark III.1. The proof below may be adapted to treat the case of $f\in L^r_{\mathrm{loc}}(\Omega)$ satisfying (18) with $r>N$.

Remark III.2. In general, there may exist solutions of (1) which are not bounded from below. For instance, take

$$f(x)=-\frac{\alpha C_0\Delta d}{d^{\alpha+1}}+\frac{C_0\alpha(\alpha+1)}{d^{\alpha+2}}|\nabla d|^2+C_0^p\alpha^p|\nabla d|^p d^{-(\alpha+1)p}-\lambda C_0 d^{-\alpha}$$

with α, $C_0>0$, $u(x)=-\frac{C_0}{d^\alpha}$ is obviously a solution of (1) and f satisfies (18) with $\beta=\max((\alpha+1)p,\alpha+2)$. And it is easy to check that any $\beta>\max(p,q)$ can be reached with a convenient α [in fact even $\beta=\max(p,q)$] may be reached provided we replace $-C_0d^{-\alpha}$ by $C_0\log d$ for $\beta=p\geqq q$. It is also worth noticing that such solutions may exist for linear equations like

$$-\Delta u+u=f \quad \text{in} \quad \Omega$$

provided f behaves like $\frac{C_1}{d^\beta}$ with $\beta\geqq 2$ near the boundary.

Proof of Theorem III.1. Even if the arguments are very much similar, we will have to consider two different cases namely $\beta\geqq p>2$ and $\beta\geqq q\geqq p$. In both cases, the strategy of proof consists in picking a point x_0 at a distance $2r$ of the boundary, working in the ball $B(x_0,r)$ rescaling the equation conveniently in order to deduce that $\liminf\{u(x)\,|\,d(x)\to 0_+\}$ is more than a fixed constant K_0 and then reiterating the argument to show that $\liminf\{u(x)\,|\,d(x)\to 0_+\}\geqq nK_0$ for all $n\geqq 1$.

Without loss of generality (add a large constant to u) we may assume that $u\geqq 0$ in Ω and that $f\geqq C_2d^{-\beta}$ for some $C_2>0$, with $\beta=\max(p,q)$. Next, let $r>0$ and let x_0 be any point in Ω such that $d(x_0)=2r$. Clearly, we have

$$-\Delta u+|\nabla u|^p+\lambda u\geqq C_3r^{-\beta} \quad \text{in} \quad B(x_0,r), \qquad u|_{\partial B(x_0,r)}\geqq 0, \tag{35}$$

where $C_3=C_2 2^{-\beta}$. Using the existence results of Lions [16], we deduce that $u\geqq \tilde u_r(x-x_0)$ in $B(x_0,r)$ where $\tilde u_r\in C^2(\overline{B(0,r)})$ solves

$$-\Delta\tilde u_r+|\nabla\tilde u_r|^p+\lambda\tilde u_r=C_3r^{-\beta} \quad \text{in} \quad B(0,r), \qquad \tilde u|_{\partial B(0,r)}=0. \tag{36}$$

Next, in the case when $1<p\leqq 2\leqq q=\beta$, we introduce $u_r(x)=r^\alpha\tilde u_r(rx)$ for $x\in B(0,1)$ where $\alpha=(2-p)/(p-1)$ so that u_r solves

$$-\Delta u_r+|\nabla u_r|^p+\lambda r^2u_r=C_3 \quad \text{in} \quad B(0,1), \qquad u_r|_{\partial B(0,1)}=0. \tag{37}$$

And using the estimates of [21], one checks easily that u_r, as r goes to 0, converges uniformly to the solution u_0 of

$$-\Delta u_0+|\nabla u_0|^p=C_3 \quad \text{in} \quad B(0,1), \qquad u_0|_{\partial B(0,1)}=0. \tag{38}$$

Observing that $u_0>0$ in $B(0,1)$ (strong maximum principle) and so $u_0(0)>0$, we deduce easily that if $p<2<q=\beta$ then u blows up at $\partial\Omega$ and $\liminf\{u(x)d(x)^\alpha \mid d(x)\to 0_+\}>0$.

Now, if $p=2=q=\beta$, the above argument only shows

$$\liminf\{u(x) \mid d(x)\to 0_+\} \geqq K_0>0, \tag{39}$$

where $K_0=u_0(0)$.

In the other case i.e. $2<p=\beta$, we introduce $u_r(x)=\tilde{u}_r(rx)$ for $x\in B(0,1)$ so that $u_r\in C^2(\overline{B(0,1)})$ solves

$$-r^{p-2}\Delta u_r+|\nabla u_r|^p+\lambda r^p u_r=C_3 \quad \text{in} \quad B(0,1), \qquad u_r|_{\partial B(0,1)}=0.$$

And using the results of Lions [21], one sees that u_r converges uniformly to the unique viscosity solution u_0 in $C(\overline{B(0,1)})$ of

$$|\nabla u_0|^p=C_3 \quad \text{in} \quad B(0,1), \qquad u_0|_{\partial B(0,1)}=0$$

which is in fact explicitely given by

$$u_0(x)=C_3^{1/p}(1-|x|).$$

Therefore, in this case also, we prove that (39) holds with $K_0=C_3^{1/p}$.

In particular, for any $\varepsilon>0$, there exists $s_\varepsilon>0$ such that for $x\in\Omega$, $d(x)<s_\varepsilon$ then $u(x)\geqq K_0-\varepsilon$. Then, we go back to (35) replacing the boundary inequality by $u|_{\partial B(x_0,r)}\geqq K_0-\varepsilon$ if $r<s_\varepsilon/2$. And we go through the above proof to deduce finally

$$\liminf\{u(x) \mid d(x)\to 0_+\} \geqq K_0+K_0-\varepsilon=2K_0-\varepsilon$$

for all $\varepsilon>0$: indeed, the limit functions u_0' now satisfy the boundary conditions $u_0'=K_0-\varepsilon$ on $\partial B(0,1)$ i.e. $u_0'=u_0+K_0-\varepsilon$. Letting ε go to 0 and iterating the above argument, Theorem III.1 is proved. □

Remark III.3. Considering $w_{\varepsilon,\delta}(x)=-\varepsilon\log(d(x)+\delta)+\delta\log d(n)-C$, we see that $u\geqq w_{\varepsilon,\delta}$ near $d\Omega$ and this proves Theorem III.1 even if $\beta\geqq 2>\mathrm{p}>1$.

III.2. Uniqueness Results

Theorem III.2. *Let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ satisfy* (18). *Then, there exists a maximum solution of* (1) *in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) which goes to $+\infty$ on $\partial\Omega$ and any $v\in L^1_{\mathrm{loc}}(\Omega)$ satisfying* (21) *satisfies $v\leqq u$ a.e. in Ω. Among all solutions of* (1) *in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) which go to $+\infty$ on $\partial\Omega$, or equivalently that are bounded from below on Ω, there exists a minimum one which is the increasing limit of sequence of subsolutions of* (1) *(i.e. satisfying* (21)*) in $W^{2,r}(\Omega)$ ($\forall r<\infty$).*

If we impose further restrictions on f, when we have the

Theorem III.3. *Let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ satisfy* (18′). *Then, there exists a unique solution of* (1) *in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) bounded from below. In addition, this solution satisfies* (19).

Proof of Theorem III.2. Let $C>0$ be a constant such that

$$f(x)\geqq Cd(x)^{-\beta}-C,$$

where $\beta=\max(p,q)$. Then, we set $\underline{w}_\delta=-M\,\mathrm{Log}(d+\delta)-K$ if $p\geqq 2\geqq q$, $\underline{w}=M(d+\delta)^{-\alpha}-K$ if $p<2<q$, where $\alpha=(q-p)/p$, M, K are positive constants

chosen in such a way that for δ small enough $\underline{w}_\delta$ is a subsolution of (1). In fact, we may find $R(\delta)\downarrow +\infty$ as $\delta\uparrow 0_+$ such that (with $a\wedge b=\inf(a,b)$):

$$+\Delta\underline{w}_\delta+|\nabla\underline{w}_\delta|^p+\lambda\underline{w}_\delta\leqq f\wedge R(\delta) \quad \text{in} \quad \Omega\,.$$

Then, using the existence results of Lions [16, 19] we deduce that there exists $\underline{u}_\delta\in W^{2,r}(\Omega)$ $(\forall r<\infty)$ solution of

$$-\Delta\underline{u}_\delta+|\nabla\underline{u}_\delta|^p+\lambda\underline{u}_\delta=f\wedge R(\delta) \quad \text{in} \quad \Omega\,, \qquad \underline{u}_\delta=\underline{w}_\delta \quad \text{on} \quad \partial\Omega\,;$$

and by the maximum principle $\underline{u}_\delta\geqq\underline{w}_\delta$ in Ω.

The remainder of the proof consists in passing to the limit as δ goes to 0 in order to build the minimum solution. To do so we need local upper bounds on $\underline{u}_\delta$: we will achieve this by building a supersolution. We first observe that it is possible to find $\Phi\in C^1(0,\infty)$ such that $\Phi(t)\to+\infty$ as $t\to 0_+$, $\Phi'(t)<0$ if $t>0$, $\Phi(t)>0$ if $t>0$ and

$$(\Phi^{-1/q})'\to 0 \quad \text{as} \quad t\to 0_+\,, \qquad f(x)\leqq\Phi(d(x)) \qquad \text{a.e. in } \Omega\,.$$

Now let $R=\sup_{\bar\Omega} d$, $C_0=\sup_{[0,R]}(\Phi^{-1/q})'$. We denote by

$$\Psi_1=\mu\Phi^{1/p}\,, \qquad \Psi(t)=\int_t^R \Psi_1(s)ds\,,$$

where μ is a positive constant to be determined. We finally set

$$\bar w(x)=\Psi(d)+K$$

where K is a positive constant to be determined. We claim next that for large μ and K, $\bar w$ is a supersolution of (1) which of course blows up at $\partial\Omega$. Indeed, we find, denoting by $C=\|\Delta d\|_\infty$, that if $d(x)\leqq\delta_0$

$$\begin{aligned}
-\Delta\bar w+|\nabla\bar w|^p+\lambda\bar w &\geqq -\Psi''(d)-C|\Psi'(d)|+|\Psi'(d)|^p\\
&=\frac{\mu}{p}\,\Phi^{\frac{1}{p}-1}\Phi'-C\mu\Phi^{1/p}+\mu^p\Phi\\
&\geqq \mu^p\Phi-C\mu\Phi^{1/p}-\mu\,\frac{C_0}{p}\,\Phi^{\frac{1}{p}-1}\,\Phi^{\frac{1}{q}+1}\\
&=\left(\mu^p-\mu\,\frac{C_0}{p}\right)\Phi-C\mu\Phi^{1/p}\geqq f,
\end{aligned}$$

if μ is large enough, say $\mu\geqq\mu_0>0$. We then fix $\mu=\mu_0$ and we consider on the set $d(x)>\delta_0$

$$-\Delta\bar w+|\nabla\bar w|^p+\lambda\bar w\geqq -M+\lambda K$$

for some constant M, and choosing $K\geqq\frac{1}{\lambda}\left(M+\sup_{\Omega_{\delta_0}}|f|\right)$ we conclude.

In particular, we see that $\underline{u}_\delta\leqq\bar w$ and thus $\underline{u}_\delta$ is bounded in $L^\infty_{\text{loc}}(\Omega)$. Furthermore, by the bounds proved in the appendix, this implies that $\underline{u}_\delta$ is also bounded in $W^{1,\infty}_{\text{loc}}(\Omega)$ and thus in $W^{2,r}_{\text{loc}}(\Omega)$ by elliptic regularity. And, letting δ go to 0, $\underline{u}_\delta$ increases to a solution of (1) $\underline{u}$ which is above $\underline{w}$. The fact that $\underline{u}$ is the minimum solution of (1) which goes to $+\infty$ on $\partial\Omega$ is an easy consequence of the fact that any such solution is above $\underline{u}_\delta$ by the maximum principle.

To prove the existence of a maximum solution of (1) going to $+\infty$ on $\partial\Omega$, we first observe that $\bar{w}_\delta=\Psi(d(x)-\delta)+K$ is also a supersolution of (1) with Ω replaced by Ω_δ. Therefore, by maximum principle, any solution of (1) is below $\bar{w}_\delta$ and, passing to the limit in δ, thus below $\bar{w}$.

To build the maximum solution, several arguments are possible. One way to do it consists in maximizing $u(x_0)$ for some fixed $x_0\in\Omega$ among all solutions of (1) bounded from below on Ω (or equivalently going to $+\infty$ on $\partial\Omega$). Then, observe that if u_1, u_2 are two such solutions then there exists another one, say u_3, above u_1 and u_2: indeed $\max(u_1,u_2)$ is a subsolution of (1) and we may solve for

$$-\Delta u_3^\delta+|\nabla u_3^\delta|^p+\lambda u_3^\delta=f \quad \text{in} \quad \Omega_\delta, \qquad u_3^\delta=\max(u_1,u_2) \quad \text{on} \quad \partial\Omega_\delta$$

the existence follows from [19]. Then $u_3^\delta\leqq\bar{w}_\delta$ and thus is bounded in $W^{2;r}_{\mathrm{loc}}(\Omega)$ by arguments we already made several times. Using several times the maximum principle, we see that u_3^δ converges (and increases) to a solution u_3 of (1) which is above u_1 and u_2. This observation implies that there exists a maximizing sequence (u_n) of solutions of (1) which maximizes $u_n(x_0)$ and which is nondecreasing. Then, since $u_n\leqq\bar{w}$, u_n converges (use again the a priori estimates) to a solution $\bar{u}$ of (1) which is bounded from below on Ω and thus blows up at $\partial\Omega$. Furthermore, the above construction of u_3 shows that the fact that $\bar{u}$ maximizes $u(x_0)$ among all solutions implies in fact that $\bar{u}$ is the maximum solution of (1).

Proof of Theorem III.3. Using the results of Theorem III.2 and their proofs, it is now easy to mimick the proofs of Theorems II.1–II.2 in order to obtain the uniqueness. Indeed, if we use (18′), we may replace the functions $\bar{w}_\delta$, $\underline{w}_\delta$ built above by the ones given by (22) provided one takes the values for C_0, α which are given in Theorem I.3. Then, this implies that, by the same proof as above, the minimum solution $\underline{u}$ and the maximum solution $\bar{u}$ of (1) going to $+\infty$ on $\partial\Omega$ satisfy

$$(C_0-\varepsilon)d(x)^{-\alpha}-C_\varepsilon\leqq\underline{u}(x)\leqq\bar{u}(x)\leqq(C_0+\varepsilon)d(x)^{-\alpha}+C_\varepsilon \quad \text{in} \quad \Omega$$

and we may now conclude using the same proof as in Theorem I.1. □

We now conclude this section with an improved uniqueness result where however no precise behaviour of the solution is given.

Theorem III.4. *Let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ satisfy*

$$C'd^{-\beta}-C'\leqq f\leqq Cd^{-\beta}+C \quad \text{for some} \quad C\geqq C'>0, \qquad \beta\geqq\max(p,q). \tag{40}$$

Then, there exists a unique solution of (1) *in $W^{2;r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) which is bounded from below. Denoting by u this solution, we have for some $M\geqq1$*

$$\frac{1}{M}d^{-\alpha}-M\leqq u\leqq Md^{-\alpha}+M \quad \textit{in} \quad \Omega,$$

where $\alpha=\dfrac{\beta}{p}-1$ if $\beta>p$, and $d^{-\alpha}$ is replaced by $|\mathrm{Log}\,d|$ if $\beta=p\geqq q$.

Proof. By similar arguments to the ones given above, the maximum solution $\bar{u}$ and the minimum solution $\underline{u}$ satisfy for some $M\geqq1$

$$\frac{1}{M}d^{-\alpha}-M\leqq u\leqq Md^{-\alpha}+M \quad \text{in} \quad \Omega.$$

Without loss of generality (adding a large constant to f, $\underline{u}$, $\bar{u}$) we may assume that $\bar{u} \geqq \underline{u} \geqq 1$, $f \geqq 1$ a.e. in Ω. Therefore, there exists $\theta \in (0,1)$ small enough such that $\underline{u} \geqq \theta \bar{u}$ in Ω. Let then $\theta_0 = \sup\{\theta \in (0,1]/\underline{u} \geqq \theta \bar{u}$ in $\Omega\}$ – we follow a uniqueness argument which was introduced in a different context by Laetsch [14]. If $\theta_0 = 1$, we are done. We thus argue by contradiction and assume that $\theta_0 < 1$. Of course, we have $\underline{u} \geqq \theta_0 \bar{u}$ in Ω. We then consider $z = \varepsilon d^{-\alpha}$ and we observe that z satisfies

$$-\Delta z + |\nabla z|^p + \lambda z \leqq \varepsilon^p d^{-\beta} + C_\varepsilon d^{-\beta+1}$$

and this is less than f for ε small enough say $\varepsilon \leqq \varepsilon_0$. We choose $\varepsilon = \varepsilon_0$. In fact $z_\delta = \varepsilon(d+\delta)^{-\alpha}$ also satisfies

$$-\Delta z_\delta + |\nabla z_\delta|^p + \lambda z_\delta \leqq f \quad \text{in} \quad \Omega .$$

And we consider $w_{\gamma,\delta} = (\theta_0 - \gamma)\bar{u} + (1 - \theta_0 + \gamma) z_\delta$; $w_{\gamma,\delta}$ satisfies for $\gamma < \theta_0$

$$-\Delta w_{\gamma,\delta} + |\nabla w_{\gamma,\delta}|^p + \lambda w_{\gamma,\delta} \leqq (\theta_0 - \gamma) f + (1 - \theta_0 + \gamma) f \equiv f \quad \text{in} \quad \Omega$$

and since $\underline{u}$, $\bar{u}$ blow up near the boundary we have $w_{\gamma,\delta} \leqq \theta_0 \bar{u} \leqq \underline{u}$ near the boundary. Therefore, by the maximum principle, $w_{\gamma,\delta} \leqq \underline{u}$ in Ω. We now let γ go to 0_+ and then δ go to 0_+ to find

$$\theta_0 \bar{u} + (1 - \theta_0) z \leqq \underline{u} \quad \text{in} \quad \Omega$$

but we obviously have $z \geqq v \bar{u}$ for some $v > 0$. Hence,

$$(\theta_0 + (1 - \theta_0) v) \bar{u} \leqq \underline{u} \quad \text{in} \quad \Omega$$

and this contradicts the definition of θ_0. □

IV. Superquadratic Hamiltonians

IV.1. Interior Gradient Bounds and Maximum Solutions

We begin with a result which gives interior gradient bounds for solutions of (1): similar bounds were first derived in [16, 19] and the proofs are recalled in the appendix. We only remark here that a sharper form of these bounds may be obtained by a simple scaling argument.

Theorem IV.1. *Let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below on Ω and satisfy*

$$|f(x)| \leqq C_1 d(x)^{-\beta} \quad \textit{for some} \quad \beta \geqq 0, \qquad C_1 \geqq 0. \tag{41}$$

Let $u \in W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r < \infty$) be a solution of (1) satisfying

$$\lambda u \geqq -C_2 \quad \textit{for some} \quad C_2 \geqq 0. \tag{42}$$

Then, we set $\gamma = \dfrac{1}{p-1}$ if $\beta \leqq q$, γ arbitrary in $\left(\dfrac{\beta}{p}, 1\right)$ if $\beta > q$ and $\gamma = \dfrac{\beta}{p}$ if $f \in W^{1,\infty}_{\mathrm{loc}}(\Omega)$ and $|\nabla f(x)| d(x)^{-\beta-1} \in L^\infty(\Omega)$. With these notations and assumptions we have

$$|\nabla u(x)| \leqq C_3 d(x)^{-\gamma} \quad \textit{in} \quad \Omega, \quad \square \tag{43}$$

where C_3 only depends on C_1, C_2, γ, β and the diameter of Ω.

Remark IV.1. The bound is optimal as it may be easily checked on simple examples like $\frac{C_0}{d^\alpha}$ if $p \leqq 2$ ($-\operatorname{Log} d$ if $p=2$) with C_0, α given as in Theorem I.1 or Theorem I.3, or $-C_0 d^\alpha$ if $p>2$ with $\alpha=\frac{p-2}{p-1}$ if $\beta \leqq q$, $\alpha=1-\beta/p$ if $\beta<p$ and C_0 is a convenient positive constant. □

Exactly as in [19], this implies of course the following result

Corollary IV.1. *Let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below on Ω and satisfy* (41). *Then, any solution $u \in W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) of* (1) *which is bounded from below belongs to $W^{1,s}(\Omega)$ with $s<p-1$ if $p>2$ and $\beta \leqq q$, $s<p/\beta$ if $p>\beta>q$ (and thus $p>2$). In addition, any such solution may be extended continuously on $\bar{\Omega}$ and $u \in C^{0,\theta}(\bar{\Omega})$ with $\theta=(p-2)/(p-1)$ if $p>2$, $\beta \leqq q$; $\theta=1-\beta'/p$ if $p>\beta>\beta' q$; and $\theta=1-\beta/p$ if $p>\beta>q$ and $f \in W^{1,\infty}_{\mathrm{loc}}(\Omega)$ satisfies $|\nabla f| d^{-\beta-1} \in L^\infty(\Omega)$.* □

We now just sketch the *proof of Theorem IV.1*: let $x_0 \in \Omega$, set $r=\frac{1}{2}d(x_0)$ and consider $v(x)=r^{-(1-\gamma)}u(x_0+rx)$ for $x \in B(0,1)$. One checks easily that v solves

$$-r^\sigma \Delta v+|\nabla v|^p+\lambda r^\nu u=r^{p\gamma} f(x_0+rx) \quad \text{in} \quad B(0,1) \tag{44}$$

with $\sigma=(p-1)\gamma-1$, $\nu=(p-1)\gamma+1$. Next, observe that

$$|r^{p\gamma} f(x_0+rx)| \leqq C_4 \quad \text{on} \quad B(0,1),$$

where C_4 depends only on C_R and β. And if $\beta=q$, then $\sigma=0$, $\nu=2$ while if $\beta>q$, $\nu=\sigma+2$ and $\sigma>0$. If $\beta \leqq q$ or if $\beta>q$ and $d \in W^{1,\infty}_{\mathrm{loc}}$, $|\nabla f| d^{-\beta-1} \in L^\infty(\Omega)$, interior estimates are available (see appendix) and we deduce from this

$$|\nabla v(0)| \leqq C_3$$

which of course yields (43). □

In the last case, we observe that

$$\|\Delta v\|_{L^m(0,0,k))} \leqq \frac{C'(m,k)}{r^\sigma} \quad \text{for all} \quad m \geqq 1, \quad k \in (0,1).$$

But then, recalling the following "standard" inequality for all $m>N$

$$\|\nabla v\|_{L^\infty(B(0,\frac{1}{2}))} \leqq C \|\nabla v\|^{1-\frac{N}{m}}_{L^m(B(0,\frac{1}{2}))} \{\|\Delta v\|_{L^m(B(0,\frac{1}{2}))}+\|\nabla v\|_{L^m(B(0,\frac{1}{2}))}\}^{\frac{N}{m}}$$

we finally obtain

$$|\nabla v(0)| \leqq C(m) r^{-N\sigma/m} \quad \text{for all} \quad m>N.$$

And this yields (43). □

Next, using these estimates and Corollary IV.1, we may now deduce easily the following

Corollary IV.2. *Let $p>2$, let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below on Ω and satisfy* (41) *with $\beta<p$. Then, there exist solutions $\underline{u}, \bar{u}$ of* (1) *in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) bounded from below such that if v is a solution of* (1) *in $W^{2,r}(\Omega)$ ($\forall r<\infty$), respectively $W^{2,r}_{\mathrm{loc}}(\Omega)$*

($\forall r<\infty$), then $\underline{u}\geqq v$ in Ω, respectively $\bar{u}\geqq v$ in Ω. Furthermore, if $v\in L^1_{\mathrm{loc}}(\Omega)$ satisfies (21) *then $v\leqq\bar{u}$ a.e. in Ω. And if $\bar{u}_\delta$, $\underline{u}_\delta$ denote the corresponding maximum solutions of* (1) *with Ω replaced by Ω_δ then*

$$\bar{u}_{\delta'}\geqq\underline{u}_{\delta'}\geqq\bar{u}_\delta\geqq\underline{u}_\delta\geqq\bar{u}\geqq\underline{u} \quad in \quad \Omega_{\delta'} \qquad for \quad 0<\delta<\delta' \tag{45}$$

and $\bar{u}_\delta$ decreases to $\bar{u}$ as δ goes to 0_+. □

Remark IV.2. A consequence of the results we will prove in the following sections is the following: assume that $f\in L^\infty(\mathbb{R}^N)$ and denote by $\bar{u}^\delta$, $\underline{u}^\delta$ the corresponding maximum solution of (1) with Ω replaced by Ω^δ then

$$\underline{u}_{\delta'}\leqq\bar{u}_{\delta'}\leqq\underline{u}_\delta\leqq\bar{u}_\delta\leqq\underline{u} \quad \text{in} \quad \Omega \quad \text{for} \quad 0<\delta<\delta' \tag{46}$$

and $\underline{u}_\delta$ increases to $\underline{u}$ as δ goes to 0_+.

Remark IV.3. We will show in Sect. V that if f behaves like $C_1 d^{-\beta}$ near the boundary with $0\leqq\beta<p$ [$\beta=0$ means $f\in L^\infty(\Omega)$] then $\underline{u}=\bar{u}$ in Ω.

Proof of Corollary IV.2. The existence of the maximum solutions $\underline{u}$, $\bar{u}$ is exactly the same as in Theorem III.2. Next, the string of inequalities in (45) follows from the definitions of $\bar{u}$, $\underline{u}$. Finally, $\bar{u}_\delta$ decreases to a solution of (1) in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) in view of the estimates given by Theorem IV.1. Therefore, the limit is below $\bar{u}$. Since on the other hand, by (45), $\bar{u}_\delta\geqq\bar{u}$, we conclude easily. □

We conclude this section by a property of $\bar{u}$, $\underline{u}$ which will be useful later on.

Proposition IV.1. *Let ω be a bounded smooth domain such that $\bar{\omega}\subset\bar{\Omega}$. Let $v\in W^{2,r}(\omega)$ ($\forall r<\infty$) (resp. $W^{2,r}_{\mathrm{loc}}(\omega)\cap C(\bar{\omega})$ ($\forall r<\infty$)) be a subsolution of* (1) *with Ω replaced by ω. If $v\leqq\underline{u}$ (resp. $v\leqq\bar{u}$) on $\partial\omega\cap\Omega$ then $v\leqq\underline{u}$ (resp. $v\leqq\bar{u}$) in ω.* □

Proof. Let $\varepsilon>0$, $v_\varepsilon=v-\varepsilon$ satisfies the same properties than v. In one case, we just consider

$$w_\varepsilon=\bar{u} \quad \text{in} \quad \Omega-\omega, \qquad =\max(\bar{u},v-\varepsilon) \quad \text{in} \quad \omega$$

and we observe that w_ε is a subsolution of (1) [in $W^{1,\infty}_{\mathrm{loc}}(\Omega)$]. Therefore, by Corollary IV.2, $w_\varepsilon\leqq\bar{u}$ and thus $v\leqq\bar{u}$ in ω by letting ε go to 0.

In the other case, the above construction has to be modified a bit since w_ε does not belong to $W^{2,r}(\Omega)$ ($\forall r<\infty$). We then consider $\beta_\varepsilon(t)=\varepsilon\beta\left(\frac{t}{\varepsilon}\right)$ where $\beta(t)=t$ if $t\geqq 0$, $\beta\in C^\infty(\mathbb{R})$, β is convex, $1\geqq\beta'(t)\geqq 0$ on $\mathbb{R}$, $\beta(t)\equiv -1$ if $t\leqq -2$. And we now introduce

$$z_\varepsilon=\bar{u} \quad \text{in} \quad \Omega-\omega, \qquad =\bar{u}+\beta_\varepsilon(v_\varepsilon-\bar{u}) \quad \text{in} \quad \omega.$$

Now, $z_\varepsilon\in W^{2,r}(\Omega)$ ($\forall r<\infty$) and we claim that z_ε is a subsolution of (1). We only have to check this claim inside ω where we find

$$\begin{aligned}\nabla z_\varepsilon&=\beta'_\varepsilon\nabla v_\varepsilon+(1-\beta'_\varepsilon)\nabla\bar{u},\\ -\Delta z_\varepsilon&=\beta'_\varepsilon(-\Delta v_\varepsilon)+(1-\beta'_\varepsilon)(-\Delta\bar{u})-\beta''_\varepsilon|\nabla v_\varepsilon-\nabla\bar{u}|^2\\ &\leqq\beta'_\varepsilon(-\Delta v_\varepsilon)+(1-\beta'_\varepsilon)(-\Delta\bar{u}),\\ z_\varepsilon&\leqq\bar{u}+\beta'_\varepsilon(v_\varepsilon-\bar{u})\end{aligned}$$

and our claim follows easily from these inequalities.

We may now complete the proof of Proposition IV.1 since, by definition, $z_\varepsilon \leqq \underline{u}$ in Ω and letting ε go to 0, remarking that β_ε converges uniformly to t^+, we obtain $v \leqq \underline{u}$ in ω. $\square$

Remark IV.4. The expert reader will notice that this result is one form of the dynamic programming principle for the associated stochastic control problem!

IV.2. An Estimate on the Boundary Behaviour

We want to show in this section some properties of $\bar{u}$, $\underline{u}$ like (17). We will be always dealing with the case $p > 2$, $f \in L^\infty_{\text{loc}}(\Omega)$ bounded from below and satisfying (41) with $\beta < p$. Hence, Corollary IV.2 and Proposition IV.1 apply. In all the results which follow in this section and in Sect. IV.3, we will not recall these assumptions.

Theorem IV.2. *The maximum solutions* $\bar{u}$, $\underline{u}$ *satisfy* (17) *with* $\alpha = -\dfrac{p-2}{p-1}$. *In addition, if* f *satisfies*

$$\liminf \{ f(x) d(x)^\theta \mid d(x) \to 0_+ \} > 0 \quad \textit{for some} \quad \theta \in (q, \beta] \tag{47}$$

then $\bar{u}$, $\underline{u}$ *satisfy* (17) *with* $\alpha = 1 - \theta/p$. $\square$

Remark IV.5. Again, this result is rather optimal since if f satisfies (41) with $\beta \leqq q$, we already know that $u \in C^{0,\alpha}(\bar{\Omega})$ and $-C_0 d^\alpha$ gives a simple example (for the ad hoc $C_0 > 0$) which shows the sharpness of (17). Similarly if $f(x)$ behaves like $C_1 d(x)^{-\beta}$ for some $q < \beta < p$ then we already know that $u \in C^{0,\alpha}(\bar{\Omega})$ and again $-C_0 d^\alpha$ shows the sharpness of (17). The only improvement we could think of would be to show (and we were unable to do it)

$$\liminf_{y \in \Omega, y \to x} \{ u(y) - u(x) \} |y - x|^{-\alpha} = -C_0, \quad \text{for all} \quad x \in \partial\Omega,$$

where $C_0 = (p-2)^{-1}(p-1)^{\frac{p-2}{p-1}}$ if $\beta < q$, solves $C_0^p \alpha^p - C_0 \alpha(1-\alpha) = C_1$ if $\beta = q$, $C_0 = \dfrac{1}{\alpha} C_1^{1/p}$ if $q < \beta < p$ at least when f behaves like $C_1 d^{-\beta}$ near the boundary.

Proof of Theorem IV.2. The proof is rather delicate so we will begin with a simpler claim than (17). But let us first give the idea of the proof: we just observe that (17) is equivalent to say that for all $x_0 \in \partial\Omega$, $u\,(=\underline{u}, \bar{u})$ $-\varepsilon|x - x_0|^\alpha$ cannot have a local minimum in $\bar{\Omega}$ at x_0 for ε small enough. To prove this fact, we will argue by contradiction and we will do so by building a subsolution on a neighbourhood of x_0 such that on the boundary of the neighbourhood it is below u while it is above u at x_0. This will contradict Proposition IV.1 proving thus our claim.

To explain how this strategy works, we will begin proving that if $\varphi \in C^{1,1}(\bar{\Omega})$ then $u - \varphi$ cannot have a local minimum on $\bar{\Omega}$ at $x_0 \in \partial\Omega$ where $u = \underline{u}$ or $\bar{u}$. Assume by way of contradiction that x_0 is a local minimum of $u - \varphi$ for some $\varphi \in C^{1,1}(\bar{\Omega})$. Then, denoting by $\xi_0 = \nabla\varphi(x_0)$, there exists $C \geqq 0$ such that

$$u(x) \geqq u(x_0) + (\xi_0, x - x_0) - C|x - x_0|^2 \quad \text{for all} \quad x \in \bar{\Omega}. \tag{48}$$

We then consider the following function defined on $\bar{\omega}$ where $\omega=\{x\in\Omega,\ d(x)<\delta\}$ where $\delta>0$ will be determined later on

$$w(x)=u(x_0)+(\xi_0,x-x_0)-C|x-x_0|^2+\mu(\delta^\alpha-d^\alpha),\qquad \forall x\in\bar{\omega} \tag{49}$$

with $\alpha=\dfrac{p-2}{p-1}$, for some $\mu>0$ to be determined. In view of (48) and (49), we have

$$w\leqq u \quad \text{on} \quad \partial\omega\cap\Omega,\qquad w(x_0)>u(x_0). \tag{50}$$

Hence, Proposition IV.1 will yield the desired contradiction if we show that w is a subsolution of (1) in ω. Therefore, we compute in ω

$$\begin{aligned}-\Delta w+|\nabla w|^p+\lambda w-f&=2NC+\alpha\mu d^{\alpha-1}\Delta d-\mu\alpha(1-\alpha)\frac{1}{d^{2-\alpha}}\\&\quad+\left|\xi_0-C(x-x_0)-\alpha\mu\frac{\nabla d}{d^{1-\alpha}}\right|^p+\lambda w-f\\&\leqq C\left(1+\frac{1}{d^{1-\alpha}}\right)-\mu\alpha(1-\alpha)\frac{1}{d^{2-\alpha}}+\left|\xi_0-C(x-x_0)-\alpha\mu\frac{\nabla d}{d^{1-\alpha}}\right|^p,\end{aligned}$$

where C denotes various constants independent of δ. Recalling that $|\nabla d|=1$, $(1-\alpha)p=2-\alpha$, we see that if δ is small enough and $(\alpha\mu)^{p-1}<1-\alpha$ [depending only on μ, $|\xi_0|$, C in (49), a lower bound on f and Ω] w is a subsolution of (1) in ω.

We now show (17): it is enough to show that the following inequality cannot hold

$$u(x)\geqq u(x_0)-\varepsilon_0|x-x_0|^\alpha-C|x-x_0|^2,\qquad \forall x\in\bar{\Omega} \tag{51}$$

for small ε_0, $\delta>0$ and for some $C\geqq 0$, where $\alpha=\dfrac{p-2}{p-1}$ or $\alpha=1-\theta/p$ if f satisfies (47). Indeed, if (51) holds, then we introduce

$$w(x)=u(x_0)-\varepsilon_0\beta_\varepsilon(|x-x_0|)-C|x-x_0|^2+\mu(\delta^\alpha-d^\alpha)\quad \text{in} \quad \omega$$

where $\omega=\{x\in\Omega/d(x)<\delta\}$, $\beta_\varepsilon(t)$ is the function defined by

$$\beta_\varepsilon(t)=\frac{\alpha}{2}\frac{|t|^2}{\varepsilon}+\frac{2-\alpha}{2}\varepsilon^{\frac{\alpha}{2-\alpha}}\quad \text{if}\quad |t|\leqq\varepsilon^{\frac{1}{2-\alpha}},\qquad =|t|^\alpha\quad \text{if}\quad |t|\geqq\varepsilon^{\frac{1}{2-\alpha}}.$$

In view of (51), (50) will hold if

$$\mu\delta^\alpha>\frac{2-\alpha}{2}\varepsilon_0\varepsilon^{\frac{\alpha}{2-\alpha}}. \tag{52}$$

Next, we compute for x in ω the following quantity

$$\begin{aligned}-\Delta w+|\nabla w|^p+\lambda w-f&=2NC+(N-1)\varepsilon_0\beta'\frac{1}{|x-x_0|}+\varepsilon_0\beta''-\mu\alpha(1-\alpha)\frac{1}{d^{2-\alpha}}\\&\quad+\alpha\mu\frac{1}{d^{1-\alpha}}\Delta d+\left|-\varepsilon_0\beta'\frac{x-x_0}{|x-x_0|}-2C(x-x_0)-\alpha\mu\frac{\nabla d}{d^{1-\alpha}}\right|^p+\lambda w-f\end{aligned}$$

(in fact this equality holds a.e. in ω), and this yields

$$-\Delta w+|\nabla w|^p+\lambda w-f\leqq C-f+\varepsilon_0 N\frac{\alpha}{\varepsilon}+C\mu\frac{1}{d^{1-\alpha}}-\mu\alpha(1-\alpha)\frac{1}{d^{2-\alpha}}$$

$$+\left|\varepsilon_0\beta'\frac{x-x_0}{|x-x_0|}+2C(x-x_0)+\alpha\mu\frac{\nabla d}{d^{1-\alpha}}\right|^p.$$

Now, if we begin by the case where we do not assume (47), then we just bound f by a constant C and we deduce

$$-\Delta w+|\nabla w|^p+\lambda w-f\leqq C+\varepsilon_0 N\frac{\alpha}{\varepsilon}+C\mu\frac{1}{d^{1-\alpha}}+\mu\alpha(1-\alpha)\frac{1}{d^{2-\alpha}}$$

$$+\left(\varepsilon_0\alpha\frac{1}{\varepsilon^{\frac{1-\alpha}{2-\alpha}}}+C+\alpha\mu\frac{1}{d^{1-\alpha}}\right)^p.$$

We then choose $\varepsilon=(t\mu\delta^\alpha\varepsilon_0^{-1})^{\frac{2-\alpha}{\alpha}}$ with $0<t<\dfrac{2}{2-\alpha}$ so that (52) holds and we obtain

$$-\Delta w+|\nabla w|^p+\lambda w-f\leqq C+C\mu\frac{1}{d^{1-\alpha}}+Nt^{-\frac{2-\alpha}{\alpha}}\alpha\mu^{-\frac{2-\alpha}{\alpha}}\delta^{-(2-\alpha)}\varepsilon_0^{2/\alpha}$$

$$+-\mu\alpha(1-\alpha)\frac{1}{d^{2-\alpha}}+\left(\alpha t^{-\frac{1-\alpha}{\alpha}}\delta^{-(1-\alpha)}\mu^{-\frac{1-\alpha}{\alpha}}\varepsilon_0^{1/\alpha}+C+\alpha\mu\frac{1}{d^{1-\alpha}}\right)^p.$$

Next, if we fix t in $\left(0,\dfrac{2}{2-\alpha}\right)$ and μ in $\left(0,(t\alpha)^{\frac{1}{p-1}}\alpha^{-1}\right)$, recalling that $d(x)<\delta$, we see that for ε_0 small enough (depending only on N, t, μ, α) we may bound the above terms by

$$C+C\mu\frac{1}{d^{1-\alpha}}-K\frac{1}{d^{2-\alpha}}$$

for some $K>0$, and then we conclude choosing δ small enough.

In the other case, that is when we assume (47), we obtain

$$-\Delta w+|\nabla w|^p+\lambda w-f\leqq C-\frac{v}{d^\theta}+N\varepsilon_0\frac{\alpha}{\varepsilon}+\left(\varepsilon_0\alpha\frac{1}{\varepsilon^{\frac{1-\alpha}{2-\alpha}}}+C+\alpha\mu\frac{1}{d^{1-\alpha}}\right)^p$$

$$\leqq C-\frac{v}{d^\theta}+N\varepsilon_0\frac{\alpha}{\varepsilon}+\left(C+\alpha\varepsilon_0\varepsilon^{\frac{\theta}{\theta+p}}+\alpha\mu d^{-\frac{\theta}{p}}\right)^p$$

and again writing $\varepsilon=(t\mu\delta^\alpha\varepsilon_0^{-1})^{\frac{2-\alpha}{\alpha}}$ with $0<t<\dfrac{2}{2-\alpha}$ so that (52) holds we deduce

$$-\Delta w+|\nabla w|^p+\lambda w-f\leqq C-\frac{v}{d^\theta}+N\alpha t^{-\frac{2-\alpha}{\alpha}}\delta^{-(2-\alpha)}\mu^{-\frac{2-\alpha}{\alpha}}\varepsilon_0^{\frac{2}{\alpha}}$$

$$+\left(C+\alpha t^{-\frac{1-\alpha}{\alpha}}\mu^{-\frac{1-\alpha}{\alpha}}\delta^{-\frac{\theta}{p}}\varepsilon_0^{\frac{1}{\alpha}}+\alpha\mu d^{-\frac{\theta}{p}}\right)^p.$$

And, if we choose t in $\left(0, \dfrac{2}{2-\alpha}\right)$, μ in $(0, v^{1/p}\alpha^{-1})$ we see that for ε_0 small enough the above terms may be bounded by

$$C - \frac{v}{d^\theta}$$

therefore w is a subsolution in ω for δ small enough and we conclude. □

We, in fact, proved the

Corollary IV.3. *The maximum solutions $\underline{u}$, $\bar{u}$ satisfy for all $x \in \partial\Omega$*

$$\liminf_{x \in \Omega, x \to x_0} \{u(x) - u(x_0)\} |x - x_0|^{-\alpha} \leqq -K < 0 \tag{53}$$

where $K = K(p, N)$ and $\alpha = 1 - 1/(p-1)$.

And if f satisfies (47), *then* (53) *holds with $\alpha = 1 - \dfrac{\theta}{p}$ and $K = K(p, \theta, N, C_1)$ where $C_1 = \liminf \{f(x) d(x)^\theta / d(x) \to 0_+\}$.*

IV.3. Infinite Neumann Conditions

Our goal in this section is to investigate the behaviour of the maximum solutions near the boundary. We suspect that the results given in Proposition IV.3 hold in full generality but we were unable to prove it.

We will first sketch the proof of

Proposition IV.2. *Let $f \in W^{1,\infty}(\Omega)$, $p > 2$.*

i) *If Ω is a ball (or if Ω is an half-space), the maximum solutions of* (1) *are Lipschitz tangentially i.e. if $\Omega = B_R$ then*

$$|u(y) - u(x)| \leqq C|y - x| \quad \forall y, x \in \bar{\Omega} \quad \textit{with} \quad |y| = |x| \tag{54}$$

and if $\Omega = \{x_N > 0\}$ then

$$|u(y) - u(x)| \leqq C|y - x| \quad \forall y, x \in \bar{\Omega} \quad \textit{with} \quad y_N = x_N \tag{55}$$

for some $C \geqq 0$, where $u = \underline{u}$ or $\bar{u}$.

ii) *If Ω is convex, then $u = \underline{u}$ or $\bar{u}$ satisfies*

$$|\nabla u - (\nabla u(x) \cdot n(x)) n(x)| \leqq C d^{-1/2} \quad \textit{in} \quad \Omega \tag{56}$$

for some $C \geqq 0$, where n is any smooth vector-field equal to the unit outward normal near $\partial\Omega$ (i.e. $n = -\nabla d$ near $\partial\Omega$). And if $2 < p < 3$, this yields

$$|u(x) - u(y)| \leqq C|x - y|^{2(p-2)/(3p-5)} \quad \forall y, x \in \bar{\Omega} \quad \textit{with} \quad d(x) = d(y). \tag{57}$$

Remark IV.6. It is proved in Lasry and Lions [15] that if Ω is convex, f is convex ($\in C(\Omega)$) and satisfies (41) then $\underline{u}$ and $\bar{u}$ are convex. In addition, if (41) holds then $\underline{u}, \bar{u} \in C^{0,1-\gamma}(\bar{\Omega})$ with $\gamma = \dfrac{1}{p-1}$ if $\beta \leqq q$ and $\gamma = \dfrac{\beta}{p}$ if $\beta > q$. This Hölder continuity combined with the convexity then implies

$$|\nabla u-(\nabla u\cdot n)n|\leqq Cd^{-\gamma/2} \quad \text{in} \quad \Omega. \tag{58}$$

This improved bound on the tangential gradient enables us, in that particular case, to follow the arguments given below. □

Proof of Proposition IV.2. i) In the case of the half-space, one simply remark that $u(\cdot+he_i)$ (for $1\leqq i\leqq N-1$) is the maximum solution of (1) with f replaced by $f(\cdot+he_i)$ hence, using as in Corollary II.1, the maximality

$$\|u(\cdot+he_i)-u(\cdot)\|_\infty\leqq\frac{1}{\lambda}\|f(\cdot+he_i)-f(\cdot)\|_\infty\leqq C|h|$$

and (55) is proved. One proves (54) similarly replacing the tangential translations by rotations.

ii) Let y be an arbitrary point in $\bar\Omega$, we set $u_t(x)=\frac{1}{t^\alpha}u(y+t(x-y))w$ for $0<t<1$, $x\in y+\frac{1}{t}(\Omega-y)=\Omega_t$ with $\alpha=\frac{p-2}{p-1}$. Observe that $\Omega\subset\Omega_t$ and that u_t solves

$$-\Delta u_t+|\nabla u_t|^p+\lambda t^2u_t=t^{2-\alpha}f(tx) \quad \text{in} \quad \Omega_t.$$

Therefore, we have for some $C\geqq 0$

$$-\Delta u_t+|\nabla u_t|^p+\lambda u_t\leqq f+C(1-t) \quad \text{in} \quad \Omega,$$

and $u_t-\frac{C}{\lambda}(1-t)$ is a subsolution of (1); hence $u_t\leqq u+\frac{C}{\lambda}(1-t)$. But this inequality immediately implies

$$(x-y,\nabla u(x))\geqq -C, \quad \forall(x,y)\in\Omega\times\bar\Omega, \tag{59}$$

which in turn yields (56) and (57). □

The improved Hölder continuity of $\underline{u}$, $\bar u$ in the tangential directions enables us to obtain the

Proposition IV.3. *Let $f\in W^{1,\infty}(\Omega)$, $p>2$. Assume that either Ω is a ball, or Ω is an half-space, or Ω is convex and $p<3$, then $\alpha=\frac{p-2}{p-1}$ the maximum solutions $\underline{u}$, $\bar u$ satisfy*

$$t^{-\alpha}\{u(x_0-tn(x_0))-u(x_0)\}\to -C_0 \quad \textit{as} \quad t\to 0_+, \quad \textit{uniformly in } x_0\in\partial\Omega \tag{60}$$

$$\nabla u(x)d(x)^{1-\alpha}\to C_0\alpha n \quad \textit{as} \quad d(x)\to 0_+, \tag{61}$$

where $C_0=(1-\alpha)^{1/(p-1)}\alpha^{-1}$.

Proof. We just sketch it. Let $x_0\in\partial\Omega$, we introduce the blown-up-functions u_t defined by $u_t(x)=t^{-\alpha}\{u(x_0+tx)-u(x_0)\}$ defined on $Q_t=(\Omega-x_0)/t$. We want to let t go to 0_+. We first observe that by Theorem IV.1 and Corollary IV.1 u_t is bounded in $L^\infty(Q_t\cap B_R)$ ($\forall R<\infty$). In addition, u_t solves

$$-\Delta u_t+|\nabla u_t|^p=t^{2-\alpha}\{f(x_0+tx)-\lambda u(x_0+tx)\} \quad \text{in} \quad Q_t.$$

And we obtain easily a priori bounds from the interior gradient estimates: therefore u_t is relatively compact and any convergent subsequence u_{t_n} converges uniformly on compact sets, as t_n goes to 0, to a solution $v \in W^{2,r}_{\mathrm{loc}}(\Pi)$ $(\forall r < \infty)$ of

$$-\Delta v + |\nabla v|^p = 0 \quad \text{in} \quad \Pi, \qquad v(x_0) = 0, \qquad |v(x)| \leqq C|x|^\alpha \quad \text{in} \quad \Pi,$$

where $\Pi = \{x \in \mathbb{R}^N / n(x_0) \cdot x < 0\}$. In addition, using Proposition IV.1, we deduce that v is above any function $w \in W^{2,r}_{\mathrm{loc}}(\omega) \cap C(\bar{\omega})$ [resp. $W^{2,r}(\omega)$ $\forall r < \infty$ if we are dealing with $\underline{u}$] satisfying

$$-\Delta w + |\nabla w|^p \leqq 0 \quad \text{in} \quad \omega, \qquad w \leqq v \quad \text{on} \quad \partial\omega \cap \Pi.$$

Finally, the estimates (54), (55) or (57) imply that v depends only on the variable $x \cdot n(x_0)$ i.e. $v(x) = \varphi(-x \cdot n(x_0))$ where φ solves

$$-\varphi'' + |\varphi'|^p = 0 \quad \text{for} \quad t > 0, \qquad \varphi(0) = 0, \qquad \varphi \in C([0, \infty)) \cap C^2(0, \infty).$$

Hence, $\varphi(t) \leqq 0$ or $\varphi(t) = C_0 \lambda^\alpha - C_0(t + \lambda)^\alpha$ on $\mathbb{R}_+$ for some $\lambda \geqq 0$. But, since v is "a maximum solution" we deduce that $\varphi \geqq \psi$ on $[0, L]$ for all $L > 0$, $\psi \in C^2([0, L])$ satisfying

$$\psi(L) \leqq \varphi(L), \qquad -\psi'' + |\psi'|^p \leqq 0 \quad \text{in} \quad (0, L).$$

And this implies by Theorem IV.2 and its proof that

$$\liminf_{t \to 0_+} \varphi(t) t^{-\alpha} < 0$$

therefore $\varphi(t) \equiv -C_0 t^\alpha$ and we conclude easily. □

We conjecture that if $p > 2$, $f \in L^\infty(\Omega)$ then (60) and (61) always hold. Of course, in view of the preceding argument, it would be enough to prove that $|u(x) - u(y)| \leqq C|x - y|^\theta$ for some $\theta < \alpha$ if $x, y \in \Omega$, $d(x) = d(y)$ but this type of estimate seems rather difficult to obtain in general.

V. Viscosity Formulation of the Boundary Conditions

V.1. Uniqueness Results

If we accept the stochastic control interpretation of the solutions built in the preceding sections, one is led (see [20] for more details) to the following formulation of the boundary condition

$$\text{for all} \quad \varphi \in C^2(\bar{\Omega}), \qquad u - \varphi \quad \text{achieves its minimum over} \quad \Omega. \tag{62}$$

Or course, an equivalent formulation in the case when $u \in C(\bar{\Omega})$ is to impose that $u - \varphi$ never has a local minimum on $\bar{\Omega}$ at a point $x_0 \in \partial\Omega$ for all $\varphi \in C^2(\bar{\Omega})$. It is quite clear that solutions considered in Sects. II and III satisfy (62), even with $\varphi \in C(\bar{\Omega})$, since they blow up at $\partial\Omega$. Similarly, the maximum solutions built in Sect. IV also satisfy (62), even with $\varphi \in C^{0,\theta}(\bar{\Omega})$ for $\theta > \alpha$, since they satisfy (17): indeed, assume for instance that $\bar{u} - \varphi$ does not achieve its minimum over Ω. Since $u, \varphi \in C(\bar{\Omega})$, there is a minimum point x_0 over $\bar{\Omega}$ of $u - \varphi$ and $x_0 \in \partial\Omega$. Then, we have for $x \in \Omega$

$$u(x) - u(x_0) \geqq \varphi(x) - \varphi(x_0)$$

therefore

$$\liminf_{x\in\Omega, x\to x_0} \{u(x)-u(x_0)\}\,|x-x_0|^{-\alpha} \geqq \liminf_{x\in\Omega, x\to x_0} \{\varphi(x)-\varphi(x_0)\}\,|x-x_0|^{-\alpha}$$

and the right-hand side is 0 since φ is smooth and $\alpha\in(0,1)$. And we reach a contradiction with (17). In other words, any solution of (1) satisfying (17) does satisfy the boundary condition in "viscosity form" given by (62).

Our goal in this section is to prove, under quite general assumptions, that there is a unique solution of (1) satisfying (62). In particular, when this holds, this will imply that with the notations of Sect. IV the maximum solutions $\bar{u}$ and $\underline{u}$ are equal.

We may now state our main result.

Theorem V.I. *There exists a unique solution of* (1) *in* $W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ *satisfying* (62) *under one of the following three sets of assumptions*

i) $1<p\leqq 2$, $f\in L^\infty_{\mathrm{loc}}(\Omega)$ *satisfies* (14) *and is bounded from below.*
ii) $1<p$, $f\in L^\infty_{\mathrm{loc}}(\Omega)$ *satisfies* (40).
iii) $2<p$, $f\in L^\infty_{\mathrm{loc}}(\Omega)$ *satisfies* (41) *with* $\beta<p$ *and is bounded from below.*

Corollary V.1. *Let* $p>2$. *Let* $f\in L^\infty_{\mathrm{loc}}(\Omega)$ *satisfy* (41) *with* $\beta<p$ *and be bounded from below. Then, the maximum solutions built in Sect. IV are equal. Furthermore, if* $f\in C(\bar{\Omega})$ *or if* $f\in C(\Omega)$ *and* $f(x)d(x)^{-\theta}\to C_1$ *as* $d\to 0_+$ *with* $0<\theta<p$, *they coincide also with the envelope of all* $C^2(\bar{\Omega})$ *subsolutions of* (1).

Remark V.1. Actually, (62) can be proved to be equivalent to

$$u-\varphi \text{ achieves its minimum in } \Omega \text{ for all quadratic functions } \varphi. \qquad (62')$$

Indeed, suppose (62′) holds and let $\varphi\in C^2(\bar{\Omega})$. Let $x_n\in\Omega$ be a minimizing sequence for $u-\varphi$ converging to some $x_0\in\bar{\Omega}$. For C large enough we have $\psi(x)<\varphi(x)\ \forall x\neq x_0$, where ψ is defined by $\psi(x)=\varphi(x_0)+\nabla\varphi(x_0)\cdot(x-x_0)-C|x-x_0|^2\ \forall x\in\Omega$. Hence, ψ is quadratic and $u(x)-\psi(x)>u(x)-\varphi(x)=\min(u-\varphi)=\min(u-\psi)$ for all $x\neq x_0$. Hence from (62′) x_0 lies in Ω.

Proof of Corollary V.1. As we already said, $\bar{u}$, $\underline{u}$ satisfy (62) and so are equal by Theorem V.1. In addition, if we denote by $\tilde{u}$ the envelope of all $C^2(\bar{\Omega})$ subsolutions of (1); we first claim that by the same arguments as in Sects. III and IV u is a solution of (1) in $W^{2,r}_{\mathrm{loc}}(\Omega)\cap C(\bar{\Omega})$ $(\forall r<\infty)$. If $f\in C^{0,\gamma}(\bar{\Omega})$ for some $\gamma>0$, the arguments indeed adapt without changes. If $f\in C(\bar{\Omega})$, we just approximate f by $f_n\in C^1(\bar{\Omega})$ such that $f_n\leqq f$, $f_n \underset{n}{\to} f$ uniformly on $\bar{\Omega}$. If $f\in C(\Omega)$ satisfies (63), we first observe that $g(x)=f(x)d(x)^\theta$ may be extended continuously to $\bar{\Omega}$ by giving it the value C_1 on $\partial\Omega$ then we approximate g by $g_n\in C^1(\Omega)$ such that $g_n\leqq g$, $g_n \underset{n}{\to} g$ uniformly on $\bar{\Omega}$, g_n is constant on $\partial\Omega$ and we consider $f_{n,R}=g_n(d(x)^{-\theta}\wedge R)$. And these approximations easily yield our claim on $\tilde{u}$.

Next, we observe that Proposition IV.1 and Theorem IV.2 may be applied or more precisely that their proofs are immediately adapted to the case of $\tilde{u}\leqq 0$ that $\tilde{u}$ satisfies (17). Hence, $\tilde{u}$ satisfies (62) and Corollary V.I is proved. □

V.2. Proofs

We begin with the *proof of Theorem V.1 in the case i)*. We denote by $\bar{u}$ the unique solution of (1) in $W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ which blows up at (see Theorems II.1 and II.2) and we consider another solution u of (1) in $W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ satisfying (62). We obviously have $u \leqq \bar{u}$ by Theorems II.1 and II.2 and we want to show that the reverse inequality also holds. The strategy of proof is quite simple: we just observe that if $v \in C^2(\bar{\Omega})$ is a subsolution of (1) then $u-v$ achieves its minimum over Ω at some point x_0 because of (62) and by the maximum principle we deduce $(u-v)(x_0) \geqq 0$ hence $u \geqq v$. Therefore, if we are able to approximate $\bar{u}$ by $C^2(\bar{\Omega})$ subsolutions of (1), we complete the uniqueness proof. Now, if $f \in C^1(\bar{\Omega})$ such that $f_n \leqq f$, $f_n \geqq -C$ for some C independent of n and f_n converges uniformly to f on compact subsets of Ω. We next denote by $\bar{u}_n$ the corresponding unique solutions of (1) (with f replaced by f_n) which blow up near the boundary (Theorem II.1) and as remarked in Sect. II we know that $\bar{u}_n$ converges uniformly on compact subsets of Ω to $\bar{u}$ and of course $\bar{u}_n \leqq \bar{u}$. Since we know now by the proof of Theorem II.1 that $\bar{u}_n$ is an increasing limit of $C^2(\bar{\Omega})$ (use the smoothness of f_n) solutions of (1) (with f replaced by f_n), the desired sequence of subsolutions of (1) in $C^2(\bar{\Omega})$ is built.

However, this argument does not work as well if we only assume (as we did in Theorem V.1) that $f \in L^\infty_{\mathrm{loc}}(\Omega)$, satisfies (14) and is bounded from below. In this general case, we approximate f by f_n given by

$$f_n = f \quad \text{in} \quad \Omega_{1/n}, \qquad = -C_0 \quad \text{in} \quad \Omega - \Omega_{1/n}, \tag{63}$$

where $C_0 \geqq 0$ is any constant such that $f \geqq -C_0$ in Ω.

Again, we consider the solutions $\bar{u}_n$ of (1) (with f replaced by f_n) which blow up near $\partial\Omega$. We know from the proof of Theorem II.1 that there exists for each $n \geqq 1$, a sequence $(\bar{u}_{n,m})_{m \geqq 1}$ satisfying

$$-\Delta \bar{u}_{n,m} + |\nabla \bar{u}_{n,m}|^p + \lambda \bar{u}_{n,m} = f_n \quad \text{in} \quad \Omega, \qquad \bar{u}_{n,m} \in W^{2,r}(\Omega)\,(\forall r < \infty),$$
$$\bar{u}_{n,m} = m \quad \text{on} \quad \partial\Omega \tag{64}$$

and $\bar{u}_{n,m} \underset{m}{\uparrow} \bar{u}_n$ uniformly on compact subsets of Ω.

Therefore, we obtain: $u \geqq \bar{u}_{n,m}$ and, passing to the limit in m, $u \geqq \bar{u}_n$. And we recall from the results and arguments of Sect. II that $\bar{u}_n$ increases to $\bar{u}$ and thus $u \geqq \bar{u}$, completing the proof of Theorem V.1 in case i).

The proof of case ii) is almost trivial: indeed, we apply (62) with $\varphi \equiv 0$ to deduce that any solution of (1) satisfying (62) is bounded from below on Ω. Thus, by Theorems III.3 and III.4, the uniqueness is proved.

Unfortunately, *the proof of case iii)* is much more complicated; in order to keep the ideas clear (or to try at least) we will begin with the case when $f \in C^1(\bar{\Omega})$ and Ω is starshaped (step 1), then we will treat the case when $f \in C^1(\bar{\Omega})$ but Ω is arbitrary (step 2) and we will conclude with the general case (step 3).

Step 1. $f \in C^1(\bar{\Omega})$, Ω is starshaped.

Without loss of generality we may assume that Ω is starshaped with respect to 0. Again, we denote by $\bar{u}$ the maximum solution of (1) in $W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r < \infty)$ or equivalently in $C^2(\Omega)$ in view of the smoothness of f – see Sect. IV. And we denote by u any other solution of (1) [in $C^2(\Omega)$] satisfying (62). Recall that $\bar{u} \in C(\bar{\Omega})$ (cf. Sect. IV) and observe that applying (62) with $\varphi = 0$, one deduces that u is bounded from below and thus u may be extended continuously to $\bar{\Omega}$ (cf. Sect. IV). Finally, $u \leqq \bar{u}$ and thus we want to show the reverse inequality.

We then introduce for $t \in (0,1)$

$$v_t(x) = t^{-\frac{p-2}{p-1}} u(tx) \quad \text{for} \quad x \in \Omega/t \supset \bar{\Omega}. \tag{65}$$

Obviously, v_t satisfies

$$-\Delta v_t+|\nabla v_t|^p+\lambda t^2 v_t=t^q f(tx) \quad \text{in} \quad \Omega/t, \qquad v_t\in C^2(\Omega/t)\cap C(\bar{\Omega}/t) \tag{66}$$

and thus in particular $v_t\in C^2(\bar{\Omega})$ and satisfies

$$-\Delta v_t+|\nabla v_t|^p+\lambda v_t\leqq f(x)+C(1-t) \quad \text{in} \quad \Omega \tag{67}$$

for some $C\geqq 0$ independent of t. In other words, $v_t-\dfrac{C}{\lambda}(1-t)$ is a $C^2(\bar{\Omega})$ subsolution of (1) and our strategy applies: $u\geqq v_t$ in $\bar{\Omega}$ and thus passing to the limit as t goes to 1 we conclude $u\geqq \bar{u}$ in $\bar{\Omega}$.

Step 2. $f\in C^1(\bar{\Omega})$, Ω arbitrary.

We first observe that by the maximum principle the minimum of $u-\bar{u}$ is achieved at the boundary. Furthermore, if $\theta\in(0,1)$, we may still assume that the minimum of $u-\theta\bar{u}$ is still achieved at the boundary. Indeed, if $u-\theta_n\bar{u}$ has an interior minimum over $\bar{\Omega}$ say at $x_n\in\Omega$ for some sequence $\theta_n \underset{n}{\to} 1$, then observing that $\theta_n\bar{u}$ satisfies

$$-\Delta(\theta_n\bar{u})+|\nabla(\theta_n\bar{u})|^p+\lambda\theta_n\bar{u}\leqq\theta_n f\leqq f+C(1-\theta_n)$$

we deduce from the maximum principle

$$\min_{\bar{\Omega}}(u-\theta_n\bar{u})\geqq-\frac{C}{\lambda}(1-\theta_n)$$

and we conclude letting n to to $+\infty$.

Therefore, let fix $\theta\in(0,1)$, we assume that $u-\theta\bar{u}$ has a minimum over $\bar{\Omega}$ at $x_0\in\partial\Omega$. Then, we remark that $u-\theta\bar{u}+(1-\theta)|x-x_0|^2$ has a *unique* maximum over $\bar{\Omega}$ at $x_0\in\partial\Omega$ and, denoting by $\tilde{u}=\theta\bar{u}+(1-\theta)|x-x_0|^2$, that $\tilde{u}$ satisfies

$$-\Delta\tilde{u}+|\nabla\tilde{u}|^p+\lambda\tilde{u}\leqq\theta f+C(1-\theta)\leqq f+C(1-\theta), \tag{68}$$

where C denotes various nonnegative constants independent of θ.

We next observe that for some small $\delta>0$, the open set $Q=(x_0,\delta)\cap\Omega$ is starshaped with respect to a point that we denote by 0 such that $d(0)\geqq\gamma>0$ where γ,δ are independent of x_0 and θ. We then consider as in step 1 the functions

$$v_t(x)=t^{-\frac{p-2}{p-1}}\tilde{u}(tx) \quad \text{for} \quad x\in Q/t, \qquad t\in(0,1)$$

and we obtain exactly as in step 1 using now (68) instead of (1)

$$-\Delta v_t+|\nabla v_t|^p+\lambda v_t\leqq f+C(1-\theta)+C(1-t) \quad \text{in} \quad Q, \qquad v_t\in C^2(\bar{\Omega}). \tag{69}$$

Let $\bar{x}$ be a minimum point of $u-v_t$ on $\bar{Q}$: because of (62), $\bar{x}\in Q$ or $\bar{x}\in\partial Q\cap\Omega$. If $\bar{x}\in Q$, we use maximum principle to deduce

$$\min_{\bar{\Omega}}(u-v_t)\geqq-\frac{C}{\lambda}(1-\theta)-\frac{C}{\lambda}(1-t)$$

and thus in particular

$$u(x_0)-t^{-\frac{p-2}{p-1}}\theta\bar{u}(x_0)\geqq-\frac{C}{\lambda}(1-\theta)-\frac{C}{\lambda}(1-t)$$

and we deduce letting t go to 1

$$\min_{\bar{\Omega}} (u-\theta\bar{u})=(u-\theta\bar{u})(x_0)\geqq -\frac{C}{\lambda}(1-\theta).$$

The conclusion follows upon letting θ go to 1.

In the other case i.e. if $\bar{x}\in\partial Q\cap\Omega$, we obtain letting t go to 1

$$(u-\tilde{u})(x_0)\geqq \min_{\partial Q\cap\Omega} (u-\tilde{u})$$

and this yields a contradiction with the fact that x_0 is the unique minimum point of $u-\tilde{u}$ over $\bar{\Omega}$.

Step 3. The general case.

We begin by observing that if $f\in C(\bar{\Omega})$ or even if f is continuous near $\partial\Omega$ the above proof is easily adapted: the only difficulty lies with the fact that $\bar{u}$, $\tilde{u}$, v_t do not belong to C^2 in general. But this can be taken care of by observing that $v_t*\varrho_\delta=v_{t,\delta}$ $\left[\text{where } \varrho_\delta=\frac{1}{\delta^N}\varrho\left(\frac{\cdot}{\delta}\right),\ \varrho\geqq 0,\ \varrho\in\mathscr{D}(\mathbb{R}^N),\ \int_{\mathbb{R}^N}\varrho dx=1,\ \operatorname{Supp}\varrho\subset B(0,1)\right]$ satisfies

$$\begin{aligned}&-\Delta v_{t,\delta}+|\nabla v_{t,\delta}|^p+\lambda v_{t,\delta}\\ &\quad\leqq f*\varrho_\delta+C(1-\theta)+\omega(1-t)\quad\text{in}\quad \{x\in Q/t,\ \operatorname{dist}(x,\partial Q/t)>\delta\},\end{aligned}$$

where ω is a modulus of continuity of f near the boundary. Taking δ small enough, we find that $v_{t,\delta}\in C^2(\bar{Q})$ and

$$-\Delta v_{t,\delta}+|\nabla v_{t,\delta}|^p+\lambda v_{t,\delta}\leqq f+C(1-\theta)+\omega(1-t)+\omega(\delta)\quad\text{in}\quad Q$$

and we conclude as before letting δ go to 0, then t go to 1 and then θ go to 1.

To obtain the uniqueness in the case of a general f, we approximate f by f_n given by (63) where $C_0\geqq 0$ is any constant such that $f\geqq -C_0$ in Ω. The above arguments show that $u\geqq u_n$ where u_n is the unique solution of (1) satisfying (62) (with f replaced by f_n). Obviously, $\bar{u}\geqq u_n$ and u_n increases to a solution in $W^{2,r}(\Omega)$ $(\forall r<\infty)\cap C(\bar{\Omega})$ of (1) and we just have to show that $\bar{u}\leqq\hat{u}$, where $\hat{u}$ denotes the limit of u_n. To this end let $\theta\in(0,1)$, let $\gamma\in(\beta,p)$, $\sigma=1-\gamma/p$ and choose $K>0$, $C>0$ so that $w=-C-Kd^\sigma$ satisfies

$$-\Delta w+|\nabla w|^p+\lambda w\leqq -C_0-\nu d^{-\gamma}\quad\text{in}\quad\Omega,\quad\text{for some}\quad \nu>0. \tag{70}$$

Then, we remark that $z=\theta\bar{u}+(1-\theta)w$ satisfies

$$-\Delta z+|\nabla z|^p+\lambda z\leqq\theta f-(1-\theta)(C_0+\nu d^{-\gamma})=g.$$

But on $\Omega_{1/n}$,

$$g=f-(1-\theta)(f+C_0+\nu d^{-\gamma})\leqq f$$

while on $\Omega-\Omega_{1/n}$

$$g\leqq\theta C(1+d^{-\beta})-(1-\theta)(C_0+\nu d^{-\gamma})$$

and thus, $g\leqq f_n$ in Ω provided n is large enough say $n\geqq n_0(\theta)$. Hence,

$$\theta\bar{u}+(1-\theta)w\leqq u_n\quad\text{if}\quad n\geqq n_0(\theta)$$

and passing to the limit in n, we deduce

$$\theta \bar{u} + (1-\theta)u \leqq \hat{u}.$$

We may now conclude letting θ go to 1. □

V.3. Applications

We want in this section to show that (17) is equivalent to (62) when $p>2$ (with appropriate conditions on f) and that (62) is stable under some passages to the limit. This together with the uniqueness proved in Theorem V.1 will yield a rather powerful stability result. We begin with the relations between (17) and (62). Recall that (17) implies trivially (62).

Theorem V.1. *Let $u \in W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$), let $p>2$, let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below and let $x_0 \in \Omega$. Assume that $u \in C(\bar{\Omega})$ satisfies* (62) *and*

$$-\Delta u + |\nabla u|^p + \lambda u \geqq f \quad \text{in} \quad \Omega, \quad \text{for some} \quad \lambda, C \geqq 0. \tag{72}$$

Then, u satisfies

$$\liminf_{x\in\Omega, x\to x_0} \{u(x)-u(x_0)\}\,|x-x_0|^{-\alpha} < 0, \quad \text{where} \quad \alpha = (p-2)/(p-1). \tag{73}$$

Proof. The proof is almost the same as the one of Theorem IV.2: with the notations of Theorem IV.2, we just have to replace ω by w^n defined exactly as w with d replaced by $d+\frac{1}{n}$. Then, $w^n \in C^2(\bar{\omega})$ is a subsolution of (1) and $u \geqq w^n$ on $\partial\omega \cap \Omega$. Therefore, by maximum principle, $u-w^n$ achieves its minimum on $\partial\Omega$ and we reach a contradiction. □

Remark V.1. Many variants of the above result and of its proof exist that we will skip here.

We now present a stability result.

Theorem V.2. *Let $(F_n)_n$ be a sequence of continuous functions on $\mathbb{M}^N \times \mathbb{R}^N \times \mathbb{R} \times \bar{\Omega}$ where $\mathbb{M}^N$ denotes the space of $N\times N$ symmetric matrices, let $u_n \in W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$) satisfy for some $C \geqq 0$ independent of n*

$$F_n(D^2u_n, Du_n, u_n, x) \geqq -C \quad \text{a.e. in} \quad \Omega. \tag{74}$$

We assume that $F_n(A,\xi,t,x) \leqq F_n(B,\xi,t,x)$ for all $\xi \in \mathbb{R}^N$, $t \in \mathbb{R}$, $x \in \Omega$, $A \geqq B$ (in the sense of symmetric matrices), F_n converges uniformly on compact subsets to $-\mathrm{tr}(A)+|\xi|^p+\lambda t$, for some $p>2$, $\lambda \geqq 0$, u_n satisfies (62) *and converges uniformly on compact subsets of Ω to some function $u \in C(\bar{\Omega})$, and that $(u_n-u)^-$ converges uniformly to 0 on Ω. Then, u satisfies* (62).

Proof. Assume by way of contradiction that $u-\varphi$ admits a minimum at $x_0 \in \partial\Omega$ for some $\varphi \in C^2(\bar{\Omega})$, without loss of generality we may assume that x_0 is the unique minimum point of $u-\varphi$. By assumption, $u_n-\varphi$ achieves its minimum over Ω at some point $x_n \in \Omega$. We remark that

$$\min_{\bar{\Omega}} (u_n-\varphi) \leqq u_n(x) - \varphi(x) \xrightarrow[n]{} u(x) - \varphi(x) \quad \text{for all} \quad x \in \Omega$$

while

$$\min_{\Omega}(u_n-\varphi)\geqq \min_{\bar{\Omega}}(u-\varphi)-\|(u_n-u)^-\|_\infty,$$

hence $u_n(x_n)-\varphi(x_n)$ converges to $u(x_0)-\varphi(x_0)$. Now if x_n (or a subsequence) converges to $\bar{x}\in\bar{\Omega}$, then

$$u_n(x_n)-\varphi(x_n)\geqq u(x_n)-\varphi(x_n)-\|(u_n-u)^-\|_\infty \underset{n}{\to} u(\bar{x})-\varphi(\bar{x})$$

and thus $\bar{x}=x_0$ by the uniqueness of the minimum.

Next, by maximum principle, we have

$$F_n(D^2\varphi(x_n), D\varphi(x_n), u_n(x_n), x_n)\geqq -C$$

and passing to the limit we find

$$-\Delta\varphi(x_0)+|\nabla\varphi(x_0)|^p+\lambda u(x_0)\geqq -C. \tag{75}$$

and we observe that we may replace φ by $\varphi+c(\delta^\alpha-(d+\delta)^\alpha)$ where $\delta>0$, $\alpha=\frac{p-2}{p-1}$, $c>0$, since $u-\varphi+c((d+\delta)^\alpha-\delta^\alpha)$ admits also a unique minimum at x_0. Therefore, we deduce from (75)

$$-\Delta\varphi(x_0)+c\alpha\delta^{-(1-\alpha)}\Delta d(x_0)-c\alpha(1-\alpha)\delta^{-(2-\alpha)}+|\nabla\varphi(x_0)-c\alpha d^{1-\alpha}\nabla d|^p\geqq -C$$

and if we choose c in such a way that $(c\alpha)^{p-1}<(1-\alpha)$, we easily reach a contradiction letting δ go to 0. □

From this stability result, we deduce the

Corollary V.2. *Let $p>2$, let $f_n\in L^\infty_{\mathrm{loc}}(\Omega)$ satisfy*

$$f_n\geqq -C,\quad f_n\leqq Cd^{-\beta}\quad \text{a.e. in}\quad \Omega,\quad \text{for some}\quad C\geqq 0,\quad \beta\in(0,p). \tag{76}$$

We denote by u_n the unique solution in $W^{2,r}_{\mathrm{loc}}(\Omega)\cap C(\bar{\Omega})$ ($\forall r<\infty$) of (1) *satisfying* (62) *and we assume that f_n converges to f weakly in $L^\infty_{\mathrm{loc}}-*$. We denote by u the unique solution in $W^{2,r}_{\mathrm{loc}}(\Omega)\cap C(\bar{\Omega})$ ($\forall r<\infty$) of* (1) *satisfying* (62). *Then, u_n converges uniformly on $\bar{\Omega}$ to u.*

Proof. By Theorem IV.1, u_n is bounded in $C^{0,\gamma}(\bar{\Omega})$ for some $\gamma>0$ and in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r<\infty$). Then, we may assume (up to subsequences) that u_n converges uniformly on $\bar{\Omega}$ to a solution u of (1) [in $W^{2,r}_{\mathrm{loc}}(\Omega)\cap C(\bar{\Omega})$ for all $r<\infty$]. By Theorem V.2, u satisfies (62) and thus $u\equiv\tilde{u}$ by Theorem V.1. □

VI. The Ergodic Problem

In this section, we want to study the questions associated with the so-called ergodic stochastic control problems with state constraints. From the p.d.e.'s viewpoint this amounts to study the behaviour of λu and u as λ goes to 0 where u is the solution of (1) considered in the preceding sections. We will perform such an analysis in the three different cases studied above. The typical result we will obtain is that λu, $u-u(x_0)$ converge uniformly on compact subsets of Ω to $u_0\in\mathbb{R}$, v solution of

$$-\Delta v+|\nabla v|^p+u_0=f\quad\text{in}\quad\Omega,\qquad v(x_0)=0 \tag{77}$$

with the same boundary conditions for v than for u. And these will uniquely determine (u_0, v). In the preceding statements and below, x_0 is any fixed point in Ω and we assume that Ω is connected.

VI.1. Subquadratic Hamiltonians

Whenever it exists, we will denote by u_λ the solution of (1) with appropriate boundary conditions and if x_0 is any fixed point in Ω we will denote by $v_\lambda(\cdot)=u_\lambda(\cdot)-u_\lambda(x_0)$. We assume throughout this section that $1<p<2$.

Theorem VI.1. *Let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below and satisfy*

$$\lim\{f(x)d(x)^{-q}/d(x)\to 0_+\}=0\,. \tag{78}$$

Let u_λ be the unique solution of (1) in $W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ such that $u_\lambda\to+\infty$ as $d\to 0_+$. Then, ∇u_λ and λu_λ are bounded in $L^\infty_{\mathrm{loc}}(\Omega)$ and λu_λ, v_λ converge uniformly on compact subsets of Ω to $u_0\in\mathbb{R}$, $v\in W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ such that $v(x_0)=0$, v satisfies (15) and

$$-\Delta v+|\nabla v|^p+u_0=f\quad in\quad \Omega\,. \tag{79}$$

In addition, if $(\tilde u_0,\tilde v)\in\mathbb{R}\times W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ satisfies (79) and $\tilde v$ goes to $+\infty$ as d goes to 0_+, then $\tilde u_0=u_0$, $\tilde v=v+C$ for some $C\in\mathbb{R}$.

Proof. The proof involves several steps, we first obtain some bounds and we pass to the limit (Step 1). Then, we show that for any solution $(\tilde u_0,\tilde v)$ as above $\tilde v$ blows up at the boundary like $C_0 d^{-\alpha}$ (Step 2). Next, we show the uniqueness of u_0 (Step 3). Finally, we conclude with the uniqueness (up to constants) of $\tilde v$ (Step 4).

Step 1. Going through the proofs of Theorems II.1 and II.2, we see that u_λ satisfies for all $\varepsilon>0$, $\lambda\in(0,1]$

$$\frac{C_0-\varepsilon}{d^\alpha}-\frac{C_\varepsilon}{\lambda}\leqq u_\lambda\leqq\frac{C_0+\varepsilon}{d^\alpha}+\frac{C_\varepsilon}{\lambda} \tag{80}$$

for some $C_\varepsilon\geqq 0$, with the usual modifications if $\alpha=0$ (i.e. $p=2$). In particular, λu_λ is bounded from below and in L^∞_{loc}. Then, using Theorem IV.1, we deduce that ∇u_λ is bounded from below. Therefore, v_λ is bounded in $W^{1,\infty}_{\mathrm{loc}}$.

We next want to show that v_λ satisfies

$$\frac{C_1}{d^\alpha}-C\leqq v_\lambda\quad\text{in}\quad\Omega,\quad\text{for some}\quad C_1\in(0,C_0),\quad C\geqq 0\,. \tag{81}$$

Observe first that v_λ satisfies

$$-\Delta v_\lambda+|\nabla v_\lambda|^p+\lambda v_\lambda+\lambda u_\lambda(x_0)=f\quad\text{in}\quad\Omega\,.$$

And if we choose C_1 in $(0,C_0)$, we obtain denoting by $z=\dfrac{C_1}{d^\alpha}$

$$-\Delta z+|\nabla z|^p-\lambda z\leqq f-\lambda u_\lambda(x_0)\quad\text{on}\quad\Omega-\Omega_\delta$$

if δ is small enough, say $\delta\leqq\delta_0$. Now, there exists a constant $M\geqq 0$ such that

$$v_\lambda\geqq M\quad\text{on}\quad\Omega_{\delta_0}\,.$$

Hence, adapting the comparison results proved in Sect. II, we deduce

$$v_\lambda \geqq -M + \frac{C_1}{d^\alpha} \quad \text{on} \quad \Omega .$$

Extracting subsequences if necessary – the convergence of the whole sequence will follow from the uniqueness –, we may now pass to the limit $\lambda u_\lambda(x_0)$ converges to a constant u_0, v_λ converges to a solution v of (79) satisfying (81) and such that $v(x_0)=0$.

Step 2. Let $(\tilde{u}_0, \tilde{v}) \in \mathbb{R} \times W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ be a solution of (79) such that $\tilde{v}$ goes to $+\infty$ as d goes to 0_+. We want to prove that $\tilde{v}$ satisfies (15). To this end, we recall that $\bar{w}_{\varepsilon,\delta} = \dfrac{C_0+\varepsilon}{(d-\delta)^\alpha}$ satisfies

$$-\Delta \bar{w}_{\varepsilon,\delta} + |\nabla \bar{w}_{\varepsilon,\delta}|^p \geqq f - \tilde{u}_0 \quad \text{in} \quad \Omega_\delta - \Omega_{\delta_0} \quad \text{if} \quad 0<\delta<\delta_0=\delta_0(\varepsilon) .$$

Then, let $M_\varepsilon = \sup\{|\tilde{v}(x)|/x \in \Omega,\ d(x)=\delta_0(\varepsilon)\}$, we deduce from the maximum principle that

$$\tilde{v} \leqq \bar{w}_{\varepsilon,\delta} + M_\varepsilon \quad \text{on} \quad \Omega_\delta - \Omega_{\delta_0}$$

and letting δ go to 0, we deduce

$$-C \leqq \tilde{v} \leqq (C_0+\varepsilon)d^{-\alpha} + M_\varepsilon \quad \text{on} \quad \Omega . \tag{82}$$

Next, we simply observe that $\tilde{v}$ satisfies

$$-\Delta \tilde{v} + |\nabla \tilde{v}|^p + \tilde{v} = g \quad \text{in} \quad \Omega , \qquad \tilde{v} \to +\infty \quad \text{as} \quad d \to 0_+$$

where $g = f - \tilde{u}_0 + \tilde{v} \in L^\infty_{\mathrm{loc}}(\Omega)$ satisfies (78) because of (82). Therefore, Theorem II.2 yields the desired behaviour of $\tilde{v}$ near $\partial\Omega$.

Step 3. We first show that if (u_0, v), $(\tilde{u}_0, \tilde{v})$ are two solutions of (79) such that $v, \tilde{v} \to +\infty$ as $d \to 0_+$ then $u_0 = \tilde{u}_0$. To do so, we adapt an argument from Lions [16, 19]. Assume for instance that $u_0 < \tilde{u}_0$ and let $\varepsilon > 0$, $\theta \in (0,1)$. Obviously, we have

$$\begin{aligned} -\Delta(\theta\tilde{v}) + |\nabla(\theta\tilde{v})|^p + \varepsilon\theta\tilde{v} &\leqq \theta f + \varepsilon\theta\tilde{v} - \theta\tilde{u}_0 \\ &\leqq f + C(1-\theta) + \varepsilon\theta\tilde{v} - \theta\tilde{u}_0 . \end{aligned}$$

Next, since v, $\tilde{v}$ behave like $C_0 d^{-\alpha}$ near $\partial\Omega$, $\theta\tilde{v} \leqq v + C_\theta$ in Ω; hence

$$\begin{aligned} -\Delta(\theta\tilde{v}) + |\nabla(\theta\tilde{v})|^p + \varepsilon\theta\tilde{v} &\leqq f + C(1-\theta) + \varepsilon v + \varepsilon C_\theta - \theta\tilde{u}_0 \\ &\leqq f + \varepsilon v - u_0 + (u_0 - \theta\tilde{u}_0) + \varepsilon C_\theta + C(1-\theta) \end{aligned}$$

while v satisfies of course

$$-\Delta v + |\nabla v|^p + \varepsilon v = f + \varepsilon v - u_0 \quad \text{in} \quad \Omega .$$

But $u_0 < \tilde{u}_0$. Therefore, for θ near 1 and ε small enough (depending on θ) $\theta\tilde{v}$ is a subsolution of the equation satisfied by v. By Theorem II.2, this implies $\theta\tilde{v} \leqq v$. Letting θ go to 1, we find $\tilde{v} \leqq v$. But, $v + C_1$, $\tilde{v} + C_2$ satisfy the same problems for arbitrary constants C_1, C_2 and we reach a contradiction.

Step 4. Uniqueness of v_0 up to a constant.

Let $C_1 \in (0, C_0)$, again we observe that

$$-\Delta\left(\frac{C_1}{d^\alpha}\right) + \left|\nabla\left(\frac{C_1}{d^\alpha}\right)\right|^p \leqq f - u_0 \quad \text{in} \quad \Omega - \Omega_\delta$$

for some small enough $\delta > 0$. Therefore, if $\theta \in (0,1)$, $w = \theta\tilde{v} + (1-\theta)\frac{C_1}{d^\alpha}$ satisfies

$$-\Delta w + |\nabla w|^p \leqq \theta(f - u_0) + (1-\theta)(f - u_0) = f \quad \text{in} \quad \Omega - \Omega_\delta .$$

And since v, $\tilde{v}$ behave like $\frac{C_0}{d^\alpha}$, $(w - v) \to -\infty$ as $d \to 0_+$. Therefore, by the maximum principle,

$$\max_{\Omega - \Omega_\delta} (w - v) = \max_{\partial\Omega_\delta} (w - v) .$$

Hence, if we let θ go to 1, we find that

$$\sup_{\Omega - \Omega_\delta} (\tilde{v} - v) = \max_{\partial\Omega_\delta} (\tilde{v} - v) .$$

On the other hand, we also deduce from the maximum principle that

$$\max_{\bar{\Omega}_\delta} (\tilde{v} - v) = \max_{\partial\Omega_\delta} (\tilde{v} - v) .$$

Therefore, any maximum point $\bar{x}$ of $\tilde{v} - v$ on $\partial\Omega_\delta$ is in fact a global maximum point of $\tilde{v} - v$ on Ω. But, since $\tilde{v} - v = \psi$ satisfies the equation

$$-\Delta\psi + B \cdot \nabla\psi = 0 \quad \text{in} \quad \Omega$$

for some B $L^\infty_{\mathrm{loc}}(\Omega; \mathbb{R}^N)$, the strong maximum principle then yields

$$\tilde{v} - v \equiv (\tilde{v} - v)(\bar{x}) \quad \text{in} \quad \Omega$$

and we conclude. □

We would like to conclude this section with a few remarks on the case $p = 2$ which make a connection between our results and the interpretation of first eigenvalues in terms of optimal stochastic control that was considered by Holland [9, 10]. Indeed, if $p = 2$ and if v solves (79) with $v \to \infty$ as $d \to 0_+$ then we may perform the well known logarithmic transformation i.e. $v = -\operatorname{Log}\varphi$ and we find

$$-\Delta\varphi + f\varphi = u_0\varphi \quad \text{in} \quad \Omega, \qquad \varphi > 0 \quad \text{in} \quad \Omega, \qquad \varphi \to 0 \quad \text{as} \quad d \to 0_+ \tag{83}$$

or in other words u_0 is the minimum eigenvalue of the operator $(-\Delta + f)$ with Dirichlet boundary conditions. And the uniqueness of u_0 corresponds to the uniqueness of an eigenvalue with a positive eigenfunction, while the uniqueness of v up to an additive constant corresponds to the uniqueness of φ up to a multiplicative constant.

VI.2. Forced Infinite Boundary Conditions

We will be now concerned with the case when f grows so fast at the boundary that u_λ automatically has to blow up at $\partial\Omega$. To simplify the presentation, we will only consider the case when f satisfies (40) and therefore, by Theorem III.4, u_λ is the unique solution of (1) which is bounded from below.

Theorem VI.2. *Let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ satisfy* (40) *and let $p>1$, we denote by u_λ the unique solution of* (1) *which is bounded from below. Then, ∇u_λ and λu_λ are bounded in $L^\infty_{\mathrm{loc}}(\Omega)$ and λu_λ, v_λ converge uniformly on compact subsets of Ω to $u_0 \in \mathbb{R}$, $v \in W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ such that $v(x_0)=0$, v satisfies* (79) *and*

$$M^{-1}d^{-\alpha}-M \leqq v \leqq Md^{-\alpha}+M \quad \text{in} \quad \Omega, \quad \text{for some} \quad M \geqq 1,$$

$$\text{where} \quad \alpha=\frac{\beta}{p}-1 \quad \text{if} \quad \beta>p \tag{84}$$

and $d^{-\alpha}$ is replaced by $|\mathrm{Log}\, d|$ if $\beta=p\geqq q$. In addition, if f satisfies (18′), *then v satisfies* (19). *Furthermore, if $(\tilde{u}_0, \tilde{v}) \in \mathbb{R} \times W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r<\infty)$ satisfies* (79) *and $\tilde{v}$ is bounded from below, then $\tilde{u}_0=u_0$, $\tilde{v}=v+C$ for some $C \in \mathbb{R}$.*

Remark VI.1. If we consider the special case $p=2$ and if we perform the same logarithmic transformation as in the preceding section $v=-\mathrm{Log}\,\varphi$, we see that we are dealing with bounded, positive solutions of (83) and that the very fact that f blows up fast enough at $\partial\Omega$ forces φ to vanish on the boundary. Again, the uniqueness part of the above result may be interpreted as a uniqueness for first eigenvalues and eigenfunctions of the operator $-\Delta+f$ where no boundary condition on φ is imposed except "φ is bounded".

Proof of Theorem VI.2. Most of the proof of Theorem VI.1 goes through in this case except for the uniqueness arguments which use the precise behaviours of v, $\tilde{v}$ near the boundary. Of course, if we assume (18′) then the proof of Theorem VI.1 applies with some rather easy adaptations. In the general case, however, we have to involve slightly more elaborate arguments to show the uniqueness part of the above result. We only prove as in Theorem VI.1 that v, $\tilde{v}$ both satisfy (84). Next, we prove that $u_0=\tilde{u}_0$. We see that the corresponding proof (Step 3) in the proof of Theorem VI.1 only uses the fact that $\theta\tilde{v} \leqq v+C_\theta$ for any two solutions (u_0, v), $(\tilde{u}_0, \tilde{v})$ and for all $\theta \in (0,1)$. But this can be deduced from Theorem III.4: indeed $w=\theta\tilde{v}$ is a subsolution of the equation

$$-\Delta w+|\nabla w|^p+w \leqq g=\theta f+\theta\tilde{v}-\tilde{u}_0 \quad \text{in} \quad \Omega .$$

But in view of (40) and (84) $\theta f+\theta\tilde{v}-\tilde{u}_0 \leqq f+C'_\theta \leqq f+v-u_0+C_\theta$ for some constants C'_θ, $C_\theta \geqq 0$. Therefore, by Theorem III.4, we deduce

$$w \leqq v+C_\theta \quad \text{in} \quad \Omega$$

and our claim is proved.

Finally, we have to show the uniqueness of v up to a constant. Then, we observe that the proof given in Step 4 of the proof of Theorem VI.1 still applies provided we take C_1 small enough, indeed the only difference comes into the verification that $w-v \to -\infty$ as $d \to 0_+$ where $w=\theta\tilde{v}+(1-\theta)\dfrac{C_1}{d^\alpha}$ with $0<\theta<1$. But the inequality we just proved shows that

$$w \leqq \frac{1+\theta}{2}\, v+C_\theta+(1-\theta)\frac{C_1}{d^\alpha}$$

and taking C_1 small enough so that $\frac{C_1}{d^\alpha} \leqq \frac{1}{4} v + C$ we find

$$w \leqq \left(\frac{1+\theta}{2} + \frac{1-\theta}{4}\right) v + C_\theta + C = \frac{3+\theta}{4} v + C_\theta + C$$

therefore $w - v \to -\infty$ as $d \to 0_+$ since $v \to +\infty$ as $d \to 0_+$.

Then, the proof of Theorem VI.1 applies and we may conclude the proof of Theorem VI.2. □

VI.3. Superquadratic Hamiltonians

We now conclude this section by examining the remaining case namely the case when $p > 2$ and $f \in L^\infty_{\mathrm{loc}}(\Omega)$, is bounded from below and satisfies (41) with $\beta < p$. Then, we know there exists a unique solution u_λ of (1) satisfying (17) or (62). As λ goes to 0_+, we obtain the

Theorem VI.3. *Let $p > 2$, let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below and satisfy* (41) *for some $\beta < p$. We denote by u_λ the unique solution of* (1) *in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r < \infty$) satisfying* (62). *Then, ∇u_λ is bounded in $L^\infty_{\mathrm{loc}}(\Omega)$ and λu_λ is bounded in $L^\infty(\Omega)$. And λu_λ, v_λ converge uniformly on $\bar{\Omega}$ to $u_0 \in \mathbb{R}$, $v \in W^{2,r}_{\mathrm{loc}}(\Omega) \cap C(\bar{\Omega})$ ($\forall r < \infty$) such that $v(x_0) = 0$, v satisfies* (17) *and* (79). *In addition, if $(\tilde{u}_0, \tilde{v}) \in \mathbb{R} \times W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r < \infty$) satisfies* (79) *and if $\tilde{v}$ satisfies* (62), *then $\tilde{u}_0 = u_0$, $\tilde{v} = v + C$ for some $c \in \mathbb{R}$.*

Proof. Clearly, λu_λ is bounded from below. Then, we may apply the local gradient bound given in Theorem IV.1: hence, ∇u_λ is bounded in $L^\infty_{\mathrm{loc}}(\Omega)$. But this bound (see Corollary IV.1) also implies that u_λ is bounded in $C^{0,\theta}(\bar{\Omega})$ for some $\theta \in (0,1)$ independent of λ. Therefore, up to subsequences, λu_λ and u_λ converge uniformly on $\bar{\Omega}$ to $u_0 \in \mathbb{R}$, $v \in W^{2,r}_{\mathrm{loc}}(\Omega) \cap C(\bar{\Omega})$ ($\forall r < \infty$) such that $v(x_0) = 0$, v satisfies (79). Next, by Theorem V.2, v satisfies (62) and therefore, by Proposition V.1, v satisfies (17).

Notice also that v is the unique solution satisfying (62) of

$$-\Delta v - |\nabla v|^p + \lambda v = g_\lambda \quad \text{in} \quad \Omega,$$

where $g_\lambda = f - u_0 + \lambda v$.

Next, using Theorem V.1, the uniqueness of u_0 follows immediately as in Step 3 of the proof of Theorem VI.1 ($\theta = 1$ is enough in this case).

Finally, we want to prove the uniqueness of v_0 up to a constant. Again, the only fact we have to prove is the following

$$\sup_{\Omega - \Omega_\delta} (\tilde{v} - v) = \max_{\partial\Omega_\delta} (\tilde{v} - v).$$

To this end, we set $\alpha = \frac{p-2}{p-1}$ and we consider for $\theta \in (0,1)$, $w = \theta\tilde{v} - (1-\theta) C_1 d^\alpha$ then for $C_1 > 0$, δ small enough (independent of θ) we have

$$-\Delta w + |\nabla w|^p \leqq f - u_0 - 1 \quad \text{in} \quad \Omega - \Omega_\delta.$$

In particular, for λ small enough, we have

$$-\Delta w + |\nabla w|^p + \lambda w \leqq g_\lambda \quad \text{in} \quad \Omega - \Omega_\delta.$$

We may now adapt without any real modification the proof of Theorem V.1 to deduce

$$\sup_{\Omega-\Omega_\delta}(w-v)=\max_{\partial\Omega_\delta}(w-v)$$

and we conclude letting θ go to 1. $\square$

VII. Optimal Stochastic Control with State Constraints

We now want to use the results obtained in the preceding sections in order to solve some optimal stochastic control problems with state constraints: a rather vague way to formulate our problem is to say that we want to "constrain a Brownian motion in a given domain Ω by controlling its drift". More precisely, we consider a system whose state is given by the solution of the following stochastic differential equation

$$dX_t=a_t dt+\sqrt{2}\,dB_t, \qquad X_0=x\in\Omega, \tag{3}$$

where B_t is a Brownian motion on a standard probability space (Ω, F, F_t, P) and where a_t is the control process i.e. a progressively measurable stochastic process taking values in $\mathbb{R}^N$ for instance. In other words

$$X_t=x+\int_0^t a_s ds+\sqrt{2}B_t$$

and we assume (at least) that $\int_0^T |a_s|ds<\infty$ a.s. $(\forall r<\infty)$.

We will say that this control a is admissible if $X_t\in\Omega\ \forall t\geqq 0$ a.s. Even if we could work with general controls of the above form, we will restrict ourselves to feedbacks or Markovian controls which are defined as follows. Let $a\in C(\Omega;\mathbb{R}^N)$, we may solve the stochastic differential equation

$$dX_t=a(X_t)dt+\sqrt{2}\,dB_t \quad \text{for} \quad 0\leqq t<\tau_x, \qquad X_0=x\in\Omega, \tag{85}$$

where τ_x is the first exit time of X_t from Ω i.e., $\tau_x=\inf\{t\geqq 0,\ X_t\notin\Omega\}$ ($\tau=+\infty$ if $X_t\in\Omega$ for all $t\geqq 0$). Thus, $a(X_t)$ is really the control but we will ignore this minor point of terminology and we will say that $a(\cdot)$ is the control (or control policy). Next, we define an admissible control as a control $a(\cdot)$ such that

$$P(\tau_x<\infty)=0 \quad \text{for all} \quad x\in\Omega. \tag{86}$$

And we will denote by $\mathscr{A}$ the class of all admissible controls.

For each a, we define a cost function

$$J_\lambda(x,a)=E\int_0^\infty \{f(X_t)+c|a(X_t)|^q\}e^{-\lambda t}dt, \tag{87}$$

where $c=c(p,q)=q^{-1}p^{-1/(p-1)}$, $\lambda>0$ is the discount factor. Observe that the running cost $f(x)+c|a|^q$ contains two terms: one which measures the cost for the state to be at x, and the other measuring the cost for using the control a. All throughout this section $f\in L^\infty_{\text{loc}}(\Omega)$ is bounded from below so that $J_\lambda(x,a)$ makes sense even if it may be infinite. This is of course a very special example but we will come back on part 2 on much more general problems for which similar results to those which follow still hold.

Finally, we want to minimize J_λ. We introduce the value function

$$u_\lambda(x) = \inf_{a \in \mathscr{A}} J_\lambda(x, a), \qquad \forall x \in \Omega. \tag{88}$$

The typical questions that one wants to solve in such problems is to determine u_λ and possibly an optimal control (here an optimal Markovian control or an optimal feedback), i.e. some a in $\mathscr{A}$ such that

$$u_\lambda(x) = J_\lambda(x, a), \qquad \forall x \in \Omega.$$

And this is precisely what we will achieve using the results of the preceding sections. Let us also observe that it is not completely obvious that $\mathscr{A} \neq \emptyset$, let alone that there exists $a \in \mathscr{A}$ such that $J(x, a) < \infty$ for $x \in \Omega$.

Finally, we will also consider the case of ergodic control which consists, roughly speaking, in taking $\lambda = 0$.

VII.1. Subquadratic Hamiltonians

Theorem VII.1. *Let $1 < p < 2$, let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below and satisfy* (78). *Then, the value function u_λ given by* (88) *is the unique solution of* (1) *in $W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r < \infty)$ such that $u_\lambda \to +\infty$ as $d \to 0_+$. Furthermore, $a_0(x) = p|\nabla u_\lambda|^{p-2}\nabla u_\lambda(x)$ is the unique optimal markovian control.*

Proof. We denote by $\tilde{u}_\lambda$ the unique solution of (1) in $W^{2,r}_{\mathrm{loc}}(\Omega)$ $(\forall r < \infty)$ such that $\tilde{u}_\lambda \to +\infty$ as $d \to 0_+$. We are first going to show that $\tilde{u}_\lambda \geqq u_\lambda$ and that $a_0 \in \mathscr{A}$. Indeed, for $\delta > 0$ let $\tau_x^\delta = \inf(t \geqq 0, X_t \notin \Omega_\delta)$, we apply Itô's formula on $[0, \tau_\delta]$ with the process X_t corresponding to the control $a_0(x) = -p|\nabla \tilde{u}_\lambda|^{p-2}\nabla u_\lambda$ and we find for all x Ω_δ

$$\tilde{u}_\lambda(x) = E \int_0^{\tau_x^\delta} \{-\Delta \tilde{u}_\lambda(X_t) + p|\nabla \tilde{u}_\lambda|^p(X_t) + \lambda \tilde{u}_\lambda(X_t)\} e^{-\lambda t} dt + E\tilde{u}_\lambda(X_{\tau_x^\delta}) e^{-\lambda \tau_x^\delta}$$

hence from the equation (1) this yields

$$\tilde{u}_\lambda(x) = E \int_0^{\tau_x^\delta} \{f(X_t) + c|a_0(X_t)|^q\} e^{-\lambda t} dt + E\tilde{u}_\lambda(X_{\tau_x^\delta}) e^{-\lambda \tau_x^\delta}, \qquad \forall x \in \Omega_\delta. \tag{89}$$

In particular, we may deduce from this quantity

$$\left(\inf_{\partial\Omega_\delta} \tilde{u}_\lambda\right) E[e^{-\lambda \tau_x^\delta}] \leqq C, \qquad \forall x \in \Omega_\delta$$

for some C independent of δ.

Now, since $\tilde{u}_\lambda \to +\infty$ as $d \to 0_+$, $\left(\inf_{\partial\Omega_\delta} \tilde{u}_\lambda\right) \to +\infty$ as $\delta \to 0_+$, therefore $E[e^{-\lambda \tau_x^\delta}] \to 0$ as $\delta \to 0_+$. Hence $E[e^{-\lambda \tau_x}] = 0$ for all $x \in \Omega$ and this precisely means that $a_0 \in \mathscr{A}$.

In addition, we may also deduce from (89) that for δ small enough and for $x \in \Omega_\delta$

$$\tilde{u}_\lambda(x) \geqq E \int_0^{\tau_x^\delta} \{f(X_t) + c|a_0(X_t)|^q\} e^{-\lambda t} dt$$

since $\tilde{u}_\lambda \geqq 0$ for $x \in \Omega - \Omega_\delta$. Now, if we let δ go to and if we use the fact that $a_0 \in \mathscr{A}$ and thus $\tau_x^\delta \to +\infty$ a.s. as $\delta \to 0_+$ for all $x \in \Omega$, this yields

$$\tilde{u}_\lambda(x) \geqq J(x, a_0), \quad \forall x \in \Omega. \tag{90}$$

We next show that $\tilde{u}_\lambda \equiv u_\lambda$. If this is the case, (90) then implies that a_0 is optimal. To show that $\tilde{u}_\lambda \equiv u_\lambda$, we first recall from section that there exist w_n subsolutions of (1) in $W^{2,r}(\Omega)$ ($\forall r < \infty$) such that $w_n \xrightarrow[n]{} \tilde{u}_\lambda$ uniformly on compact subsets of Ω. Therefore, if $a \in \mathscr{A}$, we find using again Itô's rule for all $x \in \Omega_\delta$

$$w_n(x) \leqq E \int_0^{\tau_x^\delta} \{f(X_t) + c|a(X_t)|^q\} e^{-\lambda t} dt + E[w_n(X_{\tau_x^\delta}) e^{-\lambda \tau_x^\delta}].$$

Now, if $J(x, a) = +\infty$, we obviously have $w_n(x) \leqq J(x, a)$, while if $J(x, a) < \infty$, we deduce from the above inequality letting δ go to 0_+

$$w_n \leqq E \int_0^\infty \{f(X_t) + c|a(X_t)|^q\} e^{-\lambda t} dt = J(x, a)$$

since $\tau_x^\delta \to +\infty$ a.s. as $\delta \to 0_+$ and thus

$$|E[w_n(X_{\tau_x^\delta}) e^{-\lambda \tau_x^\delta}]| \leqq \sup_{\bar{\Omega}} |w_n| E[e^{-\lambda \tau_x^\delta}] \to 0 \quad \text{as} \quad \delta \to 0_+.$$

Therefore, letting n go to $+\infty$, we finally deduce for all $x \in \Omega$, $a \in \mathscr{A}$

$$\tilde{u}_\lambda(x) \leqq J(x, a)$$

and our claim is proved. $\square$

The uniqueness of the optimal control is a bit technical but very simple to understand so we just sketch the argument: assume that a is optimal then applying Itô's rule we find for all $\delta > 0$, $x \in \Omega_\delta$

$$\begin{aligned} u_\lambda(x) &= E \int_0^{\tau_x^\delta} \{f(X_t) - a(X_t) \cdot \nabla u_\lambda(X_t) - |\nabla u_\lambda(X_t)|^p\} e^{-\lambda t} dt + E u_\lambda(X_{\tau_x^\delta}) e^{-\lambda \tau_x^\delta} \\ &\leqq E \int_0^{\tau_x^\delta} \{f(X_t) + c|a(X_t)|^q\} e^{-\lambda t} + E u_\lambda(X_{\tau_x^\delta}) e^{-\lambda \tau_x^\delta}. \end{aligned}$$

But recalling that $u_\lambda(x) = J(x, a)$ for all $x \in \Omega$ and using the Markov property of X_t, we deduce that the above right-hand side is also equal to $u_\lambda(x)$. Therefore, the equality yields that for all $x \in \Omega_\delta$

$$a(X_t) = a_0(X_t) \quad \text{for all} \quad t \in (0, \tau_x^\delta) \quad \text{a.s.},$$

(where X_t is the solution corresponding to a) and letting δ go to 0_+ we finally find that for all $x \in \Omega$, $a(x) = a_0(x)$ (recall that a, a_0 are continuous on Ω). $\square$

VII.2. Superquadratic Hamiltonians

Theorem VII.2. *Let $f \in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below, satisfy* (41) *for some $\beta < p$ and let $p > 2$. Then, the value function u_λ given by* (88) *is the unique solution of* (1) *in $W^{2,r}_{\mathrm{loc}}(\Omega)$ ($\forall r < \infty$) satisfying* (62).

Proof. We will approximate the Hamiltonian $|\xi|^p$ as follows: let $R \geqq 1$, consider some Hamiltonian H_R such that H_R is convex on $\mathbb{R}^N$, $H_R(\xi)=|\xi|^p$ if $|\xi| \leqq R$, $H_R(\xi)|\xi|^{-\alpha}$ is constant for $|\xi|$ large where $\frac{\alpha}{\alpha-1} > \beta$ and $1<\alpha<2$, H_R increases uniformly on compact subsets to $|\xi|^p$ as R goes to $+\infty$. And we denote by $L_R(\eta)$ the following convex function

$$-L_R(\eta)=\inf_{\xi \in \mathbb{R}^N} (\eta \cdot \xi + H_R(\xi))$$

so that $L_R(\eta) \geqq c|\eta|^q$ and L_R decreases uniformly on compact subsets to $c|\eta|^q$.

Then, because $H_R = C_R|\xi|^\alpha$ for $|\xi|$ large, it is not difficult to adapt the results and the proofs of Sect. II to show that there exists a unique solution u_R of

$$-\Delta u_R + H_R(\nabla u_R) + \lambda u_R = f(x) \quad \text{in} \quad \Omega, \qquad u_R \in W^{2,r}_{\text{loc}}(\Omega) \quad (\forall r<\infty)$$

such that $u_R \to +\infty$ as $d \to 0_+$.

And exactly as in the preceding section, we can check that

$$u_R(x) = \inf_{a \in \mathscr{A}} E \int_0^\infty \{f(X_t) + L_R(a(X_t))\} e^{-\lambda t} dt, \qquad \forall x \in \Omega.$$

Of course, u_R decreases to the value function u_λ given by (88). On the other hand, we remark that we may choose $H_R \in C^2(\mathbb{R}^N)$ such that

$$|D^2 H_R(\xi)|\,|\xi|^2, \qquad |DH_R(\xi)|\,|\xi| \leqq C_0(H_R(\xi)+1), \qquad \forall \xi \in \mathbb{R}^N,$$

and

$$H_R(\xi) \geqq |\xi|^\alpha, \qquad \forall \xi \in \mathbb{R}^N$$

for some C_0 independent of R. And we may adapt the a priori estimates in the appendix (see also part 2) to deduce that u_R is bounded in $W^{1,\infty}_{\text{loc}}$ and thus in $W^{2,r}_{\text{loc}}(\Omega)$ $(\forall r<\infty)$. Hence, $u_\lambda \in W^{2,r}_{\text{loc}}(\Omega)$ $(\forall r<\infty)$ and solves (1). But then by Corollary IV.1, u_λ extends continuously to $\bar{\Omega}$ and we may apply Theorem V.2 to deduce that u_λ is the unique solution of (1) satisfying (62). $\square$

The question of the optimality of the control $a_0 = -p|\nabla u_\lambda|^{p-2}\nabla u_\lambda$ is much more delicate: in fact, if an optimal control exists, by a similar proof to the one made in the preceding section, it has to be a_0 and if we know that $a_0 \in \mathscr{A}$ then a_0 is the optimal Markovian control. Hence, the main problem is whether $a_0 \in \mathscr{A}$. We know how to prove that $a_0 \in \mathscr{A}$ only when (61) (or some easy variants) holds and we refer the reader to Sect. IV.3 where a few cases when (61) holds are given. Indeed, if (61) holds then we deal with a diffusion X_t satisfying

$$dX_t = \sqrt{2}\,dB_t + a_0(X_t)dt,$$

where a_0 satisfies

$$a_0(x)d(x) \to -\mu n \quad \text{as} \quad d(x) \to 0_+, \tag{91}$$

with $\mu = q > 1$. Then, we claim that for any diffusion process of the above form, if (91) holds and $\mu > 1$, then X_t never leaves Ω with probability 1, while if (91) holds and $\mu<1$, X_t hots $\partial\Omega$ in finite time with propability 1. Indeed, if $\mu>1$, we apply Itô's rule with $-\log d(x)$ and we find for all $T<\infty$

$$-\log d(x)=E\left[-\log d(X_{\tau\wedge T})+\int_0^{\tau\wedge T}\left\{\frac{\Delta d}{d}-\frac{|\nabla d|^2}{d^2}-\frac{a\cdot\Delta d}{d}\right\}(X_s)ds\right]$$

(in fact we should replace τ by τ_δ for $\delta>0\ldots$). And we observe that

$$\frac{\Delta d}{d}-\frac{|\nabla d|^2}{d^2}-\frac{a\cdot\nabla d}{d}$$

behaves like $(\mu-1)\dfrac{1}{d^2}$ near $\partial\Omega$ and so this quantity is bounded from below on Ω. Hence, we obtain

$$E[-\log d(X_{\tau\wedge T})]\leqq CT-\log d(x)$$

therefore for all $x\in\Omega$, $P(\tau<T)=0$ and we conclude since this holds for all $T<\infty$.

On the other hand if $\mu<1$ by a simple argument, showing that $E[\tau_x]\leqq C$ for all $x\in\Omega$ is easily done if we prove the existence of a supersolution of

$$-\Delta z-a\cdot\nabla z\geqq\varepsilon \quad \text{in} \quad \Omega, \qquad \text{for some} \quad \varepsilon>0, \quad z\in C(\bar\Omega), \quad z=0 \quad \text{on} \quad \partial\Omega.$$

But this is achieved by considering for $\mu_0\in(\mu,1)$ the function

$$z_1=d^{1-\mu_0}(1-\mu_0)^{-1}-d^2\{2(\mu_0+1)\}^{-1}$$

which satisfies in $\Omega-\Omega_\delta$ for some δ small enough

$$\begin{aligned}-\Delta z_1-a\cdot\nabla z_1&=\mu_0 d^{-\mu_0-1}-a\cdot\nabla d d^{-\mu_0}+\frac{1}{\mu_0+1}\{1+a\cdot\nabla d d\}\\&+\left\{\frac{1}{\mu_0+1}-d^{-\mu_0}\right\}\Delta d\geqq K>0\end{aligned}$$

for some $K>0$. Then, we consider the solution z_2 of

$$-\Delta z_2-a\cdot\nabla z_2=1 \quad \text{in} \quad \Omega_\delta, \qquad z_2=0 \quad \text{on} \quad \partial\Omega_\delta.$$

Finally, we set $z=z_1$ in $\Omega-\Omega_\delta$, $=z_1|_{\partial\Omega_\delta}+\gamma z_2$ in $\bar\Omega_\delta$ where γ is small enough so that $\gamma\dfrac{\partial z_2}{\partial n}\geqq\dfrac{\partial z_1}{\partial n}$ on $\partial\Omega_\delta$. It is then easy to check that z satisfies the desired inequality with $\varepsilon=\min(K,\gamma)$.

VII.3. Forced Constraints

We first observe that by the results and methods of the preceding sections (and the interior estimates given in the Appendix), for any $p>1$ and for any $f\in L^\infty_{\text{loc}}(\Omega)$ bounded from below there exists a solution $\underline{u}\in W^{2,r}_{\text{loc}}(\Omega)$ $(\forall r<\infty)$ of (1) such that for all $v\in W^{2,r}(\Omega)$ $(\forall r<\infty)$ satisfying

$$-\Delta v+|\nabla v|^p+\lambda v\leqq f \quad \text{in} \quad \Omega \tag{92}$$

then $v \leqq \underline{u}$ in Ω. Of course, if $1<p\leqq 2$ and $f(x)\leqq Cd(x)^{-1}$ then (see Sect. II) $\underline{u}\to +\infty$ as $d\to 0_+$ and $\underline{u}$ is the minimum such solution, if $p>2$ and $f(x)\leqq Cd(x)^{-\beta}$ for some $\beta<p$ then $\underline{u}$ is the unique solution of (1) satisfying (62) (see Sects. IV and V), while if $f(x)\geqq cd(x)^{-\beta}-C$ for some $c>0$, $\beta\geqq \max(p,q)$ then $\underline{u}\to +\infty$ as $d\to 0_+$ and $\underline{u}$ is the minimum solution of (1) bounded from below [and we have uniqueness if f behaves like $C_1 d(x)^{-\beta}$]. In fact, if $1<p\leqq 2$, then $\underline{u}\to +\infty$ as $d\to 0_+$ and $\underline{u}$ is the minimum such solution.

We then have the following

Proposition VII.1. *Let $1<p\leqq 2$, or let $p>2$ and $f\geqq cd^{-\beta}-C$ for some $c>0$, $C\geqq 0$, $\beta\geqq p$. Then, the value function u_λ given by (88) is the above ("minimum explosive") solution $\underline{u}$. In addition, $a_0(x)=-p|\nabla u|^{p-2}\nabla u$ is the unique optimal Markovian control.*

In fact, since $\underline{u}\to\infty$ as $d\to 0_+$, the proof is exactly the same as the proof of Theorem VI.1: one shows that $\underline{u}\geqq u_\lambda$ and $a_0\in\mathscr{A}$, then $\underline{u}\leqq u_\lambda$ and a_0 is the unique optimal Markovian control.

Remark VII.1. These results show that for any f bounded from below the formula (88) yields a finite function (locally bounded) on Ω. This may be proved directly by a tedious probabilistic construction of a control $\bar{a}\in\mathscr{A}$ such that $J(x,\bar{a})<\infty$ for all $x\in\Omega$.

VII.4. Ergodic Control

We now want to explain in this section the control problems associated with the asymptotic problems solved in Sect. VI. We begin with the cases when solutions go to $+\infty$ as $d(x)$ goes to 0_+.

Theorem VII.3. *Let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below and satisfy (78), let $1<p\leqq 2$. We denote by (v,u_0) the solutions given by Theorem VI.1. Then, we have the following equalities: for any $a\in\mathscr{A}$, let θ_a be a stopping time bounded by some arbitrary $T\geqq 0$ (independent of a), then*

$$v(x)=\inf_{a\in\mathscr{A}} E\int_0^{\theta_a}\{f(X_t)+c|a(X_t)|^q\}dt+v(X_{\theta_a})-\theta_a u_0\,,\qquad \forall x\in\Omega\,, \tag{93}$$

$$u_0=\lim_{T\to\infty}\inf_{a\in\mathscr{A}} E\frac{1}{T}\int_0^T\{f(X_t)+c|a(X_t)|^q\}dt\,,\qquad \forall x\in\Omega \tag{94}$$

and the control $a_0=-p|\nabla v|^{p-2}\nabla v$ belongs to $\mathscr{A}$ and is the unique optimal Markovian control where optimal means that (93)–(94) are equalities when we choose $a=a_0$.

Theorem VII.4. *Let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ satisfy (40) and let $p>1$. Denoting by (v,u_0) the solutions given by Theorem VI.2, Theorem VII.3 still holds.*

Since the proof of Theorem VII.4 is very similar to the one of Theorem VII.3 we will only prove the latter.

We first deduce from Itô's formula that if X^0_t denotes the process corresponding to the choice a_0 then for all $\delta>0$, $x\in\Omega$

$$v(x)=E\int_0^{\theta_0\wedge\tau^\delta_x}\{f(X^0_t)+c|a_0(X^0_t)|^q\}dt+v(X^0_{\theta_0\wedge\tau^\delta_x})-\theta_0\wedge\tau^\delta_x u_0\,, \tag{95}$$

where θ_0 stands for θ_{a_0} and τ_x^δ is the first exit time from Ω_δ. In particular for $\theta_0=T$, we deduce

$$E(v(X^0_{T\wedge\tau_x^\delta}))\leqq v(x)+CT, \quad \text{for some} \quad C\geqq 0.$$

Therefore, recalling that v is bounded from below, we obtain

$$\left(\inf_{\partial\Omega_\delta} v\right) P[\tau_x^\delta\leqq T]\leqq v(x)+C(1+T)$$

and since $v\to+\infty$ as $d\to 0_+$, we deduce that $a_0\in\mathscr{A}$.

In addition, if we pass to the limit in (95) as δ goes to 0_+, we find for all $x\in\Omega$

$$v(x)=E\int_0^{\theta_0}\{f(X_t^0)+c|a_0(X_t^0)|^q\}dt-\theta_0u_0+\lim_{\delta\to 0_+} E[v(X^0_{\theta_0\wedge\tau_x^\delta})]$$

and

$$\lim_{\delta\to 0_+} E[v(X^0_{\theta_0\wedge\tau_x^\delta})]\geqq \lim_{\delta\to 0_+}\{E[(v+C)(X^0_{\theta_0})\mathbb{1}_{\theta_0\leqq\tau_x^\delta}]-C\},$$

where $C\leqq \inf_\Omega v$, and this last expectation increases to $E[v(X^0_{\theta_0})]$. Hence, we finally obtain for all $x\in\Omega$

$$v(x)\geqq E\int_0^{\theta_0}\{f(X_t^0)+c|a_0(X_t^0)|^q\}dt+v(X^0_{\theta_0})-\theta_0u_0. \tag{96}$$

And taking $\theta_0=T$, we also deduce for all $x\in\Omega$

$$u_0\geqq \varlimsup_{T\to\infty}\frac{1}{T}E\int_0^T\{f(X_T^0)+c|a_0(X_t^0)|^q\}dt \tag{97}$$

since $E\frac{1}{T}v(X_T^0)\geqq -\frac{C}{T}\to 0$ as $T\to\infty$.

To complete the proof of Theorem VII.3, we basically need to prove the complementary inequalities in (93)–(94). This will be achieved by first introducing some approximated problem: let (v^δ, u_0^δ) be the solution in $W^{2,r}_{\text{loc}}(\Omega^\delta)\times\mathbb{R}$ $(\forall r<\infty)$ of (79) with Ω replaced by Ω^δ such that $v^\delta(x_0)=0$, $v^\delta\to+\infty$ as $d(x)\to 0_+$. With the techniques of Sect. VI one can show that $v^\delta\uparrow v$ as δ goes to 0_+ and converges uniformly on compact subsets of Ω, while $u_0^\delta\downarrow u_0$ as $\delta\downarrow 0_+$.

Using Itô's formula, we immediately obtain for all $x\in\Omega$

$$v^\delta(x)\leqq \inf_{a\in\mathscr{A}} E\int_0^{\theta_a}\{f(X_t)+c|a(X_t)|^q\}dt+v^\delta(X_{\theta_a})-\theta_a u_0^\delta$$

and letting δ go to 0_+ we deduce since $v^\delta\uparrow v$ as $\delta\downarrow 0_+$

$$v(x)\leqq \inf_{a\in\mathscr{A}} E\int_0^{\theta_a}\{f(X_t)+c|a(X_t)|^q\}dt+v(X_{\theta_a})-\theta_a u_0$$

and this combined with (96) yields (93). In addition taking $\theta_a=T$, we also deduce

$$u_0^\delta\leqq \inf_{a\in\mathscr{A}} E\frac{1}{T}\int_0^T\{f(X_t)+c|a(X_t)|^q\}dt+\frac{2}{T}\sup_{\bar\Omega}|v^\delta|$$

hence

$$u_0^\delta \leqq \varlimsup_{T\to\infty} \inf_{a\in\mathscr{A}} E\frac{1}{T}\int_0^T \{f(X_t)+c|a(X_t)|^q\}dt$$

and letting δ go to 0, the resulting inequality combined with (97) yields (94). This also shows the optimality of a_0 and the uniqueness is easy to prove as in the preceding sections.

By the same truncation argument as in the proof of Theorem VII.2, one deduces the

Theorem VII.5. *Let $p>2$, let $f\in L^\infty_{\mathrm{loc}}(\Omega)$ be bounded from below and satisfy* (40). *We denote by (v, u_0) the solutions given by Theorem VI.2. Then, the identities* (93)–(94) *still hold.*

Appendix: On Some Local Gradient Bounds

We want to show here some local gradient bounds for solutions of

$$-\varepsilon\Delta u+|\nabla u|^p+\lambda u=f \quad \text{in} \quad \Omega, \qquad u\in W^{2,r}_{\mathrm{loc}}(\Omega) \quad (\forall r<\infty), \tag{A.1}$$

where $0<\varepsilon\leqq 1$, $0\leqq\lambda\leqq 1$, $1<p<\infty$, and $f\in L^\infty(\Omega)$ or even $f\in W^{1,\infty}(\Omega)$ and Ω is a bounded open set in $\mathbb{R}^n$. These bounds are obtained by the method introduced in [16, 19]. For related local bounds concerning different equations, we refer to Bombieri et al. [4], Ladyzhenskaya and Ural'tseva [13], Simon [26–28]. Our main result is the

Theorem A.1. *For any $\delta>0$, we set $\Omega_\delta=\{x\in\Omega/\mathrm{dist}(x,\partial\Omega)>\delta\}$.*

1) *Let $f\in W^{1,\infty}(\Omega)$, then we have for all $\delta>0$*

$$|\nabla u(x)|\leqq C_\delta \quad \textit{if} \quad x\in\Omega_\delta, \tag{A.2}$$

where C_δ depends only on bounds on $|\nabla f|$, lower bounds on $\lambda u-f$, δ, and p.

2) *Let $f\in L^\infty(\Omega)$, then we have for all $r<\infty$, $\delta>0$*

$$\|\nabla u\|_{L^r(\Omega_\delta)}\leqq C_\delta, \tag{A.3}$$

where C_δ depends only on bounds on f, lower bounds on $f-\lambda u$, δ, p, and r.

Proof. We begin with case 1) i.e. when $f\in W^{1,\infty}(\Omega)$. In both cases, we will ignore the fact that u is not assumed to be smooth and we will thus skip the tedious approximation argument required to make the proof below complete. Then, let $\theta\in(0,1)$ to be determined later on and let $\varphi\in\mathscr{D}(\Omega)$, $0\leqq\varphi\leqq 1$ in Ω, $\varphi\equiv 1$ on Ω_δ, be such that

$$|\Delta\varphi|\leqq C\varphi^\theta, \qquad |\nabla\varphi|^2\leqq C\varphi^{1+\theta} \quad \text{in} \quad \Omega$$

for some C (depending only on δ, θ).

We next consider $w=|\nabla u|^2$ and we compute easily on $\mathrm{Supp}\,\varphi$

$$\left.\begin{aligned}&-\varepsilon\Delta(\varphi w)+p|\nabla u|^{p-2}\nabla u\cdot\nabla(\varphi w)+2\lambda\varphi w+2\varepsilon\varphi|D^2u|^2+2\varepsilon\frac{\nabla\varphi}{\varphi}\cdot\nabla(\varphi w)\\&=2\varphi\nabla f\cdot\nabla u+p|\nabla u|^{p-2}(\nabla u\cdot\nabla\varphi)w-\varepsilon(\nabla\varphi)w+2\varepsilon\frac{|\nabla\varphi|^2}{\varphi}w.\end{aligned}\right\} \tag{A.4}$$

Then, let $x_0 \in \Omega$ be a maximum point of φw: we may assume that $x_0 \in \operatorname{Supp} \varphi$ and by the classical maximum principle we deduce from (A.4) the following inequality where all functions are taken at x_0

$$2\varepsilon\varphi|D^2u|^2 \leqq C\varphi w^{1/2} + C\varphi^\theta w^{\frac{p+1}{2}} + C\varepsilon\varphi^\theta w. \tag{A.5}$$

Now, from Cauchy-Schwarz inequality and (A.1)

$$|D^2u|^2 \geqq \frac{1}{N}(\Delta u)^2 \geqq \frac{1}{N\varepsilon^2}(|\nabla u|^p + \lambda u - f)^2 \geqq \frac{1}{N\varepsilon^2}(|\nabla u|^p - C)^{+2}$$

and this combined with (A.5) yields

$$\varphi w^p \leqq C + C\varepsilon\varphi w^{1/2} + C\varepsilon\varphi^\theta w^{\frac{p+1}{2}} + C\varepsilon^2\varphi^\theta w. \tag{A.6}$$

Now, choosing $\theta \geqq \frac{3-p}{2}$, we deduce easily

$$\max_{\Omega} \varphi w = \varphi w(x_0) \leqq C.$$

In case 2), i.e. when $f \in L^\infty(\Omega)$ we use integral estimates as follows: let $m \geqq 1$, we multiply (A.4) by $(\varphi w)^m$ and we find

$$\begin{aligned}
&\varepsilon m \int |\nabla(\varphi w)|^2(\varphi w)^{m-1}dx + p\int |\nabla u|^{p-2}\nabla u \cdot \nabla(\varphi w)(\varphi w)^m dx \\
&\quad + \varepsilon \int \varphi |D^2u|^2(\varphi w)^m dx + \frac{1}{N\varepsilon}\int \varphi(|\nabla u|^p - C)^{+2}\varphi^m w^m dx \\
&\quad + 2\varepsilon \int \varphi^{m-1}\nabla\varphi \cdot \nabla(\varphi^{m+1}w^{m+1})(m+1)^{-1}dx \\
&\leqq 2\int \varphi^{m+1}w^m \nabla f \cdot \nabla u\, dx + C_p \int w^{\frac{p+1}{2}}\varphi^{\theta m}w^m dx + C\varepsilon \int \varphi^{m+\theta}w^m dx.
\end{aligned}$$

We now want to bound the following terms

$$\begin{aligned}
&2\int \varphi^{m+1}w^m \nabla f \cdot \nabla u\, dx \leqq 2C\int \varphi^{m+1}w^m|D^2u|dx + m\int \varphi|\nabla(\varphi w)|(\varphi w)^{m+1} \\
&\qquad \times |\nabla u| dx + C\int \varphi^{m+\theta}w^{m+1/2}dx \\
&\quad \leqq \varepsilon \int \varphi^{m+1}w^m|D^2u|^2dx + \frac{C}{\varepsilon}\int \varphi^{m+1}w^m dx + \varepsilon\frac{m}{2}\int |\nabla(\varphi w)|^2(\varphi w)^{m-1}dx \\
&\qquad + C\frac{m}{\varepsilon}\int \varphi^{m+1}w^m dx + C\int \varphi^{m+\theta}w^{m+1/2}dx;
\end{aligned}$$

$$\begin{aligned}
&p\int |\nabla u|^{p-2}\nabla u \cdot \nabla(\varphi w)(\varphi w)^m dx \leqq \frac{\varepsilon m}{2}\int |\nabla(\varphi w)|^2(\varphi w)^{m-1}dx \\
&\qquad + \frac{p^2}{\varepsilon m}\int \varphi^{m+1}w^{m+p}dx;
\end{aligned}$$

$$2\varepsilon(m+1)^{-1}\int \varphi^{-1}\nabla\varphi \cdot \nabla(\varphi^{m+1}w^{m+1})dx \leqq \frac{C\varepsilon}{m+1}\int \varphi^{m+\theta}w^{m+1}dx.$$

And collecting all these bounds, we finally deduce

$$\frac{1}{N\varepsilon}\int \varphi(|\nabla u|^p - C)^{+2}\varphi^m w^m dx \leqq C_p \int w^{m+\frac{p+1}{2}} \varphi^{m+\theta} dx + C\varepsilon \int \varphi^{m+\theta} w^m dx$$

$$+ C\int \varphi^{m+\theta} w^{m+1/2} dx + \frac{C}{\varepsilon}\int \varphi^{m+1} w^m dx + C\frac{m}{\varepsilon}\int \varphi^{m+1} w^m dx$$

$$+ \frac{C\varepsilon}{m+1}\int \varphi^{m+\theta} w^{m+1} dx + \frac{p^2}{\varepsilon m}\int \varphi^{m+1} w^{m+p} dx.$$

To get rid of the last term, we choose m in $\left]\frac{p^2}{N}, \infty\right[$ and we find

$$\int \varphi^{m+1} w^{m+p} dx \leqq C + C\int w^{m+\frac{p+1}{2}} \varphi^{m+\theta} dx.$$

And we conclude choosing $\theta \geqq (m+p)^{-1}\{(p+1)/2 + m(3-p)/2\}$.

Acknowledgement. We wish to thank the referee for useful comments and suggestions.

References

1. Amann, H., Crandall, M.G.: On some existence theorems for semilinear elliptic equations. Ind. Univ. Math. J. **27**, 779–790 (1978)
2. Bensoussan, A., Lions, J.L.: Applications des inéquations variationnelles en contrôle stochastique. Paris: Dunod 1978
3. Bensoussan, A., Lions, J.L.: Contrôle impulsionnel et inéquations quasi-variationnelles. Paris: Dunod 1982
4. Bombieri, E., De Giorgi, E., Miranda, M.: Una maggiorazioni a priori relativa alle ipersuperfici minimali non parametriche. Arch. Rat. Mech. Anal. **32**, 255–267 (1969)
5. Bony, J.M.: Principe du maximum dans les espaces de Sobolev. Cr. Acad. Sci. Paris **265**, 233–236 (1967)
6. Brézis, H.: Semilinear equations in $\mathbb{R}^n$ without condition ar infinity. Appl. Math. Opt. **12**, 271–282 (1984)
7. Crandall, M.G., Lions, P.L.: Remarks on unbounded viscosity solutions of Hamilton-Jacobi equations. Ill. J. Math. **31**, 665–688 (1987)
8. Fleming, W.H., Rishel, R.: Deterministic and stochastic optimal control. Berlin Heidelberg New York: Springer 1975
9. Holland, C.J.: A new energy characterization of the smallest eigenvalue of the Schrödinger equation. Comm. Pure Appl. Math. **30**, 755–765 (1977)
10. Holland, C.J.: A minimum principle for the principal eigenvalue for second-order linear elliptic equations with natural boundary conditions. Comm. Pure Appl. Math. **31**, 509–519 (1978)
11. Krylov, N.V.: Controlled diffusion processes. Berlin Heidelberg New York: Springer 1980
12. Krylov, N.V.: Nonlinear second-order elliptic and parabolic equations. Moscou: Nauk 1985 (in Russian)
13. Ladyshenskaya, O., Ural'tseva, N.: Local estimates for gradients of solutions of nonuniformly elliptic and parabolic equations. Comm. Pure Appl. Math. **23**, 677–703 (1970)
14. Laetsch, T.: On the number of solutions of boundary value problems with convex nonlinearities. J. Math. Anal. Appl. **35**, 389–404 (1971)
15. Lasry, J.M., Lions, P.L.: In preparation
16. Lions, P.L.: Résolution de problèmes elliptiques quasilinéaires. Arch. Rat. Mech. Anal. **74**, 336–353 (1980)

17. Lions, P.L.: On the Hamilton-Jacobi-Bellman equations. Acta Applicandae **1**, 17–41 (1983)
18. Lions, P.L.: Some recent results in the optimal control of diffusion processes. In: Stochastic Analysis, Proceedings in the Taniguchi International Symposium on Stochastic Analysis, Katata, Kyoto (1982). Tokyo: Kinokuniya 1984
19. Lions, P.L.: Quelques remarques sur les problèmes elliptiques quasilinéaires du second ordre. J. Anal. Math. **45**, 234–254 (1985)
20. Lions, P.L.: Viscosity solutions of Hamilton-Jacobi equations and boundary conditions. Preprint
21. Lions, P.L.: Generalized solutions of Hamilton-Jacobi equations. London: Pitman 1982
22. Lions, P.L.: Neumann type boundary conditions for Hamilton-Jacobi equations. Duke Math. J. **52**, 793–820 (1985)
23. Lions, P.L.: A remark on Bony maximum principle. Proc. Am. Math. Soc. **88**, 503–508 (1983)
24. Lions, P.L., Trudinger, N.S.: Linear oblique derivative problems for the uniformly elliptic Hamilton-Jacobi-Bellman equations. Math. Z. **191**, 4–15 (1985)
25. Lions, P.L., Trudinger, N.S.: In preparation
26. Simon, L.: Interior gradient bounds for nonuniformly elliptic equations. Ind. Univ. Math. J. **25**, 821–855 (1976)
27. Simon, L.: Boundary regularity for solutions of the nonparametric least area problem. Ann. Math. **103**, 429–455 (1976)
28. Simon, L.: Global estimates of Hölder continuity for a class of divergence form equations. Arch. Rat. Mech. Anal. **56**, 223–272 (1974)
29. Soner, M.H.: Optimal control with state-space constraints. I. SIAM J. Control Optim. (1986)
30. Urbas, J.I.E.: Elliptic equations of Monge-Ampère type. Thesis, Univ. of Canberra, Australia, 1984

Received March 3, 1988

Math. Ann. 283, 631–643 (1989)

Mathematische
Annalen

Continuous Fields of C^*-Algebras Coming from Group Cocycles and Actions

Marc A. Rieffel*

Department of Mathematics, University of California, Berkeley, CA 94720, USA

Recently I have been attempting to formulate a suitable C^*-algebraic framework for the subject of deformation quantization [3, 19]. Continuous fields of C^*-algebras provide one of the key elements for this framework. The main examples of deformation quantizations which I have constructed up to now in this C^*-algebra framework come from letting either cocycles on groups, or actions of groups, vary. It has thus become necessary to show that one obtains in this way fields of C^*-algebras that are indeed continuous. Since this material is of a general nature, and can be useful in other situations [4–6, 11–13] it has seemed appropriate to give a separate exposition of it, in the present article.

Section 1 of this article contains a review of the published results on continuous fields which we will need, as well as a discussion of the fact that the approach which we will take involves treating upper and lower semi-continuity separately. In Sect. 2 we discuss the continuity of fields of C^*-algebras which arise from varying cocycles on groups, while in Sect. 3 we do the same for actions which vary.

1. Generalities About Continuous Fields

One slight novelty of our approach to continuous fields of C^*-algebras is that we view the continuity of fields as arising from the combination of upper and lower semi-continuity, and we observe that for various constructions these two forms of semi-continuity arise from fairly different considerations. Vaguely speaking, upper semi-continuity tends to arise from the universality of the constructions (as is evident in the extensive literature concerning upper semi-continuous fields, referenced in [9]); whereas lower semi-continuity tends to arise from the presence of natural fields of representations on Hilbert space.

In treating upper semi-continuity, we will find it very useful to use the relationship with disintegration over central subalgebras which has been widely discussed (see references in [9]). We now review briefly some of the facts which we

* Supported in part by National Science Foundation grant DMS 8601900

need. For any C^*-algebra A we let $M(A)$ denote its multiplier algebra. For any locally compact space Ω we let $C_\infty(\Omega)$ denote the algebra of complex-valued continuous functions on Ω vanishing at infinity.

To phrase the results succinctly we make the following definition, motivated by terminology of Fell, as clarified by the second half of Proposition 1 of [18] and, for uper semi-continuous fields, by Proposition 2.3 of [9].

1.1. Definition. Let $\{A_\omega\}$ be a field of C^*-algebras over a locally compact space Ω, and let A be a C^*-algebra (for the supremum norm) of cross-sections of $\{A_\omega\}$. We will say that S is *maximal* if

1) The evaluation maps π_ω from A to each A_ω are surjective.

2) For each $a \in A$ the function $\omega \mapsto \|\pi_\omega(a)\|$ is upper semi-continuous, and vanishes at infinity on Ω.

3) The pointwise product of any element of A with any element of $C_\infty(\Omega)$ is again in A.

It follows from the results in Chap. 2 of [9], notably Proposition 2.3, that a maximal A determines a unique (upper semi-continuous) continuity structure on the field, and is in turn uniquely determined by this continuity structure. We will use the usual imprecise notation of referring to just $\{A_\omega\}$ as a (semi-) continuous field, where the continuity structure is to be inferred from the context. The next proposition is the slight extension of Theorem 2.4 of [9] from compact to locally compact Ω.

We include the brief proof, of the first part along lines due to Varela.

1.2. Proposition. *Let A be a C^*-algebra, and let C be a C^*-subalgebra of the center of $M(A)$. Let Ω denote the spectrum of C, so that we can view C as $C_\infty(\Omega)$. For each $\omega \in \Omega$ let I_ω denote the ideal of elements of C which vanish at ω, let $J_\omega = AI_\omega$ (closure of linear span), and let $A_\omega = A/J_\omega$. Let π_ω denote the homomorphism from A to A_ω. Then for any $a \in A$ the function*

$$\omega \mapsto \|\pi_\omega(a)\|$$

is upper semi-continuous, so that the field $\{A_\omega\}$ of C^*-algebras over Ω is upper semi-continuous. If A is an essential C-module, that is, $A = CA$ (closure of linear span), then A is identified with the maximal algebra of cross sections of $\{A_\omega\}$ by the map $a \mapsto \{\pi_\omega(a)\}$.

Proof. For given $a \in A$ let $\omega_0 \in \Omega$ and $\varepsilon > 0$ be given. Then by the definition of the quotient norm there is a finite sum $t = \sum f_i a_i$ with $f_i \in I_{\omega_0}$ and $a_i \in A$ such that

$$\|\pi_{\omega_0}(a)\| \geqq \|a - t\| - \varepsilon .$$

Choose $g \in C$ with $\|g\| \leqq 1$ such that $g \equiv 1$ in a neighborhood N of ω_0, but such that all the gf_i are small enough so that $\|gt\| < \varepsilon$. Then

$$\begin{aligned}\|\pi_{\omega_0}(a)\| &\geqq \|a-t\| - \varepsilon \geqq \|g(a-t)\| - \varepsilon \geqq \|ga\| - 2\varepsilon \\ &= \|a - (1-g)a\| - 2\varepsilon .\end{aligned}$$

Since $l - g$ is in I_ω for $\omega \in N$, it follows that

$$\|\pi_{\omega_0}(a)\| \geqq \|\pi_\omega(a)\| - 2\varepsilon$$

for $\omega \in N$, which shows the desired upper semi-continuity. The last statement of the proposition is clear from Theorem 2.4 of [9] if Ω is compact. But one reduces immediately to the compact case by adjoining the identity element of $M(A)$ to C (so forming the one-point compactification of Ω), and by noting that the condition $A = CA$ says exactly that the fibre over the point at infinity is $\{0\}$. Q.E.D.

We now give a converse to the above proposition which is a slightly specialized version of Proposition 1.2 of [9].

1.3. Proposition. *Let $\{A_\omega\}$ be a field of C^*-algebras over locally compact Ω, and let A be a maximal C^*-algebra of cross-sections. For each ω let J_ω denote the kernel of the evaluation map π_ω from A to A_ω. Then*

$$J_\omega = I_\omega A$$

(closure of linear span) where I_ω denotes the ideal of elements of $C_\infty(\Omega)$ which vanish at ω.

Proof. Let $\omega_0 \in \Omega$ and $a \in J_{\omega_0}$ be given. Since $\omega \mapsto \|\pi_\omega(a)\|$ is upper semi-continuous and $\pi_{\omega_0}(a) = 0$, there is, for any given $\varepsilon > 0$, a compact neighborhood N of ω_0 such that $\|\pi_\omega(a)\| < \varepsilon$ for $\omega \in N$. Let f be an element of I_{ω_0} taking values between 0 and 1 such that $f(\omega) = 1$ whenever $\|\pi_\omega(a)\| \geqq \varepsilon$ (which is possible since a vanishes at infinity). Let g be an element of $C_\infty(\Omega)$ taking values between 0 and $1/\varepsilon$ which on the union of N with the set where $f(\omega) \geqq \varepsilon$ agrees with $(\sup(f, \varepsilon))^{-1}$, so that $0 \leqq fg \leqq 1$, and $(fg)(\omega) = 1$ if $f(\omega) \geqq \varepsilon$. By hypothesis, $ga \in A$. But a quick calculation shows that

$$\|a - f(ga)\| \leqq \varepsilon .$$

Thus we have approximated a within ε by an element of $I_\omega A$ as desired. Q.E.D.

In contrast to the situation for upper semi-continuity, there appears to be little general discussion in the literature of structures which result in lower semi-continuity of fields of C^*-algebras; and indeed a precise formulation appears difficult to find. But, in vague terms, one can view the approach which we will use as follows. Consider again a subalgebra $C = C_\infty(\Omega)$ in the center of $M(A)$, and assume that $A = CA$. If we take a faithful representation, ϱ, of A, then ϱ extends to $M(A)$ and so to C. We can now try to disintegrate ϱ over C into a field $\{\varrho_\omega\}$ of representations of A. Under favorable circumstances, one can hope that the field $\{\varrho_\omega(A)\}$ of C^*-algebras will give a decomposition of A. Examples show that in this situation the continuity which one is likely to obtain is lower semi-continuity, that is, for each $a \in A$ the function $\omega \mapsto \|\varrho_\omega(a)\|$ will be lower semi-continuous.

The following example illustrates well the above ideas, and suggests the phenomena which may arise also in the non-commutative case.

1.4. Example. Let Ω denote the interval $[-1, 1]$ of the real line R, let S be the interval $[1, \infty)$, and let $M = \Omega \times S$. Let $A = C_b(M)$, the algebra of continuous bounded functions on M. We wish to examine how A can be expressed as a field of C^*-algebras over Ω. Let $C = C(\Omega)$, viewed as consisting of functions on M constant in the S direction. Thus $C \subset M(A) = A$. Following the method of Propositions 1.2 and 1.3, we form I_ω and $A_\omega = A/I_\omega A$, which gives a field which is upper semi-continuous. Let ϕ be the function in A which has value 0 for $\omega \leqq 0$, has value 1 if both $\omega > 0$ and $s \geqq 1/\omega$, and for each fixed $s > 0$ interpolates linearly for ω between 0 and $1/s$. Thus ϕ has value 0 everywhere on $\{0\} \times S$, but for any $\omega > 0$ there are values of s such that $\phi(\omega, s) = 1$. Then it is easily seen that $\omega \mapsto \|\pi_\omega(\phi)\|$ is not lower semi-continuous at $\omega = 0$. On the other hand, A has the evident faithful representation on $L^2(M)$, where planar Lebesgue measure is used on M. This representation can be disintegrated in a very natural way by the representations ϱ_ω on $L^2(S)$ defined by

$$(\varrho_\omega(\psi)\xi)(s) = \psi(\omega, s)\xi(s)$$

for $\psi \in A$ and $\xi \in L^2(S)$. It is easily seen that $\omega \mapsto \|\varrho_\omega(\psi)\|$ is lower semi-continuous, but for the special function ϕ defined above, $\omega \mapsto \|\varrho_\omega(\phi)\|$ is not upper semi-continuous at $\omega = 0$.

2. Fields from Cocycles

We will now discuss fields of C^*-algebras which arise when one varies two-cocycles on a locally compact group G, as this will be useful in [27]. The direction of my investigation of this topic was originally motivated by Theorem 1 of [2]. We will actually consider a slightly more general situation than needed for [27], namely that in which there is a fixed action, α, of G on a C^*-algebra A, and the cocycles take values in the group $UZM(A)$ of unitary elements of the center of the multiplier algebra, $M(A)$, of A. This extra generality is useful in treating certain C^*-algebras which have arisen recently in solid state physics [4], and in particular in the quantum Hall effect. (See [5] and Theorem 12 of [6], as well as [30]; also see [11, 13] for a different direction.)

For the case in which G is discrete, the corresponding twisted group C^*-algebras are defined and studied in thorough detail in [31]. On the other hand, the literature concerning the case in which G is not discrete is in a somewhat less satisfactory state. There is much discussion of the twisted L^1-algebras under various hypotheses. (See [23] and [28] and the references they give.) A treatment of twisted crossed product C^*-algebras has very recently been given for the separable case in [24], but I am not aware of any treatment of the twisted C^*-algebras in the general case. It seems fairly clear that all could be recast into Fell's elegant framework of C^*-algebraic bundles [14]. But this would require an extensive discussion, which is not appropriate in the present paper, both because the extra generality is not needed for our immediate purposes in [27], and because in applications one is usually presented with cocycles rather than a bundle, and the coyclces usually have sufficient continuity to make it apparent that all works smoothly (and the connection with Fell's theory is known by [8] and [20]). Consequently, we will here be somewhat cavalier about these technicalities, and will work as though we know that all works smoothly at the C^*-algebra level. However, for ease of exposition in dealing with questions of measurability, we will assume that G, if it is not discrete, is second countable. Again, with more work this hypothesis could probably be substantially eliminated by using the techniques in [16] and [20]. Frequently we will only discuss explicitly the second countable case, letting the reader check [31] for the general discrete case. Finally, let us mention that some treatments (see [20] and its references) do not even require the cocycles to be center-valued, but [31] does require this, so we restrict to this case.

We will throughout let Ω be a locally compact space which will serve as the base space for our fields, and also at this point as the parameter space for cocycles. We let $C_\infty(\Omega, A)$ denote the C^*-algebra of continuous A-valued functions on Ω vanishing at infinity. The following definition is more-or-less that made in Sect. 1 of [22], if we view the cocycle σ as having values in $UZM(C_\infty(\Omega, A))$.

2.1. Definition. By a *continuous field over Ω of α-cocycles* of G we will mean a function σ on $G \times G \times \Omega$ with values in $UZM(A)$ such that

1) If we fix $\omega \in \Omega$ then σ is a normalized α-cocycle on G, that is,

$$\sigma(x, yz; \omega)\alpha_x(\sigma(y, z; \omega)) = \sigma(xy, z; \omega)\sigma(x, y; \omega)$$

for x, y, $z \in G$ (where α here is the extension of α to $M(A)$), and

$$\sigma(x, e; \omega) = 1 = \sigma(e, x; \omega)$$

where e denotes the identity element of G.

2) If we fix x, $y \in G$, then σ is continuous on Ω.

3) For any $f \in C_\infty(\Omega, A)$ the function

$$(x, y) \mapsto f(\cdot)\sigma(x, y; \cdot)$$

from $G \times G$ to $C_\infty(\Omega, A)$ is Bochner measurable.

An example which is important for [27], consists of letting A be the complex numbers, letting $G = \mathbb{R} \times \mathbb{R}$ and $\Omega = \mathbb{R}$, and defining σ by

$$\sigma((r, s), (t, u); \omega) = \exp(i\omega(ru - st)).$$

Returning to the general case, we recall (Theorem 2.2 of [8]) that the corresponding twisted L^1-algebra is defined as follows. Choose a left Haar measure on G. For Φ, $\Psi \in L^1(G, C_\infty(\Omega, A))$ one defines the twisted convolution by

$$(\Phi * \Psi)(x) = \int \Phi(y)\alpha_y(\Psi(y^{-1}x))\sigma(y, y^{-1}x; \cdot)dy$$

and the involution by

$$\Phi^*(x) = \alpha_x(\Phi(x^{-1})^*)\sigma(x, x^{-1}; \cdot)^*\Delta(x^{-1}),$$

where Δ is the modular function of G. We will denote the resulting Banach $*$-algebra by $L^1(G, \Omega, A, \sigma)$, leaving the action α to be understood.

For any $\omega \in \Omega$ let I_ω denote the ideal in $C_\infty(\Omega)$ consisting of functions vanishing at ω, and let J_ω denote the corresponding ideal of $C_\infty(\Omega, A)$. Then α can be viewed as giving an action on J_ω, and σ can be viewed as having values on $UZM(J_\omega)$, so that exactly as above we can form $L^1(G, J_\omega, \sigma)$, which is a closed ideal of $L^1(G, \Omega, A, \sigma)$. Finally, let σ_ω denote $\sigma(\cdot, \cdot; \omega)$, a cocycle with values in $UZM(A)$, so that we can define $L^1(G, A, \sigma_\omega)$. There is an evident evaluation homomorphism from $L^1(G, \Omega, A, \sigma)$ to $L^1(G, A, \sigma_\omega)$ whose kernel is exactly $L^1(G, J_\omega, \sigma)$. From Satz 1 of [21] one sees that $L^1(G, A, \sigma_\omega)$ has the corresponding quotient norm. Thus from the exact sequence

$$0 \to J_\omega \to C_\infty(\Omega, A) \to A \to 0$$

we obtain, in the strongest sense, the exact sequence

$$0 \to L^1(G, J_\omega, \sigma) \to L^1(G, \Omega, A, \sigma) \to L^1(G, A, \sigma_\omega) \to 0,$$

for each fixed $\omega \in \Omega$.

We now define $C^*(G, \Omega, A, \sigma)$ to be the enveloping C^*-algebra of $L^1(G, \Omega, A, \sigma)$, and in the same way we define $C^*(G, J_\omega, \sigma)$ and $C^*(G, A, \sigma_\omega)$. Now from 2.29 of [31] we have:

2.2. Proposition. *Let B be a Banach $*$-algebra possessing a bounded approximate identity, and let J be a closed $*$-ideal (2-sided) of B also possessing a bounded approximate identity. Then, letting $C^*(\ )$ denote enveloping C^*-algebras, we obtain the short exact sequence of C^*-algebras*

$$0 \to C^*(J) \to C^*(B) \to C^*(B/J) \to 0.$$

However, the situation as to whether every $L^1(G, \Omega, A, \sigma)$ has a bounded approximate identity is somewhat murky. According to the discussion in the first paragraph on p. 311 of [28], it can be shown that if the action α is trivial this is always so. But it would be nice to have a proof of this in print, for one might well then be able to see that the proof also works for non-trivial α. (But careful – in the notation of [28] α is the cocycle, non-trivial.) At any rate, it seems highly likely that almost all specific examples of significant interest will be smooth at the identity as defined in Definition 2 of [7], and in this case the combination of Theorem 1 of [7] with the results of [14], notably Proposition 8.2, shows that approximate identities exist. (Alternately, [20] can be combined with [14]. See also Remark 2.6 of [24].) So we will now proceed as if this were always so. Thus from Proposition 2.2 we obtain:

2.3. Proposition. *With notation and caveats as above, the sequence*

$$0 \to C^*(G, J_\omega, \sigma) \to C^*(G, \Omega, A, \sigma) \to C^*(G, A, \sigma_\omega) \to 0$$

is exact for every $\omega \in \Omega$.

Now it is clear that $C_\infty(\Omega)$ acts as an algebra of (central) multipliers on $L^1(G, \Omega, A, \sigma)$, and that

$$L^1(G, J_\omega, \sigma) = L^1(G, \Omega, A, \sigma) I_\omega.$$

It is easily seen from this that $C_\infty(\Omega)$ also acts as an algebra of central multipliers of $C^*(G, \Omega, A, \sigma)$, and that

$$C^*(G, J_\omega, \sigma) = C^*(G, \Omega, A, \sigma) I_\omega.$$

We thus find that we are exactly in the right position to apply Proposition 1.2 to obtain:

2.4. Theorem. *With notation and caveats as above, let σ be a continuous field over Ω of α-cocycles on the discrete, or second countable locally compact, group G. Then the field of C^*-algebras $\{C^*(G, A, \sigma_\omega)\}$ over Ω, with the continuity structure coming from $C^*(G, \Omega, A, \sigma)$, is upper semi-continuous.*

We now turn to the question of lower semi-continuity. With notation as above, pick any faithful representation of A on a Hilbert space H, and any positive Borel measure on Ω of full support, and consider the corresponding faithful representation, p, of $C_\infty(\Omega, A)$ on $L^2(\Omega, H)$. We can then induce (see [23] or [8]) this to a representation of $L^1(G, \Omega, A, \sigma)$. The corresponding C^*-norm and completion give, by definition, the reduced C^*-algebra $C_r^*(G, \Omega, A, \sigma)$. Again, $C_\infty(\Omega)$ acts as an algebra of multipliers, and there is a natural choice of disintegration of the representation with respect to $C_\infty(\Omega)$ given by composing the evaluation homomorphism π_ω with the representations of $L^1(G, A, \sigma_\omega)$ on $L^2(G, H)$ obtained by inducing the representation of A on H. More specifically, define a twisted covariant pair (U, μ) of representations of $(G, C_\infty(\Omega, A))$ on $L^2(G, L^2(\Omega, H))$ by

$$(U_y\eta)(x) = p(\sigma(x^{-1}, y))\eta(y^{-1}x)$$
$$(\mu(a)\eta)(x) = p(\alpha_x^{-1}(a))\eta(x)$$

for $x, y \in G$ and $a \in A$. It is easily verified that (U, μ) satisfies Definition 2.1 of [8], aside from questions of measurability in the non-separable case. Consequently, by Theorem 3.3 of [8] the integrated form, ϱ, of (U, μ), defined by

$$\varrho(\Phi)\eta = \int \mu(\Phi(y))U_y\eta\, dy$$

for $\Phi \in L^1(G, \Omega, A, \sigma)$, defines a $*$-representation of $L^1(G, \Omega, A, \sigma)$, the induced representation. We let ϱ_ω denote the corresponding representation of $L^1(G, \Omega, A, \sigma)$ or $C_r^*(G, \Omega, A, \sigma)$ on $L^2(G, H)$ obtained by the same formulas as above except evaluating at ω. It is clear that the range of ϱ_ω on $C_r^*(G, \Omega, A, \sigma)$ will be essentially $C_r^*(G, A, \sigma_\omega)$, so that we are considering the field $\{C_r^*(G, A, \sigma_\omega)\}$. At this juncture the measure originally chosen on Ω is seen to be irrelevant – we only introduced it to put matters into the general framework sketched earlier.

2.5. Theorem. *With notation and caveats as above, let σ be a continuous field over Ω of α-cocycles on the discrete, or second countable locally compact, group G. Then the field of C^*-algebras $\{C_r^*(G, A, \sigma_\omega)\}$ over Ω, with the continuity structure coming from $C_r^*(G, \Omega, A, \sigma)$, is lower semi-continuous.*

Proof. We show first that for any $\Phi \in L^1(G, \Omega, A, \sigma)$, the function $\omega \mapsto \varrho_\omega(\Phi)$ is continuous for the strong operator topology from $L^2(G, H)$. By the usual 3ε argument, it suffices to show this for $\Phi \in L_c^\infty(G, \Omega, A)$, the space of bounded measurable functions of compact support on G with values in $C_\infty(\Omega, A)$, since it is dense in $L^1(G, \Omega, A, \sigma)$. Let $\xi \in L^2(G, H)$, and let $\omega_0 \in \Omega$. For fixed $x \in G$, as ω converges to ω_0 the function

$$y \mapsto \mu(\Phi(y, \omega))p(\sigma(x^{-1}, y, \omega))\xi(y^{-1}x)$$

converges pointwise to the corresponding function for ω_0, and is dominated by the function

$$y \mapsto \|\Phi(y)\|_\infty \; \|\xi(y^{-1}x)\|$$

in $L^2(G)$. Thus, by the dominated convergence theorem, $(\varrho_\omega(\Phi)\xi)(x)$ converges to $(\varrho_{\omega_0}(\Phi))\,\xi(x)$ for each $x \in G$. But these functions are in turn dominated by the function

$$\|\Phi(\cdot)\|_\infty * \|\xi\|\,,$$

which is in $L^2(G)$, where $*$ denotes here ordinary convolution on G. Thus, again by the dominated convergence theorem, $\varrho_\omega(\Phi)\xi$ converges in $L^2(G, H)$ to $\varrho_{\omega_0}(\Phi)\xi$. Thus

$\varrho_\omega(\Phi)$ converges to $\varrho_{\omega_0}(\Phi)$ in the strong operator topology. But it is easily seen that this strong operator topology convergence implies that $\omega \mapsto \|\varrho_\omega(\phi)\|$ is lower semi-continuous. Q.E.D.

Suppose now that G is amenable. For G discrete it is proven in 5.1 of [31] that reduced and full twisted crossed product C^*-algebras then coincide, while in Theorem 7.7.7 of [25] this is shown for crossed product algebras with arbitrary G. By combining the techniques used in the proofs of these two cases, it seems very likely that one can extend this result to various varieties of more general twisted group C^*-algebras. But since this matter is not of central importance to the present paper, and since there are many other situations where the reduced and full twisted crossed product C^*-algebras coincide, we will content ourselves with simply formulating:

2.6. The Amenability Hypothesis. *We will say that a collection $\{G, A, \sigma, \alpha\}$ consisting of an action α of a locally compact group G on a C^*-algebra A together with a corresponding cocycle σ, satisfies the amenability hypothesis if $C_r^*(G, A, \alpha, \sigma) = C^*(G, A, \alpha, \sigma)$.*

Combining this with Theorems 2.4 and 2.5, we obtain:

2.7. Corollary. *With notation and caveats as earlier, assume that each $\{G, A, \sigma_\omega\}$ satisfies the amenability hypothesis. Then the field $\{C^*(G, A, \sigma_\omega)\}$ over Ω is continuous.*

When G is discrete and $A = \mathbb{C}$ (so α is trivial) one has the following universal formulation of the above results. Let $\Gamma = \Gamma_G$ be the set of normalized cocycles on G with values in the set T of complex numbers of modulus one. Equip Γ with the topology of pointwise convergence. Since the conditions defining a normalized cocycle are closed conditions, Γ is compact by Tychonoff's theorem. Then the tautological field $\gamma \mapsto \gamma$ is a continuous field over Γ of cocycles on G. Combining Theorems 2.4 and 2.5 with Corollary 2.7 and the fact that, according to [31], the amenability hypothesis is always true for amenable discrete groups, one obtains:

2.8. Corollary. *Let G be discrete, and let Γ be the compact space of all normalized T-valued cocycles on G. Then*

a) *$\{C^*(G, \gamma)\}$ is an upper semi-continuous field on Γ.*
b) *$\{C_r^*(G, \gamma)\}$ is a lower semi-continuous field on Γ.*
c) *If G is amenable, then $\{C^*(G, \gamma)\} = \{C_r^*(G, \gamma)\}$ is a continuous field over Γ.*

It is clear that the above corollary remains true if Γ is replaced by any closed subset of Γ. Because the set of skew bicharacters of an Abelian discrete group is such a closed subset, one obtains immediately from the above corollary a proof of the statement in [12] that non-commutative tori form a continuous field (and a proof of the special case of this statement applying to irrational and rational rotation C^*-algebras, used in [13], which is also handled by [2]).

Actually, it seems reasonable to me to conjecture that $\{C_r^*(G, \gamma)\}$ is a continuous field even when G is not amenable (and a corresponding fact for non-discrete G). But an attempt to imitate for $\{C_r^*(G, \gamma)\}$ the proof of Theorem 2.4 founders on the fact that it seems not to be known whether the exact sequence used in that proof is also exact for the reduced algebras. Antony Wassermann pointed out to me that the results of [17] show that in a closely related situation such exactness actually does fail. Specifically, if G is the free group on two generators, then $C^*(G)$ is not an exact C^*-algebra, in the sense that there are exact sequences

$$0 \to I \to A \to Q \to 0$$

of C^*-algebras for which

$$0 \to I \underset{\min}{\otimes} C^*(G) \to A \underset{\min}{\otimes} C^*(G) \to Q \underset{\min}{\otimes} C^*(G) \to 0$$

is not exact. But for this to happen A must be non-commutative (and it deals with the *full* group C^*-algebra), while in the conjecture indicated above the relevant algebra $C(\Gamma)$ is commutative. We also mention that no example seems to be known of a discrete group G whose *reduced* algebra $C_r^*(G)$ is not exact, although it is likely that such examples exist. If G is again the free group on two generators, then Proposition 5.3 of [10] and the discussion following show that $C_r^*(G)$ is exact. (But see Proposition 4.5 of [32].) In general, it is, of course, also interesting to ask if $\{C^*(G, \gamma)\}$ also might always be continuous.

3. Fields from Actions

We now turn to similar questions for crossed product algebras [25], where now we let the action vary. There is some overlap here with the results in [29], but the techniques are quite different.

3.1. Definition. By an *upper semi-continuous field of actions on C^*-algebras* for a locally compact group G we mean an upper semi-continuous field $\{A_\omega\}$ of C^*-algebras over a locally compact space Ω, with maximal C^*-algebra of sections A, and evaluation maps π_ω, together with an action α of G on A which carries the kernel K_ω of each π_ω into itself, and so induces an action on each A_ω. If the field $\{A_\omega\}$ is actually continuous, then we will say that the field of actions is continuous.

3.2. Theorem. *Let α be an upper semi-continuous field of actions of G on the field $\{A_\omega\}$ of C^*-algebras. For each $\Phi \in L^1(G, A)$ and each $\omega \in \Omega$, let $\|\Phi\|_\omega$ denote the norm of the image of Φ in the crossed product algebra $C^*(G, A_\omega; \alpha)$. Then the function*

$$\omega \mapsto \|\Phi\|_\omega$$

is upper semi-continuous, so that $\{C^*(G, A_\omega; \alpha)\}$ is an upper semi-continuous field of C^*-algebras over Ω. Furthermore, $C^*(G, A, \alpha)$ is identified with this field's maximal C^*-algebra of sections.

Proof. Much as in the proof of Proposition 2.3, it follows from Proposition 2.2 together with Satz 1 of [21] that the sequence

$$0 \to C^*(G, K_\omega; \alpha) \to C^*(G, A; \alpha) \to C^*(G, A_\omega; \alpha) \to 0$$

is exact. As before, for $\omega \in \Omega$ let I_ω denote the ideal in $C_\infty(\Omega)$ of functions vanishing at ω. By Proposition 1.3 we have $K_\omega = I_\omega A$. Now it is clear from the hypotheses that the elements of $C_\infty(\Omega) \subsetneqq M(A)$ are invariant under α. It follows easily that

$$C^*(G, K_\omega; \alpha) = I_\omega C^*(G, A; \alpha),$$

so that

$$C^*(G, A_\omega; \alpha) = C^*(G, A; \alpha)/I_\omega C^*(G, A; \alpha).$$

But then the upper semi-continuity and the identification of $C^*(G, A, \alpha)$ as the maximal C^*-algebra of sections follow from Proposition 1.2. Q.E.D.

We need a corresponding lower semi-continuity result. I have not seen how to obtain this without imposing stronger hypotheses on the field involved. These can be motivated by observing that a key feature in the proof of Theorem 2.5 was that all the fiber algebras ended up acting on the same fixed Hilbert space, so that the strong operator topology could be employed.

3.3. Definition. Let $\{A_\omega\}$ be an upper semi-continuous field of C^*-algebras over Ω, with A its maximal C^*-algebra of sections. We will say that the field is *Hilbert continuous* if there is a fixed Hilbert space H and, for each $\omega \in \Omega$, a faithful representation σ_ω of A_ω on H (which thus also defines a representation of A) such that for all $a \in A$ the function

$$\omega \mapsto \sigma_\omega(\pi_\omega(a))$$

is continuous for the strong operator topology.

Since the norm on operators is lower semi-continuous with respect to the strong operator topology, it follows that a Hilbert-continuous field is, in fact, continuous (since the σ_ω are faithful).

3.4. Question. How can one characterize those continuous fields of C^*-algebras which are actually Hilbert-continuous?

In preparation for the last part of the next theorem, we remark that there are many situations where $C^*(G, A, \alpha) = C_r^*(G, A, \alpha)$ even when G is not amenable. (See Theorem 4.5 and Proposition 4.8 of [1].)

3.5. Theorem. *Let Ω, $\{A_\omega\}$, A, K_ω, G and α be as in Theorem 3.2, except assume in addition that the field is Hilbert-continuous. For each $\phi \in C_c(G, A)$ and each $\omega \in \Omega$ let $\|\phi\|_\omega^r$ denote the norm of the image of ϕ in $C_r^*(G, A_\omega, \alpha)$. Then*

a) *For each $\phi \in C_c(G, A_\omega, \alpha)$ the function $\omega \mapsto \|\phi\|_\omega^r$ is lower semi-continuous.*

b) *Suppose that $C_r^*(G, A_\omega, \alpha) = C^*(G, A_\omega, \omega)$ for all ω (for instance if G is amenable). Then the function*

$$\omega \mapsto \|f\|_\omega = \|f\|_\omega^r$$

is continuous, $C_r^*(G, A, \alpha) = C^*(G, A, \alpha)$, and the field $\{C^*(G, A_\omega, \alpha)\}$, which is again a Hilbert-continuous field, has $C^*(G, A, \alpha)$ as its maximal C^*-algebra of sections.

Proof. For each $\omega \in \Omega$ let $C_\omega = C_r^*(G, A_\omega, \alpha)$. Let H and $\{\sigma_\omega\}$ be as in the definition of a Hilbert-continuous field. For each $\omega \in \Omega$ let ϱ_ω be the representation of C_ω on $L^2(G, H)$ obtained by inducing the representation σ_ω to C_ω, so that, essentially by definition, $\|\phi\|_\omega^r = \|\varrho_\omega(\phi)\|$ for ϕ in $C_c(G, A)$. We will show that for any ϕ in $C_c(G, A)$ the function $\omega \mapsto \varrho_\omega(\phi)$ is strong operator continuous. Now for any $\xi \in L^2(G, H)$ we have, by definition (see [25]),

$$\varrho_\omega(\phi)\xi = \int_G \tilde{\sigma}_\omega(\phi(y)) L_y \xi \, dy,$$

where L denotes the operator of left translation by y, and where for $a \in A$ we define $\tilde{\sigma}_\omega(a)$ by

$$(\tilde{\sigma}_\omega(a)\xi)(x) = \sigma_\omega(\alpha_x^{-1}(a))\xi(x).$$

But

$$\|\tilde{\sigma}(\phi(y))L_y\xi\| \leqq \|\phi(y)\| \; \|\xi\|,$$

and the right-hand side is integrable since ϕ is continuous of compact support. Furthermore, it is easily seen by using the Lebesgue dominated convergence theorem that for each fixed y the integrands converge as ω goes to ω_0, by the assumption on the σ_ω's. Hence, again by the Lebesgue dominated convergence theorem, $\varrho_\omega(\phi)\xi$ converges to $\varrho_{\omega_0}(\phi)\xi$ as ω goes to ω_0. This shows the strong operator continuity of the representations ϱ_ω as functions of ω. But from this, the lower semi-continuity of $\|\varrho_\omega(\phi)\|$ for any given ϕ follows immediately. Now σ_ω is a faithful representation of A_ω, by assumption, and so ϱ_ω is a faithful representation of C_ω, essentially by the definition of the reduced crossed product [25]. It follows that $\|\phi\|_\omega$ is a lower semi-continuous function of ω for each ϕ, which proves a).

If $C_r^*(G, A_\omega, \alpha) = C^*(G, A_\omega, \alpha)$ for each ω, then, of course, $\|\phi\|_\omega^r = \|\phi\|_\omega$ for each ω, so that these are both continuous. Furthermore, the proof of part a) given above provides the fixed Hilbert space $L^2(G, H)$ and the representations ϱ_ω which play the role of the new σ_ω's to give the Hilbert-continuity of the field. We next show that $C_r^*(G, A, \alpha) = C^*(G, A, \alpha)$. Now $C_r^*(G, A, \alpha)$ is a quotient of $C^*(G, A, \alpha)$. Thus it suffices to show that every irreducible representation of $C^*(G, A, \alpha)$ factors through $C_r^*(G, A, \alpha)$. So let σ be an irreducible representation of $C^*(G, A, \alpha)$. Since $C_\infty(\Omega)$ is in the center of the multiplier algebra $M(A)$ of A and is α-invariant, it is in the center of $M(C^*(G, A, \alpha))$. The extension of σ to $M(C^*(G, A, \alpha))$ must carry $C_\infty(\Omega)$ to scalar multiples of the identity operator since σ is irreducible. Thus σ on $C_\infty(\Omega)$ corresponds to evaluation at some point, say ω of Ω. Let I_ω denote, as in the proof of Theorem 2.4, the ideal of functions in $C_\infty(\Omega)$ vanishing at ω. Then, as in the proof of that theorem, the kernel K_ω of the homomorphism of A onto A_ω is $I_\omega A$. Clearly $I_\omega A$ is in the kernel of σ (extended to $M(C^*(G, A, \alpha))$), and so $C^*(G, K_\omega, \alpha)$ is in the kernel of σ. Thus σ can be considered to be a representation of the quotient, that is, of $C^*(G, A_\omega, \alpha)$. But by assumption $C^*(G, A_\omega, \alpha) = C_r^*(G, A_\omega, \alpha)$. Now $C_r^*(G, A_\omega, \alpha)$ is a quotient of $C_r^*(G, A, \alpha)$, even if the corresponding kernel is in general hard to describe. Thus σ gives a representation of $C_r^*(G, A, \alpha)$. But it is easily seen that on the dense subalgebra $C_c(G, A)$, this σ, as representation of $C_r^*(G, A, \alpha)$, agrees with the original σ, so that the original σ factors through $C_r^*(G, A, \alpha)$ as desired. Thus $C_r^*(G, A, \alpha) = C^*(G, A, \alpha)$. Finally, the last part of Theorem 3.2 applies to show that $C^*(G, A, \alpha)$ is exactly the maximal C^*-algebra of sections of the field $\{C^*(G, A_\omega, \alpha)\}$. Q.E.D.

3.6. Corollary. *Let Ω be a locally compact space, and for each $\omega \in \Omega$ let α^ω be an action of the locally compact group G on a fixed C^*-algebra B, such that for each $x \in G$ and $b \in B$ the function $\omega \mapsto \alpha_x^\omega(b)$ is norm continuous on Ω. For each $\phi \in C_c(G, B)$ let $\|\phi\|_\omega$ and $\|\phi\|_\omega^r$ denote the norms of ϕ in $C^*(G, B, \alpha^\omega)$ and $C_r^*(G, B, \alpha^\omega)$ respectively.*

a) *For each ϕ the function $\omega \mapsto \|\phi\|_\omega$ is upper semi-continuous.*

b) *For each ϕ the function $\omega \mapsto \|\phi\|_\omega^r$ is lower semi-continuous.*

c) *Hence if $C_r^*(G, B, \alpha^\omega) = C^*(G, B, \alpha^\omega)$ for all ω (for instance if G is amenable), then the function $\omega \mapsto \|\phi\|_\omega = \|\phi\|_\omega^r$ is continuous. Furthermore, in this case $\{C^*(G, B, \alpha^\omega)\}$ is a Hilbert-continuous field of C^*-algebras, and if $A = C_\infty(\Omega, B)$, then $\{\alpha^\omega\}$ defines an action α of G on A such that $C_r^*(G, A, \alpha) = C^*(G, A, \alpha)$, and $C^*(G, A, \alpha)$ can be identified with the maximal C^*-algebra for $\{C^*(G, B, \alpha^\omega)\}$.*

Proof. Of course, the action α on A is defined by

$$(\alpha_x(f))(\omega) = \alpha_x^\omega(f(\omega))$$

for $f \in A$. A simple argument using the hypotheses on $\{\alpha^\omega\}$ shows that $\alpha_x(f) \in A$. A straightforward uniform continuity argument on compact subsets of Ω shows that α is strongly continuous, so defines an action of G on A. For each $\omega \in \Omega$ evaluation at ω gives a homomorphism of A onto B which carries α to α^ω. Thus we are exactly in the setting of Theorem 3.2, so that part a) follows immediately from Theorem 3.2.

We are dealing here essentially with the constant field over Ω with fiber B. This is clearly a Hilbert-continuous field. Thus, parts a) and b) of Theorem 3.4 immediately imply parts b) and c) here. Q.E.D.

We remark that Lemma 2 of [11] is an immediate consequence of the above corollary.

There seems little doubt that there is a joint generalization of the two main results of this section to the case in which one has cocycles and actions both of which are varying. Since we have no present use in mind for such a result, it seemed best to wait until a significant application arises before treating this situation, but presumably the proof can be modeled on the proofs given above.

As indicated in the introduction, interesting applications of the results of this section will be given in [27]. A relation between the results of this section and proper actions will be given in [26].

References

1. Anantharaman-Delaroche, C.: Systèm dynamiques non commutatifs et moyenabilité. Math. Ann. **279**, 297–315 (1987)
2. Anderson, J., Paschke, W.: The rotation algebra. MSRI preprint
3. Bayen, F., Flato, M., Fronsdal, C., Lichnerowicz, A., Sternheimer, D.: Deformation theory and quantization. I, II. Ann. Phys. **110**, 61–110; 111–151 (1978)
4. Bellissard, J.: K-theory of C^*-algebras in solid state physics. Statistical mechanics and field theory, mathematical aspects, (Lect. Notes Phys., Vol. 257, 99–156.) Berlin Heidelberg New York: Springer 1986
5. Belissard, J.: Ordinary quantum Hall effect and non-commutative cohomology. Proceedings Bad Schandau Conference on Localization, Leipzig: Teubner (to appear)
6. Belissard, J.: C^*-algebras in solid state physics, $2D$ electrons in a uniform magnetic field. Preprint
7. Busby, R.C.: On the equivalence of twisted group algebras and Banach *-algebraic bundles. Proceedings A.M.S. **37**, 142–148 (1973)
8. Busby, R.C., Smith, H.A.: Representations of twisted group algebras, Trans. A.M.S. **149**, 503–537 (1970)

9. Dupré, M.J., Gillette, R.M.: Banach bundles, Banach modules and automorphisms of C^*-algebras. Research Notes in Math. **92**. London: Pitman 1983
10. Effros, E.G., Haagerup, U.: Lifting problems and local reflexivity for C^*-algebras. Duke Math. J. **52**, 103–128 (1985)
11. Elliott, G.A.: Gaps in the spectrum of an almost periodic Schrödinger operator. C.R. Math. Rep. Acad. Sci. Canada **4**, 255–259 (1982)
12. Elliott, G.A.: On the K-theory of the C^*-algebra generated by a projective representation of a torsion-free discrete abelian group. Operator algebras and group representations, vol. 1, pp. 157–184. London: Pitman 1984
13. Elliott, G.A.: Gaps in the spectrum of an almost periodic Schrödinger operator. II. Geometric methods in operator algebras, Araki, H., Effros, E.G. (eds.), pp. 181–191. London: Longman 1986
14. Fell, J.M.G.: An extension of Mackey's method to Banach *-algebraic bundles. Mem. A.M.S. **90** (1969)
15. Fell, J.M.G.: Induced representations and Banach *-algebraic bundles, (Lecture Notes Math. **582**). Berlin Heidelberg New York: Springer 1977
16. Kehlet, E.T.: Cross sections for quotient maps of locally compact groups. Math. Scand. **55**, 152–160 (1984)
17. Kirchberg, E.: The Fubini theorem for exact C^*-algebras. J. Oper. Theory **10**, 3–8 (1983)
18. Lee, R.-Y.: On the C^*-algebras of operator fields. Indiana U. Math. J. **25**, 303–314 (1976)
19. Lichnerowicz, A.: Deformations and quantization. Geometry and physics. Modugno, M., ed., pp. 103–116. Bologna: Pitagory 1983
20. Leinert, M.: Fell-Bündel und verallgemeinerte L^1-Algebren. J. Funct. Anal. **22**, 323–345 (1976)
21. Leptin, H.: Verallgemeinerte L^1-Algebren. Math. Ann. **159**, 51–76 (1965)
22. Leptin, H.: Verallgemeinerte L^1-Algebren und projektive Darstellungen lokal kompakter Gruppen. Invent. Math. **3**, 257–281 (1967)
23. Leptin, H.: Darstellungen verallgemeinerter L^1-Algebren. II. Lecture Notes Math. 247, 251–307. Berlin Heidelberg New York: Springer 1972
24. Packer, J.A., Raeburn, I.: Twisted crossed products of C^*-algebras. Preprint
25. Pedersen, G.K.: C^*-algebras and their automorphism groups. Lond. Math. Soc. Monographs 14. London: Academic Press 1979
26. Rieffel, M.A.: Proper actions of groups on C^*-algebras. Preprint
27. Rieffel, M.A.: Deformation quantization for Heisenberg manifolds. Preprint
28. Smith, H.A.: Central twisted group algebras. Trans. A.M.S. **238**, 309–320 (1978)
29. Williams, D.P.: The structure of crossed products by smooth actions. J. Aust. Math. Soc. Ser. A (to appear)
30. Xia, J.: Geometric invariants of the quantum Hall effect. Preprint
31. Zeller-Meier, G.: Produits croisés d'une C^*-algèbre par un group d'automorphisms. J. Math. Pures Appl. **47**, 101–239 (1968)
32. Renault, J.: A groupoid approach to C^*-algebras. (Lecture Notes Math. 793). Berlin Heidelberg New York: Springer 1980

Received July 21, 1988

Math. Ann. 283, 645–655 (1989)

Mathematische Annalen

Common Fixed Points of Commuting Holomorphic Maps

Marco Abate
Scuola Normale Superiore, Piazza Cavalieri 7, I-56100 Pisa, Italy

0. Introduction

In 1964, Shields proved the following theorem:

Theorem 0.1 (Shields [S]). *Let Δ be the unit disk in $\mathbf{C}$ and $\mathscr{F}$ a family of holomorphic functions of Δ into itself, continuous up to the boundary of Δ. Assume that $f \circ g = g \circ f$ for every $f, g \in \mathscr{F}$. Then there exists a point $\tau \in \bar{\Delta}$ such that $f(\tau) = \tau$ for every $f \in \mathscr{F}$.*

In other words, a *commuting family* $\mathscr{F}$ of holomorphic functions (i.e., such that $f \circ g = g \circ f$ for every $f, g \in \mathscr{F}$) continuous up to the boundary always admits a *common fixed point* (i.e., a point τ such that $f(\tau) = \tau$ for all $f \in \mathscr{F}$) in $\bar{\Delta}$. This result is a feature of the holomorphic structure, and not some sort of consequence of Brouwer's theorem: indeed, there are examples of commuting continuous functions mapping the closed unit interval $[-1, 1] \subset \mathbf{R}$ into itself without common fixed points (see Boyce [Bo] and Huneke [Hu]).

The first generalization of Shields' result to several complex variables is due to Eustice [E], who in 1972 proved the same fact for holomorphic maps of the bidisk $\Delta^2 = \Delta \times \Delta \subset \mathbf{C}^2$. Shortly later, Suffridge [Su] in 1974 found a proof of the same result for the unit ball B^n of $\mathbf{C}^n$ (actually, Suffidge constructed a common fixed point for two commuting maps, but his proof can be adapted to a generic family).

Shields' proof of Theorem 0.1 relies on the main fact of iteration theory of holomorphic functions of Δ into itself, the Wolff-Denjoy theorem:

Theorem 0.2 (Wolff [W 1, 2, 3], Denjoy [D]). *Let $f : \Delta \to \Delta$ be a holomorphic function, $f \neq \mathrm{id}_\Delta$. Assume that f is not an automorphism of Δ with exactly one fixed point. Then the sequence $\{f^k\}$ of iterates of f converges, uniformly on compact sets, to a constant function $\tau \in \bar{\Delta}$.*

In their proofs, Eustice and Suffridge used some weak version of the iteration theory on the bidisk and the ball, respectively, together with Shields' theorem to build up an induction argument; in particular, in Suffridge's proof the fact that the fixed point set of a holomorphic map of B^n into itself is biholomorphic to a ball of smaller dimension plays a fundamental rôle. Furthermore, it should be remarked that the whole strength of the iteration theory in the ball (due to Hervé [H]) yields an easier proof of Suffridge's result.

Recently ([A 1, 2]), the iteration theory has been developed in strongly convex domains, and the aim of this paper is to use it to extend Shields' theorem to this situation. Actually, a common fixed point for two commuting maps was already constructed in [A 1]; here we shall prove the complete generalization of Theorem 0.1. As in Eustice's and Suffridge's proofs, we shall use in a fundamental way the structure of the fixed point set of a holomorphic map, as described by Vigué [Vi 1, 2]. Unfortunately, it is not clear whether the fixed point set of a holomorphic map of a convex domain into itself is biholomorphic to a convex domain of smaller dimension; we only know how to describe it by means of complex geodesics, a concept introduced by Vesentini [V 1, 2] and mainly studied by Lempert [L 1, 2] and Royden and Wong [RW]. So, the first section of this paper is devoted to a review of some facts about complex geodesics, probably already known but lacking in suitable bibliographical references. For the sake of simplicity, we shall prove several results for strongly convex domains with C^3 boundary, but they probably hold in domains with C^2 boundary too. The proof of the main theorem is the bulk of the second section of the paper, and it was already announced in [A 3].

We shall denote by $\mathrm{Hol}(M, N)$ the space of holomorphic maps from the complex manifold M into the complex manifold N, endowed with the compact-open topology; by $\|\cdot\|$ the usual euclidean norm on $\mathbf{C}^n$, and by ω the Poincaré distance on Δ. For every $z, w \in \mathbf{C}^n$ we set

$$\langle z, w\rangle = \sum_{j=1}^{n} z_j w_j;$$

in particular, $\langle z, \bar{w}\rangle$ is the usual hermitian product on $\mathbf{C}^n$. Finally, if $D \subset\subset \mathbf{C}^n$ has at least C^2 boundary, we shall denote by $\mathbf{n}_x$ the exterior unit normal to D at $x \in \partial D$.

I would like to thank László Lempert for an illuminating conversation regarding Proposition 1.8.

1. Complex Geodesics

Let $D \subset\subset \mathbf{C}^n$ be a bounded domain in $\mathbf{C}^n$. For every $z \in D$, we identify the tangent space $T_z D$ with $\mathbf{C}^n$, as usual. The *Kobayashi metric* $\kappa_D \colon TD \to \mathbf{R}^+$ is defined by

$$\forall z \in D \quad \forall v \in T_z D \quad \kappa_D(z; v) = \inf\{|\xi| \mid \exists \varphi \in \mathrm{Hol}(\Delta, D) \colon \varphi(0) = z,\ \varphi'(0) = v\}.$$

The *Kobayashi distance* $k_D \colon D \times D \to \mathbf{R}^+$ on D is the integrated form of κ_D:

$$\forall z_0, z_1 \in D \quad k_D(z_0, z_1)$$
$$= \inf\left\{ \int_0^1 \kappa_D(\gamma(t); \gamma'(t)) \,\middle|\, \gamma \text{ is a } C^1 \text{ curve in } D \text{ connecting } z_0 \text{ and } z_1 \right\}.$$

For general properties of the Kobayashi metric and distance consult [K 1, 2] and references therein.

Lempert [L 1] has shown that if D is convex then the Kobayashi distance is given by

$$\forall z_0, z_1 \in D \quad k_D(z_0, z_1) = \inf\{\omega(0, \zeta) \mid \exists \varphi \in \mathrm{Hol}(\Delta, D) : \varphi(0) = z_0,\ \varphi(\zeta) = z_1\}.$$

A *complex geodesic* φ is a holomorphic map from Δ into D which is an isometry for the Poincaré distance on Δ and the Kobayashi distance on D, i.e.,

$$\forall \zeta_1, \zeta_2 \in \Delta \quad k_D(\varphi(\zeta_1), \varphi(\zeta_2)) = \omega(\zeta_1, \zeta_2).$$

Remark that if a complex geodesic φ extends continuously to $\partial\Delta$, then $\varphi(\partial\Delta) \subset \partial D$.

A *geodesic disk* is the image of a complex geodesic. Vesentini [V 2] has shown that a complex geodesic is determined up to an automorphism of Δ by its image. We shall say that a complex geodesic φ is *passing through* two given points $z_0, z_1 \in D$ if z_0 and z_1 belong to the image of φ; it is clear that, up to an automorphism of Δ, we can assume $\varphi(0) = z_0$ and $\varphi(\zeta_0) = z_1$ for some $\zeta_0 \in (0, 1)$. Analogously, we shall say that a complex geodesic φ is *tangent* to a direction $v \in \mathbf{C}^n$ at the point $z_0 \in D$ if $\varphi(0) = z_0$ and $\varphi'(0) = \lambda v$ for some $\lambda > 0$.

The main facts about complex geodesics in convex domains are summarized in:

Theorem 1.1. *Let $D \subset\subset \mathbf{C}^n$ be a bounded convex domain. Then:*

(i) (Lempert [L 1], Royden and Wong [RW]) *For any two points $z_0, z_1 \in D$ there exists a complex geodesic passing through z_0 and z_1;*

(ii) (Lempert [L 1], Royden and Wong [RW]) *For any point $z_0 \in D$ and any direction $v \in \mathbf{C}^n$ there exists a complex geodesic tangent to v at z_0;*

(iii) (Lempert [L 1]) *If D is strongly convex with C^2 boundary, every complex geodesic extends to a $C^{0,1/2}$ map of $\bar{\Delta}$ into $\bar{D}$;*

(iv) (Lempert [L 1]) *More generally, if D is strongly convex with C^r boundary ($r = 3, \ldots, \infty$), every complex geodesic extends to a C^{r-2} map of $\bar{\Delta}$ into $\bar{D}$;*

(v) (Vesentini [V 2], Royden and Wong [RW]) *A holomorphic map $\varphi : \Delta \to D$ is a complex geodesic iff there are $\zeta_0, \zeta_1 \in \Delta$ such that $k_D(\varphi(\zeta_0), \varphi(\zeta_1)) = \omega(\zeta_0, \zeta_1)$;*

(vi) (Vesentini [V 2], Royden and Wong [RW]) *A holomorphic map $\varphi : \Delta \to D$ is a complex geodesic iff $\kappa_D(\varphi(0); \varphi'(0)) = 1$.*

Lempert gave the following characterization of complex geodesics in convex domains with C^2 boundary:

Theorem 1.2 (Lempert [L 1]). *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^2 domain. Then a holomorphic map $\varphi : \Delta \to D$ is a complex geodesic iff it extends to a $C^{0,1/2}$ map of $\bar{\Delta}$ into $\bar{D}$ and there exists a continuous function $p : \partial\Delta \to \mathbf{R}^+$ such that the map $\zeta \mapsto \zeta p(\zeta)\overline{\mathbf{n}_{\varphi(\zeta)}}$ defined on $\partial\Delta$ extends continuously to a holomorphic map $\tilde{\varphi} : \Delta \to \mathbf{C}^n$. Furthermore, p can be chosen so that on Δ we have*

$$\langle \varphi', \tilde{\varphi} \rangle \equiv 1. \tag{1.1}$$

A holomorphic map $\varrho : D \to D$ is a *holomorphic retraction* if $\varrho^2 = \varrho$. In particular, $\varrho(D)$ is a complex submanifold of D (see Rossi [R] or Cartan [C]), and coincides with the set of fixed points of ϱ. There is a strong connection between complex geodesics and holomorphic retractions:

Theorem 1.3 (Lempert [L2]). *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^r domain, ($r=2,3,\ldots,\infty$), and $\varphi \in \mathrm{Hol}(\Delta, D)$ a complex geodesic. Then there exists a holomorphic function $F: D \to \Delta$ such that*

$$\forall z \in D \quad \langle z-\varphi(F(z)), \tilde{\varphi}(F(z))\rangle = 0. \tag{1.2}$$

In particular, $F \circ \varphi = \mathrm{id}_\Delta$, and $\varphi \circ F$ is a holomorphic retraction of D onto $\varphi(\Delta)$. Furthermore, F extends C^{r-2} to ∂D.

The function F is called the *left inverse* of the complex geodesic φ.

The rest of this section is devoted to the study of the uniqueness of complex geodesics. First of all, in strongly convex domains there exists a unique geodesic disk passing through two given points:

Proposition 1.4 (Lempert [L1], Royden and Wong [RW]). *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^2 domain. Then*

(i) *For any two distinct points z_0, $z_1 \in D$ there exists a unique geodesic disk passing through z_0 and z_1;*

(ii) *For any point $z_0 \in D$ and direction $v \in \mathbf{C}^n$, $v \neq 0$, there exists a unique geodesic disk tangent to v at z_0.*

Proof. Since the proof of (ii) is quite similar to the proof of (i), we shall describe in detail the latter only. The existence is Theorem 1.1 (i). Assume now that $\varphi_0, \varphi_1: \Delta \to D$ are two complex geodesics passing through z_0 and z_1; up to automorphisms of Δ we can assume that $\varphi_0(0)=\varphi_1(0)=z_0$ and $\varphi_0(\zeta_0)=\varphi_1(\zeta_0)=z_1$ for some $\zeta_0 \in \Delta$. For $\lambda \in [0,1]$ set $\varphi_\lambda=(1-\lambda)\varphi_0+\lambda\varphi_1$. Clearly, every φ_λ is a holomorphic map of Δ into D; moreover, $\varphi_\lambda(0)=z_0$ and $\varphi_\lambda(\zeta_0)=z_1$ for every $\lambda \in [0,1]$. Then, by Theorem 1.1 (v), every φ_λ is a complex geodesic; in particular, $\varphi_\lambda(\partial\Delta) \subset \partial D$ for every $\lambda \in [0,1]$. But D is strongly convex; hence $\varphi_\lambda|_{\partial\Delta}$ does not depend on λ. In particular, $\varphi_0|_{\partial\Delta}=\varphi_1|_{\partial\Delta}$ and, since bounded holomorphic maps are completely determined by their boundary values, $\varphi_0 \equiv \varphi_1$. q.e.d.

Given $z_0, z_1 \in D$, the unique complex geodesic φ such that $\varphi(0)=z_0$ and $\varphi(\zeta_0)=z_1$ for some $\zeta_0 \in (0,1)$ will be called the *normalized complex geodesic* passing through z_0 and z_1.

Now we want to discuss the uniqueness of complex geodesics passing through given points in the boundary of D. To do so, we shall need the concept of angular derivative. Let f be a holomorphic map of Δ into itself, with radial limit $\tau \in \bar{\Delta}$ at $\sigma \in \partial\Delta$; then the *angular derivative* of f at σ is defined by

$$f'(\sigma)=\lim_{t\to 1}\frac{\tau-f(t\sigma)}{(1-t)\sigma}. \tag{1.3}$$

It turns out (see for instance Burckel [Bu]) that the limit (1.3) always exists (and it can be $+\infty$); moreover, if $\tau=\sigma$ then $f'(\sigma) \in (0,+\infty]$ and, if it is finite, it coincides with the non-tangential limit of f' at σ. Conversely, if f extends C^1 to $\bar{\Delta}$, then the angular derivative is always finite and it is clearly given by the value of f' at σ.

The facts we need about the angular derivative are summarized in:

Theorem 1.5. *Let $f: \Delta \to \Delta$ be holomorphic. Then*

(i) (Herzich [He]) *If $f(0)=0$ and the radial limit of f at 1 is 1, then $f'(1) \geqq 1$. Moreover, $f'(1)=1$ iff $f=\mathrm{id}_\Delta$.*

(ii) (Behan [B]) *If the radial limit of f at 1 is 1, and at -1 is -1, then* $f'(1)f'(-1)\geqq 1$. *Moreover,* $f'(1)f'(-1)=1$ *iff f is an automorphism of Δ fixing* 1 *and* -1.

To compute angular derivatives, we shall use the

Lemma 1.6. *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^3 domain, $x \in \partial D$ and $\varphi: \Delta \to D$ a complex geodesic such that $\varphi(1)=x$. Denote by $F: D \to \Delta$ the left inverse of φ. Then*

$$\forall v \in \mathbf{C}^n \quad dF_x(v)=\frac{\langle v, \overline{\mathbf{n}_x}\rangle}{\langle \varphi'(1), \overline{\mathbf{n}_x}\rangle}=\langle v, \tilde{\varphi}(1)\rangle . \tag{1.4}$$

Proof. Since for any $\zeta \in \Delta$ we have $dF_{\varphi(z)}(\varphi'(\zeta))=1$ and

$$\ker dF_{\varphi(\zeta)}=\{v \in \mathbf{C}^n | \langle v, \tilde{\varphi}(\zeta)\rangle=0\},$$

it is clear that [by (1.1)]

$$dF_{\varphi(\zeta)}(v)=\langle v, \tilde{\varphi}(\zeta)\rangle .$$

Then letting $\zeta \to 1$ and recalling Theorem 1.2, we get (1.4). q.e.d.

Now we can prove

Proposition 1.7. *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^3 domain. Then for any $z_0 \in D$ and $x \in \partial D$ there exists a unique complex geodesic $\varphi: \Delta \to D$ such that $\varphi(0)=z_0$ and $\varphi(1)=x$.*

Proof. First of all the existence. Let $\{z_k\} \subset D$ be a sequence converging to x; denote by φ_k the normalized complex geodesic passing through z_0 and z_k. Since D is taut, up to a subsequence we can assume that $\{\varphi_k\}$ converges to a holomorphic map $\varphi: \Delta \to D$. Clearly $\varphi(0)=z_0$; moreover for all $\zeta \in \Delta$

$$k_D(z_0, \varphi(\zeta))=\lim_{k \to \infty} k_D(z_0, \varphi_k(\zeta))=\omega(0, \zeta),$$

and φ is a complex geodesic. Then it extends $C^{0,1/2}$ to ∂D, and clearly $\varphi(1)=x$.

Assume now that ψ is another complex geodesic with $\psi(0)=z_0$ and $\psi(1)=x$. Denote by F (respectively G) the left inverse of φ (respectively ψ). We claim that

$$F \circ \psi=\mathrm{id}_\Delta . \tag{1.5}$$

In fact, by Lemma 1.6

$$(F \circ \psi)'(1)=dF_x(\psi'(1))=\frac{\langle \psi'(1), \overline{\mathbf{n}_x}\rangle}{\langle \varphi'(1), \overline{\mathbf{n}_x}\rangle}.$$

Analogously,

$$(G \circ \varphi)'(1)=\frac{\langle \varphi'(1), \overline{\mathbf{n}_x}\rangle}{\langle \psi'(1), \overline{\mathbf{n}_x}\rangle}=\frac{1}{(F \circ \psi)'(1)}.$$

Then, by Theorem 1.5 (i), $(F \circ \psi)'(1)=(G \circ \varphi)'(1)=1$ and therefore (1.5) is proven.

Now, (1.5) means that $\langle \psi-\varphi, \tilde{\varphi}\rangle \equiv 0$ on $\bar{\Delta}$. In particular,

$$\langle \psi(\sigma)-\varphi(\sigma), \overline{\mathbf{n}_{\varphi(\sigma)}}\rangle=0$$

for every $\sigma \in \partial\Delta$. Since D is strongly convex, this implies that $\psi \equiv \varphi$ on $\partial\Delta$, and hence everywhere. q.e.d.

Using a similar argument, we can prove the uniqueness of the geodesic disk passing through two given boundary points:

Proposition 1.8. *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^3 domain. Then for any pair of distinct points $x_1, x_2 \in \partial D$ there exists a unique (up to automorphisms of Δ) complex geodesic φ such that $\varphi(1) = x_1$ and $\varphi(-1) = x_2$.*

Proof. We begin with the existence. Let $\{z_k\} \subset D$ be a sequence converging to x_2, and denote by φ_k a complex geodesic such that $\varphi_k(1) = x_1$, $z_k \in \varphi_k((-1, 1))$ and

$$\|\varphi_k(0) - x_1\| < \frac{\|x_2 - x_1\|}{2}. \tag{1.6}$$

Since D is bounded, up to a subsequence we can assume that $\{\varphi_k\}$ converges to a holomorphic map $\varphi : \Delta \to \mathbf{C}^n$. Since D is strongly convex, either $\varphi(\Delta) \subset D$ or φ is a constant contained in ∂D. The last possibility cannot occur [by (1.6)]; so $\varphi \in \mathrm{Hol}(\Delta, D)$, and it is clear that φ is as desired.

Assume now that ψ is another complex geodesic with $\psi(1) = x_1$ and $\psi(-1) = x_2$, and denote again by F (respectively G) the left inverse of φ (respectively ψ). We claim that this time

$$F \circ \psi \in \mathrm{Aut}(\Delta). \tag{1.7}$$

Indeed, by Lemma 1.6

$$(F \circ \psi)'(1) \cdot (F \circ \psi)'(-1) = \frac{\langle \psi'(1), \overline{\mathbf{n}_{x_1}} \rangle}{\langle \varphi'(1), \overline{\mathbf{n}_{x_1}} \rangle} \cdot \frac{\langle \psi'(-1), \overline{\mathbf{n}_{x_2}} \rangle}{\langle \varphi'(-1), \overline{\mathbf{n}_{x_2}} \rangle}.$$

For the same reason,

$$(G \circ \varphi)'(1) \cdot (G \circ \varphi)'(-1) = \frac{1}{(F \circ \psi)'(1) \cdot (F \circ \psi)'(-1)}.$$

Then Theorem 1.5 (ii) yields

$$(F \circ \psi)'(1) \cdot (F \circ \psi)'(-1) = 1,$$

and then (1.7). Hence, up to compose ψ with an automorphism of Δ, we can assume that $F \circ \psi = \mathrm{id}_\Delta$, and the assertion follows as in the proof of Proposition 1.7. q.e.d.

Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^3 domain: for every $z_0 \in D$ and $w_0 \in \bar{D}$, with $z_0 \neq w_0$, let φ_{z_0, w_0} denote the unique complex geodesic such that

$$\varphi_{z_0, w_0}(0) = z_0 \quad \text{and} \quad \varphi_{z_0, w_0}(\tanh[k_D(z_0, w_0)]) = w_0,$$

where $\tanh[k_D(z_0, w_0)] = 1$ if $w_0 \in \partial D$.

Lemma 1.9. *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^3 domain. Then the map $(z_0, w_0) \mapsto \varphi_{z_0, w_0}$ from $D \times \bar{D}$ minus the diagonal to $\mathrm{Hol}(\Delta, D) \cap C^0(\bar{\Delta}, \bar{D})$ is continuous.*

Proof. Let $\{(z_k, w_k)\}$ be a sequence in $D \times \bar{D}$ (with $z_k \neq w_k$ for all $k \in \mathbf{N}$) converging to $(z_0, w_0) \in D \times \bar{D}$ with $z_0 \neq w_0$; we have to show that $\varphi_k = \varphi_{z_k, w_k}$ tends to φ_{z_0, w_0}. Since

D is taut and $\varphi_k(0)=z_k\to z_0\in D$, it suffices to show that the only limit point of $\{\varphi_k\}$ is φ_{z_0,w_0}.

Let $\{\varphi_{k_\nu}\}$ be a subsequence of $\{\varphi_k\}$ converging to a map φ. Since D is taut, $\varphi\in\mathrm{Hol}(\Delta,D)$ is a complex geodesic with $\varphi(0)=z_0$. Put

$$s_k=\tanh(k_D(z_k,w_k));$$

obviously, $s_k\to s_0=\tanh(k_D(z_0,w_0))$. Therefore

$$\varphi(s_0)=\lim_{\nu\to\infty}\varphi_{k_\nu}(s_{k_\nu})=\lim_{\nu\to\infty}w_{k_\nu}=w_0,$$

and φ must be φ_{z_0,w_0}. q.e.d.

Now we can define the representation of D onto the unit ball B^n of $\mathbf{C}^n$ introduced by Lempert [L1]. Fix $z_0\in D$, and define $\Phi_{z_0}:\bar{D}\to\overline{B^n}$ by setting $\Phi_{z_0}(z_0)=0$ and

$$\Phi_{z_0}(z)=\tanh(k_D(z_0,z))\frac{\varphi'_{z_0,z}(0)}{\|\varphi'_{z_0,z}(0)\|}. \tag{1.8}$$

In other words, $\Phi_{z_0}(z)$ is the point w of B^n such that $w/\|w\|$ is parallel to $\varphi'_{z_0,z}$ and $k_{B^n}(0,w)=k_D(z_0,z)$. Then

Proposition 1.10. *Let $D\subset\subset\mathbf{C}^n$ be a strongly convex C^3 domain, and $z_0\in D$. Then the map $\Phi_{z_0}:\bar{D}\to\overline{B^n}$ defined in* (1.8) *is a homeomorphism of $\bar{D}$ with $\overline{B^n}$ such that*

(i) *$\Phi_{z_0}(z_0)=0$ and $\Phi_{z_0}(\partial D)=(\partial B^n)$;*

(ii) *for any $z\in D$ we have $k_D(z_0,z)=\omega(0,\|\Phi_{z_0}(z)\|)$;*

(iii) *for any $z\in\bar{D}\setminus\{z_0\}$ we have*

$$\varphi_{z_0,z}(\zeta)=\Phi_{z_0}^{-1}\left(\zeta\frac{\Phi_{z_0}(z)}{\|\Phi_{z_0}(z)\|}\right).$$

Furthermore, Φ_{z_0} depends continuously on z_0.

Proof. Since every $\varphi_{z_0,z}$ is holomorphic, by Lemma 1.9 $\varphi'_{z_0,z}(0)$ depends continuously on z_0 and z, and so Φ_{z_0} is continuous and depends continuously on z_0. Since $\bar{D}$ is compact, it suffices to show that Φ_{z_0} is bijective, and to verify (iii).

Φ_{z_0} is injective: assume $\Phi_{z_0}(z_1)=\Phi_{z_0}(z_2)$. Then $k_D(z_0,z_1)=k_D(z_0,z_2)$ and

$$\frac{\varphi'_{z_0,z_1}(0)}{\|\varphi'_{z_0,z_1}(0)\|}=\frac{\varphi'_{z_0,z_2}(0)}{\|\varphi'_{z_0,z_2}(0)\|}.$$

Now, φ_{z_0,z_1} and φ_{z_0,z_2} are complex geodesics; hence, by Theorem 1.1 (vi),

$$\kappa_D(z_0;\varphi'_{z_0,z_1}(0))=1=\kappa_D(z_0;\varphi'_{z_0,z_2}(0)),$$

and so $\|\varphi'_{z_0,z_1}(0)\|=\|\varphi'_{z_0,z_2}(0)\|$. Thus $\varphi'_{z_0,z_1}(0)=\varphi'_{z_0,z_2}(0)$; hence, by Proposition 1.4 (ii), $\varphi_{z_0,z_1}=\varphi_{z_0,z_2}$ and then $z_1=z_2$.

Φ_{z_0} is surjective: let $w=rx\in\overline{B^n}$, where $r\in(0,1]$ and $x\in\partial B^n$. Choose a complex geodesic $\varphi\in\mathrm{Hol}(\Delta,D)$ such that $\varphi(0)=z_0$ and $\varphi'(0)=\lambda x$ for some $\lambda>0$. Clearly, $\varphi=\varphi_{z_0,z}$ for some $z\in\bar{D}$; we can also assume $\tanh(k_D(z_0,z))=r$. Then

$$\Phi_{z_0}(z)=\tanh(k_D(z_0,z))\cdot\frac{\lambda x}{\lambda}=w.$$

Finally, we are left to show that

$$\Phi_{z_0}(\varphi_{z_0,z}(\zeta))=\zeta\frac{\Phi_{z_0}(z)}{\|\Phi_{z_0}(z)\|}$$

for all $\zeta\in\Delta^*$. Fix $\zeta_0\in\Delta^*$; by definition,

$$\tanh[k_D(z_0,\varphi_{z_0,z}(\zeta_0))]=|\zeta_0|$$

and

$$\frac{\Phi_{z_0}(z)}{\|\Phi_{z_0}(z)\|}=\frac{\varphi'_{z_0,z}(0)}{\|\varphi'_{z_0,z}(0)\|}.$$

Now let $\tau=\zeta_0/|\zeta_0|\in\partial\Delta$, and define $\psi\in\mathrm{Hol}(\Delta,D)$ by $\psi(\zeta)=\varphi_{z_0,z}(\tau\zeta)$. ψ is a complex geodesic such that $\psi(0)=z_0$ and $\psi(|\zeta_0|)=\varphi_{z_0,z}(\zeta_0)$; therefore

$$\Phi_{z_0}(\varphi_{z_0,z}(\zeta_0))=|\zeta_0|\frac{\psi'(0)}{\|\psi'(0)\|}=\zeta_0\frac{\varphi'_{z_0,z}(0)}{\|\varphi'_{z_0,z}(0)\|},$$

and we are done. q.e.d.

It should be remarked that Lempert [L 1], using different methods, proved that if D has C^r boundary ($r=4,\dots,\infty$) then Φ_{z_0} is a C^{r-3}-diffeomorphism of $\bar{D}\backslash\{z_0\}$ onto $\bar{B}^n\backslash\{0\}$, depending C^{r-3} on z_0.

2. Common Fixed Points

We can now move toward the main theorem of this paper; we are left to recall few facts about iteration theory and the structure of fixed point sets. If f is a holomorphic map of a domain D into itself, we shall denote by $\mathrm{Fix}(f)$ the set of fixed points of f in D. Remark that, by definition, $\mathrm{Fix}(f)$ is always contained in D, even if f is continuous up to the boundary of D and has fixed points there. Moreover, the set of fixed points of f in $\bar{D}$ is in general strictly greater of the closure in $\bar{D}$ of $\mathrm{Fix}(f)$ [consider for instance the map $f\colon B^2\to B^2$ given by $f(z,w)=(z^3,w)$].

The structure of $\mathrm{Fix}(f)$ in convex domains is quite well understood:

Theorem 2.1 (Vigué [Vi 1, 2]). *Let $D\subset\subset\mathbf{C}^n$ be a bounded convex domain, and $f\in\mathrm{Hol}(D,D)$. Then:*

(i) $\mathrm{Fix}(f)$ *is a (possibly empty) closed connected submanifold of D;*

(ii) *for any pair of distinct points $z_1,z_2\in\mathrm{Fix}(f)$ there exists a geodesic disk passing through z_1 and z_2 contained in* $\mathrm{Fix}(f)$.

It should be remarked that if z_1 and z_2 are two distinct points in the topological closure $\overline{\mathrm{Fix}(f)}\subset\bar{D}$ of $\mathrm{Fix}(f)$, then again we can find a geodesic disk passing through z_1 and z_2; it suffices to use the construction of geodesic disks passing through given boundary points described in the proofs of Propositions 1.7 and 1.8. Combining this with the results of the previous section we get:

Corollary 2.2. *Let $D\subset\subset\mathbf{C}^n$ be a strongly convex C^2 domain, and take $f_1,\dots,f_\mu\in\mathrm{Hol}(D,D)$ such that $F=\mathrm{Fix}(f_1)\cap\dots\cap\mathrm{Fix}(f_\mu)\neq\emptyset$. Then $\bar{F}=\overline{\mathrm{Fix}(f_1)\cap\dots\cap\mathrm{Fix}(f_\mu)}\subset\bar{D}$ is homeomorphic to a compact convex subset of $\mathbf{C}^n$.*

Proof. Take $z_0 \in F$. If $F=\{z_0\}$, the assertion is obvious. Otherwise, Proposition 1.10 and Theorem 2.1 show that

$$\bar{F}=\Phi_{z_0}^{-1}(\overline{B^r}\cap V),$$

for a suitable linear subspace V of $\mathbf{C}^n$, and we are done. q.e.d.

The main facts of iteration theory we need are summarized in:

Theorem 2.3 (Abate [A 1, 2]). *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^2 domain, and $f \in \mathrm{Hol}(D,D)$. Then:*

(i) *If $\mathrm{Fix}(f)=\emptyset$, then the sequence of iterates $\{f^k\}$ of f converges, uniformly on compact sets, to a constant map $x \in \partial D$.*

(ii) *If $\mathrm{Fix}(f)\neq\emptyset$, then there is a subsequence $\{f^{k_\nu}\}$ of iterates of f converging, uniformly on compact sets, to a holomorphic retraction ϱ_f of D onto a submanifold M_f of D. Moreover, ϱ_f does not depend on the particular subsequence, but only on f, and $f|_{M_f}$ is an automorphism of M_f.*

If $\mathrm{Fix}(f)\neq\emptyset$, the holomorphic retraction ϱ_f is called the *limit retraction* of f.

Finally we can prove our main theorem:

Theorem 2.4. *Let $D \subset\subset \mathbf{C}^n$ be a strongly convex C^3 domain, and $\mathscr{F} \subset \mathrm{Hol}(D,D) \cap C^0(\bar{D})$ a family of commuting holomorphic maps. Then there exists a point $x \in \bar{D}$ such that $f(x)=x$ for all $f \in \mathscr{F}$.*

Proof. Assume first that there is $f \in \mathscr{F}$ without fixed points in D. Then, by Theorem 2.3 (i), there exists a point $x \in \partial D$ such that the sequence of iterates of f converges to x. Hence for any $g \in \mathscr{F}$ we have

$$g(x)=\lim_{k\to\infty} g(f^k(z))=\lim_{k\to\infty} f^k(g(z))=x,$$

where z is any point of D, and we are done.

So assume that $\mathrm{Fix}(f)\neq\emptyset$ for every $f \in \mathscr{F}$. The first key observation now is that if $f, g \in \mathrm{Hol}(D,D)$ commute, then g sends $\mathrm{Fix}(f)$ into itself. In fact, if $z \in \mathrm{Fix}(f)$ then

$$f(g(z))=g(f(z))=g(z).$$

The second key observation is that if $f, g \in \mathscr{F}$ then ϱ_f and ϱ_g commute. In particular, $\varrho_f \circ \varrho_g$ is a holomorphic retraction, and moreover

$$\varrho_f \circ \varrho_g(D)=\varrho_f(D)\cap\varrho_g(D).$$

Indeed, $\varrho_f \circ \varrho_g(D)$ is contained in $\varrho_f(D)\cap\varrho_g(D)$ for ϱ_f and ϱ_g commute; on the other hand, $\varrho_f(D)\cap\varrho_g(D)$ is clearly contained in $\mathrm{Fix}(\varrho_f\circ\varrho_g)=\varrho_f\circ\varrho_g(D)$. An induction argument then shows that for any $f_1,\dots,f_\mu \in \mathscr{F}$ the map $\varrho_{f_1}\circ\dots\circ\varrho_{f_\mu}$ is a holomorphic retraction of D onto

$$\varrho_{f_1}\circ\dots\circ\varrho_{f_\mu}(D)=\varrho_{f_1}(D)\cap\dots\cap\varrho_{f_\mu}(D).$$

In particular, the intersection on the right-hand side is always a non-empty closed connected submanifold of D. Choose $f_1,\dots,f_\mu \in \mathscr{F}$ so that the dimension of $\varrho_{f_1}(D)\cap\dots\cap\varrho_{f_\mu}(D)$ is minimal; then for any $f \in \mathscr{F}$ we should have

$$\varrho_{f_1}(D)\cap\dots\cap\varrho_{f_\mu}(D)\subset\varrho_f(D). \tag{2.1}$$

Set $\varrho=\varrho_{f_1}\circ\ldots\circ\varrho_{f_\mu}$ and $M=\varrho(D)$. Then ϱ is a holomorphic retraction of D onto M commuting with every $f\in\mathscr{F}$. Moreover, for every $f\in\mathscr{F}$ we have $M\subset\varrho_f(D)$, by (2.1), and $f(M)\subset M$, for $M=\mathrm{Fix}(\varrho)$. Now, $f|_{\varrho_f(D)}$ is an automorphism of $\varrho_f(D)$; hence $f(M)=M$ and $f|_M\in\mathrm{Aut}(M)$. Finally, for any $f\in\mathscr{F}$ we have $\mathrm{Fix}(f)\cap M\neq\emptyset$; indeed, since ϱ commutes with f, $\varrho(\mathrm{Fix}(f))\subset\mathrm{Fix}(f)$ and so

$$\mathrm{Fix}(f)\cap M=\varrho(\mathrm{Fix}(f))\neq\emptyset\,.$$

We shall denote $\mathrm{Fix}(f)\cap M$ by F_f; since it is easy to check that $F_f=\mathrm{Fix}(\varrho\circ f)$, every F_f is again a closed connected submanifold of D, invariant under the action of any $g\in\mathscr{F}$.

Now, $k_M=k_D|_{M\times M}$; in fact

$$\forall z_1,z_2\in M\quad k_D(z_1,z_2)\leqq k_M(z_1,z_2)=k_M(\varrho(z_1),\varrho(z_2))\leqq k_D(z_1,z_2)\,.$$

In particular, every complex geodesic in M is a complex geodesic in D; moreover, since $M=\mathrm{Fix}(\varrho)$, for every two distinct points in $\bar{M}$ passes a unique complex geodesic.

If $\mathscr{F}$ has a common fixed point in ∂D, our work is clearly finished. So assume that there are no common fixed points of $\mathscr{F}$ in the boundary of D; we want to construct a common fixed point in the interior of D. First of all we claim that

$$\forall f_1,\ldots,f_\mu\in\mathscr{F}\quad F_{f_1}\cap\ldots\cap F_{f_\mu}\neq\emptyset\,. \tag{2.2}$$

We argue by induction on μ. For $\mu=1$ is clear. Assume $F_{f_1}\cap\ldots\cap F_{f_{\mu-1}}\neq\emptyset$, and take $f_\mu\in\mathscr{F}$. By Corollary 2.2, $\overline{F_{f_1}\cap\ldots\cap F_{f_{\mu-1}}}$ is homeomorphic to a compact convex subset of $\mathbf{C}^n$. Since it is also invariant under f_μ, by Brouwer's theorem f_μ has a fixed point in $\overline{F_{f_1}\cap\ldots\cap F_{f_{\mu-1}}}$. We have to show that f_μ actually has a fixed point in $F_{f_1}\cap\ldots\cap F_{f_{\mu-1}}$. Assume it does not. If f_μ had a unique fixed point $x\in\overline{F_{f_1}\cap\ldots\cap F_{f_{\mu-1}}}\cap\partial D$, then $\{x\}$ should be invariant under any $f\in\mathscr{F}$, that is x should be a common fixed point of $\mathscr{F}$ contained in ∂D, against our assumption. So f_μ should have at least two distinct fixed points $x,y\in\overline{F_{f_1}\cap\ldots\cap F_{f_{\mu-1}}}\cap\partial D$. Let $\varphi:\bar{\Delta}\to\bar{D}$ be the unique complex geodesic passing through x and y; clearly, $\varphi(\bar{\Delta})\subset\bar{M}$. But $f_\mu|_M$ is an automorphism of M; hence $f_\mu\circ\varphi$ is again a complex geodesic passing through x and y. By Proposition 1.8, $f_\mu\circ\varphi(\Delta)=\varphi(\Delta)$, that is f_μ sends $\varphi(\Delta)$ into itself, and without fixed points, for $\varphi(\Delta)\subset F_{f_2}\cap\ldots\cap F_{f_{\mu-1}}$. But then the sequence of iterates of $f_\mu|_{\varphi(\Delta)}$ should converge to a point of the boundary, by Theorem 0.2, and this is impossible, for f_μ has fixed points in D, by assumption. The contradiction arises from assuming $F_{f_1}\cap\ldots\cap F_{f_\mu}=\emptyset$; hence $F_{f_1}\cap\ldots\cap F_{f_\mu}\neq\emptyset$, as claimed.

So we have proven that, if $\mathscr{F}$ has no common fixed points in ∂D, then (2.2) holds. In particular, the intersection of every finite subset of the family $\{\overline{F_f}|f\in\mathscr{F}\}$ is not empty; since $\bar{D}$ is compact, this implies that the intersection of the whole family is not empty, and every element in this intersection is a common fixed point of $\mathscr{F}$. q.e.d.

References

[A 1] Abate, M.: Horospheres and iterates of holomorphic maps. Math. Z. **198**, 225–238 (1988)
[A 2] Abate, M.: Iterates and semigroups on taut manifolds. To appear in Atti delle Giornate di Geometria Analitica a Analisi Complessa, Rocca di Papa, 1988. Cosenza: Mediterranean Press 1989

[A 3] Abate, M.: Converging semigroups of holomorphic maps. To appear in Rend. Accad. Naz. Lincei **82** (1988)
[B] Behan, D.F.: Commuting analytic functions without fixed points. Proc. Am. Math. Soc. **37**, 114–120 (1973)
[Bo] Boyce, W.M.: Commuting functions with no common fixed points. Trans. Am. Math. Soc. **137**, 77–92 (1969)
[Bu] Burckel, R.B.: An introduction to classical complex analysis. New York: Academic Press 1979
[C] Cartan, H.: Sur les rétractions d'une variété. C.R. Acad. Sci. Paris **303**, 715–716 (1986)
[D] Denjoy, A.: Sur l'itération des fonctions analytiques. C.R. Acad. Sci. Paris **182**, 255–257 (1926)
[E] Eustice, D.J.: Holomorphic idempotents and common fixed points on the 2-disk. Mich. Math. J. **19**, 347–352 (1972)
[H] Hervé, M.: Quelques propriétés des applications analytiques d'une boule à m dimensions dans elle-même. J. Math. Pures Appl. **42**, 117–147 (1963)
[He] Herzig, A.: Die Winkelderivierte und das Poisson-Stieltjes Integral. Math. Z. **46**, 129–156 (1940)
[Hu] Huneke, J.P.: On common fixed points of commuting continuous functions on an interval. Trans. Am. Math. Soc. **139**, 371–381 (1969)
[K 1] Kobayashi, S.: Hyperbolic manifolds and holomorphic mappings. New York: Dekker 1970
[K 2] Kobayashi, S.: Intrinsic distances, measures, and geometric function theory. Bull. Am. Math. Soc. **82**, 357–416 (1976)
[L 1] Lempert, L.: La métrique de Kobayashi et la réprésentation des domaines sur la boule. Bull. Soc. Math. France **109**, 427–474 (1981)
[L 2] Lempert, L.: Holomorphic retracts and intrinsic metrics in convex domains. Anal. Math. **8**, 257–261 (1982)
[R] Rossi, H.: Vector fields on analytic spaces. Ann. Math. **78**, 455–467 (1963)
[RW] Royden, H.L., Wong, P.M.: Carathéodory and Kobayashi metrics on convex domains. Preprint (1983)
[S] Shields, A.L.: On fixed points of commuting analytic functions. Proc. Am. Math. Soc. **15**, 703–706 (1964)
[Su] Suffridge, T.J.: Common fixed points of commuting holomorphic maps of the hyperball. Mich. Math. J. **21**, 309–314 (1974)
[V 1] Vesentini, E.: Variations on a theme of Carathéodory. Ann. Scuola Norm. Sup. Pisa **6**, 39–68 (1979)
[V 2] Vesentini, E.: Complex geodesics. Compos. Math. **44**, 375–394 (1981)
[Vi 1] Vigué, J.-P.: Géodésiques complexes et points fixes d'applications holomorphes. Adv. Math. **52**, 241–247 (1984)
[Vi 2] Vigué, J.-P.: Points fixes d'applications holomorphes dans un domaine borné convexe de $\mathbb{C}^n$. Trans. Am. Math. Soc. **289**, 345–353 (1985)
[W 1] Wolff, J.: Sur l'itération des fonctions holomorphes dans une région, et dont les valeurs appartiennent à cette région. C.R. Acad. Sci. Paris **182**, 42–43 (1926)
[W 2] Wolff, J.: Sur l'itération des fonctions bornées. C.R. Acad. Sci. Paris **182**, 200–201 (1926)
[W 3] Wolff, J.: Sur une généralisation d'un théorème de Schwarz. C.R. Acad. Sci. Paris **182**, 918–920 (1926)

Received July 25, 1988

Math. Ann. 283, 657–678 (1989)

Mathematische Annalen

Laws of Large Numbers of Hypergroups on $\mathbb{R}_+$

Hansmartin Zeuner*

Mathematisches Institut der Universität Tübingen, Auf der Morgenstelle 10, D-7400 Tübingen, Federal Republic of Germany

1. Introduction

Being aware of the difficulties in formulating and proving a law of large numbers for random walks on an arbitrary locally compact group, it seems hopeless to try the same attempt on the even more general structure of a hypergroup. However, since these difficulties arise in part from the complicated geometric structure of many of the groups considered, one might expect that it is possible to obtain theorems on hypergroups which are of particularly simple geometry. This has successfully been done by Eymard, Roynette, Gallardo, and Ries (see [8, 15, 14]) in the case of the hypergroups on $\mathbb{N}$ related to the Gegenbauer polynomials. Whereas on the real line (with the usual topology) there is exactly one structure as a topological group, there is an abundant collection of hypergroups on the half line $\mathbb{R}_+$ (see Chébli [5], Zeuner [28]). In the case of Chébli-Trimèche hypergroups enough analytical tools are developed to prove the law of large numbers and the central limit theorem. The latter will be studied in the forthcoming paper [29].

If $X_1, X_2, \ldots$ are i.i.d. random variables with values in a group, the corresponding random walk is the sequence $S_1, S_2, \ldots$ defined by $S_n := X_n X_{n-1} \ldots X_1$. Since the operation on a hypergroup is only defined in terms of the convolution of measures, the random walk $(S_n : n \geqq 1)$ can only be defined by its distribution and not as a function of $(X_n : n \geqq 1)$. The notion of *concretization* and a randomized multiplication is introduced in 3.3 in order to obtain an explicit construction of $(S_n : n \geqq 1)$ in terms of $(X_n : n \geqq 1)$ for every 2^{nd} countable locally compact hypergroup.

As in the classical case, the moments of a random variable are introduced, both to formulate the conditions under which a particular limit theorem holds, and to calculate the actual value of the limit. This has to be done by a modified definition to fit with the hypergroup operation. The first and second moments are in very close connection with the notion of the dispersion of a probability measure used in Faraut [9] and Trimèche [24]. As to be expected by the results of Guivarc'h [16],

* Die Arbeit wurde mit Unterstützung eines Stipendiums des Wissenschaftsausschusses der NATO durch den DAAD ermöglicht

two different situations occur depending on the parameter ϱ (see 2.2) which determines the growth of the hypergroup. If $\varrho > 0$ the hypergroup is of exponential growth and the expectation of every nonzero random variable is strictly positive. This result includes the symmetric spaces of rank one of non-compact type and corresponds to Guivarc'h [16], corollaire on page 77. If $\varrho = 0$ the expectation of every random variable is 0 and so is the limit of $\frac{1}{n} S_n$ in probability. This result should be compared with [16, théorème 3, p. 72].

In both cases a strong law of large numbers will be proved. Apart from the different and less general situation in this article the main difference with the results in [16] is the fact that for hypergroups on $\mathbb{R}_+$ the law of large numbers can be formulated without the use of a gauge function and the a.s. limit of $\frac{1}{n} S_n$ can be calculated explicitly.

2. Preliminaries

2.1. Let K be a *hypergroup* in the sense of Jewett [20]; this means that K is a locally compact space with an associative convolution $(x, y) \mapsto \varepsilon_x * \varepsilon_y \in \mathscr{M}^1(K)$ (the space of probability measures on K) such that there exist a neutral element $e \in K$ and an inversion $x \mapsto x^\vee$ satisfying certain conditions (see [19, 20, 23] for details). In the cases considered in this article (except in the third paragraph) K will be *Hermitian* (i.e. $x^\vee = x$ for all $x \in K$); in particular this implies the commutativity of K.

The *dual* $\hat{K}$ of the Hermitian hypergroup K is the space of all real valued multiplicative functions φ on K with $\varphi(e) = \|\varphi\|_\infty = 1$ [20, 6.3]. For every probability measure P on K the Fourier transform $\mathscr{F}P$ is the continuous real-valued function $\varphi \mapsto \mathscr{F}P(\varphi) := \int \varphi \, dP$ on $\hat{K}$. It is a well known fact that the uniqueness theorem and the continuity theorem for the Fourier transform are valid for many commutative hypergroups [3, 19, 20].

2.2. In the sequel we consider the class of *Chébli-Trimèche hypergroups* on $K := \mathbb{R}_+$: For every function A on $\mathbb{R}_+$ (which turns out to be the Lebesgue density of a Haar measure of K) satisfying $A(0) = 0$, A strictly increasing and unbounded, $\frac{A'}{A}$ decreasing on $\mathbb{R}_+^*$, and $\frac{A'(x)}{A(x)} = \frac{\alpha}{x} + B(x)$ on a neighbourhood of 0 (where $\alpha > 0$ and B is an odd $\mathscr{C}^\infty$-function on $\mathbb{R}$), there exists a unique hypergroup structure on $\mathbb{R}_+$ such that

$$\frac{\partial}{\partial x}\left(A(x)A(y) \frac{\partial}{\partial x} (\int f \, d\varepsilon_x * \varepsilon_y) \right) = \frac{\partial}{\partial y}\left(A(x)A(y) \frac{\partial}{\partial y} (\int f \, d\varepsilon_x * \varepsilon_y) \right)$$

for every even $\mathscr{C}^\infty$-function f on $\mathbb{R}$ and $x, y \in \mathbb{R}_+$. The neutral element of this hypergroup is 0 and the inversion is the identity mapping.

The growth of this hypergroup is determined by the number $\varrho := \frac{1}{2} \lim_{x \to \infty} \frac{A'(x)}{A(x)} \geqq 0$: If $\varrho > 0$ then we obtain $A(x) \geqq A(1) \cdot e^{2\varrho(x-1)}$ for $x \geqq 1$ and so the hypergroup is of exponential growth; if $\varrho = 0$ then $(\mathbb{R}_+, *)$ is exponentially bounded. The multiplicative functions are precisely the solutions φ_λ ($\lambda \in \mathbb{C}$) of the differential equation

$$\varphi''_\lambda + \frac{A'}{A}\varphi'_\lambda + (\varrho^2+\lambda^2)\varphi_\lambda = 0, \qquad \varphi_\lambda(0)=1,\ \varphi'_\lambda(0)=0$$

and the dual is $\hat{K} = \{\varphi_\lambda : \lambda \in \mathbb{R}_+ \cup i[0,\varrho]\}$. In the following we will identify $\hat{K}$ with the set of parameters $\mathbb{R}_+ \cup i[0,\varrho]$. The proof of the preceding results can be found in [5, 28].

2.3. The most important technical tool used in this article is the Laplace representation for the multiplicative functions φ_λ ($\lambda \in \mathbb{C}$) proved in [5, Proposition I–IV]: For every $x \in \mathbb{R}_+$ there exists a probability measure v_x on $[-x, x]$ such that

$$\varphi_\lambda(x) = \int e^{-t(\varrho+i\lambda)} v_x(dt) \quad \text{for} \quad x \in \mathbb{R}_+,\ \lambda \in \mathbb{C}.$$

Furthermore the measure τ_x with the density $t \mapsto e^{-\varrho t}$ with respect to v_x is a symmetric subprobability measure on $\mathbb{R}$ which depends continuously on x [in the weak topology on $\mathscr{M}^b(\mathbb{R})$]. Therefore τ may be considered as a sub-Markovian kernel from $\mathbb{R}_+$ into $\mathbb{R}$ and it follows from the Laplace representation for φ_λ that for every $P \in \mathscr{M}^1(\mathbb{R}_+)$ we have

$$\mathscr{F}P(\lambda) = \widehat{\tau P}(\lambda) \quad \text{for all } \lambda \in \mathbb{R}_+$$

where $\tau P(A) = \int \tau_x(A) P(dx)$ for every Borel measurable subset A of $\mathbb{R}$, and ^ denotes the usual Fourier transform on the real line (which should be well distinguished from the Fourier transform $\mathscr{F}$ of $\mathbb{R}_+$ considered as a hypergroup).

3. Concretizations of Hypergroups

The main problem which makes it difficult to state probabilistic results on a hypergroup K is the fact that the definition does not allow us to define the "product" of two independent random variables X and Y with values in K as a K-valued random variable $X \cdot Y$ directly. It is clear, however, that the distribution of this product – if it exists – should be $P_X * P_Y$. It is the purpose of this paragraph to construct such a random variable, unifying the different approaches which have been made in concrete examples.

3.1. Definition. Let $(K, *)$ be a hypergroup (not necessarily commutative), μ a probability measure on a compact set M and $\Phi: K \times K \times M \to K$ be Borel-measurable. (M, μ, Φ) is called a *concretization* of $(K, *)$ if

$$\mu\{\Phi(x, y, \cdot) \in A\} = (\varepsilon_x * \varepsilon_y)(A) \quad \text{for} \quad x, y \in K,\ A \in \mathfrak{B}(K).$$

Here $\mathfrak{B}(K)$ denotes the Borel σ-field of K.

3.2. Examples

3.2.1. Let G be a locally compact group and $*$ the convolution defined by the group operation. If we define $\Phi(x, y, 1) := xy$ for $x, y \in G$ then $(\{1\}, \varepsilon_1, \Phi)$ is a concretization of $(G, *)$.

3.2.2. More generally let H be a compact subgroup of G and $(G//H, *)$ the double coset hypergroup (see Jewett [20]). Then (H, ω_H, Φ) is a concretization of $(G//H, *)$ if we define $\Phi(x, y, h) := H\varphi(x)h\varphi(y)H$ where $\varphi: G//H \to G$ is measurable and satisfies $x = H\varphi(x)H$ for every $x \in G//H$ (if G is locally compact, metrizable, and separable the existence of φ follows from [4]).

3.2.3. Let $K := \mathbb{R}_+$, $\varepsilon_x * \varepsilon_y := \frac{1}{2}\varepsilon_{|x-y|} + \frac{1}{2}\varepsilon_{x+y}$, $M := \{-1, 1\}$, $\mu := \frac{1}{2}\varepsilon_{-1} + \frac{1}{2}\varepsilon_1$, and $\Phi(x, y, \lambda) := |x + \lambda y|$. Then (M, μ, Φ) is a concretization of $(\mathbb{R}_+, *)$.

3.2.4. Let $\alpha > -\frac{1}{2}$ and $(\mathbb{R}_+, *)$ be the hypergroup defined in [22, 18] (see also [11]) by

$$\varepsilon_x * \varepsilon_y := c_\alpha \int_{-1}^{1} \varepsilon_{\sqrt{x^2+y^2-2\lambda xy}} (1-\lambda^2)^{\alpha-1/2} d\lambda \qquad (x, y \in \mathbb{R}_+)$$

with $c_\alpha := \dfrac{\Gamma(\alpha+1)}{\Gamma(\alpha+\frac{1}{2}) \cdot \sqrt{\pi}}$. A concretization of this hypergroup is given by $M := [-1, 1]$, $\mu := g \cdot \lambda_{[-1,1]}$ (where $g(\eta) := c_\alpha \cdot (1-\eta^2)^{\alpha-1/2}$ and $\lambda_{[-1,1]}$ denotes the Lebesgue measure on $[-1, 1]$) and $\Phi(x, y, \lambda) := \sqrt{x^2+y^2-2\lambda xy}$.

3.2.5. Let $\alpha > -\frac{1}{2}$ and $([0, \pi], *)$ the hypergroup defined in [2] by

$$\varepsilon_x * \varepsilon_y := c_\alpha \int_{-1}^{1} \varepsilon_{\arccos(\cos x \cos y + \lambda \sin x \sin y)} (1-\lambda^2)^{\alpha-1/2} d\lambda \quad \text{for} \quad x, y \in [0, \pi].$$

A concretization of this hypergroup is given by $M := [-1, 1]$, $\mu := g \cdot \lambda_{[-1,1]}$ as in 3.2.4 and

$$\Phi(x, y, \lambda) := \arccos(\cos x \cos y + \lambda \sin x \sin y).$$

3.2.6. If we choose M and μ as in 3.2.4 and

$$\Phi(x, y, \lambda) := \operatorname{arch}(\cosh x \cosh y + \lambda \sinh x \sinh x) \quad \text{for} \quad x, y, \in \mathbb{R}_+, \ \lambda \in [-1, 1],$$

we obtain a concretization of the hyperbolic hypergroup [21, 26, 27].

3.2.7. Let $\alpha > \beta > -\frac{1}{2}$ and consider the hypergroup operation on $\mathbb{R}_+$ defined in [12]. In this case a concretization is given by

$$M := [0, 1] \times [0, \pi], \qquad \mu := g \cdot \lambda_M^2$$

[with $g(r, \vartheta) := c_{\alpha,\beta}(1-r^2)^{\alpha-\beta-1} r^{2\beta+1} (\sin\vartheta)^{2\beta}$], and

$$\Phi(x, y, (r, \vartheta)) := \operatorname{arch}\left[\frac{1}{2}(1+\cosh x)(1+\cosh y) + \frac{r^2}{2}(1-\cosh x)(1-\cosh y) + r\cos\vartheta \sinh x \sinh y\right].$$

3.2.8. Consider the convolution

$$\varepsilon_x * \varepsilon_y := \sum_{s=0}^{x \wedge y} \frac{|x-y|+2s+1}{(x+1)(y+1)} \varepsilon_{|x-y|+2s} (x, y \in \mathbb{N})$$

on the nonnegative integers $\mathbb{N}$ (compare [8] and [15]). In [15] a concretization of this hypergroup is given by $M := \mathbb{S}^2$ (the sphere in $\mathbb{R}^3$), μ the uniform distribution on $\mathbb{S}^2$ and

$$\Phi(x, y, D) := |x-y| + 2\lfloor \tfrac{1}{2}\{\|(x+1)e_0 + (y+1)D\| - |x-y|\}\rfloor$$

(where e_0 is a fixed unit vector in $\mathbb{R}^3$ and $\|\ \|$ the Euclidean norm).

3.3. Definition. In the sequel let (M, μ, Φ) be a concretization of the hypergroup $(K, *)$ and $(\Omega, \mathfrak{A}, P)$ be a probability space. If X and Y are K-valued random variables and if Λ is an M-valued random variable, independent from (X, Y) and satisfying $P_\Lambda = \mu$ we define

$$X \overset{\Lambda}{\cdot} Y := \Phi(X, Y, \Lambda).$$

This is a K-valued random variable.

More generally let $(X_n : n \geqq 1)$ be a sequence of K-valued random variables and $(\Lambda_n : n \geqq 1)$ be a sequence of M-valued random variables with $P_{\Lambda_n} = \mu$ for $n \geqq 1$ and such that $X_1, \Lambda_1, X_2, \Lambda_2, \ldots$ are independent. Then we define ${}_\Lambda\prod_{j=1}^{n}$ recursively by

$${}_\Lambda\prod_{j=1}^{0} X_j := e \quad \text{and} \quad {}_\Lambda\prod_{j=1}^{n} X_j := X_n \overset{\Lambda_n}{\cdot} {}_\Lambda\prod_{j=1}^{n-1} X_j$$

for $n \geqq 1$.

It is clear that $\left({}_\Lambda\prod_{j=1}^{n} X_j : n \in \mathbb{N}\right)$ is a (non homogeneous) Markov chain, the transition kernel being

$$P\left\{{}_\Lambda\prod_{j=1}^{n} X_j \in A \,\middle|\, {}_\Lambda\prod_{j=1}^{n-1} X_j = x\right\} = (P_{X_n} * \varepsilon_x)(A) \quad P\text{-a.s.}$$

If the hypergroup is commutative we will write $X \overset{\Lambda}{+} Y$ instead of $X \overset{\Lambda}{\cdot} Y$ and ${}_\Lambda\sum_{j=1}^{n}$ instead of ${}_\Lambda\prod_{j=1}^{n}$.

3.4. Proposition. *Let (X, Y, Λ) be independent and $P_\Lambda = \mu$. Then $P_{X \overset{\Lambda}{\cdot} Y} = P_X * P_Y$.*

Proof.

$$\begin{aligned} P_{X \overset{\Lambda}{\cdot} Y}(A) &= P\{\Phi(X, Y, \Lambda) \in A\} \\ &= \int\int P\{\Phi(x, y, \Lambda) \in A\} P_X(dx) P_Y(dy) \\ &= \int\int \mu\{\Phi(x, y, \cdot) \in A\} P_X(dx) P_Y(dy) \\ &= \int\int \varepsilon_x * \varepsilon_y(A) P_X(dx) P_Y(dy) \\ &= (P_X * P_Y)(A) \quad \text{for} \quad A \in \mathfrak{B}(K). \quad \square \end{aligned}$$

3.5. *Remark.* It is clear that a concretization is not uniquely determined by the hypergroup and hence the same is true for $X \overset{\Lambda}{\cdot} Y$. However, the following proposition shows that the joint distribution of X, Y, and $X \overset{\Lambda}{\cdot} Y$ does not depend on the choice of the concretization.

3.6. Proposition. *Let X, Y, X', and Y' be K-valued random variables with $P_{(X,Y)} = P_{(X',Y')}$. Furthermore let (M, μ, Φ) and (M', μ', Φ') be concretizations of $(K, *)$ and*

Λ, Λ' be M-valued resp. M'-valued random variables such that Λ is independent of (X, Y) and Λ' is independent of (X', Y') with $P_\Lambda = \mu$ and $P_{\Lambda'} = \mu'$. Then $P_{(X,Y,X\overset{\Lambda}{*}Y)} = P_{(X',Y',X'\overset{\Lambda'}{*}Y')}$.

Proof. For every $A, B, C \in \mathfrak{B}(K)$ we have

$$\begin{aligned}
P_{(X,Y,X\overset{\Lambda}{*}Y)}(A\times B\times C) &= \int_{A\times B} P\{\Phi(x,y,\Lambda)\in C\} P_{(X,Y)}(d(x,y)) \\
&= \int_{A\times B} \varepsilon_x * \varepsilon_y(C) P_{(X',Y')}(d(x,y)) \\
&= \int_{A\times B} P\{\Phi'(x,y,\Lambda')\in C\} P_{(X',Y')}(d(x,y)) \\
&= P_{(X',Y',X'\overset{\Lambda'}{*}Y')}(A\times B\times C). \quad \square
\end{aligned}$$

3.7. Corollary. *If $(X_n, \Lambda_n : n\in\mathbb{N})$ are independent with $P_{\Lambda_n} = \mu$ then the distribution of $\left(X_n, {}^{\Lambda}\prod_{j=1}^{n} X_j : n\geqq 1\right)$ does not depend on the concretization (M, μ, Φ) or on the choice of $(\Lambda_n : n\geqq 1)$.*

3.8. Proposition. *Let $(K, *)$ be a (locally compact) hypergroup with countable base of topology. Then there exists a mapping $\Phi: K\times K\times [0,1]\to K$ such that $([0,1], \lambda_{[0,1]}, \Phi)$ is a concretization of $(K, *)$.*

Proof. We will only treat the case that K is not countable (if K is at most countable we may construct Φ in the same way as below without worrying about measurability). It follows from the assumptions that there exists a bimeasurable bijection $\psi: K\to[0,1]$. The induced mapping $\mu\mapsto\psi(\mu)$ from $\mathscr{M}^1(K)$ onto $\mathscr{M}^1([0,1])$ is Borel measurable and hence so is the mapping $p:[0,1]^2\to\mathscr{M}^1([0,1])$ defined by

$$p(x,y) := \psi(\varepsilon_{\psi^{-1}(x)} * \varepsilon_{\psi^{-1}(y)}).$$

It is a well known fact that for every $\mu\in\mathscr{M}^1[0,1]$ there is a unique left continuous increasing function $\varphi_\mu:[0,1]\to[0,1]$ such that $\varphi_\mu(\lambda_{[0,1]}) = \mu$, namely

$$\varphi_\mu(\eta) = 0 \vee \sup\{z\in[0,1] : \mu([0,z]) < \lambda\}.$$

We will prove that the mapping $\Phi_0 : [0,1]^3\to[0,1]$ defined by $\Phi_0(x,y,\lambda) := \varphi_{p(x,y)}(\lambda)$ for all $x, y, \lambda\in[0,1]$ is Borel measurable. It follows from the left sided continuity of φ_μ that it suffices to show that for every $\lambda\in[0,1]$ the mapping

$$\mu\mapsto\varphi_\mu(\lambda) = 0\vee\sup\{z\in[0,1] : \mu([0,z[) < \lambda\}$$

is Borel measurable. This, however, is a consequence of

$$\{\mu\in\mathscr{M}^1([0,1]) : \varphi_\mu(\lambda)\leqq\varepsilon\} = \{\mu\in\mathscr{M}^1([0,1]) : \int 1_{[0,\varepsilon]} d\mu \geqq \lambda\}.$$

The mapping $\Phi: K\times K\times[0,1]\to K$ can therefore be defined by

$$\Phi(h,k,\lambda) := \psi^{-1}(\Phi_0(\psi(h), \psi(k), \lambda)). \quad \square$$

3.9. We are now considering the special cases $K = \mathbb{R}_+$ and $K = [0,1]$ – see [1, 5, 28]. It follows from [28] that we may suppose without loss of generality that

$$\min\operatorname{supp}(\varepsilon_x * \varepsilon_y) = |x-y| \quad \text{for} \quad x, y\in K$$

and

$$\max \operatorname{supp}(\varepsilon_x * \varepsilon_y) = x+y \quad \text{if} \quad x, y \in K \text{ (and } x+y \leqq 1 \text{ in the case } K=[0,1]).$$

Since in these cases there is no need for a Borel isomorphism ψ in the proof of 3.8, we get the following additional properties of Φ:

$$\Phi(x,y,0) = |x-y| \quad \text{for} \quad x,y \in K$$

and

$$\Phi(x,y,1) = x+y \quad \text{for} \quad x,y \in K \text{ (and } x+y \leqq 1 \text{ in the case } K=[0,1]).$$

Furthermore $\Phi(x,0,\lambda) = \Phi(0,x,\lambda) = x$. Every hypergroup on $\mathbb{R}_+$ or $[0,1]$ is commutative [28, Corollary 2.4] and hence

$$\Phi(x,y,\lambda) = \Phi(y,x,\lambda) \quad \text{for} \quad x,y \in K, \ \lambda \in [0,1].$$

It is easy to prove that for every $\lambda \in [0,1]$ the mapping $\Phi(\cdot,\cdot,\lambda): K \times K \to K$ is lower semicontinuous. Under the additional assumptions that $\varepsilon_x * \varepsilon_y$ is diffuse for $x,y>0$ and $\operatorname{supp}(\varepsilon_x * \varepsilon_y) = [|x-y|, x+y]$ (which happens to be true if $(\mathbb{R}_+, *)$ is a Chébli-Trimèche hypergroup as shown by Trimèche [25, Sect. 8]), $\Phi(\cdot,\cdot,\lambda)$ is continuous for every $\lambda \in [0,1]$.

4. Moments

The usual definition of moments of higher order on a locally compact group [16, 18] depends on the choice of the gauge function $\delta_V(x) := \inf\{n \geqq 1 : x \in V^n\}$ and essentially only states whether the moment of order α ($\alpha > 0$) exists or not. On Chébli-Trimèche hypergroups however, the moment of order n of a probability measure can be defined for every integer $n \geqq 1$ in a unique way fitting (in the sense of 4.14) with the convolution structure of the hypergroup.

From now on let $(K, *)$ be a Chébli-Trimèche hypergroup on $\mathbb{R}_+$ (see 2.2). It is proved in [5] that $\varphi_\lambda(x)$ is an analytic function of λ. The derivations of $\varphi_\lambda(x)$ with respect to λ will be the most important tool to define moments for each probability measure on $\mathbb{R}_+$ in a way which is consistent with the convolution structure.

4.1. Definition. For every $\lambda \in \mathbb{C}$, $x \in \mathbb{R}_+$, and $n \geqq 0$ let

$$\varphi_{n,\lambda}(x) := \left(\frac{\partial}{\partial \mu}\right)^n \varphi_{\lambda + i\mu}(x)|_{\mu=0} \quad \text{and} \quad m_n(x) := \varphi_{n, i\varrho}(x).$$

Some elementary properties of the functions $\varphi_{n,\lambda}$ and m_n will be proved first.

4.2. Let L be the differential operator on $\mathbb{R}_+$ defined by $Lf(x) = -f''(x) - \dfrac{A'(x)}{A(x)} f'(x)$ for $x>0$ and $f \in \mathscr{C}^2(\mathbb{R}_+)$ with $f'(0)=0$. By differentiating the differential equation $\varphi_\lambda = (\varrho^2 + \lambda^2)\varphi_\lambda$, $\varphi_\lambda(0)=1$, $\varphi'_\lambda(0)=0$ with respect to λ we obtain

$$L\varphi_{n,\lambda} = (\varrho^2 + \lambda^2)\varphi_{n,\lambda} + 2in\lambda\varphi_{n-1,\lambda} - n(n-1)\varphi_{n-2,\lambda}, \quad \varphi_{n,\lambda}(0) = \varphi'_{n,\lambda}(0) = 0$$

and especially

$$Lm_n = -2n\varrho m_{n-1} - n(n-1)m_{n-2}, \quad m_n(0) = m_n'(0) = 0 \quad \text{for} \quad n \geqq 1$$

(with $m_0(x) = 1$ for every $x \in \mathbb{R}_+$).

4.3. It follows from the Laplace representation (2.3) that

$$\varphi_{n,\lambda}(x) = \int_{-x}^{x} t^n e^{-t(\varrho+i\lambda)} \nu_x(dt) = \int_{-x}^{x} t^n e^{-it\lambda} \tau_x(dt)$$

and

$$m_n(x) = \int_{-x}^{x} t^n \nu_x(dt) \quad \text{for} \quad x \in \mathbb{R}_+,\ \lambda \in \mathbb{C},\ n \geqq 1.$$

4.4. If $\lambda \in i[0,\varrho]$ $\varphi_{n,\lambda}$ is real valued for every $n \geqq 1$ since φ_λ is real valued. For $\lambda \in \mathbb{R}_+$ $\varphi_{n,\lambda}$ is real valued if n is even and $i\varphi_{n,\lambda}$ is real if n is odd. This follows since $\varphi_\lambda(x)$ is an analytic function of λ and φ_λ is real for $\lambda \in \mathbb{R}_+$.

It follows from $m_n(x) = \int_0^x t^n(e^{\varrho t} + (-1)^n e^{-\varrho t})\tau_x(dt)$ that $m_n \geqq 0$ for every $n \geqq 1$.

4.5. We now have to study the two cases $\varrho = 0$ and $\varrho > 0$ separately. We begin with the case $\varrho = 0$. It is clear that $m_n = 0$ if n is odd.

4.6. Lemma. *Let $\varrho = 0$, $\lambda \in \mathbb{R}_+$, and $n \in \mathbb{N}$. Then*

a) $m_{2k} \leqq 1 + m_{2n}$ *for every* $k < n$,
b) $|\varphi_{2n,\lambda}| \leqq m_{2n}$, *and*
c) $|\varphi_{2n-1,\lambda}| \leqq 1 + m_{2n}$.

Proof.

a) $$m_{2k}(x) = \int t^{2k} \nu_x(dt) \leqq \int (1 + t^{2n}) \nu_x(dt) = 1 + m_{2n}(x),$$

b) $$|\varphi_{2n,\lambda}(x)| \leqq \int |t^{2n} e^{-it\lambda}| \nu_x(dt) = \int t^{2n} \nu_x(dt) = m_{2n}(x),$$

c) $$|\varphi_{2n-1,\lambda}(x)| \leqq \int |t^{2n-1} e^{-it\lambda}| \nu_x(dt) = \int |t|^{2n-1} \nu_x(dt)$$
$$\leqq \int (1 + t^{2n}) \nu_x(dt) = 1 + m_{2n}(x). \quad \square$$

4.7. Theorem. *Let P be a probability measure on $\mathbb{R}_+$ and $n \geqq 1$. Then the following conditions are equivalent:*

(i) $\int m_{2n} dP$ *is finite,*
(ii) $\mathscr{F}P$ *is* $2n-1$ *times differentiable,* $\mathscr{F}P^{(2n-1)}(0) = 0$ *and* $\mathscr{F}P^{(2n)}(0)$ *exists.*

In both cases $\mathscr{F}P^{(k)}(\lambda) = i^k \int \varphi_{k,\lambda} dP$ *for all* $k \leqq 2n$, $\lambda \in \mathbb{R}_+$. *In particular* $\mathscr{F}P^{(2k)}(0) = \int m_{2k} dP$.

Proof. "i ⇒ ii": By 4.6 a) and induction $\mathscr{F}P^{(2n-1)}$ exists. From 4.6 c) and b) we obtain

$$\left| \frac{\varphi_{2n-2,\lambda} - \varphi_{2n-2,\mu}}{\lambda - \mu} \right| \leqq m_{2n} + 1$$

and

$$\left| \frac{\varphi_{2n-1,\lambda} - \varphi_{2n-1,\mu}}{\lambda - \mu} \right| \leqq m_{2n}$$

and it follows from the dominated convergence theorem that $\mathcal{F}P^{(2n-1)}(\lambda)$ and $\mathcal{F}P^{(2n)}(\lambda)$ exist and equal $i^{2n-1}\cdot\int\varphi_{2n-1,\lambda}dP$ and $i^{2n}\cdot\int\varphi_{2n,\lambda}dP$ respectively. In particular

$$\mathcal{F}P^{(2n-1)}(0)=\int\varphi_{2n-1,0}dP=\int m_{2n-1}dP=0.$$

"ii $\Rightarrow$ i": Because of $\mathcal{F}P^{(2n-1)}(0)=0$ the $2n$-th derivative $\mathcal{F}P^{(2n)}(0)$ equals

$$2\lim_{h\to 0}\frac{1}{h^2}(\mathcal{F}P^{(2n-2)}(h)-\mathcal{F}P^{(2n-2)}(0)).$$

Since $\varphi_{2n-2,\lambda}\leqq m_{2n-2}$ by 4.6 b) we may apply Fatou's lemma to obtain

$$\begin{aligned}
0&\leqq\int m_{2n}dP\\
&=-\int\left(\frac{\partial}{\partial h}\right)^2\varphi_{2n-2,h}|_{h=0}dP\\
&=2\int\lim_{h\to 0}\frac{1}{h^2}(m_{2n-2}-\varphi_{2n-2,h})dP\\
&\leqq 2\liminf_{h\to 0}\frac{1}{h^2}\int m_{2n-2}-\varphi_{2n-2,h}dP\\
&=2\liminf_{h\to 0}\frac{1}{h^2}(\mathcal{F}P^{(2n-2)}(0)-\mathcal{F}P^{(2n-2)}(h))\\
&=-2\mathcal{F}P^{(2n)}(0)<\infty.\quad\square
\end{aligned}$$

4.8. *Remark.* The condition $\mathcal{F}P^{(2n-1)}(0)=0$ – which does not occur in the usual formulation of this theorem on $\mathbb{R}$ – cannot be dismissed. For example in the case of Kingman's hypergroups [$A(x)=x^{2\alpha+1}$, $\alpha>-\frac{1}{2}$, see 3.2.4] the (Cauchy type) distribution P with density

$$x\mapsto\frac{2\Gamma(\alpha+\frac{3}{2})}{\sqrt{\pi}\,\Gamma(\alpha+1)}\,\frac{x^{2\alpha+1}}{(1+x^2)^{\alpha+3/2}}$$

has the Fourier transform $\lambda\mapsto e^{-\lambda}$ (see [7, 8.6(4)]) which is infinitely often differentiable on $\mathbb{R}_+$ but $\int m_2dP=\infty$.

Let us now suppose that $\varrho>0$. Then m_n is strictly positive on $\mathbb{R}_+^*$ for every $n\geqq 0$. In view of (7.6) and (5.4) this fact corresponds to the corollary on p. 77 in [16].

4.9. Lemma. *Let $\varrho>0$ and $n\geqq 0$. Then*

a) $\varphi_{n,i\lambda}>0$ *if* $\lambda\in[0,\varrho]$,

b) $\varphi_{n,i\lambda}\leqq\varphi_{n,i\mu}$ *if* $0\leqq\lambda\leqq\mu\leqq\varrho$, *and*

c) $\dfrac{m_n-\varphi_{n,i\lambda}}{\varrho-\lambda}\leqq\dfrac{m_n-\varphi_{n,i\mu}}{\varrho-\mu}$ *if* $0\leqq\lambda\leqq\mu<\varrho$.

Proof. a) $\varphi_{n,i\lambda}(x)=\int t^ne^{-t(\varrho-\lambda)}\nu_x(dt)>0$.

b)
$$\begin{aligned}
\varphi_{n,i\lambda}(x)&=\int_{-x}^{x}t^ne^{-t(\varrho-\lambda)}\nu_x(dt)=\int_0^x t^n(e^{t\lambda}+(-1)^ne^{-t\lambda})\tau_x(dt)\\
&\leqq\int_0^x t^n\cdot(e^{t\mu}+(-1)^ne^{-t\mu})\tau_x(dt)\\
&=\varphi_{n,i\mu}(x)
\end{aligned}$$

since sinh and cosh are increasing functions.

c) $$\frac{m_n(x)-\varphi_{n,i\lambda}(x)}{\varrho-\lambda}=\int_0^x t^n\frac{(e^{t\varrho}+(-1)^n e^{-t\varrho})-(e^{t\lambda}+(-1)^n\cdot e^{-t\lambda})}{\varrho-\lambda}\tau_x(dt)$$
$$\leqq\int_0^x t^n\frac{(e^{t\varrho}+(-1)^n e^{-t\varrho})-(e^{t\mu}+(-1)^n e^{-t\mu})}{\varrho-\mu}\tau_x(dt)$$
$$=\frac{m_n(x)-\varphi_{n,i\mu}(x)}{\varrho-\mu}$$

since sinh and cosh are both convex functions. □

4.10. Lemma. *Let* $\varrho>0$. *Then* $m_n(x)\leqq\left(\frac{2}{\varrho}\right)^n+\varrho m_{n+1}(x)$ for all $x\geqq 0$, $n\in\mathbb{N}$.

Proof. If n is odd we conclude from $\sinh y\leqq y\cosh y$ for all $y\geqq 0$ that

$$m_n(x)=\int_0^x t^n\cdot\sinh(t\varrho)\tau_x(dt)$$
$$\leqq\varrho\int_0^x t^{n+1}\cosh(t\varrho)\tau_x(dt)$$
$$=\varrho m_{n+1}(x)\quad\text{for all } x\geqq 0.$$

If n is even we use the inequality $\cosh y\leqq y\sinh y+1_{[0,2[}(y)$ to obtain $y^n\cosh y\leqq y^{n+1}\sinh y+2^n$ which implies

$$m_n(x)=\int_0^x t^n\cosh(t\varrho)\tau_x(dt)$$
$$\leqq\int_0^x(\varrho t^{n+1}\sinh(t\varrho)+(2/\varrho)^n)\tau_x(dt)$$
$$\leqq\varrho m_{n+1}(x)+(2/\varrho)^n\quad\text{for all } x\geqq 0.\quad\square$$

4.11. Theorem. *Let* $\varrho>0$, $n\geqq 1$, *and* P *be a probability measure on* $\mathbb{R}_+$. *Then the following conditions are equivalent:*

(i) $\int m_n dP$ *is finite,*

(ii) $\lambda\mapsto\mathscr{F}P(i\lambda)$ *is* n *times differentiable on* $[0,\varrho]$.

In both cases $\left(\frac{\partial}{\partial\lambda}\right)^k\mathscr{F}P(i\lambda)=\int\varphi_{k,i\lambda}dP$ *for all* $k\leqq n$, $\lambda\in[0,\varrho]$. *In particular* $\mathscr{F}P^{(k)}(0)=\int m_k dP$.

Proof. "i ⇒ ii": It follows from 4.10 and induction that $f:\lambda\mapsto\int\varphi_{i\lambda}dP$ is $(n-1)$ times differentiable on $[0,\varrho]$ and $f^{(n-1)}(\lambda)=\int\varphi_{n-1,i\lambda}dP$. The mean value theorem and 4.9 b) imply that

$$\varphi_{n-1,i\lambda}(x)-\varphi_{n-1,i\mu}(x)|\leqq|\lambda-\mu|\cdot m_n(x).$$

By Lebesgue's theorem we obtain that $f^{(n-1)}$ is differentiable and

$$\left(\frac{\partial}{\partial\lambda}\right)^n\mathscr{F}P(i\lambda)=f^{(n)}(\lambda)=\int\varphi_{n,i\lambda}dP.$$

"*ii* ⇒ i": By induction the first $n-1$ moments $\int m_k dP$ ($k \leqq n-1$) are finite and $f^{(n-1)}(\lambda) = \int \varphi_{n-1,i\lambda} dP$. It follows from 4.9c) that the difference quotients $\frac{m_{n-1} - \varphi_{n-1,i\lambda}}{\varrho - \lambda}$ ($\lambda < \varrho$) approach m_n increasingly as $\lambda \nearrow \varrho$ and hence by the theorem of monotone convergence

$$\int m_n dP = \frac{\partial}{\partial \lambda} \int \varphi_{n-1,i\lambda} dP|_{\lambda=\varrho}$$

exists and equals $f^{(n)}(\varrho)$ which is finite. □

4.12. *Remark.* Let $\varrho > 0$ and $\int m_n dP$ be finite. Then it follows from 4.3 and 4.9 b) that $|\varphi_{n,\lambda}(x)| \leqq m_n(x)$ for all $x \geqq 0$ and $\lambda \in \mathbb{C}$ such that $|\Im \lambda| \leqq \varrho$. Therefore the function $\lambda \mapsto \int \varphi_\lambda dP$ is n times differentiable in this strip and in particular $\eta \mapsto \int \varphi_{\eta + i\varrho} dP$ is n times differentiable on $\mathbb{R}$. This fact will be used later.

4.13. *Remark.* Let $\varrho \geqq 0$. Then a sufficient condition for $E(m_n(X))$ being finite (where X is a $\mathbb{R}_+$-valued random variable) is $E(X^n) < \infty$. This follows from the inequality $m_n(x) \leqq x^n$ for all $x \geqq 0$ which is a consequence of the fact that the measure ν_x in 4.3 is supported by $[-x, x]$.

4.14. Theorem. *Let $\varrho > 0$, X and Y be independent $\mathbb{R}_+$-valued random variables such that $E(m_n(X))$ and $E(m_n(Y))$ are finite. Then $E(m_n(X \overset{A}{+} Y))$ is finite and*

$$E(m_n(X \overset{A}{+} Y)) = \sum_{k=0}^{n} \binom{n}{k} E(m_k(X)) E(m_{n-k}(Y)).$$

Proof. This follows from the fact that the product of two n times differentiable functions is again n times differentiable, Theorem 4.11, and Leibniz's rule. □

5. Expectation

5.1. In this paragraph the special properties of the function m_1 will be considered. We will assume $\varrho > 0$ throughout the whole paragraph. The function m_1 has already been defined in [9, 24] under the name "forme quadratique généralisée". It will be used to define a modified expectation for every $\mathbb{R}_+$-valued random variable consistent with the hypergroup structure (see 5.6).

5.2. *Examples.* If A is of the form $A(x) = (\sinh x)^\alpha$ for some $\alpha > 0$, the function m_1 can be written down in closed form for some values of α. According to Faraut [9], $m_1(x) = 2 \ln \cosh \frac{x}{2}$ if $\alpha = 1$ and $m_1(x) = x \coth x - 1$ if $\alpha = 2$. If $\alpha = 3/2$ one calculates $m_1(x) = 2 \ln \cosh \frac{x}{2} + \frac{1}{2}\left(\tanh \frac{x}{2}\right)^2$ ($x \geqq 0$). If $A(x) = (\cosh x)^2$ then $m_1(x) = x \tanh x$.

5.3. *Remark.* By integrating the differential equation for m_1 (see 4.2), one obtains

$$m_1(x) = 2\varrho \int_0^x \frac{1}{A(y)} \int_0^y A(z) dz\, dy \quad \text{for} \quad x \geqq 0.$$

This formula has been used in [13].

5.4. Definition. Let $(\mathbb{R}_+, *)$ be a Chébli-Trimèche hypergroup with the corresponding function m_1. Then for every $\mathbb{R}_+$-valued random variable X, $E_*(X) := E(m_1(X))$ is called the *$*$-expectation* of X.

5.5. *Remark.* Although $E_*(X)=0$ holds for every random variable in the case $\varrho=0$, the $*$-expectation does not loose its entire sense as can be seen from Theorem 8.4. The notions of "dispersion" and "variance" are also used for $E_*(X)$ by some authors (see [9, 13, 24]). Theorems 7.6, 7.7, and 8.4, as well as the central limit theorems in the forthcoming paper [29] are the motivation to call $E_*(X)$ the "expectation" of X and to reserve the expression "variance" to the corresponding number related with the second moment function m_2.

5.6. Proposition. *Let A have the distribution μ and be independent from (X, Y). Then $E_*(X \overset{A}{+} Y) = E_*(X) + E_*(Y)$.*

The proof follows from 4.14. □

5.7. Lemma. *If $\varrho>0$ then* $\lim_{x\to\infty} \frac{m_1(x)}{x} = 1$.

Proof. Suppose that m_1'' takes negative values. Since $m_1''(0) = \frac{2\varrho}{\alpha+1} > 0$ there exists $x_0 > 0$ with $m_1'(x_0) > 0$, $m_1''(x_0) < 0$, and $m_1'''(x_0) < 0$. This would imply that $m_1' \cdot \frac{A'}{A}$ and m_1'' are strictly decreasing in a neighbourhood of x_0. But this is impossible since $m_1'' + m_1' \frac{A'}{A} = 2\varrho$ by 4.2. From this contradiction we conclude that $m_1'' \geqq 0$ and m_1' is increasing. Suppose now that $\beta := \lim_{x\to\infty} m_1'(x) < 1$. This implies $m_1'' = 2\varrho - \frac{A'}{A} \cdot m_1'$ $> 2\varrho - \frac{A'}{A}\beta$. When $\frac{A'}{A}(x)$ is close enough to 2ϱ the last number becomes strictly positive and hence $m_1''(x)$ is bounded away from 0 for large enough x. This is a contradiction with $\sup\{m_1'(x) : x \geqq 0\} = \beta < 1$. On the other hand, from $2\varrho m_1'$ $\leqq m_1'' + \frac{A'}{A} m_1' = 2\varrho$ we obtain $m_1' \leqq 1$ and hence $m_1'(x) \nearrow 1$ as $x \to \infty$. This implies $\frac{m_1(x)}{x} \nearrow 1$. □

5.8. Corollary. *Let $\varrho>0$ and X be a $\mathbb{R}_+$-valued random variable. Then $E_*(X)$ is finite if and only if $E(X)$ (the expectation of X in the usual sense) is finite.*

5.9. Proposition. *Let $\varrho>0$ and X be a $\mathbb{R}_+$-valued random variable with $0 \leqq E_*(X)$ $\leqq +\infty$. Then*

$$\frac{\partial}{\partial\lambda} E(\varphi_{i\lambda}(X))|_{\lambda=\varrho} = E_*(X).$$

Proof. If $E_*(X)$ is finite this is Theorem 4.11. If $E_*(X) = \infty$ this follows from 4.9 and the theorem of monotone convergence. □

6. Variance

We will now explore the properties of the function m_2 to obtain a modification of the notion of variance in a similar way as for the expectation in the last paragraph.

6.1. *Examples.* If $A(x)=x^\alpha$ ($\alpha \geqq 0$) we obtain $m_2(x)=\dfrac{1}{\alpha+1}x^2$. If $A(x)=(\sinh x)^2$, $m_2(x)=x^2+2-2x\cosh x$ [26, p. 191] and in the case $A(x)=(\cosh x)^2$ we have $m_2(x)=x^2$.

6.2. Lemma. *$m_2(x)^2 \leqq m_2(x) \leqq x^2$ for every $x \geqq 0$.*

Proof. The first inequality follows from 4.3 and Jensen's inequality; the second has already been proved in 4.13. □

6.3. Corollary. *Let $\varrho>0$. Then $\lim\limits_{x\to\infty}\dfrac{1}{x^2}m_2(x)=1$.*

6.4. Lemma. *If $\varrho=0$ then m_2 is a convex function and $\lim\limits_{x\to\infty}\dfrac{m_2(x)}{x}=+\infty$.*

Proof. The convexity of m_2 follows in the same way as the convexity of m_1 in the first part of the proof of 5.7. The assumption that m_2' is bounded leads to a contradiction since it implies

$$\lim_{x\to\infty} m_2''(x)=2-\lim_{x\to\infty} m_2'(x)\frac{A'(x)}{A(x)}=2.$$

Hence m_2' is unbounded. But from $m_2'(x)\nearrow\infty$ as $x\to\infty$ follows $\lim\limits_{x\to\infty}\dfrac{m_2(x)}{x}=+\infty$ by the mean value theorem. □

6.5. Lemma. *Suppose that $\varrho=0$ and $\left\{x\dfrac{A'(x)}{A(x)}: x>0\right\}$ is bounded. Then there exists $\gamma>0$ such that $m_2(x)\geqq\gamma x^2$ for $x\geqq 0$.*

Proof. It follows from the differential equation $m_2''+\dfrac{A'}{A}m_2'=2$ that the function ψ defined on $\mathbb{R}_+$ by

$$\psi(x)=\begin{cases} m_2'(x)/x & \text{for } x>0\\ m_2''(0)=\dfrac{2}{\alpha+1} & \text{for } x=0\end{cases}$$

satisfies the differential equation

$$x\psi'(x)+\left(x\frac{A'(x)}{A(x)}+1\right)\psi(x)=2, \qquad \psi(0)=\frac{2}{\alpha+1}.$$

Let b be an upper bound for $x\dfrac{A'(x)}{A(x)}$ ($x>0$). Then for every x such that $\psi(x)<\dfrac{2}{b+1}$ we obtain $\psi'(x)>0$ and so ψ is certainly bounded from below by 2γ where $\gamma:=\min\left(\dfrac{1}{b+1},\dfrac{1}{\alpha+1}\right)$. But this implies $m_2'(x)\geqq 2\gamma x$ and hence the result. □

6.6. Definition. In order to define the $*$-*variance* for every Chébli-Trimèche hypergroup $(\mathbb{R}_+, *)$ with corresponding functions m_1 and m_2 we introduce the function $v: \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, $v(x,\xi) := m_2(x) - 2\xi m_1(x) + \xi^2$ $(x, \xi \geqq 0)$ which is non-negative by 6.2. For every $\mathbb{R}_+$-valued random variable X such that $E(m_2(X)) < \infty$ the function $\xi \mapsto E(v(X,\xi))$ on $\mathbb{R}$ takes its minimum at $\xi = E_*(X)$, this value being

$$V_*(X) := E(v(X, E_*(X))) = E(m_2(X)) - E(m_1(X))^2 \geqq 0.$$

If $E(m_2(X)) = \infty$ we define $V_*(X) = \infty$. $V_*(X)$ is called the $*$-*variance* of X.

6.7. *Remark.* In the case $\varrho = 0$ $V_*(X)$ equals $E(m_2(X))$; this number is called the "dispersion" of X in [24].

6.8. *Remark.* $V_*(X)$ is strictly positive unless $X = 0$ P-a.s. At the first look it is surprising that the $*$-variance of a constant $X \neq 0$ does not equal zero. But it reflects the fact that $X \overset{A}{+} Y$ is random even if $X > 0$ and $Y > 0$ are deterministic.

6.9. Proposition. *If* $\varrho > 0$ *or if* $\left\{ x \cdot \dfrac{A'(x)}{A(x)} : x > 0 \right\}$ *is bounded,* $V_*(X)$ *exists if and only if* $E(X^2) < \infty$.

Proof. This is a consequence of 6.3 in the first case, and 6.2 and 6.5 in the second. $\square$

6.10. Proposition. *Let X and Y be independent* $\mathbb{R}_+$*-valued random variables. Then*

$$V_*(X \overset{A}{+} Y) = V_*(X) + V_*(Y).$$

Both sides of this equation may be infinite.

Proof. If $E(m_2(X))$ or $E(m_2(Y))$ are infinite it follows from 4.14 that $E(m_2(X \overset{A}{+} Y))$ and hence $V_*(X \overset{A}{+} Y)$ equals $+\infty$. Let us therefore suppose that $V_*(X) < \infty$ and $V_*(Y) < \infty$. It follows from 4.14 that

$$\begin{aligned} V_*(X \overset{A}{+} Y) &= E(m_2(X \overset{A}{+} Y)) - E(m_1(X \overset{A}{+} Y))^2 \\ &= E(m_2(X)) + 2E(m_1(X))E(m_1(Y)) + E(m_2(Y)) \\ &\quad - E(m_1(X))^2 - 2E(m_1(X))E(m_1(Y)) - E(m_1(Y))^2 \\ &= V_*(X) + V_*(Y). \quad \square \end{aligned}$$

6.11. *Remark.* If X and Y are not independent but only $*$-*uncorrelated* (in the sense that

$$E(m_1(X)m_1(Y)) = E_*(X)E_*(Y))$$

then the assertion of Proposition 6.10 remains valid.

7. Laws of Large Numbers in the Case of Exponential Growth

Recall that (M, μ, Φ) denotes a fixed concretization of a Chébli-Trimèche hypergroup $(\mathbb{R}_+, *)$.

7.1. Proposition. *Let X, Y, and A be independent* $\mathbb{R}_+$*-,* $\mathbb{R}_+$*-, and M-valued random variables such that* $P_A = \mu$.

a) *If* $E_*(X)$ *and* $E_*(Y)$ *are finite, then* $E(m_1(X \overset{A}{+} Y)|X) = m_1(X) + E_*(Y)$ *P-almost surely.*

b) *If* $V_*(X)$ *and* $V_*(Y)$ *are finite, then*

$$E(v((X \overset{A}{+} Y), E_*(X \overset{A}{+} Y))|X) = v(X, E_*(X)) + V_*(Y) \quad P\text{-a.s.}$$

c) *If* $|\Im\lambda| \leqq \varrho$, *then*

$$E(\varphi_\lambda(X \overset{A}{+} Y)|X) = \varphi_\lambda(X) \cdot E(\varphi_\lambda(Y)) \quad P\text{-a.s.}$$

Proof. a) Let $A \in \mathfrak{B}(\mathbb{R}_+)$. Then by 5.6 it follows

$$\begin{aligned}
E(1_{\{X\in A\}} \cdot m_1(X \overset{A}{+} Y)) &= E(m_1((1_{\{X\in A\}}X) \overset{A}{+} Y)) - E(1_{\{X\notin A\}} m_1(Y)) \\
&= E(m_1(1_{\{X\notin A\}}X)) + E_*(Y) - P\{X \notin A\} E_*(Y) \\
&= E(1_{\{X\in A\}} \cdot [m_1(X) + E_*(Y)]) .
\end{aligned}$$

b) For every $A \in \mathfrak{B}(\mathbb{R}_+)$ we conclude from 4.14 that

$$\begin{aligned}
E(1_{\{X\in A\}} \cdot m_2(X \overset{A}{+} Y)) &= E(m_2(1_{\{X\in A\}}X) \overset{A}{+} Y)) - E(1_{\{X\notin A\}} m_2(Y)) \\
&= E(m_2(1_{\{X\in A\}}X)) + 2E(m_1(1_{\{X\in A\}}X))E_*(Y) \\
&\quad + E(m_2(Y)) - P\{X \notin A\} E(m_2(Y)) \\
&= E(1_{\{X\in A\}} \cdot [m_2(X) + 2m_1(X)E_*(Y) + E(m_2(Y))]) .
\end{aligned}$$

Therefore $E(m_2(X \overset{A}{+} Y)|X) = m_2(X) + 2m_1(X)E_*(Y) + E(m_2(Y))$ P-almost surely and hence

$$\begin{aligned}
E(v(X \overset{A}{+} Y, E_*(X \overset{A}{+} Y))|X) &= E(m_2(X \overset{A}{+} Y)|X) \\
&\quad - 2E_*(X \overset{A}{+} Y)E(m_1(X \overset{A}{+} Y)|X) + E_*(X \overset{A}{+} Y)^2 \\
&= m_2(X) + 2m_1(X)E_*(Y) + E(m_2(Y)) \\
&\quad - 2(E_*(X) + E_*(Y))(m_1(X) + E_*(Y)) \\
&\quad + (E_*(X) + E_*(Y))^2 \\
&= m_2(X) - 2m_1(X)E_*(X) + E_*(X)^2 \\
&\quad + E(m_2(Y)) - E_*(Y)^2 \\
&= v(X, E_*(X)) + V_*(Y) \quad P\text{-a.s.}
\end{aligned}$$

c) Since φ_λ is a bounded multiplicative function we obtain for every $A \in \mathfrak{B}(\mathbb{R}_+)$

$$\begin{aligned}
E(1_{\{X\in A\}} \cdot \varphi_\lambda(X \overset{A}{+} Y)) &= E(\varphi_\lambda((1_{\{X\in A\}}X) \overset{A}{+} Y)) - E(1_{\{X\notin A\}} \varphi_\lambda(Y)) \\
&= E(\varphi_\lambda(1_{\{X\in A\}}X))E(\varphi_\lambda(Y)) - P\{X \notin A\} E(\varphi_\lambda(Y)) \\
&= (E(1_{\{X\in A\}}\varphi_\lambda(X)) + P\{X \notin A\})E(\varphi_\lambda(Y)) \\
&\quad - P\{X \notin A\} E(\varphi_\lambda(Y)) \\
&= E(1_{\{X\in A\}} \cdot \varphi_\lambda(X)E(\varphi_\lambda(Y))) . \quad \square
\end{aligned}$$

7.2. *Notation.* For the rest of this article we suppose that $X_1, X_2, \ldots, A_1, A_2, \ldots$ are independent $\mathbb{R}_+$- resp. M-valued random variables such that $P_{A_n} = \mu$ for every $n \geqq 1$. It follows from 3.3 that the process $(S_n : n \geqq 0)$ where $S_n := {}^{A}\sum_{j=1}^{n} X_j$ is a (non homogeneous) Markov chain.

7.3. Corollary. a) *If* $E_*(X_n) < \infty$ *resp.* $V_*(X_n) < \infty$ *for* $n \geqq 1$ *then* $(m_1(S_n) : n \in \mathbb{N})$ *resp.* $(v(S_n, E_*(S_n)) : n \in \mathbb{N})$ *are submartingales with respect to the canonical filtration.*

b) *If* $\lambda \in i[0, \varrho]$ *then* $(\varphi_\lambda(S_n) : n \in \mathbb{N})$ *is a supermartingale.*

Proof. From 7.1 a) we obtain for every $n \geqq 1$

$$\begin{aligned} E(m_1(S_n)|S_{n-1}) &= E(m_1(X_n \overset{A}{+} S_{n-1})|S_{n-1}) \\ &= m_1(S_{n-1}) + E_*(X_n) \geqq m_1(S_{n-1}) \quad P\text{-a.s.} \end{aligned}$$

The other assertions can be proved in the same way. □

Note that this corollary holds for any hypergroup K and $\mathbb{R}_+$-valued functions m_1 and m_2 on K such that 4.14 and $m_2 \geqq m_1^2$ hold.

For the rest of this paragraph we suppose $\varrho > 0$. In view of $A(x) \geqq A(1) \cdot e^{2\varrho(x-1)}$ for $x \geqq 1$ this implies that $(\mathbb{R}_+, *)$ is of exponential growth.

7.4. Theorem. *Let* $(X_n : n \geqq 1)$ *be an independent series of* $\mathbb{R}_+$*-valued random variables such that* $\sum_{n=1}^{\infty} \frac{1}{n^2} V_*(X_n) < \infty$. *Then*

$$\frac{1}{n}(S_n - m_1^{-1}(E_*(S_n))) \to 0 \quad P\text{-a.s.}$$

Proof. Let $s_n := E_*(S_n)$. It follows from 6.2 and 7.3 a), that $(v(S_n, s_n) : n \geqq 1)$ is a positive submartingale. Furthermore, the assumptions and 6.10 imply that

$$\sum_{n=1}^{\infty} \frac{1}{n^2} E(v(S_n, s_n) - v(S_{n-1}, s_{n-1})) < \infty .$$

Hence by Chow's law of large numbers [6] and 6.2 we obtain

$$\left(\frac{m_1(S_n) - s_n}{n}\right)^2 \leqq \frac{1}{n^2} v(S_n, s_n) \to 0 \quad P\text{-a.s.}$$

Since $(m_1^{-1})'(t) \searrow 1$ as $t \to \infty$ there is a number $a > 0$ such that $|m_1^{-1}(x) - m_1^{-1}(y)| \leqq 2|x - y| + a$ for all $x, y \in \mathbb{R}_+$. Therefore $\frac{m_1(S_n) - s_n}{n} \to 0$ implies $\lim_{n \to \infty} \frac{1}{n}(S_n - m_1^{-1}(s_n)) = 0$. □

7.5. *Remarks.*

7.5.1. If in the situation of the preceding theorem we assume additionally that $\frac{1}{n} E_*(S_n)$ is bounded we obtain

$$\lim_{n \to \infty} \frac{1}{n}(S_n - E_*(S_n)) = 0 \quad P\text{-a.s.}$$

This is a consequence of

$$\frac{1}{n}(m_1^{-1}(E_*(S_n)) - E_*(S_n)) = \frac{1}{n}E_*(S_n)\cdot\left(\frac{m_1^{-1}(E_*(S_n))}{E_*(S_n)} - 1\right) \to 0$$

(compare 5.7).

7.5.2. If in the situation of the preceding theorem $\eta := \lim_{n\to\infty} \frac{1}{n} E_*(S_n)$ exists, then

$$\lim_{n\to\infty} \frac{1}{n} S_n = \eta \quad P\text{-a.s.}$$

7.5.3. Under additional assumptions on the function A we can obtain $m_1(x) = x + O(1)$ for $x \to \infty$. Then the conclusion of Theorem 7.4 may be written as $\lim_{n\to\infty} \frac{1}{n}(S_n - E_*(S_n)) = 0$ P-a.s.

7.6. Corollary. *Let $(X_n : n \geqq 1)$ be an i.i.d. sequence of integrable random variables. Then*

$$\frac{1}{n} S_n \to E_*(X_1) \quad P\text{-a.s.}$$

Proof. Let $a > 0$ be arbitrary, consider the truncated variables $X_n^a := 1_{\{X_n < na\}} \cdot X_n$, and define $S_0^a := 0$, $S_n^a := S_{n-1}^a \overset{A_n}{+} X_n^a$, $s_n^a := E_*(S_n^a)$ for $n \geqq 1$, using the same A_n's as in the definition of S_n. By Lemma 6.2 we obtain

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n^2} V_*(X_n^a) &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=0}^{\infty} E(1_{\{aj \leqq X_n < a(j+1)\}} v(X_n^a, E_*(X_n^a))) \\
&\leqq \sum_{n=1}^{\infty} \frac{1}{n^2} \left\{ \sum_{j=0}^{n-1} P\{aj \leqq X_n < a(j+1)\} (E_*(X_n)^2 + a^2(j+1)^2) \right. \\
&\qquad \left. + \sum_{j=n}^{\infty} P\{aj \leqq X_n < a(j+1)\} \cdot E_*(X_n)^2 \right\} \\
&\leqq \sum_{n=1}^{\infty} \frac{1}{n^2} \left\{ \sum_{j=0}^{n-1} P\{aj \leqq X_1 < a(j+1)\} \cdot a^2(j+1)^2 \right. \\
&\qquad \left. + \sum_{j=0}^{\infty} P\{aj \leqq X_1 < a(j+1)\} E_*(X_1)^2 \right\} \\
&= \sum_{j=0}^{\infty} P\{aj \leqq X_1 < a(j+1)\} \\
&\qquad \times \left\{ a^2(j+1)^2 \cdot \sum_{n=j+1}^{\infty} \frac{1}{n^2} + E_*(X_1)^2 \cdot \sum_{n=1}^{\infty} \frac{1}{n^2} \right\} \\
&\leqq \sum_{j=0}^{\infty} 2a^2(j+1) P\{aj \leqq X_1 < a(j+1)\} + \frac{\pi^2}{6} \cdot E_*(X_1)^2 \\
&\leqq 2a^2 + 2aE(X_1) + \frac{\pi^2}{6} E_*(X_1)^2 < \infty.
\end{aligned}$$

On the other hand it follows from $\lim_{n\to\infty} E_*(X_n^a)=E_*(X_1)$ that

$$\lim_{n\to\infty} \frac{1}{n} s_n^a = E_*(X_1).$$

Hence 5.5.2 implies $\lim_{n\to\infty} \frac{1}{n} S_n^a = E_*(X_1)$ P-a.s. for every $a>0$.

The probability of $\Omega_a := \{X_n < na \text{ for all } n \geqq 1\}$ is

$$\begin{aligned} P(\Omega_a) &= 1 - P\{X_n \geqq na \text{ for some } n \geqq 1\} \\ &\geqq 1 - \sum_{n=1}^{\infty} P\{X_n \geqq na\} \\ &\geqq 1 - \frac{1}{a} E(X_1). \end{aligned}$$

Since in the definition of S_n^a the same A_n's were used as in the construction of S_n we obtain that $S_n = S_n^a$ on Ω_a and hence $\lim_{n\to\infty} \frac{1}{n} S_n = E_*(X_1)$ P-a.s. on Ω_a. Since $P(\Omega_a)\to 1$ as $a\to\infty$ the corollary is proved. □

7.7. Theorem. *Let* $(X_n : n \geqq 1)$ *be an i.i.d. sequence of random variables with* $E_*(X_1) = +\infty$. *Then* $\frac{1}{n} S_n \to \infty$ *P-a.s.*

Proof. Let a be an arbitrary positive number. We will show that $P\left\{\frac{1}{n} S_n < a \text{ i.o.}\right\} = 0$. This can be done by proving

$$\sum_{n\geqq 1} P\left\{\frac{1}{n} S_n < a\right\} < \infty$$

and using the Borel-Cantelli lemma.

Consider the functions $\lambda \mapsto E(\varphi_{i(\varrho-\lambda)}(X_1))$ and $\lambda \mapsto e^{-a\lambda}$. Since the derivations at 0 of these functions are $-\infty$ (by 5.9) and $-a$ there exists $\lambda \in]0, \varrho[$ such that

$$0 < e^{a\lambda} \cdot E(\varphi_{i(\varrho-\lambda)}(X_1)) < 1.$$

Therefore from $\varphi_{i(\varrho-\lambda)}(x) = \int_{-x}^{x} e^{-t\lambda} v_x(dt) \geqq e^{-\lambda x}$ it follows

$$\begin{aligned} P\left\{\frac{1}{n} S_n < a\right\} &= P\{\varphi_{i(\varrho-\lambda)}(S_n) > \varphi_{i(\varrho-\lambda)}(an)\} \\ &\leqq P\{\varphi_{i(\varrho-\lambda)}(S_n) > e^{-\lambda an}\} \\ &\leqq e^{\lambda an} \cdot E(\varphi_{i(\varrho-\lambda)}(S_n)) \\ &= (e^{\lambda a} \cdot E(\varphi_{i(\varrho-\lambda)}(X_1))^n \end{aligned}$$

and finally

$$\sum_{n\geqq 1} P\left\{\frac{1}{n} S_n < a\right\} \leqq \sum_{n\geqq 1} (e^{\lambda a} E(\varphi_{i(\varrho-\lambda)}(X_1)))^n < \infty. \quad \square$$

7.8. *Remark.* In 7.2 we have only considered the case of a random walk starting at the neutral element 0 of $(\mathbb{R}_+, *)$. However, 7.6 and 7.7 (and clearly 7.4) remain valid if the starting point is arbitrarily distributed. A short look at the proofs of 7.6 and 7.7 shows that it suffices to suppose that $(X_2, X_3, \ldots)$ are identically distributed (X_1 even does not need to be integrable): $X_1 = S_1$ can then be considered as the starting point of the random walk $(S_n : n \geqq 1)$.

7.9. *Remark.* Let $(\mathbb{R}_+, *)$ be the Sturm-Liouville hypergroup with $A(x) = (\cosh x)^2$ (see [28, Example 2.5c]). This is not a Chébli-Trimèche hypergroup in the sense of 2.2 since $A(0) \neq 0$. However, the assertions of 7.4, 7.5.3, 7.6, and 7.7 remain valid. In this case $\varrho = 1$, $m_1(x) = x \tanh x$, $m_2(x) = x^2$, and $\varphi_\lambda(x) = \dfrac{\cos \lambda x}{\cosh x}$ $(x \geqq 0,\ \lambda \in \mathbb{C})$ and it is easily checked that the facts used in the proofs of 7.4, 7.6, and 7.7 also hold in this situation.

8. Laws of Large Numbers in the Case of Exponential Boundedness

8.1. In this paragraph we suppose that $\varrho = 0$. This implies $E_*(X) = 0$ for every random variable and therefore we expect the law of large numbers to be of a particularly simple form. For example if (in the terminology of 6.11) the variances $V_*(X_j) = E(m_2(X_j))$ are bounded by some constant $b > 0$ and the variables X_j are pairwise $*$-uncorrelated we obtain for every $\varepsilon > 0$

$$\begin{aligned} P\left\{\frac{1}{n} S_n \geqq \varepsilon\right\} &= P\{m_2(S_n) \geqq m_2(n\varepsilon)\} \\ &\leqq \frac{V_*(S_n)}{m_2(n\varepsilon)} \\ &\leqq \frac{nb}{m_2(n\varepsilon)} \to 0 \end{aligned}$$

by 6.4 and hence

$$\frac{1}{n} S_n \to 0 = E_*(X_j) \quad \text{in probability}.$$

8.2. However, the proof of a *strong* law becomes more difficult and requires some restrictions concerning the function m_2. For the rest of this paragraph we have to suppose that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $m_2(\varepsilon x) \geqq \delta m_2(x)$ for every $x \geqq 0$.

8.3. *Examples*

8.3.1. If $\left\{x \dfrac{A'(x)}{A(x)} : x > 0\right\}$ is bounded, then 8.2 holds. This follows from 6.5. This criterium is useful if $\dfrac{A'}{A}$ decreases fast. The opposite case is considered in the following example.

8.3.2. Suppose that there is a $c > 0$ with $\dfrac{A'(2x)}{A(2x)} \geqq c \cdot \dfrac{A'(x)}{A(x)}$ for $x > 0$. Then 8.2 holds.

Proof. We consider the function $\varphi:\mathbb{R}_+\to\mathbb{R}$ defined by $\varphi(x):=4m_2\left(\frac{x}{2}\right)-2cm_2(x)$ for $x\geqq 0$. From the convexity of m_2 (6.4) and 4.2 we obtain $m_2'\left(\frac{x}{2}\right)\cdot\frac{A'(x/2)}{A(x/2)}\leqq 2$ and hence

$$\left[2\frac{A'(x)}{A(x)}-\frac{A'(x/2)}{A(x/2)}\right]m_2'\left(\frac{x}{2}\right)\geqq(2c-1)\frac{A'(x/2)}{A(x/2)}m_2'\left(\frac{x}{2}\right)\geqq 4c-2.$$

An easy calculation yields

$$L\varphi(x)=4c-2+\left[\frac{A'(x/2)}{A(x/2)}-2\frac{A'(x)}{A(x)}\right]m_2'\left(\frac{x}{2}\right)\leqq 0.$$

Therefore the assumption $\varphi'(x_0)<0$ leads to $\varphi''(x_0)>0$ and hence $\varphi'(0)<0$ which is a contradiction to $\varphi'(0)=2m_2'(0)-2cm_2'(0)=0$. This implies $m_2\left(\frac{x}{2}\right)>\frac{c}{2}m_2(x)$ for every $x\geqq 0$.

Now let $\varepsilon>0$. Then there is an $n\in\mathbb{N}$ with $2^{-n}<\varepsilon$ and we obtain

$$m_2(\varepsilon x)\geqq m_2(2^{-n}x)\geqq\left(\frac{c}{2}\right)^n m_2(x)$$

for every $x\geqq 0$. □

8.4. Theorem. *Suppose that* 8.2 *holds. Let* $(X_n:n\geqq 1)$ *be a series of independent random variables such that*

$$\sum_{n\geqq 1}\frac{1}{m_2(n)}V_*(X_n)<\infty.$$

Then $\frac{1}{n}S_n\to 0$ *P-almost surely.*

Proof. Since it follows from the assumption and 6.10 that

$$\sum_{n=1}^{\infty}\frac{1}{m_2(n)}E(m_2(S_n)-m_2(S_{n-1}))<\infty$$

we may apply Chow's law of large numbers [6] in order to obtain

$$\lim_{n\to\infty}\frac{1}{m_2(n)}m_2(S_n)=0\quad P\text{-a.s.}$$

But if for some $\varepsilon>0$ $S_n>\varepsilon n$ happens infinitely often with positive probability and if $\delta(\varepsilon)$ is chosen according to 8.2 this would imply $\frac{m_2(S_n)}{m_2(n)}\geqq\frac{m_2(\varepsilon n)}{m_2(n)}\geqq\delta(\varepsilon)$ infinitely often with positive probability. Hence the assertion of the theorem. □

8.5. *Remark.* Even if the sequence $(X_n:n\geqq 1)$ is i.i.d. with $V_*(X_n)<\infty$ the condition of the preceding theorem not necessarily holds since $\sum_{n\geqq 1}\frac{1}{m_2(n)}$ may be infinite. Since it follows from 6.4 that $\frac{A'}{A}m_2'\leqq 2$ and hence $m_2(x)\leqq 2x\frac{A(x)}{A'(x)}$ this happens for example if $\frac{A'(x)}{A(x)}\sim\frac{1}{\ln x}$ for $x\to\infty$.

8.6. Corollary. *Suppose that* $\left\{x \cdot \frac{A'(x)}{A(x)} : x>0\right\}$ *is bounded. Then for every i.i.d. sequence* $(X_n : n \geqq 1)$ *of integrable* $\mathbb{R}_+$*-valued random variables* $\frac{1}{n} S_n \to 0$ *P-almost surely.*

Proof. Let $a>0$. As in the proof of 7.6 we consider the truncated variables $X_n^a := 1_{\{X_n < an\}} \cdot X_n$ and define $S_0^a := 0$, $S_n^a := S_{n-1}^a \overset{A_n}{+} X_n^a$ for $n \geqq 1$. Let γ be defined as in Lemma 6.5. Since X_1 is integrable,

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{m_2(n)} V_*(X_n^a) &\leqq \frac{1}{\gamma} \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=0}^{\infty} E(1_{\{aj \leqq X_n^a < a(j+1)\}} m_2(X_n^a)) \\
&\leqq \frac{1}{\gamma} \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{j=0}^{n-1} P\{aj \leqq X_1 < a(j+1)\} a^2(j+1)^2 \\
&= \frac{1}{\gamma} \sum_{j=0}^{\infty} P\{aj \leqq X_1 < a(j+1)\} \cdot a^2(j+1)^2 \sum_{n=j+1}^{\infty} \frac{1}{n^2} \\
&\leqq \frac{2}{\gamma} \sum_{j=0}^{\infty} P\{aj \leqq X_1 < a(j+1)\} \cdot a^2(j+1) \\
&\leqq \frac{2}{\gamma}(aE(X_1) + a^2) < \infty
\end{aligned}$$

we obtain from 8.3.1 and the preceding theorem that $\frac{1}{n} S_n^a \to 0$ P-a.s. for every $a>0$. The rest of the proof is identical with 7.6. □

8.7. *Remark.* By the same argument as in 7.8. we see that 8.6 is valid even if the starting point of the random walk $(S_n : n \in \mathbb{N})$ is not 0 but arbitrary.

8.8. *Remark.* Suppose that $\left\{x \cdot \frac{A'(x)}{A(x)} : x>0\right\}$ is bounded and let $0<\beta<2$. Then for every i.i.d. sequence $(X_n : n \geqq 1)$ of $\mathbb{R}_+$-valued random variables such that $E(X_1^\beta)$ is finite, $\frac{1}{n^{1/\beta}} S_n \to 0$ P-almost surely. A similar result has been proved by Gallardo and Ries [15].

Proof. It is a straightforward generalization of Theorem 8.4 that for every independent sequence $(Y_n : n \geqq 1)$ such that $\sum_{n=1}^{\infty} \frac{1}{n^{2/\beta}} V_*(Y_n) < \infty$ we obtain $\frac{1}{n^{1/\beta}} {}^{A}\!\sum_{j=1}^{n} Y_j \to 0$ P-almost surely. If we choose Y_n to be the truncated variable $X_n 1_{\{X_n \leqq an^{1/\beta}\}}$ it follows as in the proof of 8.6 that $\frac{1}{n^{1/\beta}} S_n$ and $\frac{1}{n^{1/\beta}} {}^{A}\!\sum_{j=1}^{n} Y_j$ tend to the same limit almost surely. □

References

1. Achour, A., Trimèche, K.: Opérateurs de translation généralisée associées à un opérateur singulier sur un interval borné. C.R. Acad. Sci. Paris Sér. A **288**, 399–402 (1979)
2. Bingham, N.H.: Random walks on spheres. Z. Wahrscheinlichkeitstheorie Verw. Geb. **22**, 169–192 (1972)

3. Bloom, W.R., Heyer, H.: The Fourier transform for probability measures on hypergroups. Rend. Mat. Appl., VII Ser. **2**, 315–334 (1982)
4. Bondar, T.V.: Borel cross sections and maximal invariants. Ann. Stat. **4**, 866–877 (1976)
5. Chébli, H.: Positivité des opérateurs de «translation généralisée» associées à un opérateur de Sturm-Liouville et quelques applications à l'analyse harmonique. Thèse, Université Louis Pasteur, Strasbourg I (1974)
6. Chow, Y.: A martingale inequality and the law of large numbers. Proc. AMS **11**, 107–111 (1960)
7. Erdélyi, A., et al.: Tables of integral transforms, Vol. II. New York Toronto London: McGraw-Hill 1954
8. Eymard, P., Roynette, B.: Marches aléatoires sur le dual de SU(2). In: Analyse harmonique sur les groupes de Lie (Lecture Notes Mathematics, Vol. 497, pp. 108–152). Berlin Heidelberg New York: Springer 1975
9. Faraut, J.: Analyse harmonique sur les paires de Guelfand et les espaces hyperboliques. Strasbourg (1975)
10. Feller, W.: An introduction to probability theory and its applications, Vol. II, 2nd edition. New York: Wiley 1970
11. Finckh, U.: Beiträge zur Wahrscheinlichkeitstheorie auf einer Kingman-Struktur. Dissertation, Tübingen (1986)
12. Flensted-Jensen, M., Koornwinder, T.: The convolution structure for Jacobi function expansions. Ark. Mat. **11**, 245–262 (1973)
13. Gallardo, L.: Exemples d'hypergroupes transientes. In: Probability measures on groups VIII. H. Heyer (ed.). Lecture Notes Mathematics Vol. 1210. Berlin Heidelberg New York: Springer 1986
14. Gallardo, L.: Comportement asymptotique des marches aléatoires associées aux polynômes de Gegenbauer. Adv. Appl. Prob. **16**, 293–323 (1984)
15. Gallardo, L., Ries, V.: La loi des grands nombres pour les marches aléatoires sur le dual de SU(2). Stud. Math. LXVI, 93–105 (1979)
16. Guivarc'h, Y.: Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire. Astérisque **74**, 47–98 (1980)
17. Haldane, J.B.S.: The addition of random vectors. Indian J. Stat. **22**, 213–220 (1960)
18. Heyer, H.: Moments of probability measures on a group. Int. J. Math. Sci. **4**, 1–37 (1981)
19. Heyer, H.: Probability theory on hypergroups: A survey. In: Probability measures on groups VII, H. Heyer (ed.). (Lecture Notes Mathematics Vol. 1064). Berlin Heidelberg New York: Springer 1984
20. Jewett, R.I.: Spaces with an abstract convolution of measures. Adv. Math. **18**, 1–101 (1975)
21. Karpelevich, F.I., Tutubalin, V.N., Shur, M.G.: Limit theorems for the compositions of distributions in the Lobachevsky plane and space. Theory Probab. Appl. **4**, 399–402 (1959)
22. Kingman, J.F.C.: Random walks with spherical symmetry. Acta Math. **109**, 11–53 (1963)
23. Spector, R.: Aperçu de la théorie des hypergroupes. In: Analyse harmonique sur les groupes de Lie (Lecture Notes Mathematics Vol. 497, pp. 643–673). Berlin Heidelberg New York: Springer 1975
24. Trimèche, K.: Probabilités indéfiniment divisibles et théorème de la limite centrale pour une convolution généralisée sur la demi-droite. C.R. Acad. Sci. Paris Sér. A **286**, 63–66 (1978)
25. Trimèche, K.: Transformation intégrale de Weyl et théorème de Paley-Wiener associé à un opérateur différentiel singulier sur $(0, \infty)$. J. Math. Pures Appl. **60**, 51–98 (1981)
26. Tutubalin, V.N.: On the limit behaviour of compositions of measures in the plane and space of Lobachevski. Theory Probab. Appl. **7**, 189–196 (1962)
27. Zeuner, H.: On hyperbolic hypergroups. In: Probability measures on groups VIII, H. Heyer (ed.). (Lecture Notes Mathematics Vol. 1210). Berlin Heidelberg New York: Springer 1986
28. Zeuner, H.: One-dimensional hypergroups. To appear in: Adv. Math. (1989)
29. Zeuner, H.: The central limit theorem for Chébli-Trimèche hypergroups. J. Theor. Probab. To appear

Received July 25, 1988

Math. Ann. 283, 679–687 (1989)

Mathematische Annalen

A $W^{1,p}$-Estimate for Solutions to Mixed Boundary Value Problems for Second Order Elliptic Differential Equations

Konrad Gröger

Karl-Weierstraß-Institut für Mathematik der Akademie der Wissenschaften der DDR, Mohrenstrasse 39, DDR-1086 Berlin, German Democratic Republic

1. Introduction

In this paper we shall prove that, under rather weak hypotheses, any solution to a mixed boundary value problem for a second order elliptic differential equation is in the Sobolev space $W^{1,p}$ for some $p>2$. Our starting point is the following result due to Meyers [6]:

If for some $q>2$ it holds the implication

$$u\in W_0^{1,2}(G), \qquad \Delta u\in W^{-1,q}(G)\Rightarrow u\in W_0^{1,q}(G), \tag{1.1}$$

then for every positive definite matrix (a_{ij}) of bounded measurable functions on G there exists a $p>2$ such that

$$u\in W_0^{1,2}(G), \qquad \sum_{i,j=1}^{N} D_i(a_{ij}D_ju)\in W^{-1,p}(G)\Rightarrow u\in W_0^{1,p}(G). \tag{1.2}$$

Here G is a bounded domain in $\mathbb{R}^N$, D_i denotes the derivative with respect to the coordinate x_i of $x=(x_1,\dots,x_N)\in\mathbb{R}^N$, and $W_0^{1,p}(G)$, $W^{-1,p}(G)$ are the usual Sobolev spaces.

The hypothesis (1.1) is satisfied for every $q\in]2,\infty[$ provided that G is a bounded domain of class C^1 (see Simader [8, Theorem 4.6]). We shall show that an analogue of Meyers' result holds if the homogeneous Dirichlet boundary condition [which is included in the requirement $u\in W_0^{1,2}(G)$] is replaced by a mixed boundary condition. Moreover, we are going to prove that results of the type (1.2) can be obtained not only for differential operators of the form

$$u\mapsto \sum_{i,j=1}^{N} D_i(a_{ij}D_ju)$$

but also for (generally nonlinear) operators of the form

$$u\mapsto \sum_{i=1}^{N} D_ib_i(\cdot,u,Du)+b_0(\cdot,u,Du);$$

here and later on Du denotes the gradient of u, and the dot indicates the dependence on the spatial variable. For precise assumptions with respect to $b=(b_0, \ldots, b_N)$ see Sect. 4. It is well known that in the case of mixed boundary conditions in general one cannot expect an analogue of (1.1) to hold for *every* $q>2$. We are able to prove, however, that under weak hypotheses there exists *some* $q>2$ such that an analogue of (1.1) holds also in that case. For this proof it is essential that we deal with second order differential operators only. ($W^{k,p}$-estimates for solutions to elliptic equations of order $2k$ in case of smooth boundary conditions were presented, for example by Nečas [7] and Krbec [5].)

Let us mention that local L^p-estimates for gradients of solutions to nonlinear problems were obtained also by other authors (see Giaquinta and Giusti [3] and the papers quoted therein). However, to our knowledge these authors made no attempt to prove global estimates for solutions to mixed boundary value problems.

The paper is organized as follows. In Sect. 2 we shall introduce the notation and some notions needed later on. In particular, we shall define a class of subsets of $\mathbb{R}^N$ called *regular*. This class turns out to be quite useful for the formulation of results on general boundary value problems. For any regular $G \subset \mathbb{R}^N$ we shall introduce spaces $W_0^{1,p}(G)$ and $W^{-1,p}(G)$ coinciding with the usual Sobolev spaces provided that G is open. We shall denote by R_q the class of all regular subsets of $\mathbb{R}^N$ for which an analogue of (1.1) holds. Section 3 contains some preliminary results. In particular, it will be shown that the validity of the relation $G \in R_q$ depends on the local properties of G only. In Sect. 4 we shall prove that the relation $G \in R_q$ implies a regularity result of type (1.2). We shall take advantage of an iteration procedure which has widely been used by Koševev in order to prove other regularity results (see [4]). In Sect. 5 we shall show that if $G \in R_q$, $q>2$, and if $\tilde{G}$ is obtained from G by a Lipschitzian transformation, then there exists a $p>2$ such that $\tilde{G} \in R_p$. From this fact it will follow that for every regular subset G of $\mathbb{R}^N$ there exists a $q>2$ such that $G \in R_q$.

2. Notations and Notation

If G is any subset of the Euclidean N-space $\mathbb{R}^N$, then we denote by $\mathring{G}$, ∂G and $\bar{G}$ the interior, the boundary, and the closure of G, respectively. We write $|x|$ for the Euclidean norm of $x \in \mathbb{R}^N$.

Assume that u is a solution to a second order elliptic differential equation in a domain $\Omega \subset \mathbb{R}^N$. Let u satisfy a Dirichlet condition on $\tilde{\Gamma} \subset \partial\Omega$ and natural boundary conditions on $\Gamma := \partial\Omega \setminus \tilde{\Gamma}$. If one wants to prove a regularity result for u, then one has to impose an appropriate "regularity condition" on Γ and $\tilde{\Gamma}$. We are going to show that it is useful to formulate all conditions and results in terms of $G := \Omega \cup \Gamma$.

Definition 1. Let G and $\tilde{G}$ be subsets of $\mathbb{R}^N$. A bijection $\Phi: G \to \tilde{G}$ will be called a *Lipschitz-transformation*, if Φ and Φ^{-1} are Lipschitzian with respect to the standard metrics of G and $\tilde{G}$.

Definition 2. We shall call $G \subset \mathbb{R}^N$ *regular*, if G is bounded and if for every $y \in \partial G$ there exist subsets U and $\tilde{U}$ of $\mathbb{R}^N$ and a Lipschitz-transformation $\Phi: U \to \tilde{U}$ such that U is an open neighbourhood of y in $\mathbb{R}^N$ and that $\Phi(U \cap G)$ is one of the following sets:

$$E_1 := \{x \in \mathbb{R}^N : |x| < 1,\ x_N < 0\},$$
$$E_2 := \{x \in \mathbb{R}^N : |x| < 1,\ x_N \leqq 0\},$$
$$E_3 := \{x \in E_2 : x_N < 0 \text{ or } x_1 > 0\}.$$

Remark 1. Apart from boundedness regularity of G means, roughly speaking, that the parts $\Gamma := G \setminus \mathring{G}$ and $\tilde{\Gamma} := \bar{G} \setminus G$ of the boundary ∂G are separated by a Lipschitzian hypersurface of ∂G.

Remark 2. In the following we shall assume always – even if this is not mentioned explicitly – that G is a regular subset of $\mathbb{R}^N$. Then G is of finite Lebesgue measure. The boundary $\partial G = \partial \mathring{G}$ is of N-dimensional Lebesgue measure 0. Therefore we are allowed to identify the spaces $L^p(G)$ and $L^p(\mathring{G})$.

Definition 3. For $1 \leqq p \leqq \infty$ we denote by $W_0^{1,p}(G)$ the closure of the set

$$\{u | \mathring{G} : u \in C_0^\infty(\mathbb{R}^N),\ \operatorname{supp} u \cap (\bar{G} \setminus G) = \emptyset\}$$

in the Sobolev space $W^{1,p}(\mathring{G})$, equipped with the standard norm of that space. The space dual to $W_0^{1,p'}(G)$ will be denoted by $W^{-1,p}(G)$; here (and later on) p' denotes the exponent conjugate to p defined by $\frac{1}{p} + \frac{1}{p'} = 1$. For the norms in $W_0^{1,p}(G)$ and $W^{-1,p}(G)$ we write $\|\cdot\|_{1,p}$ and $\|\cdot\|_{-1,p}$, respectively. If necessary we indicate the dependence of these norms on G by an additional index. By J_G we denote the duality map of the Hilbert space $W_0^{1,2}(G)$.

Remark 3. If G is open, then our definition of $W_0^{1,p}(G)$ coincides with the usual one. If G is closed then $W_0^{1,p}(G) = W^{1,p}(\mathring{G})$.

Remark 4. If this should not lead to misunderstandings we write $W_0^{1,p}$, $W^{-1,p}$, and J instead of $W_0^{1,p}(G)$, $W^{-1,p}(G)$, and J_G, respectively.

Remark 5. Let $1 \leqq p \leqq q \leqq \infty$. Then $W_0^{1,q} \hookrightarrow W_0^{1,p}$, and $W_0^{1,q}$ is dense in $W_0^{1,p}$ (the sign $\hookrightarrow$ means that the imbedding is continuous). Therefore we have $W^{-1,p'} \hookrightarrow W^{-1,q'}$.

Remark 6. From the formula

$$\langle Ju, v \rangle = \int_G (uv + Du \cdot Dv)\,dx, \quad \text{for} \quad u, v \in W_0^{1,2},$$

it follows easily that J maps $W_0^{1,p}$, $p > 2$, into $W^{-1,p}$ and that $J | W_0^{1,p}$ is continuous as a map from $W_0^{1,p}$ into $W^{-1,p}$. Throughout this paper, for $p \geqq 2$, we shall use M_p as an abbreviation for

$$\sup\{\|u\|_{1,p} : u \in W_0^{1,p},\ \|Ju\|_{-1,p} \leqq 1\}.$$

Note that $M_2 = 1$.

Definition 4. For $2 \leqq q < \infty$ we denote by R_q the class of all regular subsets G of $\mathbb{R}^N$ for which J_G maps $W_0^{1,q}(G)$ *onto* $W^{-1,q}(G)$.

Remark 7. If G is a bounded domain of class C^1 then $G \in \bigcap_{q \geqq 2} R_q$. This follows easily from a result stated by Simader (see [8, Theorem 4.6]). As mentioned in the introduction we shall show in Sect. 5 that for every regular G there exists a $q>2$ such that $G \in R_q$.

Remark 8. In view of the Open Mapping Theorem the relation $G \in R_q$ implies that $M_q < \infty$.

Let us introduce some further notation. By Y_p, $1<p<\infty$, we denote the space $L^p(G; \mathbb{R}^{N+1})$, equipped with its standard norm. The space dual to Y_p will be identified with $Y_{p'}$. Moreover, let $L \in \mathscr{L}(W_0^{1,2}; Y_2)$ be defined by $Lu := (u, Du)$, $u \in W_0^{1,2}$. Obviously, $J = L^* L$. It is easy to check that L maps $W_0^{1,p}$, $p>2$, continuously into Y_p and that L^* maps Y_p, $p>2$, continuously into $W^{-1,p}$.

3. Preliminary Results

Lemma 1. *Let $G \in R_q$ for some $q>2$. Then $G \in R_p$ for $2 \leqq p \leqq q$ and $M_p \leqq M_q^\theta$ if $\frac{1}{p} = \frac{1-\theta}{2} + \frac{\theta}{q}$.*

Proof. 1. Let $P \in \mathscr{L}(Y_2; Y_2)$ be defined by $P := LJ^{-1}L^*$. Since $G \in R_q$ the operator P maps Y_q continuously into itself. It is easy to check that $\|P\|_{\mathscr{L}(Y_2;Y_2)} = 1$ and that $\|P\|_{\mathscr{L}(Y_q;Y_q)} \leqq M_q$. In view of the well-known Riesz-Thorin Interpolation Theorem (see, e.g., Bergh and Löfström [1]) this implies that P maps Y_p continuously into itself and that $\|P\|_{\mathscr{L}(Y_p;Y_p)} \leqq M_2^{1-\theta} M_q^\theta$ provided that $\frac{1}{p} = \frac{1-\theta}{2} + \frac{\theta}{q}$, $\theta \in [0,1]$.

2. Let $2 \leqq p \leqq q$, and let $f \in W^{-1,p}$ be fixed. We define

$$\forall v \in W_0^{1,2} : z(Lv) := \langle f, v \rangle .$$

Because v is uniquely determined by Lv, this definition makes sense. Since $z(Lv) \leqq \|f\|_{-1,p} \|v\|_{1,p'}$, z is a continuous linear functional on a subspace of $Y_{p'}$. By the Hahn-Banach Theorem z can be extended to a functional on $Y_{p'}$ (again denoted by z) with the same norm. Thus, $z \in Y_p$ and $\|z\|_{Y_p} = \|f\|_{-1,p}$. Moreover, $L^* z = f$, because

$$\forall v \in W_0^{1,2} : \langle L^* z, v \rangle = \langle z, Lv \rangle = \langle f, v \rangle .$$

For $u := J^{-1} f$ we have $Lu = LJ^{-1}L^* z = Pz \in Y_p$ and $\|Lu\|_{Y_p} \leqq M_q^\theta \|z\|_{Y_p} = M_q^\theta \|f\|_{-1,p}$. Consequently, $u \in W^{1,p}(\mathring{G})$ and $\|u\|_{1,p} \leqq M_q^\theta \|f\|_{-1,p}$. To show that u is in $W_0^{1,p}$ we proceed as follows. We choose a sequence (f_n) from $W^{-1,q}$ converging in $W^{-1,p}$ to f. (Note that $W_0^{1,p'} \hookrightarrow W_0^{1,q'}$ with dense imbedding and that $W_0^{1,p'}$ is reflexive as a subspace of a reflexive space. This implies that $W^{-1,q}$ is dense in $W^{-1,p}$.) Let $u_n := J^{-1} f_n$. Then $u_n \in W_0^{1,q} \subset W_0^{1,p}$ and $\|u_n - u_m\|_{1,p} \leqq M_q^\theta \|f_n - f_m\|_{-1,p}$. Hence (u_n) converges in $W_0^{1,p}$. Its limit must be u since $J^{-1} : W^{-1,2} \to W_0^{1,2}$ is continuous. Thus, J^{-1} maps $W^{-1,p}$ continuously into $W_0^{1,p}$ and

$$M_p = \sup\{\|u\|_{1,p} : u \in W_0^{1,p},\ \|Ju\|_{-1,p} \leqq 1\} \leqq M_q^\theta .$$

Lemma 2. *Let $\{U_0, \ldots, U_r\}$ be an open covering of G, and let $q \geqq 2$. If $U_i \cap G \in R_q$, $i=0, \ldots, r$, then $G \in R_q$.*

Proof. Let $I_0 := \{\frac{1}{2}\}$, $I_k := \left[\frac{1}{2} - \frac{k}{N}, \frac{1}{2} - \frac{k-1}{N}\right[$, $k=1, \ldots, l$, where l denotes the largest integer such that $\frac{1}{2} > \frac{l}{N}$, and let $I_{l+1} := \left]0, \frac{1}{2} - \frac{l}{N}\right[$. We shall prove the assertion for $\frac{1}{q} \in I_k$ by induction with respect to k. For $k=0$, i.e. $q=2$, the assertion is trivial. Now let the assertion be proved for all q such that $\frac{1}{q} \in I_k$ for some $k \leqq l$. We want to prove the assertion under the hypothesis $\frac{1}{q} \in I_{k+1}$. Then we can choose p such that $\frac{1}{p} \in I_k$ and $\frac{1}{p} \leqq \frac{1}{q} + \frac{1}{N}$. In view of Lemma 1 we have $U_i \cap G \in R_p$, $i=0, \ldots, r$. By our induction hypothesis this implies that $G \in R_p$. Let $f \in W^{-1,q}(G)$ and $u := J^{-1} f$. Because of $G \in R_p$ we obtain $u \in W_0^{1,p}(G) \subset L^q(G)$ (the inclusion follows from Sobolev's Imbedding Theorem). We choose a partition of unity $\{\varphi_0, \ldots, \varphi_r\}$ subordinate to the covering $\{U_0, \ldots, U_r\}$. We want to show that each of the functions $\varphi_i u$, $i=0, \ldots, r$, is an element of $W_0^{1,q}(G)$. We have $\varphi_i u \in W_0^{1,p}(U_i \cap G)$ and (in view of the choice of p)

$$\begin{aligned}
\forall v \in W_0^{1,2}(U_i \cap G):& \int_{U_i \cap G} (\varphi_i u v + D(\varphi_i u) \cdot Dv) dx \\
&= \int_{U_i \cap G} (u \varphi_i v + Du \cdot D(\varphi_i v) + (u Dv - v Du) \cdot D\varphi_i) dx \\
&= \langle f, \varphi_i v \rangle + \int_{U_i \cap G} (u Dv - v Du) \cdot D\varphi_i dx \\
&\leqq c(\|f\|_{-1,q} + \|u\|_{1,p}) \|v\|_{1,q', U_i \cap G}.
\end{aligned}$$

In the dual pairing $\langle f, \varphi_i v \rangle$ the function $\varphi_i v$ is to be interpreted in the usual way as a function defined on G vanishing on $G \setminus U_i$. The estimate shows that $J_{U_i \cap G}(\varphi_i u) \in W^{-1,q}(U_i \cap G)$. Consequently, $\varphi_i u \in W_0^{1,q}(U_i \cap G)$ and $u = \varphi_0 u + \ldots + \varphi_r u \in W_0^{1,q}(G)$. This result completes the proof of Lemma 2.

Lemma 3. *For every $q \geqq 2$ the sets E_1 and E_2 are in the class R_q.*

Proof. Let $E := \{x \in \mathbb{R}^N : |x| < 1\}$, and let $q \geqq 2$ and $i \in \{1, 2\}$ be fixed. For any $u \in W_0^{1,2}(E_i)$ we define

$$(Su)(x) := \begin{cases} u(x) & \text{for} \quad x \in E_i, \\ (-1)^i u(x', -x_N) & \text{for} \quad x = (x', x_N) \in E \setminus E_i. \end{cases}$$

Clearly, if $u \in W_0^{1,q}(E_i)$ then $Su \in W_0^{1,q}(E)$. We fix $f \in W^{-1,q}(E_i)$ and set $u := J_{E_i}^{-1} f$. Then $u \in W_0^{1,2}(E_i)$ and

$$\begin{aligned}
\forall v \in W_0^{1,2}(E): \langle J_E Su, v \rangle &= \int_E ((Su)v + DSu \cdot Dv) dx \\
&= \int_{E_i} (uw + Du \cdot Dw) dx = \langle J_{E_i} u, w \rangle = \langle f, w \rangle,
\end{aligned}$$

where

$$w(x) := v(x) + (-1)^i v(x', -x_N) \quad \text{for} \quad x = (x', x_N) \in E_i.$$

[Note that $w \in W_0^{1,2}(E_i)$.] Since $\|w\|_{1,q',E_i} \leqq 2\|v\|_{1,q',E}$ we have $\langle f, w\rangle = \langle g, v\rangle$ for some $g \in W^{-1,q}(E)$. Because $E \in R_q$ this implies that $Su \in W_0^{1,q}(E)$. In view of the definition of $W_0^{1,q}(E_i)$ we obtain $u = Su|E_i \in W_0^{1,q}(E_i)$. This completes the proof.

Remark 9. In the same manner one can prove that $E_4 := \{x \in E_2 : x_1 > 0\}$ is in R_q for every $q \geqq 2$.

4. Boundary Value Problems

Let b be a function satisfying the following hypotheses:

$$\left.\begin{gathered} b: G \times \mathbb{R}^{N+1} \to \mathbb{R}^{N+1},\ b(\cdot, 0) \in L^q(G; \mathbb{R}^{N+1}) \text{ for some } q > 2, \\ b(\cdot, \xi) \text{ is measurable for every } \xi \in \mathbb{R}^{N+1}; \\ (b(x,\xi) - b(x,\eta)) \cdot (\xi - \eta) \geqq m|\xi - \eta|^2,\ m > 0, \\ |b(x,\xi) - b(x,\eta)| \leqq M|\xi - \eta|,\ M < \infty, \text{ for } x \in G,\ \xi,\ \eta \in \mathbb{R}^{N+1}. \end{gathered}\right\} \tag{4.1}$$

Of course, here the dot indicates the Euclidean scalar product in $\mathbb{R}^{N+1}$, and $|\xi|$ is the Euclidean norm of $\xi \in \mathbb{R}^{N+1}$. We define $A : W_0^{1,2} \to W^{-1,2}$ setting

$$\forall v \in W_0^{1,2} : \langle Au, v\rangle := \int_G b(\cdot, Lu) \cdot Lv\, dx, \tag{4.2}$$

where L is the operator introduced at the end of Sect. 2.

Remark 10. The operator A is strongly monotone and Lipschitzian (cf. [2, Chap. III]).

Remark 11. The hypotheses (4.1) are satisfied in particular, if

$$b_j(\cdot, \xi) = \sum_{i=1}^N a_{ij} \xi_i, \qquad j = 1, \ldots, N,$$

$$b_0(\cdot, \xi) = a_0 \xi_0, \quad \text{for} \quad \xi = (\xi_0, \ldots, \xi_N) \in \mathbb{R}^{N+1},$$

provided that $a_{ij} \in L^\infty(G)$, $i, j = 1, \ldots, N$, and $a_0 \in L^\infty(G)$ are such that

$$\sum_{i,j=1}^N a_{ij}(x) \xi_i \xi_j \geqq m|\xi|^2,\ a_0(x) \geqq m,\ m > 0, \quad \text{for} \quad x \in G,\ \xi \in \mathbb{R}^N.$$

In that case (4.2) reads as follows:

$$\forall v \in W_0^{1,2} : \langle Au, v\rangle = \int_G \left(\sum_{i,j=1}^N a_{ij} D_i u D_j v + a_0 uv \right) dx.$$

Remark 12. For $p \in [2, q]$ the operator A maps $W_0^{1,p}$ continuously into $W^{-1,p}$. Indeed, if u, $\bar{u} \in W_0^{1,p}$, then

$$\forall v \in W_0^{1,2} : \langle Au, v\rangle \leqq \int_G (M|Lu| + |b(\cdot, 0)|)\, |Lv|\, dx \leqq c\|v\|_{1,p'},$$

$$\langle Au - A\bar{u}, v\rangle \leqq \int_G M|L(u - \bar{u})|\, |Lv|\, dx \leqq M\|u - \bar{u}\|_{1,p} \|v\|_{1,p'}.$$

The next theorem deals with the question whether A maps $W_0^{1,p}$ *onto* $W^{-1,p}$.

Theorem 1. *Let $G \in R_q$. Suppose that* (4.1) *holds and that A is defined by* (4.2). *Then A maps $W_0^{1,p}$ onto $W^{-1,p}$ provided that $p \in [2, q]$ and $M_p k < 1$, where $k := (1 - m^2/M^2)^{1/2}$. If $p \in [2, q]$ and $M_p k < 1$, then*

$$\|A^{-1}f - A^{-1}g\|_{1,p} \leqq mM^{-2} M_p (1 - M_p k)^{-1} \|f - g\|_{-1,p} \quad \text{for} \quad f, g \in W^{-1,p}.$$

For $p \in [2, q]$ the inequality $M_p k < 1$ is satisfied, if

$$\frac{1}{p} > \frac{1}{2} - \left(\frac{1}{2} - \frac{1}{q}\right) \frac{|\log k|}{\log M_q}.$$

Proof. Let $t := mM^{-2}$ and let $(By)(x) := y(x) - tb(x, y(x))$ for $y \in Y_2$. It is easy to check that (4.1) implies that B, restricted to Y_p, $p \in [2, q]$, is a Lipschitzian mapping from Y_p into itself, where k is a Lipschitz constant of this mapping (cf. [2, Chap. III, Lemma 3.1]). Let $f \in W^{-1,p}$, $p \in [2, q]$, and let

$$Q_f u := J^{-1}(L^* BLu + tf) = u - tJ^{-1}(Au - f), \quad u \in W_0^{1,p}.$$

Then Q_f is a Lipschitzian mapping from $W_0^{1,p}$ into itself, and $M_p k$ is a Lipschitz constant of Q_f. This follows from the fact that $G \in R_p$ (cf. Lemma 1) and from the properties of L and B. Thus, the requirement $M_p k < 1$ guarantees $Q_f : W_0^{1,p} \to W_0^{1,p}$ to be strictly contractive. By definition of Q_f the fixed point $u \in W_0^{1,p}$ of Q_f is a solution to $Au = f$. Hence A maps $W_0^{1,p}$ onto $W^{-1,p}$. Since $A : W_0^{1,2} \to W^{-1,2}$ is invertible, the fixed point u of Q_f is the unique solution to $Au = f$. If $f, g \in W^{-1,p}$ are given and u, v are the fixed points of Q_f, Q_g, respectively, then

$$\begin{aligned} \|u - v\|_{1,p} = \|Q_f u - Q_g v\|_{1,p} &\leqq M_p k \|u - v\|_{1,p} + \|Q_f v - Q_g v\|_{1,p} \\ &\leqq M_p k \|u - v\|_{1,p} + M_p t \|f - g\|_{-1,p}. \end{aligned}$$

Hence

$$\|u - v\|_{1,p} \leqq t M_p (1 - M_p k)^{-1} \|f - g\|_{-1,p}.$$

The last assertion of the theorem follows from $M_p \leqq M_q^\theta$, where θ is defined by $\frac{1}{p} = \frac{1-\theta}{2} + \frac{\theta}{q}$ (cf. Lemma 1).

Remark 13. Let the hypotheses of Theorem 1 be satisfied, and let $p \in [2, q]$ be such that $M_p k < 1$. Furthermore, let F be any mapping from $W_0^{1,2}$ into $W^{-1,p}$. Then from $Au = Fu$, $u \in W_0^{1,2}$, it follows that $u \in W_0^{1,p}$. This is an immediate consequence of Theorem 1. An example for F is given by

$$\langle Fu, v \rangle := \int_G d_0(\cdot, u) v \, dx + \int_\Gamma d_1(\cdot, u) v \, d\sigma \quad \text{for} \quad v \in W_0^{1,2},$$

where $\Gamma := G \setminus \mathring{G}$, and $d_0 : G \times \mathbb{R} \to \mathbb{R}$, $d_1 : \Gamma \times \mathbb{R} \to \mathbb{R}$ are Carathéodory functions satisfying appropriate growth conditions. (Γ is to be equipped with the standard surface measure.) In this case $Au = Fu$ means that

$$-\sum_{i=1}^N D_i b_i(\cdot, Lu) + b_0(\cdot, Lu) = d_0(\cdot, u) \quad \text{in} \quad \mathring{G},$$

$$\sum_{i=1}^N b_i(\cdot, Lu) \nu_i = d_1(\cdot, u) \quad \text{on} \quad \Gamma, \qquad u = 0 \quad \text{on} \quad \partial G \setminus \Gamma,$$

where $\nu=(\nu_1, \ldots, \nu_N)$ is the outer unit normal at a point of Γ. The term $\sum_{i=1}^{N} b_i(\cdot, Lu)\nu_i$ is defined as an element of the Sobolev space $W^{-1/2,2}(\Gamma)$. This example shows that Theorem 1 can be used to prove $W^{1,p}$-estimates for solutions to rather general boundary value problems for second order elliptic differential equations.

Remark 14. An analogue of Theorem 1 holds for *systems* of second order equations. This follows easily from the fact that, for any $n \in \mathbb{N}$, the duality map of $W_0^{1,2}(G; \mathbb{R}^n)$ is "diagonal".

Remark 15. If there exists a function $\varphi: G \times \mathbb{R}^{N+1} \to \mathbb{R}$ such that

$$b(x,\xi)\cdot\eta = \frac{d}{dt}\varphi(x, \xi+t\eta)\Big|_{t=0} \quad \text{for} \quad \xi, \eta \in \mathbb{R}^{N+1},\ x \in G,$$

then the number k in Theorem 1 may be replaced by the strictly smaller number $(M-m)/(M+m)$ (cf. [2, Chap. III, Lemma 4.14]).

5. Lipschitz-Transformations of Sets $G \in R_q$

Let Φ be a Lipschitz-transformation from $G \subset \mathbb{R}^N$ onto $\tilde{G} \subset \mathbb{R}^N$. If G is regular then $\tilde{G}$ is regular as well, but from $G \in R_q$ it does not necessarily follow that $\tilde{G} \in R_q$. It holds, however, the following

Theorem 2. *Suppose that $G \in R_q$ for some $q>2$ and that Φ is a Lipschitz-transformation from G onto $\tilde{G} \subset \mathbb{R}^N$. Then there exists a $p>2$ such that $\tilde{G} \in R_p$.*

Proof. Let $Tu := u \circ \Phi^{-1}$, for $u \in W_0^{1,2}(G)$. By means of the chain rule one can easily prove that T maps $W_0^{1,p}(G)$, $p \geqq 2$, continuously onto $W_0^{1,p}(\tilde{G})$. Let T^* be the adjoint operator of $T: W_0^{1,2}(G) \to W_0^{1,2}(\tilde{G})$. Standard calculations show that Theorem 1 is applicable to the operator $A := T^* J_{\tilde{G}} T: W_0^{1,2}(G) \to W^{-1,2}(G)$ (cf. Remark 11). Consequently, there exists a $p>2$ such that A maps $W_0^{1,p}(G)$ onto $W^{-1,p}(G)$. Since T^* maps $W^{-1,p}(\tilde{G})$ into $W^{-1,p}(G)$ (this follows from the properties of T), the operator $J_{\tilde{G}}^{-1} = TA^{-1}T^*$ maps $W^{-1,p}(\tilde{G})$ into $W_0^{1,p}(\tilde{G})$. This shows that $\tilde{G} \in R_p$.

Theorem 3. *If $G \subset \mathbb{R}^N$ is regular, then $G \in \bigcup_{q>2} R_q$.*

Proof. In view of Lemma 2 it suffices to find an open covering $\{U_0, \ldots, U_r\}$ of G such that $U_i \cap G \in \bigcup_{q>2} R_q$, $i=0, \ldots, r$. Since ∂G is compact there exist open sets $U_1, \ldots, U_r$ and Lipschitz-transformations $\Phi_1, \ldots, \Phi_r$ such that $\partial G \subset \bigcup_{i=1}^{r} U_i$ and $\Phi_i(U_i \cap G) \in \{E_1, E_2, E_3\}$ (cf. Definition 2). One can find an open set of class C^1 such that $U_0 \subset \mathring{G}$ and $G \subset \bigcup_{i=0}^{r} U_i$. Then $U_0 \cap G = U_0 \in \bigcap_{q \geqq 2} R_q$ (cf. Remark 7). Theorem 2 shows that $U_i \cap G \in \bigcup_{q>2} R_q$, $i=1, \ldots, r$, if $E_i \in \bigcup_{q>2} R_q$, $i=1,2,3$. From Lemma 3 we know already that $E_i \in \bigcap_{q \geqq 2} R_q$, $i=1,2$. Moreover, elementary considerations show that there exists a Lipschitz-transformation mapping E_3 onto E_2. (One can also show that there exists a Lipschitz-transformation from E_3 onto E_4, cf. Remark 9.) Therefore, once more using Theorem 2, we find that $E_3 \in \bigcup_{q>2} R_q$.

References

1. Bergh, J., Löfström, J.: Interpolation spaces. Berlin Heidelberg New York: Springer 1976
2. Gajewski, H., Gröger, K., Zacharias, K.: Nichtlinear Operatorgleichungen und Operatordifferentialgleichungen. Berlin: Akademie-Verlag 1974
3. Giaquinta, M., Giusti, E.: On the regularity of minima of variational integrals. Acta Math. **148**, 31–46 (1982)
4. Košelev, A.I.: Regularity of solutions to elliptic equations and systems (in Russian). Moscow: Nauka 1986
5. Krbec, M.: On L^p-estimates for solutions of elliptic boundary value problems. Comm. Math. Univ. Carol. **17**, 363–375 (1976)
6. Meyers, N.G.: An L^p-estimate for the gradient of solutions of second order elliptic divergence equations. Ann. Scuola Norm. Sup. Pisa **17**, 189–206 (1963)
7. Nečas, J.: Sur la régularité des solutions faibles des équations elliptiques non linéaires. Comm. Math. Univ. Carol. **9**, 365–414 (1968)
8. Simader, C.G.: On Dirichlet's boundary value problem. Lecture Notes Mathematics. Vol. 268. Berlin Heidelberg New York: Springer 1972

Received July 29, 1988

Math. Ann. 283, 689–700 (1989)

Mathematische Annalen

Thetareihen und modulare Spitzenformen zu den Hilbertschen Modulgruppen reell-quadratischer Körper. II

Carl Friedrich Hermann

Universität Mannheim, Seminargebäude A 5, D-6800 Mannheim, Bundesrepublik Deutschland

Im Buch von van der Geer [6] wird gezeigt, daß aus der Existenz eines effektiven und modularen multikanonischen Divisors auf der Hilbertschen Modulfläche $Y_\gamma(D)$ die Minimalität des Modells $Y_\gamma^0(D)$ folgt. Es wird vermutet, daß ein solcher Divisor immer existiert. Falls γ das Geschlecht von $(\sqrt{D})$ ist, wurde diese Vermutung in [3] für zwei Serien von Diskriminanten $D\equiv 1 \bmod 8$ mit Hilfe von auf $H\times H_-$ definierten Γ_D-Spitzenformen vom Gewicht zwei bewiesen [$\Gamma_D=\mathrm{Sl}_2(\mathfrak{o}_D)$.] Im vorliegenden Artikel untersuchen wir, wann geeignete Produkte der in [3] studierten modularen Thetareihen effektive multikanonische Divisoren auf $Y_-(D)$ liefern. Die Bedingungen, die hierbei auftreten, betreffen die Fourierentwicklungen und Nullstellen dieser Thetareihen und sind für Spitzenformen vom Gewicht zwei automatisch erfüllt.

Wir übernehmen die Notation von [3].

Sei a ein quadratfreier Teiler von D. Dann ist $\Pi^2(D,a)$ (s. [3, 3.6]) eine modulare Γ_D-Spitzenform vom Gewicht 10. Die Hurwitz-Maaß-Erweiterung G_D von Γ_D operiert transitiv auf der Menge $\{\Pi(D,a),\ a|D\}$ [3, Satz. 2.] Deshalb ist die Differentialform $\omega^*(\Pi^2(D,a))$ entweder für alle Teiler von D holomorph auf $Y_-(D)$ oder für keinen. Es kommt aber vor, daß $\omega^*\left(\prod_{a|D}(\Pi^2(D,a)\right)$ holomorph ist, obwohl $\omega^*(\Pi^2(D,a))$ für alle $a|D$ Pole hat. Falls $D\not\equiv 5 \bmod 8$, zerfällt $\Pi(D,a)$ in ein Produkt zweier Modulformen. Für $D\equiv 1\bmod 8$ bzw. $D\equiv 0 \bmod 4$ sind $\theta^8(D,a)$ bzw. $\Delta_4^2(D,a)$ modulare Spitzenformen vom Gewicht 4, [3, 3.6]. Zwei Teiler a und $\chi^2\frac{d}{a}$, wobei χ durch $\chi=2$, falls $(d,a)\equiv(3,2) \bmod 4$ und $\chi=1$ sonst definiert ist, sind als äquivalent anzusehen. Wir setzen $\Pi(D)=\prod_{a|D}(\Pi^2(D,a))$, sowie $\Delta(D)=\prod_{a|D}\Delta_4^2(D,a)$, falls $D\equiv 0 \bmod 4$ und $\theta(D)=\prod_{a|D}\theta^8(D,a)$, falls $D\equiv 1 \bmod 8$, wobei sich die Produkte über je $2^{t(D)-1}$ nicht äquivalente Teiler erstrecken. $\Pi(D)$ ist eine modulare Spitzenform vom Gewicht $5\cdot 2^{t(D)}$, und $\Delta(D)$ bzw. $\theta(D)$ sind modulare Spitzenformen vom Gewicht $2^{t(D)+1}$. [$t(D)$ bezeichnet die Anzahl der Primteiler von D.]

Es gilt

Satz 1. a) *$\omega^*(\Pi(D))$ ist genau dann holomorph auf $Y_-(D)$, wenn D kongruent* 6 mod 9 *ist, und es in $\mathbb{Q}(\sqrt{D})$ keine Zahl der Norm -2 gibt.*

b) $(D \equiv 0 \bmod 4)$

$\omega^*(\Delta(D))$ ist genau dann holomorph auf $Y_-(D)$, wenn es in $\mathbb{Q}(\sqrt{D})$ keine Zahl der Norm $-N$ mit $N \in \{1,2,3\}$ gibt.

c) $(D \equiv 1 \bmod 8)$

$\omega^*(\theta(D))$ ist genau dann holomorph auf $Y_-(D)$, wenn es zu jedem

$$N \in \left\{\frac{D-k^2}{4n^2} \in \mathbb{N},\ k, n, \in \mathbb{Z},\ k \equiv n \equiv 1 \bmod 2\right\}$$

einen Primteiler von D gibt, der nicht durch die quadratische Form X^2+4Ny^2 darstellbar ist.

Wir stellen den Beweis von Satz 1 zurück und ziehen zunächst eine Folgerung. Auf $Y_-(D)$ existieren jedenfalls dann effektive und modular multikanonische Divisoren, wenn $Y_-(D) = Y_-^0(D)$ gilt, und D eine der Kongruenzen $D \equiv 6 \bmod 9$, $D \equiv 0 \bmod 4$ oder $D \equiv 1 \bmod 8$ erfüllt.

Beispiele für a) sind $D = 60, 69, 105$. Alle Diskriminanten $D = 4m$, wobei m eine quadratfreie Zahl mit $m \equiv -1 \bmod 24$ ist, genügen beispielsweise der Bedingung b). Für Diskriminanten $D \equiv 1 \bmod 8$ ist die Aussage von Satz 1 komplizierter. Aus $n \equiv 1 \bmod 2$ folgt $N \equiv 0 \bmod 2$, also ist $\omega^*(D)$ jedenfalls dann holomorph, falls D einen Teiler a mit $a \not\equiv 1 \bmod 8$ besitzt in Übereinstimmung mit [3] Satz 6. Weitere Beispiele für c) sind etwa die Diskriminanten $D = 17p$, wobei p eine Primzahl mit $p = a^2+8b^2 = c^2+16d^2$, $a, b, c, d \in 2\mathbb{Z}+1$ ist. Denn es gibt dann keine ungerade Zahlen k, n mit $\dfrac{D-k^2}{4n^2} \in \{2,4\}$, aber aus $x^2+Ny^2=17$ und $N \equiv 0 \bmod 2$ folgt $N \in \{2,4\}$. Aus einer modularen Spitzenform $f \in \left[\Gamma_D, \frac{k}{2}, v\right]$, $k \in \mathbb{Z}$, erhält man auf folgende Weise weitere modulare Spitzenformen. Sei $\mathfrak{M} = \alpha\mathfrak{o}_D + \beta\mathfrak{o}_D$ ein ganzes $\mathfrak{o}_D$-Ideal, $\mathfrak{p}$ ein Primideal der Norm p und $\lambda \in \mathbb{Q}(\sqrt{D})$ eine total-positive Zahl mit $\mathfrak{M}^2 = (\lambda)\mathfrak{p}$. Es gibt Zahlen η, $\xi \in \mathfrak{M}^{-1}$ mit

$$M_{\mathfrak{M}} = \begin{pmatrix} \alpha & \eta \\ \beta & \xi \end{pmatrix} \in \mathrm{Sl}_2(\mathbb{Q}(\sqrt{D}))\,.$$

Setze

$$I = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} (\sqrt{\lambda})^{-1} & 0 \\ 0 & \sqrt{\lambda} \end{pmatrix}$$

und

$$T_p(\mathfrak{M}, \lambda, f) = \left(f \underset{\frac{k}{2}}{|} M\Lambda\right) \prod_{j=0}^{p-1} \left(f \underset{\frac{k}{2}}{|} M\Lambda I T^j\right).$$

Direkte Rechnung unter Benutzung der bekannten Tatsache, daß Γ_D von den speziellen Matrizen I und $\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$, $\alpha \in \mathfrak{o}_D$ erzeugt wird, ergibt

$$T_p(\mathfrak{M}, \lambda, f) \in \left[\Gamma_D, k\frac{p+1}{2}, v^*\right]$$

mit einem Γ_D-Charakter v^*. $T_p(\mathfrak{M}, \lambda, f)$ ist genau dann modular, wenn f modular ist. Falls f ein Produkt von Thetareihen der Art [3,(9)] ist, dann ist auch $T_p(\mathfrak{M}, \lambda, f)$ eine Thetareihe. Dies folgt aus [3, Satz 3]. Leider wurde dort ein Term vergessen. Korrigiert und leicht verallgemeinert lautet dieser Satz

Satz 2. *Sei* $\mathfrak{M} = \mathfrak{o}_D\alpha + \mathfrak{o}_D\beta$ *ein (gebrochenes) Ideal in* $\mathbb{Q}(\sqrt{D})$ *und*

$$M_{\mathfrak{M}} = \begin{pmatrix} \alpha & \eta \\ \beta & \xi \end{pmatrix} \in \mathrm{Sl}_2(\mathbb{Q}(\sqrt{D}))$$

wobei $\eta, \xi \in \mathfrak{M}^{-1}$. *Für jeden quadratfreien Teiler* a *von* D *gibt es eine Abbildung* $t(a, \mathfrak{M}): C(\mathfrak{a}_a) \to C(\mathfrak{a}_a\mathfrak{M})$, *so daß für alle* $z \in H \times H_-$ *gilt*

$$(\theta^8(D, a, \mathfrak{o}_D, \mu, v) \mathop{|}_{4} M_{\mathfrak{M}})(z) = (\mathrm{Norm}(\mathfrak{M}))^4 \theta^8(D, \mathfrak{a}_a\mathfrak{M}, t(a, \mathfrak{M})(\mu, v)) \left(\frac{z}{a}\right).$$

Für $D \equiv 1 \bmod 8$ *gilt* $t(a, \mathfrak{M})(1,1) = (1,1)$ *und für* $D \equiv 0 \bmod 4$

$$t(a, \mathfrak{M})(C_k(\mathfrak{a}_a)) = C_k(\mathfrak{a}_a\mathfrak{M}), \qquad k = 4, 6.$$

Beweis. Wegen $M_{\lambda\mathfrak{M}} = M_{\mathfrak{M}} \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$ genügt es, Satz 3 für ein zu $\mathfrak{M}$ äquivalentes Ideal zu beweisen. In jeder Klasse gibt es ein Ideal $\mathfrak{J}$ mit den in [3] Satz 3 vorausgesetzten Eigenschaften. Also gibt es Matrizen

$$M \in \mathrm{Sp}_2(\mathbb{Z}) \quad \text{und} \quad \Omega = \begin{pmatrix} U & 0 \\ 0 & U^{t^{-1}} \end{pmatrix} \in \mathrm{Sp}_2(\mathbb{Q})$$

mit $\det U = \mathrm{Norm}\,\mathfrak{J}$ und $\phi(a, \mathfrak{o}_D)(M_{\mathfrak{J}}) = M\Omega$. Aus

$$(\theta_m^8 | M\Omega)(Z) = (\theta_m^8 | \Omega)(Z) = (\det U)^4 \theta_m^8(\Omega(Z))$$

folgt dann Satz 3. Der Faktor $(\det U)^4 = (\mathrm{Norm}\,\mathfrak{J})^4$ war in [3] vergessen worden.

Falls $\mathfrak{M}$ ein Hauptideal (α_p) ist, wobei $\alpha_p \in \mathfrak{o}_D$, $\alpha_p \gg 0$ und $N(\alpha_p) = p$, gilt

$$T_p((\alpha_p), \alpha_p, f) = f(\alpha_p z) \prod_{x=0}^{p-1} f\left(\bar{\alpha}_p\left(\frac{z+x}{p}\right)\right).$$

In [3] wurde (in Fall $D \equiv 1 \bmod 8$) ein hinreichendes Kriterium dafür angegeben, daß $\theta(D, a_1) T_2((\alpha_2), \alpha_2, \theta(D, a_2))$ eine Spitzenform vom Gewicht zwei zur vollen Gruppe Γ_D ist. Das verallgemeinern wir nun und setzen dazu voraus, daß $D = u^2 + v^2 = s^2 + pt^2$ gilt (p eine Primzahl), wobei $u, s \in 2\mathbb{N}+1$, $v \in 4\mathbb{N}$, $t \in 2\mathbb{N}$, und daß es in $\mathbb{Q}(\sqrt{D})$ eine Einheit $\varepsilon > 0$ negativer Norm gibt. Setze $\gamma_1 = \frac{u+\sqrt{D}}{2}$, $\mathfrak{J}_1 = \frac{v}{2}\mathbb{Z} + \gamma_1\mathbb{Z}$, $\gamma_p = \frac{s+\sqrt{D}}{2}$ und $\mathfrak{J}_p = \frac{t}{2}\mathbb{Z} + \gamma_p\mathbb{Z}$. Es gilt $\mathfrak{J}_1^2 = (\gamma_1)$ und $\mathfrak{J}_p^2 = \mathfrak{p}\left(\frac{\gamma_p}{p}\right)$ mit einem Primideal $\mathfrak{p}$ der Norm p. Setze $\vartheta(D, p)\,(z) = T_p(\mathfrak{J}_p, -\bar{\varepsilon}\gamma_p, \theta(D, 1)(z)$. Aus Satz 2 folgt (Bezeichnungen wie in [3] S. 332–333)

$$\vartheta(D, p)(z) = \theta(D, 1, \mathfrak{J}_p, 1, 1)\left(\frac{\varepsilon(-\bar{\gamma}_p)z}{(t/2)^2}\right) \prod_{x=0}^{p-1} \theta(D, 1, \bar{\mathfrak{J}}_p, 1, 1)\left(\frac{(-\bar{\varepsilon})\gamma_p(z+x)}{(t/2)^2 p}\right). \tag{1}$$

Die Charakteristik $(1,1) \in \mathfrak{J}_1 \bmod 2\mathfrak{J}_1$ wird durch γ_1 repräsentiert. Falls $\frac{t}{2} \equiv 1 \bmod 2$ wird $(1,1) \in \mathfrak{J}_p \bmod 2\mathfrak{J}_p$ durch $\frac{t}{2}$ repräsentiert. Dies kommt nur für

$p=2$ vor, und der Charakter der Modulform $\theta(D,1)\vartheta(D,2)$ ist dann nicht trivial. Falls $t\equiv 0 \bmod 4$ wird $(1,1)\in\mathfrak{J}_p \bmod 2\mathfrak{J}_p$ durch γ_p repräsentiert. Das Transformationsverhalten der Thetareihen auf der rechten Seite von (1) unter Translationen $z\mapsto z+\alpha$, $\alpha\in\mathfrak{o}_D$, erhält man leicht durch direkte Rechnung unter Verwendung von [3] Formel (12). Es gilt

$$\theta(D,1)(z+\bar{\varepsilon}\alpha)=(-1)^{S(\alpha)\frac{v}{4}}e^{\frac{\pi i}{4}S(\alpha)}\theta(D,1)(z)$$

und

$$\vartheta(D,p)(z+\bar{\varepsilon}\alpha)=e^{\frac{\pi i}{4}(p+1)S(\alpha)}\vartheta(D,p)(z). \tag{2}$$

Ein Γ_D-Charakter ist genau dann trivial, wenn er auf der Untergruppe

$$\left\{\begin{pmatrix}1 & \alpha\\ 0 & 1\end{pmatrix}, \alpha\in\mathfrak{o}_D\right\}$$

trivial ist. Aus (2) folgt daher

Satz 3. *Sei D die Diskriminante eines reell-quadratischen Zahlkörpers, dessen sämtliche Teiler kongruent* $1 \bmod 8$ *sind. Dann gibt es Zahlen $u,s\in 2\mathbb{N}+1$ und $v,t\in 2\mathbb{N}$ mit $D=u^2+v^2=s^2+2t^2$ und eine durch diese bestimmte modulare Γ_D-Spitzenform $\theta(D,1)\vartheta(D,2)$ vom Gewicht zwei zu einem Charakter C mit $C^4=1$, der genau dann trivial ist, wenn $t\equiv 0 \bmod 4$ und $v\equiv 4 \bmod 8$ gilt.*

Die ersten 10 Primzahlen, die die Bedingungen von Satz 3 erfüllen, sind

41, 137, 313, 409, 457, 521, 569, 761, 809, 857.

Ein einfaches Kriterium dafür, wann die zu den modularen Spitzenformen der Art

$$f=\theta(D,1)^r\vartheta(D,p)\in\left[\Gamma_D,\frac{p+1+r}{2},1\right]$$

gehörenden $\frac{p+1+r}{4}$-kanonischen Divisoren effektiv sind, habe ich außer für $(r,p)=(1,2)$ nicht gefunden. Allerdings kann $\mathrm{Div}(\omega^*(f))$ in jedem Einzelfall explizit berechnet werden. (s. dazu die Beispiele weiter unten). Aus Satz 3 und [6, VII 7.12] folgt

Satz 4. *$Y_-^0(D)$ ist minimal, falls* $D=u^2+v^2=s^2+2t^2$ wobei $v\equiv 4 \bmod 8$ *und* $t\equiv 0 \bmod 4$.

Beweis von Satz 1. Für modulares $f\in[\Gamma_D,2k]$ gilt

$$\mathrm{Div}(\omega^*(f))=\mathfrak{F}(f)+\mathfrak{S}(f)+\mathfrak{E}(f)$$

mit $\mathfrak{F}(f)=\sum_{N,i}a_N^{(i)}F_N^{(i)}$, $a_N^{(i)}\geqq 0$. Der Index i numeriert hier die verschiedenen Komponenten der Kurve F_N. Der Träger des Divisors $\mathfrak{S}(f)$ setzt sich aus rationalen Kurven, die durch die Auflösung der Spitzen entstehen, zusammen, und der Träger von $\mathfrak{E}(f)$ besteht aus rationalen Kurven, die von der Auflösung der elliptischen Fixpunkte herkommen.

Für $D=5, 8, 12$ sind $\mathrm{Div}(\omega^*(\Pi(D))$ bzw. $\mathrm{Div}(\omega^*(\Delta(D)))$ nicht effektiv. Falls $D>12$ gibt es nur elliptische Fixpunkte der Ordnung 2 oder 3. Ihre Auflösung wird im Buch [6] von van der Geer detailliert beschrieben (Kap. II 3). Falls in der Auflösung eines Fixpunktes nur (-2)-Kurven vorkommen, gibt es keine Fortsetzungsprobleme [6, 3.3]. Sei E eine (-3)-Kurve, die zu einem Fixpunkt e der Ordnung 3 gehört. Jede Modulform $f\in[\Gamma_D, k]$ mit $k\not\equiv 0 \bmod 3$ verschwindet in e. Für alle $a|D$ liegt demnach e auf einer Nullstellenkurve von $\Pi(D,a)$.

Hilfssatz 1. *Zwei irreduzible Nullstellenkurven von $\Pi(D,a)$ schneiden sich im Endlichen nicht.*

Dies folgt aus [3, Satz 4] und einem Resultat von Franke, wonach zwei verschiedene Humbertsche Flächen der Diskriminante 1 auf der Siegelschen Halbebene keinen Schnittpunkt haben [1, 3.3.4]. Im folgenden bezeichne f eine der Modulformen

$$f\in\{\Pi^2(D,a), \Delta_4^2(D,a), \theta^8(D,a)\}.$$

Aus Hilfssatz 1 folgt, daß die Nullstellenkurven von f auf $Y_-(D)$ nicht singulär sind, und daß E genau eine dieser Kurven schneidet. Wir folgern daraus, daß $\omega^*(\Delta(D))$ und $\omega^*(\theta(D))$ über allen elliptischen Fixpunkten holomorph sind, und daß dies für $\omega^*(\Pi(D))$ genau dann gilt, wenn in der Auflösung der Fixpunkte nur (-2)-Kurven vorkommen. Nach [5, 2.1 (7)] ist dies für Diskriminanten $D\equiv 6 \bmod 9$ und nur für diese der Fall.

Wir setzen $\delta(D)=0$, falls es in $\mathfrak{o}_D$ eine Einheit negativer Norm gibt, und $\delta(D)=1$ sonst. Die Hurwitz-Maaß-Erweiterung G_D von Γ_D operiert auf der Menge der Spitzen und zerlegt sie in Orbits mit $h_1(D)=2^{t(D)-1-\delta(D)}$ Elementen. Es gibt $h_1(D)$ Teiler a_j von D und $h_2(D)=h(D)(h_1(D))^{-1}$ ganze Ideale $\mathfrak{J}_i$, so daß die $h(D)$ Ideale $\mathfrak{a}_{a_j}\mathfrak{J}_i$, $j=1,\dots,h_1(D)$, $i=1,\dots,h_2(D)$ alle Idealklassen von $\mathbb{Q}(\sqrt{D})$ repräsentieren. Der Divisor $\mathfrak{S}(f)$ zerfällt in eine Summe

$$\mathfrak{S}(f)=\sum_{j=1}^{h_1(D)}\sum_{i=1}^{h_2(D)}\mathfrak{S}(\mathfrak{a}_{a_j}\mathfrak{J}_i, f). \tag{3}$$

Wegen [3, Satz 2] genügt es, $\mathfrak{S}(\mathfrak{J}_i, f)$ zu berechnen. Mit $B_{i,k}$, $k\in\mathbb{Z}$, seien die Randpunkte der Menge $\{\alpha\in\mathfrak{J}_i^{-2},\ \alpha>0,\ \bar{\alpha}<0\}$ bezeichnet. Es gilt $\mathfrak{J}_i^{-2}=\mathbb{Z}B_{i,k}+\mathbb{Z}B_{i,k+1}$. Über die Gleichungen

$$\begin{aligned} 2\pi i z_1 &= B_{i,k-1}\log u_{i,k}+B_{i,k}\log v_{i,k},\\ 2\pi i(-z_2) &= \bar{B}_{i,k-1}\log u_{i,k}+\bar{B}_{i,k}\log v_{i,k} \end{aligned} \tag{4}$$

werden Koordinaten $(z_1, -z_2)\in H\times H_-$ in Koordinaten $(u_{i,k}, v_{i,k})$ auf $Y_-(D)$ umgerechnet (s. [4, 2.3] oder [6, Kap. II]).

Die Länge der Kettenbruchentwicklung einer zu $\mathfrak{J}_i^{-2}$ gehörenden reduzierten quadratischen Irrationalzahl sei $2^{-\delta(D)}l_i$ und die Geraden S_{ik} seien durch die Gleichungen $v_{i,k}=u_{i,k+1}=0$ gegeben. Wir setzen

$$\min(D,a,\mu,i,k)=\min\left\{S\left(\frac{g^2B_{i,k}}{a\sqrt{D}}\right),\ g\in\mathfrak{a}_a\mathfrak{J}_i,\ g\equiv\mu \bmod 2\mathfrak{o}_a\mathfrak{J}_i\right\}$$

und $\Omega(i,j)=(\mathfrak{a}_{a_j}\mathfrak{J}_i/2\mathfrak{a}_{a_j}\mathfrak{J}_i)\backslash\{0\}$.

Indem wir (4) in die Fourierentwicklung $(f|M_{\mathfrak{J}_i})(z)$ einsetzen, erhalten wir

$$\mathfrak{S}(\mathfrak{J}_i, f) = \sum_{k=0}^{l_1-1} c(i,k)(f) S_{ik} \tag{5}$$

mit

$$c(i,k)(\Pi^2(D,a)) = \tfrac{1}{2}\Big(\sum_{\mu\in\Omega(i,j)} \min(D,a,\mu,i,k)\Big) - 5,$$

$$c(i,k)(\Delta^2(D,a)) = \tfrac{1}{2}(\min(D,a,n,i,k)) - 2$$

und

$$c(i,k)(\theta^8(D,a)) = (\min(D,a,\tilde{1},i,k)) - 2,$$

wobei n das in [3, 3.2] eingeführte Element $n \in \Omega(i,j)$ ist, und $\tilde{1}$ den vorderen Teil der Charakteristik $(1,1) \in C(\mathfrak{a}_{a_j}\mathfrak{J}_i)$ bezeichnet.
Sei $\mathfrak{J}$ ein $\mathfrak{o}_D$-Ideal und $\lambda \in \mathfrak{J}^{-2}$ eine Zahl mit $\lambda > 0$ und $\bar{\lambda} < 0$. Nach (5) muß man zur Berechnung von $\mathfrak{S}(f)$ die Minima der Abbildungen

$$Q(a,\mathfrak{J},\lambda) : (\mathfrak{a}_a\mathfrak{J})^2 \setminus \{0\} \to \mathbb{N},$$

$$g \mapsto S\left(g^2 \frac{\lambda}{a\sqrt{D}}\right)$$

für $\mathfrak{J} = \mathfrak{J}_i$ und $\lambda = B_{ik}$ unter den Nebenbedingungen $g \equiv \mu \bmod 2\mathfrak{a}_a\mathfrak{J}_i$ bestimmen. Zum Beweis von Satz 1 genügt es aber, $Q(a_j, \mathfrak{J}, \lambda)$ für verschiedene Teiler a_j von D zu vergleichen. Zu $Q(a,\mathfrak{J},\lambda)$ gehört nach Wahl einer $\mathbb{Z}$-Basis von $\mathfrak{a}_a\mathfrak{J}$ eine positiv definite quadratische Form der Diskriminante $4N(\lambda)m^2$, wobei $m = \operatorname{Norm}\mathfrak{J}$. Wir setzen o.B.d.A. voraus, daß $\mathfrak{J}$ relativ prim zu $(2D)$ ist. Dann gibt es eine Zahl $\omega(\mathfrak{J}) = \frac{b+\sqrt{D}}{2}$ mit $N(\omega(\mathfrak{J})) \equiv 0 \bmod 2dm^2$ und $\mathfrak{o}_a\mathfrak{J} = \mathbb{Z}am + \mathbb{Z}\omega(\mathfrak{J})$. Setze $Q(a,\mathfrak{J},\lambda)(x,y) = S\left((xam + y\omega)^2 \frac{\lambda}{a\sqrt{D}}\right)$. Die Wahl von $\{am, \omega(\mathfrak{J})\}$ als Basis von $\mathfrak{a}_a\mathfrak{J}$ hat den Vorteil, daß für zwei quadratfreie Teiler a_1 und a_2 von D die Formel

$$Q(a_2,\mathfrak{J},\lambda)\left(\sqrt{\frac{a_1}{a_2}}\,x, \sqrt{\frac{a_2}{a_1}}\,y\right) = Q(a_1,\mathfrak{J},\lambda)(x,y) \tag{6}$$

gilt. Im folgenden bezeichnen wir die quadratische Form $Q(x,y) = ax^2 + bxy + cy^2$ auch einfach mit Q oder auch mit $[a,b,c]$ und setzen $(a,b,c) = \operatorname{cont} Q$.

Aus (6) erhält man nach einigen elementaren Rechnungen

Hilfssatz 2. *$Q(a_2,\mathfrak{J},\lambda)$ ist genau dann* $\mathrm{Sl}_2(\mathbb{Z})$*-äquivalent zu $Q(a_1,\mathfrak{J},\lambda)$, wenn es in der Ordnung der Diskriminante*

$$4N(\lambda)m^2(\operatorname{cont} Q(a_1,\mathfrak{J},\lambda), \operatorname{cont} Q(a_2,\mathfrak{J},\lambda))^{-1}$$

von $\mathbb{Q}(\sqrt{N(\lambda)})$ ein Element der Norm $\dfrac{a_1a_2}{(a_1,a_2)^2}$ gibt.

Der quadratischen Form $Q = [a,b,c]$ ordnen wir die symmetrische Matrix

$$\varphi(Q) = \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$$

zu. Die Abbildung

$$\varphi(a,\mathfrak{J}):\mathfrak{J}^{-2}\to M_2(\mathbb{Q}),$$
$$\lambda\mapsto\varphi(Q(a,\mathfrak{J},\lambda))$$

ist ein Homomorphismus von $\mathbb{Z}$-Moduln mit $\varphi(a,\mathfrak{J})(\mathfrak{J}^{-2})\subset M_2(\mathbb{Z})$ (s. auch [3, (12)]). Die Matrix $\varphi(a,\mathfrak{J}_i)(B_{ik})$ is primitiv, weil $\{B_{ik},B_{i,k+1}\}$ eine Basis von $\mathfrak{J}_i^{-2}$ ist. Also gilt $\operatorname{cont}Q(a,\mathfrak{J}_i,B_{ik})\in\{1,2\}$. Zu jeder positiv definiten quadratischen Form Q gibt es bekanntlich eine zu Q äquivalente Form $Q_{\text{red}}=[a,b,c]$ mit 1) $a,c>0$ und 2) $-a<b\leqq a\leqq c$ (s. z.B. [7, 13]). Es gilt

$$\min\{Q(x,y),\,(x,y)\equiv(1,0)\bmod 2\}=a,$$
$$\min\{Q(x,y),\,(x,y)\equiv(0,1)\bmod 2\}=c$$

und

$$\min\{Q(x,y),\,(x,y)\equiv(1,1)\bmod 2\}=a-b+c.$$

Wir kommen nun zurück auf Formel (5) und setzen $Q(a,\mathfrak{J}_i,B_{ik})_{\text{red}}=[r,2s,t]$. Eine einfache Rechnung ergibt $c(i,k)(\Pi^2(D,a))=r-s+t-5$. Also gilt

$$c(i,k)(\Pi^2(D,a))<0$$
$$\Rightarrow\ Q(a,\mathfrak{J}_i,B_{ik})_{\text{red}}\in\{[1,0,1],\,[1,0,2],\,[1,0,3],\,[2,2,2],\,[1,0,4],\,[2,2,3]\}.$$

Falls es in $\mathbb{Q}(\sqrt{D})$ ein Element der Norm -2 gibt, dann auch ein B_{ik} mit $N(B_{ik})=-2$ und $Q(a,\mathfrak{J}_i,B_{ik})_{\text{red}}=[1,0,2]$ für alle $a|D$. Die in Satz 1a) genannten Bedingungen sind also notwendig. Aus $D\equiv 6\bmod 9$ folgt, daß es in $\mathbb{Q}(\sqrt{D})$ kein Element der Norm -1, -3 oder -4 gibt. Es gibt genau zwei $\mathrm{Sl}_2(\mathbb{Z})$-Äquivalenzklassen positiv definiter quadratischer Formen der Determinante 5. Sie werden durch $[1,0,5]$ und $[2,2,3]$ repräsentiert. Sei $B_{ik}\in\mathfrak{J}_i^{-2}$ eine Zahl mit $N(B_{ik})m_i^2=-5$. Weil $x^2+5y^2=3$ keine ganzzahlige Lösung hat, folgt aus Hilfssatz 2, daß $Q(a,\mathfrak{J}_i,B_{ik})$ und $Q\left(\frac{3a}{(3,a)^2},\mathfrak{J}_i,B_{ik}\right)$ nicht äquivalent sind, also ist die Multiplizität von S_{ik} in $\mathfrak{S}(\pi(D))$ gleich 0 und Satz 1a) ist bewiesen.

Der Beweis der Teile b) und c) von Satz 1 erfolgt weitgehend analog. Diskriminanten $D\equiv 0\bmod 4$ haben immer mindestens zwei Teiler. Die Pole von $\omega^*(\Delta_4^2(D,a))$ werden durch Nullstellen von $\omega^*\left(\Delta_4^2\left(D,\frac{2a}{(2,a)^2}\right)\right)$ kompensiert, außer wenn $\text{Norm}\,(\mathfrak{J}_i^2)N(B_{ik})\in\{-1,-2,-3\}$. Im Fall $D\equiv 1\bmod 8$ ist $\operatorname{Div}(\omega^*(\theta^8(D,a)))$ für kein $a|D$ effektiv. Falls $\omega^*(\theta^8(D,a_j)),\,j=1,2$, eine gemeinsame Polkurve S_{ik} haben, gibt es $M_j\in\mathrm{Sl}_2(\mathbb{Z})$ mit

$$\varphi(D,a_j)(B_{ik})=M_j\begin{pmatrix}1&0\\0&N\end{pmatrix}M_j^t,$$

wobei die Matrizen M_j noch einer Kongruenzbedingung mod 2 genügen müssen, die von der Wahl einer Basis von $\mathfrak{a}_{a_j}\mathfrak{J}_i\bmod 2\mathfrak{a}_{a_j}\mathfrak{J}_i$ abhängt. Aus [3, Satz 4] folgt $N\in\left\{\frac{D-k^2}{4n^2},\,k,n\in\mathbb{Z}\right\}$ und aus Hilfssatz 2 folgt, daß es Zahlen $x,y\in\mathbb{Z}$ gibt mit

$x^2+Ny^2=\frac{a_1a_2}{(a_1,a_2)^2}$. Die oben erwähnte Kongruenzbedingung erzwingt $y\equiv 0 \bmod 2$ und $n\equiv 1 \bmod 2$. Daher ist S_{ik} kein Pol von $\left(\theta^8(D,a)\theta^8\left(D,\frac{ap}{(a,p)^2}\right)\right)$, falls p nicht durch die quadratische Form x^2+4Ny^2 dargestellt wird. Nun sei eine Zahl $N=\frac{D-r^2}{4n^2}$, $N,r,n\in\mathbb{Z}$, $n\equiv 1 \bmod 2$, $r^2<D$ vorgegeben. Setze $\alpha_N=\frac{r+\sqrt{D}}{2}$. Wir können o.B.d.A. annehmen, daß $\mathfrak{J}:=\mathbb{Z}n+\mathbb{Z}'\alpha_N$ eines der Ideale $\mathfrak{J}_i$, $i\in\{1,\ldots,h_2(D)\}$ ist. Setze $\beta_N=\frac{-\alpha_N}{n^2}$. Es gilt

$$\varphi(1,\mathfrak{J}_i)(\beta_N)=M\begin{pmatrix}N & 0\\ 0 & 1\end{pmatrix}M^t$$

mit $M\in\Gamma_0(2)$ und daher

$$\min\left\{S\left(\frac{g^2\beta_N}{\sqrt{D}}\right),\ g\in\mathfrak{J}_i,\ g\equiv n \bmod 2\mathfrak{J}_i\right\}=1\,.$$

Daraus folgt, daß es ein $k\in\mathbb{Z}$ gibt mit $\beta_N=B_{ik}$ und $c(i,k)(\theta^8(D,1))=-1$. Wenn jeder Teiler von D durch x^2+4Ny^2 darstellbar ist, gilt wegen Hilfssatz 2 $c(i,k)(\theta^8(D,a))=-1$ für alle $a|D$. Also tritt S_{ik} in $\mathrm{Div}(\omega^*(\theta(D))$ mit der Vielfachheit $-2^{t(D)}$ auf.

Ich möchte nun anhand einiger Beispiele illustrieren, wie man

$$\mathrm{Div}(\omega^*(f)),\quad f\in\{\Pi^2(D,a),\ \Delta_4^2(D,a),\ \theta^8(D,a)\}$$

explizit berechnen kann. Zunächst einige allgemeine Bemerkungen. Auf den Nullstellenkurven von f liegt kein Fixpunkt der Ordnung 3 und der zweiten Art, also geben diese Fixpunkte in $\mathfrak{C}(f)$ keinen Beitrag. Sei $v_2(f)$ die Multiplizität, mit der die zum Fixpunkt e der Ordnung 2 gehörende (-2)-Kurve in $\mathfrak{C}(f)$ auftritt. Dann gilt $v_2(f)=0$, falls e nicht auf $\mathfrak{F}(f)$ liegt und (wegen Hilfssatz 1) $v_2(\Pi^2(D,a))=v_2(\Delta_4^2(D,a))=1$, $v_2(\theta^8(D,a))=4$ sonst. Sei E eine (-3)-Kurve, die zu einem Fixpunkt der Ordnung 3 gehört. Dann enthält $\mathfrak{C}(f)$ den Beitrag $v_3(f)E$ mit $v_3(\Pi^2(D,a))=-1$, $v_3(\Delta_4^2(D,a))=0$ und $v_3(\theta^8(D,a))=2$.
Nach [3] Satz 4 ist jede Nullstellenkurve von $\Pi^2(D,a)$ in $\bigcup F_N$, $N\in\left\{\frac{D-k^2}{4m^2}\in\mathbb{N}\right\}$ enthalten. Setze $\alpha_N=\frac{k+\sqrt{D}}{2}$ und $\mathfrak{M}=\mathfrak{o}_D m+\mathfrak{o}_D\alpha_N$. Die schief-Hermitesche Matrix

$$B_N:=\frac{1}{m}M_{\mathfrak{M}}\begin{pmatrix}0 & \bar{\alpha}_N\\ -\alpha_N & 0\end{pmatrix}\bar{M}_{\mathfrak{M}}t$$

($M_{\mathfrak{M}}$ wie in Satz 2) ist primitiv. Die durch B_N definierte Komponente von F_N sei mit $F_N^{(1)}$ bezeichnet. Aufgrund von Satz 2 verschwindet $\theta(D,a,\mathfrak{o}_D,\mu,\nu)$ auf $F_N^{(1)}$ genau dann, wenn $\theta(D,a,\mathfrak{M},t(\mu,\nu))$ auf $\{(\alpha_N,\bar{\alpha}_N)z,z\in H\}\subset H\times H_-$ verschwindet.

Mit Hilfe dieses Kriteriums sowie [3, Satz 2] und den Überlegungen von [3], Beweis von Satz 4, kann man $\mathfrak{F}(f)$ leicht berechnen. Man muß dazu nur feststellen, welche der quadratischen Formen $Q\left(a,\mathfrak{M},\frac{\alpha_N}{m^2}\right)$ diagonalisierbar sind.

In unseren Beispielen gibt es in $\mathfrak{o}_D$ nur je ein Idealgeschlecht. Zur Bestimmung von $\mathfrak{S}(f)$ genügt es daher, die Minima der quadratischen Formen $Q(a, \mathfrak{o}_D, B_k)$ wobei

$$\{\alpha \in \mathfrak{o}_D,\ \alpha > 0,\ \bar{\alpha} < 0\} = \ldots + \mathbb{N}B_{-1} + \mathbb{N}B_0 + \mathbb{N}B_1 + \ldots,$$

zu berechnen. Falls die Grundeinheit von $\mathbb{Q}(\sqrt{D})$ positiv ist, gibt es zu jedem a ein $\tilde{a}$, so daß $Q(a, \mathfrak{I}, \lambda)$ äquivalent ist zu $Q(\tilde{a}, \mathfrak{I}, \varepsilon\lambda)$. Deshalb und wegen $Q(a, \mathfrak{I}, \lambda) = Q(a, \bar{\mathfrak{I}}, \bar{\lambda})$ genügt es, $Q(a, \mathfrak{o}_D, B)$ für $k = 0, 1, \ldots, \frac{1}{2}(2^{-\delta(D)} l_- + \mu)$ zu reduzieren ($\mu = 1$, falls $l \equiv 1 \bmod 2$ und $\mu = 2$ sonst. (l_- bezeichnet die Länge des zu $(\sqrt{D})$ gehörenden Zykels). Zur Illustration der einzelnen Rechenschritte behandeln wir das erste Beispiel ziemlich ausführlich.

<u>$D = 28$</u>

$$\left(\frac{1}{\sqrt{7}}\right) = \mathbb{Z} + \mathbb{Z}\omega,\quad \omega = \frac{7+\sqrt{7}}{7} = [[\overline{b_0, \ldots, b_4}]] = [[\overline{2, 2, 3, 3, 2}]],\quad l_- = 10,\quad \delta(28) = 1,$$

$B_{-1} = 1 + \sqrt{7}$, $B_0 = \sqrt{7}$, $B_{k+1} = b_k B_k - B_{k-1}$.

$$Q(a, \mathfrak{o}_{28}, x + y7\bar{\omega}) = \left[\frac{14}{a}x + \frac{42}{a}y,\ 2x,\ -ay\right], \qquad a \in \{1, 2\}.$$

Setze $\varphi(Q(a, \mathfrak{o}_{28}, B_k)) = R(a, k)\varphi(Q(a, \mathfrak{o}_{28}, B_k)_{\mathrm{red}})R(a, k)^t$. Es gilt

k	0	1	2
b_k	2	2	3
B_k	$\sqrt{7}$	$\sqrt{7}-1$	$\sqrt{7}-2$
$Q(1, \mathfrak{o}_{28}, B_k)_{\mathrm{red}}$	$[1, 0, 7]$	$[1, 0, 6]$	$[1, 0, 3]$
$Q(2, \mathfrak{o}_{28}, B_k)_{\mathrm{red}}$	$[2, 2, 4]$	$[2, 0, 3]$	$[2, 2, 2]$
$R(1, k) \bmod 2$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$
$R(2, k) \bmod 2$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$	$\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$

Die Charakteristik $n \in \mathfrak{a}_a \bmod 2\mathfrak{a}_a$ wird für $a = 1$ durch $n = (1, 0)\begin{pmatrix}7+\sqrt{7}\\ 1\end{pmatrix}$ und für $a = 2$ durch $n = (0, 1)\begin{pmatrix}7+\sqrt{7}\\ 2\end{pmatrix}$ repräsentiert. Daher gilt

$$\min\left\{S\left(g^2 \frac{B_k}{\sqrt{D}}\right),\ g \in \mathfrak{o}_{28},\ g \equiv 7 + \sqrt{7} \bmod 2\right\}$$
$$= \min\{r_k x^2 + 2s_k xy + t_k y^2,\ (x, y) \equiv (1, 0)R(1, k) \bmod 2\},$$

wobei $[r_k, 2s_k, t_k] = Q(1, \mathfrak{o}_{28}, B_k)_{\mathrm{red}}$. Mit Hilfe von Formel (5) und der Tabelle errechnet man

$$\mathfrak{S}(\Delta_4^2(28, 1)) = 2S_0 + S_1 - S_3 - S_4 - S_5 - S_6 - S_7 + S_9.$$

Analog dazu findet man

$$\mathfrak{S}(\Delta_4^2(28,2)) = -S_0 - S_1 - S_2 + S_4 + 2S_5 + S_6 - S_8 - S_9 .$$

Die Komponenten $F_7^{(1)}$ bzw. $F_3^{(1)}$ seien durch die Bedingungen $\{(B_0, \bar{B}_0)z,\ z \in H\} \subset F_7^{(1)}$ bzw. $\{(B_2, \bar{B}_2)z, z \in H\} \subset F_3^{(1)}$ definiert. Es gilt

$$\mathfrak{F}(\Delta_4^2(28,a)) = 2(F_3^{(a)} + F_7^{(a)}), \qquad a \in \{1,2\}$$

sowie $\mathfrak{F}\left(\dfrac{\Pi(28,1)}{\Delta_4(28,1)}\right) = \mathfrak{F}\left(\dfrac{\Pi(28,2)}{\Delta_4(28,2)}\right) = F_6$. Letzteres impliziert

$$\Delta_6 = : \frac{\Pi(28,1)}{\Delta_4(28,1)} \in \mathbb{C} \frac{\Pi(28,2)}{\Delta_4(28,2)} .$$

Man kann $\Delta_6 \in [\tilde{G}_{28}, 3, 1]$ zeigen. Es gilt

$$\operatorname{Div}(\omega^*(\Delta(28))) = 2(F_3 + F_7) + S_0 - S_2 - S_3 + S_5 - S_7 - S_8 + 2E^{(1)} + E^{(2)} + E^{(3)},$$

wobei die $E^{(i)}$ zu den Fixpunkten der Ordnung 2 gehören, die auf F_7 liegen. Weiter gilt

$$\operatorname{Div}(\omega^*(\Delta_6^2)) = 2F_6 + S_0 + S_1 - S_2 - S_3 + S_4 + S_5 + S_6 - S_7 - S_8 + S_9 - E_3^{(1)} - E_3^{(2)},$$

wobei die $E_3^{(i)}$ die zu den beiden Fixpunkten der Ordnung drei und der ersten Art gehörenden (-3)-Kurven sind. Das geometrische Geschlecht von $Y_-(28)$ ist eins. Bis auf einen konstanten Faktor gibt es also genau eine Γ_{28}-Spitzenform vom Gewicht zwei. Sie sei mit Ω bezeichnet.

Satz 5. *Ω ist eine symmetrische* Γ_{28}-Spitzenform vom Gewicht zwei mit $\Omega(\varepsilon_0(z)) = \Omega(z)$ *und* $\mathfrak{F}(\Omega) = F_3$.

Beweis. Die Einschränkung von Ω auf F_3 liefert $\mathfrak{F}(\Omega) \geqq F_3$. Daraus ergibt sich

$$\mathfrak{F}(T_2(\alpha_2, \Omega)) \geqq F_6, \qquad (\alpha_2 = 9 + \sqrt{7})$$

und wegen $\mathfrak{F}(\Delta_6) = F_6$ sogar $\mathfrak{F}(T_2(\alpha_2, \Omega)) = F_6$, woraus $\mathfrak{F}(\Omega) \leqq F_3$ folgt. Wäre Ω schiefsymmetrisch oder nicht invariant unter $z \mapsto \varepsilon_0 z$, so würde es auf F_6 Nullstellen von Ω geben, was wegen $F_3 \cap F_6 = \emptyset$ nicht möglich ist.

Folgerung. $\operatorname{Div}(\omega^*(\Omega))$ ist ein effektiver und modularer kanonischer Divisor auf $Y_-(28)$.

$\underline{D = 60}$

$\omega = \dfrac{15 + \sqrt{15}}{15} = [[\overline{2,2,2,3,2,2}]]$, $B_{-1} = \sqrt{15}\omega$, $B_0 = \sqrt{15}$. Die zwei Idealklassen in $\mathbb{Q}(\sqrt{60})$ können durch $\mathfrak{o}_{60}$ und $\mathfrak{o}_2$ repräsentiert werden. Dazu gehören zwei Zykel von je zwölf rationalen Kurven S_{ik}, $i \in \{1,2\}$, $k = 0, \ldots, 11$. Es gilt $\operatorname{Div}(\omega^*(\Pi(60))) = 4(F_{15} + F_{14}) + 2F_{11} + 8F_6$

$$\sum_{i=1}^{2} (20(S_{i0} + S_{i6}) + 18(S_{i1} + S_{i5} + S_{i7} + S_{i7} + S_{i11})$$
$$+ 11(S_{i2} + S_{i4} + S_{i8} + S_{i10}) + 4(S_{i3} + S_{i9}))$$

und

$$\operatorname{Div}(\omega^*(\Delta(60)))=4F_{15}+2F_{11}+\sum_{i=1}^{2}(8(S_{i0}+S_{i6})+6(S_{i1}+S_{i5}+S_{i7}+S_{i11})$$
$$+4(S_{i2}+S_{i4}+S_{i8}+S_{i10})\,.$$

In den beiden letzten Beispielen bezeichnen $E^{(i)}$, $i=1,\dots,4$ (-2)-Kurven, die von Fixpunkten der Ordnung zwei herkommen. Es gilt $E^{(i)}\cdot F_i=1$, $i=1,2$ und $E^{(3)}\cdot F_N=E^{(4)}\cdot F_N=1$ mit $N=5$ im Fall $D=41$ und $N=26$ falls $D=113$. E_3 ist eine (-3)-Kurve mit $E_3\cdot F_1=1$.

$\underline{D=41}$

$$\omega=\frac{7+\sqrt{41}}{2}=[[\overline{7,4,2,3,2,2,2,2,3,2,4}]],\ B_{-1}=\varepsilon_0\omega,\ B_0=\varepsilon_0\,.$$

$$\operatorname{Div}(\omega^*(\Pi(41)))=2(F_1+F_2+F_4+F_8+F_{10})-3S_0-2(S_1+S_{10})-2(S_2+S_9)$$
$$+4(S_4+S_7)+6(S_5+S_6)-E_3+E^{(1)}+E^{(2)}+E^{(3)}+E^{(4)}\,,$$
$$\operatorname{Div}(\omega^*(\theta(41)))=8(F_1+F_2)+S_1+S_{10}-(S_3+S_8+S_4+S_7+S_5+S_6)$$
$$+2E_3+4(E^{(1)}+E^{(2)})\,,$$
$$\operatorname{Div}(\omega^*(\theta(41,1)\vartheta(41,2)))=4F_1+2F_2+F_4+E_3+2E^{(1)}+E^{(2)}\,.$$

$\underline{D=113}$

$$\omega=\frac{11+\sqrt{113}}{2}=[[\overline{11,6,2,4,2,2,2,3,\underbrace{2,\dots,2}_{8\times},3,2,2,2,4,2,6}]]\,,$$
$$B_{-1}=\varepsilon_0\omega,\ B_0=\varepsilon_0\,,$$

$$\operatorname{Div}(\omega^*(\theta(113)))=8(F_1+F_2+F_4+F_7)+S_1+S_{22}+6(S_2+S_{21})$$
$$+3(S_3+S_{20})+2(S_4+S_{19})$$
$$+(S_5+S_{18})-(S_7+S_{16}+S_8+S_{15}+S_9+S_{14}+S_{10}+S_{13}+S_{11}+S_{12})$$
$$+2(E_3+E_3^{(2)}+E_3^{(3)})+4(E^{(1)}+E^{(2)})\,.$$
$$\operatorname{Div}(\omega^*(\theta(113,1)\vartheta(113,2))^2)=8F_1+8F_2+4F_4+2F_7$$
$$+2F_8+2F_{14}+S_1+S_{22}+2(S_2+S_{21})$$
$$+(S_3+S_{20})+2(S_4+S_{19})+3(S_5+S_{18})+2(S_6+S_{17})+S_7+S_{16}$$
$$+S_8+S_{15}+S_9+S_{14}+S_{10}+S_{13}+S_{11}+S_{12}+2E_3\,.$$
$$\operatorname{Div}(\omega^*(\vartheta(113,7)))=8F_1+2F_7+F_{14}+F_{28}+F_{49}-(S_1+S_{22})$$
$$-(S_3+S_{20})+S_5+S_{18}+S_6+S_{17}+S_7+S_{16}+3(S_8+S_{15})$$
$$+4(S_9+S_{14})+5(S_{10}+S_{13})+6(S_{11}+S_{12})+2E_3+4E^{(1)}\,.$$

Die Summe der beiden 2-kanonischen modularen Divisoren $\operatorname{Div}(\omega^*(\theta(113)))$ und $\operatorname{Div}(\omega^*(\vartheta(113,7)))$ ist effektiv.

Literatur

1. Franke, H.G.: Kurven in Hilbertschen Modulflächen und Humbertsche Flächen im Siegelraum. Bonn. Math. Schr. **104** (1978)
2. Hausmann, W.: Kurven auf Hilbertschen Modulflächen. Bonn. Math. Schr. **123** (1980)
3. Hermann, C.F.: Thetareihen und modulare Spitzenformen zu den Hilbertschen Modulgruppen reell-quadratischer Körper. Math. Ann. **277**, 327–344 (1987)
4. Hirzebruch, F.: Hilbert modular surfaces. L'Enseignement Math. **19**, 183–281 (1973)
5. Hirzebruch, F., Zagier, D.: Classification of Hilbert modular surfaces. In: Complex analysis and algebraic geometry, pp. 43–77, Iwanami Shoten – Cambridge: University Press 1977
6. Van der Geer, G.: Hilbert modular surfaces. Berlin Heidelberg New York: Springer 1987
7. Zaiger, D.B.: Zetafunktionen und quadratische Körper. Berlin Heidelberg New York: Springer 1981

Eingegangen am 11. November 1987; revidiert am 21. März 1988 und 31. August 1988