Werk

Titel: Monatshefte für Mathematik
Verlag: Springer
Jahr: 1998
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN362162050_0125
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN362162050_0125

Übergeordnetes Werk

Werk Id: PPN362162050
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN362162050

Terms and Conditions

Contact

Niedersächsische Staats- und Universitätsbibliothek Göttingen
Georg-August-Universität Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen
Germany
Email: gdz@sub.uni-goettingen.de

Monatshefte für Mathematik

Vol. 125, 1998

ISSN 0026-9255

SpringerWienNewYork

Inhaltsverzeichnis / Contents

Abstracted/Indexed in:
Current Contents, SCI, ASCA, MATH Database,
ISI/COMPUMATH, Mathematical Reviews,
Zentralblatt für Mathematik/Mathematics Abstracts

Mh. Math. 125, 1–14 (1998)

Monatshefte für
Mathematik

Printed in Austria

Convergence of Continued Fraction Type Algorithms and Generators

By

Cor Kraaikamp, Delft, and **Ronald Meester,** Utrecht

(Received 25 March 1996; in final form 16 August 1996)

Abstract. The concept of convergence of continued fraction type algorithms has been defined a number of times in the literature. We investigate the relation between these definitions, and show that they do not always coincide. We relate the definitions to the question whether or not the natural partition of the underlying dynamical system is a generator. It turns out that the 'right' definition of convergence is equivalent to this partition being a generator. The second definition of convergence is shown to be equivalent only under extra conditions on the transformation. These extra conditions are typically found to be satisfied when the second definition is used in the literature.

1. Motivation

There are many ways to approximate real numbers (vectors) by a sequence of numbers (vectors) arising from an algorithm. Three typical examples of the algorithms we have in mind are the continued fraction expansion, the binary expansion of a real number and the (modified) Jacobi-Perron algorithm. Since we want to look at the question of convergence from a particular point of view, and also since future definitions are motivated by phenomena in these particular examples, we start this paper by looking at these three well-known examples in some detail, although we shall skip proofs at this point.

Example 1. Every irrational number x in the unit interval has a unique continued fraction expansion which we write as $x = [a_1(x), a_2(x), \ldots]$. The finite expansions $[a_1(x), a_2(x), \ldots, a_n(x)]$ are called the *convergents* of x. We can define a dynamical system which generates these expansions as follows (see e.g. BILLINGSLEY 1965). We denote the integer part of x by $[x]$, and the fractional part by $\{x\}$. Define a transformation T: $[0, 1) \to [0, 1)$ by

$$T(x) = \begin{cases} \left\{\frac{1}{x}\right\}, & x \neq 0; \\ 0, & x = 0. \end{cases}$$

The transformation T is many to one, and the continued fraction expansion of x can be retrieved from the orbit of x under T as follows: $a_1(x) = [1/x]$, $a_2(x) = [1/T(x)]$ and in general $a_n(x) = [1/T^{n-1}(x)]$. Another (of course equivalent) way of retrieving the expansion from the orbit proceeds via the definition of the following

1991 Mathematics Subject Classification: 11J70, 11K50, 28D05
Key words: Continued fractions, expansion, generators, convergence

partition of $[0, 1)$: $D_k = \left[\frac{1}{k+1}, \frac{1}{k}\right)$, for $k = 1,2,\ldots$ Now it is easy to check that $a_n(x) = k_0$ if and only if $T^{n-1}(x) \in D_{k_0}$.

Let T_k be the transformation T restricted to D_k. Then T_k maps D_k onto $[0, 1)$ injectively, and for $x \in D_k$ one clearly has

$$T_k(x) = \frac{1 - kx}{x}.$$

With any transformation S of the form

$$S(x) = \frac{a_{10} + a_{11}x}{a_{00} + a_{01}x}$$

we can associate a matrix $A = (a_{ij})$ which we assume is nonsingular, and which is normalised as to satisfy $|\det A| = 1$. The transformations T_k are of this form, and we denote the associated matrix by $A(k)$. The inverse of $A(k)$ is denoted $B(k)$. Looking at the first n iterates of x under T, we see that $T^n(x) = T_{a_n(x)} \cdots T_{a_1(x)}(x)$. One can now check that the matrix associated with T^n is just $A(a_n(x)) \cdots A(a_1(x))$ with inverse

$$B^{(n)}(x) = B(a_1(x)) \cdots B(a_n(x)).$$

It turns out that the matrix $B^{(n)}(x) = (b^{(n)}(x)_{ij})$ contains all information about the n-th convergent of x. More precisely, we have as n-th convergent

$$[a_1(x), a_2(x), \ldots, a_n(x)] = \frac{b^{(n)}(x)_{10}}{b^{(n)}(x)_{00}},$$

and it follows (see e.g. KRAAIKAMP 1991) that

$$x = \lim_{n\to\infty} \frac{b^{(n)}(x)_{10}}{b^{(n)}(x)_{00}}. \tag{1}$$

This fact motivated the definition of convergence of more general algorithms, which we shall give in the next section. For future reference we note at this point that it turns out that for $n \geqslant 2$ we have

$$[a_1(x), a_2(x), \ldots, a_{n-1}(x)] = \frac{b^{(n)}(x)_{11}}{b^{(n)}(x)_{01}},$$

i.e., the 'previous' convergent is given by the appropriate quotient in the second column of $B^{(n)}(x)$.

Example 2. Our second example is the ordinary binary expansion. We consider the transformation $S : [0, 1) \to [0, 1)$ defined by

$$S(x) = \begin{cases} 2x, & x \in [0, \frac{1}{2}), \\ 2x - 1, & x \in [\frac{1}{2}, 1). \end{cases}$$

We write $D_0 = [0, \frac{1}{2})$ and $D_1 = [\frac{1}{2}, 1)$. Writing the binary expansion of a point x as $x_1, x_2, \ldots$, we see that $x_n = 0$ if $S^{n-1}(x) \in D_0$ and $x_n = 1$ if $S^{n-1}(x) \in D_1$. To see the convergence of the binary expansion from the point of view explained above,

we write S_i for the transformation S restricted to $D_i, i = 0, 1$. The matrices $A(0)$ and $A(1)$ corresponding to S_0 and S_1 are

$$\begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & \sqrt{2} \end{pmatrix},$$

respectively. Now if we want to compute the second stage approximation of a number x whose first two digits are 01, we compute, just as above,

$$B^{(2)}(x) := (A(1)A(0))^{-1} = \begin{pmatrix} 2 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

The second convergent to x is then given by

$$\frac{b^{(2)}(x)_{10}}{b^{(2)}(x)_{00}} = \frac{1}{4}.$$

In fact, it is easy to check that

$$B^{(n)}(x) = \begin{pmatrix} (\sqrt{2})^n & 0 \\ r_n(x) & (\sqrt{2})^{-n} \end{pmatrix},$$

for suitable $r_n(x)$. From this it is easy to prove by induction that the n-th convergent $b^{(n)}(x)_{10}/b^{(n)}(x)_{00}$ to x is just the rational number given by the first n binary digits of x. In this case, the quotient $b^{(n)}(x)_{11}/b^{(n)}(x)_{01}$ corresponding to the second column equals ∞.

Example 3. The archetypal example of a multi-dimensional generalisation of the regular continued fraction expansion is the so-called Jacobi-Perron algorithm; see BERNSTEIN (1971), BRENTJES (1981), LAGARIAS (1993) and SCHWEIGER (1973). Here we will briefly discuss the modified Jacobi-Perron algorithm, as introduced in PODSYPANIN (1977); see also ITO *et. al.* (1993). This algorithm yields – and this is a common feature of all such multi-dimensional continued fraction algorithms – rational vectors $\left(\frac{p_n}{q_n}, \frac{r_n}{q_n}\right)$ with the same denominator for simultaneous approximation of any two numbers x and y in the unit interval.

Let E be the unit square. For $(x, y) \in E$ we define the map $T : E \to E$ by

$$T(x,y) = \begin{cases} \left(\frac{y}{x}, \left\{\frac{1}{x}\right\}\right), & \text{if } y \leqslant x,\ x > 0, \\ \left(\left\{\frac{1}{y}\right\}, \frac{x}{y}\right), & \text{if } x < y, \\ (0,0), & \text{if } x = y = 0. \end{cases}$$

Writing $(x_n, y_n) = T^n(x, y), n \geqslant 0$, and defining two sequences of digits $(a_n)_n$ and $(\varepsilon_n)_n$ by

$$(a_n, \varepsilon_n) = \begin{cases} \left(\left[\frac{1}{x_{n-1}}\right], 0\right), & \text{if } y_{n-1} \leqslant x_{n-1},\ x_{n-1} > 0, \\ \left(\left[\frac{1}{y_{n-1}}\right], 1\right), & \text{if } x_{n-1} < y_{n-1}, \end{cases}$$

it is not hard to check that

$$\begin{pmatrix} 1 \\ x \\ y \end{pmatrix} = \theta_n \prod_{k=1}^{n} \begin{pmatrix} a_k & \varepsilon_k & 1-\varepsilon_k \\ 1-\varepsilon_k & 0 & \varepsilon_k \\ \varepsilon_k & 1-\varepsilon_k & 0 \end{pmatrix} \begin{pmatrix} 1 \\ x_n \\ y_n \end{pmatrix} = \theta_n B^{(n)} \begin{pmatrix} 1 \\ x_n \\ y_n \end{pmatrix},$$

where

$$\theta_n = \prod_{k=0}^{n-1} \max(x_k, y_k) \in [0, 1].$$

We can define matrices $B(k)$ as before, and denote the suitable product $B^{(n)} = B^{(n)}(x, y)$ by

$$B^{(n)} = \begin{pmatrix} q_n & q_n' & q_n'' \\ p_n & p_n' & p_n'' \\ r_n & r_n' & r_n'' \end{pmatrix}.$$

We therefore have

$$x = \frac{p_n + x_n p_n' + y_n p_n''}{q_n + x_n q_n' + y_n q_n''}, \quad y = \frac{r_n + x_n r_n' + y_n r_n''}{q_n + x_n q_n' + y_n q_n''}.$$

Replacing (x_n, y_n) by $(0, 0)$ yields the rational approximations p_n/q_n and r_n/q_n to x resp. y.

Among other things, PODSYPANIN (1977) shows that if $(x, y) \in E$ and at least one of x and y is irrational, then

$$\lim_{n\to\infty} \frac{p_n}{q_n} = x, \quad \lim_{n\to\infty} \frac{r_n}{q_n} = y.$$

As far as the other columns are concerned, if we write $(q_{-2}, p_{-2}, r_{-2})^t = (0, 0, 1)$, $(q_{-1}, p_{-1}, r_{-1})^t = (0, 1, 0)$ and $(q_0, p_0, r_0)^t = (1, 0, 0)$, then Podsypanin shows that all columns of $B^{(n)}$ are of the form $(q_m, p_m, r_m)^t$ for some $m \geqslant -2$. As the referee pointed out, this fact does not immediately imply that the quotients p_n'/q_n', r_n'/q_n', p_n''/q_n'' and r_n''/q_n'' also converge to the appropriate point. However, from SCHWEIGER (1978) we know that for some (explicit) T-invariant measure ρ, which is equivalent to Lebesgue measure, (E, ρ, T) forms an ergodic system. It follows then from the ergodic theorem that for almost all $(x, y) \in E, \varepsilon_n = \varepsilon_{n+1}$ infinitely often. From Podsypanin's Lemma 1 it now follows that for ρ almost all (and therefore also for Lebesgue almost all) points in E, all ratios p_n/q_n, p_n'/q_n' and p_n''/q_n'' converge to the same limit, and similarly for the other ratios. □

The convergence phenomena described in these examples are the basis for the classical definition of a convergent algorithm. In fact, at least three different definitions appear in the literature. The first definition only requires ratio's as in (1) to converge, without specifying the limit; see for example SCHWEIGER (1991). The second definition requires in addition that these ratio's converge to the 'right' point; see for instance SCHWEIGER (1973), LAGARIAS (1993), and BERNSTEIN (1971).

The third definition is also called topological convergence of the algorithm, and comes down to requiring that the natural partition associated with the transformation is a generator from the ergodic-theoretical point of view. (For definitions see the next section.) This definition is mentioned (among others) in SCHWEIGER (1996).

The purpose of this paper is to try to understand and clarify the relation between these different definitions. In the next section we give the general setup, all necessary definitions and our results. The proofs are given in the last section.

2. General Setup and Results

In this section we present the general setup. For more details on what follows, see e.g. SCHWEIGER (1991). Let $D \subset \mathbf{R}^d$ be compact, containing the origin and convex. We write μ for Lebesgue measure. We assume that D admits a countable partition $D = U_k D(k)$, each of whose atoms is assumed to satisfy $\mu(\partial D(k)) = 0$, where ∂ denotes boundary. In almost all examples, the $D(k)$'s will be suitable polygons.

Let $T : D \to D$ be a transformation with the following properties:

(i) $T(D(k)) = D$ for all k,

(ii) T restricted to $D(k)$ is written T_k, is injective and given via

$$T_k(x_1, \ldots, x_d) = (y_1, \ldots, y_d),$$

where

$$y_i = \frac{a_{i0} + a_{i1}x_1 + \cdots + a_{id}x_d}{a_{00} + a_{01}x_1 + \cdots + a_{0d}x_d}, \quad i = 1, \ldots, d, \tag{2}$$

(iii) The matrix $A = A(k) = (a_{ij})$ is invertible, and we can assume that $|\det A| = 1$.

The map T_k has an inverse $V_k : D \to D(k)$ given by $V_k(y_1, \ldots, y_d) = (x_1, \ldots, x_d)$, where

$$x_i = \frac{b_{i0} + b_{i1}y_1 + \cdots + b_{id}y_d}{b_{00} + b_{01}y_1 + \cdots + b_{0d}y_d}, \quad i = 1, \ldots, d,$$

and $B = B(k) = (b_{ij})$ is the inverse matrix of A. Furthermore, one can check that if $y = T^n x$, and $T^m x \in D(k_{m+1})$, for $0 \leqslant m < n$, then

$$x_i = \frac{b_{i0}^{(n)} + b_{i1}^{(n)}y_1 + \cdots + b_{id}^{(n)}y_d}{b_{00}^{(n)} + b_{01}^{(n)}y_1 + \cdots + b_{0d}^{(n)}y_d}, \quad i = 1, \ldots, d,$$

where $B^{(n)} = B^{(n)}(x) = (b_{ij}^{(n)}) = B(k_1)B(k_2)\cdots B(k_n)$. We denote the set $\{x \in D;\ T^m x \in D(k_{m+1}),\ 0 \leqslant m < n\}$ by $D(k_1, \ldots, k_n)$. For x in this set, the inverse of T^n is denoted by $V^{(n)} = V_{(k_1,\ldots,k_n)}$, and we sometimes write $D(k_1, \ldots, k_n) = D^{(n)}(x)$. Note that $D^{(n)}(x)$ is convex since T maps straight lines onto straight lines. Also note that $T^n(D^{(n)}(x)) = D$ for all n. Finally, the *name* of x is the appropriate sequence $(k_1, k_2, \ldots)$, i.e. the name of x is the enumeration of the atoms of the underlying partition visited by the orbit of x. The name of x is almost surely infinite.

Remarks 1. It is not hard to see that the determinant of the Jacobian of the transformation in (2) is equal to

$$JT(x) = \frac{\det(A)}{(a_{00} + a_{01}x_1 + \cdots + a_{0d}x_d)^{d+1}}.$$

2. If the linear transformation in $\mathbf{R}^{d+1}$ corresponding to $A(k)$ maps $(1, x_1, x_2, \ldots, x_d)$ onto $(y_0, y_1, \ldots, y_d)$ for $(x_1, \ldots, x_d) \in D(k)$, then we see that

$$T_k(x_1, \ldots, x_d) = \left(\frac{y_1}{y_0}, \ldots, \frac{y_d}{y_0}\right).$$

A similar statement holds for compositions of T.

In order to define the convergence of an algorithm, we need of course to get rid of those points whose name is finite. One could either look only at infinite names (see SCHWEIGER 1991, BERNSTEIN 1971, PODSYPANIN 1977, and LAGARIAS 1993), or allow an exceptional set of measure zero. The latter is the approach followed in this paper. Therefore we have the following definitions:

Definition 1.

$$C = \left\{ x \in D; \lim_{n\to\infty} \left(\frac{b^{(n)}(x)_{10}}{b^{(n)}(x)_{00}}, \ldots, \frac{b^{(n)}(x)_{d0}}{b^{(n)}(x)_{00}} \right) \quad \text{exists} \right\};$$

$$F = \left\{ x \in D; \lim_{n\to\infty} \left(\frac{b^{(n)}(x)_{10}}{b^{(n)}(x)_{00}}, \ldots, \frac{b^{(n)}(x)_{d0}}{b^{(n)}(x)_{00}} \right) = x \right\}.$$

In this paper we want to relate these sets to a well known concept from ergodic theory.

Definition 2. A finite or countable partition of D is called *a generator* with respect to T it for all x outside a set of (Lebesgue) measure zero we have

$$\bigcap_{n=1}^{\infty} D^{(n)}(x) = \{x\}.$$

In words, no two points outside some exceptional set of measure zero have orbits which visit the same atoms of the partition in the same order. This means that we can distinguish between points by just looking at which atoms of the partition are visited after consecutive applications of the transformation T.

Next consider the following three statements, writing μ for d-dimensional normalised Lebesgue measure, so that $\mu(D) = 1$.

(A) $\mu(C) = 1$;
(B) $\mu(F) = 1$;
(C) The partition $D(k)$ is a generator.

As mentioned previously, either (A) and (B) are sometimes used as the definition of convergence of an algorithm. Obviously, (B) is the 'right' definition since one wants to know which number is approximated by the algorithm. It is not

true in general though that (A) and (B) are equivalent, as can be seen from our main results which we state now. For this we need another definition.

Definition 3. The transformation T is said to be *expanding* if

$$|JT(x)| \geqslant 1$$

for all $x \in D$.

Theorem 1. *(i) Statements (B) and (C) are equivalent.*

(ii) *Suppose $d = 1$ and T is expanding. Then all three statements (A), (B) and (C) are equivalent.*

(iii) *If T is not expanding or $d \geqslant 2$, then there is an example such that (A) is strictly weaker than (B).*

Although in higher dimensions the situation is not as nice as in the case $d = 1$, we can still provide sufficient conditions for convergence to the 'right' point to take place. It turns out that convergence of appropriate quotients in other than the first column is important here.

Definition 4.

$$C^* = \left\{ \begin{array}{l} x \in D; \lim_{n\to\infty} \left(\dfrac{b^{(n)}(x)_{1i}}{b^{(n)}(x)_{0i}}, \ldots, \dfrac{b^{(n)}(x)_{di}}{b^{(n)}(x)_{0i}} \right) \text{ exists and is} \\ \text{independent of } i \end{array} \right\}.$$

Theorem 2. Suppose that $\mu(C^*) = 1$ and that T is expanding. Then $\mu(F) = 1$.

Remark. It might seem that problems arise in the literature whenever the 'wrong' definition of convergence is used. However, to our knowledge, in all places where this wrong definition is used, the extra requirements mentioned in Theorem 1(iii) or Theorem 2 are satisfied. To illustrate Theorem 2, note that it follows from the last paragraph of Example 3 that $\mu(C^*) = 1$ in that case.

3. Proofs

We start with some notation used throughout this section. When $d = 1$, we write

$$B^{(n)}(x) = \begin{pmatrix} q_n & q'_n \\ p_n & p'_n \end{pmatrix},$$

and when $d = 2$ we write

$$B^{(n)}(x) = \begin{pmatrix} q_n & q'_n & q''_n \\ p_n & p'_n & p''_n \\ r_n & r'_n & r''_n \end{pmatrix}.$$

Note that of course $q_n = q_n(x)$ and similarly for the other quantities. (Typically we supress the dependence on x whenever possible.) We write all proofs as if $d = 2$, except of course when $d = 1$ is required. We start with a geometrical lemma needed in the proof of Theorem 1(i).

Lemma 1. *If the partition* $\{D(k)\}$ *forms a generator, then for almost all* $x \in D$ *we have that*

$$\bigcap_{n=1}^{\infty} \overline{D^{(n)}(x)} = \{x\}, \tag{3}$$

where $\bar{A}$ *denotes the closure of* A.

Proof. Suppose that $y \in \bigcap_{n=1}^{\infty} \overline{D^{(n)}(x)}$ and suppose that $y \neq x$. Choose any point $z \neq x, y$ on the straight-line segment between x and y. Note that $z \in \bigcap_{n=1}^{\infty} \overline{D^{(n)}(x)}$ since intersections of convex sets are convex, and closures of convex sets are also convex, whence $\overline{D^{(n)}(x)}$ is a convex set for all n.

If $z \in D^{(n)}(x)$ for all n, then $z = x$ since $\{D(k)\}$ forms a generator, and this is a contradiction (we assume that x is not in the exceptional set in Definition 2).

If $z \notin D^{(N_0)}(x)$ for some N_0, then it must be the case that $z \in \partial D^{(N_0)}(x)$. We claim that it now follows that

$$x \in \partial D^{(N_0)}(x). \tag{4}$$

This is enough since this means that all points x for which (3) fails are contained in the union of the boundaries of all atoms, a set of measure zero by assumption.

It remains to prove (4). This is easy though: the points x, y are contained in $\overline{D^{(N_0)}(x)}$, and if x is not on its boundary, then we can form a little ball around x which is completely contained in the interior of $D^{(N_0)}(x)$. By convexity, all line segments between points in this ball and y are contained in $D^{(N_0)}(x)$, which implies that the intermediate point z is in the interior of $D^{(N_0)}(x)$, a contradiction.

Proof of Theorem 1(i). We first show that (B) implies (C). If two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ have the same name, then for all n we have

$$\frac{p_n(x)}{q_n(x)} = \frac{p_n(y)}{q_n(y)} \quad \text{and} \quad \frac{r_n(x)}{q_n(x)} = \frac{r_n(y)}{q_n(y)}.$$

Assumption (B) implies that

$$\lim_{n\to\infty} \frac{p_n(x)}{q_n(x)} = x_1 \quad \text{and} \quad \lim_{n\to\infty} \frac{p_n(y)}{q_n(y)} = y_1,$$

giving that $x_1 = y_1$. Similarly, we have $x_2 = y_2$ and we are done.

For the reverse implication, we fix $x \in D$, and consider $p_n = p_n(x)$ etc. Now consider an arbitrary point $c = (c_1, c_2) \in D$ and note that it follows from Remark 2 in Section 2 that

$$V^{(n)}(c) = \left(\frac{p_n + c_1 p_n' + c_2 p_n''}{q_n + c_1 q_n' + c_2 q_n''}, \frac{r_n + c_1 r_n' + c_2 r_n''}{q_n + c_1 q_n' + c_2 q_n''} \right). \tag{5}$$

The partition $D(k)$ generates, and it follows from Lemma 1 that for almost all x, $\bigcap_n \overline{D^{(n)}(x)} = \{x\}$, and we can assume that the fixed x is not in the exceptional set. Since the sets $\overline{D^{(n)}(x)}$ are compact and connected (this follows from convexity), it follows that their diameters converges to zero when $n \to \infty$. But $V^{(n)}(c)$ has the same first n digits as x, and it follows that

$$\lim_{n\to\infty} V^{(n)}(c) = x. \tag{6}$$

Note that this is true for all $c \in D$. Next we choose three points u, v, w in D such that there exist constants α_i so that

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ u_1 \\ u_2 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ v_1 \\ v_2 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ w_1 \\ w_2 \end{pmatrix}.$$

(This is always possible as long as D has positive Lebesgue measure. The elementary proof of this fact is left to the reader.)

We can now write

$$\begin{aligned} \begin{pmatrix} q_n \\ p_n \\ r_n \end{pmatrix} &= B^{(n)} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \alpha_1 \begin{pmatrix} q_n + u_1 q_n' + u_2 q_n'' \\ p_n + u_1 p_n' + u_2 p_n'' \\ r_n + u_1 r_n' + u_2 r_n'' \end{pmatrix} + \alpha_2 \begin{pmatrix} q_n + v_1 q_n' + v_2 q_n'' \\ p_n + v_1 p_n' + v_2 p_n'' \\ r_n + v_1 r_n' + v_2 r_n'' \end{pmatrix} \\ &\quad + \alpha_2 \begin{pmatrix} q_n + w_1 q_n' + w_2 q_n'' \\ p_n + w_1 p_n' + w_2 p_n'' \\ r_n + w_1 r_n' + w_2 r_n'' \end{pmatrix}. \end{aligned}$$

From this, together with (5), (6) and the fact that the α_i's do not depend on n, it follows that

$$\frac{p_n(x)}{q_n(x)} \to x_1 \quad \text{and} \quad \frac{r_n(x)}{q_n(x)} \to x_2. \qquad \square$$

For the proof of Theorem 1(ii) we need the following lemma.

Lemma 2. *Suppose T is expanding. Then for all names $a = (a_1, a_2, \ldots)$ it is the case that*

$$A_a := \{x \in D; \text{the name of } x \text{ is } a\}$$

satisfies $\mu(A_a) = 0$.

Proof. Suppose that $\mu(A_a) > 0$. First we assume that a is aperiodic. This implies that $T^i(A_a) \cap T^j(A_a) = \emptyset$ for all $i \neq j$. Using the fact that T is expanding, we see that for all i

$$\mu(T^{i+1}(A_a)) = \int_{T^i(A_a)} |JT| d\mu \geqslant \mu(T^i(A_a)),$$

which is impossible when $\mu(A_a) > 0$.

If a is periodic with period k we have that $A_a = T^k(A_a)$. It follows that $\mu(A) = \mu(T^i(A_a))$ for all i. On the other hand we have

$$\mu(T(A_a)) = \int_{A_a} |JT| d\mu,$$

and hence

$$\int_{A_a} (|JT| - 1)d\mu = 0.$$

Since $|JT(x)| \geqslant 1$ this implies that $|JT(x)| = 1$ on A_a. This is impossible if $\mu(A_a) > 0$ according to the formula for $JT(x)$ in Remark 1 in Section 2.

Finally, we also have to deal with the case in which a is eventually periodic. In this case there exists an index k for which $a^* = (a_{k+1}, a_{k+2}, \ldots)$ is a periodic word. Notice that $\mu(A_a^*) = \mu(T^k(A_a)) \geqslant \mu(A_a) > 0$ which is impossible by the previous case.

Proof of Theorem 1(ii). We need to show that (A) implies (B). Fix $x \in D$, and let $V^{(n)}$ be the inverse operator corresponding to x. Since T is expanding, we have from Lemma 2 that $\mu(A_a) = 0$ for all names a. In particular, the set $V^{(n)}(D)$, which is the set of points whose first n digits agree with the name of x, must satisfy

$$\mu(V^{(n)}(D)) \to 0,$$

when $n \to \infty$. This comes down to

$$\lim_{n\to\infty} \int_D \left| \frac{1}{(q_n + yq_n')^2} \right| d\mu(y) = 0,$$

which yields, using Fatou's lemma, that

$$\int_D \liminf_{n\to\infty} \left| \frac{1}{(q_n + yq_n')^2} \right| d\mu(y) = 0.$$

This then finally gives that $\limsup_{n\to\infty} |q_n + yq_n'| = \infty$ for almost all $y \in D$, which allows us to choose $c \in D$ and a subsequence $n_1, n_2 \ldots$ so that

$$\lim_{k\to\infty} |q_{n_k} + cq_{n_k}'| = \infty. \tag{7}$$

Next we consider the parallellogram $\mathscr{P}_n$ in the plane spanned by the vectors $(q_n(x), p_n(x))$ and $(cq_n'(x), cp_n'(x))$. The requirement on the determinant of the matrices $B^{(n)}$ tells us that the volume of $\mathscr{P}_n$ does not depend on n and is equal to $|c|$. This, together with (7) and the assumption that

$$\lim_{n\to\infty} \frac{p_n}{q_n}$$

exists and is equal to ξ, say, implies that

$$\lim_{k\to\infty} \frac{p_{n_k} + cp'_{n_k}}{q_{n_k} + cq'_{n_k}} = \xi.$$

To see this properly, note that the line through the origin and the vertex (q_{n_k}, p_{n_k}) of $\mathscr{P}_{n_k}$ converges to a fixed line through the origin. In addition the vertex $(q_{n_k} + cq'_{n_k}, p_{n_k} + cp'_{n_k})$ goes to infinity according to (7). Finally, since the absolute value of the determinant of the Jacobian of the inverse map $V^{(n)}$ evaluated in 0 equals $1/q_n^2$, we have by the expandingness of T that $|q_n| \geqslant 1$; therefore the points (q_{n_k}, P_{n_k}) are

bounded away from the origin. These three facts imply the result, but only in two dimensions; in higher dimensions, this is the point where the proof breaks down.

Let $R := \sup\{y;\, y \in D\}$. Since the parallellogram spanned by (q_n, p_n) and (Rq'_n, Rp'_n) also has fixed volume independent of n, and is obtained from $\mathscr{P}_n$ by multiplying the appropriate sides by R/c, we also have that

$$\lim_{k\to\infty} \frac{p_{n_k} + Rp'_{n_k}}{q_{n_k} + Rq'_{n_k}} = \xi. \tag{8}$$

For $r := \inf\{y;\, y \in D\}$, a similar argument shows that

$$\lim_{k\to\infty} \frac{p_{n_k} + rp'_{n_k}}{q_{n_k} + rq'_{n_k}} = \xi. \tag{9}$$

The point $(q_{n_k} + T^{n_k}(x)q'_{n_k},\ p_{n_k} + T^{n_k}(x)p'_{n_k})$ lies on the straight line segment between $(q_{n_k} + rq'_{n_k},\ p_{n_k} + rp'_{n_k})$ and $(q_{n_k} + Rq'_{n_k},\ p_{n_k} + Rp'_{n_k})$. Hence it follows from (8) and (9) that

$$\lim_{k\to\infty} \frac{p_{n_k} + T^{n_k}(x)p'_{n_k}}{q_{n_k} + T^{n_k}(x)q'_{n_k}} = \xi.$$

But according to Remark 2 in Section 2 we have

$$\frac{p_{n_k} + T^{n_k}(x)p'_{n_k}}{q_{n_k} + T^{n_k}(x)q'_{n_k}} = V^{(n_k)}T^{n_k}(x) = x,$$

and it follows that $\xi = x$, as required. □

Proof of Theorem 1(iii). First we give an example in $d = 1$ without the expanding property, and for which $\mu(C) = 1$ but $\mu(F) < 1$.

Let $D = [0, 2)$ and let $D(1) = [0, 1), D(2) = [1, 2)$. We define T as follows: on $D(1)$ we have

$$T(x) = \frac{\frac{1}{2}x}{2 - \frac{7}{4}x},$$

and on $D(2)$ we have

$$T(x) = -2 + 2x.$$

The corresponding inverse matrices $B(1)$ and $B(2)$ are

$$\begin{pmatrix} 1/2 & 7/4 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \sqrt{2} & 0 \\ \sqrt{2} & 1/\sqrt{2} \end{pmatrix},$$

respectively. It is not hard to see that $[0, \frac{6}{7}]$ is an absorbing set, so that almost all points eventually end up in this set. But notice that when we apply the matrix $B(1)$, the ratio p_n/q_n does not change. It follows that

$$\lim_{n\to\infty} \frac{p_n(x)}{q_n(x)}$$

exists almost surely. On the other hand, $D(1)$ and $D(2)$ do not form a generator, so $\mu(F) < 1$ and we are done.

Next we give an example of a two-dimensional algorithm which is expanding, for which $\mu(C) = 1$ but $\mu(F) < 1$. Let $D(1) = \{(x, y); x \geqslant 0, x + y \leqslant 1, y \geqslant 2x\}$, $D(2) = \{(x, y); x \leqslant y \leqslant 2x, x + y \leqslant 1\}$, and $D = D(1) \cup D(2)$. Define $T : D \to D$ by

$$T(x, y) = \begin{cases} \left(\frac{x}{1-x}, \frac{y-x}{1-x}\right), & \text{on } D(1), \\ \left(\frac{y-x}{1-x}, \frac{x}{1-x}\right), & \text{on } D(2). \end{cases}$$

It is easy to verify by direct calculations that T is expanding. On the other hand, we also see by a simple calculation that both $B(1)$ and $B(2)$ have as first column $(1, 0, 0)^t$ which clearly implies that $\mu(C) = 1$. Finally, straight lines through the origin are mapped onto straight lines through the origin, and from this it follows immediately that $D(1)$ and $D(2)$ cannot form a generator. □

Proof of Theorem 2. We will only consider the case $d = 2$; as before the case $d \geqslant 3$ is proved similarly. The proof of Theorem 1(ii) will be followed closely, the new ingredient being stronger assumptions on the behaviour of the columns of $B(k)$.

We choose some fixed $x \in D$, giving quantities like q_n, $V^{(n)}$ etc. Again due to the fact that T is expanding we have for almost all $x \in D$

$$\int_D \liminf_{n\to\infty} |JV^{(n)}(y_1, y_2)| \, d\mu(y_1, y_2) = 0,$$

from which (using Remark 1 in Section 2)

$$\limsup_{n\to\infty} |q_n + y_1 q_n' + y_2 q_n''| = \infty \quad \text{almost surely.}$$

Now choose $(c_1, c_2) \in D$ and a subsequence $n_1, n_2, \ldots$ such that

$$\lim_{k\to\infty} |q_{n_k} + c_1 q_{n_k}' + c_2 q_{n_k}''| = \infty.$$

The parallellopipedum spanned by $(q_n, p_n, r_n), c_1(q_n', p_n', r_n')$ and $c_2(q_n'', p_n'', r_n'')$ has fixed volume $|c_1 c_2|$, independent of n. By assumption one has that for some suitable ξ_1 and ξ_2,

$$\lim_{k\to\infty} \frac{p_{n_k}}{q_{n_k}} = \lim_{k\to\infty} \frac{p_{n_k}'}{q_{n_k}'} = \lim_{k\to\infty} \frac{p_{n_k}''}{q_{n_k}''} = \xi_1$$

and

$$\lim_{k\to\infty} \frac{r_{n_k}}{q_{n_k}} = \lim_{k\to\infty} \frac{r_{n_k}'}{q_{n_k}'} = \lim_{k\to\infty} \frac{r_{n_k}''}{q_{n_k}''} = \xi_2,$$

and therefore

$$\lim_{k\to\infty} \frac{p_{n_k} + c_1 p_{n_k}' + c_2 p_{n_k}''}{q_{n_k} + c_1 q_{n_k}' + c_2 q_{n_k}''} = \xi_1 \text{ and } \lim_{k\to\infty} \frac{r_{n_k} + c_1 r_{n_k}' + c_2 r_{n_k}''}{q_{n_k} + c_1 q_{n_k}' + c_2 q_{n_k}''} = \xi_2. \tag{10}$$

Define $r_1 = \inf\{y_1; (y_1, y_2) \in D\}, R_1 = \sup\{y_1; (y_1, y_2) \in D\}$ and similarly for r_2 and R_2. For $c_i \in \{r_i, R_i\}$, $i = 1, 2$, (c_1, c_2) is not necessarily a point in D, but clearly (10) still holds.

The point with coordinates

$$q_{n_k} + (T^{n_k}(x))_1 q'_{n_k} + (T^{n_k}(x))_2 q''_{n_k},\ p_{n_k} + (T^{n_k}(x))_1 p'_{n_k} + (T^{n_k}(x))_2 p''_{n_k}$$

and $r_{n_k} + (T^{n_k}(x))_1 r'_{n_k} + (T^{n_k}(x))_2 r''_{n_k}$ lies in the quadrangle with vertices

$$(q_{n_k} + t_1 q'_{n_k} + t_2 q''_{n_k},\ p_{n_k} + t_1 p'_{n_k} + t_2 p''_{n_k},\ r_{n_k} + t_1 r'_{n_k} + t_2 r''_{n_k}),$$

where (t_1, t_2) is any of the four mentioned possibilities $(r_1, r_2), (r_1, R_2)$, $(R_1, r_2), (R_1, R_2)$. But then (10) also holds for $(c_1, c_2) = T^{n_k}(x)$, i.e.

$$\lim_{k\to\infty} \frac{p_{n_k} + (T^{n_k}(x))_1 p'_{n_k} + (T^{n_k}(x))_2 p''_{n_k}}{q_{n_k} + (T^{n_k}(x))_1 q'_{n_k} + (T^{n_k}(x))_2 q''_{n_k}} = \xi_1 = \lim_{n\to\infty} \frac{p_n}{q_n}$$

and

$$\lim_{k\to\infty} \frac{r_{n_k} + (T^{n_k}(x))_1 r'_{n_k} + (T^{n_k}(x))_2 r''_{n_k}}{q_{n_k} + (T^{n_k}(x))_1 q'_{n_k} + (T^{n_k}(x))_2 q''_{n_k}} = \xi_2 = \lim_{n\to\infty} \frac{r_n}{q_n}.$$

As in the proof of Theorem 1(ii) it now follows that $x = (\xi_1, \xi_2)$. □

Final Remarks 1. Clearly, as is shown by the counterexample from the proof of Theorem 1(iii), the fact that T is expanding is not enough to guarantee that for $d \geqslant 2$ the columns of the matrices $B(k)$ all converge, and converge to the same point in D. (The i-th column of $B(k)$ converges to $x \in D$ if $\left(\frac{b^{(n)}(x)_{1i}}{b^{(n)}(x)_{0i}}, \ldots, \frac{b^{(n)}(x)_{di}}{b^{(n)}(x)_{0i}}\right) \to x$.) We need this to obtain (10) above. This is a sufficient condition, but certainly not a necessary one, as the second example from Section 1 showed; there the first column converges to the 'right' point, while the second column yields a convergent which always equals ∞.

2. There are multi-dimensional continued fraction algorithms (the most important example being the Jacobi-Perron algorithm) which do not satisfy the condition $T(D(k)) = D$ for all k. In Yuri (1986), this condition is replaced by the requirement that there is a finite collection $\{U_0, \ldots, U_N\}$ of subsets of D, all with positive Lebesgue measure, and with the property that $T^n(D(k_1, \ldots, k_n)) \in \{U_0, \ldots, U_N\}$ for all finite names $(k_1, \ldots, k_n)$. A map T satisfying this condition is called a multidimensional mapping with *finite range structure*. It is not hard to check that for such maps, all results in this paper still go through if we assume that all U_i's are convex, and we leave this to the reader.

Acknowledgement. We thank Fritz Schweiger for pointing out a mistake in an earlier version of Theorem 1(i) and for helping us with the current proof.

References

[1] Bernstein L (1971) The Jacobi-Perron Algorithm. Its Theory and Application. Lect Notes Math **207**: New York: Springer

[2] Billingsley P (1965) Ergodic Theory and Information, New York: Wiley

[3] Brentjes AJ (1981) Multi-dimensional Continued Fraction Algorithms. Amsterdam: Math Centre

[4] Ito S, Keane MS, Ohtsuki M (1993) Almost everywhere exponential convergence of the modified Jacobi-Perron algorithm. Ergodic Th and Dyn Sys **13**, 319–334

[5] KRAAIKAMP C (1991) A new class of continued fractions. Acta Arithmetica **LVII**, 1–39
[6] LAGARIAS JC (1993) The quality of the diophantine approximations found by the Jacobi-Perron algorithm and related algorithms, Mh Math **115**: 299–328
[7] PODSYPANIN EV (1977) A generalization of the algorithm for continued fractions related to the algorithm of Viggo Brunn. Studies in Number Theory (LOMI), 4. Zap Naucn Sem Leningrad Otdel Mat Inst Steklov **67**: 184–194. English translation: J Soviet Math **16**: 885–893 (1981)
[8] SCHWEIGER F (1973) The metrical theory of the Jacobi-Perron algorithm. Lect Notes Math **334**. Berlin Heidelberg New York: Springer
[9] SCHWEIGER F (1978) A modified Jacobi-Perron algorithm with explicitly given invariant measure. Lect Notes Math **729**. Berlin-Heidelberg New York: Springer
[10] SCHWEIGER F (1991) Invariant measures for maps of continued fraction type. J Number Theory **39**: 162–174
[11] SCHWEIGER F (1996) Multidimensional continued fractions. Preliminary version
[12] YURI M (1986) On a Bernoulli property for multidimensional mapping with finite range structure. Tokyo J Math **9**: 457–485

COR KRAAIKAMP
Department of Mathematics
Delft University of Technology
Mekelweg 4
2628 CD Delft
The Netherlands
e-mail: cork@twi.tudelft.nl

RONALD MEESTER
Department of Mathematics
University of Utrecht
P.O. Box 80.010
3508 TA Utrecht
The Netherlands
e-mail: meester@math.ruu.nl

Mh. Math. 125, 15–24 (1998)

Monatshefte für
Mathematik

Printed in Austria

On the Value Distribution of a Certain Type of Differential Polynomials

By

Ping Li, Hefei, and **Chung-Chau Yang***, Hong Kong

(Received 12 December 1995; in revised form 10 March 1997)

Abstract. Some quantitative estimations on the value distribution of the function $afg^n (n \geqslant 2)$ are obtained, where g is a linear differential polynomial in f and a is a small function of f.

1. Introduction

Let f be a transcendental meromorphic function. A complex value a is said to be a Picard value of f, if and only if, $f(z) - a$ has at most finitely many zeros. In 1967, W. K. HAYMAN [6] conjectured that the only possible Picard value of $f^n f'$ is zero. He [7] proved the case $n \geqslant 3$ in 1959, and CLUNIE [4] proved the case $n = 1$ for entire functions in 1962. Many studies related to Hayman's conjecture have been published, see [1]–[5], [10]–[15] and [19]. In [1], W. BERGWEILER and A. EREMENKO confirmed the conjecture for meromorphic functions of finite order. Very shortly after this, H. H. CHEN and M. L. FANG [2] were able to resolve the conjecture completely by utilizing the normal-family technique and the finit-order case. This conjecture was also resolved by L. ZALCMAN and A. ERENMENKO independently, using essentially the same normal family argument to reduce the infinite-order case to the finite-order case. Further, it was conjectured by C. C. YANG that for any transcendental meromorphic function f, $ff^{(k)} (k \geqslant 1)$ can only have zero as its Picard value. Relating to the conjecture, the following results were obtained.

Theorem A. [17] *Let f be a transcendental entire function and n, k be non-negative integers with $n \geqslant 2$. Then the only possible Picard value of $(f^{(k)})^n f$ is the value zero.*

Theorem B. [14] *Let f be a transcendental meromorphic function, and n, k be positive integers with $n > 9e + 1$. Then the only possible Picard value of $(f^{(k)})^n f$ is the value zero.*

Recently, G. D. SONG and Z. F. ZHANG [12] improved the above two theorems and obtained the following result.

1991 Mathematics Subject Classification: 30D35
Key words: Meromorphic functions, differential polynomial, small functions, Picard value
* The research was partially supported by a U.G.C grant of Hong Kong

Theorem C. *Let f be a transcendental meromorphic function, and n, k be positive integers with $n \geqslant 2$. Then $af(f^{(k)})^n$ assumes all nonzero finite complex values infinitely often, where $a \neq 0$ is a small function of f.*

In this paper, we have generalised the above results by obtaining some quantitative estimations on the zeros of $afg^n - 1$ with $n \geqslant 2$, where $g = \sum_{i=0}^{k} a_i f^{(i)}$, and $a, a_i, i = 0, 1, \ldots, k$ $(aa_k \not\equiv 0)$ are small functions of f

It is assumed that the reader is familiar with the basic theory of Nevanlinna value distribution and its standard symbols and notations such as $T(r,f)$, $N(r,f)$, $\bar{N}(r,f)$, $m(r,f)$, and so on; see, eg. [8]. Particularly, $S(r,f)$ will be used to denote any quantity that satisfies $S(r,f) = o(T(r,f))$ as $r \to \infty$ and $r \notin E$ with E being a set of $r \in (0, \infty)$ of finite linear measure. As usual, a meromorphic function α is said to be a small function of f if $T(r,f) = S(r,f)$.

2. Main Results and Lemmas

Lemma 1. [16] *Let f be a meromorphic function satisfying $f^{(k)} \not\equiv 0$. Then*

$$N\left(r, \frac{1}{f^{(k)}}\right) \leqslant N\left(r, \frac{1}{f}\right) + k\bar{N}(r,f) + S(r,f).$$

Lemma 2. [18] *Let f_1 and f_2 be two non-zero meromorphic functions. Then*

$$N(r, f_1 f_2) - N\left(r, \frac{1}{f_1 f_2}\right) = N(r, f_1) + N(r, f_2) - N\left(r, \frac{1}{f_1}\right) - N\left(r, \frac{1}{f_2}\right).$$

Theorem 1. *Let f be a transcendental meromorphic function and $n \geqslant 2, k \geqslant 1$ be integers. Let $\psi = afg^n - 1$, where $g = \sum_{i=0}^{k} a_i f^{(i)} \not\equiv 0$, and $a, a_i, i = 0, 1, \ldots, k$, $(aa_0 \not\equiv 0)$ are small functions of f. Furthermore, let $h = -\frac{a'}{a} - n\frac{g'}{g}$,*

$$D_0 = 1, D_i = D'_{i-1} + hD_{i-1}, \quad i = 1, 2, \cdots, k, \quad \text{and} \quad D = \sum_{i=0}^{k} a_i D_i.$$

If $Dg' - Dg' - hDg \not\equiv 0$, then

$$T(r,f) \leqslant C_{n,k} N\left(r, \frac{1}{\psi}\right) + S(r,f),$$

where $C_{n,k} = \dfrac{(2nk + 2n + 1)(3n + k - 4)}{n^2 k - 2nk + n^2 - 3n + 3}$.

Theorem 2. *Let ψ, g, a, a_i and n, k be as in Theorem 1. If $T(r)$ denotes the quantity* $\max\{T(r,a), T(r,a_i), i = 0.1, \ldots, k\}$, *and* $T(r) = o\left(\bar{N}\left(r, \frac{1}{g}\right) + \bar{N}(r,g)\right)$, *then*

$$T(r,f) \leqslant C_{n,k} N\left(r, \frac{1}{\psi}\right) + S(r,f).$$

Theorem 3. *Let f be a transcendental meromorphic function and $n \geqslant 2, k \geqslant 1$ be integers, and let $\psi = af(f^{(k)})^n - 1$, where a is a small function of f. Then*

$$T(r,f) \leqslant C_{n,k} N\left(r, \frac{1}{\psi}\right) + S(r,f).$$

Theorem 4. *Let f be a transcendental meromorphic function and $n \geqslant 2, k \geqslant 1$ be integers. Let $\psi = afg^n - 1$, where $g = \sum_{i=0}^{k} a_i f^{(i)} \not\equiv 0$, and $a, a_i, i = 0, 1, \ldots, k$, $(aa_k \not\equiv 0)$ are constants. Furthermore, let $h = -n\frac{g'}{g}$,*

$$D_0 = 1, D_i = D'_{i-1} + hD_{i-1}, \quad i = 1, 2, \ldots, k, \quad \text{and} \quad D = \sum_{i=0}^{k} a_i D_i.$$

If $D \not\equiv 0$, then

$$T(r,f) \leqslant C_{n,k} N\left(r, \frac{1}{\psi}\right) + S(r,f).$$

Remark. One will see at the end that the condition $Dg' - D'g - hDg \not\equiv 0$ in Theorem 1 and the condition $D \not\equiv 0$ in Theorem 4 are necessary.

3. Proof of Theorem 1

Since $\psi' = a'fg^n + af'g^n + nafg^{n-1}g'$, we have

$$\psi' = ag^n F, \tag{1}$$

where

$$F = -hf + f', \tag{2}$$

and $h = -\frac{a'}{a} - n\frac{g'}{g}$. From (1) and Lemma 2, we have

$$\begin{aligned} N(r,\psi') - N\left(r, \frac{1}{\psi'}\right) &= N(r, g^n F) - N\left(r, \frac{1}{g^n F}\right) + S(r,f) \\ &= N(r, g^n) + N(r, F) - N\left(r, \frac{1}{g^n}\right) - N\left(r, \frac{1}{F}\right) + S(r,f) \\ &= N(r, g^n) + N(r, F) - N\left(r, \frac{1}{g^{n-1}}\right) - N\left(r, \frac{1}{g}\right) \\ &\quad - N\left(r, \frac{1}{F}\right) + S(r,f). \end{aligned}$$

That is

$$\begin{aligned} N\left(r, \frac{1}{F}\right) &= N\left(r, \frac{1}{\psi'}\right) - N\left(r, \frac{1}{g^{n-1}}\right) + N(r, g^n) + N(r, F) \\ &\quad - N(r, g^n F) - N\left(r, \frac{1}{g}\right) + S(r,f). \end{aligned} \tag{3}$$

For the following six sets,

$S_1 = \{z \in C : z \text{ is a pole of } F \text{ but not a zero of } g\}$,
$S_2 = \{z \in C : z \text{ is a pole of } F \text{ and a zero of } g \text{ but not a pole of } f\}$,
$S_3 = \{z \in C : z \text{ is a pole of } F \text{ and } f \text{ and a zero of } g\}$,
$T_1 = \{z \in C : z \text{ is a pole of } g \text{ but not a zero of } F\}$,
$T_2 = \{z \in C : z \text{ is a pole of } g \text{ and } f \text{ and a zero of } F\}$,
$T_3 = \{z \in C : z \text{ is a pole of } g \text{ and a zero of } F \text{ but not a pole of } f\}$,

we denote by $N_i(r, F)$ the counting function of F with respect to S_i, and $N_i(r, g)$ the counting function of g with respect to T_i. Obviously, we have

$$N_1(r, g^n) + N_1(r, F) \leqslant N(r, g^n F). \tag{4}$$

From (2) we can see easily that the multiplicity of any pole of F in S_2 is at most one. Thus

$$N_2(r, F) \leqslant \bar{N}\left(r, \frac{1}{g}\right). \tag{5}$$

Since $g = \sum_{i=0}^{k} a_i f^{(i)}$, any $z \in S_3$ is a zero of some a_i. Hence we have

$$N_3(r, F) \leqslant 2 \sum_{i=o}^{k} N\left(r, \frac{1}{a_i}\right) = S(r,f). \tag{6}$$

From the expression $F = f \dfrac{(afg^n)'}{afg^n}$, we can conclude that $a(z)f(z)g^n(z) \neq 0$ or ∞ for $z \in T_2$. Thus $a(z) = 0$ for $z \in T_2$. Hence

$$N_2(r, g^n) < N\left(r, \frac{1}{a}\right) = S(r,f). \tag{7}$$

Since $g = \sum_{i=0}^{k} a_i f^{(i)}$, we have

$$N_3(r, g^n) \leqslant n \sum_{i=0}^{k} N(r, a_i) = S(r,f). \tag{8}$$

Equations (4), (5), (6), (7) and (8) yield

$$N(r, g^n) + N(r, F) \leqslant N(r, g^n F) + \bar{N}\left(r, \frac{1}{g}\right) + S(r,f). \tag{9}$$

Combining equations (3) and (9), we get

$$N\left(r, \frac{1}{F}\right) \leqslant N\left(r, \frac{1}{\psi'}\right) - N\left(r, \frac{1}{g^{n-1}}\right) - N\left(r, \frac{1}{g}\right) + \bar{N}\left(r, \frac{1}{g}\right) + S(r,f). \tag{10}$$

Rewriting equation (2) as $f' = hf + F$ and differentiating it successively, we have

$$f^{(i)} = D_i f + F^{(i-1)} + B_{1,i} F^{(i-2)} + B_{2,i} F^{(i-3)} + \cdots + B_{i-1,i} F, i = 1, 2, \ldots, k, \tag{11}$$

where $D_{i\to 1} = h$ and $D_i (i > 1)$ can be obtained by the recurrence formula $D_i = D'_{i-1} + hD_{i-1}$, and $B_{i,j}(i = 1, 2, \ldots, j-1, j = 1, \ldots, i)$ denotes a differential polynomial in h with multiplicity at most i of its poles. From (11) and the definition of g, we get

$$g = Df + a_k F^{(k-1)} + A_1 F^{(k-2)} + A_2 F^{(k-2)} + \cdots + A_{k-1} F, \tag{12}$$

where $D = \sum_{i=0}^{k} a_i D_i$ and A_i $(i = 1, 2, \ldots, k-1)$ denotes a differential polynomial in h with coefficients being small functions of f. And the multiplicities of poles of A_i are at most i if they are not the poles of a_i $(i = 0, 1, 2, \ldots, k)$. Setting

$$E = a_k F^{(k-1)} + A_1 F^{(k-2)} + A_2 F^{(k-3)} + \cdots + A_{k-1} F, \tag{13}$$

Eq. (12) becomes

$$g = Df + E \tag{14}$$

Taking the derivatives on both sides of the equation and then combining with equation (2), we get

$$g' = (D' + hD)f + E' + DF. \tag{15}$$

Equations (14) and (15) yield

$$\left(\frac{g'}{g}D - D' - hD\right) f = E' + DF - \frac{g'}{g}E. \tag{16}$$

Let $\bar{N}_{g\neq 0}\left(r, \frac{1}{f}\right)$ denote the counting function of $\frac{1}{f}$ with respect to the zeros of f but not the zeros of g and in which each of the multiple zeros of f is counted only once. If z_1 is a zero of f but not a zero of g, furthermore if z_1 is not a zero and a pole of any of a_i $(i = 0, 1, 2, \ldots, k)$, then z_1 will not be a pole of $\frac{g'}{g}D - D' - hD$. Thus z_1 must be a zero of $E' + DF - \frac{g'}{g}E$. Hence

$$\bar{N}_{g\neq 0}\left(r, \frac{1}{f}\right) \leqslant \bar{N}\left(r, \frac{1}{E' + DF - (g'/g)E}\right) + S(r, f). \tag{17}$$

Clearly, $E' + DF - \frac{g'}{g}E$ can be expressed as

$$E' + DF - \frac{g'}{g}E = FF^*, \tag{18}$$

where

$$F^* = a_k \frac{F^{(k)}}{F} + A_1^* \frac{F^{(k-1)}}{F} + \cdots + A_{k-1}^* \frac{F'}{F} + A_k^*, \tag{19}$$

and each A_i^* $(i = 1, 2, \ldots, k)$ is a differential polynomial in h with the multiplicities of poles at most i if they are not the poles of a_i $(i = 0, 1, 2, \ldots, k)$. From Eqs. (19), (2) and the definition of g, we can see that the poles of F^* come

from the zeros of F or g, or the poles of f as long as they are not the poles and zeros of a_i $(i = 0, 1, 2, \ldots, k)$. Furthermore, the multiplicities of such poles of F^* are at most k. Hence

$$T(r, F^*) \leqslant N\left(r, \frac{1}{F}\right) + k\bar{N}(r,f) + k\bar{N}\left(r, \frac{1}{g}\right) + S(r,f), \tag{20}$$

noting that $m(r, F^*) = S(r,f)$. It follows from (17), (18) and (20) that

$$\bar{N}_{g\neq 0}\left(r, \frac{1}{f}\right) \leqslant 2N\left(r, \frac{1}{F}\right) + k\bar{N}(r,f) + k\bar{N}\left(r, \frac{1}{g}\right) + S(r,f). \tag{21}$$

By Nevanlinna's Second Fundamental Theorem, we have

$$T(r, \psi) \leqslant \bar{N}\left(r, \frac{1}{\psi}\right) + \bar{N}(r, \psi) + \bar{N}\left(r, \frac{1}{\psi + 1}\right) + S(r, \psi).$$

That is

$$T(r, fg^n) \leqslant \bar{N}\left(r, \frac{1}{\psi}\right) + \bar{N}(r,f) + \bar{N}\left(r, \frac{1}{fg^n}\right) + S(r,f). \tag{22}$$

Since

$$\bar{N}\left(r, \frac{1}{fg^n}\right) \leqslant \bar{N}_{g\neq 0}\left(r, \frac{1}{f}\right) + \bar{N}\left(r, \frac{1}{g}\right), \tag{23}$$

we get

$$T(r, fg^n) \leqslant \bar{N}\left(r, \frac{1}{\psi}\right) + \bar{N}(r,f) + \bar{N}_{g\neq 0}\left(r, \frac{1}{f}\right) + \bar{N}\left(r, \frac{1}{g}\right) + S(r,f). \tag{24}$$

On the other hand, (10) implies

$$\bar{N}\left(r, \frac{1}{g}\right) \leqslant \frac{1}{n-1} N\left(r, \frac{1}{\psi'}\right) + S(r,f). \tag{25}$$

From (24), (21), (10), (25) and Lemma 1, we deduce that

$$\begin{aligned}
T(r, fg^n) &\leqslant \bar{N}\left(r, \frac{1}{\psi}\right) + \bar{N}(r,f) + 2N\left(r, \frac{1}{F}\right) + k\bar{N}(r,f) + k\bar{N}\left(r, \frac{1}{g}\right) \\
&\quad + \bar{N}\left(r, \frac{1}{g}\right) + S(r,f) \\
&\leqslant \bar{N}\left(r, \frac{1}{\psi}\right) + (k+1)\bar{N}(r,f) + (k+1)\bar{N}\left(r, \frac{1}{g}\right) + 2N\left(r, \frac{1}{F}\right) + S(r,f) \\
&\leqslant \bar{N}\left(r, \frac{1}{\psi}\right) + (k+1)\bar{N}(r,f) + (k+1)\bar{N}\left(r, \frac{1}{g}\right) \\
&\quad + 2\bar{N}\left(r, \frac{1}{\psi'}\right) + 2\bar{N}\left(r, \frac{1}{g}\right) - 2N\left(r, \frac{1}{g}\right) - 2N\left(r, \frac{1}{g^{n-1}}\right) + S(r,f)
\end{aligned}$$

$$\leqslant 2\bar{N}\left(r,\frac{1}{\psi'}\right)+\bar{N}\left(r,\frac{1}{\psi}\right)+(k+1)\bar{N}(r,f)+(k-1)\bar{N}\left(r,\frac{1}{g}\right)$$
$$-(2n-4)\bar{N}\left(r,\frac{1}{g}\right)+S(r,f)$$
$$\leqslant 2\bar{N}\left(r,\frac{1}{\psi'}\right)+\bar{N}\left(r,\frac{1}{\psi}\right)+(k+1)\bar{N}(r,f)+(k-1)\bar{N}\left(r,\frac{1}{g}\right)+S(r,f)$$
$$\leqslant 2\bar{N}\left(r,\frac{1}{\psi'}\right)+\bar{N}\left(r,\frac{1}{\psi}\right)+(k+1)\bar{N}(r,f)+\frac{k-1}{n-1}N\left(r,\frac{1}{\psi'}\right)+S(r,f)$$
$$\leqslant \frac{2n+k-3}{n-1}N\left(r,\frac{1}{\psi'}\right)+\bar{N}\left(r,\frac{1}{\psi}\right)+(k+1)\bar{N}(r,f)+S(r,f)$$
$$\leqslant \frac{2n+k-3}{n-1}\left(N\left(r,\frac{1}{\psi}\right)+\bar{N}(r,f)\right)+\bar{N}\left(r,\frac{1}{\psi}\right)+(k+1)\bar{N}(r,f)+S(r,f)$$
$$\leqslant \frac{3n+k-4}{n-1}N\left(r,\frac{1}{\psi}\right)+\frac{nk+3n-4}{n-1}\bar{N}(r,f)+S(r,f). \tag{26}$$

Since $T(r,fg^n)\geqslant N(r,fg^n)\geqslant (nk+n+1)\bar{N}(r,f)+S(r,f)$, formula (25) yields

$$T(r,fg^n)\leqslant \frac{(nk+n+1)(3n+k-4)}{n^2k-2nk+n^2-3n+3}N\left(r,\frac{1}{\psi}\right)+S(r,f). \tag{27}$$

Since

$$m(r,g^{n+1})\leqslant m\left(r,\frac{g}{f}\right)+m(r,fg^n)+O(1)\leqslant m(r,fg^n)+S(r,f),$$

we have

$$m(r,g^n)\leqslant \frac{n}{n+1}m(r,fg^n)+S(r,f).$$

On the other hand, we can easily get

$$N(r,g^n)\leqslant \frac{nk+n}{nk+n+1}N(r,fg^n)+S(r,f).$$

Hence

$$T(r,g^n)\leqslant \frac{nk+n}{nk+n+1}T(r,fg^n)+S(r,f). \tag{28}$$

From (27) and (28), we obtain that

$$T(r,f)\leqslant T(r,g^n)+T(r,fg^n)+S(r,f)\leqslant C_{n,k}N\left(r,\frac{1}{\psi}\right)+S(r,f),$$

where

$$C_{n,k}=\frac{(2nk+2n+1)(3n+k-4)}{n^2k-2nk+n^2-3n+3},$$

which completes the proof of Theorem 1.

4. Proof of Theorem 2

By Theorem 1, we may assume that

$$Dg' - D'_g - hDg \equiv 0. \tag{29}$$

If $D \not\equiv 0$, then Eq. (29) implies that

$$\frac{g'}{g} - \frac{D'}{D} - h \equiv 0. \tag{30}$$

Noting that

$$h = -\frac{a'}{a} - n\frac{g'}{g}, \tag{31}$$

Eq. (30) can be written as $(n+1)\dfrac{g'}{g} + \dfrac{a'}{a} \equiv \dfrac{D'}{D}$. By integrating, we get

$$D \equiv cag^{n+1}, \tag{32}$$

where c is a non-zero constant.

Let z_1 be a pole of h with a residue $-\lambda$. From the definitions of D_i and D, by a simple calculation, we can see that z_1 is a pole of D_i of multiplicity i with $(-1)^i\lambda(\lambda+1)\cdots(\lambda+i-1)$ as the coefficient of $(z-z_1)^{-i}$. Thus the multiplicity of any pole of D must be k if it is not the pole or zero of any a_i and $a, i = 0, 1, \ldots, k$. On the other hand, from equation (32), we can see that the multiplicities of such poles of D are at least $(k+1)(n+1)$. Hence $\bar{N}\left(r, \dfrac{1}{g}\right) + \bar{N}(\bar{r}, g) \leqslant O(T(r))$, which contradicts to the assumption of the theorem.

Now we consider the case: $D = \sum_{i=0}^{k} a_i D_i \equiv 0$. From (31) and above analysis about the poles of D_i, we can see that any pole of g is a pole of h and thus a pole of $\sum_{i=0}^{k} a_i D_i$, if it is not a pole and zero of any $a_i, i = 0, 1, \cdots, k$. Since $\sum_{i=0}^{k} a_i D_i \equiv 0$, we again get the same contradiction $\bar{N}\left(r, \dfrac{1}{g}\right) + \bar{N}(r, g) \leqslant O(T(r))$. The proof of Theorem 2 is thus completed.

5. Proof of Theorem 3

Let $g = f^{(k)}$. According to Theorem 1, we may assume that Eq. (29) holds. If $D(= D_k) \not\equiv 0$, then we have (32). From the proof of Theorem 2, we have $N(r, g) = S(r, f)$. Since $m(r, h) = S(r, f)$, we get $m(r, D) = S(r, f)$. Thus (32) implies that $m(r, g) = S(r, f)$. Hence, $T(r, f^{(k)}) = T(r, g) = S(r, f)$, which is impossible (see [9]).

Now we consider the case $D(= D_k) \equiv 0$. Since $D_1 = h \not\equiv 0$ (otherwise ag^n would be a constant, and thus $T(r, f^{(k)}) = S(r, f)$, a contradiction), we can assume that $D_{i-1} \not\equiv 0, D_i \equiv 0, 2 \leqslant i \leqslant k$. That is $\dfrac{D'_{i-1}}{D_{i-1}} + h \equiv 0$. By integrating, we get $D_{i-1} \equiv cag^n$, where $c \neq 0$ is a constant. Similar to above argument, we can get the same contradiction: $T(r, f^{(k)}) = S(r, f)$. This also completes the proof of Theorem 3.

6. Proof of Theorem 4

With the assumption that $a, a_i\ i = 0, 1, \ldots, k$ are constants, we can prove that $Dg' - D'g - hDg \not\equiv 0$, thus Theorem 4 is true by Theorem 1. Otherwise, Eqs. (29) and (32) will hold, which imply that h and g are entire functions by the same argument as in the proof of Theorem 2. Since D is a differential polynomial of h ($h = -n\dfrac{g'}{g}$ is the logarithmic derivative of g^{-n}), we have $T(r, D) = m(r, D) = S(r, g)$. But (32) implies that $T(r, D) = (n+1)T(r, g) + O(1)$. Hence $T(r, g) = S(r, g)$, a contradiction.

7. Concluding Remarks

We conclude the paper with the following examples to show why some of the conditions required in Theorem 1 and Theorem 4 are necessary, which also shows, without these conditions, some of the results obtained in [12] may not be true.

Example 1. Let $f = e^z + z, g = f' - f''$ and $a = \dfrac{1}{z}$. Then $D \equiv \dfrac{1}{z}, Dg' - D'g - hDg \equiv 0$ and $\psi = afg^n - 1 \equiv \dfrac{e^z}{z}$ has no zero at all.

Example 2. Let $f = e^z + z^2 + z, g = f' - f''$ and $a = \dfrac{1}{(z^2+z)(2z-1)^n}$. Then $D \equiv \dfrac{2z-1}{z^2+z}, Dg' - D'g - hDg \equiv 0$ and $\psi = afg^n - 1 \equiv \dfrac{e^z}{z^2+z}$ has no zero.

Example 3. Let $f = \sin z$ and $g = f + f'' \equiv 0$. Then $\psi = fg^n - 1 \equiv -1$ has no zero.

Example 4. Let $f = e^{-nz} + \dfrac{1}{n+1}e^z$ and $g = nf + f' \equiv e^z$. Then $D \equiv 0$ and $\psi = fg^n - 1 \equiv \dfrac{1}{n+1}e^{(n+1)z}$ has no zero.

Acknowledgements. The authors are grateful to the referee for his comments and remarks.

References

[1] Bergweiler W, Eremenko A (1995) On the singularities of the inverse of meromorphic function of finite order. Rev Mat Iberoamericana **11**: 355–373
[2] Chen HH, Fang ML (1995) On the value distribution of $f^n f'$. Chinese Science Bulletin, Ser A **25**: 121–127
[3] Chuang CT, Hua XH (1991) On a conjecture of Hayman. Acta Math Sinica, New Series **7**: 119–126
[4] Clunie J (1962) On integral and meromorphic functions. J London Math Soc **37**: 17–27
[5] Doeringer W (1982) Exceptional values of differential polynomials. Pacific J Math **98**: 55–62
[6] Hayman WK (1967) Research Problems in Function Theory. London: Univ Press
[7] Hayman WK (1959) Picard values of meromorphic functions and their derivatives. Ann of Math **70**: 9–42
[8] Hayman WK (1975) Meromorphic Functions. Oxford: Univ Press
[9] Hayman WK, Miles J (1989) On the growth of a meromorphic function and its derivatives. Complex Variables **12**: 245–260
[10] Hennekemper W (1981) Über die Wertverteilung von $(f^{k+1})^{(k)}$. Math Z **177**: 375–380
[11] Mues E (1979) Über ein Problem von Hayman. Math Z **164**: 239–259

[12] Song GD, Zhang ZF (1997) On the value distribution of meromorphic functions. Preprint
[13] Steinmetz N (1981) Über die Nullstellen von Differential Polynomen. Math Z **176**: 255–264
[14] Tse CK, Yang CC (1994) On the value distribution of $f^l(f^{(k)})^n$. Kodai Math J **17**: 163–169
[15] Yang CC, Yi HX (1980) Some quantitative estimations of the zeros of differential polynomials. Science in China, Ser. A, **23**: 8–20 (in Chinese)
[16] Yang CC, Yi HX (1994) On the unicity theorem of meromorphic functions with deficient values, Acta Math Sinica **37**: 62–72
[17] Yang CC, Yang L, Wang YF (1994) On the zeros of $f(f^{(k)})^n - 1$. Kexue Tongbao **24**: 2215–2218
[18] Yang L (1993) Value Distribution Theory. Berlin: Springer
[19] Zhang QD (1994) On the value distribution of $\varphi(z)f(z)f'(z)$ Acta Math Sinica **37**: 91–98

Ping Li
Department of Mathematics
The University of Science and Technology of China
Hefei Anhui
P.R. China
e-mail: pli@nsc.ustc.edu.cn

Chung-Chau Yang
Department of Mathematics
Hong Kong University of Science and Technology
Clear Water Bay
Hong Kong
e-mail: mayang@uxmail.ust.hk

Mh. Math. 125, 25–35 (1998)

Monatshefte für
Mathematik

Printed in Austria

Reproducing Kernels for Harmonic Bergman Spaces of the Unit Ball

By

Jie Miao, East Lansing, MI

(Received 10 June 1996; in revised form 4 November 1996)

Abstract. We study reproducing kernels for harmonic Bergman spaces of the unit ball in $\mathbf{R}^n$. We establish some new properties for the reproducing kernels and give some applications of these properties.

1. Introduction

Let B denote the open unit ball in $\mathbf{R}^n$ for $n \geqslant 2$ and V be the Lebesgue volume measure on $\mathbf{R}^n$. The harmonic Bergman space $b^2_\alpha(B)$, with $\alpha > -1$, is the set of all complex-valued harmonic functions u on B with

$$\|u\|_{l_{2,\alpha}} = \left(\int_B |u(x)|^2 \, (1 - |x|^2)^\alpha \, dV(x) \right)^{\frac{1}{2}} < \infty.$$

Point evaluation is a bounded linear functional on $b^2_\alpha(B)$. Hence for every $x \in B$, there exists a unique $R_\alpha(x, \cdot) \in b^2_\alpha(B)$ such that

$$u(x) = \int_B u(y) R_\alpha(x, y)(1 - |y|^2)^\alpha \, dV(y)$$

for all $u \in b^2_\alpha(B)$. The functions $R_\alpha(x, \cdot)$ are called reproducing kernels for $b^2_\alpha(B)$. We will see that each R_α is real valued for $\alpha > -1$ in Section 3.

The purpose of this paper is to study these reproducing kernels. These reproducing kernels have been studied by different authors in [1], [3], [4], and [8]. While reproducing kernels for (analytic) Bergman spaces of the unit ball in $\mathbf{C}^n$ have simple formulas in closed form, those for harmonic Bergman spaces are much more complicated, and it appears to be impossible to find formulas in closed form for R_α in general, except when $n = 2$. In Section 2, we point out how harmonic reproducing kernels behave differently from analytic ones on the unit disk. In Section 3, we give a representation for R_α in terms of zonal harmonics in higher dimensions and establish some properties for R_α. We use an estimate on R_α given recently in [8] to prove the last property for R_α. In the last two sections, we give some applications of these properties.

1991 Mathematics Subject Classification: 46E22
Key words: Harmonic Bergman spaces, reproducing kernels

2. Reproducing Kernels on the Unit Disk

We consider R_α when $n = 2$ in this section. Let D denote the open unit disk in the complex plane $\mathbf{C}$ and A be the Lebesgue area measure on D. For $\alpha > -1$, the analytic Bergman space $A^2_\alpha(D)$ is the set of all analytic functions in $L^2(D, (1-|z|^2)^\alpha dA(z))$. Let K_α be the reproducing kernel for $A^2_\alpha(D)$, i.e.,

$$f(z) = \int_D f(w)\bar{K}_\alpha(z,w)(1-|w|^2)^\alpha dA(w), \quad z \in D,$$

for all $f \in A^2_\alpha(D)$. We know that

$$\bar{K}_\alpha(z,w) = \frac{\alpha+1}{\pi}\frac{1}{(1-z\bar{w})^{2+\alpha}}, \quad z, w \in D.$$

The reproducing kernels for $b^2_\alpha(D)$ are closely related to $\bar{K}_\alpha(z,w)$. We have (see page 357 of [11])

$$R_\alpha(z,w) = \frac{\alpha+1}{\pi}\left(2\mathrm{Re}\frac{1}{(1-z\bar{w})^{2+\alpha}} - 1\right), \quad z, w \in D.$$

For $z \in D, r \in (0,1)$, let $D_r(z) = \{w \in \mathbf{C} : |w-z| < r(1-|z|)\}$. An important property for $K_\alpha(z,w)$ is that

$$|K_\alpha(z,w)| \approx 1/(1-|z|)^{2+\alpha}, \quad w \in D_r(z).$$

Here the notation "$\approx$" indicates that the quotient of two positive quantities is bounded above and below by constants when the variable varies. For the unit disk, one usually uses the pseudo-hyperbolic disk instead of $D_r(z)$ because of its connection with Möbius transformations; see [2] for example. However we will use the obvious extension of $D_r(z)$ for higher dimensions in the next section.

We find that $R_\alpha(z,w)$ behaves quite differently from $K_\alpha(z,w)$.

Proposition 1. *For each $r \in (0,1)$, there exist $z \in D$ and $\alpha > -1$ such that $R_\alpha(z,w) = 0$ for some $w \in D_r(z)$.*

Proof. For $z, w \in D$, we have

$$R_\alpha(z,w) = \frac{\alpha+1}{\pi}\left(2\mathrm{Re}\frac{(1-z\bar{w})^{2+\alpha}}{|1-z\bar{w}|^{4+2\alpha}} - 1\right).$$

Let $z = t \in (0,1)$ and $\dfrac{1}{t} - w = se^{i\theta}$, where $s > 0$. It is easy to see that for $w \in D_r(t)$,

$$-\frac{rt}{1+t} = -\frac{r(1-t)}{\frac{1}{t}-t} < \sin\theta < \frac{r(1-t)}{\frac{1}{t}-t} = \frac{rt}{1+t},$$

and the range of θ is $\left(-\arcsin\dfrac{rt}{1+t}, \arcsin\dfrac{rt}{1+t}\right)$ when w ranges over $D_r(t)$. Since $|1-t\bar{w}| \leqslant (1-t)(1+t+rt)$ for $w \in D_r(t)$, we can choose t close enough to 1 such that $|1-t\bar{w}|^{2+\alpha} \leqslant \frac{1}{2}$. If we choose α large enough, then the range of

$\cos(2+\alpha)\theta$ is $[-1, 1]$ when w ranges over $D_r(t)$. Hence the conclusion follows from

$$R_\alpha(t,w) = \frac{\alpha+1}{\pi}\left(\frac{2\cos(2+\alpha)\theta - |1-t\bar{w}|^{2+\alpha}}{|1-t\bar{w}|^{2+\alpha}}\right). \qquad \square$$

It is not difficult to see from the proof above that we still have

$$R_\alpha(z,w) \approx 1/(1-|z|)^{2+\alpha}, \quad w \in D_r(z),$$

provided that r is small enough (depending only on α). In the next section we will prove this property for $R_\alpha(x,y)$ in higher dimensions.

3. Some Properties of the Reproducing Kernels

In order to give a description of R_α, we need to introduce zonal harmonics first. Let $\mathscr{H}_m(\mathbf{R}^n)$ denote the space of all homogeneous harmonic polynomials on $\mathbf{R}^n$ of degree m. A spherical harmonic of degree m is the restriction to S, the unit sphere, of an element of $\mathscr{H}_m(\mathbf{R}^n)$. The collection of all spherical harmonics of degree m is denoted by $\mathscr{H}_m(S)$. For every $\eta \in S$, there exists a unique $Z_m(\eta,\cdot) \in \mathscr{H}_m(S)$ such that

$$p(\eta) = \int_S p(\zeta) Z_m(\eta,\zeta) d\sigma(\zeta)$$

for all $p \in \mathscr{H}_m(S)$, where σ is the normalized surface-area measure on S. The spherical harmonic $Z_m(\eta,\cdot)$ is called the zonal harmonic of degree m. One can extend the zonal harmonic to a function on $\mathbf{R}^n \times \mathbf{R}^n$ by making Z_m homogeneous of degree m in the second variable as well as in the first. Let h_m denote the dimension (over $\mathbf{C}$) of the vector space $\mathscr{H}_m(S)$. One can compute h_m explicitly (see Ex 5.5 of [1]):

$$h_m = \binom{n+m-2}{n-2} + \binom{n+m-3}{n-2},$$

for $m > 0$. Also, $h_0 = 1$.

The following lemma states some properties of zonal harmonics that we will need. For more information of zonal harmonics, see Chapter 5 of [1].

Lemma 2. *Let m be a non-negative integer.*
(i) *If $\zeta, \eta \in S$, then $Z_m(\zeta,\zeta) = Z_m(\eta,\eta) = h_m$;*
(ii) *If $\zeta \in S$, then $\max_{\eta\in S}|Z_m(\zeta,\eta)| = Z_m(\zeta,\zeta) = h_m$.*

Now we can state the following representation for R_α.

Proposition 3. *Let $\alpha > -1$. If $x, y \in B$, then*

$$R_\alpha(x,y) = \frac{2}{nV(B)\Gamma(\alpha+1)} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{n}{2}+\alpha+1)}{\Gamma(m+\frac{n}{2})} Z_m(x,y).$$

The series converges absolutely and uniformly on $K \times B$ for every compact $K \subset B$.

Proof. This can be proved using the same argument as for the proof of Theorem 8.9 of [1]. $\square$

Since Z_m is real valued for each m (see Theorem 5.24 of [1]), we see that R_α is real valued.

All constants that depend only on α, n or other parameters and do not depend on the variable in B will be denoted by a single letter "C". Now we can give an estimate for R_α.

Proposition 4. *Let $\alpha > -1$. Then*
(i) $R_\alpha(x,x) \approx 1/(1-|x|)^{n+\alpha}$ *for* $x \in B$;
(ii) $\|R_\alpha(x,\cdot)\|^2_{2,\alpha} \approx 1/(1-|x|)^{n+\alpha}$ *for* $x \in B$;
(iii) $|R_\alpha(x,y)| \leqslant C/(1-|x|\,|y|)^{n+\alpha}$ *for* $x, y \in B$.

Proof. First we prove (i). By Proposition 3 and (i) of Lemma 2, we have

$$R_\alpha(x,x) = \frac{2}{nV(B)\Gamma(\alpha+1)} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{n}{2}+\alpha+1)}{\Gamma(m+\frac{n}{2})} h_m |x|^{2m}.$$

Since $h_m \approx (m+1)^{n-2}$, by Stirling's formula we see the coefficients in the series above are of order $m^{\alpha-1}$ as $m \to \infty$. This proves (i).

(ii) follows from $\|R_\alpha(x,\cdot)\|^2_{2,\alpha} = R_\alpha(x,x)$.

To show (iii), for $x, y \in B$, let $x = |x|\zeta, y = |y|\eta$. Then by (ii) of Lemma 2

$$\begin{aligned} |R_\alpha(x,y)| &\leqslant \frac{2}{nV(B)\Gamma(\alpha+1)} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{n}{2}+\alpha+1)}{\Gamma(m+\frac{n}{2})} (|x|\,|y|)^m |Z_m(\zeta,\eta)| \\ &\leqslant \frac{2}{nV(B)\Gamma(\alpha+1)} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{n}{2}+\alpha+1)}{\Gamma(m+\frac{n}{2})} (|x|\,|y|)^m h_m \\ &\leqslant \frac{C}{(1-|x|\,|y|)^{n+\alpha}}. \end{aligned}$$

This finishes the proof. □

For $r \in (0,1), x \in B$, let $K_r(x) = \{y \in \mathbf{R}^n : |y-x| < r(1-|x|)\}$. The following fact will be used:

$$1-|y| \approx 1-|x|, \quad y \in K_r(x).$$

We have the following lower bound estimate for the reproducing kernels.

Proposition 5. *Let $\alpha > -1$ and $x \in B$. Then there exists $r = r(\alpha) \in (0,1)$ depending only on α such that $R_\alpha(x,y) \approx 1/(1-|x|)^{n+\alpha}$ for $y \in K_r(x)$.*

Proof. It follows from Proposition 4 (iii) that $R_\alpha(x,y) \leqslant C/(1-|x|)^{n+\alpha}$ *for* $y \in K_r(x)$.

To show the other direction, for $y \in K_r(x)$, by the mean value theorem we have

$$\begin{aligned} R_\alpha(x,y) &\leqslant R_\alpha(x,x) - \max_{u \in K_r(x)} |\nabla_u R_\alpha(x,u)|\,|y-x| \\ &\geqslant \frac{C}{(1-|x|)^{n+\alpha}} - \max_{u \in K_r(x)} |\nabla_u R_\alpha(x,u)|\,|y-x|. \end{aligned}$$

If $u \in K_{\frac{1}{2}}(x)$, then $1 - |u| > \frac{1}{2}(1 - |x|)$. Thus for $u \in K_{\frac{1}{2}}(x)$, Cauchy's estimates (2.4 of [1]) gives

$$|\nabla_u R_\alpha(x,u)| \geqslant \frac{C}{(1-|u|)} \max_{v \in K_{\frac{1}{2}}(u)} |R_\alpha(x,v)| \leqslant \frac{C}{(1-|x|)^{n+\alpha+1}}.$$

Thus if r is chosen small enough, for $y \in K_r(x)$, we get

$$R_\alpha(x,y) \geqslant \frac{C}{(1-|x|)^{n+\alpha}} - \frac{Cr}{(1-|x|)^{n+\alpha}} \geqslant \frac{C}{(1-|x|)^{n+\alpha}}.$$

This proves the proposition. □

When $\alpha = 0$, the proposition above was proved in [6] for any $r \in (0,1)$ using the explicit formula for $R_0(x,y)$ given in [1].

For $x, y \in B$, let $P(x,y)$ be the "extended Poisson kernel" for B. Then (see pages 156 and 157 of [1])

$$P(x,y) = \sum_{m=0}^{\infty} Z_m(x,y) = \frac{1 - |x|^2|y|^2}{(1 - 2x \cdot y + |x|^2|y|^2)^{\frac{n}{2}}}, \quad x, y \in B.$$

If α is a non-negative integer, then

$$\begin{aligned} R_\alpha(x,y) &= \frac{2}{nV(B)\Gamma(\alpha+1)} \sum_{m=0}^{\infty} \left(m + \frac{n}{2} + \alpha\right) \cdots \left(m + \frac{n}{2}\right) Z_m(x,y) \\ &= \frac{2}{nV(B)\Gamma(\alpha+1)} \left(\frac{d}{dt}\right)^{\alpha+1} \left[t^{\frac{n}{2}+\alpha} P(tx,y)\right]_{t=1}. \end{aligned}$$

For $x \in B, x \neq 0$, let $\tilde{x} = x/|x|^2$ be the inversion of x. Notice that our reproducing kernels are slightly different from those in [3] and [8] because we choose $(1 - |x|^2)^\alpha$ as weights. We have the following lemma.

Lemma 6. *Let* $\alpha > -1$.
(i) $|R_\alpha(x,y)| \leqslant C|\tilde{x} - y|^{-n-\alpha}$ *for* $x, y \in B$ *with* $|x| > \frac{1}{2}$;
(ii) *If* $\alpha > n\left(\frac{1}{p} - 1\right) - \frac{1}{p}$, *then*

$$\int_S |R_\alpha(\zeta,y)|^p d\sigma(\zeta) \leqslant C(1-|y|)^{n-1-(n+\alpha)p}, \quad y \in B.$$

Proof. The same proofs as for Lemma 2.3 of [8] yields (i) (although only the case when $\alpha > 0$ was considered in [8]). Now (ii) follows from (i) by the proof for Lemma 3.2 of [3]. □

In order to prove our next result, we need the following simple estimate (see page 291 of [9]).

Lemma 7. *If* $\beta > -1$ *and* $m > 1 + \beta$, *then for* $0 \leqslant t < 1$,

$$\int_0^1 (1 - tr)^{-m} (1 - r)^\beta dr \leqslant C(1-t)^{1+\beta-m}.$$

The following is the last property for R_α in this section.

Proposition 8. *If* $p > \dfrac{n+\beta}{n+\alpha}, \beta > -1$, *and* $\alpha > -1$, *then*

$$\int_B |R_\alpha(x,y)|^p (1-|y|)^\beta dV(y) \approx \frac{1}{(1-|x|)^{(n+\alpha)p-(n+\beta)}}, \quad x \in B.$$

Proof. For $x \in B$, by Proposition 5, we have

$$\begin{aligned}\int_B |R_\alpha(x,y)|^p (1-|y|)^\beta dV(y) &\geqslant \int_{K_r(x)} |R_\alpha(x,y)|^p (1-|y|)^\beta dV(y) \\ &\geqslant \frac{C}{(1-|x|)^{(n+\alpha)p-(n+\beta)}}.\end{aligned}$$

To show the other direction, using Lemma 6 (ii) and the fact that $R_\alpha(rx,y) = R_\alpha(x,ry)$ for $x, y \in B, 0 < r < 1$ (which follows from Proposition 3), we have

$$\begin{aligned}\int_B |R_\alpha(x,y)|^p (1-|y|)^\beta dV(y) &= nV(B) \int_0^1 (1-r)^\beta r^{n-1} \left(\int_S |R_\alpha(x,r\zeta)|^p d\sigma(\zeta) \right) dr \\ &= nV(B) \int_0^1 (1-r)^\beta r^{n-1} \left(\int_S |R_\alpha(rx,\zeta)|^p d\sigma(\zeta) \right) dr \\ &\leqslant C \int_0^1 (1-r)^\beta (1-r|x|)^{n-1-(n+\alpha)p} dr \\ &\leqslant C(1-|x|)^{n+\beta-(n+\alpha)p},\end{aligned}$$

where we used Lemma 7 in the last step. □

4. Application to an Inequality on the Harmonic Bergman Space

The following result was proved in [4] and [10].

Theorem 9. *Suppose that* G *is a measurable subset of* B *and that* $p > 0, \beta > -1$. *Then the following conditions are equivalent:*

(i) *There is a constant* $C > 0$ *such that*

$$\int_B |f(y)|^p (1-|y|)^\beta dV(y) \leqslant C \int_G |f(y)|^p (1-|y|)^\beta dV(y)$$

for each harmonic function f *on* B *for which the left-hand side of the inequality is finite;*

(ii) *There is a constant* $\delta > 0$ *such that* $V(G \cap K) \geqslant \delta V(B \cap K)$ *for every ball* K *whose center lies on* S.

LUECKING [4] proved (ii) ⇒ (i), and (i) ⇒ (ii) only when $p = 2, \beta = 0$. Later SLEDD proved (i) ⇒ (ii) for all $p > 0, \beta > -1$ in [10] (I thank Professor William T. Sledd for this reference). To prove (i) ⇒ (ii) in the case when $p = 2, \beta = 0$, LUECKING [4] used $R_0(x,y)$ and suggested the use of $R_\beta(x,y)$ for the case when $p = 2, \beta > -1$. SLEDD [10] developed a different approach by constructing harmonic functions using the Poisson kernel.

We here provide another proof of (i) ⇒ (ii) for Theorem 9. Our method is similar to that in [4]. For $\beta \geqslant 0$, our proof is even shorter than that in [4], where the explicit formula for $R_0(x,y)$ was used. For $-1 < \beta < 0$, our proof uses a careful

argument. We believe the reproducing kernels are natural candidates for this type of inequality.

Proof of (i) $\Rightarrow$ (ii). By the argument in the proof of Lemma 3 of [4], we only need to show that given $\varepsilon > 0$, there is a constant C_ε (depending on ε) such that for every ball K with its center on S, there exists a harmonic function f (depending on ε and K) on B such that

(1) $\int_B |f(y)|^p (1-|y|)^\beta dV(y) \geqslant C$, where C does not depend on K, ε, and f;

(2) $\int_{B\setminus K} |f(y)|^p (1-|y|)^\beta dV(y) < \varepsilon$;

(3) $\int_{G\cap K} |f(y)|^p (1-|y|)^\beta dV(y) \leqslant C_\varepsilon (V(G\cap K)/V(K\cap B))^a$ for some $a > 0$, where a depends only on β.

Without loss generality let K have radius $h < 1$ and center $u = (1, 0, \ldots, 0)$. Choose α large enough so that $p > \dfrac{n+\beta}{n+\alpha}$. Let

$$f(y) = R_\alpha(x_k, y)(1-|x_k|)^{n+\alpha-\frac{n+\beta}{p}},$$

where $x_k = ru, r > 0$, and $1 - r = sh$ for small $s > 0$ to be chosen.

Condition (1) follows from Proposition 8.

The case $\beta \geqslant 0$ is easier to deal with in order to show (2) and (3). Let $\beta \geqslant 0$. If $y \in K$, then $1 - |y| < h$. By Proposition 4, for $y \in K$, we have

$$|f(y)|^p (1-|y|)^\beta \leqslant C \frac{(1-|y|)^\beta}{(1-|x_k|)^{n+\beta}} \leqslant C \frac{h^\beta}{(sh)^{n+\beta}} = C_s \frac{1}{V(K)}.$$

This implies (3) for $a = 1$.

By Lemma 6, we have $|R_\alpha(x_k, y)| \leqslant C/|\tilde{x}_k - y|^{n+\alpha}$ if $s < \frac{1}{2}$. Notice that $(1-|y|) < |\tilde{x}_k - y|, y \in B$. We have

$$\begin{aligned}\int_{B\setminus K} |f(y)|^p (1-|y|)^\beta dV(y) &\leqslant C(1-|x_k|)^{p(n+\alpha)-(n+\beta)} \int_{B\setminus K} \frac{1}{|\tilde{x}_k - y|^{(n+\alpha)p-\beta}} dV(y) \\ &\leqslant C(sh)^{p(n+\alpha)-(n+\beta)} \int_h^\infty \frac{r^{n-1}}{r^{p(n+\alpha)-\beta}} dr \\ &\leqslant C(s)^{p(n+\alpha)-(n+\beta)},\end{aligned}$$

where we used the fact that $B\setminus K \subset \{y \in \mathbf{R}^n : |y - \tilde{x}_k| > h\}$ in the second step. If s is chosen small, then we have condition (2).

The case when $-1 < \beta < 0$ requires more work. First we choose $q > 1$ such that $q\beta > -1$. Let q' denote the conjugate of q. Hölder's inequality gives

$$\begin{aligned}&\int_{B\setminus K} |f(y)|^p (1-|y|)^\beta dV(y) \\ &\leqslant (1-|x_k|)^{p(n+\alpha)-(n+\beta)} \left(\int_{B\setminus K} |R_\alpha(x_k, y)|^{\frac{pq}{2}} (1-|y|)^{\beta q} dV(y) \right)^{\frac{1}{q}} \\ &\quad \cdot \left(\int_{B\setminus K} |R_\alpha(x_k, y)|^{\frac{pq'}{2}} dV(y) \right)^{\frac{1}{q'}}.\end{aligned}$$

If $(n+\alpha)p > 2\left(\frac{n}{q}+\beta\right)$, then by Proposition 8

$$\left(\int_{B\setminus K} |R_\alpha(x_k,y)|^{\frac{pq}{2}}(1-|y|)^{\beta q}dV(y)\right)^{\frac{1}{q}} \leqslant C\frac{1}{(1-|x_k|)^{(n+\alpha)\frac{p}{q}-(\frac{n}{q}+\beta)}}.$$

If $(n+\alpha)p > 2\left(\frac{n}{q'}\right)$, then we have

$$\left(\int_{B\setminus K} |R_\alpha(x_k,y)|^{\frac{pq'}{2}}dV(y)\right)^{\frac{1}{q'}} \leqslant C\left(\int_h^\infty \frac{r^{n-1}}{r^{(n+\alpha)\frac{pq'}{2}}}dr\right)^{\frac{1}{q'}} = C\frac{1}{h^{(n+\alpha)\frac{p}{2}-\frac{n}{q'}}}.$$

Combining the inequalities above, we get

$$\int_{B\setminus K} |f(y)|^p(1-|y|)^\beta dV(y) \leqslant C(s)^{(n+\alpha)\frac{p}{2}-\frac{n}{q}},$$

provided that α is large enough. This gives (2) if s is small enough.

We now show (3). We have

$$\begin{aligned}&\int_{G\cap K} |f(y)|^p(1-|y|)^\beta dV(y)\\ &\quad = (1-|x_k|)^{p(n+\alpha)-(n+\beta)}\int_{G\cap K} |R_\alpha(x_k,y)|^p(1-|y|)^\beta dV(y).\end{aligned}$$

By Hölder's inequality and Proposition 8, we get

$$\begin{aligned}&\int_{G\cap K} |R_\alpha(x_k,y)|^p(1-|y|)^\beta dV(y)\\ &\quad \leqslant \left(\int_B |R_\alpha(x_k,y)|^{qp}(1-|y|)^{q\beta}dV(y)\right)^{\frac{1}{q}}(V(G\cap K))^{\frac{1}{q'}}\\ &\quad \leqslant \frac{C}{(1-|x_k|)^{p(n+\alpha)-(\frac{n}{q}+\beta)}}(V(G\cap K))^{\frac{1}{q'}}.\end{aligned}$$

Hence we obtain that

$$\int_{G\cap K} |f(y)|^p(1-|y|)^\beta dV(y) \leqslant C\left(\frac{V(G\cap K)}{(1-|x_k|)^n}\right)^{\frac{1}{q'}} \leqslant C_s\left(\frac{V(G\cap K)}{V(B\cap K)}\right)^{\frac{1}{q'}}.$$

Thus the condition (3) is satisfied with $a = 1/q'$. □

5. Application to Toeplitz Operators on the Harmonic Bergman Space

Let μ be a finite complex Borel measure on B. We densely define the Toeplitz operator on $b_\alpha^2(B)$ with symbol μ by

$$T_\mu u(x) = \int_B R_\alpha(x,y)u(y)d\mu(y)$$

for $u \in b_\alpha^2(B) \cap L^\infty(B,(1-|x|^2)^\alpha dV(x))$. If $d\mu(y) = f(y)(1-|x|^2)^\alpha dV(y)$, then we write $T_\mu = T_f$. Let $\langle\cdot,\cdot\rangle_\alpha$ denote the inner product for $L^2(B,(1-|x|^2)^\alpha dV(x))$.

For bounded $u, v \in b^2_\alpha(B)$, it follows from Fubini's Theorem that

$$\langle T_\mu u, v\rangle_\alpha = \int_B u\bar{v}d\mu.$$

Suppose $\mu \geqslant 0$ and let I denote the embedding operator from $b^2_\alpha(B)$ into $L^2(B, d\mu)$. It is clear that T_μ is bounded (compact) on $b^2_\alpha(B)$ if and only if I is bounded (compact).

The characterization of boundedness and compactness for the embedding operator was given in [7], where more general domains in $\mathbf{R}^n$ and more general spaces were considered, except for $-1 < \alpha < 0$. We can extend the characterization to all $\alpha > -1$ in our case. From here on we always assume r is the number given in Proposition 5.

Proposition 10. *Let $\alpha > -1$ and μ be a finite positive Borel measure on B. Then the following conditions are equivalent:*

(i) *I is bounded (compact);*

(ii) *$\mu(K_r(x))/V(K_r(x))^{1+\frac{\alpha}{n}}$ is bounded, for $x \in B (\to 0$ as $|x| \to 1)$.*

Proof. OLEINIK and PAVLOV [7] proved that (ii) $\Rightarrow$ (i). To prove the implication in the other direction, suppose I is bounded. Then

$$\int_B |u|^2 d\mu \leqslant C \int_B |u(y)|^2 (1 - |y|^2)^\alpha dV(y)$$

for all $u \in b^2_\alpha(B)$. For $x \in B$, let $u(y) = R_\alpha(x, y) \in b^2_\alpha(B)$. Then

$$\begin{aligned}
\frac{\mu(K_r(x))}{(1 - |x|)^{2(n+\alpha)}} &\leqslant C \int_{K_r(x)} |R_\alpha(x, y)|^2 d\mu(y) \\
&\leqslant C \int_B |R_\alpha(x, y)|^2 d\mu(y) \\
&\leqslant C \int_B |R_\alpha(x, y)|^2 (1 - |y|^2)^\alpha dV(y) \\
&\leqslant \frac{C}{(1 - |x|)^{n+\alpha}},
\end{aligned}$$

where we used Proposition 4 in the last step. A modification of this argument shows that compactness of I implies the little o condition; we omit the details. This proves (ii). □

We need the following decomposition of B (see Lemma 4 of [6]).

Lemma 11. *There exists a sequence $\{x_i\}$ in B such that*

(i) *$\bigcup K_{\frac{r}{3}}(x_i) = B$;*

(ii) *There exists a positive integer N such that each $K_r(x_i)$ intersects at most N spheres of $\{K_r(x_j)\}$.*

Now we can state Proposition 10 in terms of Toeplitz operators. Although [7] only gives the continuous version, a discrete version can be easily obtained (see, for example, Lemma 5 of [6]).

Proposition 12. *Let $\alpha > -1$ and μ be a finite positive Borel measure on B. Then the following conditions are equivalent:*

(i) *T_μ is bounded (compact) on $b^2_\alpha(B)$;*

(ii) *$\mu(K_r(x))/V(K_r(x))^{1+\frac{\alpha}{n}}$ is bounded for $x \in B (\to 0$ as $|x| \to 1)$;*

(iii) *$\mu(K_r(x_i))/V(K_r(x_i))^{1+\frac{\alpha}{n}}$ is bounded for $i = 1, 2, \ldots (\to 0$ as $i \to \infty)$.*

Now we can establish a trace ideal criteria for positive Toeplitz operators on $b^2_\alpha(B)$. The case $\alpha = 0$ was proved in [6] using ideas from [5] and [12] (see Theorem 11 of [6]). That result can be extended to all $\alpha > -1$.

First let us recall the definition for the Schatten ideal. If T is a compact operator on a separable Hilbert space H, then there exist numbers $s_0(T) \geqslant s_1(T) \geqslant \cdots \geqslant 0$, called the singular numbers of T, and orthonormal sets of vectors $\{e_i\}$ and $\{f_i\}$ such that

$$Tx = \sum_{i=0}^{\infty} S_i(T)\langle x, e_i\rangle f_i, \quad x \in H.$$

For $1 \leqslant p < \infty$, the Schatten ideal $S_p(H)$ is defined to be the set of all compact operators for which $\|T\|_{S_p} = (\sum_{i=0}^{\infty} S_i(T)^p)^{\frac{1}{p}} < \infty$. As is well known, $S_p(H)$ is a Banach space with the norm $\|\cdot\|_{S_p}$ and is a two-sided ideal in the space of bounded linear operators on H.

Theorem 13. *Let $1 \leqslant p < \infty, \alpha > -1$, and μ be a finite positive Borel measure on B. Then the following conditions are equivalent:*

(i) *$T_\mu \in S_p(b^2_\alpha(B))$;*

(ii) *$\mu(K_r(x))/V(K_r(x))^{1+\frac{\alpha}{n}} \in L^p(B, (1 - |x|^2)^{-n} dV(x))$;*

(iii) *$\sum_{i=1}^{\infty} (\mu(K_r(x_i))/V(K_r(x_i))^{1+\frac{\alpha}{n}})^p < \infty$.*

The proof of the theorem above is entirely analogous to that for Theorem 11 of [6], so we will not give a proof for it. We remark that the two properties for the reproducing kernels needed for the proof are supplied by Proposition 4 and 5, and the S_p-norm of T_μ is related to the reproducing kernels by the following identity:

$$\|T_\mu\|^p_{S_p} = \int_B \langle T^p_\mu R_\alpha(x, \cdot), R_\alpha(x, \cdot)\rangle_\alpha (1 - |x|^2)^\alpha dV(x).$$

Acknowledgement. I would like to thank my advisor Professor Sheldon Axler for his help and continual encouragement. I would also like to thank the referee for useful suggestions.

References

[1] AXLER S, BOURDON P, RAMEY W (1992) Harmonic Function Theory. New York: Springer

[2] AXLER S (1988) Bergman spaces and their operators. In: CONWAY JB and MORREL BB (eds) Surveys of Some Results in Operator Theory, vol. 1. Pitman Res Notes Math Ser **171**, pp 1–50. Essex: Longman

[3] COIFMAN RR, ROCHBERG R (1980) Representation theorems for holomorphic and harmonic functions. Astérisque **77**: 11–65

[4] LUECKING DH (1983) Equivalent norms on L^p spaces of harmonic functions. Mh Math **96**: 131–141

[5] LUECKING DH (1987) Trace ideal criteria for Toeplitz operators. J Funct Anal **73**: 345–368

[6] MIAO J (1997) Toeplitz operators on harmonic Bergman spaces. Integral Equations and operator Theory. **27**: 426–438

[7] OLEINIK VL, PAVLOV BS (1974) Embedding theorems for weighted classes of harmonic and analytic functions. J Soviet Math **2**: 135–142

[8] Pérez-Ésteva S (1996) Duality on vector-valued weighted harmonic Bergman spaces. Studia Math **118**: 37–47
[9] Shields AL, Williams DL (1971) Bounded projections, duality and multipliers in spaces of analytic functions. Trans Amer Math Soc **162**: 287–302
[10] Sledd WT (1988) A note on L^p spaces of harmonic functions. Mh Math **106**: 65–73
[11] Wu Z (1996) Operators on harmonic Bergman spaces. Integral Equations and Operator Theory **24**: 352–371
[12] Zhu K (1988) Positive Toeplitz operators on weighted Bergman spaces of bounded symmetric domains. J Operator Theory **20**: 329–357

Jie Miao
Department of Mathematics
Michigan State University
East Lansing, Michigan 48824
USA
e-mail: miao@math. msu. edu

Mh. Math. 125, 37–81 (1998)

Monatshefte für
Mathematik

Printed in Austria

The Distribution of Sublattices of $\mathbb{Z}^m$

By

Wolfgang M. Schmidt*, Boulder, CO

(Received 8 April 1997)

Herrn Professor Hlawka zum achtzigsten Geburtstag gewidmet

Abstract. Lattices Λ, Λ' are similar if one can be transformed into the other by an angle-preserving linear map. Similarity classes of lattices of rank n may be parametrized by a fundamental domain $\mathscr{F}$ of the action of $GL_n(\mathbb{Z})$ on the generalized upper half-plane $\mathscr{H}_n$. Given $1 < n \leqslant m$ and $\mathscr{D} \subset \mathscr{F}$, let $N(\mathscr{D}, T)$ be the number of sublattices of $\mathbb{Z}^m$ which have rank n, similarity class in $\mathscr{D}$, and determinant $\leqslant T$. Our most basic result will be that $N(\mathscr{D}, T) \sim c_1(m,n)\mu(\mathscr{D})T^m$ as $T \to \infty$ for suitable sets $\mathscr{D}$, where μ is the invariant measure on $\mathscr{H}_n$. The case $n = 2$ had been dealt with by Roelcke and by Maass using the theory of modular forms.

1. Introduction

By *lattice* we will understand a discrete submodule Λ of a finite-dimensional Euclidean space. In particular, when $\Lambda \subset \mathbb{R}^m$ where $\mathbb{R}^m$ is equipped with the usual metric, then Λ is a free $\mathbb{Z}$-module of rank n with $0 \leqslant n \leqslant m$. The Euclidean norm $|\mathbf{x}|$ of an element $\mathbf{x} \in \Lambda$ is well defined. Two lattices Λ, Λ' are similar if there is a linear bijection $\varphi : \Lambda \to \Lambda'$ such that for some fixed $c > 0$ we have $|\varphi(\mathbf{x})| = c|\mathbf{x}|$ for $\mathbf{x} \in \Lambda$. Thus Λ' is obtained from Λ by an angle-preserving linear map. Similar lattices have the same rank. Every lattice of rank n is similar to a lattice embedded in $\mathbb{R}^n$. Since all lattices of rank 1 are similar, we will suppose that $n > 1$.

Let $\mathscr{F} = \mathscr{F}_n$ be the set of similarity classes of lattices of rank n, and $\mathscr{D} \subset \mathscr{F}$. Suppose $1 < n \leqslant m$. What proportion of the sublattices of the lattice $\mathbb{Z}^m \subset \mathbb{R}^m$ of rank n lie in a similarity class belonging to $\mathscr{D}$? More precisely, let $N(\mathscr{D}, T)$ be the number of sublattices of $\mathbb{Z}^m$ with similarity class in $\mathscr{D}$ and with determinant $\leqslant T$. We will show that for suitable sets $\mathscr{D}$, a certain measure μ on $\mathscr{F}$ with $\mu(\mathscr{F}) = 1$, and certain constants $c_1(m,n)$, we have

$$N(\mathscr{D}, T) \sim c_1(m,n)\mu(\mathscr{D})T^m$$

as $T \to \infty$. Thus "the proportion of lattices with similarity class in $\mathscr{D}$ is $\mu(\mathscr{D})$".

Let $\tilde{O}_n$ be the group of matrices $K = (\mathbf{k}_1, \ldots, \mathbf{k}_n) \in GL_n(\mathbb{R})$ whose columns $\mathbf{k}_1, \ldots, \mathbf{k}_n$ have $|\mathbf{k}_1| = \cdots = |\mathbf{k}_n| \neq 0$ and inner products $\mathbf{k}_i\mathbf{k}_j = 0$ for $i \neq j$. It is the product of the orthogonal group O_n and the group of nonzero multiples of the

1991 Mathematics Subject Classification: 11H06, 11H99, 11H55
Key words: Distribution of lattices, generalized upper half plane
* Supported in part by NSF-DMS-9401426

identity matrix. When $X = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in GL_n(\mathbb{R})$, we may write uniquely

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{k}_1, \\ \mathbf{x}_j &= x_{1j}\mathbf{k}_1 + \cdots + x_{j-1,j}\mathbf{k}_{j-1} + y_j\mathbf{k}_j \quad (2 \leqslant j \leqslant n) \end{aligned} \tag{1.1}$$

where $K = (\mathbf{k}_1, \dots, \mathbf{k}_n) \in \tilde{O}_n$ and where $y_2, \dots, y_n$ are positive. Thus uniquely

$$X = KZ \tag{1.2}$$

where $K \in \tilde{O}_n$ and

$$Z = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1n} \\ 0 & y_2 & \cdots & x_{2n} \\ \cdots & & & \\ 0 & 0 & \cdots & y_n \end{pmatrix} \quad \text{with } y_2, \dots, y_n > 0. \tag{1.3}$$

The matrices Z as in (1.3) make up the generalized upper half-plane $\mathscr{H} = \mathscr{H}_n$. When $Z \in \mathscr{H}$ and $M \in GL_n(\mathbb{R})$, we may write ZM in the form (1.2), i.e., we may write uniquely $ZM = KZ_M$ with $K \in \tilde{O}_n$ and $Z_M \in \mathscr{H}$. Clearly $Z_{MN} = (Z_M)_N$. Thus $GL_n(\mathbb{R})$ acts on $\mathscr{H}$: to M corresponds the map $Z \mapsto Z_M$. In particular $GL_n(\mathbb{Z})$, being a subgroup of $GL_n(\mathbb{R})$, acts on $\mathscr{H}$. By $\mathscr{F}$ we will denote a fundamental domain for the action of $GL_n(\mathbb{Z})$ on $\mathscr{H}$. We will write μ for the invariant measure on $\mathscr{H}$ (invariant under the action of $GL_n(\mathbb{R})$), normalized such that $\mu(\mathscr{F}) = 1$. (This measure will be described explicitly in Section 5).

Now suppose that $1 < n \leqslant m$, and let $\mathbf{x}, \mathbf{x}_1, \dots, \mathbf{k}, \mathbf{k}_1, \dots$ denote column vectors with m entries. Let $\mathscr{X}$ denote the set of n-tuples $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ of linearly independent columns. Let $\mathscr{K}$ consist of $K = (\mathbf{k}_1, \dots, \mathbf{k}_n) \in \mathscr{X}$ having $|\mathbf{k}_1| = \cdots = |\mathbf{k}_n|$ and $\mathbf{k}_i\mathbf{k}_j = 0$ when $i \neq j$. Thus $\mathscr{X}, \mathscr{K}$ generalize $GL_n(\mathbb{R})$, $\tilde{O}_n$ respectively, but for $m > n$ they are no longer groups. Every $X = (\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathscr{X}$ has uniquely (1.1) and (1.2) where $K = (\mathbf{k}_1, \dots, \mathbf{k}_n) \in \mathscr{K}$ and $Z \in \mathscr{H}$.

When Λ is a lattice of rank n in $\mathbb{R}^m$ with basis $X = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ (i.e., a basis with vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$), the general basis will be XM with $M \in GL_n(\mathbb{Z})$. When $X = KZ$ as in (1.2), then $XM = KZM = KLZ_M = K'Z_M$ with $L \in \tilde{O}_n$ and $KL = K' \in \mathscr{K}$. Hence there is a map from lattices of rank n onto the set $\mathscr{H}/GL_n(\mathbb{Z})$ of orbits of $GL_n(\mathbb{Z})$ in $\mathscr{H}$, therefore a map onto a given fundamental domain $\mathscr{F}$. Lattices Λ, Λ' are similar precisely if they possess respective bases X, X' such that the $(n \times n)$-matrices $(\mathbf{x}_i\mathbf{x}_j)$ and $(\mathbf{x}_i'\mathbf{x}_j')$ are proportional, and when $X = KZ$, $X' = K'Z'$ this will happen if and only if $Z = Z'$. Therefore the lattices Λ, Λ' will be similar precisely if they have the same image in $\mathscr{H}/GL_n(\mathbb{Z})$, hence the same image in $\mathscr{F}$. Similarity classes of lattices are parametrized by the elements of a fundamental domain $\mathscr{F}$.

Call $\mathscr{D} \subset \mathscr{H}$ *lean* if no two of its elements have the same image in $\mathscr{H}/GL_n(\mathbb{Z})$. This happens precisely when $\mathscr{D}$ is contained in some fundamental domain $\mathscr{F}$. When $\mathscr{D}$ is lean, let $N(\mathscr{D}, T)$ be the number of lattices $\Lambda \subset \mathbb{Z}^m$ with similarity class corresponding to an element of $\mathscr{D}$, and of determinant $\leqslant T$. (The determinant – also called covolume – of a lattice with basis $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is the square root of $\det(\mathbf{x}_i\mathbf{x}_j)_{1 \leqslant i,j \leqslant n}$.)

For $a > 0$, $b > 0$, let $\mathscr{H}(a,b)$ consist of $Z \in \mathscr{H}$ (as given by (1.3)) with

$$y_{i+1} \geqslant a y_i \quad (1 \leqslant i < n), \tag{1.4}$$

$$|x_{ij}| \leqslant b y_i \quad (1 \leqslant i < j \leqslant n), \tag{1.5}$$

where here and throughout we use the notation

$$\boxed{y_1 = 1} \tag{1.6}$$

(The sets $\mathscr{H}(a,b)$ are related to "Siegel sets"; see [10,§4.4,(4.23)]). We will see in Section 5 that $\mathscr{H}(\sqrt{3/4}, 1/2)$ contains a fundamental domain $\mathscr{F}$. Further let $\mathscr{H}(a,b,c)$ consist of $Z \in \mathscr{H}(a,b)$ with $y_n \leqslant c$. Then $\mathscr{H}(a,b,c)$ is compact. The generalized half-plane $\mathscr{H}_n$ may be identified with the set of points $(x_{12}, y_2, x_{13}, x_{23}, y_3, \ldots, y_n) \in \mathbb{R}^h$ (where $h = 2+3+\cdots+n$) with positive $y_2, \ldots, y_n$. A set $\mathscr{D} \subset \mathscr{H}(a,b)$ is Jordan-measurable if for any c the set $\mathscr{D} \cap \mathscr{H}(a,b,c)$ is Jordan-measurable in $\mathbb{R}^h$, i.e., if its indicator function is Riemann-integrable. In this case $\mu(\mathscr{D})$ exists, and $\mu(\mathscr{D}) = \lim_{c\to\infty} \mu(\mathscr{D} \cap \mathscr{H}(a,b,c))$.

Theorem 1. *Suppose $1 < n \leqslant m$, and let $\mathscr{D} \subset \mathscr{H}(a,b)$ be lean and Jordan-measurable. Then as $T \to \infty$,*

$$N(\mathscr{D}, T) \sim c_1(m,n)\mu(\mathscr{D})T^m \tag{1.7}$$

with

$$c_1(m,n) = \frac{1}{m}\binom{m}{n}\frac{V_{m-n+1}\cdots V_m}{V_1 V_2 \cdots V_n}\zeta(2)\cdots\zeta(n), \tag{1.8}$$

where V_l is the volume of the unit ball in $\mathbb{R}^l$.

Remark. When $\mu(\mathscr{D}) = 0$, (1.7) is to be interpreted as $N(\mathscr{D}, T) = o(T^m)$. This convention applies throughout.

Note that for a fundamental domain $\mathscr{F}$, $N(\mathscr{F}, T) = N(T)$, say, counts all the lattices $\Lambda \subset \mathbb{Z}^m$ of rank n and determinant $\leqslant T$. In this special case the asymptotic formula (1.7) had been proved in [7]. The case $n = 2$ of Theorem 1 had been done by Roelcke [6] and again (in more generality) by Maass [5], using modular forms. I am grateful to William Duke for pointing out these references. Our arguments will be elementary.

The condition in Theorem 1 that $\mathscr{D}$ be lean may be removed if more generally $N(\mathscr{D}, T)$ counts lattices Λ of determinant $\leqslant T$ with multiplicity t if there are t distinct elements $Z_1, \ldots, Z_t$ in $\mathscr{D}$ such that Λ has bases $K_1 Z_1, \ldots, K_t Z_t$ with $K_i \in \mathscr{K}$.

Remark. An open set $\mathscr{D}$ need not be Jordan-measurable. Whereas $\mu(\mathscr{D})$ exists as a Lebesgue integral, (1.7) is not necessarily correct: Let $\mathscr{D}$ be lean and open with $\mu(\mathscr{D}) > 0$. Let $\mathscr{S}$ be the set of points in $\mathscr{D}$ corresponding to lattices $\Lambda \subset \mathbb{Z}^m$ of rank n. For each Z in the denumerable set $\mathscr{S}$, let $\mathscr{U}(Z)$ be an open neighborhood contained in $\mathscr{D}$ such that $\Sigma_{Z\in\mathscr{S}}\mu(\mathscr{U}(Z)) < \mu(\mathscr{D})$. Then $\mathscr{D}' = \cup_{Z\in\mathscr{S}} \mathscr{U}(Z)$ is lean and open with $\mu(\mathscr{D}') < \mu(\mathscr{D})$, yet $N(\mathscr{D}', T) = N(\mathscr{D}, T)$. Therefore (1.7) cannot

hold for both $\mathscr{D}$ and $\mathscr{D}'$. The hypothesis in [6, Satz] that $\mathscr{D}$ be open should be replaced by the condition that $\mathscr{D}$ be Jordan-measurable.

Given a lattice $\Lambda \subset \mathbb{Z}^m$, let $S(\Lambda)$ be the subspace of $\mathbb{R}^m$ spanned by Λ, and $\tilde{\Lambda} = S(\Lambda) \cap \mathbb{Z}^m$. Then $\tilde{\Lambda}$ is a lattice, and $\Lambda \subset \tilde{\Lambda}$ with finite index. The lattice Λ is called *primitive* if $\Lambda = \tilde{\Lambda}$. There is a 1–1 correspondence between primitive lattices and rational subspaces of $\mathbb{R}^m$. Let $P(\mathscr{D}, T)$ where $\mathscr{D}$ is lean be the number of primitive lattices $\Lambda \subset \mathbb{Z}^m$ with similarity class in $\mathscr{D}$, and determinant $\leqslant T$.

Theorem 2. *Suppose $1 < n < m$, and let $\mathscr{D}$ be as in Theorem 1. Then as $T \to \infty$,*

$$P(\mathscr{D}, T) \sim c_2(m,n)\mu(\mathscr{D})T^m \tag{1.9}$$

with

$$c_2(m,n) = \frac{1}{m}\binom{m}{n}\frac{V_{m-n+1}\cdots V_m}{V_1 V_2 \cdots V_n} \cdot \frac{\zeta(2)\cdots\zeta(n)}{\zeta(m-n+1)\cdots\zeta(m)}. \tag{1.10}$$

Again, the case when $\mathscr{D}$ is a fundamental domain had been proved in [7].

Let $\mathscr{G}_n^m$ be the Grassmann variety of n-dimensional subspaces of $\mathbb{R}^m$. Let ν be the invariant (under the action of the orthogonal group O_m) measure on $\mathscr{G}_n^m$, normalized so that $\nu(\mathscr{G}_n^m) = 1$. When $\mathscr{D} \subset \mathscr{H}(a,b)$ is lean and $\mathscr{E} \subset \mathscr{G}_n^m$, let $N(\mathscr{D}, \mathscr{E}, T)$ be the number of lattices $\Lambda \subset \mathbb{Z}^m$ of determinant $\leqslant T$ with similarity class in $\mathscr{D}$ and $S(\Lambda) \in \mathscr{E}$.

Theorem 3. *Suppose $1 < n < m$, $\mathscr{D} \subset \mathscr{H}(a,b)$ is lean and Jordan-measurable and $\mathscr{E} \subset \mathscr{G}_n^m$ is Jordan-measurable.* Then as $T \to \infty$,*

$$N(\mathscr{D}, \mathscr{E}, T) \sim c_1(m,n)\mu(\mathscr{D})\nu(\mathscr{E})T^m.$$

This generalizes Theorem 1. The case $n = 2$ is due to MAASS [5]. A similar result could be proved for the number $P(\mathscr{D}, \mathscr{E}, T)$ (defined in an obvious manner) which counts only primitive lattices.

Our theorems could very easily be generalized to the situation where $\mathbb{Z}^m$ is replaced by a fixed lattice $\Gamma \subset \mathbb{R}^m$ of rank m and determinant 1, and where, e.g., $N(\mathscr{D}, T)$ is replaced by the number $N_\Gamma(\mathscr{D}, T)$ of sublattices Λ of Γ belonging to $\mathscr{D}$ and having determinant $\leqslant T$. The asymptotic formulas remain the same.

2. The Error Term

For a wide class of domains $\mathscr{D}$ we will prove (1.7) with an error term $\ll T^{m-1/n}$.

Generalizing $\mathscr{H}(a,b)$, we define $\mathscr{H}(\mathbf{a},b) = \mathscr{H}(a_1,\ldots,a_{n-1},b)$ for positive $a_1,\ldots,a_{n-1},b$ to consist of $Z \in \mathscr{H}$ with

$$y_{i+1} \geqslant a_i y_i \quad (1 \leqslant i < n), \tag{2.1}$$

$$|x_{ij}| \leqslant b y_i \quad (1 \leqslant i < j \leqslant n), \tag{2.2}$$

* That is, for every $\varepsilon > 0$ there are continuous functions f_1, f_2 on $\mathscr{G}_n^m$ with $f_1 \leqslant \iota \leqslant f_2$ where ι is the indicator function of $\mathscr{E}$ having $\int (f_2(S) - f_1(S))d\nu(S) < \varepsilon$.

where again we set $y_1 = 1$. We put

$$\Pi(\mathbf{a}) = \prod_{i=1}^{n-1} a_i^{-(i-1/n)(n-i)}. \tag{2.3}$$

Recall that $\mathscr{H}_n$ may be considered to lie in $\mathbb{R}^h$ with $h = 2 + 3 + \cdots + n$. A set $\mathscr{B} \subset \mathbb{R}^h$ will be called γ-*special* if it is in the image of a map $\Psi : \mathscr{A} \to \mathbb{R}^h$ where $\mathscr{A} \subset \mathbb{R}^{h-1}$ and where

$$|\boldsymbol{\alpha} - \boldsymbol{\alpha}'| \leqslant |\Psi(\boldsymbol{\alpha}) - \Psi(\boldsymbol{\alpha}')| \leqslant \gamma|\boldsymbol{\alpha} - \boldsymbol{\alpha}'| \tag{2.4}$$

for $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathscr{A}$. The norms here are the Euclidean norms in $\mathbb{R}^{h-1}$ and $\mathbb{R}^h$. A set $\mathscr{B}$ will be called (t, γ)-special if it is the union of t γ-special sets. A subset of such a set is again (t, γ)-special.

Theorem 4. *Suppose* $\gamma \geqslant 1$,

$$a_i \geqslant a > 0 \quad (1 \leqslant i < n), \tag{2.5}$$

and $\mathscr{D} \subset \mathscr{H}(\mathbf{a}, b)$ *is lean, and has* (t, γ)*-special boundary* $\partial\mathscr{D}$. *Then*

$$N(\mathscr{D}, T) = c_1(m, n)\mu(\mathscr{D})T^m + O(t\gamma^{3h}\Pi(\mathbf{a})T^{m-1/n}).$$

Here and throughout, constants implied in $O(\cdots)$ and $\ll$ depend only on parameters m, n, a, b. The length of the paper stems partially from the fact that not only do we insist on an error term $\ll T^{m-1/n}$, but on a factor such as $\Pi(\mathbf{a})$ in the error term, which becomes small when some a_i $(1 \leqslant i < n)$ is large.

Let $\lambda_1, \ldots, \lambda_n$ be the successive minima in the sense of Minkowski of a lattice Λ of rank n. Given numbers $a_i \geqslant 1$ $(1 \leqslant i < n)$, the lattices with

$$\lambda_{i+1}/\lambda_i \geqslant a_i \tag{2.6}$$

make up a set $\mathscr{D}(\mathbf{a}) = \mathscr{D}(a_1, \ldots, a_{n-1})$ of similarity classes.

Theorem 5. $\mathscr{D} = \mathscr{D}(\mathbf{a})$ *may be realized as a lean, Jordan-measurable subset of* $\mathscr{H}$. *We have*

(*i*)
$$\mu(\mathscr{D}) \gg\ll \prod_{i=1}^{n-1} a_i^{-i(n-i)},$$

(*ii*)
$$N(\mathscr{D}, T) = c_1(m, n)\mu(\mathscr{D})T^m + O\Big(\prod(\mathbf{a})T^{m-1/n}\Big).$$

The constant in $O(\cdots)$ *depends on m only.*

In particular, when some a_i is large, the proportion of lattices with (2.6) is small. This answers a question of E. Burger. When $n = 2$ and $\mathscr{D} = \mathscr{D}(a_1)$, then (as seen in Section 9) $\mu(\mathscr{D}) = (6/\pi)\arcsin(1/2a_1)$.

When $1 \leqslant i < n$ and $\alpha > 0$, set

$$S(i, \alpha, T) = \sum \left(\frac{\lambda_{i+1}}{\lambda_i}(\Lambda)\right)^{\alpha},$$

where the sum is over lattices $\Lambda \subset \mathbb{Z}^m$ of rank n and determinant $\leqslant T$. One may deduce from Theorem 5 by partial integration that

$$S(i, \alpha, T) \sim c_3(m, n, i, \alpha) T^m$$

as $T \to \infty$ when $\alpha < i(n-i)$, but that

$$S(i, \alpha, T)/T^m \to \infty$$

when $\alpha \geqslant i(n-i)$.

When $n = 2$, write $x_{12} = x$, $y_2 = y$ for convenience. The elements $Z \in \mathscr{H}$ with

$$0 \leqslant x \leqslant 1/2, \quad x^2 + y^2 \geqslant 1 \tag{2.7}$$

make up a fundamental domain $\mathscr{F}_1$. (Z corresponds to points $x + iy$ in the classical upper half-plane. Our $\mathscr{F}_1$, being a fundamental domain for the action of $GL_2(\mathbb{Z})$, is roughly half a more familiar fundamental domain for the action of $SL_2(\mathbb{Z})$.) We will consider domains $\mathscr{D} \subset \mathscr{F}_1$ whose boundary is a simple rectifiable curve, parametrized as $(x(\alpha), y(\alpha))$ where α signifies the arc length. We will suppose that α runs through I, where either I is an interval, say $I = [u, v]$, and $(x(u), y(u)) = (x(v), y(v))$, so that the curve is closed, or $I = \mathbb{R}$ and $y(\alpha) \to \infty$, as $|\alpha| \to \infty$. The possibilities for $\mathscr{D}$ are indicated by the following figures.

We will suppose that the integral in

$$c_4(\mathscr{D}) = \int_I y(\alpha)^{-3/2} d\alpha + \left(\min_\alpha y(\alpha)\right)^{-3/2} \tag{2.8}$$

is finite. Note that $c_4(\mathscr{D})$ is $\ll$ the integral when I has length $\geqslant 1$.

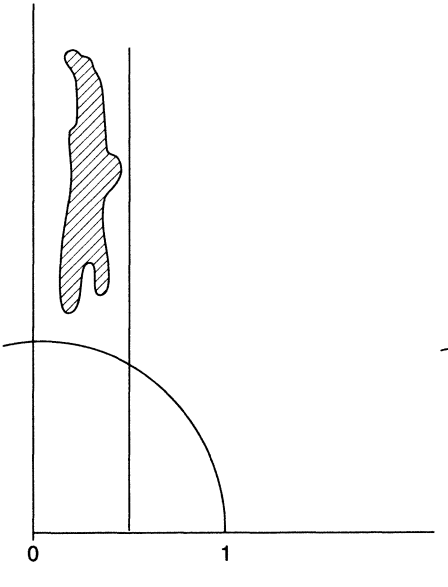


Figure 1

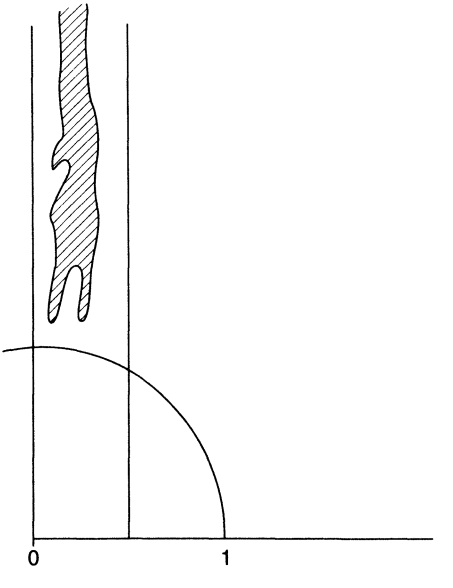


Figure 2

Theorem 6. *Under the above assumption*

$$N(\mathscr{D}, T) = c_1(m,2)\mu(\mathscr{D})T^m + O(c_4(\mathscr{D})T^{m-1/2}),$$

where the constant in $\ll$ depends on m only.

3. The Plan of the Paper

In the next two sections we will discuss the invariant measure μ and fundamental domains. While these are essentially known, they are usually stated in the language of quadratic forms rather than the generalized upper half-plane. Our exposition will be self-contained, requiring no prior knowledge of the subject.

Next, we will derive consequences of Theorem 4. We will derive Theorem 1 in Section 6, Theorem 2 in Sections 7, 8, Theorem 5 in Section 9.

In Section 10 we will formulate Theorem 7, which gives a better understanding of the error term. Roughly speaking, a point Z on the boundary of $\mathscr{D}$ contributes with weight $y_2^{-2+1/n} \cdots y_n^{-n+1/n}$ to the error. Theorem 6 is an almost immediate consequence of Theorem 7, and Theorem 4 will be derived in Section 11. The implications between the theorems are as follows.

$$\begin{array}{ccccccc} 7 & \longrightarrow & 4 & \longrightarrow & 1 & \longrightarrow & 2 \\ & \searrow & & \searrow & & & \\ & & 6 & & 5 & & \end{array}$$

Finally in sections 12–17 we prove Theorem 7. In the last section, Section 18, we will outline a proof of Theorem 3.

Some of our notations are as follows. O_m is the orthogonal group, $\tilde{O}_m$ the orthogonal group composed with nonzero multiplications.

$GL_m(\mathbb{R})$	$\tilde{O}_m$	O_m
$\mathscr{X}$	$\mathscr{K}$	$\mathscr{O}$

$\mathscr{X}, \mathscr{K}, \mathscr{O}$ consist respectively of $(m \times n)$-matrices $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ which can be extended to matrices $(\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_m)$ in $GL_m(\mathbb{R})$, $\tilde{O}_m$, O_m.

$\mathscr{D}$ is a subset of the upper half-plane $\mathscr{H}$. Further $\mathscr{D}^{(K)}, \mathscr{D}^{[K]}$ consist respectively of points in $\mathscr{D}$ with $y_2 \cdots y_n \leqslant K$ and $y_2 \cdots y_n = K$. On the other hand, $\mathscr{D}^*$ consists of $X \in \mathscr{X}$ of the form $X = KZ$ with $K \in \mathscr{K}$, $Z \in \mathscr{D}$, so that $\mathscr{D}^* \subset \mathbb{R}^{mn}$. We will write $[\mathscr{D}]$ for the set of lattices $\Lambda \subset \mathbb{Z}^m$ with a basis KZ where $K \in \mathscr{K}$, $Z \in \mathscr{D}$, and $[\mathscr{D}, T]$ for the lattices in $[\mathscr{D}]$ of determinant $\leqslant T$.

When $\mathbf{x} \in \mathbb{R}^m$, we set $C_\delta(\mathbf{x})$ for the closed cube centered at $\mathbf{x}$ with sides of length δ, and parallel to the coordinate axes. The cubes $C_\delta(\mathbf{x})$ with $\mathbf{x} \in \delta\mathbb{Z}^m$ cover $\mathbb{R}^m$, and their interiors are disjoint. It will be convenient to set $\langle \mathbf{x} \rangle = C_1(\mathbf{x})$. A map Ψ between subsets of Euclidean spaces is γ-*Lipschitz* if

$$|\Psi(\boldsymbol{\alpha}) - \Psi(\boldsymbol{\alpha}')| \leqslant \gamma |\boldsymbol{\alpha} - \boldsymbol{\alpha}'| \tag{3.1}$$

for $\boldsymbol{\alpha}, \boldsymbol{\alpha}'$ in its domain. Given a set $\mathscr{A}$ in Euclidean space we write $\mathscr{A}_\varepsilon$ for the "rounded" set consisting of points $\boldsymbol{\alpha} + \boldsymbol{\beta}$ with $\boldsymbol{\alpha} \in \mathscr{A}$ and $|\boldsymbol{\beta}| < \varepsilon$. The volume of the unit ball in $\mathbb{R}^l$ is denoted by V_l, and

$$h(t) = -1 + (1 + 2 + \cdots + t). \tag{3.2}$$

The sets $\mathscr{H}(a,b)$ and $\mathscr{H}(\mathbf{a},b)$ have been defined by (1.4), (1.5) and by (2.1), (2.2). We will assume throughout that (2.5) holds, which entails $\mathscr{H}(\mathbf{a},b) \subset \mathscr{H}(a,b)$.

4. The Invariant Measure

With Z given by (1.3), let $\mathfrak{z}_j$ for $2 \leqslant j \leqslant n$ be the column vector with entries $x_{1j}, \dots, x_{j-1,j}, y_j$, so that $\mathfrak{z}_j$ lies in $\mathbb{R}^j$. By abuse of notation we will write $Z = (\mathfrak{z}_2, \dots, \mathfrak{z}_n)$. When Z' or Z^* or Z° lie in $\mathscr{H}$, we will write x'_{ij}, y'_i or x^*_{ij}, y^*_i or x°_{ij}, y°_i for their respective entries, and we define vectors $\mathfrak{z}'_j$ or $\mathfrak{z}^*_j$ or $\mathfrak{z}^\circ_j$ accordingly. $Z = (\mathfrak{z}_2 \dots, \mathfrak{z}_n)$ may be interpreted as a point in $\mathbb{R}^h$ where

$$h = h(n) = 2 + 3 + \cdots + n. \tag{4.1}$$

Write $d\mathfrak{z}_j = dx_{1j} \cdots dx_{j-1,j} dy_j$ $(2 \leqslant j \leqslant n)$. Let μ_0 be the measure on $\mathscr{H}$ with volume element

$$\frac{d\mathfrak{z}_2}{y_2^2} \cdots \frac{d\mathfrak{z}_n}{y_n^n}. \tag{4.2}$$

Lemma 1. *μ_0 is invariant under the action of $GL_n(\mathbb{R})$.*

Proof. $GL_n(\mathbb{R})$ is generated by

(a) diagonal matrices diag $(1, \dots, 1, \lambda, 1, \dots, 1)$ where $\lambda \neq 0$ is in the l-th row with $1 < l \leqslant n$.

(b) matrices

$$\begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ & \ddots & & & & & \\ 0 & & 1 & \dots & \alpha & \dots & 0 \\ & & & \ddots & & & \\ 0 & & 0 & & 1 & \dots & 0 \\ & & & & & \ddots & \\ 0 & & 0 & & 0 & & 1 \end{pmatrix}$$

where α is in the i-th row and j-th column with $1 \leqslant i < j \leqslant n$.

(c) permutation matrices.

Multiplication of $Z \in \mathscr{H}$ on the right by a matrix M of type (a) has the effect of replacing $\mathfrak{z}_l$ by $\lambda \mathfrak{z}_l$, so that in Z_M one has to replace $\mathfrak{z}_l$ by $|\lambda| \mathfrak{z}_l$. Both $d\mathfrak{z}_l$ and y_l^l are multiplied by $|\lambda^l|$, so that the volume element (4.2) remains unchanged. If we multiply Z by a matrix (b), then $\mathfrak{z}_j$ is replaced by $\mathfrak{z}_j + \alpha \mathfrak{z}'_i$ (where $\mathfrak{z}'_i$ is $\mathfrak{z}_i$ augmented by $j - i$ zeros, so that $\mathfrak{z}'_i \in \mathbb{R}^j$). Since $i < j$, both $d\mathfrak{z}_j$ (for fixed $\mathfrak{z}_i$) and y_j are unchanged.

To deal with (c) we proceed as follows. When $X \in GL_n(\mathbb{R})$, let X^t be its transpose, and set $[X] = X^t X$. Then $[X]$ is symmetric and positive definite, i.e., the quadratic form with coefficient matrix $[X]$ is positive definite. Further $[X_1] = [X_2]$ precisely if $X_1 \in O_n X_2$ were O_n is the orthogonal group. When S is symmetric and positive definite, set $S' = c^{-1} S$, where c is the left upper entry of S; then S' is symmetric and positive definite, with 1 in the left upper corner. Let $\mathscr{S}$ be the set of

such matrices. We have $[X]' \in \mathscr{S}$, and $[X_1]' = [X_2]'$ precisely when $X_1 \in \tilde{O}_n X_2$. The matrices in $\mathscr{S}$ are in 1–1-correspondence with the orbits of $\tilde{O}_n$ in $GL_n(\mathbb{R})$, hence in 1–1-correspondence with points $Z \in \mathscr{H}$. When $Z \in \mathscr{H}$ corresponds to $S = [Z]'$, then Z_M (which lies in the same orbit of $\tilde{O}_n$ as ZM) corresponds to $[ZM]' = c^{-1}M^tSM$ where c is the entry in the left upper corner of M^tSM.

A matrix $S = (s_{ij}) \in \mathscr{S}$ is determined by the $h = h(n)$ entries

$$s_{12}, \ldots, s_{1n}, s_{22}, s_{23}, \ldots, s_{2n}, \ldots, s_{nn}. \tag{4.3}$$

The l-th entry in (4.3) where $1 \leqslant l \leqslant h$ depends only on the first l entries in

$$x_{12}, \ldots, x_{1n}, y_2, x_{23}, \ldots, x_{2n}, \ldots, y_n, \tag{4.4}$$

so that the Jacobian

$$\begin{aligned}\frac{\partial(s_{12}, \ldots, s_{nn})}{\partial(x_{12}, \ldots, y_n)} &= \left(\frac{\partial s_{12}}{\partial x_{12}} \cdots \frac{\partial s_{1n}}{\partial x_{1n}}\right)\left(\frac{\partial s_{22}}{\partial y_2}\frac{\partial s_{23}}{\partial x_{23}} \cdots \frac{\partial s_{2n}}{\partial x_{2n}}\right) \cdots \left(\frac{\partial s_{nn}}{\partial y_n}\right) \\ &= (1)(2y_2^{n-1})(2y_3^{n-2}) \cdots (2y_n) = 2^{n-1}y_2^{n-1}y_3^{n-2} \cdots y_n.\end{aligned}$$

Therefore the volume element (4.2) becomes

$$2^{1-n}(y_2y_3 \cdots y_n)^{-n-1} ds_{12} \cdots ds_{nn} = 2^{1-n}(\det S)^{-(n+1)/2} ds_{12} \cdots ds_{nn}. \tag{4.5}$$

(See also [10, §4.1, (1.16)] or [8, p. 58].)

It will suffice to consider the permutation matrices M_l $(2 \leqslant l \leqslant n)$ whose action (by multiplication on the right) interchanges the 1st and l-th columns. When $[Z]' = S$, the matrix $M_l^tSM_l$ is obtained from S by interchanging the 1st and the l-th columns, and the 1st and l-th rows. Therefore $[ZM_l]' = s_{ll}^{-1}[ZM_l] = s_{ll}^{-1}M_l^tSM_l = S'$, say. In the transition from S to S', the entries (4.3) are replaced by

$$1/s_{ll}, s_{12}/s_{ll}, \ldots, \widehat{s_{ll}/s_{ll}}, \ldots, s_{nn}/s_{ll}$$

in some order, where the hatted entry is deleted. This substitution has Jacobian $s_{ll}^{-2}s_{ll}^{-(h-1)} = s_{ll}^{-(h+1)}$. Further

$$\begin{aligned}(\det S')^{-(n+1)/2} ds'_{12} \cdots ds'_{nn} &= s_{ll}^{n(n+1)/2}(\det S)^{-(n+1)/2} s_{ll}^{-h-1} ds_{12} \cdots ds_{nn} \\ &= (\det S)^{-(n+1)/2} ds_{12} \cdots ds_{nn}.\end{aligned}$$

Thus the volume element (4.5) is invariant under replacing S by S'.

It turns out that a fundamental domain $\mathscr{F}$ (for the action of $GL_n(\mathbb{Z})$) has

$$\mu_0(\mathscr{F}) = \frac{2}{(n-1)!}\frac{\zeta(2) \cdots \zeta(n)}{V_1 \cdots V_n} = c_5(n), \tag{4.6}$$

say. Therefore the invariant measure μ with $\mu(\mathscr{F}) = 1$ is given by

$$\mu = c_5(n)^{-1}\mu_0. \tag{4.7}$$

We will give three arguments for (4.6), the first one depending on a well known computation as presented in [10], the second one depending on my earlier work [7], the third one depending only on the present work.

Let $\mathscr{M}_n^*$ be the image of a fundamental domain $\mathscr{F}$ under the map $Z \mapsto [Z]'$, interpreted as a subset of $\mathbb{R}^h$, and $\mathscr{M}_n$ the cone over $\mathscr{M}_n^*$ consisting of multiples tS with $t > 0$, $S \in \mathscr{M}_n^*$. By [10, §4.4, Theorem 4], the volume of the set of $Y \in \mathscr{M}_n$ with $\det Y \leqslant 1$ is $V = (2^n/n(n+1))c_5(n)$. On the other hand with $Y = tS$ and $S \in \mathscr{S}$, the Jacobian $\partial Y/\partial(t,S) = t^h$. Further $t^n \det S = \det Y \leqslant 1$ gives $t \leqslant t_0 := (\det S)^{-1/n}$, and $\int_0^{t_0} t^h dt = (2/n(n+1))(\det S)^{-(n+1)/2}$, so that

$$V = (2/n(n+1)) \int_{\mathscr{M}_n^*} (\det S)^{-(n+1)/2} dS = 2^{n-1}(2/n(n+1))\mu_0(\mathscr{F})$$

by (4.5). Comparison of our formulae gives (4.6).

Next, let $N(T)$ be the number of sublattices of $\mathbb{Z}^m$ of rank n and determinant $\leqslant T$, so that $N(T) = N(\mathscr{F}, T)$ for a fundamental domain $\mathscr{F}$. In Section 6 we will show that $N(T) \sim (2mn)^{-1} c_6(m,n)\mu_0(\mathscr{F})T^m$ with

$$c_6(m,n) = \prod_{l=m-n+1}^{m} (lV_l). \tag{4.8}$$

But in [7] I had shown that $N(T) \sim c_1(m,n)T^m$. Comparison gives $\mu_0(\mathscr{F}) = 2mnc_1(m,n)/c_6(m,n) = c_5(n)$.

Finally, let $P(T) = P_{mn}(T)$ be the number of primitive sublattices of $\mathbb{Z}^m$ of rank n and determinant $\leqslant T$. In Section 8 we will see that for $1 \leqslant n < m$

$$P_{mn}(T) \sim (c_6(m,n)/(2mn\zeta(m-n+1)\cdots\zeta(m)))\mu_0(\mathscr{F})T^m. \tag{4.9}$$

To a primitive lattice $\Lambda \subset \mathbb{Z}^m$ of rank n corresponds the primitive lattice $\Lambda^\perp$ of rank $m-n$ consisting of integer points orthogonal to Λ, and it is easily seen that $\det \Lambda^\perp = \det \Lambda$ (see, e.g., [7, Corollary to Lemma 1]). Therefore $P_{mn}(T) = P_{m,m-n}(T)$; in particular $P_{n+1,n}(T) = P_{n+1,1}(T)$, and the latter is half the number of primitive integer points in $\mathbb{R}^{n+1}$ of norm $\leqslant T$, so that it is well known to be

$$\sim (2\zeta(n+1))^{-1} V_{n+1} T^{n+1}.$$

Comparison with the case $m = n+1$ of (4.9) again yields $\mu_0(\mathscr{F}) = c_5(n)$.

A matrix $Z \in \mathscr{H}$ may also be written as

$$Z = \begin{pmatrix} 1 & \xi_{12} & \xi_{13} & \dots & \xi_{1n} \\ 0 & \eta_2 & \eta_2\xi_{23} & \dots & \eta_2\xi_{2n} \\ 0 & 0 & \eta_2\eta_3 & \dots & \eta_2\eta_3\xi_{3n} \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & \eta_2\eta_3\cdots\eta_n \end{pmatrix} \tag{4.10}$$

with positive $\eta_2, \dots, \eta_n$. The l-th entry in (4.4), where $1 \leqslant l \leqslant h$, depends only on the first l entries among

$$\xi_{12}, \dots, \xi_{1n}, \eta_2, \xi_{23}, \dots, \zeta_{2n}, \dots, \eta_n.$$

It is therefore easily checked that the Jacobian

$$\frac{\partial(x_{12},\dots,y_n)}{\partial(\xi_{12},\dots,\eta_n)} = y_2^{n-1}y_3^{n-2}\cdots y_{n-1}^2,$$

and the volume element (4.2) becomes

$$\left(\prod_{i=1}^{n-1}\frac{d\eta_{i+1}}{\eta_{i+1}^{1+i(n-i)}}\right)d\xi_{12}\cdots d\xi_{n-1,n}. \tag{4.11}$$

Lemma 2.

$$\mu(\mathscr{H}(\mathbf{a},b)) \gg\ll \mu_0(\mathscr{H}(\mathbf{a},b)) \gg\ll \prod_{i=1}^{n-1} a_i^{-i(n-i)}.$$

Proof. In terms of the variables in (4.10), $\mathscr{H}(\mathbf{a},b)$ is given by

$$\begin{aligned} |\xi_{ij}| &\leqslant b \qquad (1\leqslant i<j\leqslant n),\\ \eta_{i+1} &\geqslant a_i \qquad (1\leqslant i<n). \end{aligned}$$

Integration over the ξ_{ij} gives $(2b)^{n(n-1)/2}$, which is $\gg\ll 1$, since the implicit constants may depend on n, b. By (4.11),

$$\mu_0(\mathscr{H}(\mathbf{a},b) \gg\ll \prod_{i=1}^{n-1}\left(\int_{a_i}^{\infty}\frac{d\eta_{i+1}}{\eta_{i+1}^{1+i(n-i)}}\right).$$

The lemma follows.

Let μ_1 be the measure on $\mathscr{H}$ with volume element

$$\frac{d\xi_{12}}{y_2^{2-1/n}}\cdots\frac{d\xi_{1n}}{y_n^{n-1/n}},$$

which is

$$\left(\prod_{i=1}^{n-1}\frac{d\eta_{i+1}}{\eta_{i+1}^{i(n-i)+i/n}}\right)d\xi_{12}\cdots d\xi_{n-1,n}$$

in terms of the coordinates in (4.10). This measure is not invariant. An argument as in the proof of Lemma 2 yields

$$\mu_1(\mathscr{H}(\mathbf{a},b)) \gg\ll \prod(\mathbf{a}) \tag{4.12}$$

with $\Pi(\mathbf{a})$ given by (2.3).

5. Fundamental Domains

Fundamental domains are best constructed by geometric arguments. Suppose Λ is a lattice of rank n and λ_1 its first minimum, i.e., the least $\lambda_1 > 0$ such that there is a lattice point $\mathbf{x}\in\Lambda$ with $|\mathbf{x}| = \lambda_1$. The following is well known.

Lemma 3. *The number of* $\mathbf{x}\in\Lambda$ *with* $|\mathbf{x}| = \lambda_1$ *is between 2 and* $c_7(n)$.

When $\Lambda \subset \mathbb{R}^m$ is of rank n, pick $\mathbf{x}_1 \in \Lambda$ with minimal nonzero norm, i.e., with $|\mathbf{x}_1| = \lambda_1$. The number of choices for $\mathbf{x}_1$ is between 2 and $c_7(n)$. Set $\mathbf{x}_1 = \mathbf{k}_1$. Write points $\mathbf{x} \in \mathbb{R}^m$ as $\mathbf{x} = x_{12}\mathbf{k}_1 + \mathbf{x}'$ where $\mathbf{x}'$ is orthogonal to $\mathbf{k}_1$. As $\mathbf{x}$ ranges through Λ, then $\mathbf{x}'$ ranges through a lattice Λ' of rank $n-1$ in the space orthogonal to $\mathbf{k}_1$. Pick a nonzero lattice point $\mathbf{h}' \in \Lambda'$ with minimal norm. The number of choices for $\mathbf{h}'$ is between 2 and $c_7(n-1)$. There is a lattice point $\mathbf{x}_2 = x_{12}\mathbf{k}_1 + \mathbf{h}'$ in Λ. By adding a suitable integer multiple of $\mathbf{x}_1 = \mathbf{k}_1$, we obtain a point with $-1/2 < x_{12} \leqslant 1/2$. Given $\mathbf{h}'$, this point is unique. Replacing $\mathbf{x}_2$, $\mathbf{h}'$ by $-\mathbf{x}_2$, $-\mathbf{h}'$ if necessary, we obtain a lattice point $\mathbf{x}_2 = x_{12}\mathbf{k}_1 + \mathbf{h}'$ with $0 \leqslant x_{12} \leqslant 1/2$. It is easily seen that except when $x_{12} = 0$ or $1/2$, the number of choices of $\mathbf{x}_2$ is between 1 and $\frac{1}{2}c_7(n-1)$. (The $1/2$ since only one of $\mathbf{h}'$, $-\mathbf{h}'$ works.) In every case, the number of choices is between 1 and $c_7(n-1)$. We now write $\mathbf{h}' = y_2\mathbf{k}_2$ with $y_2 > 0$ and $|\mathbf{k}_2| = |\mathbf{k}_1|$, so that $\mathbf{x}_2 = x_{12}\mathbf{k}_1 + y_2\mathbf{k}_2$.

Next, write points $\mathbf{x} \in \mathbb{R}^m$ as $\mathbf{x} = x_{13}\mathbf{k}_1 + x_{23}\mathbf{k}_2 + \mathbf{h}''$ where $\mathbf{h}''$ is orthogonal to both $\mathbf{k}_1, \mathbf{k}_2$. As $\mathbf{x}$ runs through Λ, then $\mathbf{h}''$ runs through a lattice Λ'' of rank $n-2$. Pick $\mathbf{h}'' \in \Lambda'' \backslash \{\mathbf{0}\}$ with $|\mathbf{h}''|$ minimal. There will be points $\mathbf{x}_3 \in \Lambda$ of the form $\mathbf{x}_3 = x_{13}\mathbf{k}_1 + x_{23}\mathbf{k}_2 + \mathbf{h}''$. By adding suitable multiples of $\mathbf{x}_1, \mathbf{x}_2$, we obtain a lattice point with $-1/2 < x_{13} \leqslant 1/2$, $-y_2/2 < x_{23} \leqslant y_2/2$. By changing $\mathbf{x}_3$, $\mathbf{h}''$ into $-\mathbf{x}_3$, $-\mathbf{h}''$ if necessary, we obtain a lattice point $\mathbf{x}_3 = x_{13}\mathbf{k}_1 + x_{23}\mathbf{k}_2 + \mathbf{h}''$ with $0 \leqslant x_{13} \leqslant 1/2$, $|x_{23}| \leqslant y_2/2$. Setting $\mathbf{h}'' = y_3\mathbf{h}_3$ with $y_3 > 0$, $|\mathbf{k}_3| = |\mathbf{k}_1|$, we have $\mathbf{x}_3 = x_{13}\mathbf{k}_1 + x_{23}\mathbf{k}_2 + y_3\mathbf{k}_3$.

And so forth. We obtain lattice points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with (1.1), where $K = \{(\mathbf{k}_1, \ldots, \mathbf{k}_n) \in \mathscr{K}$. The number of choices for $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is between 2 and $c_8(n)$. Clearly X is a basis of Λ, and $X = KZ$ (i.e., (1.2)) where Z given by (1.3) lies in $\mathscr{H}$. We have*

$$0 \leqslant x_{1j} \leqslant 1/2 \quad (2 \leqslant j \leqslant n), \tag{5.1}$$

$$|x_{ij}| \leqslant y_i/2 \quad (2 \leqslant i < j \leqslant n). \tag{5.2}$$

Furthermore, $\mathbf{x}_i$ for $1 \leqslant i \leqslant n$ has the property P_i that among lattice points outside the space generated by $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$ (i.e., by $\mathbf{k}_1, \ldots, \mathbf{k}_{i-1}$), it has minimal distance from this space.

Let us discuss the property P_i. For a basis as in (1.1), P_n is automatically satisfied. For $1 \leqslant i < n$, it means that every lattice point

$$u_i\mathbf{x}_i + u_{i+1}\mathbf{x}_{i+1} + \cdots + u_n\mathbf{x}_n$$

with $(u_i, u_{i+1}, \ldots, u_n) \in \mathbb{Z}^{n-i+1}$ and $(u_{i+1}, \ldots, u_n) \neq (0, \ldots, 0)$ has distance $\geqslant y_i|\mathbf{k}_i| = y_i|\mathbf{k}_1|$ from the space generated by $\mathbf{k}_1, \ldots, \mathbf{k}_{i-1}$; here we use again the notation $y_1 = 1$ as in (1.6). The condition is that

$$\begin{aligned}(u_iy_i + u_{i+1}x_{i,i+1} + \cdots + u_nx_{in})^2 + (u_{i+1}y_{i+1} + u_{i+2}x_{i+1,i+2} + \cdots + u_nx_{i+1,n})^2 + \\ + \cdots + (u_{n-1}y_{n-1} + u_nx_{n-1,n})^2 + u_n^2y_n^2 \geqslant y_i^2.\end{aligned} \tag{5.3}$$

* Observe that here and throughout, the x_{ij} are not the Cartesian coordinates of the points $\mathbf{x}_j$.

In particular, taking $(u_i, u_{i+1}, \cdots, u_n) = (0, 1, 0, \ldots, 0)$, we obtain $x_{i,i+1}^2 + y_{i+1}^2 \geqslant \geqslant y_i^2$, and since $|x_{i,i+1}| \leqslant y_i/2$, we have

$$y_{i+1} \geqslant \sqrt{3/4}\, y_i \quad (1 \leqslant i < n). \tag{5.4}$$

Here again we have set $y_1 = 1$. Thus

$$y_j \geqslant c_9(n) y_i \quad \text{for } j \geqslant i. \tag{5.5}$$

We claim that the conditions (5.3) have to be checked only for coefficients $u_i, \ldots, u_n$ of modulus $\leqslant c_{10}(n)$. In fact we claim that by (5.1), (5.2), (5.5),

$$\textit{the left-hand side of } (5.3) \textit{ is } \geqslant c_{11}(n) y_i^2 \max(u_i^2, \ldots, u_n^2). \tag{5.6}$$

The left-hand side of (5.3) is $\geqslant c_9^2(n) y_i^2 u_n^2$ by (5.5). One shows by induction, down from $j+1$ to j $(i \leqslant j < n)$, that the left hand side of (5.3) is $\geqslant n^{2j-2n} c_9(n)^2 y_i^2 u_j^2$: If $|u_j| < n \max(|u_{j+1}|, \ldots, |u_n|)$, the induction step is trivial. Otherwise by (5.1), (5.2)

$$\begin{aligned} &|u_j y_j + u_{j+1} x_{j,j+1} + \cdots + u_n x_{jn}| \\ &\quad \geqslant |u_j y_j| - n^{-1}(n-j)|u_j y_j|/2 > |u_j y_j|/2 \geqslant \frac{1}{2} c_9(n) |u_j| y_i, \end{aligned}$$

so that the left hand side of (5.3) is $> 2^{-2} c_9(n)^2 y_i^2 u_j^2$.

Let $\mathscr{F}_1$ be the set of $Z \in \mathscr{H}$ which satisfy (5.1), (5.2), and (5.3) for $1 \leqslant i < n$ and tuples $(u_i, u_{i+1}, \ldots, u_n)$ with $(u_{i+1}, \ldots, u_n) \neq (0, \ldots, 0)$ and $|u_j| \leqslant c_{10}(n)$ $(i \leqslant j \leqslant n)$. We have proved.

Lemma 4. *Every lattice Λ in $\mathbb{R}^m$ of rank n has a basis $X = KZ$ with $K \in \mathscr{K}$ and $Z \in \mathscr{F}_1$.*

Our arguments above show that in the lemma the number of choices for $Z \in \mathscr{F}_1$ and $K \in \mathscr{K}$ is $\leqslant c_{12}(n)$. The interior $\mathscr{F}_1^0$ of $\mathscr{F}_1$ consists of Z satisfying all the relations (5.1), (5.2), (5.3) with strict inequality. We claim that *when Λ has a basis KZ with $K \in \mathscr{K}$ and $Z \in \mathscr{F}_1^0$, then $Z \in \mathscr{F}_1$ is unique, and there are only 2 choices for K, namely K, $-K$.* When the relations (5.3) hold with strict inequality, then in $\mathbf{x}_i = x_{1i}\mathbf{k}_1 + \cdots + x_{i-1,i}\mathbf{k}_{i-1} + y_i \mathbf{k}_i$, the choice of y_i, $\mathbf{k}_i$ is unique, except that perhaps we may replace $\mathbf{k}_i$ by $-\mathbf{k}_i$, hence $\mathbf{x}_i$ by $-\mathbf{x}_i$. When $i = 1$, there are in fact these 2 choices. But when $i > 1$ and (5.1) holds with $x_{1i} \neq 0$, such a replacement is not possible. Also, when $i > 1$, and (5.1), (5.2) hold with strict inequality, then $x_{1i}, \ldots, x_{i-1,i}$ are determined.

Since for bases KZ with $Z \in \mathscr{F}_1^0$ this Z is unique, there are sets $\mathscr{F}$ with $\mathscr{F}_1^0 \subset \mathscr{F} \subset \mathscr{F}_1$ such that for every Λ there is a basis KZ with $K \in \mathscr{K}$ and unique $Z = Z(\Lambda) \in \mathscr{F}$. Further $Z(\Lambda) = Z(\Lambda')$ precisely when Λ, Λ' are similar. Thus $\mathscr{F}$ parametrizes similarity classes of lattices. Since similarity classes of lattices correspond to the orbits of $GL_n(\mathbb{Z})$ in $\mathscr{H}$, we have

Lemma 5. *$\mathscr{F}$ is a fundamental domain for the action of $GL_n(\mathbb{Z})$ on $\mathscr{H}$.*

Except for its boundary, $\mathscr{F}_1$ is a fundamental domain. Our domain $\mathscr{F}_1$ is almost the same as one by GRENIER ([2] or [10, §4.4.3]). Our conditions (5.1), (5.2) could be rewritten as $|x_{ij}| \leqslant y_i/2$ $(1 \leqslant i < j \leqslant n)$ (with $y_1 = 1$) and $x_{1j} \geqslant 0$ $(2 \leqslant j \leqslant n)$. In Grenier's domain, $x_{1j} \geqslant 0$ is replaced by $x_{j-1,j} \geqslant 0$.

When $Z \in \mathscr{H}$, let $G(Z)$ be the group of $M \in GL_n(\mathbb{Z})$ with $Z_M = Z$, and write $g(Z) = |G(Z)|$.

Lemma 6. *$g(Z)$ is invariant under the action of $GL_n(\mathbb{Z})$. When Λ has a basis KZ with $K \in \mathscr{K}$, then it has exactly $g = g(Z)$ bases $K_1Z, \dots, K_gZ$ with $K_1, \dots, K_g \in \mathscr{K}$. We have $2 \leqslant g(Z) \leqslant c_{12}(n)$, and $g(Z) = 2$ when Z lies in the interior of a fundamental domain.*

Proof. $G(Z_N)$, where $N \in GL_n(\mathbb{Z})$, is conjugate to $G(Z)$, so that $g(Z)$ is invariant. If $K_1Z, \dots, K_gZ$ with $K_i \in \mathscr{K}$ are distinct bases of a lattice Λ, then $K_iZ = K_1ZM_i$ with $M_i \in GL_n(\mathbb{Z})$. But $ZM_i = L_iZ_{M_i}$ with $L_i \in \tilde{O}_n$, so that $K_iZ = K_i^*Z_{M_i}$ with $K_i^* = K_1L_i \in \mathscr{K}$. By the uniqueness of the representation (1.2), $Z = Z_{M_i}$, so that $M_1, \dots, M_g \in G(Z)$. Conversely when $M_1, \dots, M_g \in G(Z)$, and when KZ with $K \in \mathscr{K}$ is a basis, then since $KZ = KZ_{M_i} = KL_iZM_i = K_iZM_i$ with $L_i \in \tilde{O}_n$, $K_i = KL_i \in \mathscr{K}$, each of $K_1Z, \dots, K_gZ$ is a basis.

From our remarks below Lemma 4 it is now clear that $g(Z) \leqslant c_{12}(n)$, and since $Z_{-I} = Z$ for the identity matrix I, we have $2 \leqslant g(Z) \leqslant c_{12}(n)$. It remains for us to show that when $g(Z) > 2$, then Z cannot lie in the interior of any fundamental domain. In view of the invariance of g, we may suppose that $Z \in \mathscr{F}_1$ (so that by what was said above, $Z \in \partial\mathscr{F}_1$). When $g(Z) > 2$, there is some $M \neq I, -I$ with $Z_M = Z$. Every neighborhood $\mathscr{U}$ of Z intersects $\mathscr{F}_1^0$, hence contains a Z' with $g(Z') = 2$, hence with $Z'_M \neq Z'$. The map $Z' \to Z'_M$ is continuous, so that when Z' tends to Z, then Z'_M tends to $Z_M = Z$. Therefore there are elements Z' with both Z', $Z'_M \in \mathscr{U}$, but $Z' \notin Z'_M$. Thus $\mathscr{U}$ cannot be contained in any fundamental domain, and Z cannot lie in the interior of any fundamental domain.

Remark. The *automorphism group* $A(\Lambda)$ of Λ consists of the orthogonal maps of the space $S(\Lambda)$ which map Λ onto itself. When Λ has basis KZ with $K \in \mathscr{K}$, $Z \in \mathscr{H}$, then $|A(\Lambda)| = |G(Z)| = g(Z)$.

When $n = 2$, it is easily seen that $\mathscr{F}_1$ is given by (2.7), and $\mathscr{F} = \mathscr{F}_1$ is a fundamental domain. Further $g(Z) = 2$ when $Z \in \mathscr{F}_1^0$, and $g(Z) = 4$ when $I \in \partial\mathscr{F}_1$, except that $g(Z) = 8$ when $(x, y) = (0, 1)$ and $g(Z) = 12$ when $(x, y) = (1/2, \sqrt{3}/2)$.

Since $Z_M = Z_{-M}$, the group $PL_n(\mathbb{Z}) = GL_n\mathbb{Z}/\{\pm I\}$ acts on $\mathscr{H}$. Let $\mathscr{F}_{1M}$ be the image of $\mathscr{F}_1$ under the map $Z \mapsto Z_M$. The sets $\mathscr{F}_{1M}$ with $M \in PL_n(\mathbb{Z})$ cover $\mathscr{H}$. The interiors of these sets do not intersect, so that these sets give a tesselation of $\mathscr{H}$. In Section 9 we will prove

Lemma 7. *Given $a > 0$, $b > 0$, only finitely many sets $\mathscr{F}_{1M}$ $(M \in PL_n(\mathbb{Z}))$ intersect $\mathscr{H}(a, b)$.*

Thus $\mathscr{H}(a, b)$ is covered by finitely many tiles of the tesselation.

6. Theorem 4 implies Theorem 1

Lemma 8. *A linear or quadratic surface in $\mathbb{R}^h$ is a $(2h, 3h)$-special set.*

In other words, it is the union of at most $2h$ γ-special sets where $\gamma = 3h$.

Proof. We will use coordinates $(\alpha_1, \ldots, \alpha_h)$ for points of $\mathbb{R}^h$. The property of being γ-special is invariant under rotations and translations. A quadratic surface may be transformed into a surface

$$a_0 + a_l\alpha_l^2 + \cdots + a_h\alpha_h^2 = 0 \tag{6.1}$$

or

$$\alpha_1 + a_l\alpha_l^2 + \cdots + a_h\alpha_h^2 = 0 \tag{6.2}$$

with nonzero $a_l, \ldots, a_h$. For (6.1), it will suffice, without loss of generality, to consider the piece $\mathscr{B}$ with

$$|a_h\alpha_h| = \max(|a_l\alpha_l|, \ldots, |a_h\alpha_h|) \quad \text{and} \quad \alpha_h \geqslant 0. \tag{6.3}$$

There are $\leqslant 2h$ pieces of this type. (6.1), (6.3) restrict $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_{h-1})$ to a certain set $\mathscr{A} \subset \mathbb{R}^{h-1}$. In $\mathscr{B}$ we have

$$\alpha_h = \sqrt{Q(\boldsymbol{\alpha})}$$

where $\boldsymbol{\alpha} \in \mathscr{A}$ and $Q(\boldsymbol{\alpha}) = (-a_h)^{-1}(a_0 + a_l\alpha_l^2 + \cdots + a_{h-1}\alpha_{h-1}^2)$. The map $\Psi : \mathscr{A} \to \mathbb{R}^h$ with $\boldsymbol{\alpha} \mapsto (\boldsymbol{\alpha}, \sqrt{Q(\boldsymbol{\alpha})})$ has $\Psi(\mathscr{A}) = \mathscr{B}$. It satisfies the left inequality in (2.4). Further

$$\begin{aligned}
|\sqrt{Q(\boldsymbol{\alpha})} - \sqrt{Q(\boldsymbol{\alpha}')}| &= |Q(\boldsymbol{\alpha}) - Q(\boldsymbol{\alpha}')|/(\sqrt{Q(\boldsymbol{\alpha})} + \sqrt{Q(\boldsymbol{\alpha}')}) \\
&= \left|\sum_{i=l}^{h-1} a_i(\alpha_i + \alpha_i')(\alpha_i - \alpha_i')\right| / (|a_h|(\alpha_h + \alpha_h')) \\
&\leqslant 2\sum_{i=1}^{h-1} |\alpha_i - \alpha_i'| \\
&\leqslant 2h|\boldsymbol{\alpha} - \boldsymbol{\alpha}'|,
\end{aligned}$$

since $|a_i\alpha_i| \leqslant |a_h\alpha_h|$, $|a_i\alpha_i'| \leqslant |a_h\alpha_h'|$ for $\boldsymbol{\alpha}, \boldsymbol{\alpha}'$ in $\mathscr{A}$. Therefore $\mathscr{B}$ is γ-special with $\gamma = 3h$.

One deals similarly with (6.2). We may restrict our attention to the piece with $|a_h\alpha_h| = \max(1, |a_l\alpha_l|, \ldots, |a_h\alpha_h|)$, $\alpha_h \geqslant 0$, and the piece with $1 = \max(1, |a_l\alpha_l|, \ldots, |a_h\alpha_h|)$. For the last piece we set $\boldsymbol{\alpha} := (\alpha_2, \ldots, \alpha_h)$ and we use the map $\boldsymbol{\alpha} \mapsto (-Q(\boldsymbol{\alpha}), \boldsymbol{\alpha})$ with $Q(\boldsymbol{\alpha}) = a_l\alpha_l^2 + \cdots + a_h\alpha_h^2$.

We now will deduce Theorem 1. Let $\mathscr{F} \subset \mathscr{F}_1$ be a fundamental domain. By Section 5, $\partial\mathscr{F}$ is bounded by a finite number of linear and quadratic surfaces, hence is a special set by Lemma 8. Since $\mathscr{F}$ is lean, Theorem 4 yields

$$N(\mathscr{F}, T) \sim c_1(m, n)\mu(\mathscr{F})T^m = c_1(m, n)T^m. \tag{6.4}$$

A *box* in $\mathscr{H}$ is a set of the type $u_i \leqslant y_i \leqslant v_i (2 \leqslant i \leqslant n)$, $u_{ij} \leqslant x_{ij} \leqslant v_{ij} (1 \leqslant i < j \leqslant n)$, where $u_2, \ldots, u_n$ are positive. When $\mathscr{Q}$ is a box, let $\mathscr{Q}' = \mathscr{Q} \cap \mathscr{F}$ be called a *truncated box*. Then $\partial\mathscr{Q}'$ is a special set, so that by Theorem 4, $N(\mathscr{Q}', T) \sim$ $\sim c_1(m, n)\mu(\mathscr{Q}')T^m$. The same holds for finite unions of truncated boxes. Suppose now that $\mathscr{D} \subset \mathscr{F}$ is bounded, so that $y_n \leqslant c$ for some c and every $Z \in \mathscr{D}$. When $\mathscr{D}$ is Jordan-measurable, there are finite unions $\mathscr{S}_1, \mathscr{S}_2$ of truncated boxes with

$\mathscr{S}_1 \subset \mathscr{D} \subset \mathscr{S}_2$ and

$$\mu(\mathscr{S}_2) - \varepsilon < \mu(\mathscr{D}) < \mu(\mathscr{S}_1) + \varepsilon.$$

It follows easily that

$$N(\mathscr{D}, T) \sim c_1(m,n)\mu(\mathscr{D})T^m. \tag{6.5}$$

More generally, when $\mathscr{D} \subset \mathscr{F}$ is Jordan-measurable, we may at least conclude that

$$N(\mathscr{D}, T) \gtrsim c_1(m,n)\mu(\mathscr{D})T^m.$$

A corresponding estimate holds for the complement $\mathscr{E}$ of $\mathscr{D}$ in $\mathscr{F}$. In view of these estimates, the formula (6.4), and the facts that $\mu(\mathscr{D}) + \mu(\mathscr{E}) = \mu(\mathscr{F}) = 1$ and $N(\mathscr{D}, T) + N(\mathscr{E}, T) = N(\mathscr{F}, T)$, the asymptotic formula (6.5) holds for $\mathscr{D}$.

So far we have supposed that $\mathscr{D} \in \mathscr{F}$. When $\mathscr{D} \subset \mathscr{H}(a,b)$, then by Lemma 7,

$$\mathscr{D} = \bigcup_{M \in \mathscr{M}} \mathscr{D}'(M) \tag{6.6}$$

where $\mathscr{M} \subset PL_n(\mathbb{Z})$ is finite and $\mathscr{D}'(M) = \mathscr{D} \cap \mathscr{F}_M$, where $\mathscr{F}_M$ is the image of $\mathscr{F}$ under $Z \mapsto Z_M$. The sets $\mathscr{D}'(M)$ may intersect in their boundaries. By deleting certain boundary points, (6.6) becomes a disjoint union. Say $\mathscr{D}'(M) = \mathscr{D}(M)_M$ where $\mathscr{D}(M) \subset \mathscr{F}$. Each $\mathscr{D}'(M)$ is Jordan-measurable, hence so is $\mathscr{D}(M)$, and $\mu(\mathscr{D}(M)) = \mu(\mathscr{D}'(M))$. We already known Theorem 1 to hold for $\mathscr{D}(M)$. Therefore,

$$\begin{aligned} N(\mathscr{D}, T) &= \sum_{M \in \mathscr{M}} N(\mathscr{D}'(M), T) = \sum_{M \in \mathscr{M}} N(\mathscr{D}(M), T) \\ &\sim c_1(m,n) \sum_{M \in \mathscr{M}} \mu(\mathscr{D}(M))T^m = c_1(m,n)\mu(\mathscr{D})T^m. \end{aligned}$$

This establishes Theorem 1 in the generalized version where $\mathscr{D}$ need not be lean.

Now actually, Theorem 4 will first be proved in the provisional form that $N(\mathscr{D}, T) \sim (2mn)^{-1}c_6(m,n)\mu_0(\mathscr{D})T^m$ when $\partial\mathscr{D}$ is a special set, so that the deliberations of the present section yield $N(\mathscr{F}, T) \sim (2mn)^{-1}c_6(m,n)T^m$. As explained in Section 4, this gives the formula (4.6) for $\mu_0(\mathscr{F})$, and hence Theorem 4 as stated in Section 2.

7. A Generalized Moebius Function

We define a function $\mu(G)$ on finite abelian groups G by the condition that

$$\sum_{H \subset G} \mu(H) = \begin{cases} 1, & \text{when } G = \{0\}, \\ 0, & \text{otherwise.} \end{cases}$$

The sum here is over the subgroups H of G. The existence and uniqueness of μ is proved by induction on the order of G.

Lemma 9. *Let G be a finite abelian group.*

(i) *If $G = G_{p_1} \oplus \cdots \oplus G_{p_r}$ with distinct primes $p_1, \ldots, p_r$, and G_{p_i}, a p_i-group $(i = 1, \ldots, r)$, then*

$$\mu(G) = \mu(G_{p_1}) \cdots \mu(G_{p_r}). \tag{7.1}$$

(ii) *Let G be a p-group, $G = C_{p^{e_1}} \oplus \cdots \oplus C_{p^{e_r}}$ where C_{p^e} denotes a cyclic group of order p^e. Then $\mu(G) = 0$ when some $e_i > 1$.*

(iii) *Suppose $G = C_p \oplus \cdots \oplus C_p = C_p^l$. Then*

$$\mu(G) = (-1)^l p^{l(l-1)/2}. \tag{7.2}$$

Proof. (i) The assertion is an immediate consequence of the fact that the subgroups of G are of the type $H = H_{p_1} \oplus \cdots \oplus H_{p_r}$ where H_{p_i} is a subgroup of G_{p_i} $(i = 1, \ldots, r)$. A function μ with (7.1) will be called *multiplicative*.

(ii) For any abelian p-group H, let H' be the subgroup of $h \in H$ with $ph = 0$. We have to show that $\mu(G) = 0$ when $G \neq G'$. We will do this by induction on the order of G. Now

$$\sum_{H \subset G} \mu(H) = \sum_{H \subset G'} \mu(H) + \sum_{\substack{H \subset G \\ H \not\subset G'}} \mu(H).$$

The first and second sums vanish, for $G \neq G'$ yields $G \neq \{0\}$, $G' \neq \{0\}$. Therefore also the third sum vanishes. The groups H occurring in the last sum have $H' \neq H$. By induction, the summands (if any) with $H \neq G$ have $\mu(H) = 0$, so that also the summand $\mu(G) = 0$.

(iii) This assertion will not be needed in the sequel and is left as an exercise.

For a finite abelian group G, let $\nu(G) = \nu_n(G)$ be the number of lattices $\Lambda \supset \mathbb{Z}^n$ with $\Lambda/\mathbb{Z}^n \sim G$.

Lemma 10. *ν is multiplicative.*

Proof. Given an overlattice Λ of $\mathbb{Z}^n$ and a prime p, let Λ_p consist of $\mathbf{g} \in \Lambda$ having $p^t \mathbf{g} \in \mathbb{Z}^n$ for some $t \in \mathbb{N}$. It is easily seen that $\Lambda = \Sigma_p \Lambda_p$ and

$$\Lambda/\mathbb{Z}^n = \sum_p \oplus (\Lambda_p/\mathbb{Z}^n),$$

where all but finitely many summands are $\{\mathbf{0}\}$. Since $\Lambda_p/\mathbb{Z}^n$ is a p-group, the assertion follows.

Let $\Gamma(l)$ be the set of finite abelian groups which are products of p-groups with $p \leqslant l$. Set

$$\zeta_{nl}(s) = \sum_{\substack{\Lambda \supset \mathbb{Z}^n \\ \Lambda/\mathbb{Z}^n \in \Gamma(l)}} (\det \Lambda)^s,$$

$$\zeta_{nl}^*(s) = \sum_{\substack{\Lambda \supset \mathbb{Z}^n \\ \Lambda/\mathbb{Z}^n \in \Gamma(l)}} \mu(\Lambda/\mathbb{Z}^n)\,(\det \Lambda)^s.$$

Also let $\zeta_l(s)$ be a partial product for the Riemann zeta function:

$$\zeta_l(s) = \prod_{p \leqslant l} \left(\sum_{t=0}^{\infty} p^{-ts} \right).$$

Lemma 11. *(i) The sum for ζ_{nl} is absolutely convergent for $s > n - 1$, and*

$$\zeta_{nl}(s) = \zeta_l(s)\zeta_l(s-1) \cdots \zeta_l(s-n+1). \tag{7.3}$$

(ii) *The sum for ζ_{nl}^* has finitely many nonzero terms, and $\zeta_{nl}^*(s) = 1/\zeta_{nl}(s)$ when $s > n-1$.*

Proof. For completeness we will reproduce arguments of [7]. Define $\sigma_n(k)$ for positive integers k, n by $\sigma_1(k) = 1$,

$$\sigma_n(k) = \sum_{d|k} d^{n-1}\sigma_{n-1}(k/d).$$

We claim that there are exactly $\sigma_n(k)$ sublattices $\Gamma \subset \mathbb{Z}^n$ of rank n and determinant k. This is clear when $n = 1$. When Γ has determinant k, its intersection with the x_1-axis will consist of multiples of a point $(d, 0, \dots, 0)$ with $d|k$. Its orthogonal projection on the $x_2, \dots, x_n$-coordinate plane will be a lattice Γ' of determinant k/d. Using induction on n, we note that there are $\sigma_{n-1}(k/d)$ choices for Γ'. Let $\mathbf{g}_2' = (0, g_{22}, \dots, g_{2n}), \dots, \mathbf{g}_n' = (0, g_{n2}, \dots, g_{nn})$ be a basis for Γ'. Then Γ will have a basis

$$\mathbf{g}_1 = (d, 0, \dots, 0), \quad \mathbf{g}_2 = (g_{21}, g_{22}, \dots, g_{2n}), \dots, \mathbf{g}_n = (g_{n1}, g_{n2}, \dots, g_{nn}).$$

For given d and Γ', a choice of integers $g_{21}, \dots, g_{n1}$ or $\hat{g}_{21}, \dots, \hat{g}_{n1}$ will give the same lattice Γ precisely when $\hat{g}_{i1} \equiv g_{i1} \pmod{d}$ for $i = 2, \dots, n$. There are then d^{n-1} distinct lattices Γ with given d, Γ', so that the total number of lattices $\Gamma \subset \mathbb{Z}^n$ of determinant k is indeed

$$\sum_{d|k} d^{n-1}\sigma_{n-1}(k/d) = \sigma_n(k).$$

There is a 1–1-correspondence between sublattices $\Gamma \subset \mathbb{Z}^n$ of determinant k, i.e., with $|\mathbb{Z}^n/\Gamma| = k$, and overlattices $\Lambda \supset \mathbb{Z}^n$ with $|\Lambda/\mathbb{Z}^n| = k$, given by the map $\Gamma \mapsto \Gamma^*$, where Γ^* is the polar lattice to Γ. (The polar lattice consists of $\mathbf{x} \in \mathbb{R}^n$ having $\mathbf{gx} \in \mathbb{Z}$ for every $\mathbf{g} \in \Gamma$ (see [1, §1.5]).) Thus there are $\sigma_n(k)$ overlattices Λ of $\mathbb{Z}^n$ with $|\Lambda/\mathbb{Z}^n| = k$.

We now turn to the proof of (i).

$$\zeta_{nl}(s) = \sum_{\substack{\Lambda \supset \mathbb{Z}^n \\ \Lambda/\mathbb{Z}^n \in \Gamma(l)}} |\Lambda/\mathbb{Z}^n|^{-s} = \sum_{k \in \Gamma(l)} \frac{\sigma_n(k)}{k^s},$$

where $\Gamma(l)$ is the semigroup generated by the primes $\leqslant l$. Since σ_n is multiplicative,

$$\zeta_{nl}(s) = \prod_{p \leqslant l} \left(\sum_{t=0}^{\infty} \sigma_n(p^t) p^{-ts} \right). \tag{7.4}$$

Now

$$\sigma_n(k) \ll (k \log\log k)^{n-1}, \tag{7.5}$$

which is trivial for $n = 1$, well known for $n = 2$ (see [3, Theorem 323]), and follows in general by induction on n. Therefore $\sigma_n(p^t) \ll p^{(n-1+\varepsilon)t}$ for $\varepsilon > 0$, so that each sum in (7.4) is absolutely convergent for $s > n-1$. Clearly, $\zeta_{1l}(s) = \zeta_l(s)$.

Since σ_n is the Dirichlet product of σ_{n-1} and $f(k) = k^{n-1}$, we obtain $\zeta_{nl}(s) = \zeta_{n-1,l}(s)\zeta_l(s-n+1)$, and (7.3) follows by induction on n.

As for (ii), $\Lambda/\mathbb{Z}^n \in \Gamma(l)$ is a sum of p-groups with $p \leqslant l$, and by Lemma 9, $\mu(\Lambda/\mathbb{Z}^n) = 0$ unless each p-group is a sum of cyclic groups of order p. Since Λ is of rank n, they will be sums of at most n such cyclic groups, so that

$$|\Lambda/\mathbb{Z}^n| \leqslant \prod_{p \in \Gamma(l)} p^n. \tag{7.6}$$

The first assertion (ii) follows.

$$\begin{aligned} 1 &= \sum_{\substack{\Delta \supset \mathbb{Z}^n \\ \Delta/\mathbb{Z}^n \in \Gamma(l)}} |\Delta/\mathbb{Z}^n|^{-s} \sum_{H \subset \Delta/\mathbb{Z}^n} \mu(H) \\ &= \sum_{\substack{\Delta \supset \mathbb{Z}^n \\ \Delta/\mathbb{Z}^n \in \Gamma(l)}} |\Delta/\mathbb{Z}^n|^{-s} \sum_{\Delta \supset \Lambda \supset \mathbb{Z}^n} \mu(\Lambda/\mathbb{Z}^n) \\ &= \sum_{\substack{\Lambda \supset \mathbb{Z}^n \\ \Lambda/\mathbb{Z}^n \in \Gamma(l)}} \mu(\Lambda/\mathbb{Z}^n)|\Lambda/\mathbb{Z}^n|^{-s} \sum_{\substack{\Delta \supset \Lambda \\ \Delta/\Lambda \in \Gamma(l)}} |\Delta/\Lambda|^{-s}. \end{aligned} \tag{7.7}$$

Since

$$\zeta_{nl}(s) = \sum_{\substack{\Delta \supset \mathbb{Z}^n \\ \Delta/\mathbb{Z}^n \in \Gamma(l)}} |\Delta/\mathbb{Z}^n|^{-s},$$

and since $\Lambda \sim \mathbb{Z}^n$, the inner sum in the last expression of (7.7) is $\zeta_{nl}(s)$. We obtain $1 = \zeta^*_{nl}(s)\zeta_{nl}(s)$.

8. Theorem 1 Implies Theorem 2

Recall that given a lattice $\Lambda \subset \mathbb{Z}^m$, we write $S(\Lambda)$ for the subspace spanned by it in $\mathbb{R}^m$, and $\tilde{\Lambda} = S(\Lambda) \cap \mathbb{Z}^m$. Then $\Lambda \subset \tilde{\Lambda} \subset \mathbb{Z}^m$, and Λ is primitive when $\Lambda = \tilde{\Lambda}$. We introduce a new parameter $l \in \mathbb{N}$ and the semigroup $\Gamma(l)$ generated by the primes $\leqslant l$. We let $\tilde{\Lambda}_l$ consist of $\mathbf{g} \in \tilde{\Lambda}$ having $k\mathbf{g} \in \Lambda$ for some $k \in \Gamma(l)$. Then $\Lambda \subset \tilde{\Lambda}_l \subset \tilde{\Lambda}$. Write $P_l(\mathscr{D}, T)$ for the number of lattices $\Lambda \in [\mathscr{D}, T]$ with $\Lambda = \tilde{\Lambda}_l$: Recall that $[\mathscr{D}, T]$ was the set of lattices $\Lambda \subset \mathbb{Z}^m$ of similarity class belonging to $\mathscr{D}$, and of determinant $\leqslant T$. Clearly

$$P(\mathscr{D}, T) \leqslant P_l(\mathscr{D}, T). \tag{8.1}$$

On the other hand, when Λ is counted by $P_l(\mathscr{D}, T)$ but not by $P(\mathscr{D}, T)$, then $|\tilde{\Lambda} : \Lambda|$ is divisible by a prime $p > l$, and hence there is a lattice Λ' with $\Lambda \subset \Lambda' \subset \tilde{\Lambda}$ and $|\Lambda' : \Lambda| = p$. Given such Λ', there are $\sigma_n(p)$ lattices $\Lambda \subset \Lambda'$ of index p. Since $\det \Lambda' = p^{-1} \det \Lambda \leqslant T/p$, the number of choices for Λ when p given is

$$\leqslant \sigma_n(p) N(T/p) \ll p^{n-1/2}(T/p)^m = p^{n-m-1/2}T^m$$

by (7.5) and by Theorem 1. We obtain

$$P_l(\mathscr{D},T) \leqslant P(\mathscr{D},T) + \sum_{p>l} \sigma_n(p) N(T/p) = P(\mathscr{D},T) + O(l^{-1/2}T^m) \tag{8.2}$$

since $m > n$.

Write $\psi_l(\Lambda) = 1$ when $\Lambda = \tilde{\Lambda}_l$, $\psi_l(\Lambda) = 0$ otherwise. By the basic property of the generalized Moebius function,

$$\sum_{\substack{\Lambda' \\ \tilde{\Lambda}_l \supset \Lambda' \supset \Lambda}} \mu(\Lambda'/\Lambda) = \psi_l(\Lambda).$$

We have

$$N(\mathscr{D},T) = \sum_{\Lambda \in [\mathscr{D},T]} 1,$$

$$P_l(\mathscr{D},T) = \sum_{\Lambda \in [\mathscr{D},T]} \psi_l(\Lambda).$$

Therefore

$$P_l(\mathscr{D},T) = \sum_{\Lambda \in [\mathscr{D},T]} \sum_{\substack{\Lambda' \\ \tilde{\Lambda}_l \supset \Lambda' \supset \Lambda}} \mu(\Lambda'/\Lambda). \tag{8.3}$$

Every overlattice $\Delta \supset \mathbb{Z}^n$ has an upper triangular basis $A = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$. Choose some such basis $A(\Delta)$. When $\Lambda \subset \mathbb{Z}^m$ is a lattice of rank n with basis $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, then every overlattice $\Lambda' \supset \Lambda$ has a unique basis $XA(\Delta)$ where Δ is an overlattice of $\mathbb{Z}^n$. When Λ has the basis KZ with $K \in \mathscr{K}$, $Z \in \mathscr{H}$, then Λ' has the basis $KZA(\Delta) = K'Z_{A(\Delta)}$ with $K' \in \mathscr{K}$. Thus when $\Lambda \in [\mathscr{D},T]$, then

$$\Lambda' \in [\mathscr{D}_{A(\Delta)}, T \det \Delta],$$

where $\mathscr{D}_{A(\Delta)}$ is the image of $\mathscr{D}$ under $Z \mapsto Z_{A(\Delta)}$. Note that $\mu(\Lambda'/\Lambda) = \mu(\Delta/\mathbb{Z}^n)$. The inner sum in (8.3) is over integer lattices $\Lambda' \supset \Lambda$ with $\Lambda'/\Lambda \in \Gamma(l)$. Therefore

$$P_l(\mathscr{D},T) = \sum_{\substack{\Delta \supset \mathbb{Z}^n \\ \Delta/\mathbb{Z}^n \in \Gamma(l)}} \mu(\Delta/\mathbb{Z}^n) N(\mathscr{D}_{A(\Delta)}, T \det \Delta). \tag{8.4}$$

Note that $A(\Delta)$ will in general not be in $GL_n(\mathbb{Z})$, hence $\mathscr{D}_{A(\Delta)}$ not lean, and $N(\mathscr{D}_{A(\Delta)}, \ldots)$ is understood in the generalized sense explained below Theorem 1: (8.4) is correct since when Λ' has bases $K_1'Z_1', \ldots, K_t'Z_t'$ with distinct $Z_1', \ldots, Z_t'$ in $\mathscr{D}_{A(\Delta)}$, these come from t distinct points Z in $\mathscr{D}$, hence when $\mathscr{D}$ is lean from t distinct lattices Λ. It is easily seen that when A is upper triangular then $\mathscr{H}(a,b)_A \subset \mathscr{H}(a',b')$ for some a', b', and therefore $\mathscr{D} \subset \mathscr{H}(a,b)$ implies $\mathscr{D}_{A(\Delta)} \subset \mathscr{H}(a',b')$.

To avoid confusion with the Moebius function, we will write the invariant measure as bold face $\boldsymbol{\mu}$ in this section. By the invariance $\boldsymbol{\mu}(\mathscr{D}_{A(\Delta)}) = \boldsymbol{\mu}(\mathscr{D})$. An appeal to Theorem 1 gives

$$N(\mathscr{D}_{A(\Delta)}, T \det \Delta) \sim c_1(m,n) \boldsymbol{\mu}(\mathscr{D}) (T \det \Delta)^m$$

as $T \to \infty$. In (8.4)) we may restrict Δ to have $|\Delta/\mathbb{Z}^n|$ bounded by the right hand side of (7.6), so that we have only finitely many summands. Then

$$P_l(\mathscr{D}, T) \sim c_1(m,n)\boldsymbol{\mu}(\mathscr{D}) \left(\sum_{\substack{\Delta \supset \mathbb{Z}^n \\ \Delta/\mathbb{Z}^n \in \Gamma(l)}} \mu(\Delta/\mathbb{Z}^n)(\det \Delta)^m \right) T^m \tag{8.5}$$

$$= c_1(m,n)\zeta_{nl}^*(m)\boldsymbol{\mu}(\mathscr{D})T^m.$$

Here $\zeta_{nl}^*(m) = (\zeta_l(m)\zeta_l(m-1)\cdots\zeta_l(m-n+1))^{-1}$ tends to $(\zeta(m)\zeta(m-1)\cdots \cdots\zeta(m-n+1))^{-1}$ as $l \to \infty$, so that $c_1(m,n)\zeta_{nl}^*(m)$ tends to $c_2(m,n)$. Theorem 2 now follows from (8.1), (8.2) (8.5).

Actually Theorem 1, using the provisional form of Theorem 4, will first be proved in the form that $N(\mathscr{D}, T) \sim (2mn)^{-1}c_6(m,n)\mu_0(\mathscr{D})T^m$, from which the present arguments gives (4.9). As explained in Section 4, this yields (4.6), and hence Theorems 1, 2 as enunciated.

9. Theorem 4 Implies Theorem 5

Again positive a, b will be fixed, and constants in $\ll$ may depend on m,n,a,b. When $Z \in \mathscr{H}(a,b)$ we have

$$y_j \gg y_i \qquad (1 \leqslant i < j \leqslant n), \tag{9.1}$$

$$|x_{ij}| \ll y_i \quad (1 \leqslant i < j \leqslant n). \tag{9.2}$$

Lemma 12. *Let Λ be a lattice with minima $\lambda_1, \ldots, \lambda_n$. Suppose KZ with $K \in \mathscr{K}$, $Z \in \mathscr{H}(a,b)$ is a basis of Λ. Then (with $y_1 = 1$)*

$$y_i \ll \lambda_i/\lambda_1 \ll y_i \qquad (1 \leqslant i \leqslant n). \tag{9.3}$$

Proof. We may replace Λ by a similar lattice with basis Z. The columns of Z constitute a basis. The i-th column of Z has length $1 = y_1$ when $i = 1$, and length

$$(x_{1i}^2 + \cdots + x_{i-1,i}^2 + y_i^2)^{1/2} \ll (y_1^2 + \cdots + y_i^2)^{1/2} \ll y_i \tag{9.4}$$

when $i > 1$, so that $\lambda_i \ll \max(y_1, \ldots, y_i) \ll y_i$. Since by Minkowski's Theorem (see, e.g., [9, Lecture III]),

$$\lambda_1 \cdots \lambda_n \gg\ll \det \Lambda = y_1 y_2 \cdots y_n,$$

we have $\lambda_i \gg\ll y_i$ $(1 \leqslant i \leqslant n)$. Since $y_1 = 1$, (9.3) follows.

Lemma 13. *Suppose a lattice Λ has a basis $Z = (\mathbf{z}_1, \ldots, \mathbf{z}_n) \in \mathscr{H}(a,b)$. Let $\lambda_1, \ldots, \lambda_n$ be its successive minima, and $\mathbf{g}_1, \ldots, \mathbf{g}_n$ independent lattice points with $|\mathbf{g}_i| = \lambda_i$ $(i = 1, \ldots, n)$. Then each $\mathbf{g}_i$ is a linear combination of $\mathbf{z}_1, \ldots \mathbf{z}_n$ with coefficients of modulus $\ll 1$.*

Proof. Our hypotheses give

$$\lambda_i \gg\ll y_i \qquad (1 \leqslant i \leqslant n) \tag{9.5}$$

as in the proof of Lemma 12. Let j in $1 \leqslant j \leqslant n$ be fixed and write

$$\mathbf{g}_j = u_1\mathbf{z}_1 + \cdots + u_n\mathbf{z}_n. \tag{9.6}$$

By (9.4), (9.5) we have $|\mathbf{z}_k| \ll y_k \ll \lambda_k$, say $|\mathbf{z}_k| \leqslant c\lambda_k$ $(k = 1, \dots, n)$. Let i be the smallest number in $1 \leqslant i \leqslant j$ with $\lambda_i \geqslant c^{i-j}\lambda_j$. Then if $i > 1$, we note that for $1 \leqslant k < i$ we have $|\mathbf{z}_k| \leqslant c\lambda_k \leqslant c\lambda_{i-1} < \lambda_i$. Therefore $\mathbf{g}_1, \dots, \mathbf{g}_{i-1}$ lie in the space S_{i-1} spanned by $\mathbf{z}_1, \dots, \mathbf{z}_{i-1}$. We may infer that $\mathbf{g}_j \notin S_{i-1}$ (which is also true when $i = 1$, with $S_0 = \{\mathbf{0}\}$), so that $|\mathbf{g}_j|$ is bounded from below by its distance from S_{i-1}. The square of this distance is the left-hand side of (5.3). Therefore this left-hand side is

$$\leqslant |\mathbf{g}_j|^2 = \lambda_j^2 \leqslant c^{2j-2i}\lambda_i^2 \ll y_i^2.$$

In (5.6) we saw that the left hand side is

$$\gg y_i^2 \max(u_i^2, \dots, u_n^2).$$

This was shown using (5.5) and (5.1), (5.2), but (9.1), (9.2) serve just as well. A comparison of our estimates yields $|u_i|, \dots, |u_n| \ll 1$.

With $\mathbf{g}_k = u_{k1}\mathbf{z}_1 + \cdots + u_{kn}\mathbf{z}_n$ $(1 \leqslant k \leqslant n)$, the tuples $(u_{ki}, \dots, u_{kn})$ for $k = i$, $i+1, \dots, n$ are independent, since $\mathbf{g}_1, \dots, \mathbf{g}_{i-1}$ lie in the space S_{i-1} spanned by $\mathbf{z}_1, \dots, \mathbf{z}_{i-1}$. Therefore when $\mathbf{g}_1, \dots, \mathbf{g}_{i-1}$ are given and these tuples are given, $\mathbf{g}_1, \dots, \mathbf{g}_{i-1}, \mathbf{g}_i, \dots, \mathbf{g}_n$ will be independent, no matter how $u_{k1}, \dots, u_{k,i-1}$ for $k = i$, $i+1, \dots, n$ are chosen. They have to be chosen to minimize $|\mathbf{g}_k|$. In particular in (9.6), given $u_i, \dots, u_n$, the $u_1, \dots, u_{i-1}$ are such that $|\mathbf{g}_j|$ is minimized.

Suppose $1 < l \leqslant i$, and we know already that $|u_l|, \dots, |u_n| \ll 1$. We will show that $|u_{l-1}| \ll 1$. Write $\mathbf{g}_j = \mathbf{g}_j' + \mathbf{g}_j''$, where $\mathbf{g}_j'$ is spanned by $\mathbf{z}_1, \dots, \mathbf{z}_{l-1}$ and $\mathbf{g}_j''$ is orthogonal to $\mathbf{z}_1, \dots, \mathbf{z}_{l-1}$. Then $|\mathbf{g}_j|^2 = |\mathbf{g}_j'|^2 + |\mathbf{g}_j''|^2$, where $|\mathbf{g}_j''|$ depends only on $u_l, \dots, u_n$. On the other hand, $|\mathbf{g}_j'|^2$ is the quadratic polynomial

$$Q(u_1, \dots, u_{l-1}) = (u_1y_1 + u_2x_{12} + \cdots + u_{l-1}x_{1,l-1} + v_1)^2 + \cdots + (u_{l-1}y_{l-1} + v_{l-1})^2$$

with $v_k = u_lx_{kl} + \cdots + u_nx_{kn}$ $(1 \leqslant k < l)$. The minimum of $Q(u_1, \dots, u_{l-1})$ is

$$\leqslant Q(0, \dots, 0) = v_1^2 + \cdots + v_{l-1}^2 \ll y_1^2 + \cdots + y_{l-1}^2 \ll y_{l-1}^2.$$

Therefore $(u_{l-1}y_{l-1} + v_{l-1})^2 \ll y_{l-1}^2$, which yields $|u_{l-1}y_{l-1}| \ll |y_{l-1}| + |v_{l-1}| \ll$ $\ll |y_{l-1}|$, so that indeed $|u_{l-1}| \ll 1$.

We now turn to the proof of Theorem 5. Let $\mathscr{F} \subset \mathscr{F}_1$ be a fixed fundamental domain. Let $\mathscr{D}(\mathbf{a}) = \mathscr{D}(a_1, \dots, a_{n-1})$ consist of $Z \in \mathscr{F}$ corresponding to lattices with (2.6). Now (2.6), in view of (9.3) (which we may apply with $a = \sqrt{3/4}$, $b = 1/2$), yields $y_{i+1} \geqslant a_i'y_i$ with $a_i' = a_i/c_{13}(n)$. Therefore

$$\mathscr{D}(a_1, \dots, a_{n-1}) \subset \mathscr{H}(a_1', \dots, a_{n-1}'; 1/2).$$

Let $\mathscr{H}^+(\mathbf{a}, b)$ be the part of $\mathscr{H}(\mathbf{a}, b)$ with $x_{1j} \geqslant 0$ $(j = 2, \dots, n)$. Then $\mathscr{H}^+(1, \dots, 1; 1/2) \subset \mathscr{F}_1$, for $1 \leqslant y_2 \leqslant \cdots \leqslant y_n$ certainly implies (5.3). Now (2.6), in view of (9.3), yields $\lambda_{i+1}/\lambda_i \geqslant c_{14}(n, a)y_{i+1}/y_i$ (where we may suppose $c_{14} = c_{14}(n, a) \leqslant 1$), so that

$$\mathscr{H}^+(a_1'', \dots, a_{n-1}''; 1/2) \subset \mathscr{D}(a_1, \dots, a_{n-1})$$

with $a_i'' = a_i/c_{14} \geqslant a_i \geqslant 1$. Note that $\mu(\mathscr{H}^+(\cdots)) = 2^{1-n}\mu(\mathscr{H}(\cdots))$. Lemma 2 (with $b = 1/2$) now gives the assertion (i) of Theorem 5.

In view of Lemma 8, Theorem 5 will follow from Theorem 4 once we show that the boundary of $\mathscr{D}(a_1, \ldots, a_{n-1})$ is contained in the union of a bounded (in terms of n) number of linear and quadratic surfaces. Part of the boundary belongs to the boundary of $\mathscr{F}_1$, which by the construction in Section 5 is clearly of this type. Other parts of the boundary are given by a condition $\lambda_{j+1} = a_j\lambda_j$ where $1 \leqslant j < n$. Thus with $\mathbf{g}_1, \ldots, \mathbf{g}_n$ as above, we have $|\mathbf{g}_{j+1}| = a_j|\mathbf{g}_j|$. By Lemma 12, there are only $c_{15}(n)$ choices for the coefficients $u_1, \ldots, u_n$ in (9.6). Also only finitely many choices for the coefficients of $\mathbf{g}_{j+1}$. When these coefficients are given, $|\mathbf{g}_j|^2$, $|\mathbf{g}_{j+1}|^2$ are quadratic polynomials in the entries y_i, x_{il} of Z. Therefore $|\mathbf{g}_{j+1}|^2 = a_j^2|\mathbf{g}_j|^2$ determines a quadratic surface for $Z \in \mathbb{R}^h$. Theorem 5 follows.

It is clear that when $n = 2$, then $\mathscr{D}(a_1) \subset \mathscr{F}_1$ consists of points Z with $0 \leqslant x \leqslant 1/2$, $x^2 + y^2 \geqslant a_1^2$, so that

$$\mu(\mathscr{D}(a_1)) = \mu_0(\mathscr{F}_1)^{-1}\mu_0(\mathscr{D}(a_1)) = \frac{6}{\pi}\arcsin\frac{1}{2a_1}. \tag{9.7}$$

We are ready to give a

Proof of Lemma 7. Suppose $Z^0 \in \mathscr{H}(a, b) \cap \mathscr{F}_{1M}$. Then $Z^0 = Z_M$ with $Z \in \mathscr{F}_1$, $M \in GL_n(\mathbb{Z})$. Thus $Z^0 = KZM$ with $K \in \tilde{O}_n$. Let Λ be the lattice with basis Z; then $\Lambda^0 = K\Lambda$ has Z_0 as a basis. Let $G = (\mathbf{g}_1, \ldots, \mathbf{g}_n)$ where $\mathbf{g}_1, \ldots, \mathbf{g}_n$ in Λ are independent with $|\mathbf{g}_i| = \lambda_i$ $(i = 1, \ldots, n)$. Such a matrix will be called Λ-*minimal*. Then $G^0 = KG = (K\mathbf{g}_1, \ldots, K\mathbf{g}_n)$ is Λ^0-minimal. By Lemma 12, $G = ZV$ where V has integer entries of modulus $\ll 1$. Similarly $G^0 = Z^0V^0$ where V^0 has entries in $\mathbb{Z}$ of modulus $\ll 1$. We have $KG = KZV$, also $KG = G^0 = Z^0V^0$, so that $Z^0 = KZV(V^0)^{-1}$. Comparison with $Z^0 = KZM$ gives $M = V(V^0)^{-1}$. There are only finitely many possibilities for V and V^0, hence finitely many possibilities for M.

10. A Better Error Term

We will formulate Theorem 7 and will show that it implies Theorem 6. As will be demonstrated in the next section, Theorem 7 also implies Theorem 4. To motivate what follows, we begin with some general observations on the number of integer points in a set.

Let $\mathscr{B} \subset \mathbb{R}^m$, $\delta > 0$. By a δ-*net* for $\mathscr{B}$ we will understand a finite set $\mathscr{N}(\mathscr{B}, \delta) \subset \mathbb{R}^m$ such that every $\mathbf{x} \in \mathscr{B}$ has distance $< \delta$ from $\mathscr{N}(\mathscr{B}, \delta)$. The most basic observation is as follows.

Lemma 14. *Let $\mathscr{S} \subset \mathbb{R}^m$ be bounded and measurable, with volume $V(\mathscr{S})$, and let $\mathscr{N}(\partial\mathscr{S}, \delta)$ be a δ-net for its boundary $\partial\mathscr{S}$. Then*

$$|N(\mathscr{S}) - V(\mathscr{S})| \underset{m,\delta}{\ll} |\mathscr{N}(\partial\mathscr{S}, \delta)|,$$

where $N(\mathscr{S}) = |\mathscr{S} \cap \mathbb{Z}^m|$, and $|\mathscr{N}|$ denotes the cardinality of a set $\mathscr{N}$.

Proof. Recall the definition of cubes $C_\delta(\mathbf{x})$ and $\langle \mathbf{x} \rangle$ in Section 3. Let $\mathscr{S}_1$ be the union of the cubes $\langle \mathbf{x} \rangle$ with $\mathbf{x} \in \mathbb{Z}^m$ which are contained in $\mathscr{S}$, and $\mathscr{S}_2$ the union of the cubes $\langle \mathbf{x} \rangle$ with $\mathbf{x} \in \mathbb{Z}^m$ which intersect $\mathscr{S}$. Let $\mathscr{B}$ be the union of all cubes $C_2(\mathbf{y})$ with $\mathbf{y} \in \partial\mathscr{S}$. Then

$$V(\mathscr{S}) - V(\mathscr{B}) \leqslant V(\mathscr{S}_1) \leqslant N(\mathscr{S}) \leqslant V(\mathscr{S}_2) \leqslant V(\mathscr{S}) + V(\mathscr{B}).$$

Clearly $\mathscr{N}(\partial\mathscr{S}, \delta)$ is a $(\delta + \sqrt{m})$-net for $\mathscr{B}$, so that

$$V(\mathscr{B}) \underset{m}{\ll} (\delta + \sqrt{m})^m |\mathscr{N}(\partial\mathscr{S}, \delta)| \underset{m,\delta}{\ll} |\mathscr{N}(\partial\mathscr{S}, \delta)|.$$

The lemma follows.

We next recall a principle due to LANG [4, §VI.2]. A set $\mathscr{B} \subset \mathbb{R}^m$ where $m > 1$ will be said to be of *finite* L-*spread* if there is a map $\Psi : C \to \mathbb{R}^m$ with $\Psi(C) \supset \mathscr{B}$, where C is a cube in $\mathbb{R}^{m-1}$ and where Ψ is Lipschitz (so that (3.1) holds for some γ). More generally, a finite union of such sets $\mathscr{B}$ will be said to be of finite L-spread. Let $\mathscr{S} \subset \mathbb{R}^m$ be bounded, and have boundary $\partial\mathscr{S}$ of finite L-spread. Then Lang's principle says that the number $N(T\mathscr{S})$ of integer points in the expanded set $T\mathscr{S}$ has $N(T\mathscr{S}) = V(\mathscr{S})T^m + O(T^{m-1})$ as $T \to \infty$, where $V(\mathscr{S})$ exists as a Jordan measure.

To have more flexibility, it is advantageous to replace a cube C by a more general set $\mathscr{A} \subset \mathbb{R}^{m-1}$. But $\mathscr{A}$ in general is not "round" enough. Recall that in Section 3 we defined $\mathscr{A}_\varepsilon$ to consist of points $\boldsymbol{\alpha} + \boldsymbol{\beta}$ with $\boldsymbol{\alpha} \in \mathscr{A}$ and $|\boldsymbol{\beta}| < \varepsilon$, so that $\mathscr{A}_\varepsilon$ consists of points having distance $< \varepsilon$ from $\mathscr{A}$. A set $\mathscr{B} \subset \mathbb{R}^m$ will be said to be of *simple spread* $\leqslant \tau$ if there is a set $\mathscr{A} \subset \mathbb{R}^{m-1}$ with

$$V(\mathscr{A}_1) \leqslant \tau \tag{10.1}$$

and a 1-Lipschitz map $\Psi : \mathscr{A}_1 \to \mathbb{R}^m$ with $\Psi(\mathscr{A}) \supset \mathscr{B}$. More generally, a set $\mathscr{B} = \mathscr{B}_1 \cup \cdots \cup \mathscr{B}_t$ will be said to be of *spread* $\leqslant \tau$ if $\mathscr{B}_1, \ldots, \mathscr{B}_t$ are respectively of simple spread $\leqslant \tau_1, \ldots, \tau_t$, and if $\tau_1 + \cdots + \tau_t \leqslant \tau$. A set is of *finite* spread if it is of spread $\leqslant \tau$ for some τ.

Lemma 15. *Suppose $\mathscr{S} \subset \mathbb{R}^m$ where $m > 1$ is bounded, and its boundary $\partial\mathscr{S}$ is of spread $\leqslant \tau$. Then the volume $V(\mathscr{S})$ exists as a Jordan measure, and the number $N(\mathscr{S})$ of integer points in $\mathscr{S}$ has*

$$|N(\mathscr{S}) - V(\mathscr{S})| \ll \tau. \tag{10.2}$$

Proof. Suppose $\mathscr{B}$ is of simple spread $\leqslant \tau$, and let $\mathscr{A}$, Ψ be as above. Suppose $0 < \delta \leqslant 1$, and let $\mathscr{C}$ be the set of cubes $C_{\delta/m}(\boldsymbol{\alpha})$ in $\mathbb{R}^{m-1}$ which intersect $\mathscr{A}$, and where $\boldsymbol{\alpha} \in (\delta/m)\mathbb{Z}^{m-1}$. These cubes are contained in $\mathscr{A}_\delta \subset \mathscr{A}_1$, and their union covers $\mathscr{A}$. For each $C \in \mathscr{C}$, pick $\boldsymbol{\alpha}_C \in C$. Let $\mathscr{N}(\mathscr{B}, \delta)$ be the set of points $\Psi(\boldsymbol{\alpha}_C)$ with $C \in \mathscr{C}$. When $\mathbf{x} \in \mathscr{B}$, say $\mathbf{x} = \Psi(\boldsymbol{\alpha})$, and say $\boldsymbol{\alpha} \in C$ with $C \in \mathscr{C}$. Then $|\boldsymbol{\alpha} - \boldsymbol{\alpha}_C| < \delta$, therefore $|\mathbf{x} - \Psi(\boldsymbol{\alpha}_C)| = |\Psi(\boldsymbol{\alpha}) - \Psi(\boldsymbol{\alpha}_C)| < \delta$. Therefore $\mathscr{N}(\mathscr{B}, \delta)$ is a δ-net for $\mathscr{B}$. Further

$$|\mathscr{N}(\mathscr{B}, \delta)| = |\mathscr{C}| \leqslant (m/\delta)^{m-1} V(\mathscr{A}_1) \ll \delta^{1-m}\tau$$

by (10.1). More generally, when $\mathscr{B}$ is of spread $\leqslant \tau$ and $0 < \delta \leqslant 1$, there is a δ-net $\mathscr{N}(\mathscr{B}, \delta)$ of cardinality $\underset{m}{\ll} \delta^{1-m}\tau$. It easily follows that there is an upper Riemann

sum for the indicator function of $\mathscr{B}$ which is $\ll \delta^m \cdot \delta^{1-m}\tau = \delta\tau$. Therefore $\mathscr{B}$ has Jordan measure zero.

Thus, when $\partial\mathscr{S}$ is of finite spread, $\mathscr{S}$ is Jordan-measurable. The hypotheses of Lemma 15 imply the existence of a net $\mathscr{N}(\partial\mathscr{S}, 1)$ of cardinality $\ll \tau$. An application of Lemma 14 with $\delta = 1$ yields (10.2).

We will now modify the definition of spread to suit subsets $\mathscr{B}$ of $\mathscr{H} = \mathscr{H}_n$. Again we consider $\mathscr{H}$ to be contained in $\mathbb{R}^h$ where $h = h(n) = 2 + 3 + \cdots + n$. Then $\mathscr{B} \subset \mathscr{H}$ will be considered to be of simple spread $\leqslant\tau$ if there is a set $\mathscr{A} \subset \mathbb{R}^{h-1}$, and a 1-Lipschitz map $\mathscr{A}_1 \to \mathscr{H}$, say $\boldsymbol{\alpha}\mapsto Z(\boldsymbol{\alpha})$, with $Z(\mathscr{A}) \supset \mathscr{B}$ and

$$\int_{\mathscr{A}_1} y_2(\boldsymbol{\alpha})^{-2+1/n} \cdots y_n(\boldsymbol{\alpha})^{-n+1/n} d\boldsymbol{\alpha} \leqslant \tau. \tag{10.3}$$

Here the $y_i(\boldsymbol{\alpha})$ (and $x_{ij}(\boldsymbol{\alpha})$) are the entries of $Z(\boldsymbol{\alpha})$. A set $\mathscr{B}$ is of spread $\leqslant\tau$ if $\mathscr{B} = \mathscr{B}_1 \cup \cdots \cup \mathscr{B}_t$ where $\mathscr{B}_i$ is of simple spread $\leqslant\tau_i$ $(i = 1, \ldots, t)$ and where $\tau_1 + \cdots + \tau_t \leqslant \tau$.

Again $a > 0$ $b > 0$ will be fixed, and $a_1, \ldots, a_{n-1}$ satisfy (2.5), i.e., $a_i \geqslant a$ $(i = 1, \ldots, n-1)$.

Theorem 7. *Let $\mathscr{D} \subset \mathscr{H}(\mathbf{a}, b)$ have boundary $\partial\mathscr{D}$ of spread $\leqslant\tau$. Then $\mathscr{D}$ is Jordan-measurable, and*

$$|N(\mathscr{D}, T) - c_1(m,n)\mu(\mathscr{D})T^m| \ll c_{16}T^{m-1/n}$$

where

$$c_{16} = a_1^{-1}\prod(\mathbf{a}) + \tau. \tag{10.4}$$

Recall the definition of $\Pi(\mathbf{a})$ in (2.3). As always, the implied constant depends only on m, n, a, b. In view of (2.5) we have

$$c_{16} \ll \prod(\mathbf{a}) + \tau \ll 1 + \tau.$$

Deduction of Theorem 6. Set $\mathscr{A} = I$, and let $\alpha \mapsto Z(\alpha)$ be the map from $\mathscr{A}$ onto $\partial\mathscr{D}$ with $Z(\alpha) = (x(\alpha), y(\alpha))$. When $\mathscr{A} = I = [u, v]$, extend this map to $\mathscr{A}_1$ by setting $Z(\alpha) = (x(u), y(u)) = (x(v), y(v))$ when $u - 1 < \alpha < u$ or $v < \alpha < v + 1$. Our map is 1-Lipschitz. Further

$$\int_{\mathscr{A}_1} y(\alpha)^{-3/2} d\alpha \leqslant \int_I y(\alpha)^{-3/2} d\alpha + 2y(u)^{-3/2} < 2c_4(\mathscr{D}),$$

so that $\partial\mathscr{D}$ is of spread $< 2c_4(\mathscr{D})$. Also, we may set $a_1 = \min_\alpha y(\alpha)$, so that $a_1^{-1}\prod(a_1) = a_1^{-3/2} < c_4(\mathscr{D})$. We may conclude that

$$c_{16} \ll c_4(\mathscr{D}).$$

11. Theorem 7 Implies Theorem 4

We first will show how a Lipschitz function defined on a subset of a Euclidean space E can be extended to a Lipschitz function on E.

Lemma 16. *Let $\mathscr{A}$ be a nonempty subset of Euclidean space E, and ψ a real-valued γ-Lipschitz function on E. For $\boldsymbol{\beta} \in E$ set*

$$\hat{\psi}(\boldsymbol{\beta}) = \inf_{\boldsymbol{\alpha} \in \mathscr{A}} (\psi(\boldsymbol{\alpha}) + 2\gamma|\boldsymbol{\alpha} - \boldsymbol{\beta}|). \tag{11.1}$$

Let $\boldsymbol{\alpha}_0 \in \mathscr{A}$ be arbitrary. Then

(i) *we may restrict the infimum to $\boldsymbol{\alpha} \in \mathscr{A}$ with $|\boldsymbol{\alpha} - \boldsymbol{\beta}| \leqslant 5|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}|$.*

(ii) $\psi(\boldsymbol{\alpha}_0) - 2\gamma|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}| \leqslant \hat{\psi}(\boldsymbol{\beta}) \leqslant \psi(\boldsymbol{\alpha}_0) + 2\gamma|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}|$. *In particular, $\hat{\psi}(\boldsymbol{\beta}) = \psi(\boldsymbol{\beta})$ when $\boldsymbol{\beta} \in \mathscr{A}$.*

(iii) *$\hat{\psi}$ is 2γ-Lipschitz.*

Proof. (i) For $\boldsymbol{\alpha} \in \mathscr{A}$,

$$\begin{aligned} &\psi(\boldsymbol{\alpha}) + 2\gamma|\boldsymbol{\alpha} - \boldsymbol{\beta}| \\ &\quad \geqslant \psi(\boldsymbol{\alpha}_0) - \gamma|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0| + 2\gamma|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0| - 2\gamma|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}| \\ &\quad = \psi(\boldsymbol{\alpha}_0) + 2\gamma|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}| + \gamma(|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0| - 4|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}|). \end{aligned} \tag{11.2}$$

Since clearly $\hat{\psi}(\boldsymbol{\beta}) \leqslant \psi(\boldsymbol{\alpha}_0) + 2\gamma|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}|$ (which is the (trivial) second inequality in (ii)), we may restrict to $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0| \leqslant 4|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}|$, hence with $|\boldsymbol{\alpha} - \boldsymbol{\beta}| \leqslant 5|\boldsymbol{\alpha}_0 - \boldsymbol{\beta}|$. In particular, when E is finite-dimensional and $\mathscr{A}$ is closed, the infimum in (11.1) is a minimum.

(ii) The first inequality in (ii) follows from (11.2). When $\boldsymbol{\beta} \in \mathscr{A}$, we may set $\boldsymbol{\alpha}_0 = \boldsymbol{\beta}$, to obtain $\hat{\psi}(\boldsymbol{\beta}) = \psi(\boldsymbol{\beta})$.

(iii) is a consequence of

$$\psi(\boldsymbol{\alpha}) + 2\gamma|\boldsymbol{\alpha} - \boldsymbol{\beta}| \leqslant \psi(\boldsymbol{\alpha}) + 2\gamma|\boldsymbol{\alpha} - \boldsymbol{\beta}'| + 2\gamma|\boldsymbol{\beta} - \boldsymbol{\beta}'|.$$

Lemma 17. *Suppose $\mathscr{A} \subset \mathbb{R}^{h-1}$ is nonempty, and $\Psi : \mathscr{A} \to \mathbb{R}^h$ is a map with* (2.4), *i.e.,*

$$|\boldsymbol{\alpha} - \boldsymbol{\alpha}'| \leqslant |\Psi(\boldsymbol{\alpha}) - \Psi(\boldsymbol{\alpha}')| \leqslant \gamma|\boldsymbol{\alpha} - \boldsymbol{\alpha}'| \tag{11.3}$$

for $\boldsymbol{\alpha}, \boldsymbol{\alpha}'$ in $\mathscr{A}$. We will suppose that $\gamma \geqslant 1$ (which is automatically true when $\mathscr{A}$ consists of more than one point). For each component ψ_i of $\Psi = (\psi_1, \dots, \psi_h)$ define $\hat{\psi}_i$ as in Lemma 16, and set $\hat{\Psi} = (\hat{\psi}_1, \dots, \hat{\psi}_h)$. Then

(a) *$\hat{\Psi}$ is a $2\gamma h$-Lipschitz map $\mathbb{R}^{h-1} \to \mathbb{R}^h$.*

(b) *Suppose $\boldsymbol{\beta}, \boldsymbol{\beta}'$ have distance $\leqslant 1/\gamma^2$ from $\mathscr{A}$, and $|\hat{\Psi}(\boldsymbol{\beta}) - \hat{\Psi}(\boldsymbol{\beta}')| \leqslant 1/\gamma$. Then*

$$|\boldsymbol{\beta} - \boldsymbol{\beta}'| \leqslant 9h/\gamma.$$

Proof. (a) Each $\hat{\psi}_i$ is 2γ-Lipschitz, therefore $\hat{\Psi}$ is $2\gamma h$-Lipschitz.

(b) Pick $\boldsymbol{\alpha}, \boldsymbol{\alpha}'$ is $\mathscr{A}$ with $|\boldsymbol{\alpha} - \boldsymbol{\beta}|, |\boldsymbol{\alpha}' - \boldsymbol{\beta}'|$ both $< 3/2\gamma^2$. Then $|\hat{\Psi}(\boldsymbol{\beta}) - \hat{\Psi}(\boldsymbol{\alpha})| \leqslant 2\gamma h|\boldsymbol{\beta} - \boldsymbol{\alpha}| < 3h/\gamma$, and a similar estimate holds for $\boldsymbol{\beta}', \boldsymbol{\alpha}'$. Since $\hat{\Psi}(\boldsymbol{\alpha}) = \Psi(\boldsymbol{\alpha})$, $\hat{\Psi}(\boldsymbol{\alpha}') = \Psi(\boldsymbol{\alpha}')$,

$$\begin{aligned} |\Psi(\boldsymbol{\alpha}) - \Psi(\boldsymbol{\alpha}')| &\leqslant |\hat{\Psi}(\boldsymbol{\alpha}) - \hat{\Psi}(\boldsymbol{\beta})| + |\hat{\Psi}(\boldsymbol{\beta}) - \hat{\Psi}(\boldsymbol{\beta}')| + |\hat{\Psi}(\boldsymbol{\beta}') - \hat{\Psi}(\boldsymbol{\alpha}')| \\ &\leqslant (3h/\gamma) + (1/\gamma) + (3h/\gamma) < 7h/\gamma, \end{aligned}$$

so that $|\alpha - \alpha'| < 7h/\gamma$ by (11.3). But then

$$|\beta - \beta'| \leqslant |\beta - \alpha| + |\alpha - \alpha'| + |\alpha' - \beta| < (3/2\gamma^2) + (7h/\gamma) + (3/2\gamma^2) < 9h/\gamma.$$

Lemma 18. *Suppose $\mathscr{B} \subset \mathscr{H}(\mathbf{a}, b)$ is γ-special with $\gamma \geqslant 1$. Then $\mathscr{B}$ is of simple spread*

$$\ll \gamma^{3h} \prod(\mathbf{a}).$$

Clearly this lemma and Theorem 7 imply Theorem 4.

Proof. Suppose $\eta > 0$, $\mathscr{A} \subset \mathbb{R}^{h-1}$, and $\alpha \mapsto Z(\alpha)$ is an η-Lipschitz map $\mathscr{A}_{1/\eta} \to \mathscr{H}$ with $Z(\mathscr{A}) \supset \mathscr{B}$ and

$$\eta^{h-1} \int_{\mathscr{A}_{1/\eta}} g(Z(\alpha)) d\alpha \leqslant \tau, \tag{11.4}$$

where

$$g(Z) = y_2^{-2+1/n} \cdots y_n^{-n+1/n}. \tag{11.5}$$

Then if $\mathscr{A}' = \eta\mathscr{A}$, so that $\mathscr{A}'_1 = \eta\mathscr{A}_{1/\eta}$, and if $Z'(\alpha) = Z(\eta^{-1}\alpha)$, then $\alpha \mapsto Z'(\alpha)$ is a 1-Lipschitz map $\mathscr{A}'_1 \to \mathscr{H}$, and (11.4) becomes

$$\int_{\mathscr{A}'_1} g(Z'(\alpha)) d\alpha \leqslant \tau.$$

We may conclude that $\mathscr{B}$ is of simple spread $\leqslant \tau$.

To prove Lemma 18, we may suppose that

$$\gamma \geqslant c_{17}(n, a, b), \tag{11.6}$$

where c_{17} will be specified below. Set

$$\eta = 2h\gamma^2. \tag{11.7}$$

Extend the map Ψ with (2.4) of the γ-special set $\mathscr{B}$ to $\hat{\Psi}$ as in Lemma 17. Rename $\hat{\Psi}(\alpha) = Z(\alpha)$, so that the map $\alpha \mapsto Z(\alpha)$, is η-Lipschitz, and $Z(\mathscr{A}) \supset \mathscr{B}$. We need to estimate (11.4).

Let $\mathscr{C}$ be the collection of cubes $C_{1/\eta}(Z^*)$ with $Z^* \in \eta^{-1}\mathbb{Z}^h$ which intersect $\mathscr{H}(\mathbf{a}, b)$. These cubes cover $\mathscr{H}(\mathbf{a}, b)$, and if $c_{17}(n, a, b)$ in (11.6) is chosen large enough, they are contained in $\mathscr{H}(\frac{1}{2}\mathbf{a}, 2b)$ (Recall (2.5)). When $Z, Z' \in C$ where $C \in \mathscr{C}$, with respective components y_i, y'_i, then $y'_i \geqslant y_i - 1/\eta \geqslant y_i/2$ if $c_{17}(n, a, b)$ in (11.6) is sufficiently large. Then $g(Z') < 2^h g(Z)$, therefore

$$g(Z') < 2^h V(C)^{-1} \int_C g(Z) d\mathfrak{z}_2 \cdots d\mathfrak{z}_n,$$

in the notation $Z = (\mathfrak{z}_2, \ldots, \mathfrak{z}_n)$ of Section 4.

Let $I(C)$ be the part of the integral in (11.4) with $Z(\alpha) \in C$. When both $Z(\alpha), Z(\alpha') \in C$, then $|Z(\alpha) - Z(\alpha')| < h/\eta < 1/\gamma$. Since α, α' both have distance $< h/\eta < 1/\gamma^2$ from $\mathscr{A}$, Lemma 17 gives $|\alpha - \alpha'| \leqslant 9h/\gamma$, so that α lies in a ball

$B(C) \subset \mathbb{R}^{h-1}$ of radius $9h/\gamma$. We obtain

$$I(C) < V(B(C)) \cdot 2^h(C)^{-1} \int_C g(Z) d\mathfrak{z}_2 \cdots d\mathfrak{z}_n.$$

Here

$$V(B(C)) \cdot 2^h V(C)^{-1} \ll (h/\gamma)^{h-1} \cdot 2^h \eta^h \ll \gamma^{1-h+2h} \leqslant \gamma^{h+1}$$

Taking the sum over the cubes $C \in \mathscr{C}$, which we have seen to lie in $\mathscr{H}(\frac{1}{2}\mathbf{a}, 2b)$, we may bound the integral in (11.4) by

$$\ll \gamma^{h+1} \mu_1 \left(\mathscr{H}\left(\frac{1}{2}\mathbf{a}, 2b\right)\right) \ll \gamma^{h+1} \prod(\mathbf{a})$$

in view of (4.12). Since $\eta \ll \gamma^2$, the left hand side of (11.4) is $\ll \gamma^{2h-2+h+1} \prod(\mathbf{a}) \leqslant \gamma^{3h} \prod(\mathbf{a})$.

12. The Plan of the Proof of Theorem 7

With a set $\mathscr{D} \subset \mathscr{H}$ we associate the set $\mathscr{D}^*$ consisting of $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathscr{X}$ of the form (1.2) with $K \in \mathscr{K}$ and $Z \in \mathscr{D}$. Thus $\mathscr{D}^* \subset \mathbb{R}^{mn}$. Given independent points $\mathbf{x}_1, \ldots, \mathbf{x}_t$ with $1 \leqslant t < n$, we define

$$\mathscr{D}^*(\mathbf{x}_1, \ldots, \mathbf{x}_t)$$

to consist of $\mathbf{x}_{t+1}, \ldots, \mathbf{x}_n$ such that $X = (\mathbf{x}_1, \ldots, \mathbf{x}_t, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_n) \in \mathscr{D}^*$. Then $\mathscr{D}^*(\mathbf{x}_1, \ldots, \mathbf{x}_t) \subset \mathbb{R}^{m(n-t)}$. Further $N(\mathscr{D}^*(\mathbf{x}_1, \ldots, \mathbf{x}_t))$ will denote the number of integer points in $\mathscr{D}^*(\mathbf{x}_1, \ldots, \mathbf{x}_t)$, and $V(\mathscr{D}^*(\mathbf{x}_1, \ldots, \mathbf{x}_t))$ its measure, if it exists.

Proposition A. *Suppose $\mathscr{D} \subset \mathscr{H}(a, b)$ is bounded and Jordan-measurable. Then $V(\mathscr{D}^*(\mathbf{x}_1))$ exists for every $\mathbf{x}_1 \neq \mathbf{0}$ as a Jordan-measure, and*

$$V(\mathscr{D}^*(\mathbf{x}_1)) = c_6(m-1, n-1)|\mathbf{x}_1|^{m(n-1)} \int_{\mathscr{D}} y_2^{m-2} \cdots y_n^{m-n} d\mathfrak{z}_2 \cdots d\mathfrak{z}_n.$$

Proposition B. *Let $\mathscr{D}$ be as in Proposition A, and $\mathbf{x}_1$ a nonzero integer point. Let $\mathscr{N}(\partial\mathscr{D}, \delta) \subset \mathscr{H}(a, b)$ be a δ-net for $\partial\mathscr{D}$ with*

$$\delta = m/(2|\mathbf{x}_1|). \tag{12.1}$$

Then

$$|N(\mathscr{D}^*(\mathbf{x}_1)) - V(\mathscr{D}^*(\mathbf{x}_1))| \ll |\mathbf{x}_1|^{(n-1)m-h} \sum_{Z \in \mathscr{N}(\partial\mathscr{D}, \delta)} y_2^{m-2} \cdots y_n^{m-n}. \tag{12.2}$$

Proposition A is easy, and the closeness of $N(\mathscr{D}^*(\mathbf{x}_1))$ and $V(\mathscr{D}^*(\mathbf{x}_1))$ as evidenced by Proposition B is not unsuspected. The proofs of Propositions A and B will be done in Section 13–15. Of course, in view of Lemma 14, the left hand side of (12.2) may be estimated in terms of nets for $\partial(\mathscr{D}^*(\mathbf{x}_1))$. However, such an approach would give only an estimate with $y_2^{m-n} \cdots y_n^{m-2}$ in place of $y_2^{m-2} \cdots y_n^{m-n}$, which is weaker because of (2.1). Hence our argument will be more elaborate.

Let $\mathscr{D}^*(T)$ consist of $X \in \mathscr{D}^*$ with $\det X \leqslant T$: here $\det X$ is the determinant of the lattice with basis X, i.e., the square root of $\det(\mathbf{x}_i\mathbf{x}_j)_{1\leqslant i,j\leqslant n}$.

Proposition C. *Suppose $\mathscr{D} \subset \mathscr{H}(a,b)$ has boundary of finite spread. Then $\mathscr{D}^*(T)$ has Jordan volume*

$$V(\mathscr{D}^*(T)) = (mn)^{-1}c_6(m,n)\mu_0(\mathscr{D})T^m. \tag{12.3}$$

Proposition D. *Let $\mathscr{D} \subset \mathscr{H}(\mathbf{a},b)$ have boundary of finite spread $\leqslant\tau$. Then the number $N(\mathscr{D}^*(T))$ of integer points in $\mathscr{D}^*(T)$ has*

$$|N(\mathscr{D}^*(T)) - V(\mathscr{D}^*(T))| \ll \left(a_1^{-1}\prod(\mathbf{a}) + \tau\right)T^{m-1/n}. \tag{12.4}$$

Here we assume (2.5), *and the implicit constant depends on m,n,a,b.*

Again Proposition C is easy, but the error estimate in Proposition D is difficult. The proof of these propositions will be completed in Section 17.

We will deduce Theorem 7. For $\mathscr{D}$ as in that theorem, Propositions C, D give

$$|N(\mathscr{D}^*(T)) - (mn)^{-1}c_6(m,n)\mu_0(\mathscr{D})T^m| \ll c_{16}T^{m-1/n} \tag{12.5}$$

with

$$c_{16} = a_1^{-1}\prod(\mathbf{a}) + \tau.$$

Since the boundary $\partial\mathscr{D}$ has $\mu(\mathscr{D}) = 0$,

$$N((\partial\mathscr{D})^*(T)) \ll c_{16}T^{m-1/n}. \tag{12.6}$$

$N(\mathscr{D},T)$ is not the same as $N(\mathscr{D}^*(T))$, since a lattice with basis KZ ($K \in \mathscr{K}$) has $g(Z)$ such bases. But when $\mathscr{D}$ is lean and Z in the interior of $\mathscr{D}$ (hence in the interior of some fundamental domain), $g(Z) = 2$ by Lemma 6. Therefore

$$\left|N(\mathscr{D},T) - \tfrac{1}{2}N(\mathscr{D}^*(T))\right| \leqslant N((\partial\mathscr{D}^*)(T)).$$

By (12.5), (12.6),

$$|N(\mathscr{D},T) - (2mn)^{-1}c_6(m,n)\mu_0(\mathscr{D})T^m| \ll c_{16}T^{m-1/n}.$$

As was pointed out in Section 4, we may deduce that $(2mm)^{-1}c_6(m,n)\mu_0(\mathscr{D}) = c_1(m,n)\mu(\mathscr{D})$, and Theorem 7 follows.

13. Proof of Proposition A

Let $\mathscr{X}_t$ for $1\leqslant t\leqslant n$ be the set of linearly independent t-tuples $(\mathbf{x}_1,\dots,\mathbf{x}_t)$ of columns with m components. Let $\mathscr{K}_t$ consist of $(\mathbf{k}_1,\dots,\mathbf{k}_t)$ with $|\mathbf{k}_1| = \cdots = |\mathbf{k}_t| \neq 0$ and $\mathbf{k}_i\mathbf{k}_j = 0$ for $i \neq j$. Thus $\mathscr{X}_n = \mathscr{X}$, $\mathscr{K}_n = \mathscr{K}$ in the terminology of the Introduction. Recall the notation $Z = (\mathfrak{z}_2,\dots,\mathfrak{z}_n)$ where $\mathfrak{z}_j \in \mathbb{R}^j$ has components $x_{1j},\dots,x_{j-1,j},y_j$. Note that $(\mathbf{x}_1,\dots,\mathbf{x}_t) \in \mathscr{X}_t$ *determines* $(\mathbf{k}_1,\dots,\mathbf{k}_t) \in \mathscr{K}_t$ *and (when* $t > 1$*)* $\mathfrak{z}_2,\dots,\mathfrak{z}_t$ *by* (1.1).

Given $\mathscr{D} \subset \mathscr{H}$ and $(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$ with $t < n$, we define $\mathscr{D}(\mathfrak{z}_2,\dots\mathfrak{z}_t)$ to consist of $(n-t)$-tuples $\mathfrak{z}_{t+1},\dots,\mathfrak{z}_n$ with $Z = (\mathfrak{z}_2,\dots,\mathfrak{z}_t,\mathfrak{z}_{t+1},\dots,\mathfrak{z}_n) \in \mathscr{D}$. Thus

$\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)\in\mathbb{R}^{h(n)-h(t)}$ where $h(t)$ is given by (3.2). When $t=1$ we interpret $\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$ to be $\mathscr{D}$. Given $(\mathbf{k}_1,\dots,\mathbf{k}_t)\in\mathscr{K}_t$ where $t<n$, let $S(\mathbf{k}_1,\dots,\mathbf{k}_t)$ consist of vectors $\mathbf{k}_{t+1}$ such that $(\mathbf{k}_1,\dots,\mathbf{k}_t,\mathbf{k}_{t+1})\in\mathscr{K}_{t+1}$. Thus it consists of vectors of norm $|\mathbf{k}_1|$ in the space orthogonal to $\mathbf{k}_1,\dots\mathbf{k}_t$; it is a sphere of radius $|\mathbf{k}_1|$ in $(m-t)$-dimensional space.

We will show that for $1\leqslant t\leqslant n-1$ and $(\mathbf{x}_1,\dots,\mathbf{x}_t)\in\mathscr{X}_t$,

$$V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t))=c_6(m-t,n-t)|\mathbf{x}_1|^{m(n-t)}\int_{\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)}(y'_{t+1})^{m-t-1}\cdots(y'_n)^{m-n}d\mathfrak{z}'_{t+1}\cdots d\mathfrak{z}'_n \tag{13.1}$$

where $\mathfrak{z}_2,\dots,\mathfrak{z}_t$ are the tuples determined by $\mathbf{x}_1,\dots,\mathbf{x}_t$. If $t=1$, this is Proposition A.

So let $\mathbf{x}_1,\dots,\mathbf{x}_t$, therefore $\mathbf{k}_1,\dots,\mathbf{k}_t$ and (when $t>1$) $\mathfrak{z}_2,\dots,\mathfrak{z}_t$ be given. Here $|\mathbf{k}_1|=\cdots=|\mathbf{k}_t|=|\mathbf{x}_1|$. A point $(\mathbf{x}'_{t+1},\dots,\mathbf{x}'_n)\in\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t)$ has

$$\mathbf{x}'_{t+1}=x'_{1,t+1}\mathbf{k}_1+\cdots+x'_{t,t+1}\mathbf{k}_t+y'_{t+1}\mathbf{k}$$

with $\mathbf{k}\in S(\mathbf{k}_1,\dots,\mathbf{k}_t)$. As $\mathfrak{z}'_{t+1}$ with coordinates $x'_{1,t+1},\dots,x'_{t,t+1},y'_{t+1}$ ranges through a volume element $d\mathfrak{z}'_{t+1}$ in $\mathbb{R}^{t+1}$, then $x'_{1,t+1}\mathbf{k}_1+\cdots+x'_{t,t+1}\mathbf{k}_t$ will range through a box in $\mathbb{R}^t$ of volume $dx'_{1,t+1}\cdots dx'_{t,t+1}|\mathbf{x}_1|^t$, and $\mathbf{y}'_{t+1}\mathbf{k}$ will range through a spherical shell in $\mathbb{R}^{m-t}$ of radius $\sim y'_{t+1}|\mathbf{x}_1|$ and thickness $dy'_{t+1}|\mathbf{x}_1|$. Therefore

$$d\mathbf{x}'_{t+1}=(m-t)V_{m-t}|\mathbf{x}_1|^m(y'_{t+1})^{m-t-1}d\mathfrak{z}'_{t+1}.$$

Since $(m-n+1)V_{m-n+1}=c_6(m-n+1,1)$, the case $t=n-1$ of (13.1) follows. When $1\leqslant t<n-1$, and (13.1) is true for $t+1$, we observe that

$$V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t))=\int V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\mathbf{x}'_{t+1}))d\mathbf{x}'_{t+1},$$

and that $c_6(m-t,n-t)=(m-t)V_{m-t}c_6(m-t-1,n-t-1)$, so that the assertion (13.1) is true for t.

Our arguments are certainly valid when $\mathscr{D}$ is a cube, or more generally when it is a finite union of cubes. Therefore Proposition A is true when $\mathscr{D}$ is bounded and Jordan measurable.

14. Auxiliary Lemmas

As always $\partial\mathscr{D}$ denotes the boundary of a set $\mathscr{D}$.

Lemma 19. *Let $\mathscr{D}\subset\mathscr{H}(a,b)$ and $(\mathbf{x}_1,\dots,\mathbf{x}_{n-1})\in\mathscr{X}_{n-1}$ be given. Then*

$$\partial(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_{n-1}))\subset(\partial\mathscr{D})^*(\mathbf{x}_1,\dots,\mathbf{x}_{n-1}).$$

Proof. Let $(\mathbf{k}_1,\dots,\mathbf{k}_{n-1})\in\mathscr{K}_{n-1}$ and (when $n>2$) $\mathfrak{z}_2,\dots,\mathfrak{z}_{n-1}$ be determined by $\mathbf{x}_1,\dots,\mathbf{x}_{n-1}$. Suppose $\mathbf{x}\in\partial(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_{n-1}))$ is written as

$$\mathbf{x}=x_{1n}\mathbf{k}_1+\cdots+x_{n-1,n}\mathbf{k}_{n-1}+y_n\mathbf{k} \tag{14.1}$$

where $y_n>0$ and $\mathbf{k}\in S(\mathbf{k}_1,\dots,\mathbf{k}_{n-1})$. For any $\varepsilon>0$ there are $\mathbf{x}^{(1)},\mathbf{x}^{(2)}\in C_\varepsilon(\mathbf{x})$

with $\mathbf{x}^{(1)} \in \mathscr{D}^*(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}), \mathbf{x}^{(2)} \notin \mathscr{D}^*(\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$. Say

$$\mathbf{x}^{(i)} = x_{1n}^{(i)}\mathbf{k}_1 + \cdots + x_{n-1,n}^{(i)}\mathbf{k}_{n-1} + y_n^{(i)}\mathbf{k}^{(i)} \quad (i = 1, 2)$$

with $y_n^{(i)} > 0$ and $\mathbf{k}^{(i)} \in S(\mathbf{k}_1, \dots, \mathbf{k}_{n-1})$. When $\mathfrak{z}_n$ respectively $\mathfrak{z}_n^{(i)}$ have coordinates $x_{1n}, \dots, x_{n-1,n}, y_n$ respectively $x_{1n}^{(i)}, \dots, x_{n-1,n}^{(i)}, y_n^{(i)}$, then $Z^{(1)} = (\mathfrak{z}_2, \dots, \mathfrak{z}_{n-1}, \mathfrak{z}_n^{(1)}) \in \in \mathscr{D}$, $Z^{(2)} = (\mathfrak{z}_2, \dots, \mathfrak{z}_{n-1}, \mathfrak{z}_n^{(2)}) \notin \mathscr{D}$. Moreover, $\mathbf{x}^{(i)} \in C_\varepsilon$ implies $|\mathfrak{z}_n^{(i)} - \mathfrak{z}_n| \leqslant \leqslant \sqrt{m}\varepsilon|\mathbf{x}_1|^{-1} (i = 1, 2)$. We may conclude that $(\mathfrak{z}_2, \dots, \mathfrak{z}_{n-1}, \mathfrak{z}_n) \in \partial\mathscr{D}$, so that $(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}, \mathbf{x}) \in (\partial\mathscr{D})^*$ and $\mathbf{x} \in (\partial\mathscr{D})^*(\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$.

Suppose $\mathscr{B} \subset \mathscr{H}$, and let $\mathscr{N}(\mathscr{B}, \delta)$ be a δ-net for $\mathscr{B}$. Given $\mathfrak{z}_2, \dots, \mathfrak{z}_t$ where $2 \leqslant t \leqslant n$, let

$$\mathscr{N}(\mathscr{B}, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_t) \tag{14.2}$$

consist of $Z' \in \mathscr{N}(\mathscr{B}, \delta)$ with

$$(\mathfrak{z}_2', \dots, \mathfrak{z}_t') \in C_{3\delta}(\mathfrak{z}_2, \dots, \mathfrak{z}_t). \tag{14.3}$$

When $t = 1$, interpret $\mathscr{N}(\mathscr{B}, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_t)$ to be $\mathscr{N}(\mathscr{B}, \delta)$.

Lemma 20. *Suppose $\mathscr{B} \subset \mathscr{H}$, and $0 < \delta \leqslant m$. Let $\mathscr{N}(\mathscr{B}, \delta) \subset \mathscr{H}(a, b)$ be a δ-net for $\mathscr{B}$, and $(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}) \in \mathscr{X}_{n-1}$. Then there is a $(2\delta|\mathbf{x}_1|)$-net $\mathscr{N}$ for $\mathscr{B}^*(\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$ of cardinality*

$$|\mathscr{N}| \ll \delta^{n-m} \sum_{Z' \in \mathscr{N}(\mathscr{B}, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_{n-1})} (y_n')^{m-n},$$

where $\mathfrak{z}_2, \dots, \mathfrak{z}_{n-1}$ are the vectors determined by $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$.

Proof. Let $(\mathbf{k}_1, \dots, \mathbf{k}_{n-1}) \in \mathscr{K}_{n-1}$ be determined by $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$. Elements $Z' \in \mathscr{H}(a, b)$ have $y_n' \geqslant a^{n-1}$, and then $\delta/y_n' \leqslant ma^{1-n}$.

$S = S(\mathbf{k}_1, \dots, \mathbf{k}_{n-1})$ is a sphere of radius $|\mathbf{x}_1|$ in space of dimension $m - (n-1)$. When $\varepsilon \leqslant ma^{1-n}|\mathbf{x}_1| \ll |\mathbf{x}_1|$, there is an ε-net $\mathscr{N}_1(S, \varepsilon)$ for S of cardinality $\ll (\varepsilon/|\mathbf{x}_1|)^{n-m}$. Let $\mathscr{N}$ consist of points

$$x_{1n}'\mathbf{k}_1 + \cdots + x_{n-1,n}'\mathbf{k}_{n-1} + y_n'\mathbf{k}$$

where $\mathfrak{z}_n'$ belongs to some $Z' = (\mathfrak{z}_2', \dots, \mathfrak{z}_{n-1}', \mathfrak{z}_n') \in \mathscr{N}(\mathscr{B}, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_{n-1}) \subset \subset \mathscr{H}(a, b)$, and where $\mathbf{k} \in \mathscr{N}_1(S, |\mathbf{x}_1|\delta/y_n')$. Then $\mathscr{N}$ has cardinality

$$|\mathscr{N}| \ll \sum_{Z' \in \mathscr{N}(B, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_{n-1})} (\delta/y_n')^{n-m}.$$

Suppose $\mathbf{x} \in \mathscr{B}^*(\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$ is given by (14.1) with $y_n > 0, \mathbf{k} \in S(\mathbf{k}_1, \dots, \mathbf{k}_{n-1})$. Then $\mathfrak{z}_n$ with components $x_{1n}, \dots, x_{n-1,n}, y_n$ has $Z = = (\mathfrak{z}_2, \dots, \mathfrak{z}_n) \in \mathscr{B}$. Pick $Z' \in \mathscr{N}(\mathscr{B}, \delta)$ with $|Z' - Z| \leqslant \delta$. Then (when $n > 2$) certainly $(\mathfrak{z}_2', \dots, \mathfrak{z}_{n-1}') \in C_{2\delta}(\mathfrak{z}_2, \dots, \mathfrak{z}_{n-1})$, so that Z' belongs to $\mathscr{N}(\mathscr{B}, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_{n-1})$. Set

$$\mathbf{x}' = x_{1n}'\mathbf{k}_1 + \cdots + x_{n-1,n}'\mathbf{k}_{n-1} + y_n'\mathbf{k},$$

where $\mathbf{k}$ is the vector in (14.1). Since $|Z' - Z| \leqslant \delta$, we have $|\mathbf{x}' - \mathbf{x}| \leqslant \delta|\mathbf{x}_1|$. There is some $\mathbf{k}' \in \mathscr{N}_1(S, |\mathbf{x}_1|\delta/y_n')$ with $|\mathbf{k}' - \mathbf{k}| \leqslant \delta|\mathbf{x}_1|/y_n'$. The point

$$\mathbf{x}'' = x_{1n}'\mathbf{k}_1 + \cdots + x_{n-1,n}'\mathbf{k}_{n-1} + y_n'\mathbf{k}'$$

lies in $\mathscr{N}$ and has $|\mathbf{x}''-\mathbf{x}'|\leqslant\delta|\mathbf{x}_1|$, therefore $|\mathbf{x}''-\mathbf{x}|\leqslant 2\delta|\mathbf{x}_1|$. We may conclude that $\mathscr{N}$ is a $(2\delta|\mathbf{x}_1|)$-net for $\mathscr{B}^*(\mathbf{x}_1,\dots,\mathbf{x}_{n-1})$.

Suppose $2\leqslant t<n$. Let $\mathscr{R}$ be a set in the space of tuples $(\mathfrak{z}'_{t+1},\dots,\mathfrak{z}'_n)\in\mathbb{R}^{h(n)-h(t)}$. For example, when $\mathscr{D}\subset\mathscr{H}$ and $\mathfrak{z}_2,\dots,\mathfrak{z}_t$ are given, the set $\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$ is of this type. When $\mathscr{R}$ is measurable, set

$$f(\mathscr{R})=\int_{\mathscr{R}}(y'_{t+1})^{m-t-1}\cdots(y'_n)^{m-n}d\mathfrak{z}'_{t+1}\cdots d\mathfrak{z}'_n. \tag{14.4}$$

Lemma 21. *Suppose $\mathscr{D}\subset\mathscr{H}$. Suppose $2\leqslant t<n$, $0<\delta\leqslant m$, and $(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0)\in C_\delta(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$. When $\mathscr{N}(\partial\mathscr{D},\delta)\subset\mathscr{H}(a,b)$ is a δ-net for $\partial\mathscr{D}$, and if $\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$ and $\mathscr{D}(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0)$ are measurable, then*

$$\begin{aligned}&|f(\mathscr{D}(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0))-f(\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t))|\\&\ll\delta^{h(n)-h(t)}\sum_{Z'\in\mathscr{N}(\partial\mathscr{D},\delta,\mathfrak{z}_2,\dots,\mathfrak{z}_t)}(y'_{t+1})^{m-t-1}\cdots(y'_n)^{m-n}.\end{aligned}\tag{14.5}$$

Proof. The cubes $C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)$ with $(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)\in\delta\mathbb{Z}^{h(n)-h(t)}$ cover $\mathbb{R}^{h(n)-h(t)}$. Therefore

$$f(\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t))=\sum_{(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)\in\delta\mathbb{Z}^{h(n)-h(t)}}f(\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)\cap C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)),$$

and an analogous formula holds for $f(\mathscr{D}(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0))$. The left hand side of (14.5) thus is (up to sign) the sum of the numbers

$$g(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)=f(\mathscr{D}(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0)\cap C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n))-f(\mathscr{D}(\mathfrak{z}_2,\dots)\cap C_\delta(\mathfrak{z}^+_{t+1},\dots)) \tag{14.6}$$

as $(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)$ runs through $\delta\mathbb{Z}^{h(n)-h(t)}$. Clearly $g(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)=0$ unless

$$\mathscr{D}(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0)\cap C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)\neq\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)\cap C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n).$$

In the latter case there is some $(\mathfrak{z}^*_{t+1},\dots,\mathfrak{z}^*_n)\in C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n))$ lying in the symmetric difference of $\mathscr{D}(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0)$ and $\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$; say $(\mathfrak{z}^*_{t+1},\dots,\mathfrak{z}^*_n)\in\in\mathscr{D}(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0)\setminus\mathscr{D}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$. Then $(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0,\mathfrak{z}^*_{t+1},\dots,\mathfrak{z}^*_n)\in\mathscr{D}$, $(\mathfrak{z}_2,\dots,\mathfrak{z}_t,\mathfrak{z}^*_{t+1},\dots,\mathfrak{z}^*_n)\notin\mathscr{D}$. Since $(\mathfrak{z}_2^0,\dots,\mathfrak{z}_t^0)\in C_\delta(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$, there is some $(\mathfrak{z}^*_2,\dots,\mathfrak{z}^*_t)\in C_\delta(\mathfrak{z}_2,\dots,\mathfrak{z}_t)$ with $Z^*=(\mathfrak{z}^*_2,\dots,\mathfrak{z}^*_t,\mathfrak{z}^*_{t+1},\dots,\mathfrak{z}^*_n)\in\partial\mathscr{D}$. In short: $g(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)=0$ unless there is a $Z^*\in\partial\mathscr{D}$ with

$$(\mathfrak{z}^*_2,\dots,\mathfrak{z}^*_t)\in C_\delta(\mathfrak{z}_2,\dots,\mathfrak{z}_t)\quad\text{and}\quad(\mathfrak{z}^*_{t+1},\dots,\mathfrak{z}^*_n)\in C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n).$$

Given such Z^*, pick $Z'\in\mathscr{N}(\partial\mathscr{D},\delta)$ with $|Z'-Z^*|\leqslant\delta$; then

$$(\mathfrak{z}'_2,\dots,\mathfrak{z}'_t)\in C_{3\delta}(\mathfrak{z}_2,\dots,\mathfrak{z}_t)\quad\text{and}\quad(\mathfrak{z}'_{t+1},\dots,\mathfrak{z}'_n)\in C_{3\delta}(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n). \tag{14.7}$$

We have

$$\begin{aligned}|g(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n)|&\leqslant f(C_\delta(\mathfrak{z}^+_{t+1},\dots,\mathfrak{z}^+_n))\\&\ll\delta^{h(n)-h(t)}(y'_{t+1})^{m-t+1}\cdots(y'_n)^{m-n},\end{aligned}$$

since $Z' \in \mathcal{N}(\partial\mathcal{D}, \delta) \subset \mathcal{H}(a, b)$ yields $y'_j \geqslant a^{j-1} \gg 1$, so that for $\mathfrak{z}_j \in C_\delta(\mathfrak{z}_j^+)$ $(t < j \leqslant n)$ we have

$$y_j \leqslant y_j^+ + \delta/2 \leqslant y'_j + 2\delta \leqslant y'_j + 2m \ll y'_j.$$

As $(\mathfrak{z}_{t+1}^+, \ldots, \mathfrak{z}_n^+)$ runs through points of $\delta\mathbb{Z}^{h(n)-h(t)}$ with $g(\mathfrak{z}_{t+1}^+, \ldots, \mathfrak{z}_n^+) \neq 0$, then Z' will run through certain elements of $\mathcal{N}(\partial\mathcal{D}, \delta)$ with $(\mathfrak{z}'_2, \ldots, \mathfrak{z}'_t) \in$ $\in C_{3\delta}(\mathfrak{z}_2, \ldots, \mathfrak{z}_t)$, i.e., it will run through certain elements of $\mathcal{N}(\partial\mathcal{D}, \delta, \mathfrak{z}_2, \ldots, \mathfrak{z}_t)$ (see definition (14.2), (14.3)). Given such Z', the number of $(\mathfrak{z}_{t+1}^+, \ldots, \mathfrak{z}_n^+) \in$ $\in \delta\mathbb{Z}^{h(n)-h(t)}$ with (14.7), i.e., with $(\mathfrak{z}_{t+1}^+, \ldots, \mathfrak{z}_n^+) \in C_{3\delta}(\mathfrak{z}'_{t+1}, \ldots, \mathfrak{z}'_n)$, is $\ll 1$. We may conclude that

$$\sum_{(\mathfrak{z}_{t+1}^+, \ldots, \mathfrak{z}_n^+) \in \delta\mathbb{Z}^{h(n)-h(t)}} |g(\mathfrak{z}_{t+1}^+, \ldots, \mathfrak{z}_n^+)| \ll \delta^{h(n)-h(t)} \sum_{Z' \in \mathcal{N}(\partial\mathcal{D}, \delta, \mathfrak{z}_2, \ldots, \mathfrak{z}_t)} (y'_{t+1})^{m-t-1} \cdots (y'_n)^{m-n}.$$

Since the left hand side of (14.5) is (up to sign) the sum of the numbers $g(\mathfrak{z}_{t+1}^+, \ldots, \mathfrak{z}_n^+)$, the lemma follows.

For want of a better place we now insert the following simple lemmas.

Lemma 22. *Let $\mathcal{S} \subset \mathbb{R}^m$, and suppose $\mathcal{S}_m$ (i.e., $\mathcal{S}_\varepsilon$ with $\varepsilon = m$) has finite volume $V(\mathcal{S}_m)$. Then the number $N(\mathcal{S})$ of integer points in $\mathcal{S}$ is $\leqslant V(\mathcal{S}_m)$.*

Proof. Recall the definition $\langle \mathbf{x} \rangle = C_1(\mathbf{x})$ of cubes. Let $\mathcal{S}'$ be the union of the cubes $\langle \mathbf{x} \rangle$ where $\mathbf{x}$ is an integer points in $\mathcal{S}$. Clearly $\mathcal{S}' \subset \mathcal{S}_m$, therefore $N(\mathcal{S}) = V(\mathcal{S}') \leqslant V(\mathcal{S}_m)$.

Corollary. *Suppose $1 \leqslant t < m$, and reals $c_1, \ldots, c_t, \tilde{c}$ are given. Let $\mathcal{S}^* \subset \mathbb{R}^m$ consist of points $(x_1, \ldots, x_t, \tilde{\mathbf{x}})$ with $\tilde{\mathbf{x}} = (x_{t+1}, \ldots, x_n)$ and*

$$|x_i - c_i| < 2m \quad (i = 1, \ldots, t),$$
$$|\,|\tilde{\mathbf{x}}| - \tilde{c}| < 2m.$$

Let $\mathcal{S}$ be obtained from $\mathcal{S}^*$ by a rotation. Then

$$N(\mathcal{S}) \ll (|\tilde{c}| + 1)^{m-t-1}.$$

Proof. $N(\mathcal{S}) \leqslant V(\mathcal{S}_m) = V(\mathcal{S}_m^*) \ll (|\tilde{c}| + 2m)^{m-t-1}$.

Lemma 23. *Suppose $t > -m$. Then*

$$\sum_{\substack{\mathbf{x} \in \mathbb{Z}^m \setminus \{0\} \\ |\mathbf{x}| \leqslant H}} |\mathbf{x}|^t = \frac{m}{m+t} V_m H^{m+t} + O(H^{m+t-1}).$$

Proof. Let $f(r)$ be the number of nonzero points $\mathbf{x} \in \mathbb{Z}^m$ with $|\mathbf{x}| \leqslant r$. Then our sum is

$$\int_0^H r^t df(r) = H^t f(H) - t \int_0^H r^{t-1} f(r) dr.$$

Since $f(r) = V_m r^m + 0(r^{m-1})$, we obtain a main term

$$V_m H^{m+t} - V_m(t/(m+t))H^{m+t} = (m/(m+t))V_m H^{m+t},$$

and (since $t + m > 0$) an error term $\ll H^{m+t-1}$.

15. Proof of Proposition B

Initially suppose that $\mathscr{D}$ is a finite union of cubes, so that all the integrals occurring below exist. Let $1 \leqslant t < n$ and $(\mathbf{x}_1, \dots, \mathbf{x}_t) \in \mathscr{X}_t$, and when $t > 1$, let $\mathfrak{z}_2, \dots, \mathfrak{z}_t$ be determined by $\mathbf{x}_1, \dots, \mathbf{x}_t$. We will suppose that $\delta = m/|\mathbf{x}_1|$ holds, rather than (12.1). Let $\mathscr{N}(\partial\mathscr{D}, \delta) \subset \mathscr{H}(a, b)$ be a δ-net for $\partial\mathscr{D}$. We will show that

$$|N(\mathscr{D}^*(\mathbf{x}_1, \dots, \mathbf{x}_t)) - V(\mathscr{D}^*(\mathbf{x}_1, \dots, \mathbf{x}_t))| \ll |\mathbf{x}_1|^{(n-t)m-h(n)+h(t)} \sum_{Z' \in \mathscr{N}(\partial\mathscr{D}, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_t)} (y'_{t+1})^{m-t-1} \cdots (y'_n)^{m-n}. \tag{15.1}$$

When $t = 1$, this becomes (12.2).

We begin with the case $t = n - 1$. Let $\mathbf{k}_1, \dots, \mathbf{k}_{n-1}$ be determined by $\mathbf{x}_1, \dots, \mathbf{x}_{n-1}$. Then $\mathscr{D}^*(\mathbf{x}_1, \dots, \mathbf{x}_{n-1})$ consists of points

$$\mathbf{x} = x_{1n}\mathbf{k}_1 + \cdots + x_{n-1,n}\mathbf{k}_{n-1} + y_n\mathbf{k}$$

with $\mathbf{k} \in \mathscr{S}(\mathbf{k}_1, \dots, \mathbf{k}_{n-1})$ with $\mathfrak{z}_n \in \mathscr{D}(\mathfrak{z}_2, \dots, \mathfrak{z}_{n-1})$ where as always $\mathfrak{z}_n$ has components $x_{1n}, \dots, x_{n-1,n}, y_n$. We apply Lemma 19, as well as Lemma 20 with $\mathscr{B} = \partial\mathscr{D}$, and with $\delta = m/|\mathbf{x}_1|$, to see that $\partial(\mathscr{D}^*(\mathbf{x}_1, \dots, \mathbf{x}_{n-1}))$ has a $(2\delta|\mathbf{x}_1|)$-net, hence $2m$-net, of cardinality

$$\ll \delta^{n-m} \sum_{Z' \in \mathscr{N}(\partial\mathscr{D}, \delta, \mathfrak{z}_2, \dots, \mathfrak{z}_{n-1})} (y'_n)^{m-n}.$$

Here

$$\delta^{n-m} \ll |\mathbf{x}_1|^{m-n} = |\mathbf{x}_1|^{(n-(n-1))m-h(n)+h(n-1)}.$$

Lemma 14 yields the case $t = n - 1$ of (15.1).

Now let $1 \leqslant t < n - 1$, and let $(\mathbf{x}_1, \dots, \mathbf{x}_t) \in \mathscr{X}_t$ be fixed. Then also $\mathbf{k}_1, \dots, \mathbf{k}_t$ and (when $t > 1$) $\mathfrak{z}_2, \dots, \mathfrak{z}_t$ are fixed. Write points $\mathbf{x}_{t+1}, \mathbf{x}^0_{t+1} \in \mathbb{R}^m$ as

$$\mathbf{x}_{t+1} = x_{1,t+1}\mathbf{k}_1 + \cdots + x_{t,t+1}\mathbf{k}_t + y_{t+1}\mathbf{k}, \tag{15.2}$$

$$\mathbf{x}^0_{t+1} = x^0_{1,t+1}\mathbf{k}_1 + \cdots + x^0_{t,t+1}\mathbf{k}_t + y^0_{t+1}\mathbf{k}^0 \tag{15.3}$$

with positive y_{t+1}, y^0_{t+1} and with $\mathbf{k}, \mathbf{k}^0 \in S(\mathbf{k}_1, \dots, \mathbf{k}_t)$. Suppose that

$$\mathbf{x}^0_{t+1} \in \langle \mathbf{x}_{t+1} \rangle.$$

Then $|x^0_{i,t+1} - x_{i,t+1}| < \frac{1}{2}m/|\mathbf{x}_1| = \delta/2$ $(1 \leqslant i \leqslant t)$, $|y^0_{t+1} - y_{t+1}| < \delta/2$. Therefore $\mathfrak{z}^0_{t+1} \in C_\delta(\mathfrak{z}_{t+1})$,

$$(\mathfrak{z}_2, \dots, \mathfrak{z}_t, \mathfrak{z}^0_{t+1}) \in C_\delta(\mathfrak{z}_2, \dots, \mathfrak{z}_t, \mathfrak{z}_{t+1}).$$

By (13.1) with $t + 1$ in place of t, by Lemma 21 with $t + 1$ in place of t, and

since $\delta = m/|\mathbf{x}_1|$,

$$|V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\mathbf{x}_{t+1})) - V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\mathbf{x}_{t+1}^0))| \ll |\mathbf{x}_1|^{m(n-t-1)}|\mathbf{x}_1|^{h(t+1)-h(n)} \sum_{Z'\in\mathscr{N}(\partial\mathscr{D},\delta,\mathfrak{z}_2,\dots,\mathfrak{z}_t,\mathfrak{z}_{t+1})} (y'_{t+2})^{m-t-2}\cdots(y'_n)^{m-n}. \tag{15.4}$$

Set

$$V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\langle\mathbf{x}_{t+1}\rangle)) = \int_{\langle\mathbf{x}_{t+1}\rangle} V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\mathbf{x}_{t+1}^0))d\mathbf{x}_{t+1}^0.$$

Then the upper bound (15.4) also holds for

$$|V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\mathbf{x}_{t+1})) - V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\langle\mathbf{x}_{t+1}\rangle))|.$$

Assuming the truth of (15.1) for $t+1$, we obtain

$$|N(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\mathbf{x}_{t+1})) - V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\langle\mathbf{x}_{t+1}\rangle))| \ll |\mathbf{x}_1|^{m(n-t-1)+h(t+1)-h(n)} \sum_{Z'\in\mathscr{N}(\partial\mathscr{D},\delta,\mathfrak{z}_2,\dots,\mathfrak{z}_t,\mathfrak{z}_{t+1})} (y'_{t+2})^{m-t-2}\cdots(y'_n)^{m-n}. \tag{15.5}$$

Now

$$N(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t)) - V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t)) = \sum_{\mathbf{x}_{t+1}\in\mathbb{Z}^m} (N(\mathscr{D}^*(\mathbf{x}_1,\dots\mathbf{x}_t,\mathbf{x}_{t+1})) - V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t,\langle\mathbf{x}_{t+1}\rangle))). \tag{15.6}$$

Here $\mathbf{x}_1,\dots,\mathbf{x}_t$, hence $\mathfrak{z}_2,\dots,\mathfrak{z}_t$ (when $t>1$) are fixed. But $\mathfrak{z}_{t+1}$ in (15.5) also depends on $\mathbf{x}_{t+1}$; write $\mathfrak{z}_{t+1} = \mathfrak{z}_{t+1}(\mathbf{x}_{t+1})$. For $Z'\in\mathscr{N}(\partial\mathscr{D},\delta,\mathfrak{z}_2,\dots,\mathfrak{z}_t)$, let $A(Z')$ be the number of points $\mathbf{x}_{t+1}$ written as (15.2) for which $Z'\in\mathscr{N}(\partial\mathscr{D},\delta,\mathfrak{z}_2,\dots,\mathfrak{z}_t,\mathfrak{z}_{t+1})$. Such $\mathbf{x}_{t+1}$ have $\mathfrak{z}_{t+1} = \mathfrak{z}_{t+1}(\mathbf{x}_{t+1}) \in C_{3\delta}(\mathfrak{z}'_{t+1})$ (see definition (14.2), (14.3)). Thus $A(Z')$ counts integer points (15.2) with

$$|x_{i,t+1} - x'_{i,t+1}| \leqslant 3\delta/2, \quad |y_{t+1} - y'_{t+1}| \leqslant 3\delta/2.$$

Since $|\mathbf{k}_1| = \cdots = |\mathbf{k}_t| = |\mathbf{x}_1|$, since $\mathbf{k}$ ranges through a sphere of radius $|\mathbf{x}_1|$, and since $(3\delta/2)|\mathbf{x}_1| = (3/2)m < 2m$, our $\mathbf{x}_{t+1}$ lies in a set $\mathscr{S}$ as in the Corollary to Lemma 22, with $\tilde{c} = |\mathbf{x}_1|y'_{t+1}$, so that

$$A(Z') = N(\mathscr{S}) \ll (|\mathbf{x}_1|y'_{t+1}+1)^{m-t-1} \ll (|\mathbf{x}_1|y'_{t+1})^{m-t-1}. \tag{15.7}$$

(Recall that $y'_j \gg 1$ for $Z'\in\mathscr{H}(a,b)$.) In view of (15.5), (15.6) we obtain

$$N(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t)) - V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t))| \ll |\mathbf{x}_1|^{m(n-t-1)+h(t+1)-h(n)} \sum_{Z'\in\mathscr{N}(\partial\mathscr{D},\delta,\mathfrak{z}_2,\dots,\mathfrak{z}_t)} A(Z')(y'_{t+2})^{m-t-2}\cdots(y'_n)^{m-n}$$
$$\ll |\mathbf{x}_1|^{m(n-t)+h(t)-h(n)} \sum_{Z'\in\mathscr{N}(\partial\mathscr{D},\delta,\mathfrak{z}_2,\dots,\mathfrak{z}_t)} (y'_{t+1})^{m-t-1}\cdots(y'_n)^{m-n}$$

on observing (15.7) and $h(t+1)-t-1=h(t)$. Thus (15.1) holds for t. It therefore holds for $1\leqslant t\leqslant n-1$.

Now drop the hypothesis that $\mathscr{D}$ be a finite union of cubes. When $\mathscr{D}$ is bounded and Jordan-measurable, there is a finite union $\mathscr{D}_1\supset\mathscr{D}$ of cubes with $V(\mathscr{D}_1^*(\mathbf{x}_1))$ arbitrarily close to $V(\mathscr{D}^*(\mathbf{x}_1))$, and such that a δ-net $\mathscr{N}(\partial\mathscr{D},\delta)$ for $\partial\mathscr{D}$ becomes a (2δ)-net $\mathscr{N}(\partial\mathscr{D}_1,2\delta)$ for $\partial\mathscr{D}_1$. With δ given by (12.1), we have $2\delta=m/|\mathbf{x}_1|$. Applying the case $t=1$ of (15.1) to $\mathscr{D}_1$, we obtain

$$N(\mathscr{D}^*(\mathbf{x}_1))\leqslant N(\mathscr{D}_1^*(\mathbf{x}_1))\leqslant V(\mathscr{D}_1^*(\mathbf{x}_1))+B$$

where

$$B\ll|\mathbf{x}_1|^{(n-1)m-h}\sum_{Z\in\mathscr{N}(\partial\mathscr{D}_1,2\delta)}y_2^{m-2}\cdots y_n^{m-n}.$$

Since $\mathscr{N}(\partial\mathscr{D}_1,2\delta)=\mathscr{N}(\partial\mathscr{D},\delta)$, and since $V(\mathscr{D}_1^*(\mathbf{x}_1))$ is arbitrarily close to $V(\mathscr{D}^*(\mathbf{x}_1))$, we obtain the desired upper bound for $N(\mathscr{D}^*(\mathbf{x}_1))$. The lower bound is derived similarly.

16. Further Auxiliary Lemmas

Set $\Pi'(\mathbf{a})=1$ when $n=2$,

$$\Pi'(\mathbf{a})=\prod_{i=2}^{n-1}a_i^{-(n-i)(i-n/(n-1))}\tag{16.1}$$

when $n>2$.

Lemma 24. *Suppose $n>2$, $a_2>0,\dots,a_{n-1}>0, K>0$. Then the integral*

$$I=\int\alpha_3^{-2}\alpha_4^{-4}\cdots\alpha_n^{-2n+4}d\alpha_3\cdots d\alpha_n$$

over the domain of $(\alpha_3,\dots,\alpha_n)\in\mathbb{R}^{n-2}$ with $\alpha_i\geqslant a_{i-1}\alpha_{i-1}/2>0$ $(i=3,\dots,n)$ where

$$\alpha_2=K/(\alpha_3\cdots\alpha_n)$$

has

$$I\ll K^{-n+3-1/(n-1)}\Pi'(\mathbf{a}).\tag{16.2}$$

Proof. When $n=3$, we have $\alpha_3^2\gg a_2\alpha_2\alpha_3=a_2K$, so that

$$I=\int\alpha_3^{-2}d\alpha_3\ll(a_2K)^{-1/2},$$

which is what we want. Suppose the case $n-1$ has been established. We have $\alpha_2=K'/(\alpha_3\cdots\alpha_{n-1})$ with $K'=K/\alpha_n$. By the case $n-1$, with K' in place of K,

$$I\ll K^{-n+4-1/(n-2)}\prod_{j=2}^{n-2}a_j^{-(n-1-j)(j-(n-1)/(n-2))}\int\alpha_n^{-l}d\alpha_n\tag{16.3}$$

with $l = 2n - 4 + (-n + 4 - 1/(n-2)) = n - 1/(n-2)$. Here

$$\alpha_n^{n-1} \gg a_{n-1}^{n-2}\alpha_{n-1}^{n-2}\alpha_n \gg a_{n-1}^{n-2}a_{n-2}^{n-3}\alpha_{n-2}^{n-3}\alpha_{n-1}\alpha_n$$
$$\gg \cdots \gg a_{n-1}^{n-2}\cdots a_3^2 a_2\alpha_2\alpha_3\cdots\alpha_n = Ka_2a_3^2\cdots a_{n-1}^{n-2},$$

so that

$$\int \alpha_n^{-l}d\alpha_n \ll (Ka_2a_3^2\cdots a_{n-1}^{n-2})^{-(l-1)/(n-1)}.$$

When we substitute this into (16.3) we obtain (16.2).

Lemma 25. *Let $\mathscr{A} \subset \mathbb{R}^{h-1}$ be bounded, $\eta > 0$, and $\boldsymbol{\alpha} \mapsto Z(\boldsymbol{\alpha})$ be an η-Lipschitz map $\mathscr{A}_{1/\eta} \to \mathscr{H}$. Suppose $\mathscr{B} \subset Z(\mathscr{A}) \cap \mathscr{H}(a,b)$, and $0 < \delta \leqslant m$. Then there is a δ-net $\mathscr{N}(\mathscr{B},\delta) \subset \mathscr{B}$ for $\mathscr{B}$ with*

$$\sum_{Z\in\mathscr{N}(\mathscr{B},\delta)} y_2^{m-2}\cdots y_n^{m-n} \ll (\eta/\delta)^{h-1}\int_{\mathscr{A}_{1/\eta}} y_2(\boldsymbol{\alpha})^{m-2}\cdots y_n(\boldsymbol{\alpha})^{m-n}d\boldsymbol{\alpha}. \tag{16.4}$$

Proof. Replacing $\mathscr{A}$ by $\eta\mathscr{A}$ and $Z(\boldsymbol{\alpha})$ by $Z(\eta^{-1}\boldsymbol{\alpha})$ we may suppose without loss of generality that $\eta = 1$. We further may replace $\mathscr{A}$ by the preimage of $\mathscr{B}$ under $\boldsymbol{\alpha} \mapsto Z(\boldsymbol{\alpha})$. Hence we may suppose that $\mathscr{B} = Z(\mathscr{A}) \subset \mathscr{H}(a,b)$. Since a δ-net is a δ'-net for $\delta' > \delta$, we may suppose that $\delta \leqslant \frac{1}{2}\tilde{a}^{n-1}$ where $\tilde{a} = \min(1,a)$.

Let $\mathscr{C}$ be the set of cubes $C_{\delta/mh}(\boldsymbol{\alpha})$ which intersect $\mathscr{A}$ and where $\boldsymbol{\alpha} \in (\delta/mh)\mathbb{Z}^{h-1}$. These cubes are contained in $\mathscr{A}_{\delta/m} \subset \mathscr{A}_1$, and their union covers $\mathscr{A}$. For $C \in \mathscr{C}$ pick $\boldsymbol{\alpha}_C \in C \cap \mathscr{A}$. Then $Z(\boldsymbol{\alpha}_C) \in \mathscr{B} \subset \mathscr{H}(a,b)$, so $y_i(\boldsymbol{\alpha}_C) \geqslant \tilde{a}^{n-1} \geqslant 2\delta$. When $\boldsymbol{\alpha} \in C$, we have $y_i(\boldsymbol{\alpha}) \geqslant y_i(\boldsymbol{\alpha}_C) - \delta \geqslant \frac{1}{2}y_i(\boldsymbol{\alpha}_C)$, so that

$$y_2(\boldsymbol{\alpha}_C)^{m-2}\cdots y_n(\boldsymbol{\alpha}_C)^{m-n} < 2^{mn}y_2(\boldsymbol{\alpha})^{m-2}\cdots y_n(\boldsymbol{\alpha})^{m-n},$$

and

$$y_2(\boldsymbol{\alpha}_C)^{m-2}\cdots y_n(\boldsymbol{\alpha}_C)^{m-n} \leqslant (mh/\delta)^{h-1}2^{mn}\int_C y_2(\boldsymbol{\alpha})^{m-2}\cdots y_n(\boldsymbol{\alpha})^{m-n}d\boldsymbol{\alpha}.$$

Therefore when $\mathscr{N}(\mathscr{B},\delta)$ is the set of points $Z(\boldsymbol{\alpha}_C)$ with $C \in \mathscr{C}$, the estimate (16.4) follows.

When $Z \in \mathscr{B} = Z(\mathscr{A})$, say $Z = Z(\boldsymbol{\alpha})$ with $\boldsymbol{\alpha} \in \mathscr{A}$, and say $\boldsymbol{\alpha} \in C, C \in \mathscr{C}$, then

$$|Z - Z(\boldsymbol{\alpha}_C)| = |Z(\boldsymbol{\alpha}) - Z(\boldsymbol{\alpha}_C)| \leqslant |\boldsymbol{\alpha} - \boldsymbol{\alpha}_C| < \delta.$$

Therefore $\mathscr{N}(\mathscr{B},\delta)$ is indeed a δ-net for $\mathscr{B}$.

Remark. Since $\mathscr{A}$ is bounded, so is $\mathscr{B}$. The cardinality of $\mathscr{N}(\mathscr{B},\delta)$ is $\ll$ the right hand side of (16.4). Therefore $\mathscr{B}$ has Jordan measure zero.

Given $\mathscr{D} \subset \mathscr{H}$, recall that $\mathscr{D}^{(K)}$ consists of $Z \in \mathscr{D}$ having

$$y_2y_3\cdots y_n \leqslant K, \tag{16.5}$$

and $\mathscr{D}^{[K]}$ of $Z \in \mathscr{D}$ with

$$y_2y_3\cdots y_n = K. \tag{16.6}$$

Lemma 26. *Suppose $0 < \delta \leqslant m$. Then there is a δ-net $\mathscr{N} = \mathscr{N}(\mathscr{H}(\mathbf{a}, b)^{[K]}, \delta)$ $\subset \mathscr{H}(a, b)$ for $\mathscr{H}(\mathbf{a}, b)^{[K]}$ with*

$$\sum_{Z \in \mathscr{N}} y_2^{m-2} \cdots y_n^{m-n} \ll \delta^{1-h} K^{m-n/(n-1)} \Pi'(\mathbf{a}),$$

where $\Pi'(\mathbf{a})$ is given by (16.1).

Proof. Write points $\boldsymbol{\alpha} \in \mathbb{R}^{h-1}$ as $\boldsymbol{\alpha} = (\alpha_3, \ldots, \alpha_n, \beta_{12}, \ldots, \beta_{n-1,n})$. For $\boldsymbol{\alpha}$ with positive $\alpha_3, \ldots, \alpha_n$ set

$$\alpha_2 = \alpha_2(\boldsymbol{\alpha}) = K/(\alpha_3 \cdots \alpha_n). \tag{16.7}$$

(When $n = 2$, then $\boldsymbol{\alpha} = (\beta_{12})$ and $\alpha_2 = \alpha_2(\boldsymbol{\alpha}) = K$.) Let $\mathscr{A} \subset \mathbb{R}^{h-1}$ consist of points $\boldsymbol{\alpha}$ with

$$\alpha_2 \geqslant a_1, \quad \alpha_i \geqslant a_{i-1}\alpha_{i-1} \quad (i = 3, \ldots, n)$$

and

$$|\beta_{1j}| \leqslant b \quad (1 \leqslant j \leqslant n), \quad |\beta_{ij}| \leqslant b\alpha_i \quad (2 \leqslant i < j \leqslant n),$$

where α_2 is given by (16.7). $\mathscr{A}$ is bounded. When η is sufficiently large in terms of n, a, b (note that (2.5) holds), then

$$\alpha_2 \geqslant a_1/2, \quad \alpha_i \geqslant a_{i-1}\alpha_{i-1}/2 \quad (i = 3, \ldots, n), \tag{16.8}$$

$$|\beta_{1j}| \leqslant 2b \quad (1 \leqslant j \leqslant n), \quad |\beta_{ij}| \leqslant 2b\alpha_i \quad (2 \leqslant i < j \leqslant n) \tag{16.9}$$

for $\boldsymbol{\alpha} \in \mathscr{A}_{1/\eta}$. Let $\boldsymbol{\alpha} \mapsto Z(\boldsymbol{\alpha})$ be the map $\mathscr{A}_{1/\eta} \to \mathscr{H}$ given by

$$y_i(\boldsymbol{\alpha}) = \alpha_i \quad (2 \leqslant i \leqslant n), \quad x_{ij}(\boldsymbol{\alpha}) = \beta_{ij} \quad (2 \leqslant i < j \leqslant n),$$

where again $\alpha_2 = \alpha_2(\boldsymbol{\alpha})$ is given by (16.7). Then $\mathscr{H}(\mathbf{a}, b)^{[K]} \subset Z(\mathscr{A})$. Further

$$\left|\frac{\partial y_2(\boldsymbol{\alpha})}{\partial \alpha_i}\right| = K/(\alpha_3 \cdots \alpha_i^2 \cdots \alpha_n) = \alpha_2/\alpha_i \ll 1 \quad (i = 3, \ldots, n),$$

so that our map $\boldsymbol{\alpha} \mapsto Z(\boldsymbol{\alpha})$ is η-Lipschitz on $\mathscr{A}_{1/\eta}$ for some $\eta = \eta(n, a, b) \ll 1$. By Lemma 25, there is a δ-net $\mathscr{N} \subset \mathscr{H}(a, b)$ for $\mathscr{H}(\mathbf{a}, b)^{[K]}$ with

$$\sum_{Z \in \mathscr{N}} y_2^{m-2} \cdots y_n^{m-n} \ll \delta^{1-h} \int_{\mathscr{A}_{1/\eta}} y_2(\boldsymbol{\alpha})^{m-2} \cdots y_n(\boldsymbol{\alpha})^{m-n} d\boldsymbol{\alpha}.$$

In view of (16.8), (16.9), integration over the β_{ij} gives $\ll \alpha_2^{n-2}\alpha_3^{n-3} \cdots \alpha_{n-1}$, so that we obtain

$$\ll \delta^{1-h} \int \alpha_2(\boldsymbol{\alpha})^{m+n-4} \alpha_3^{m+n-6} \cdots \alpha_n^{m-n} d\alpha_3 \cdots d\alpha_n$$

$$= \delta^{1-h} K^{m+n-4} \int \alpha_3^{-2}\alpha_4^{-4} \cdots \alpha_n^{-2n+4} d\alpha_3 \cdots d\alpha_n,$$

where the last integral is over the domain specified in Lemma 24, and is to be interpreted as 1 when $n = 2$. We obtain the desired conclusion by substituting the estimate of that lemma.

Now suppose $\mathscr{D} \subset \mathscr{H}(\mathbf{a}, b)$ has boundary $\partial\mathscr{D}$ of spread $\leqslant\tau$. Say $\partial\mathscr{D} = \mathscr{B}_1 \cup \cdots \cup \mathscr{B}_t$ where $\mathscr{B}_j$ is of simple spread τ_j, and $\tau_1 + \cdots + \tau_t \leqslant \tau$. Then

$$\partial(\mathscr{D}^{(K)}) \subset \mathscr{B}_1^{(K)} \cup \cdots \cup \mathscr{B}_t^{(K)} \cup \mathscr{H}(\mathbf{a}, b)^{[K]}.$$

Let $\mathscr{A}(j), \boldsymbol{\alpha} \mapsto Z^{(j)}(\boldsymbol{\alpha})$ be the set and the map evidencing that $\mathscr{B}_j$ is of spread $\leqslant\tau_j$ $(j = 1, \ldots, t)$. Let $\mathscr{A}(j)^{(K)}$ consist of $\boldsymbol{\alpha} \in \mathscr{A}(j)$ with $Z^{(j)}(\boldsymbol{\alpha}) \subset \mathscr{B}_j^{(K)}$. Set

$$I(j)^{(K)} = \int_{(\mathscr{A}(j)^{(K)})_1} (y_2^{(j)}(\boldsymbol{\alpha}))^{m-2} \cdots (y_n^{(j)}(\boldsymbol{\alpha}))^{m-n} d\boldsymbol{\alpha}. \tag{16.10}$$

Since the integral over $\mathscr{A}(j)_1$ occurring in the definition of spread is finite, and since each $y_i^{(j)}(\boldsymbol{\alpha})$ is bounded when $\boldsymbol{\alpha} \in \mathscr{A}(j)^{(K)}$, hence bounded when $\boldsymbol{\alpha} \in (\mathscr{A}(j)^{(K)})_1$, the volume of $(\mathscr{A}(j)^{(K)})_1$ is bounded, so that $I(j)^{(K)}$ is finite. Moreover $\mathscr{A}(j)^{(K)}$ is bounded. By combining Lemmas 25, 26 we see that for $0 < \delta \leqslant m$, there is a δ-net $\mathscr{N} = \mathscr{N}(\partial(\mathscr{D}^{(K)}), \delta) \subset \mathscr{H}(a, b)$ for $\partial(\mathscr{D}^{(K)})$ with

$$\sum_{Z \in \mathscr{N}} y_2^{m-2} \cdots y_n^{m-n} \ll \sum_{j=1}^{t} \delta^{1-h} I(j)^{(K)} + \delta^{1-h} K^{m-n/(n-1)} \Pi'(\mathbf{a}). \tag{16.11}$$

In particular, $\mathscr{D}^{(K)}$ is Jordan-measurable, so that also $\mathscr{D}$ is Jordan-measurable.

We now set $\delta = m/(2|\mathbf{x}_1|)$ and apply Proposition B, to obtain

$$\begin{aligned} &|N(\mathscr{D}^{(K)*}(\mathbf{x}_1)) - V(\mathscr{D}^{(K)*}(\mathbf{x}_1))| \\ &\ll |\mathbf{x}_1|^{(n-1)m-1} \left(\sum_{j=1}^{t} I(j)^{(K)} + K^{m-n/(n-1)} \Pi'(\mathbf{a}) \right). \end{aligned} \tag{16.12}$$

Let μ_1 be the measure on $\mathscr{H}$ introduced in Section 4.

Lemma 27. *Suppose $\mathscr{D} \subset \mathscr{H}(a, b)$ has boundary $\partial\mathscr{D}$ of spread $\leqslant\tau$. Then*

$$\mu_1(\mathscr{D}) \leqslant 2b\tau.$$

Proof. For a set $\mathscr{S} \subset \mathscr{H} \subset \mathbb{R}^h$, let $\mathscr{S}^p$ be its projection on the $\hat{x}_{12}, y_2, x_{13}, x_{23}, y_3, \ldots, y_n$-coordinate plane $\mathscr{P}$. Since $|x_{12}| \leqslant b$ for $Z \in \mathscr{H}(a, b)$,

$$\mu_1(\mathscr{D}) \leqslant 2b \int_{\mathscr{D}^p} \frac{dy_2}{y_2^{2-1/n}} \frac{d\mathfrak{z}_3}{y_3^{3-1/n}} \cdots \frac{d\mathfrak{z}_n}{y_n^{n-1/n}}.$$

When $\partial\mathscr{D} = \mathscr{B}_1 \cup \cdots \cup \mathscr{B}_t$ and when $\mathscr{A}(j), Z^{(j)} (j = 1, \ldots, t)$ are as above, then

$$\mathscr{D}^p = \bigcup_{j=1}^{t} \mathscr{B}_j^p \subset \bigcup_{j=1}^{t} (Z^{(j)}(\mathscr{A}(j)_1))^p.$$

Since the map $\boldsymbol{\alpha} \mapsto (y_2^{(j)}(\boldsymbol{\alpha}), x_{13}^{(j)}(\boldsymbol{\alpha}), \ldots, y_n^{(j)}(\boldsymbol{\alpha}))$ into $\mathscr{P}$ is 1-Lipschitz, hence cannot increase volumes,

$$\begin{aligned} &\int_{(Z^{(j)}(\mathscr{A}(j)_1))^p} y_2^{-2+1/n} \cdots y_n^{-n+1/n} dy_2 d\mathfrak{z}_3 \cdots d\mathfrak{z}_n \\ &\leqslant \int_{\mathscr{A}(j)_1} y_2^{(j)}(\boldsymbol{\alpha})^{-2+1/n} \cdots y_n^{(j)}(\boldsymbol{\alpha})^{-n+1/n} d\boldsymbol{\alpha} \leqslant \tau_j. \end{aligned}$$

We obtain

$$\mu_1(\mathscr{D}) \leqslant 2b \sum_{j=1}^{t} \tau_j \leqslant 2b\tau.$$

17. Proof of Propositions C and D

When $X = (\mathbf{x}_1, \dots, \mathbf{x}_n) = KZ$ as in (1.2), then $\det X = |\mathbf{x}_1|^n y_2 \cdots y_n$. Therefore $X \in \mathscr{D}^*(T)$ implies

$$y_2 \cdots y_n \leqslant T/|\mathbf{x}_1|^n = K(\mathbf{x}_1), \tag{17.1}$$

say. Thus for given $\mathbf{x}_1$, we have $Z \in \mathscr{D}^{(K)}$ with $K = K(\mathbf{x}_1)$. Therefore

$$V(\mathscr{D}^*(T)) = \int V(\mathscr{D}^{(K)*}(\mathbf{x}_1))d\mathbf{x}_1, \tag{17.2}$$

with $K = K(\mathbf{x}_1)$ given by (17.1). We have $y_2 \cdots y_n \geqslant a_1^{n-1} \cdots a_{n-1}$ in $\mathscr{H}(\mathbf{a}, b)$, so that we may restrict the integral (17.2) to the ball

$$|\mathbf{x}_1| \leqslant (T/a_1^{n-1} \cdots a_{n-1})^{1/n} = L, \tag{17.3}$$

say. By Proposition A, and interchanging the order of integration

$$\begin{aligned} V(\mathscr{D}^*(T)) &= c_6(m-1, n-1) \int |\mathbf{x}_1|^{mn-m} \left(\int_{\mathscr{D}^{(K)}} y_2^{m-2} \cdots y_n^{m-n} d\mathfrak{z}_2 \cdots d\mathfrak{z}_n \right) d\mathbf{x}_1 \\ &= c_6(m-1, n-1) \int_{\mathscr{D}} \left(\int |\mathbf{x}_1|^{mn-m} d\mathbf{x}_1 \right) y_2^{m-2} \cdots y_n^{m-n} d\mathfrak{z}_2 \cdots d\mathfrak{z}_n, \end{aligned} \tag{17.4}$$

where the inner integral is over $\mathbf{x}_1 \in \mathbb{R}^m$ with

$$|\mathbf{x}_1|^n \leqslant T/y_2 \cdots y_n, \tag{17.5}$$

and equals $(V_m/n)(T/y_2 \cdots y_n)^m$. Further, since by (4.8), $(V_m/n)c_6(m-1, n-1) = (mn)^{-1} c_6(m, n)$, Proposition C follows.

We now turn to Proposition D. In analogy to (17.2),

$$N(\mathscr{D}^*(T)) = \sum_{\mathbf{x}_1 \in \mathbb{Z}^m \setminus \{\mathbf{0}\}} N(\mathscr{D}^{(K)*}(\mathbf{x}_1))$$

with $K = K(\mathbf{x}_1)$. We may restrict the sum to $\mathbf{x}_1$ with (17.3). By Proposition B, and in view of Proposition A, the sum is close to

$$\begin{aligned} \sum_{\mathbf{x}_1 \in \mathbb{Z}^m \setminus \{\mathbf{0}\}} V(\mathscr{D}^{(K)*}(\mathbf{x}_1)) &= c_6(m-1, n-1) \sum_{\mathbf{x}_1} |\mathbf{x}_1|^{mn-m} \int_{\mathscr{D}^{(K)}} y_2^{m-2} \cdots y_n^{m-n} d\mathfrak{z}_2 \cdots d\mathfrak{z}_n \\ &= c_6(m-1, n-1) \int_{\mathscr{D}} \left(\sum_{\mathbf{x}_1} |\mathbf{x}_1|^{mn-m} \right) y_2^{m-2} \cdots y_n^{m-n} d\mathfrak{z}_2 \cdots d\mathfrak{z}_n, \end{aligned} \tag{17.6}$$

where the sum inside the integral is over nonzero integer points with (17.5). By Lemma 23, this inner sum is

$$(V_m/n)(T/y_2\cdots y_n)^m + O((T/y_2\cdots y_n)^{m-1/n}).$$

Comparison with (17.4) shows that (17.6) equals

$$V(\mathscr{D}^*(T)) + O(\mu_1(\mathscr{D})T^{m-1/n}).$$

Here $\mu_1(\mathscr{D}) \ll \tau$ by Lemma 27.

By (16.12),

$$N(\mathscr{D}^*(T)) = V(\mathscr{D}^*(T)) + O\left(A + B + \sum_{j=1}^{t} C(j)\right)$$

with

$$A = \tau T^{m-1/n},$$
$$B = \Pi'(\mathbf{a}) \sum_{\mathbf{x}_1} |\mathbf{x}_1|^{mn-m-1} K(\mathbf{x}_1)^{m-n/(n-1)},$$
$$C(j) = \sum_{\mathbf{x}_1} |\mathbf{x}_1|^{mn-m-1} I(j)^{(K(\mathbf{x}_1))},$$

where the sums are over nonzero integer points with (17.3).

One finds

$$B = \Pi'(\mathbf{a})T^{m-n/(n-1)} \sum_{0<|\mathbf{x}_1|\leqslant L} |\mathbf{x}_1|^{-m-1+n^2/(n-1)}$$
$$\ll \Pi'(\mathbf{a})T^{m-n/(n-1)}L^{(n^2/(n-1))-1}$$

by Lemma 23, and this equals

$$\Pi'(\mathbf{a})(a_1^{n-1}\cdots a_{n-1})^{-(n/(n-1))+1/n}T^{m-1/n} = a_1^{-1}\Pi(\mathbf{a})T^{m-1/n}$$

by (17.3) and the definitions (2.3), (16.1) of $\Pi(\mathbf{a}), \Pi'(\mathbf{a})$. When $\boldsymbol{\alpha} \in (\mathscr{A}(j)^{(K)})_1$, say $|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0| \leqslant 1$ where $\boldsymbol{\alpha}_0 \in \mathscr{A}(j)^{(K)}$, then $|Z^j(\boldsymbol{\alpha}) - Z^{(j)}(\boldsymbol{\alpha}_0)| \leqslant 1$, so that $y_i^{(j)}(\boldsymbol{\alpha}) \leqslant$ $\leqslant y_i^{(j)}(\boldsymbol{\alpha}_0) + 1 \ll y_i^{(j)}(\boldsymbol{\alpha}_0)$, therefore

$$y_2^{(j)}(\boldsymbol{\alpha})\cdots y_n^{(j)}(\boldsymbol{\alpha}) \ll y_2^{(j)}(\boldsymbol{\alpha}_0)\cdots y_n^{(j)}(\boldsymbol{\alpha}_0) \leqslant K = T/|\mathbf{x}_1|^n.$$

By (16.10)

$$C(j) = \int_{\mathscr{A}(j)_1} \left(\sum |\mathbf{x}_1|^{mn-m-1}\right)(y_2^{(j)}(\boldsymbol{\alpha}))^{m-2}\cdots(y_n^{(j)}(\boldsymbol{\alpha}))^{m-n}d\boldsymbol{\alpha},$$

where the sum is over nonzero integer points $\mathbf{x}_1$ with $|\mathbf{x}_1|^n y_2^{(j)}(\boldsymbol{\alpha})\cdots y_n^{(j)}(\boldsymbol{\alpha}) \ll T$, hence by Lemma 23 is $\ll T^{m-1/n}(y_2^{(j)}(\boldsymbol{\alpha})\cdots y_n^{(j)}(\boldsymbol{\alpha}))^{-m+1/n}$. Therefore, noting that $\mathscr{B}_j$ had simple spread $\leqslant \tau_j$, we get

$$C(j) \ll \tau_j T^{m-1/n}.$$

Combining our estimates, and recalling that $\tau_1 + \cdots + \tau_t \leqslant \tau$, we obtain

$$|N(\mathscr{D}^*(T)) - V(\mathscr{D}^*(T))| \ll (a_1^{-1}\Pi(\mathbf{a}) + \tau)T^{m-1/n}. \tag{17.7}$$

18. Outline of a Proof of Theorem 3

Let $1 < n < m$. Let ν_m be the Haar measure on the orthogonal group O_m, normalized so that $\nu_m(O_m) = 1$. Since O_m consists of matrices $U = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ with $\mathbf{u}_i\mathbf{u}_j = \delta_{ij}(1 \leqslant i,j \leqslant m)$, the measure ν_m induces by projection a measure ν_{mn} on the space $\mathscr{O}$ of matrices $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ with $\mathbf{u}_i\mathbf{u}_j = \delta_{ij}(1 \leqslant i,j \leqslant n)$, where again the columns $\mathbf{u}_i$ lie in $\mathbb{R}^m$.

Let f be a real-valued function on $\mathscr{O}$ with $f(U) = f(-U)$. Let $\mathscr{D} \subset \mathscr{H}$ be lean. A lattice Λ with basis KZ where $Z \in \mathscr{D}$ and $K \in \mathscr{K}$, has $g = g(Z)$ bases $K_1Z, \ldots, K_gZ$ with $K_i \in \mathscr{K}$. Here $K_i = t_iU_i$ with $t_i > 0$ and $U_i \in \mathscr{O}$. When $g(Z) = 2$ there are only two such U_i, e.g., $U, -U$, so that $F(\Lambda) := f(U)$ is well defined; when $g(Z) > 2$, set $F(\Lambda) = 0$, say. We will show that when f is continuous,

$$\sum_{\Lambda \in [\mathscr{D},T]} F(\Lambda) \sim c_1(m,n)\left(\int_{\mathscr{O}} f(U)d\nu_{mn}(U)\right)\mu(\mathscr{D})T^m. \tag{18.1}$$

Let Sp^j be a sphere in j-dimensional space, and σ^j the spherical measure on Sp^j with $\sigma^j(\mathrm{Sp}^j) = 1$. When $0 < \eta < 1/2$ and $\mathbf{u} \in \mathrm{Sp}^j$, let $\mathrm{Sp}^j(\eta, \mathbf{u})$ be the spherical cap on Sp^j with center $\mathbf{u}$ and $\sigma^j(\mathrm{Sp}^j(\eta, \mathbf{u})) = \eta$. Given $V = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \in \mathscr{O}$ and small $\eta > 0$, we construct a "multicap" $C(\eta, V) \subset \mathscr{O}$ as follows. A point $U = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ lies in $C(\eta, V)$ if $\mathbf{u}_1 \in \mathrm{Sp}^m(\eta, \mathbf{v}_1)$ where Sp^m is the sphere of radius 1 in $\mathbb{R}^m$, and $\mathbf{u}_j \in \mathrm{Sp}^{m-j+1}(\eta, \mathbf{v}^{(j)})(2 \leqslant j \leqslant n)$, where $\mathrm{Sp}^{m-j+1} = S(\mathbf{u}_1, \ldots, \mathbf{u}_{j-1})$ is the sphere orthogonal to $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$, and $\mathbf{v}^{(j)}$ is the point on this sphere closest to $\mathbf{v}_j$. Then $\nu_{mn}(C(\eta, V)) = \eta^n$. Given η, V, set

$$h_\eta(V, U) = \begin{cases} 1/2 \text{ if } U \in C(\eta, V) \text{ or } U \in C(\eta, -V), \\ 0 \text{ otherwise.} \end{cases}$$

Clearly, $h_\eta(V, U) = h_\eta(WV, WU)$ when $W \in O_m$. We have

$$\int_{\mathscr{O}} h_\eta(V, U)d\nu_{mn}(V) = \int_{\mathscr{O}} h_\eta(V, U)d\nu_{mn}(U) = \eta^n. \tag{18.2}$$

Define $H_{\eta V}(\Lambda)$ analogously to $F(\Lambda)$, such that, e.g., $H_{\eta V}(\Lambda) = h_\eta(V, U)$ when Λ has a basis KZ with $Z \in \mathscr{D}, g(Z) = 2$, and $K = tU$. We will show that

$$\sum_{\Lambda \in [\mathscr{D},T]} H_{\eta V}(\Lambda) \sim c_1(m,n)\eta^n\mu(\mathscr{D})T^m, \tag{18.3}$$

uniformly in V.

Let us derive (18.1) from (18.3). When f is continuous in $\mathscr{O}$, it is uniformly continuous. Given $\varepsilon > 0$, pick $\delta > 0$ such that $|f(V) - f(U)| < \varepsilon$ when $|V - U| < \delta$. Pick $\eta > 0$ so small that $U \in C(\eta, V)$ implies $|V - U| < \delta$, whence

$|f(V)-f(U)|<\varepsilon$. Set

$$f^0(U)=\eta^{-n}\int_{\mathscr{O}} f(V)h_\eta(V,U)d\nu_{mn}(V).$$

Here $h_\eta(V,U)=0$ unless $|V-U|<\delta$ or $|V+U|<\delta$, so that $h_\eta(V,U)=0$ unless $|f(V)-f(U)|<\varepsilon$, on recalling that $f(U)=f(-U)$. Therefore by (18.2)

$$\begin{aligned}|f^0(U)-f(U)|&=\eta^{-n}\Big|\int_{\mathscr{O}}(f(V)-f(U))h_\eta(V,U)d\nu_{mn}(V)\Big|\\&<\varepsilon\eta^{-n}\int_{\mathscr{O}}h_\eta(V,U)d\nu_{mn}(V)=\varepsilon.\end{aligned}\tag{18.4}$$

An interchange of summation and integration gives

$$\sum_{\Lambda\in[\mathscr{D},T]}F^0(\Lambda)=\eta^{-n}\int_{\mathscr{O}}f(V)\left(\sum_{\Lambda\in[\mathscr{D},T]}H_{\eta V}(\Lambda)\right)d\nu_{mn}(V).\tag{18.5}$$

The sum inside the integral may be estimated by (18.3), which holds uniformly in V. Therefore (18.5) is asymptotically equal to the right hand side of (18.1). In view of (18.4), which yields $|F^0(\Lambda)-F(\Lambda)|<\varepsilon$, and since our argument can be carried out for any $\varepsilon>0$, (18.1) follows.

To prove (18.3), let η, V be fixed. Given $\mathscr{D}\subset\mathscr{H}$, let $\mathscr{D}^0$ (as an analog to $\mathscr{D}^*$) consist of $X=(\mathbf{x}_1,\dots,\mathbf{x}_n)=KZ$ with $Z\in\mathscr{D}$ and $K=tU$ with $t>0, U\in\in C(\eta,V)$. Further $\mathscr{D}^0(\mathbf{x}_1,\dots,\mathbf{x}_t)$ is to consist of $(\mathbf{x}_{t+1},\dots,\mathbf{x}_n)$ such that $(\mathbf{x}_1,\dots,\mathbf{x}_t,\mathbf{x}_{t+1},\dots,\mathbf{x}_n)\in\mathscr{D}^0$, and $\mathscr{D}^0(T)$ to consist of $X\in\mathscr{D}^0$ with $\det X\leqslant T$. Then

$$\begin{aligned}V(\mathscr{D}^0(\mathbf{x}_1,\dots,\mathbf{x}_t))&=\eta^{n-t}V(\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_t)),\\V(\mathscr{D}^0(T))&=\eta^nV(\mathscr{D}^*(T)).\end{aligned}\tag{18.6}$$

Suppose $\mathscr{D}$ is as in Proposition B. Let $\mathbf{x}_1\in\mathbb{Z}^m\backslash\{\mathbf{0}\}$, and suppose that

$$|\mathbf{x}_1|^n y_2\cdots y_n\leqslant T\tag{18.7}$$

for every $Z\in\mathscr{D}$. Then for $1\leqslant t<n, \delta=m/|\mathbf{x}_1|$,

$$|N(\mathscr{D}^0(\mathbf{x}_1,\dots,\mathbf{x}_t))-V(\mathscr{D}^0(\mathbf{x}_1,\dots,\mathbf{x}_t))|\ll E_t+(T/|\mathbf{x}_1|\cdots|\mathbf{x}_t|)^{m-1/(n-t)},\tag{18.8}$$

where E_t is the right hand side of (15.1): When $t=n-1$, this follows from the fact that $\mathscr{D}^0(\mathbf{x}_1,\dots,\mathbf{x}_{n-1})$ as compared to $\mathscr{D}^*(\mathbf{x}_1,\dots,\mathbf{x}_{n-1})$ has an extra piece on its boundary (part of a certain cone), and since $|\mathbf{x}_1|\cdots|\mathbf{x}_{n-1}||\mathbf{x}_n|\ll\ll|\mathbf{x}_1|^n y_2\cdots y_n\leqslant T$, this part is contained in the ball $|\mathbf{x}_n|\ll(T/|\mathbf{x}_1|\cdots|\mathbf{x}_{n-1}|)$, hence contributes $\ll(T/|\mathbf{x}_1|\cdots|\mathbf{x}_{n-1}|)^{m-1}$ to the error. In the step from t to $t-1$ we note that $|\mathbf{x}_1|\cdots|\mathbf{x}_{t-1}||\mathbf{x}_t|^{n-t+1}\ll|\mathbf{x}_1|\cdots|\mathbf{x}_n|\leqslant T$, so that by Lemma 23

$$\begin{aligned}\sum_{\mathbf{x}_t}(T/|\mathbf{x}_1|\cdots|\mathbf{x}_{t-1}||\mathbf{x}_t|)^{m-1/(n-t)}&\ll(T/|\mathbf{x}_1|\cdots|\mathbf{x}_{t-1}|)^{m-1/(n-t)+1/(n-t)(n-t+1)}\\&=(T/|\mathbf{x}_1|\cdots|\mathbf{x}_{t-1}|)^{m-1/(n-t+1)}.\end{aligned}$$

The case $t = 1$ of (18.8) yields the analogue of Proposition B.

Following (17.1) we set $K = K(\mathbf{x}_1) = T/|\mathbf{x}_1|^n$, and then

$$|N(\mathscr{D}^{(K)0}(\mathbf{x}_1)) - V(\mathscr{D}^{(K)0}(\mathbf{x}_1))| \ll E' + (T/|\mathbf{x}_1|)^{m-1/(n-1)},$$

where E' is the right hand side of (16.12). The sum of $(T/|\mathbf{x}_1|)^{m-1/(n-1)}$ over $\mathbf{x}_1$ with $|\mathbf{x}_1|^n \ll T$ gives $\ll T^{m-1/n}$. Finally we get

$$|N(\mathscr{D}^0(T)) - V(\mathscr{D}^0(T))| \ll E + T^{m-1/n}, \tag{18.9}$$

where E is the right hand side of (17.7). This is the analogue of Proposition D. For every lattice Λ with basis $KZ, Z \in \mathscr{D}$, except for lattices where $g(Z) > 2$, we have two possible $K = tU \in \mathscr{K}$, hence two possible $U, -U$ in $\mathscr{O}$, so that $H_{\eta V}(\Lambda) = 1/2$ precisely when $U \in C(\eta, V)$ or $\in C(\eta, -V)$. Therefore the left hand side of (18.3) differs from $N(\mathscr{D}^0(T))$ by $\ll T^{m-1/n}$, and (18.3) follows from (4.7), (12.3), (18.6), (18.9).

We now turn to Theorem 3. An n-dimensional subspace $S \subset \mathbb{R}^m$ may be written as $S = U\mathbb{R}^n$ with $U \in \mathscr{O}$, but U will not be unique. The invariant (under O_m) measure $\nu = \nu_n^m$ on the Grassmann variety $\mathscr{G}_n^m$ may be defined by

$$\int_{\mathscr{G}_n^m} h(S)d\nu(S) = \int_{\mathscr{O}} h(U\mathbb{R}^n)d\mathbf{v}_{mn}(U)$$

for every continuous function h on $\mathscr{G}_n^m$. When h is such a function, define $f(U) = h(U\mathbb{R}^n)$, so that

$$\int_{\mathscr{O}} f(U)d\mathbf{v}_{mn}(U) = \int_{\mathscr{G}_n^m} h(S)d\nu(S).$$

A lattice with basis $KZ = tUZ$ spans $S(\Lambda) = U\mathbb{R}^n$, so that $F(\Lambda) = f(U) = h(S(\Lambda))$ unless $g(Z) > 2$. From (18.1) we deduce that

$$\sum_{\Lambda \in [\mathscr{D},T]} h(S(\Lambda)) \sim c_1(m,n)\left(\int_{\mathscr{G}_n^m} h(S)d\nu(S)\right)\mu(\mathscr{D})T^m.$$

Since the indicator function of a Jordan-measurable set on $\mathscr{G}_n^m$ can be suitably approximated by a continuous function h, Theorem 3 follows.

References

[1] CASSELS JWS (1959) An Introduction to the Geometry of Numbers. Berlin Heidelberg New York: Springer

[2] GRENIER D (1988) Fundamental domains for the general linear group. Pacific J Math **132**: 293–317

[3] HARDY GH, WRIGHT EM (1954) An Introduction to the Theory of Numbers, 3rd edn. Oxford: Clarendon Press

[4] LANG S (1994) Algebraic Number Theory, 2nd edn. Berlin Heidelberg New York: Springer

[5] MAASS H (1959) Über die Verteilung der zweidimensionalen Untergitter in einem euklidischen Gitter. Math Ann **137**: 319–327

[6] ROELCKE W (1956) Über die Verteilung der Klassen eigentlich assoziierter zweireihiger Matrizen, die sich durch eine positiv-definite Matrix darstellen lassen. Math Ann **131**: 260–277

[7] SCHMIDT WM (1968) Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height. Duke Math J **35**: 327–340

[8] SELBERG A (1956) Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. J Indian Math Soc **20**: 46–87
[9] SIEGEL CL (1988) In: CHANDRASEKHARAN K (ed) Lectures on the Geometry of Numbers. Berlin Heidelberg New York: Springer
[10] TERRAS A (1988) Harmonic Analysis on Symmetric Spaces and Applications II. Berlin Heidelberg New York: Springer

WOLFGANG M. SCHMIDT
Department of Mathematics
University of Colorado
Boulder, CO 80309-0395
USA

Buchbesprechungen – Book Reviews

Bosma, W., van der Poorten, A. (Eds.): *Computational Algebra and Number Theory* (Mathematics and Its Applications, Vol. 325). XIV, 321 pp. Kluwer, Dordrecht Boston London, 1995. Cloth £ 99,–.

This book consists of a number of articles that have been solicited on the occasion of a meeting on Computational Algebra and Number Theory, held at Sydney University in 1992. Areas represented are algorithms in groups and their generalizations, computations in number fields, algebra of polynomials and series, combinatorics and graph theory. The articles are of interest to students and practitioners working in Computational Algebra or Number Theory, but include sufficient introductory material to be accessible to nonexpert readers as well.

J. GRASSBERGER, Wien

Vieweg, E.: *Quasi-Projective Moduli for Polarized Manifolds* (Ergebnisse der Mathematik und ihrer Grenzgebiete, Bd. 30). VIII, 320 pp. Springer, Berlin Heidelberg New York, 1995. Cloth DM 164,–; öS 1.153,40.

This is a modern and comprehensive monograph on moduli schemes for a wide class of non-singular varieties. The starting point was Riemann's result that the isomorphism class of a Riemann surfaces of genus g depends on $3g - 3$ parameters. Following the lead of Mumford each moduli space is also furnished with an ample vector bundle. The main theme are construction methods for quotients of schemes by group actions, and for the ample vector bundle the study of base change and positivity properties for direct images of certain sheaves.

P. MICHOR, Wien

Nomitzu, K., Sasaki, T.: *Affine Differential Geometry* (Cambridge Tracts in Mathematics). XIV, 263 pp. Cambridge University Press, Cambridge, 1994. Cloth £ 35,–.

This is a self-contained account on affine differential geometry from a contemporary viewpoint. It covers on the one hand the classical theory, which has been started by Tzitzécia in 1908 and culminated in Blaschke, W.: Vorlesungen über Differentialgeometrie II, Affine Differentialgeometrie, Springer, Berlin 1923, and on the other hand, also developments that have happened in the last decade are included. There are 10 pages of references, a good index, and a list of symbols.

A. KRIEGL, Wien

Kassel, C.: *Quantum Groups* (Graduate Texts in Mathematics, Vol. 155). 88 Figs., XII, 531 pp. Springer, Berlin Heidelberg New York, 1995. Cloth DM 88,–; öS 616,20.

This book offers a clear and easily readable introduction to the theory of quantum groups as well as to applications of this theory to low dimensional topology. To follow the presentation, only a basic knowledge of linear algebra and topology is necessary. The book splits into four parts: The first part discusses the algebraic background of quantum group theory and gives a detailed description of the quantum SL_2-group and the corresponding deformed enveloping algebra. Part two deals with solution of the Yang–Baxter equations and the quantum double. The third part is devoted to applications to knot theory. In the final

part, the author describes Drinfeld's work on the monodromy of the Knizhnik–Zamolodchikov equations and Kontsevich's universal knot invariant.

A. Cap, Wien

Joseph, A.: *Quantum Groups and Their Primitive Ideals* (Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Bd. 29). 2 Figs, X, 383 pp. Springer, Berlin Heidelberg New York, 1995. Cloth DM 220,–; öS 1544.40.

This is a monograph on the theory of quantum groups, which is rather moderate in its prerequisites. Background information on Hopf algebras and algebraic groups is provided in two introductory chapters. Then the author constructs quantum groups starting from Cartan matrices, hereby also obtaining many results on quantum analogs of Kac–Moody algebras, and discusses the Rosso form and highest weight modules. The next two chapters are devoted to special bases, the crystal and the global bases. The last four chapters are devoted to structure theorems and the primitive spectrum of quantum enveloping algebras and structure theorems and the prime spectrum of quantum versions of the algebra of functions on a group. In an appendix, the author treats related subjects, in particular the most important motivations and techniques for the construction of quantum groups.

A. Cap, Wien

Alperin, J. L., Bell, R. B.: *Groups and Representations* (Graduate Texts in Mathematics, Vol. 162). X, 194 pp. Springer, Berlin, Heidelberg New York, 1995. DM 40,–; öS 277,–.

This is an introduction to the theory of discrete groups and their representations. Assuming a good knowledge of linear algebra and the very basic facts about finite groups (which are actually recalled in an introductory chapter), the authors first discuss in detail general linear groups over finite fields. Keeping these in mind as central examples, the authors then develop basic group theory and basic representation theory, including the necessary facts from the theory of algebras, up to character theory. The book contains more than 200 exercises with hints for solutions. These are partly real exercises, partly rather problems which go beyond the theory developed in the text.

A. Cap, Wien

Kirillov, A. A. (Ed): *Representation Theory and Noncommutative Harmonic Analysis II.* Homogeneous Spaces, Representations and Special Functions (Encycopaedia of Mathematical Sciences, Vol. 59). 2 Figs., 266 pp. Springer, Berlin Heidelberg New York, 1995. Cloth DM 165,–; öS 1154,40.

This is the second volume of a series of books on representation theory and noncommutative harmonic analysis. It falls into two parts (of almost equal size): The first part, due to V. F. Molchanov, deals with harmonic analysis on homogeneous spaces, mainly concentrating on the case of semisimple symmetric spaces. The second part (due to A. U. Klimyk and N. Ya. Vilenkin) deals with relations between representation theory and the theory of special functions. In both parts proofs are usually omitted, but there are many references to the available literature. They offer a nice possibility to get a view of available results and an introduction to the field in a quite condensed and quick form.

A. Cap, Wien

Kowalenko, V., et al.: *Generalised Euler-Jacobi Inversion Formula and Asymptotics Beyond All Orders* (London Mathematical Society Lecture Note Series, Vol. 214). X, 129 pp. Cambridge University Press, Cambridge, 1995. Softcover £ 19,95.

The book deals with asymptotic expansions of sums of the form $\sum_{n=1}^{\infty} e^{-an^{\alpha}} n^{r-1} (\log n)^m$, where $a \to 0, \alpha, r$ and m are given, $a > 0$, Re $r \geqslant 1, m \in \mathbb{N}$ and α is a positive rational number. The asymptotics are carried through very accurately, in particular for some special

values of α and $r = 1, m = 0$. The book is of considerable value for the number theorist and for the analyst as well.

J. SCHOISSENGEIER, Wien

Triebel, H.: *Interpolation Theory, Function Spaces, Differential Operators.* 532 pp. Johann Ambrosius Barth, Heidelberg Leipzig, 1995. Cloth DM 197,–; öS 1389,–.

This is a monograph on interpolation theory for Banach spaces and applications of this theory to problems related to elliptic differential operators. The first part of the book treats in detail general interpolation theory. In the second part, the author applies this theory to the study of various types of Lebesgue–Besov spaces (with or without weights, for spaces, half spaces, and domains), and thus in particular Sobolev spaces. In the last part, the author uses these spaces to study regular and strongly degenerate elliptic differential operators, as well as Legendre and Tricomi differential operators. It should be remarked that this book is a minor modification of a Russian version which appeared in 1980, some recent new developments in the theory are discussed in a short appendix.

A. CAP, Wien

Klingenberg, W. P. A.: *Riemannian Geometry* (de Gruyter Studies in Mathematics, Vol. 1). X, 410 pp. Walter de Gruyter, Berlin New York, 1995. Cloth DM 158,–.

This is the second edition of a monograph on Riemannian geometry. It consists of three parts. Part one contains foundational material, which is developed for manifolds modelled on separable Hilbert spaces. The second part deals with complete manifolds and various Hilbert manifolds of curves in a Riemannian manifold. Finally, part three deals with the geodesic flow on the cotangent bundle of a Riemannian manifold. This last part also contains the only major addition to the first edition, a proof that any surface of genus zero has infinitely many geometrically distinct closed geodesics.

A. CAP, Wien

Lang, S.: *Differential and Riemannian Manifolds.* 364 pp. Springer, New York Berlin Heidelberg, 1995. Cloth DM 92,–; öS 655,20.

This is the third edition of the book *Differential Manifolds*. It contains an introduction to the basic concepts of differential geometry (based on Banach manifolds), differential topology, and differential equations (from the global analytic point of view). It is a self contained book covering modern differential geometry. Where a subject is not covered widely, many references are provided. – The changes to the previous version include new sections on Riemannian and pseudo-Riemannian geometry (up to Hopf–Rinow and Hadamard–Cartan theorems), a completely rewriting of the sections on sprays, and many more references.

H. SCHICHL, Wien

Duistermaat, J. J.: *The Heat Kernel Lefschetz Fixed Point Formula for the Spin-c Dirac Operator.* (Progress in Nonlinear Differential Equations and Their Applications, Vol. 18). 247 pp. Birkhäuser, Basel Berlin Boston, 1996. DM 70,–; öS 496,–.

In this book manifolds with an almost complex structure are considered, carrying also a corresponding spin-c Diarc operator, which allows even in the non-Kählerian case to get local formulas for the Riemann–Roch number or the holomorphic Lefschetz number. The author uses the heat kernels theory of Berline, Getzler and Vargne to establish fundamental concepts of the theory of symplectic manifolds and gives applications to symplectic geometry.

F. HASLINGER, Wien

Bunke, U., Olbrich, M.: *Selberg Zeta and Theta Functions.* A Differential Operator Approach (Mathematical Research Volume 83). 5 Figs., 168 pp. Akademie Verlag, Berlin, 1995. DM 65,–; öS 468,–.

This is a research report devoted to the spectral geometry of locally homogeneous vector bundles over locally symmetric spaces of rank one. The main subjects of the book are the Selberg zeta function and the theta function associated to such a bundle, and the relations between these functions and their singularities and the spectrum of geometric differential operators. Moreover, the authors discuss the Ruelle zeta function. Several examples and two alternative descriptions of the singularities of the Selberg zeta function are given.

A. Cap, Wien

Devaney, R. L. (Ed.): *Complex Dynamical Systems.* (Proceedings of Symposia in Applied Mathematics, Vol. 49). IX, 209 pp. American Mathematical Society, Providence, Rhode Island, 1994. Cloth US $ 36,–.

This collection of lectures has arisen from a course on "Complex Dynamical Systems: The Mathematics Behind the Mandelbrot and Julia Sets" held at the annual meeting of the AMS in Cincinnati, Ohio, in 1994. It contains interesting survey articles on the classical work of Julia and Fatou as well as on more recent work on the dynamics of quadratic and cubic polynomials, on Yoccoz puzzles and tableaux, on the spider algorithm and on the dynamics of entire transcendental functions. Much of the book is accessible to anyone with a background on complex analysis. Several impressive color plates are included.

F. Haslinger, Wien

van Lieshout, M. N. M.: *Stochastic Geometry Models in Image Analysis and Spatial Statistics.* 171 pp. Centrum voor Wiskunde, Amsterdam, 1995. Hfl 40,–.

In this revision of his thesis from 1994 van Lieshout presents a survey of continuous Markov and Gibbs processes basing on the ideas of Geman & Geman and Besag. These processes provide a basic collection of models for image segmentation, object recognition and the classification of images. Therefore, iterative algorithms and an easy example is presented in each section. ML estimation and MAP estimators are compared and their relation to the Hough transform is described. The survey starts with Markov spatial processes (object and area interaction processes) and continues with a Bayesian approach to object recognition, i.e., Besag's ICM, fixed temperature sampling (spatial birth-and-death processes) and Geman & Gemans stochastic annealing. The second half concentrates on spatial clustering presenting absolute continuous cluster processes, general Poisson processes, ML and MAP estimates, and on Markov properties of clustering processes. The link between Markov processes and cluster models is shown, furthermore, the relation of Poisson cluster processes to Markov point processes. The booklet is well written and easy readable. It concentrates on the presentation of different stochastic processes without going into detail. Although iterative algorithms are presented some more informative examples would have shown the applicability of these models in statistical pattern recognition.

C. Cenker, Wien

Borkar, V. S.: *Probability Theory.* An Advanced Course (Universitext). XIV, 138 pp. Springer, Berlin Heidelberg New York, 1995. DM 50,–; öS 350,–.

Eine minimalistische Darstellung der maßtheoretischen Grundlagen der Wahrscheinlichkeitstheorie, die beim Leser recht weitgehende Kenntnisse der Analysis und der Wahrscheinlichkeitstheorie voraussetzt und in äußerst geraffter Form die wichtigsten Resultate über Räume von Wahrscheinlichkeitsmaßen, Martingale und Grenzwertsätze präsentiert. Durch dieses „streamlining" gelingt es, auch neue Resultate einzubeziehen, doch liefert

das Buch in erster Linie einen Rahmen für eine eingehendere Beschäftigung mit Wahrschein-lichkeitstheorie und stochastischen Prozessen.

K. SIGMUND, Wien

Bandt, C., Graf, S. Zähle, M. (Ed.): *Fractal Geometry and Stochastics* (Progress in Probability, Vol. 37). XI, 245 pp. Birkhäuser, Basel Berlin Boston, 1995. Cloth DM 120,–; öS 861,40.

This collection of invited papers has arisen from a conference held at Finsterbergen, Germany, in 1994. It contains interesting contributions to fractal sets and measures, iterated function systems, random fractals, fractals and dynamical systems and harmonic analysis on fractals. It will be useful to researchers and graduate students.

F. HASLINGER, Wien

Roache, P. J.: *Elliptic Marching Methods and Domain Decomposition.* 190 pp. CRC Press, Boca Raton New York London, 1995. Cloth US $ 69,95.

Marching methods for elliptic problems may be considered as relatives to the (better known) shooting methods for boundary value problems with ODEs. Due in part of their basic inherent instability (stability only on small domains) they lend themselves to be used in conjunction with domain decomposition. That much for the explanation of the title. The book itself makes an interesting reading on its topic and may well stimulate further thoughts on the development of those relatively little known methods and their relative merits with respect to, say, multigrid.

H. MUTHSAM, Wien

Bolognesi, T., van de Lagemaat, J., Vissers, C. (Eds.): *LOTOSphere: Software Development with LOTOS.* 488 pp. Kluwer, Boston Dordrecht London, 1995. Cloth US $ 115,–.

This book contains a collection of articles, which evolved from the LOTOSphere project. This project aimed towards establishing the usage of formal specification techniques, especially using the formal specification language LOTOS, in software engineering. The book is divided into six parts. The first contains an introduction and overview to the objectives of the LOTOSphere project. The second part introduces a discipline for generating LOTOS specifications for communication systems of the OSI family and techniques for transforming generic specifications. Part three describes analysis tools: SMILE, LITE, LOLA, and TOPO for analysing and testing LOTOS specifications. The next section talks about implementations generated from LOTOS specifications, and the following two parts introduce the graphical enhancement G-LOTOS and enhancements aiming towards easier to use abstract data types. For readers used to LOTOS this book provides a view of the state of the art in the field. Since, unfortunately, it lacks an introduction to LOTOS itself it cannot be used for learning software development with LOTOS.

H. SCHICHL, Wien

Purkert, W.: *Brückenkurs Mathematik für Wirtschaftswissenschaftler.* 436 S. Teubner, Stuttgart Leipzig, 1995. Brosch. DM 47,–; öS 326,–.

Dieser Brückenkurs „hat die Aufgabe, den Übergang von der Schule an die Hochschule" – was die Mathematikkenntnisse betrifft – zu erleichtern. Zu diesem Zweck wird (in Auswahl) sowohl der Schulstoff (Rechnen, Potenzen und Wurzeln, Logarithmen, Folgen und Reihen) als auch der Basisstoff der Einführungsvorlesungen (Differential- und Integralrechung, Lineare Algebra) in kompakter Form – aber natürlich möglichst leicht faßlich – zusammengefaßt. Dazu werden – unter Verzicht auf Formalismen (wie z. B. die Sprechweise der Mengenlehre) – jeweils die Begriffe, Aussagen und Methoden

anschaulich, aber – was dem Autor zu verdanken ist – fachlich korrekt erklärt einige Beispiele vorgerechnet und (wie in Schulbüchern) weitere analoge Aufgaben zum Trainieren bereitgestellt. Das Buch wendet sich zwar in erster Linie an Studenten der Wirtschaftswissenschaften – die Beispiele sind daher oft als Aufgaben der Wirtschaftsmathematik formuliert (Zinsenberechnung, Geldwirtschaft, Produktionsdaten etc.) –, aber die Erfahrung zeigt, daß auch viele Studenten der Naturwissenschaften (ja sogar der Mathematik) eine derartige Überbrückungschilfe gut brauchen könnten, auch wenn darin manche Themen (z. B. die Winkelfunktionen) fehlen.

P. SCHMITT, Wien

Chambers, L. (Ed.): *Genetic Algorithms.* Practical Handbook of Applications (Vol. I). 555 pp. CRC Press, Boca Raton New York London, 1995. Cloth US $ 69,95.

This volume contains 13 articles dedicated to various aspects of genetic algorithms and their applications for solving optimization problems. The book provides the reader with both an introduction to the theory of genetic algorithms as well as with many worked-out programs (in part on the accompanying diskette) for solving concrete problems.

R. BÜRGER, Wien

Lidl, R., Niederreiter, H.: *Finite Fields.* Encyclopedia of Mathematics and its Applications. XIV, 755 pp. Cambridge University Press, Cambridge, 1997. Cloth US $ 95,–.

This is the second, unchanged edition of "Finite Fields" published first by Addison-Wesley in 1983. For a review of the first edition see Mh. Math. **100** (1985), p. 165.

H. MITSCH, Wien

Verleger: Springer-Verlag KG, Sachsenplatz 4–6, A-1201 Wien. — Herausgeber: Prof. Dr. Karl Sigmund, Institut für Mathematik der Universität Wien, Strudlhofgasse 4, A-1090 Wien. — Redaktion: Strudlhofgasse 4, A-1090 Wien. — Satz und Umbruch: Thomson Press (India) Ltd., New Delhi. — Offsetdruck: Druckerei Eugen Ketterl Ges.m.b.H., A-1180 Wien. Verlagsort: Wien. — Herstellungsort: Wien. — Printed in Austria.

Offenlegung gemäß § 25 Abs. 1 bis 3 Mediengesetz:
Unternehmensgegenstand: Verlag von wissenschaftlichen Büchern und Zeitschriften. An der Springer-Verlag KG ist beteiligt: Verlassenschaft nach Dr. Konrad F. Springer, Sachsenplatz 4–6, A-1201 Wien, als Kommanditist zu 52,38%. Geschäftsführer: Prof. Dr. Dietrich Götze. Ing. Wolfram F. Joos, Dipl.-Kfm. Claus Michaletz und Rudolf Siegle, alle Sachsenplatz 4–6, A-1201 Wien.

Mh. Math. 125, 89–109 (1998)

Monatshefte für
Mathematik

Printed in Austria

Fourier Transforms of Schwartz Functions on Chébli-Trimèche Hypergroups

By

Walter R. Bloom and **Zengfu Xu,** Murdoch

(Received 7 February 1996; in revised form 13 January 1997)

Abstract. The Fourier transform for Schwartz spaces on Chébli-Trimèche hypergroups is studied, leading to results on approximation to the identity for functions and distributions on the half-line. In particular it is shown that the heat and Poisson kernels on the half-line form approximate units in various function spaces. A characterization of the convolution of a tempered distribution and a Schwartz function is also given.

1. Introduction

A hypergroup $(K, *)$ is a locally compact space with a certain generalized convolution structure $*$ on its measure space (see [3, Chapter 1] for the definition). Let ε_x be the point measure at $x \in K$. Then the convolution $\varepsilon_x * \varepsilon_y$ of the two point measures ε_x and ε_y is a probability measure on K with compact support. Unlike for the case of groups this convolution is not necessarily a point measure. The convolution between point measures extends naturally to all bounded measures on the hypergroup. In place of natural left translation of a function f by x, available in the group case, the generalized (left) translation is introduced on a hypergroup via

$$T_x f(y) := \int_K f(z)(\varepsilon_x * \varepsilon_y)(dz).$$

For every hypergroup admitting a Haar measure m the convolution of two functions f and g is defined as

$$f * g(x) := \int_K f(y) T_x g(y) m(dy).$$

The notion of an abstract algebraic hypergroup has its origins in the studies of F. Marty and H. S. Wall in the 1930s, and harmonic analysis on hypergroups dates back to J. Delsarte's and B. M. Levitan's work during the 1930s and 1940s, but the substantial development had to wait till the 1970s when Dunkl [6], Spector [11] and Jewett [10] put hypergroups in the right setting for harmonic analysis. There have been many fruitful developments of the theory of hypergroups and their applications in analysis, probability theory and approximation theory (see [3]).

1991 Mathematics Subject Classification: 43A62, 43A15, 43A32
Key words: Fourier transform, Schwartz spaces, hypergroup

Many examples of hypergroup structures on the half-line $\mathbf{R}_+ = [0, \infty[$ arise from Sturm-Liouville boundary value problems where the solutions coincide with the characters of the hypergroup in question. Chébli-Trimèche hypergroups (see Definition 1.6 below) are a class of such "one-dimensional" hypergroups on $\mathbf{R}_+$ with the convolution structure related to the following second order differential operators:

$$L = L_{A,x} := -\frac{d^2}{dx^2} - \frac{A'(x)}{A(x)}\frac{d}{dx} \tag{1.1}$$

where the function A is continuous on $\mathbf{R}_+$, twice continuously differentiable on $\mathbf{R}_+^* =]0, \infty[$, and satisfies the following conditions (see [16], [17]):

(1.2) $A(0) = 0$ and $A(x) > 0$ for $x > 0$;

(1.3) A is increasing and unbounded;

(1.4) $\dfrac{A'(x)}{A(x)} = \dfrac{2\alpha + 1}{x} + B(x)$ on a neighbourhood of 0 where $\alpha > -\dfrac{1}{2}$ and B is an odd C^∞-function on $\mathbf{R}$;

(1.5) $\dfrac{A'(x)}{A(x)}$ is a decreasing C^∞-function on $\mathbf{R}_+^*$, and hence $\rho := \dfrac{1}{2}\lim\limits_{x \to +\infty} \dfrac{A'(x)}{A(x)} \geqslant 0$ exists. Such a function A is called a *Chébli-Trimèche function.*

Definition 1.6. A hypergroup $(\mathbf{R}_+, *)$ is called a *Chébli-Trimèche hypergroup* if there exists a Chébli-Trimèche function A such that for any real-valued function f on $\mathbf{R}_+$ that is the restriction of an even nonnegative C^∞-function on $\mathbf{R}$ the generalized translation $u(x, y) = T_x f(y)$ is the solution of the following Cauchy problem:

$$\begin{cases} (L_{A,x} - L_{A,y})u(x, y) = 0, \\ u(x, 0) = f(x),\ u_y(x, 0) = 0, x > 0. \end{cases}$$

We denote by $(\mathbf{R}_+, *(A))$ the Chébli-Trimèche hypergroup associated with A.

Remark. In particular, if the function A is of the form $A(x) := x^{2\alpha+1}$ with $\alpha > -\frac{1}{2}$ then $(\mathbf{R}_+, *(A))$ is the Bessel-Kingman hypergroup. If $A(x) := \sinh^{2\alpha+1} x \cosh^{2\beta+1} x$ with $\alpha \geqslant \beta \geqslant -\frac{1}{2}$ and $\alpha \neq -\frac{1}{2}$ then $(\mathbf{R}_+, *(A))$ is the Jacobi hypergroup.

The hypergroup $(\mathbf{R}_+, *(A))$ is commutative with neutral element 0 and the identity mapping as the involution. Haar measure m on $(\mathbf{R}_+, *(A))$ is given by $m := A\lambda_{\mathbf{R}_+}$ where $\lambda_{\mathbf{R}_+}$ is the Lebesgue measure on $\mathbf{R}_+$. For any $x, y \in \mathbf{R}_+$ the probability measure $\varepsilon_x * \varepsilon_y$ is m-absolutely continuous with

$$\operatorname{supp}(\varepsilon_x * \varepsilon_y) \subset [|x - y|, x + y], \tag{1.7}$$

and the multiplicative functions on $(\mathbf{R}_+, *(A))$ are just the solutions $\varphi_\lambda (\lambda \in \mathbf{C})$ of the differential equation

$$L\varphi_\lambda(x) = (\lambda^2 + \rho^2)\varphi_\lambda(x), \quad \varphi_\lambda(0) = 1, \quad \varphi'_\lambda(0) = 0. \tag{1.8}$$

The dual space $\mathbf{R}_+^\wedge$ can be identified with the parameter set $\mathbf{R}_+ \cup i[0, \rho]$. For $0 < p \leqslant \infty$ the Lebesgue space $L^p(\mathbf{R}_+, m)$ on $(\mathbf{R}_+, *(A))$ is defined as usual; we denote by $\|f\|_{p,A}$ the L^p-norm of $f \in L^p(\mathbf{R}_+, *(A))$.

This paper deals with the (generalized) Fourier transform (see Definition 2.1) for Schwartz functions on Chébli-Trimèche hypergroups. In Section 2 we collect some basic properties of the Fourier transform on the hypergroups. In Section 3 we establish some estimates for characters that are essential to investigating the Fourier transform on some function spaces, particularly those on Schwartz spaces, which is the subject of Section 4. Finally, as applications we give results concerning approximations to the identity both for functions in various spaces and for distributions.

Throughout the paper we shall denote by $\mathbb{1}_E$ the characteristic function of the subset E of $\mathbf{R}_+$. Also we shall use C to denote a positive constant which value may vary from line to line. Dependence of such constants upon parameters of interest will be indicated through the use of subscripts.

2. The Fourier Transform

Definition 2.1. For $f \in L^1(\mathbf{R}_+, m)$ the *(generalized) Fourier transform* of f is given by

$$\mathscr{F}f(\lambda) = \hat{f}(\lambda) := \int_{\mathbf{R}_+} f(x)\varphi_\lambda(x)A(x)dx.$$

When $A(x) = x^{n-1}$ the (generalized) Fourier transform is reduced to the classical Fourier transform for radial functions on euclidean space $\mathbf{R}^n$, and for $A(x) = (\sinh x)^r(\cosh x)^s$ where (r, s) are certain pairs of non-negative integers the transform is just the spherical transform on rank 1 non-compact Riemannian symmetric spaces (see [8]).

We now collect some basic properties of the Fourier transform.

Theorem 2.2. (Levitan-Plancherel; see [3, Theorem 2.2.13]). *There exists a unique positive and tempered measure σ on $\mathbf{R}_+^\wedge$ with support $[\rho^2, \infty[$ such that the Fourier transform induces an isometric isomorphism from $L^2(\mathbf{R}_+, m)$ onto $L^2(\mathbf{R}_+^\wedge, \sigma)$, and for any $f \in L^1(\mathbf{R}_+, m) \cap L^2(\mathbf{R}_+, m)$*

$$\int_{\mathbf{R}_+} |f(x)|^2 A(x)dx = \int_{\mathbf{R}_+^\wedge} |\hat{f}(\lambda)|^2 \sigma(d\lambda).$$

The inverse is given by

$$f(x) = \int_0^\infty \hat{f}(\lambda)\varphi_\lambda(x)\sigma(d\lambda).$$

Remark. If the function $\dfrac{A'(x)}{A(x)}$ satisfies the following additional conditions:

$$\begin{cases} \text{there exists } \delta > 0 \text{ such that for all } x \in [x_0, +\infty[\\ \dfrac{A'(x)}{A(x)} = 2\rho + e^{-\delta x}D(x), \qquad \text{if } \rho > 0, \\ \dfrac{A'(x)}{A(x)} = \dfrac{2\alpha + 1}{x} + e^{-\delta x}D(x), \quad \text{if } \rho = 0 \\ \text{where } D \text{ is a } C^\infty\text{-function bounded together with its derivatives,} \end{cases} \tag{2.3}$$

we have $\sigma(d\lambda) = |c(\lambda)|^{-2}d\lambda$, where $|c(\lambda)|^{-2}$ is continuous on $[0, +\infty[$. Using the proof of [4, Proposition 3.7] (see also [15, Theorem 3.72]) TRIMÈCHE has obtained the following estimates (see [14]): There exist positive constants k, k_1, k_2 such that

(i) If $\rho \geqslant 0$ and $\alpha > -\frac{1}{2}$ then

$$k_1|\lambda|^{2\alpha+1} \leqslant |c(\lambda)|^{-2} \leqslant k_2|\lambda|^{2\alpha+1}, \lambda \in \mathbf{C}, |\lambda| > k.$$

(ii) If $\rho > 0$ and $\alpha > -\frac{1}{2}$ then

$$k_1|\lambda|^2 \leqslant |c(\lambda)|^{-2} \leqslant k_2|\lambda|^2, \lambda \in \mathbf{C}, |\lambda| \leqslant k.$$

(iii) If $\rho = 0$ and $\alpha > 0$ then

$$k_1|\lambda|^{2\alpha+1} \leqslant |c(\lambda)|^{-2} \leqslant k_2|\lambda|^{2\alpha+1}, \lambda \in \mathbf{C}, |\lambda| \leqslant k.$$

The following result is obvious.

Theorem 2.4. (i) $\mathscr{F}(T_x f)(\lambda) = \varphi_\lambda(x)\mathscr{F}f(\lambda)$.
(ii) $\mathscr{F}(f * g)(\lambda) = \mathscr{F}f(\lambda)\mathscr{F}g(\lambda)$.
(iii) Let $f \in L^p(\mathbf{R}_+, Adx), g \in L^q(\mathbf{R}_+, Adx)$ *where* $p, q, r \in [1, \infty]$ *satisfy* $p^{-1} + q^{-1} - 1 = r^{-1}$. *Then* $f * g \in L^r(\mathbf{R}_+, Adx)$ *and* $\|f * g\|_{r,A} \leqslant \|f\|_{p,A}\|g\|_{q,A}$.
(iv) $(L^1(\mathbf{R}_+, Adx), *)$ *is a commutative Banach algebra.*

3. Estimates for Characters

We begin with some basic properties of the characters.

Lemma 3.1. ([5], [13]) (i) *For each* $\lambda \in \mathbf{C}, \varphi_\lambda$ *is an even* C^∞*-function and* $\lambda \mapsto \varphi_\lambda(x)$ *is analytic.*
(ii) *For each* $\lambda \in \mathbf{C}, \varphi_\lambda$ *has an integral representation* (*i.e. the Laplace representation*)

$$\varphi_\lambda(x) = \int_{-x}^{x} e^{(i\lambda-\rho)t}\nu_x(dt), \quad x \in \mathbf{R}_+ \tag{3.2}$$

where ν_x *is a probability measure on* $\mathbf{R}$ *supported in* $[-x, x]$. *Also, for each* $x > 0$ *there exists a non-negative even continuous function* $K(x, \cdot)$ *supported in* $[-x, x]$ *such that for all* $\lambda \in \mathbf{C}$

$$\varphi_\lambda(x) = \int_0^x K(x, t)\cos \lambda t\, dt. \tag{3.3}$$

Lemma 3.4. *Let* $\lambda = \xi + i\eta$. *Then for all* $x \in \mathbf{R}_+$ *and* $|\eta| \leqslant \rho$
(i) $|\varphi_\lambda(x)| \leqslant 1$,
(ii) $|\varphi_\lambda(x)| \leqslant \varphi_{i\eta}(x) \leqslant e^{|\eta|x}\varphi_0(x)$,
(iii) $e^{-\rho x} \leqslant \varphi_0(x) \leqslant C(1 + x)e^{-\rho x}$,
(iv) *for any* $k \in \mathbf{N}_0$ *we have*

$$\left|\frac{\partial^k}{\partial \lambda^k}\varphi_\lambda(x)\right| \leqslant x^k\varphi_{i\eta}(x).$$

Proof. Conclusion (i) can be found in [5]. By [1] we have for all $x, \lambda \in \mathbf{R}_+$

$$|\varphi_\lambda(x)| \leqslant C(1+x)e^{-\rho x}.$$

Thus (ii), (iii) and (iv) follow from (3.2). □

To establish some further estimates for φ_λ and their derivatives we assume in addition that for each $k \in \mathbf{N}$, $\left(\frac{A'(x)}{A(x)}\right)^{(k)}$ is bounded for large $x \in \mathbf{R}_+$, and that there is some $\beta > 0$ such that

$$A(x) \leqslant A(1)x^\beta e^{2\rho x} \tag{3.5}$$

for all large $x \in \mathbf{R}_+$. It should be noted that this assumption is automatically satisfied if $\frac{A'(x)}{A(x)}$ satisfies (2.3).

Lemma 3.6. *Let $\lambda = \xi + i\eta$ and $k \in \mathbf{N}_0$.*
(i) *For $x \in \mathbf{R}_+$ sufficiently large we have*

$$|\varphi_\lambda(x)| \leqslant C_A x^{(\beta/2)+1} A(x)^{-1/2} e^{|\eta|x}.$$

(ii) *If $\eta = 0$ then there exists $m_k \in \mathbf{N}_0$ such that for all $x \in \mathbf{R}_+$*

$$\left|\frac{d^k}{dx^k}\varphi_\lambda(x)\right| \leqslant C_{A,k}(1+x)^2|\lambda^2+\rho^2|^{m_k}\varphi_0(x).$$

Proof. Conclusion (i) follows immediately from Lemma 3.4 and (3.5). To prove (ii) we first fix $k \in \mathbf{N}_0$, and in view of (1.1) and (1.8) observe that

$$\varphi'_\lambda(x) = -\frac{\lambda^2+\rho^2}{A(x)}\int_0^x \varphi_\lambda(t)A(t)dt. \tag{3.7}$$

Thus using Lemma 3.4

$$\begin{aligned} |\varphi'_\lambda(x)| &\leqslant \frac{|\lambda^2+\rho^2|}{A(x)}\int_0^x (1+t)e^{|\eta|t}e^{-\rho t}A(t)dt \\ &\leqslant C|\lambda^2+\rho^2|(x+1)^2 e^{|\eta|x}\varphi_0(x) \end{aligned} \tag{3.8}$$

and

$$\frac{d^k}{dx^k}\varphi_\lambda(x) + \frac{d^{k-2}}{dx^{k-2}}\left(\frac{A'(x)}{A(x)}\varphi'_\lambda(x)\right) + (\lambda^2+\rho^2)\frac{d^{k-2}}{dx^{k-2}}\varphi_\lambda(x) = 0. \tag{3.9}$$

By the assumption on A there are constants $C_{A,k} > 0$ and $R_{A,k} > 0$ such that

$$\left|\frac{d^i}{dx^i}\left(\frac{A'(x)}{A(x)}\right)\right| \leqslant C_{A,k}, \quad x \geqslant R_{A,k}, \quad i = 0, 1, \ldots, k. \tag{3.10}$$

Thus we obtain (ii) for $x > R_{A,k}$ by (3.8) – (3.10) and Lemma 3.4(ii) using induction. For $x \leqslant R_{A,k}$ conclusion (ii) follows readily from the fact that there are

$C_k > 0$ and $n_k \in \mathbf{N}_0$ such that (see [1])

$$\left|\frac{d^k}{dx^k}\varphi_\lambda(x)\right| \leqslant C_k|\lambda|^{n_k}\varphi_0(x), \quad x \in [0, R_{A,k}]$$

together with Lemma 3.4(ii) again. □

Lemma 3.11. *Let* $\varepsilon = \dfrac{2}{p} - 1$ *with* $0 < p \leqslant 2$. *Then for any* $k \in \mathbf{N}_0$ *there exists a positive constant* $C_{A,k}$ *such that for any* $x > 0$ *and* $\lambda = \xi + i\eta$ *with* $|\eta| \leqslant \varepsilon\rho$

$$\left|\frac{\partial}{\partial x}\frac{\partial^k}{\partial\lambda^k}\varphi_\lambda(x)\right| \leqslant C_{A,k}(1+|\lambda|)^2 x^k(1+x)(1+e^{2(1/p-1)\rho x})$$

and

$$\left|\frac{\partial^2}{\partial x^2}\frac{\partial^k}{\partial\lambda^k}\varphi_\lambda(x)\right| \leqslant \begin{cases} C_{A,k}(1+|\lambda|)^2 x^{k-1}(1+x)^2(1+e^{2(1/p-1)\rho x}), & k \geqslant 1, \\ C_{A,0}(1+|\lambda|)^2(1+x)^2(1+e^{2(1/p-1)\rho x}), & k = 0. \end{cases}$$

Proof. In view of [7, (2.11) and Corollary 2.1] we observe that

$$\frac{\partial^i}{\partial x^i}\frac{\partial^k}{\partial\lambda^k}\varphi_\lambda(x) = \frac{\partial^k}{\partial\lambda^k}\frac{\partial^i}{\partial x^i}\varphi_\lambda(x), \quad i = 1, 2. \tag{3.12}$$

Using (3.2) and Lemma 3.4 we obtain for $k \in \mathbf{N}_0$ and $\lambda = \xi + i\eta$ with $|\eta| \leqslant \varepsilon\rho$

$$\left|\frac{\partial^k}{\partial\lambda^k}\varphi_\lambda(x)\right| \leqslant x^k e^{\varepsilon\rho x}\varphi_0(x) \leqslant Cx^k(1+x)e^{2(1/p-1)\rho x}. \tag{3.13}$$

By (1.5) we observe that $0 \leqslant \dfrac{A'(x)}{A(x)} \leqslant C\left(1 + \dfrac{1}{x}\right)$. Thus the lemma follows readily from (3.12), (1.8), (3.13), (3.7) and (1.3). □

4. Fourier Transform on Schwartz Spaces

In this section we investigate the Fourier transform of functions in Schwartz spaces. Analogous to the case of Riemannian symmetric spaces (see [2]) we show that the Fourier transform is an isomorphism between the generalized Schwartz spaces and the extended Schwartz spaces. We first introduce some function classes on $\mathbf{R}_+$.

Definition 4.1. For each $a > 0$ put

$$\mathscr{D}_a(\mathbf{R}) := \{g \in C^\infty(\mathbf{R}) : g \text{ is even and supported in } [-a, a]\},$$

and define

$$\mathscr{D}_a(\mathbf{R}_+) := \{f : f = g \text{ on } \mathbf{R}_+ \text{ for some } g \in \mathscr{D}_a(\mathbf{R})\}.$$

We also write $\mathscr{D}_*(\mathbf{R}_+) = \bigcap_{a>0}\mathscr{D}_a(\mathbf{R}_+)$.

Definition 4.2. For $a > 0$ we denote by H_a the set of functions h that are even and entire on $\mathbf{C}$ and satisfy $\sup_{\lambda\in\mathbf{C}}(1+|\lambda|^2)^m|h(\lambda)|e^{-a|Im(\lambda)|} < \infty, m \in \mathbf{N}_0$, and we set

$$H_* := \bigcup_{a>0} H_a.$$

Definition 4.3. For $0 < p \leqslant 2$ denote by $\mathscr{S}_p(\mathbf{R}_+, *(A))$ the *(generalized) Schwartz space* defined by

$$\mathscr{S}_p(\mathbf{R}_+, *(A)) := \{f : f = g \text{ on } \mathbf{R}_+, g \in \mathscr{S}_p(\mathbf{R})\}$$

where

$$\mathscr{S}_p(\mathbf{R}) := \{g \in C^\infty(\mathbf{R}) : g \text{ is even and } \mu^p_{k,l}(g) < \infty, k, l \in \mathbf{N}_0\}$$

with

$$\mu^p_{k,l}(g) := \sup_{x \in \mathbf{R}_+} (1+x)^l \varphi_0(x)^{-2/p} |g^{(k)}(x)|. \tag{4.4}$$

$\mathscr{S}_p(\mathbf{R}_+, *(A))$ is topologized by means of the seminorms $\mu^p_{k,l}$.

If $A(x) = x^{n-1} (n \in \mathbf{N}, n \geqslant 2)$ then $\varphi_0(x) = 1$ and $\mathscr{S}_p(\mathbf{R}_+, *(A))$ is just the set of all radial Schwartz functions on n-dimensional Euclidean space R^n. If $A(x) = (\sinh x)^r (\cosh x)^s$ with $r \geqslant s$ and $r, s \in \mathbf{N}$ then

$$\varphi_0(x) = {}_2F_1\left(\frac{r+s}{4}, \frac{r+s}{4}, \frac{1}{2}(r+1), -\sinh^2 x\right)$$

where ${}_2F_1$ is the Gaussian hypergeometric function, and $\mathscr{S}_p(\mathbf{R}_+, *(A))$ is the spherical Schwartz space on rank 1 noncompact symmetric space for certain pairs (r, s).

Definition 4.5. Let $\varepsilon = \dfrac{2}{p} - 1$ for $0 < p \leqslant 2$ and denote by $\mathscr{S}(\mathscr{F}_\varepsilon)$ the *extended Schwartz space* defined by all functions h that are even and analytic in the interior of $\mathscr{F}_\varepsilon$, and such that h together with all its derivatives extend continuously to $\mathscr{F}_\varepsilon$ and satisfy $\tau^{(\varepsilon)}_{k,l}(h) < \infty, k, l \in \mathbf{N}_0$. Here

$$\mathscr{F}_\varepsilon := \{z \in \mathbf{C} : |\mathrm{Im}\,(z)| \leqslant \varepsilon\rho\}, \quad \tau^{(\varepsilon)}_{k,l}(h) := \sup_{\lambda \in \mathscr{F}_\varepsilon} (1 + |\lambda|)^l |h^{(k)}(\lambda)|.$$

The classical Paley-Wiener theorem can be extended to the hypergroup $(\mathbf{R}_+, *(A))$.

Theorem 4.6. (Paley-Wiener theorem, see [5], [13]). *$f \in \mathscr{D}_a(\mathbf{R}_+)$ if and only if $\hat{f} \in H_a$. The Fourier transform $\mathscr{F}$ is an isomorphism between $\mathscr{D}_*(\mathbf{R}_+)$ and H_*.*

Let $\mathscr{F}_0$ be the classical Fourier transform on $\mathbf{R}$.

Theorem 4.7. (see [13]). *$\mathscr{F}_0$ is an isomorphism between $\mathscr{D}_*(R_+)$ and H_*.*

Definition 4.8. (see [13]) The *Abel transform* of $f \in \mathscr{D}_*(\mathbf{R}_+)$ is defined by

$$\mathscr{A}f(x) = \int_x^\infty f(y) K(y, x) A(y) dy \tag{4.9}$$

where $K(y, x)$ is as in (3.3).

Lemma 4.10. *(i) The Abel transform is an isomorphism of $\mathscr{D}_*(\mathbf{R}_+)$ onto itself.* (ii) *$f \in \mathscr{D}_a(\mathbf{R}_+)$ if and only if $\mathscr{A}f \in \mathscr{D}_a(\mathbf{R}_+)$.*

Proof. Conclusion (i) is given in [13], and (ii) is obvious from the definition of $\mathscr{A}$ and the expression of the inverse Abel transform $\mathscr{A}^{-1}$ (see [13]). □

Lemma 4.11. (see [13]). *For $f \in \mathscr{D}_*(\mathbf{R}_+)$ we have $\mathscr{F}f = \mathscr{F}_0(\mathscr{A}f)$.*

Using standard methods the following result is immediate.

Lemma 4.12. *$\mathscr{D}_*(\mathbf{R}_+)$ is dense in $L^p(\mathbf{R}_+, Adx)$ for all $p \in [1, \infty[$.*

Lemma 4.13. *$\mathscr{D}_*(\mathbf{R}_+)$ is dense in $\mathscr{S}_p(\mathbf{R}_+, *(A))$ for all $p \in [1, \infty[$.*

Proof. The lemma can be proved in the same way as for noncompact symmetric spaces (see [9, p. 254]). □

The following result is immediate from Definition 4.3.

Lemma 4.14. *For $q \geqslant p$ we have $\mathscr{S}_p(\mathbf{R}_+, *(A)) \subset L^q(\mathbf{R}_+, Adx)$.*

Definition 4.15. Let $\varepsilon = \dfrac{2}{p} - 1$ with $0 < p \leqslant 2$. We denote by $\mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$ the space of the restrictions to $\mathbf{R}_+$ of the functions in $\mathscr{S}_{\varepsilon\rho}(\mathbf{R})$ where

$$\mathscr{S}_{\varepsilon\rho}(\mathbf{R}) := \{g \in C^\infty(\mathbf{R}) : g \text{ is even and } \nu_{k,l}^{(\varepsilon)}(g) < \infty\}$$

with

$$\nu_{k,l}^{(\varepsilon)}(g) := \sup_{t \in \mathbf{R}_+} (1+t)^l e^{\varepsilon\rho t} |g^{(k)}(t)|.$$

Lemma 4.16. *The classical Fourier transform $\mathscr{F}_0$ is an isomorphism between $\mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$ and $\mathscr{S}(\mathscr{F}_\varepsilon)$.*

Proof. We may (and do) assume that $\varepsilon\rho > 0$. Let $g \in \mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$. Then for any $k, l \in \mathbf{N}_0, \lambda = \xi + i\eta \in \mathscr{F}_\varepsilon$

$$\begin{aligned}
&(|\lambda| + 1)^l |\mathscr{F}_0(g)^{(k)}(\lambda)| \\
&\quad = \sum_{j=0}^{l} \binom{l}{j} \left| \mathscr{F}_0\left(\frac{d^j}{dt^j}(t^k g(t))\right)(\lambda) \right| \\
&\quad \leqslant C_{k,l} \sum_{j=0}^{l} \int_{-\infty}^{\infty} |g^{(j)}(t)| (|t| + 1)^k e^{-\eta t} dt \\
&\quad \leqslant C_{k,l} \sum_{j=0}^{l} \int_{-\infty}^{\infty} (|g^{(j)}(t)| (|t| + 1)^{k+2} e^{\varepsilon\rho t}) e^{-\varepsilon\rho t} (\cosh \eta t) (|t| + 1)^{-2} dt \\
&\quad \leqslant C_{k,l,\varepsilon,\rho} \sum_{j=0}^{l} \nu_{j,k+2}^{(\varepsilon)}(g).
\end{aligned}$$

Hence $\mathscr{F}_0$ maps $\mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$ into $\mathscr{S}(\mathscr{F}_\varepsilon)$ continuously. The classical Plancherel theorem implies that $\mathscr{F}_0 : \mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+) \to \mathscr{S}(\mathscr{F}_\varepsilon)$ is injective.

It remains to show that $\mathscr{F}_0^{-1}$ maps $\mathscr{S}(\mathscr{F}_\varepsilon)$ into $\mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$ continuously. For any $h \in \mathscr{S}(\mathscr{F}_\varepsilon)$ and $k, l \in \mathbf{N}_0$ let $h_{kl}(\lambda) = \left(1 + \dfrac{d}{d\lambda}\right)^l (\lambda^k h(\lambda))$. Then for $t \in \mathbf{R}_+$ we

have by Cauchy's integral theorem

$$\begin{aligned}(t+1)^l e^{\varepsilon\rho t}|\mathscr{F}_0^{-1}(h)^{(k)}(t)| &= C_0 e^{\varepsilon\rho t}|\mathscr{F}_0^{-1}(h_{kl})(t)| \\ &= C_0 e^{\varepsilon\rho t}\left|\int_{-\infty+i\varepsilon\rho}^{\infty+i\varepsilon\rho} h_{kl}(\lambda)e^{i\lambda t}d\lambda\right| \\ &\leqslant C_{\varepsilon,\rho}\sum_{j=0}^{l}\tau_{j,k+2}^{(\varepsilon)}(h)\end{aligned}$$

and this completes the proof of the lemma. □

Lemma 4.17. *H_* is dense in $\mathscr{S}(\mathscr{F}_\varepsilon)$.*

Proof. By Lemma 4.16 for any $h \in \mathscr{S}(\mathscr{F}_\varepsilon)$ there exists $g \in \mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$ such that $\mathscr{F}_0(g) = h$. It can be proved in the same way as in Lemma 4.13 that $\mathscr{D}_*(\mathbf{R}_+)$ is dense in $\mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$, and hence there is a sequence $(g_n) \subset \mathscr{D}_*(\mathbf{R}_+)$ such that

$$g_n \to g \quad \text{in } \mathscr{S}_{\varepsilon\rho},\ n \to \infty.$$

Once again appealing to Lemma 4.16 we have

$$\mathscr{F}_0(g_n) \to \mathscr{F}_0(g) = h \quad \text{in } \mathscr{S}(\mathscr{F}_\varepsilon),\quad n \to \infty.$$

But Theorem 4.7 shows that $\mathscr{F}_0(g_n) \in H_*$, and thus the lemma is proved. □

Lemma 4.18. *Suppose that $f \in C^\infty(\mathbf{R})$ and f is even. Then*

(i) *for any $l \in \mathbf{N}$, $L^l f(x)$ extends to an even C^∞-function on $\mathbf{R}$ where $L^0 f = f$ and $L^l f = L(L^{l-1}f)$;*

(ii) *there exists $\delta > 0$ such that for any $l \in \mathbf{N}$ there are $s_l \in \mathbf{N}_0$ and $C_{l,\alpha}(B) > 0$ satisfying the following: for each $x \in [0,\delta]$ there exist $\xi_j = \xi_j(x,l) \in [0,x]$ such that*

$$|L^l f(x)| \leqslant C_{l,\alpha}(B)\left(\sum_{i=1}^{2l}\sum_{j=0}^{s_l}|f^{(i)}(\xi_j)| + \sum_{i=1}^{2l}|f^{(i)}(x)|\right)$$

where $C_{l,\alpha}(B)$ is a constant depending only on l, α and $\max_{x\in[0,\delta]}|B^{(k)}(x)|$ for $k = 0,1,\ldots,r_l$, and B is the function given in (1.4);

(iii) *for any $l \in \mathbf{N}$ there exists a constant $C_{l,\delta}(A) > 0$ such that*

$$|L^l f(x)| \leqslant C_{l,\delta}(A)\sum_{i=1}^{2l}|f^{(i)}(x)|, \quad x \geqslant \delta.$$

Proof. By (1.4) there exists $\delta > 0$ such that

$$\frac{A'(x)}{A(x)} = \frac{2\alpha+1}{x} + B(x), \quad x \in [0,\delta]$$

where B is an odd C^∞-function on $\mathbf{R}$ and $\alpha > -\frac{1}{2}$. Hence

$$Lf(x) = -f''(x) - \frac{2\alpha+1}{x}f'(x) - B(x)f'(x), \quad x \in [0,\delta].$$

Observe that $\lim_{x\to 0} Lf(x) = -(2\alpha+2)f''(0)$. Define $Lf(0) = -(2\alpha+2)f''(0)$. Then Lf extends to an even and continuous function on $\mathbf{R}$. Put

$$\phi(x) = \frac{1}{x} f'(x), \quad x \neq 0, \quad x \in [-\delta, \delta]$$

and

$$\phi_k(x) = \sum_{j=0}^{k} (-1)^j j! \binom{k}{j} x^{k-j} f^{(k+1-j)}(x), \quad k \in \mathbf{N}_0, \ x \in [-\delta, \delta].$$

Thus

$$\phi^{(k)}(x) = x^{-k-1}\phi_k(x), \quad x \neq 0. \tag{4.19}$$

Observe that

$$\phi_k'(x) = x^k f^{(k+2)}(x) \tag{4.20}$$

and $\phi_k(0) = 0$. Hence by the mean-value theorem there exist $\zeta_k = \zeta_k(x) \in [0, x]$ if $x \geqslant 0$ and $\xi_k \in [x, 0]$ if $x < 0$ such that

$$\phi_k(x) = x\xi_k^k f^{(k+2)}(\xi_k).$$

Therefore by (4.19)

$$\phi^{(k)}(x) = \left(\frac{\xi_k}{x}\right)^k f^{(k+2)}(\xi_k). \tag{4.21}$$

By (4.20) it is easy to see that $\lim_{x\to 0} \phi^{(k)}(x) = \dfrac{f^{(k+2)}(0)}{k+1}$. Thus ϕ extends to an even C^∞-function on $\mathbf{R}$ and we have

$$\begin{aligned}(Lf)^{(k)}(x) = &-f^{(k+2)}(x) - (2\alpha+1)\phi^{(k)}(x) \\ &- \sum_{j=0}^{k} \binom{k}{j} B^{(j)}(x) f^{(k+1-j)}(x), \quad x \in [-\delta, \delta].\end{aligned}$$

Therefore Lf extends to an even C^∞-function on $\mathbf{R}$ and (i) follows by induction.

Let $\psi(x) = B(x)f'(x)$. To prove (ii) we observe that from (4.21)

$$|\phi^{(k)}(x)| \leqslant |f^{(k+2)}(\xi_k)|, \quad 0 \leqslant \xi_k \leqslant x \leqslant \delta$$

and

$$|\psi^{(k)}(x)| \leqslant C_k(B) \sum_{j=1}^{k+1} |f^{(j)}(x)|, \quad 0 \leqslant x \leqslant \delta$$

where $C_k(B)$ is a constant depending only on k and $\max_{0\leqslant x\leqslant\delta} |B^{(s)}(x)|, 0 \leqslant s \leqslant k+1$. Hence for any $k \in \mathbf{N}_0$

$$|(Lf)^{(k)}(x)| \leqslant C'_{k,\alpha}(B)\left(\sum_{j=1}^{k+2} |f^{(j)}(x)| + f^{(k+2)}(\xi_k)\right), \quad 0 \leqslant \xi_k \leqslant x \leqslant \delta. \tag{4.22}$$

This implies that (ii)) holds for $l = 1$, and the full result follows by induction. Conclusion (iii) follows immediately from (1.5), (3.10) and (1.1). □

Consider the seminorm $\tilde{\tau}_{k,l}^{(\varepsilon)}$ defined on $\mathscr{S}(\mathscr{F}_\varepsilon)$ by

$$\tilde{\tau}_{k,l}^{(\varepsilon)}(h) = \sup_{\lambda \in \mathscr{F}_\varepsilon} \left| \frac{d^k}{d\lambda^k}((\lambda^2 + \rho^2)^l h(\lambda)) \right|, \quad h \in \mathscr{S}(\mathscr{F}_\varepsilon).$$

Lemma 4.23. *The two seminorms $\tau_{k,l}^{(\varepsilon)}$ (see Definition 4.5) and $\tilde{\tau}_{k,l}^{(\varepsilon)}$ on $\mathscr{S}(\mathscr{F}_\varepsilon)$ are equivalent.*

Proof. The lemma follows from a straightforward calculation. □

With the above results we are now able to investigate the Fourier transform on Schwartz spaces.

Proposition 4.24. *The Fourier transform $\mathscr{F}$ maps $\mathscr{S}_p(\mathbf{R}_+, *(A))(0 < p \leqslant 2)$ into $\mathscr{S}(\mathscr{F}_\varepsilon)\left(\varepsilon = \frac{2}{p} - 1\right)$ continuously and is injective.*

Proof. Let $f \in \mathscr{S}_p(\mathbf{R}_+, *(A))$ and $\lambda \in \mathscr{F}_\varepsilon, \lambda = \xi + i\eta$. By Lemma 3.4(ii), (iii) and (3.5) we have

$$\int_{\mathbf{R}_+} |f(x)\varphi_\lambda(x)|A(x)dx$$

$$\leqslant C_\alpha \int_{\mathbf{R}_+} |f(x)|\varphi_0(x)^{-2/p}(1+x)^{2/p+\beta+1} e^{(\varepsilon-(2/p)+1))\rho x} dx < \infty$$

and hence $\hat{f}$ is well-defined. We recall from Lemma 3.1 that $\lambda \mapsto \varphi_\lambda(x)$ is analytic on $\mathbf{C}$. Thus $\hat{f}$ is analytic in the interior of $\mathscr{F}_\varepsilon$ and continuous on $\mathscr{F}_\varepsilon$, and by Lemma 3.4 and (3.5) all the derivatives of $\hat{f}$ extend continuously to $\mathscr{F}_\varepsilon$.

To prove that $\mathscr{F} : f \to \hat{f}$ is continuous from $\mathscr{S}_p(\mathbf{R}_+, *(A))$ into $\mathscr{S}(\mathscr{F}_\varepsilon)$ let $\varphi_{\lambda,l}(x) = L^l\varphi_\lambda(x)$. Then by (1.1) we have for $k, l \in \mathbf{N}_0$ and $f \in \mathscr{S}_p(\mathbf{R}_+, *(A))$

$$\frac{d^k}{d\lambda^k}((\lambda^2 + \rho^2)^l \hat{f}(\lambda)) = \frac{d^k}{d\lambda^k}\left(\int_0^\infty (-f(x))d(A(x)\varphi'_{\lambda,l-1}(x))\right).$$

By (3.8)

$$|\varphi'_{\lambda,l-1}(x)| \leqslant C_\alpha(|\lambda|^2 + \rho^2)^l(1+x)^2 e^{|\eta|x}\varphi_0(x)$$

for $\lambda \in \mathscr{F}_\varepsilon$ and all large $x \in \mathbf{R}_+$, and also by Lemma 3.4

$$|\varphi_{\lambda,l-1}(x)| \leqslant (|\lambda|^2 + \rho^2)^{l-1} e^{|\eta|x}\varphi_0(x).$$

Therefore

$$\frac{d^k}{d\lambda^k}((\lambda^2 + \rho^2)^l \hat{f}(\lambda)) = \frac{d^k}{d\lambda^k}\int_0^\infty Lf(x)\varphi_{\lambda,l-1}(x)A(x)\,dx.$$

Appealing to Lemma 4.18 an induction argument then gives

$$\frac{d^k}{d\lambda^k}((\lambda^2 + \rho^2)^l \hat{f}(\lambda)) = \int_0^\infty L^l f(x)\frac{d^k}{d\lambda^k}\varphi_\lambda(x))A(x)\,dx.$$

Hence using Lemmas 4.18 and 3.4 together with (1.4) we obtain for $\lambda \in \mathscr{F}_\varepsilon$

$$\left|\frac{d^k}{d\lambda^k}((\lambda^2+\rho^2)^l \hat{f}(\lambda))\right| \leqslant \int_0^\delta \left| L^l f(x) \frac{d^k}{d\lambda^k}\varphi_\lambda(x)\right| A(x)dx + \int_\delta^\infty \left| L^l f(x) \frac{d^k}{d\lambda^k}\varphi_\lambda(x)\right| A(x)dx \leqslant C^{(1)}_{l,k,p}(A) \sum_{i=1}^{2l} \mu^p_{i,0}(f) + C^{(2)}_{l,k,p,r}(A) \sum_{i=1}^{2l} \mu^p_{i,r}(f)$$

where $r \in \mathbf{N}$ with $r > k + 2 + \frac{2}{p}$. Therefore by Lemma 4.23 we have for $f \in \mathscr{S}_p(\mathbf{R}_+, *(A))$

$$\tau^{(\varepsilon)}_{k,l}(\hat{f}) \leqslant C_{l,k,p}(A)\left(\sum_{i=0}^{2l} \mu^p_{i,r}(f) + \sum_{i=0}^{2l} \mu^p_{i,0}(f)\right). \tag{4.25}$$

Thus $\mathscr{F}$ maps $\mathscr{S}_p(\mathbf{R}_+, *(A))$ into $\mathscr{S}(\mathscr{F}_\varepsilon)$ continuously. Finally Theorem 2.2 implies that $\mathscr{F}$ is injective. □

Proposition 4.26. *The inverse Fourier transform* $\mathscr{F}^{-1} : H_* \to \mathscr{D}_*(\mathbf{R}_+)$ *given by*

$$\mathscr{F}^{-1}h(x) := \int_0^\infty h(\lambda)\varphi_\lambda(x)\sigma(d\lambda)$$

is continuous under the topologies of $\mathscr{S}(\mathscr{F}_\varepsilon)$ *and* $\mathscr{S}_p(\mathbf{R}_+, *(A))$.

Proof. The proposition can be proved in a similar way to that in [2] making use of the Paley-Wiener theorem for the generalized and the classical Fourier transforms (Theorems 4.6 and 4.7), the support conservation property for the Abel transform and the relation between the Fourier transform on the hypergroup, the classical Fourier transform and the Abel transform (Lemmas 4.10, 4.11). This allows us to apply elementary Fourier analysis using the estimates for characters (Lemmas 3.6(ii) and 3.4) together with [3, Theorem 2.2.36]. □

The main result of this section is the following theorem:

Theorem 4.27. *Let* $0 < p \leqslant 2$ *and* $\varepsilon = \frac{2}{p} - 1$. *Then the Fourier transform* $\mathscr{F}$ on $(\mathbf{R}_+, *(A))$ *is an isomorphism between* $\mathscr{S}_p(\mathbf{R}_+, *(A))$ *and* $\mathscr{S}(\mathscr{F}_\varepsilon)$. *The inverse transform is given by*

$$\mathscr{F}^{-1}h(x) = \int_0^\infty h(\lambda)\varphi_\lambda(x)\sigma(d\lambda).$$

Proof. The theorem follows immediately from Lemmas 4.13 and 4.17 and Propositions 4.24 and 4.26. □

We conclude with a property of the Abel transform.

Lemma 4.28. *For* $f \in \mathscr{S}_p(\mathbf{R}_+, *(A))$ *the Abel transform* $\mathscr{A}f$ *given by* (4.9) *is well-defined.*

Proof. By [13] the kernel $K(t,x)$ in (4.9) satisfies

$$K(t,x)A(t) = t^{1/2-\alpha}A(t)^{1/2}S(t,x) + C_\alpha t^{1/2-\alpha}A(t)^{1/2}(t^2-x^2)^{\alpha-1/2}\mathbb{1}_{[0,t]}(x) \tag{4.29}$$

where

$$S(t,x) = \frac{1}{\pi}t^{\alpha-1/2}A(t)^{1/2}\int_0^\infty \psi(t,\lambda)\cos\lambda x\, d\lambda \tag{4.30}$$

and

$$\psi(t,\lambda) = \varphi_\lambda(t) - c_0 t^{\alpha+1/2}A(t)^{-1/2}j_\alpha(\lambda t) \tag{4.31}$$

where $j_\alpha = 2^\alpha\Gamma(\alpha+1)x^{-\alpha}J_\alpha(x)$, J_α is the Bessel function of order α and Γ is the Gamma function. Write

$$S(t,x) = \pi^{-1}t^{\alpha-1/2}A(t)^{1/2}\left(\int_0^1 \psi(t,\lambda)\cos\lambda x\, d\lambda + \int_1^\infty \psi(t,\lambda)\cos\lambda x\, d\lambda\right) := I_1 + I_2.$$

By (4.31), Lemma 3.4, (3.5) and properties of the Bessel function we see there exist constants $R_0 > 1$ and $C_{A,R_0} > 0$ such that

$$|I_1| \leqslant C_{A,R_0}(t^{\alpha+1/2+\beta/2} + t^{2\alpha+1}), \quad t \geqslant R_0.$$

To estimate I_2 we first observe that (3.5) implies

$$\int_1^t F(y)dy \leqslant C_A \ln t, \quad t \geqslant R_0 \tag{4.32}$$

where $F(y) = \dfrac{A'(y)}{A(y)} - 2\rho$. On the other hand, from (3.8) and [13, Theorem 3.2] we have

$$|\psi(t,\lambda)| \leqslant CA(t)^{-1/2}\lambda^{-\alpha-3/2}\tilde{\chi}(t)\exp\left(\frac{C_1}{\lambda}\tilde{\chi}(t)\right)$$

where

$$\tilde{\chi}(t) = \int_0^t |\chi(r)|dr$$

and

$$\chi(t) = \frac{1}{4}F(t)^2 + \rho F(t) + \frac{1}{2}F'(t) + \frac{\frac{1}{4}-\alpha^2}{t^2}.$$

By (1.4), χ is continuous on $\mathbf{R}_+$. Thus by (1.5) and (4.32)

$$|\tilde{\chi}(t)| \leqslant C_{A,R_0}(1 + \rho\ln t), \quad t \geqslant R_0.$$

Hence

$$|I_2| \leqslant C_{A,R_0}t^{\alpha+1/2+C_1\rho}, \quad t \geqslant R_0.$$

Therefore there exist constants C and k depending only on A, β and R_0 such that

$$K(t,x)A(t) \leqslant Ct^kA(t)^{1/2}, \quad t \geqslant R_0, \quad x \leqslant t.$$

Thus if $x \geqslant R_0$ then by Lemma 3.4

$$\int_x^\infty |f(t)|K(t,x)A(t)dt \leqslant C_1 \mu_{0,r}^p(f)$$

with $r \in \mathbf{N}$ satisfying $r > 2k+3$ and $C_1 > 0$ depending only on A, β and R_0. By [13, Proposition 4.3] there exists a constant $C(R_0) > 0$ such that

$$|S(t,x)| \leqslant C(R_0)t^{\alpha+1/2}, \quad 0 \leqslant x \leqslant t \leqslant R_0.$$

Hence by (4.29) and (3.5) we have for $x \leqslant R_0$

$$\begin{aligned}\int_x^\infty |f(t)|K(t,x)A(t)dt &\leqslant \int_x^{R_0} |f(t)|K(t,x)A(t)dt + \int_{R_0}^\infty |f(t)|K(t,x)A(t)dt \\ &\leqslant C_2 \mu_{0,r}^p(f)\end{aligned}$$

and this completes the proof of the lemma. □

Theorem 4.33. *The Abel transform $\mathscr{A}$ is an isomorphism between $\mathscr{S}_p(\mathbf{R}_+, *(A))$ and $\mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$.*

Proof. For any $f \in \mathscr{S}_p(\mathbf{R}_+, *(A))$ there exists by Lemma 4.13 a sequence $(f_n) \subset \mathscr{D}_*(\mathbf{R}_+)$ such that $|f_n| \leqslant |f|$ and $f_n \to f$ in $\mathscr{S}_p$ as $n \to \infty$. By Theorem 4.27 and Lemma 4.16, $\mathscr{F}_0^{-1} \circ \mathscr{F}$ is an isomorphism between $\mathscr{S}_p(\mathbf{R}_+, *(A))$ and $\mathscr{S}_{\varepsilon\rho}(\mathbf{R}_+)$. The theorem now follows immediately from Lemma 4.11. □

5. Some Applications

We now give some applications of Theorem 4.27 to approximations of the identity. Suppose that $\phi \in L^1(\mathbf{R}_+, Adx)$ such that $\|\phi\|_{1,A} = 1$, and set

$$\phi_t(x) = \frac{A\left(\frac{x}{t}\right)}{tA(x)}\phi\left(\frac{x}{t}\right), \quad t > 0, \ x \in \mathbf{R}_+. \tag{5.1}$$

Clearly $\|\phi_t\|_{1,A} = \|\phi\|_{1,A} = 1$. We need the following basic lemma.

Lemma 5.2. *Let $0 < p \leqslant 2$.*

(i) *If $\phi_1, \phi_2 \in \mathscr{S}_p(\mathbf{R}_+, *(A))$ then $\phi_1 * \phi_2 \in \mathscr{S}_p(\mathbf{R}_+, *(A))$.*

(ii) *If $\phi \in \mathscr{S}_p(\mathbf{R}_+, *(A))$ then for any $x \in \mathbf{R}_+, T_x\phi \in \mathscr{S}_p(\mathbf{R}_+, *(A))$ and for any $k, l \in \mathbf{N}_0$ there exist $r, s \in \mathbf{N}_0$ such that*

$$\mu_{k,l}^p(T_x\phi) \leqslant C_{k,l,r,s,}(1+x)x^r e^{2(1/p-1)\rho x} \sum_{i=0}^{2s}\sum_{j=0}^{r} \mu_{i,j}^p(\phi).$$

(iii) *If $\phi \in \mathscr{S}_p(\mathbf{R}_+, *(A))$ and $x, y \in \mathbf{R}_+, x < y$ then for any $k, l \in \mathbf{N}_0$ there exist $r, s \in \mathbf{N}_0$ such that*

$$\mu_{k,l}^p(T_x\phi - T_y\phi) \leqslant C_{k,l,r,s,p}(y-x)(1+y)y^r(1+e^{2(1/p-1)\rho y}) \sum_{i=0}^{2s+4}\sum_{j=0}^{r} \mu_{i,j}^p(\phi).$$

Proof. Clearly (i) follows from Theorems 2.4(ii) and 4.27. To prove (ii) we first observe that by Lemma 4.13 there is a sequence (g_n) in $\mathscr{D}_*$ such that for any $k, l \in \mathbf{N}, \mu_{k,l}^p(g_n - \phi) \to 0$ as $n \to \infty$. By Theorem 4.27 for any $r, s \in \mathbf{N}_0$,

$\tau_{r,s}^{(\varepsilon)}(\hat{g}_n - \hat{\phi}) \to 0$ as $n \to \infty$. Thus in view of Proposition 4.26 we deduce that for any $k, l \in \mathbf{N}_0$ there exist $r, s \in \mathbf{N}_0$ such that for any $\phi \in \mathcal{S}_p(\mathbf{R}_+, *(A))$

$$\mu_{k,l}^p(\phi) \leqslant C_{k,l,r,s} \sum_{j=0}^{r} \tau_{j,s}^{(\varepsilon)}(\hat{\phi}). \tag{5.3}$$

Appealing to (3.13) we obtain for any $x \in \mathbf{R}_+$ and $j, s \in \mathbf{N}_0$

$$\begin{aligned} \tau_{j,s}^{(\varepsilon)}(\mathcal{F}(T_x\phi)) &:= \sup_{\lambda \in \mathcal{F}_\varepsilon} (1+|\lambda|)^s \left| \frac{\partial^j}{\partial \lambda^j} (\varphi_\lambda(x)\hat{\phi}(\lambda)) \right| \\ &\leqslant C_{j,s}(1+x)x^j e^{2(1/p-1)\rho x} \sum_{q=o}^{j} \tau_{q,s}^{(\varepsilon)}(\hat{\phi}). \end{aligned} \tag{5.4}$$

Observe that by Theorem 2.4(i), Lemma 3.1(i) and Theorem 4.27 the function $\mathcal{F}(T_x\phi) = \varphi_\lambda(x)\hat{\phi}(\lambda)$ is even and analytic in the interior of $\mathcal{F}_\varepsilon$. Therefore $\mathcal{F}(T_x\phi) \in \mathcal{S}(\mathcal{F}_\varepsilon)$. Consequently (ii) follows immediately from Theorem 4.27, (5.3), (5.4) and (4.25).

It remains to show (iii). Applying the mean-value theorem together with Lemma 3.11 we have for any $n \in \mathbf{N}_0$

$$\begin{aligned} \left| \frac{\partial^n}{\partial \lambda^n} (\varphi_\lambda(x) - \varphi_\lambda(y)) \right| &= \left| \frac{\partial^n}{\partial \lambda^n} \varphi_\lambda(x) - \frac{\partial^n}{\partial \lambda^n} \varphi_\lambda(y)) \right| \\ &\leqslant C_{A,n}(y-x)(1+y)y^n(1+|\lambda|)^2(1+e^{2(1/p-1)\rho y}). \end{aligned} \tag{5.5}$$

By Theorem 2.4(i) we observe $\mathcal{F}(T_x\phi - T_y\phi) = (\varphi_\lambda(x) - \varphi_\lambda(y))\hat{\phi}(\lambda)$. Therefore we deduce (iii) from (5.3), (5.5) and (4.25), and hence the lemma is proved. □

Theorem 5.6. *Suppose* $\phi \in L^1(\mathbf{R}_+, Adx)$ *and* $\int_0^\infty \phi(x)A(x)dx = 1$, *and* ϕ_t *is as in (5.1). Then*

(i) *for* $f \in C_c(\mathbf{R}_+)$, $\|f * \phi_t - f\|_\infty \to 0$ *as* $t \to 0^+$;

(ii) *for* $f \in L^p(\mathbf{R}_+, Adx)(1 \leqslant p < \infty)$, $\|f * \phi_t - f\|_{p,A} \to 0$ *as* $t \to 0^+$;

(iii) *if* $f, \phi \in \mathcal{S}_q(\mathbf{R}_+, *(A))(0 < q \leqslant 2)$ *then* $f * \phi_t \to f$ *in* $\mathcal{S}_q(\mathbf{R}_+, *(A))$ *as* $t \to 0^+$.

Proof. If $f \in C_c(\mathbf{R}_+)$ then there exists $M > 0$ such that $\operatorname{supp}(f) \subset [0, M]$. By [3, Lemma 1.2.1] we deduce that $u_f(x,y) = T_x f(y)$ is continuous on $[0, M+1] \times [0,1]$. From (1.7) we see that $\operatorname{supp}(T_y f) \subset [0, M+1]$ for $y \in [0,1]$. Hence for $1 \leqslant p \leqslant \infty$

$$\|T_y f - f\|_{p,A} \leqslant \left\{ \int_0^{M+1} A(x)dx \right\}^{1/p} \sup_{x \in [0, M+1]} |T_y f(x) - f(x)|, \quad 0 \leqslant y \leqslant 1. \tag{5.7}$$

By (5.1) we observe

$$\begin{aligned} (f * \phi_t)(x) - f(x) &= \int_0^\infty (T_x f(y) - f(x)))\phi_t(y)A(y)dy \\ &= \int_0^\infty (T_{ty} f(x) - f(x))\phi(y)A(y)dy. \end{aligned} \tag{5.8}$$

Thus (i) and (ii) follow from (5.7) and (5.8) together with an application of Lemma 4.12 and the Lebesgue dominated convergence theorem. Finally using Lemma 5.2(ii) we have

$$T_y f \to f \text{ in } \mathscr{S}_q(\mathbf{R}_+, *(A)) \quad (y \to 0^+) \tag{5.9}$$

and we obtain (iii) from (5.8) and (5.9). □

Let h_t be the heat kernel and P_t the Poisson kernel defined (see [1]) by

$$h_t(x) := \int_0^\infty e^{-t(\lambda^2+\rho^2)} \varphi_\lambda(x) \sigma(d\lambda)$$

and

$$P_t(x) := \int_0^\infty e^{-t(\lambda^2+\rho^2)^{1/2}} \varphi_\lambda(x) \sigma(d\lambda).$$

Theorem 5.10. (i) *Let* $0 < q \leqslant 2$ *and* $f \in \mathscr{S}_q(\mathbf{R}_+, *(A))$. *Then* $h_t \in \mathscr{S}_q(\mathbf{R}_+, *(A))$ $(t > 0)$ *and*

$$h_t * f \to f \text{ in } \mathscr{S}_q(\mathbf{R}_+, *(A)) (t \to 0^+). \tag{5.11}$$

(ii) *Let* $\rho > 0$ *and* $f \in \mathscr{S}_1(\mathbf{R}_+, *(A))$. *Then* $P_t \in \mathscr{S}_1(\mathbf{R}_+, *(A)) (t > 0)$ *and*

$$P_t * f \to f \text{ in } \mathscr{S}_1(\mathbf{R}_+, *(A)) \ (t \to 0^+). \tag{5.12}$$

Proof. Observe that $\mathscr{F}(h_t)(\lambda) = e^{-t(\lambda^2+\rho^2)}$ and $\mathscr{F}(P_t)(\lambda) = e^{-t(\lambda^2+\rho^2)^{1/2}}$ are even and analytic in $\mathbf{C}$ and in the interior of $\mathscr{F}_1$ respectively. A direct calculation gives that for $0 < t \leqslant 1$

$$\left| \frac{d^k}{d\lambda^k} e^{-t(\lambda^2+\rho^2)} \right| \leqslant C_k t (1 + |\lambda|)^k |e^{-t(\lambda^2+\rho^2)}|, \quad k \geqslant 1 \tag{5.13}$$

and for $|\mathrm{Im}\,(\lambda)| \leqslant \rho$

$$\left| \frac{d^k}{d\lambda^k} e^{-t(\lambda^2+\rho^2)^{1/2}} \right| \leqslant C_k t (1 + |\lambda|)^k |e^{-t(\lambda^2+\rho^2)^{1/2}}|, \quad \rho > 0, \ k \geqslant 1.$$

Hence $\mathscr{F}(h_t) \in \mathscr{S}(\mathscr{F}_\varepsilon)\left(\varepsilon = \frac{2}{q} - 1\right)$ and $\mathscr{F}(P_t) \in \mathscr{S}(\mathscr{F}_1)$. It follows using Theorem 4.27 that $h_t \in \mathscr{S}_q(\mathbf{R}_+, *(A))$ and $P_t \in \mathscr{S}_1(\mathbf{R}_+, *(A))$.

It remains to show (5.11) and (5.12). For any $k, l \in \mathbf{N}_0$ we have by (5.13)

$$\begin{aligned} \tau_{k,l}^{(\varepsilon)}(\hat{f}(\lambda)(1 - e^{-t(\lambda^2+\rho^2)})) &\leqslant \sup_{\lambda \in \mathscr{F}_\varepsilon} (1 + |\lambda|)^l |1 - e^{-t(\lambda^2+\rho^2)}| |\hat{f}^{(k)}(\lambda)| \\ &\quad + C_k t \sup_{\lambda \in \mathscr{F}_\varepsilon} (1 + |\lambda|)^{k+l} \sum_{j=0}^{k-1} |\hat{f}^{(j)}(\lambda)| \\ &:= \sigma_1 + \sigma_2. \end{aligned}$$

Given $\eta > 0$ there exists $R > 0$ such that $(1 + |\lambda|)^l |\hat{f}^{(k)}(\lambda)| < \eta$ for $\lambda \in \mathscr{F}_\varepsilon, |\lambda| > R$ since $\hat{f} \in \mathscr{S}(\mathscr{F}_\varepsilon)$. For $|\lambda| \leqslant R, \lambda = \xi + i\eta \in \mathscr{F}_\varepsilon$ we have $|1 - e^{-t(\lambda^2+\rho^2)}| \leqslant C_{R,\varepsilon,\rho} t$. Also we observe for all $\lambda \in \mathscr{F}_\varepsilon, |1 - e^{-t(\lambda^2+\rho^2)}| \leqslant C_{\varepsilon,\rho}$. Therefore we can find $\delta > 0$ such that for $0 < t < \delta$, $\sigma_1 < C_0 \eta$ and $\sigma_2 < C_0 \eta$

where C_0 is a positive constant depending only on f, k, l, ε and ρ. Thus we deduce $\lim_{t\to 0^+} \tau_{k,l}^{(\varepsilon)}(\hat{f}(\lambda)(1 - e^{-t(\lambda^2+\rho^2)})) = 0$, and (5.11) now follows from Theorem 2.4(ii) and (5.3). Similarly we can obtain (5.12). □

Remark. If $\rho = 0$ then $\mathscr{F}(P_t)(\lambda) = e^{-t|\lambda|}$ which is not differentiable at $\lambda = 0$. So $\mathscr{F}(P_t) \notin \mathscr{S}(\mathscr{F}_1) = \mathscr{S}_0(\mathbf{R}_+)$, and by Theorem 4.27, $P_t \notin \mathscr{S}_1(\mathbf{R}_+, *(A))$.

We can establish a result about approximations to the identity for distributions as well. We first introduce distributions on Chébli-Trimèche hypergroups and discuss some of their properties.

Definition 5.14. For $0 < q \leqslant 2$ a *q-distribution* on $\mathbf{R}_+$ is a continuous linear functional on $\mathscr{S}_q(\mathbf{R}_+, *(A))$. The totality of q-distributions on $\mathbf{R}_+$ is denoted by $\mathscr{S}'_q(\mathbf{R}_+, *(A))$.

As usual the class $\mathscr{S}'_q(\mathbf{R}_+, *(A))$ is given the structure of a complex vector space by the natural definition of addition and multiplication by complex numbers. We observe the following obvious inclusions:

$$\mathscr{D}_* \subset \mathscr{S}_q(\mathbf{R}_+, *(A)) \subset \mathscr{S}'_q(\mathbf{R}_+, *(A)), \quad 0 < q \leqslant 2$$

and

$$\mathscr{S}_1(\mathbf{R}_+, *(A)) \subset L^p(\mathbf{R}_+, Adx) \subset \mathscr{S}'_1(\mathbf{R}_+, *(A)), \quad 1 \leqslant p \leqslant \infty.$$

Here a function f is regarded as a q-distribution in the sense that

$$u_f(\phi) = (f, \phi) := \int_0^\infty f(x)\phi(x)A(x)dx, \quad \phi \in \mathscr{S}_q(\mathbf{R}_+, *(A))$$

defines a continuous linear functional on $\mathscr{S}_q(\mathbf{R}_+, *(A))$. However we use f to denote both the function and the corresponding q-distribution. Clearly every locally integrable function f satisfying $|f(x)| \leqslant Mx^k$ for large x (where $k \in \mathbf{N}_0$ and $M > 0$) can be regarded as a distribution in $\mathscr{S}'_1(\mathbf{R}_+, *(A))$. The elements of $\mathscr{S}'_1(\mathbf{R}_+, *(A))$ are called tempered distributions.

Lemma 5.15. *Let $0 < q \leqslant 2$ and u a linear functional on $\mathscr{S}_q(\mathbf{R}_+, *(A))$. Then $u \in \mathscr{S}'_q(\mathbf{R}_+, *(A))$ if and only if there exist a constant $C > 0$ and $k \in \mathbf{N}_0$ such that*

$$|u(\phi)| \leqslant C \sum_{j=0}^{k} \mu_{j,k}(\phi), \quad \phi \in \mathscr{S}_q(\mathbf{R}_+, *(A)).$$

Proof. The lemma follows from the general result [12, Lemma 1.1]. □

Definition 5.16. The convolution of $u \in \mathscr{S}'_q(\mathbf{R}_+, *(A))$ and $\phi \in \mathscr{S}_q(\mathbf{R}_+, *(A))$ is a q-distribution defined by

$$(u * \phi)(\psi) := u(\phi * \psi), \quad \psi \in \mathscr{S}_q(\mathbf{R}_+, *(A)).$$

Lemma 5.2 shows that this convolution is well-defined.

Theorem 5.17. *Suppose $u \in \mathscr{S}'_q(\mathbf{R}_+, *(A))$ and $\phi \in \mathscr{S}_q(\mathbf{R}_+, *(A))$. Let $f(x) := u(T_x\phi)$ for $x \in \mathbf{R}_+$. Then*

(i) $f \in C^1(\mathbf{R}_+)$ *(the space of continuous functions on* $\mathbf{R}_+$ *that are continuously differentiable on* $]0,\infty[$*);*

(ii) *there exist constants* $C_\phi > 0$ *and* $r \in \mathbf{N}_0$ *such that*

$$|f(x)| \leqslant C_\phi(1+x)x^r e^{2(1/q-1)\rho x}, \quad x \in \mathbf{R}_+;$$

(iii) $u * \phi$ *is a function and* $(u*\phi)(x) = f(x)$.

Proof. Fixing $x \in \mathbf{R}_+$ we have for $h \in \mathbf{R}$ and $h + x > 0$

$$\frac{f(x+h)-f(x)}{h} = u\left(\frac{T_{x+h}\phi - T_x\phi}{h}\right).$$

By Theorem 4.27 it is easy to see that

$$\mathscr{F}\left(\frac{\partial}{\partial x}T_x\phi\right)(\lambda) = \varphi'_\lambda(x)\hat{\phi}(\lambda). \tag{5.18}$$

In view of Theorem 2.4(i) and (5.18) we have for any $r, s \in \mathbf{N}_0$

$$\begin{aligned} &\tau^{(\varepsilon)}_{r,s}\left(\mathscr{F}\left(\frac{T_{x+h}\phi - T_x\phi}{h} - \frac{\partial}{\partial x}T_x\phi\right)\right) \\ &\quad = \sup_{\lambda \in \mathscr{F}_\varepsilon}(1+|\lambda|)^s \left|\frac{\partial^r}{\partial \lambda^r}\left(\left(\frac{\varphi_\lambda(x+h)-\varphi_\lambda(x)}{h} - \varphi'_\lambda(x)\right)\hat{\phi}(\lambda)\right)\right|. \end{aligned} \tag{5.19}$$

We assume $x > 0$ and $0 < h \leqslant 1$ (the other cases can be treated similarly). Applying Taylor's expansion together with (3.12) and Lemma 3.11 there exists $\xi \in]x, x+h[$ such that

$$\begin{aligned} &\left|\frac{\partial^n}{\partial \lambda^n}\left(\frac{\varphi_\lambda(x+h)-\varphi_\lambda(x)}{h} - \varphi'_\lambda(x)\right)\right| \\ &\quad = \left|\frac{h}{2}\left(\frac{\partial^n}{\partial \lambda^n}\varphi_\lambda\right)''(\xi)\right| \\ &\quad \leqslant C_{A,n}h(1+|\lambda|^2)(1+x)^{n+2}(1+e^{2(1/q-1)\rho x}). \end{aligned} \tag{5.20}$$

Therefore by (5.3), (5.19) and (5.20) we deduce that for any $k, l, \in \mathbf{N}_0$

$$\mu^q_{k,l}\left(\frac{T_{x+h}\phi - T_x\phi}{h} - \frac{\partial}{\partial x}T_x\phi\right) \to 0 \quad (h \to 0).$$

Thus f is differentiable and

$$f'(x) = u\left(\frac{\partial}{\partial x}T_x\phi\right).$$

Setting $\phi_{1,x} = \dfrac{\partial}{\partial x}T_x\phi$ we have by Theorem 2.4(i) and (5.18)

$$\mathscr{F}(\phi_{1,x} - \phi_{1,x_0})(\lambda) = (\varphi'_\lambda(x) - \varphi'_\lambda(x_0))\hat{\phi}(\lambda).$$

Hence from (5.3), (3.12), Lemma 3.11 and the mean-value theorem we deduce that $\phi_{1,x} \to \phi_{1,x_0}$ in $\mathscr{S}_q(\mathbf{R}_+, *(A))$ as $x \to x_0$. Consequently $f' \in C(\mathbf{R}_+)$ and this proves (i). Conclusion (ii) follows immediately from Lemmas 5.15 and 5.2(ii).

To prove (iii) we need only show that for each $\psi \in \mathscr{S}_q(\mathbf{R}_+, *(A))$

$$(u * \phi)(\psi) = \int_0^\infty u(T_x\phi)\psi(x)A(x)dx. \tag{5.21}$$

Both sides of (5.21) make sense because of (ii), Definition 5.16 and Lemma 5.2(i). By (i), $f(x) = u(T_x\phi)$ is continuous on $\mathbf{R}_+$. For any $b > 0$ let (s_n) be the sequence of the Riemann sums approximating $\int_0^b f(x)\psi(x)\, m(dx)$, viz.

$$s_n = \sum_{j=1}^n f(x_{jn})\psi(x_{jn})m(\Delta_{jn})$$

where $\Delta_{jn} = \left[\frac{j-1}{n}b, \frac{j}{n}b\right]$ and m is the (normalized) Haar measure on $(\mathbf{R}_+, *(A))$. Then $s_n = u(\sigma_n)$ where $\sigma_n = \sum_{j=1}^n \psi(x_{jn})T_{x_{jn}}\phi m(\Delta_{jn})$. By the integral mean-value theorem we observe that for $k \in \mathbf{N}_0$ there exist $\tilde{x}_{jn} \in \Delta_{jn}$ such that

$$\begin{aligned}\int_0^b \psi(x)\frac{\partial^k}{\partial y^k}T_x\phi(y)m(dx) &= \sum_{j=1}^n \int_{\Delta_{jn}} \psi(x)\frac{\partial^k}{\partial y^k}T_x\phi(y)m(dx)\\ &= \psi(\tilde{x}_{jn})\frac{\partial^k}{\partial y^k}T_{\tilde{x}_{jn}}\phi(y)m(\Delta_{jn}).\end{aligned}$$

Thus for any $k, l \in \mathbf{N}_0$

$$\begin{aligned}&\mu_{k,l}^q\left(\sigma_n - \int_0^b \psi(x)T_x\phi\, m(dx)\right)\\ &= \sup_{y\in\mathbf{R}_+}(1+y)^l\varphi_0(y)^{-2/q}\left|\sigma_n^{(k)}(y) - \int_0^b \psi(x)\frac{\partial^k}{\partial y^k}T_x\phi(y)m(dx)\right|\\ &\leqslant \sup_{y\in\mathbf{R}_+}(1+y)^l\varphi_0(y)^{-2/q}\sum_{j=1}^n m(\Delta_{jn})\\ &\quad\times\left|\psi(x_{jn})\frac{\partial^k}{\partial y^k}T_{x_{jn}}\phi(y) - \psi(\tilde{x}_{jn})\frac{\partial^k}{\partial y^k}T_{\tilde{x}_{jn}}\phi(y)\right|\\ &\leqslant M\sum_{j=1}^n m(\Delta_{jn})\mu_{k,l}^q(T_{x_{jn}}\phi - T_{\tilde{x}_{jn}}\phi)\\ &\quad+\sum_{j=1}^n m(\Delta_{jn})\mu_{k,l}^q(T_{\tilde{x}_{jn}}\phi)|\psi(x_{jn}) - \psi(\tilde{x}_{jn})|\end{aligned} \tag{5.22}$$

where $M = \sup_{x\in\mathbf{R}_+}|\psi(x)|$. Observe now that $x_{jn}, \tilde{x}_{jn} \in \Delta_{jn} \subset [0, b]$. Hence by (5.22), Lemma 5.2 (ii), (iii) and the fact that ψ is uniformly continuous on $\mathbf{R}_+$ we deduce that for each $b > 0$

$$\sigma_n \to \int_0^b \psi(x)T_x\phi(y)m(dx) \text{ in } \mathscr{S}_q(\mathbf{R}_+, *(A)) \quad (n \to \infty)$$

which implies (note that $y \mapsto \int_0^b \psi(x) T_x \phi(y) m(dx) \in \mathscr{S}_q(\mathbf{R}_+, *(A))$

$$u(\sigma_n) \to u\left(\int_0^b \psi(x) T_x \phi \, m(dx)\right) \quad (n \to \infty).$$

Since $(\mathbf{R}_+, *(A))$ is commutative we have $T_x\phi(y) = T_y\phi(x)$. Thus by the definition of the convolution we obtain for each $b > 0$

$$\int_0^b u(T_x\phi)\psi(x) m(dx) = u((\mathbb{1}_{[0,b]}\psi) * \phi). \tag{5.23}$$

Finally by (5.23) we observe

$$\begin{aligned}
&\left| \int_0^\infty \psi(x) u(T_x\phi) m(dx) - u(\psi * \phi) \right| \\
&\quad = \left| \int_b^\infty \psi(x) u(T_x\phi) m(dx) - u((\mathbb{1}_{[b,\infty[}\psi) * \phi) \right| \\
&\quad \leqslant \int_b^\infty |\psi(x) u(T_x\phi)| m(dx) - |u((\mathbb{1}_{[b,\infty[}\psi) * \phi)|.
\end{aligned}$$

The last integral tends to 0 as $b \to \infty$ since by (ii), $\psi(x)u(T_x\phi) \in L^1(\mathbf{R}_+, Adx)$. To prove $u((\mathbb{1}_{[b,\infty[}\psi) * \phi) \to 0$ as $b \to \infty$ it suffices to show $(\mathbb{1}_{[b,\infty[}\psi) * \phi \to 0$ in $\mathscr{S}_q(\mathbf{R}_+, *(A))$ as $b \to \infty$. But for any $k, l \in \mathbf{N}_0$

$$\mu_{k,l}^q((\mathbb{1}_{[b,\infty[}\psi) * \phi) \leqslant \int_b^\infty \psi(t) \mu_{k,l}^q(T_t\phi) m(dt).$$

By Lemma 5.2(ii) the last integral tends to 0 as $b \to \infty$. We therefore obtain (5.21) and this completes the proof of the theorem. □

Theorem 5.24. *Suppose* $\phi \in L^1(\mathbf{R}_+, Adx) \cap \mathscr{S}_q(\mathbf{R}_+, *(A))(0 < q \leqslant 2)$ *and* $\int_0^\infty \phi(x) m(dx) = 1$. *Then for any* $u \in \mathscr{S}'_q(\mathbf{R}_+, *(A)), u * \phi_t \to u$ *in* $\mathscr{S}'_q(\mathbf{R}_+, *(A))$ *as* $t \to 0^+$.

Proof. The theorem follows from Definition 5.16 and Theorem 5.6(iii). □

The authors are grateful to the referee for the many helpful comments on an earlier version of this paper.

References

[1] Achour A, Trimèche K (1983) La g-fonction de Littlewood-Paley associée à un opérateur différentiel singulier sur $(0, \infty)$. Ann Inst Fourier (Grenoble) **33**: 203–226

[2] Anker J-P (1991) The spherical Fourier transform of rapidly decreasing functions. A simple proof of a characterization due to Harish-Chandra, Helgason, Trombi, and Varadarajan. J Funct Anal **96**: 331–349

[3] Bloom WR, Heyer H (1995) Harmonic Analysis of Probability Measures on Hypergroups. Berlin New York: Walter de Gruyter

[4] Bloom WR, Xu Z (1995) The Hardy-Littlewood maximal function for Chébli–Trimèche hypergroups. Contemporary Math **183**: 45–69

[5] Chébli H (1974) Sur un théorème de Paley-Wiener associé à la décomposition spectrale d'un opérateur de Sturm-Liouville sur $]0, \infty[$. J Funct Anal **17**: 447–461

[6] Dunkl CF (1973) The measure algebra of a locally compact hypergroup. Trans Amer Math Soc **179**: 331–348

[7] Fitouhi A (1989) Heat "polynomials" for a singular differential operator on $]0,\infty[$. Constr Approx **5**: 241–270
[8] Flensted-Jensen M (1972) Paley-Wiener type theorems for a differential operator connected with symmetric spaces. Ark Mat **10**: 143–162
[9] Gangolli R, Varadarajan VS (1988) Harmonic Analysis of Spherical Functions on Real Reductive Groups. Berlin New York: Springer
[10] Jewett RI (1975) Spaces with an abstract convolution of measures. Adv Math **18**: 1–101
[11] Spector R (1975) Aperçu de la théorie des hypergroupes. In: Analyse harmonique sur les groupes de Lie (Sém. Nancy-Strasbourg, 1973–1975), pp 643–673. Lect Notes Math **497**. Berlin: Springer
[12] Szmydt Z (1977) Fourier Transformation and Linear Differential Equations. Warszawa: Polish Scientific Publ
[13] Trimèche K (1981) Transformation intégrale de Weyl et théorème de Paley-Wiener associés à un opérateur différentiel singulier sur $(0,\infty)$. J Math Pures Appl (9) **60**: 51–98
[14] Trimèche K (1995)) Inversion of the J. L. Lions transmutation operators using generalized wavelets. Faculty of Sciences of Tunis (preprint)
[15] Xu Z (1994) Harmonic analysis on Chébli–Trimèche hypergroups. PhD Thesis, Murdoch University, Australia
[16] Zeuner H (1989) Laws of large numbers for hypergroups on $\mathbf{R}_+$. Math Ann **283**: 657–678
[17] Zeuner H (1989) One-dimensional hypergroups. Adv Math **76**: 1–18

W. R. Bloom and Zengfu Xu
Division of Science
Murdoch University
Murdoch, WA 6150
Australia
e-mail: bloom@murdoch.edu.au
zengfu@murdoch.edu.au

Mh. Math. 125, 111–126 (1998)

Monatshefte für
Mathematik

Printed in Austria

The Normal Density of Prime Ideals in Small Regions

By

M. D. Coleman, Manchester

(Received 9 April 1996; in revised form 12 August 1996)

Abstract. For a Gaussian prime $\omega \in \mathbb{Q}(i)$ define $\rho(\omega) = \min|\omega - \omega'|$ where ω' runs through Gaussian primes satisfying $|\omega'| > |\omega|$. We prove that, subject to the Riemann Hypothesis for appropriate L-functions

$$\sum_{N\omega \leqslant x} \frac{\rho^4(\omega)}{N\omega} \ll \log^3 x,$$

which generalises a result due to Selberg (Archiv for Mathematik og Naturvidenskab **47** (1943) 87–105).

1. Introduction

In 1943 Selberg proved, subject to the Riemann Hypothesis, that

$$\sum_{p_n \leqslant x} \frac{(p_n - p_{n-1})^2}{p_n} \ll \log^3 x.$$

For a Gaussian prime $\omega \in \mathbb{Q}(i)$ define $\rho(\omega) = \min|\omega - \omega'|$ where ω' runs through Gaussian primes satisfying $|\omega'| > |\omega|$. We will prove

Theorem 1. *Assume the Extended Riemann Hypothesis in* $\mathbb{Q}(i)$. *Then*

$$\sum_{N\omega \leqslant x} \frac{\rho^4(\omega)}{N\omega} \ll \log^3 x. \tag{1}$$

The Extended Riemann Hypothesis will be discussed in Section 2. A result similar to (1) can be given in a general number field but we will delay the definition of ρ to the next section. The (almost) trivial result in this area is

Lemma 2.

$$\sum_{N\omega \leqslant x} \rho^2(\omega) \asymp x. \tag{2}$$

Here $A \asymp B$ means $A \ll B \ll A$. We have the condition $|\omega'| > |\omega|$ in the definition of ρ to ease comparison with Selbergs' result, but it does complicate proofs. From (2) it can be deduced that given a positive function Φ satisfying

1991 Mathematics Subject Classification: 11R44
Key words: Gaussian Primes, Hecke L-functions

$\Phi(x) \to \infty$ as $x \to \infty$, then $\rho(\omega) < \Phi(|\omega|)\ \log^{1/2} |\omega|$ for almost all ω, a result stronger than can be deduced from Theorem 1. Yet the method that gives (1) leads to non-trivial results for almost all $z \in \mathbb{C}$. These are given in Corollary 4 below.

Let $\mathcal{O}_k$ denotes the ring of integers of K a number field of degree n. Let $M \subseteq \mathcal{O}_K$ be a full module in K. Given $\gamma \in \mathcal{O}_K$ consider $N_{K/\mathbb{Q}}(\gamma + \mu)$ for all $\mu \in M$. Let J be the largest positive divisior of these rational integers. Then, if $\{\alpha_i\}_{1 \leqslant i \leqslant n}$ is a $\mathbb{Z}$-basis of M, we can define

$$f_\gamma(\mathbf{x}) = J^{-1} N_{K/\mathbb{Q}}(\gamma + \alpha(\mathbf{x}))$$

for $\mathbf{x} \in \mathbb{R}^n$, where $\alpha(\mathbf{x}) = \sum_{i=1}^n x_i \alpha_i$. These have been studied in [7] and [2]. We can give a conditional form of Theorem 9 of [2] as

Theorem 3. *Assume the Extended Riemann Hypothesis in K. Let $\Phi(x)$ be a monotonically increasing function with $\Phi(x) \to \infty$ as $x \to \infty$. Then for almost all $\mathbf{x} \in \mathbb{R}^n$ there exist $\mathbf{m} \in \mathbb{Z}^n$ such that $|f_\gamma(\mathbf{m})|$ is prime and*

$$\|\mathbf{m} - \mathbf{x}\|_2 \ll \Phi(\|\mathbf{x}\|_2)\ \log^{2/n} \|\mathbf{x}\|_2.$$

Here $\|\mathbf{x}\|_2 = (\sum_{i=1}^n x_i^2)^{1/2}$. The proof of this result follows closely the methods in [2] and it is necessary to have that paper at hand to follow.

An immediate consequence of Theorem 3 is that we obtain

Corollary 4. *Assume the Extended Riemann Hypothesis in $\mathbb{Q}(i)$. Let $\Phi(x)$ be as above. Then for almost all $z \in \mathbb{C}$ there exists a Gaussian prime ω such that*

$$|\omega - z| \ll \Phi(|z|)\ \log |z|.$$

This should be compared with the unconditional result of [2], that for all $\varepsilon > 0$

$$\exists\, \omega : |\omega - z| \ll |z|^{1/6+\varepsilon}$$

for almost all z. Theorem 5.1 [1] is the conditional result that

$$\exists\, \omega : |\omega - z| \ll |z|^{1/2}\ \log |z|$$

for all z while Corollary 3 of [3] is the unconditional result that

$$\exists\, \omega : |\omega - z| \ll |z|^{11/20+\varepsilon}$$

for all z.

2. The Extended Riemann Hypothesis

Let K be a number field of degree $n = r_1 + 2r_2$, $\mathfrak{f}$ a fixed integral ideal and $I_\mathfrak{f}$ the group of fractional ideals of K whose prime decomposition contains no prime factors of $\mathfrak{f}$. Let

$$P_{\mathfrak{f}\infty} = \{(\alpha) \in I_\mathfrak{f},\ \alpha \in K^*,\ \alpha \equiv 1\ (\mathrm{mod}\, \mathfrak{f}),\ \alpha > 0\},$$

and let χ denote a narrow ideal class character mod $\mathfrak{f}$, that is a character on $I_\mathfrak{f}/P_{\mathfrak{f}\infty}$. Then an arbitrary Groessencharakter mod $\mathfrak{f}$ is of the form $\chi\lambda^{\mathbf{m}}$ with $\mathbf{m} \in \mathbb{Z}^{n-1}$,

$$\lambda^{\mathbf{m}} = \lambda_1^{m_1} \ldots \lambda_{n-1}^{m_{n-1}}$$

where $\{\lambda_i\}$ forms a basis for the torsion-free characters on $I_{\mathfrak{f}}$ whose values on any $\mathfrak{a} \in P_{\mathfrak{f}\infty}$ depends only on the $\alpha \in K^*$, $\alpha \equiv 1 \pmod{\mathfrak{f}}$, $\alpha > 0$ that exists such that $\mathfrak{a} = (\alpha)$.

For $\mathfrak{a} \in I_{\mathfrak{f}}$ define

$$\psi(\mathfrak{a}) = (\psi_j(\mathfrak{a})) \in \mathbb{R}^{n-1}/\mathbb{Z}^{n-1} = \mathbf{T}^{n-1}$$

by

$$\lambda_j(\mathfrak{a}) = e^{2\pi i \psi_j(\mathfrak{a})}, \quad 1 \leqslant j \leqslant n-1.$$

Given $I \in I_{\mathfrak{f}}/P_{\mathfrak{f}\infty}$, $\psi_0 \in \mathbf{T}^{n-1}$ and $0 \leqslant l \leqslant \frac{1}{2}$ set

$$\mathscr{S}(x, \psi_0, l) = \{\mathfrak{a} \in I_{\mathfrak{f}},\ x(1-l) \leqslant N\mathfrak{a} \leqslant x(1+l), |\psi_j(\mathfrak{a}) - \psi_{oj}| < l,\ 1 \leqslant j \leqslant n-1\}.$$

Further, define

$$\pi_I(x, \psi_0, l) = |\{\mathfrak{p} \in I,\ \mathfrak{p} \in \mathscr{S}(x, \psi_0, l)\}|,$$

where $\mathfrak{p}$ denotes a prime ideal of K and

$$\theta_I(x, \psi_0, l) = \sum_{\mathfrak{p} \in \mathscr{S}(x, \psi_0, l)} \log N\mathfrak{p}.$$

We will express this in terms of the Hecke L-function in K defined by

$$L(s, \chi\lambda^{\mathbf{m}}) = \sum_{(\mathfrak{a}, \mathfrak{f})=1} \chi\lambda^{\mathbf{m}}(\mathfrak{a}) N\mathfrak{a}^{-s},$$

for Re $s > 1$, where the sum is over integral ideals co-prime to $\mathfrak{f}$. This has a meromorphic continuation to $\mathbb{C}$: the continuation is entire unless $\chi\lambda^{\mathbf{m}} \equiv 1$ on $I_{\mathfrak{f}}$, that is $\chi = \chi_0$ and $\mathbf{m} = \mathbf{0}$, when it has a simple pole at $s = 1$ and no other singularity. When χ is a primitive character we have a functional equation

$$L(s, \chi\lambda^{\mathbf{m}}) = w(\chi) A^{1-2s} G(1-s, \chi\lambda^{\mathbf{m}}) L(1-s, \bar{\chi}\bar{\lambda}^{\mathbf{m}})$$

where $|w| = 1, A^2 = |d|\pi^{-n}2^{-r_2}N\mathfrak{f}$ (d is the discriminant of K) and G is a quotient of gamma factors. We can use this to conclude that $L(s, \chi\lambda^{\mathbf{m}})$ has zeros (so called trivial zeros) where the gamma factors in $G(1-s, \chi\lambda^{\mathbf{m}})$ have poles. These all lie in the half plane Re $s \leqslant 0$. Classical methods based on the order of growth of the L-functions show the existence of further zeros in the critical strip $0 <$ Re $s < 1$. The Extended Riemann Hypothesis of Section 1 is that, for all Groessencharaktere the zeros of $L(s, \chi\lambda^{\mathbf{m}})$ in the critical strip lie on the line Re $s = \frac{1}{2}$.

The main result of this paper is

Theorem 5. *Subject to the Extended Riemann Hypothesis we have*

$$\int_{\mathbf{T}^{n-1}} \int_1^{l^{-A}} \left| \frac{\theta_I(x, \psi, l) - (2l)^n x/h(\mathfrak{f})}{x} \right|^2 dx\, d\psi \ll_A l^n \log(1/l)$$

for any $A > 0$. Here $h(\mathfrak{f})$ is the order of $I_{\mathfrak{f}}/P_{\mathfrak{f}\infty}$.

This is an analogue of a result implicit in [9] (see p. 169). It should be compared with Theorem 3 of [2], an unconditional result.

Following the argument on pp. 170 and 171 of [9] we can prove

Theorem 6. *Assume the Extended Riemann Hypothesis. Let $L(x)$ be a positive function decreasing to zero such that $x^{1/n}L(x)$ is increasing and $xL^n(x)/\log^2 x \to \infty$ as $x \to \infty$. Then*

$$\pi(x, \psi, L(x)) \sim \frac{(2L(x))^n x}{h(\mathfrak{f})}$$

for all (x, ψ) except for those in the exceptional set E satisfying

$$\int_{\substack{\mathbf{T}^{n-1} \\ (x,\psi)\in E}} \int_1^X dx\, d\psi \ll \left(\frac{\log^2 X}{XL^n(X)}\right)^{\frac{1}{4}} X.$$

So, almost all small regions contain the expected number or, in Selbergs terminology, the normal density of prime ideals. We will not give details of the proof of Theorem 6 but rather of Theorem 3 which follows similar lines but is complicated by the existence of fundamental units in the number field.

A result similar to Theorem 1 can be given for a general number field on defining

$$\rho(\mathfrak{p}) = (N\mathfrak{p})^{1/n} \max \{l : \mathscr{S}(N\mathfrak{p}, \psi(\mathfrak{p}), l) = 1\},$$

when, subject to the Extended Riemann Hypothesis,

$$\sum_{N\mathfrak{p}\leqslant x} \frac{\rho^{2n}(\mathfrak{p})}{N\mathfrak{p}} \ll \log^3 x.$$

We only give full details of Theorem 1 since ρ is defined sligthly differently, in that result, to the above.

3. Mean Value Results

Fundamental to the proof of Theorem 5 is

Theorem 7. *Assume the Extended Riemann Hypothesis in K. Then, for any character χ mod $\mathfrak{f}$,*

$$\sum_{\substack{\mathbf{m} \\ M_i < m_i \leqslant M_i + W}} \int_T^{T+W} \left|\frac{L'}{L}(\sigma + it, \chi\lambda^{\mathbf{m}})\right|^2 dt \ll \frac{W^n}{(\sigma - \frac{1}{2})^2} \frac{\log^2 V}{\log^2 W} \tag{4}$$

holds for $\sigma > \frac{1}{2} + \frac{8}{\log W^n}$ where $V = T^2 + M_1^2 + \cdots + M_{n-1}^2 + W^2$.

This should be compared with Lemma 4 of [9] whose proof we follow closely.

Proof. If χ is not primitive then it is induced by χ^*, a primitive character mod $\mathfrak{q}$, where $\mathfrak{q}|\mathfrak{f}$ and $\chi^*\lambda^{\mathbf{m}}$ is a primitive Groessencharakter mod $\mathfrak{q}$ (see [6]). Further

$$L(s, \chi\lambda^{\mathbf{m}}) = L(s, \chi^*\lambda^{\mathbf{m}}) \prod_{\mathfrak{p}|\mathfrak{f}} (1 - \chi^*\lambda^{\mathbf{m}}(\mathfrak{p})N^{-s}\mathfrak{p}).$$

In the logarithmic derivative in (4) these additional Euler factors are easily estimated so we can assume that the χ in (4) is primitive.

Let $x > 1$ and define

$$\Lambda_x(\mathfrak{a}) = \begin{cases} \Lambda(\mathfrak{a}) & \text{for } 1 \leqslant N\mathfrak{a} \leqslant x \\ \Lambda(\mathfrak{a}) \dfrac{\log x^2/N\mathfrak{a}}{\log x} & \text{for } x \leqslant N\mathfrak{a} \leqslant x^2. \end{cases}$$

Then

$$\sum_{N\mathfrak{a} \leqslant x^2} \frac{\Lambda_x(\mathfrak{a})\chi\lambda^{\mathbf{m}})(\mathfrak{a})}{N\mathfrak{a}^s} = \frac{1}{2\pi i \log x} \int_{c-i\infty}^{c+i\infty} \frac{x^{z-s} - x^{2(z-s)}}{(z-s)^2} \frac{L'}{L}(z, \chi\lambda^{\mathbf{m}})dz \tag{5}$$

where $c = \max(2, 1+\sigma)$.

As noted by Duke ([5], p. 221) the arguments used for Dirichlet L-series (see [4]) show that there exist a sequence of contours consisting of straight lines connecting $(c + iT_j)$, $(c_j + iT_j)$, $(c_j - iT_j)$ and $(c - iT_j)$, where $c_j < -j, j < T_j < j+1$, and on which

$$\frac{L'}{L}(z, \chi\lambda^{\mathbf{m}}) \ll \log^2 j. \tag{6}$$

The sequence of contours depends on $\mathbf{m}$. If we require a sequence that works simultaneously for all $\|\mathbf{m}\| < W$, the bound in (6) increases. But for any single value of $\mathbf{m}$ we can move the path of integration in (5) to the left, passing over poles at the following points:

$$z = s \quad \text{with residue} \quad \frac{-L'}{L}(s, \chi\lambda^{\mathbf{m}}) \log x,$$

$$z = 1 \quad \text{with residue} \quad -\frac{(x^{1-s} - x^{2(1-s)})}{(1-s)^2},$$

(though only when $\chi = \chi_0, \mathbf{m} = \mathbf{0}$), and the zeros $\rho_{\mathbf{m}\chi}$, both trivial and non-trivial, of $L(z, \chi\lambda^{\mathbf{m}})$ with residues

$$\frac{x^{\rho_{\mathbf{m}\chi}-s} - x^{2(\rho_{\mathbf{m}\chi}-s)^2}}{(s - \rho_{\mathbf{m}\chi})^2} \tag{7}$$

The trivial zeros of $L(z, \chi\lambda^{\mathbf{m}})$ are the poles of the gamma factors in

$$H(s, \chi\lambda^{\mathbf{m}}) = \prod_{q=1}^{r_1} \Gamma\left(\frac{1}{2}(s + a_q - ib_q)\right) \prod_{q=r_1+1}^{r_1+r_2} \Gamma\left(s + \frac{1}{2}|a_q| - ib_q\right)$$

where $a_1, \ldots, a_{r_1} \in \{0, 1\}$ are determined by χ and $a_{r_1+1}, \ldots, a_{r_1+r_2} \in \mathbb{Z}$, $b_1, \ldots, b_{r_1+r_2} \in \mathbb{R}$ depend only on $\mathbf{m}$. If s is not equal to a trivial zero the sum of (7) over such zeros is absolutely convergent. So, if $\operatorname{Re} s > \frac{1}{2}$ say, the sum is bound independently of the $\{a_i\}$, $\{b_i\}$, i.e. χ and $\mathbf{m}$. Thus, using the Extended

Riemann Hypothesis by writing $\rho_{\mathbf{m}\chi} = \frac{1}{2} + i\gamma_{\mathbf{m}\chi}$,

$$\frac{L'}{L}(s, \chi\lambda^{\mathbf{m}}) = -\sum_{N\mathfrak{a} \leqslant x^2} \frac{\Lambda_x(\mathfrak{a})\chi\lambda^{\mathbf{m}}(\mathfrak{a})}{N\mathfrak{a}^s} + O\left(\delta_{\mathbf{m},\mathbf{0}} \frac{x}{1+t^2}\right) + O(1) + \frac{2w}{\log x} \sum_{\gamma_{\mathbf{m}\chi}} \frac{1}{(\sigma - \frac{1}{2})^2 + (t - \gamma_{\mathbf{m}\chi})^2}, \tag{8}$$

for some $|w| < 1$, where $\delta_{\mathbf{m},\mathbf{0}} = 1$ if $\chi = \chi_0$ and $\mathbf{m} = \mathbf{0}$, zero otherwise and the constants implied by the O-notation are independent of $\mathbf{m}$. We require the analogue of a result proved by Ingham for the Riemann zeta function. This is

$$\operatorname{Re} \frac{L'}{L}(\sigma + it, \chi\lambda^{\mathbf{m}}) = O(\log V) + \sum_{\gamma_{\mathbf{m}\chi}} \frac{\sigma - \frac{1}{2}}{(\sigma - \frac{1}{2})^2 + (t - \gamma_{\mathbf{m}\chi}^2)}, \tag{9}$$

valid for $|t| < V$, $\|\mathbf{m}\| < V$. For this we can follow the proof of Lemma 16 in [10]. As there $H(s, \chi\lambda^{\mathbf{m}})L(s, \chi, \lambda^{\mathbf{m}})$ is an integral function of order 1 with zeros $\rho_{\mathbf{m}\chi}$ and no others. Thus

$$\frac{d}{ds}\left(\frac{H'}{H}(s, \chi\lambda^{\mathbf{m}}) + \frac{L'}{L}(s, \chi\lambda^{\mathbf{m}})\right) = -\sum_{\rho_{\mathbf{m}\chi}} \frac{1}{(s - \rho_{\mathbf{m}\chi})^2}.$$

Integrate this from s to s' where $\operatorname{Re} s' = \frac{1}{2}$. Note from the functional equation that

$$\log A + \frac{H'}{H}(s', \chi\lambda^{\mathbf{m}}) + \frac{L'}{L}(s', \chi\lambda^{\mathbf{m}})$$

is purely imaginary, being the negative of its complex conjugate. So, on taking real parts,

$$\operatorname{Re} \frac{L'}{L}(s, \chi\lambda^{\mathbf{m}}) = \sum_{\rho_{\mathbf{m}\chi}} \frac{\sigma - \frac{1}{2}}{(\sigma - \frac{1}{2})^2 + (t - \gamma_{\mathbf{m}\chi})^2} - \log A - \operatorname{Re} \frac{H'}{H}(s, \chi\lambda^{\mathbf{m}})$$

having used $\beta_{\mathbf{m}\chi} = \frac{1}{2}$. Here, $H'/H(s, \chi\lambda^{\mathbf{m}})$ can be written as a linear sum of

$$\frac{\Gamma'}{\Gamma}\left(\frac{1}{2}(s + a_q - ib_q)\right) \quad \text{and} \quad \frac{\Gamma'}{\Gamma}\left(s + \frac{1}{2}|a_q| - ib_q\right)$$

each of which is $\ll \log(e + t^2 + a_q^2 + b_q^2)$. But as in Section 2 of [2],

$$\sum_{q=1}^{r_1+r_2} (a_q^2 + b_q^2) \ll \sum_{i=1}^{n-1} m_i^2 \ll V^2.$$

Hence

$$\frac{H'}{H}(s, \chi\lambda^{\mathbf{m}}) \ll \log V^2$$

as required.

Combining (8) and (9) gives

$$\frac{L'}{L}(s,\chi\lambda^{\mathbf{m}}) \ll \left|\sum_{N\mathfrak{a}\leqslant x^2}\frac{\Lambda_x(\mathfrak{a})\chi\lambda^{\mathbf{m}}(\mathfrak{a})}{N\mathfrak{a}^s}\right| + \delta_{\mathbf{m},\mathbf{0}}\frac{x}{1+t^2} + 1 + \frac{\log V}{(\sigma-\frac{1}{2})\log x}, \tag{10}$$

valid for $\sigma > \frac{1}{2} + \frac{4}{\log x}$. Applying a mean value result, Lemma 1 [3], we have

$$\begin{aligned}\sum_{\substack{\mathbf{m}\\ M_i<m_i\leqslant M_i+W}}\int_T^{T+W}&\left|\sum_{N\mathfrak{a}\leqslant x^2}\frac{\Lambda_x(\mathfrak{a})\chi\lambda^{\mathbf{m}}(\mathfrak{a})}{N\mathfrak{a}^{\sigma+it}}\right|^2 dt\\ &\ll (x^2+W^n)\sum_{N\mathfrak{a}\leqslant x^2}\left|\frac{\Lambda_x(\mathfrak{a})\chi\lambda^{\mathbf{M}}(\mathfrak{a})}{N\mathfrak{a}^{\sigma+iT}}\right|^2\\ &\ll W^n\sum_{\mathfrak{a}}\frac{\Lambda^2(\mathfrak{a})}{N\mathfrak{a}^{2\sigma}}\end{aligned}$$

on choosing $x^2 = W^{n/2}$, say. And

$$\begin{aligned}\sum_{\mathfrak{a}}\frac{\Lambda^2(\mathfrak{a})}{N\mathfrak{a}^{2\sigma}} < \sum_{\mathfrak{a}}\frac{\Lambda(\mathfrak{a})\log N\mathfrak{a}}{N\mathfrak{a}^{2\sigma}} &= \left\{\frac{d}{dz}\left(\frac{\zeta_K'}{\zeta_K}(z)\right)\right\}_{z=2\sigma}\\ &= \left\{\frac{1}{(z-1)^2}+O(1)\right\}_{z=2\sigma} \ll \frac{1}{(\sigma-\frac{1}{2})^2}.\end{aligned}$$

In this way we see that the bound in Theorem 7 arises from the last term in (10).

Proof of Theorem 5. Implicit in $\theta_l(z,\psi,l)$ are the characteristic functions

$$f_0(y) = \begin{cases}1 & |y|<l\\ 0 & l\leqslant|y|\leqslant\frac{1}{2}\end{cases},$$

extended to $\mathbb{R}$ by periodicity and

$$g_0(y) = \begin{cases}1 & x(1-l)<y<x(1+l)\\ 0 & \text{otherwise.}\end{cases}$$

These will be approximated by continuous functions $f_\pm$ and $g_\pm$ satisfying $0\leqslant f_-\leqslant f\leqslant f_+$ and $0\leqslant g_-\leqslant g\leqslant g_+$. In this way we obtain lower and upper bounds for $\theta_l(x,\psi,l)$. The same method was used in Section 4 of [1] though it was not made explicit there that f_- is always non-negative. As in [1] the approximations can be chosen to differ from f_0 on a set in $[-\frac{1}{2},\frac{1}{2}]$ of length $\ll\Delta$ and from g_0 on a set of length $\ll\Delta x$. We will take $\Delta = l^B$ with B to be chosen. This is a far smaller choice (at least when $l<1/\log x$) than in previous applications of this method. Further, the approximations to f_0 have a fourier series $\sum_{-\infty}^{\infty}a_m e^{-2\pi imy}$ where

$$a_m \ll \begin{cases}a_0 = 2l+O(\Delta)\\ 1/|m|,\ m\neq 0\end{cases} \tag{11}$$

and, given C, D,

$$a_m \ll l^C/|m|^D \quad \text{for } |m| > \Delta^{-1}\log^2 l,\ l < l_0(B,C,D). \tag{12}$$

Also the approximations to g_0 can be expressed as Mellin transforms $\hat{g}(s) = x^s\hat{h}(s)$ where

$$\begin{aligned} \hat{h}(1) &= 2l + O(\Delta) \\ \hat{h}(\sigma + it) &\ll \begin{cases} l \\ 1/|t|, \quad t \neq 0 \end{cases} \end{aligned} \tag{13}$$

and, given E, F,

$$\hat{h}(\sigma + it) \ll l^E/|t|^F, \quad \text{for } |t| > \Delta^{-1}\log^3 1/l,\ l < l_0(B,E,F). \tag{14}$$

So we now have $\theta_L \leqslant \theta_I(x,\psi,l) \leqslant \theta_U$ where both θ_L and θ_U are of the form

$$\theta(x,\psi,l) = \frac{1}{2\pi i h(\mathfrak{f})}\sum_{\chi}\bar{\chi}(I)\sum_{\mathbf{m}} a_{\mathbf{m}} e^{-2\pi i \mathbf{m}\psi}\int_{2-i\infty}^{2+i\infty}\hat{g}(s)\sum(s,\chi\lambda^{\mathbf{m}})ds \tag{15}$$

with $a_{\mathbf{m}} = \prod_{i=1}^{n-1} a_{m_i}$ and where

$$\begin{aligned} \sum(s,\chi\lambda^{\mathbf{m}}) &= \sum_{\mathfrak{p}}\frac{\chi\lambda^{\mathbf{m}}(\mathfrak{p})\log N\mathfrak{p}}{N\mathfrak{p}^s} \\ &= -\frac{L'}{L}(s,\chi\lambda^{\mathbf{m}}) - g_{\mathbf{m}\chi}(s). \end{aligned}$$

Here $g_{\mathbf{m}\chi}(s)$ is regular for $\sigma > \frac{1}{2}$ and in this half-plane $g_{\mathbf{m}\chi}(s) \ll (\sigma - 1/2)^{-1}$ uniformly in $\mathbf{m}$ and χ. Moving the line of integration in (15) to Re $s = \sigma, \frac{1}{2} < \sigma \leqslant \frac{3}{4}$, and writing $x = e^\tau$ gives

$$\begin{aligned} \frac{\theta(e^\tau,\psi,l) - a_0\hat{g}(1)/h(\mathfrak{f})}{e^{\sigma\tau}} &= \frac{1}{2\pi i}\int_{-\infty}^{\infty} e^{it\tau}\hat{h}(\sigma+it)\frac{1}{h(\mathfrak{f})} \\ &\quad \times \sum_{\chi}\bar{\chi}(I)\sum_{\mathbf{m}} a_{\mathbf{m}} e^{-2\pi i\mathbf{m}\psi}\sum(\sigma+it,\chi\lambda^{\mathbf{m}})dt. \end{aligned}$$

Then by Parseval's Theorem

$$\begin{aligned} &\int_0^\infty \left|\frac{\theta(e^\tau,\psi,l) - a_0\hat{g}(1)/h(\mathfrak{f})}{e^{\sigma\tau}}\right|^2 d\tau \\ &\ll \frac{1}{h^2(\mathfrak{f})}\int_{-\infty}^{\infty}|\hat{h}(\sigma+it)|^2\left|\sum_{\chi}\bar{\chi}(I)\sum_{\mathbf{m}} a_{\mathbf{m}} e^{-2\pi i\mathbf{m}\psi}\sum(\sigma+it,\chi\lambda^{\mathbf{m}})\right|^2 dt. \end{aligned}$$

Hence

$$\begin{aligned} &\sum_I\int_{\mathbf{T}^{n-1}}\int_1^\infty\left|\frac{\theta(x,\psi,l) - a_0\hat{g}(1)/h(\mathfrak{f})}{x^\sigma}\right|^2\frac{dx}{x}d\psi \\ &\ll \frac{1}{h(\mathfrak{f})}\int_{-\infty}^{\infty}|\hat{h}(\sigma+it)|^2\sum_{\chi}\sum_{\mathbf{m}}|a_{\mathbf{m}}|^2\left|\sum(\sigma+it,\chi\lambda^{\mathbf{m}})\right|^2 dt. \end{aligned} \tag{16}$$

From (12) and (14) we see that the integral and each summation over a co-ordinate of $\mathbf{m}$ can be cut-off at $\pm M$, where $M = \Delta^{-1}\log^3 1/l = l^{-B}\log^3 1/l$, with negligible error. This value of M is larger than in previous applications of this method so we decompose the remaining region into the union of

$$C(\mathbf{q}) = \{(t,\mathbf{m}) : q_0 W \leqslant t < (q_0+1)W,\ q_i W \leqslant m_i < (q_i+1)W,\ 1 \leqslant i \leqslant n-1\}$$

with $\mathbf{q} \in \mathbb{Z}^n$, $\|\mathbf{q}\| < M/W$ where $W = 1/l$. In each new region we use (11) and (13) obtaining

$$\begin{aligned}
&\sum_l \int_{\mathbf{T}^{n-1}} \int_1^\infty \frac{|\theta(x,\psi,l) - a_0\hat{g}(1)/h(\mathfrak{f})|^2}{x^{1+2\sigma}}\,dx\,d\psi \\
&\ll \sum_{\|\mathbf{q}\|<M/W} \frac{1}{W^{2n}} \prod_{i=0}^{n-1}\left(\frac{1}{1+q_i^2}\right)\frac{1}{h(\mathfrak{f})}\sum_\chi \\
&\quad \times \int \sum_{(t,\mathbf{m})\in C(\mathbf{q})} \left(\left|\frac{L'}{L}(\sigma+it,\chi\lambda^{\mathbf{m}})\right|^2 + |g_{\mathbf{m}\chi}(\sigma+it)|^2\right)dt \\
&\ll \sum_{\|\mathbf{q}\|<M/W} \frac{1}{W^{2n}} \prod_{i=0}^{n-1}\left(\frac{1}{q_i^2+1}\right)\left(\frac{W_n}{(\sigma-\frac{1}{2})^2}\frac{\log^2 M}{\log^2 W} + \frac{W^n}{(\sigma-\frac{1}{2})^2}\right)
\end{aligned} \tag{17}$$

by Theorem 7 and (16). We have now seen the need for the mean-value in (4) to be over intervals displaced from the origin. Fortunately this displacement effects the bound only as $\log^2 V$ so the result can be used if there is truncation at some M with $\log^2 M$ comparable to $\log^2 W$ as is the case. For then (17) is

$$\ll \frac{1}{W^n(\sigma-\frac{1}{2})^2}.$$

Choosing $\sigma = \frac{1}{2} + 8/\log W^n$ and noting that $x^{-16/\log W^n} \geqslant e^{-16A/n}$ for $1 \leqslant x \leqslant W^A$ gives

$$\int_{\mathbf{T}^{n-1}} \int_1^{l^{-A}} \frac{|\theta(x,\psi,l) - a_0\hat{g}(1)/h(\mathfrak{f})|^2}{x^2}\,dx\,d\psi \ll_A \frac{\log^2 W}{W^n}.$$

Here θ can be either θ_L or θ_U and so the required result follows on writing

$$\frac{a_0\hat{g}(1)}{h(\mathfrak{f})} = \frac{(2l)^n x}{h(\mathfrak{f})}\left(1 + O(l^B)\right)$$

where the error term is sufficiently small if $2B > A$ as we not assume. Thus we see the necessity of choosing Δ as small as we did.

4. Proof of Theorem 1

Given an ideal $\mathfrak{a}$ in $\mathbb{Q}(i)$ we can choose a generator α with argument in $(-\pi/4, \pi/4]$. The basis for the group of Groessencharaktere (with conductor (1)) consists of the single element $\lambda(\mathfrak{a}) = (\alpha/|\alpha|)^4$. So $\psi(\mathfrak{a}) = \dfrac{2}{\pi}\arg\alpha$.

Throughout we consider only Gaussian primes with argument in $(-\pi/4, \pi/4]$. Given such a prime ω associate with it the region

$$\mathscr{R}(\omega, \rho(\omega)) = \left\{ z : |z - \omega| < \frac{\rho(\omega)}{2}, |z| > |\omega| \right\}.$$

By the definition of ρ this region contains no Gaussian primes. We write

$$\mathscr{S}(y, \psi, l) \subseteq \mathscr{R}(\omega, \rho(\omega)) \tag{18}$$

if, for all $\mathfrak{a} \in \mathscr{S}(y, \psi, l)$, we can write $\mathfrak{a} = (\alpha)$ with $\alpha \in \mathscr{R}(\omega, \rho(\omega))$. We first show that, for an appropriate l, the inclusion (18) holds for all (y, ψ) in

$$I(\omega) := \{(y, \psi) : |y - \delta^2 N\omega| < lN\omega, \ |\psi - \psi(\omega)| < l\}$$

for some δ. It will be important that the $I(\omega)$ are disjoint and vol $I \gg$ vol $\mathscr{R}$. The difficulty is that the $\mathscr{R}(\omega, \rho(\omega))$ need not be pairwise disjoint.

As an intermediate step define

$$\mathscr{M}(\omega) = \{z : z = re^{2\pi i\theta}\omega, |r - \delta| < r(\omega)/500, 2\pi|\theta| < r(\omega)/500\}$$

where $r(\omega) = \rho(\omega)/|\omega|$ and $\delta = 1 + r(\omega)/10$. It is easily seen that $\mathscr{M}(\omega) \subseteq \subseteq \mathscr{R}(\omega, \rho(\omega))$.

Let ω' satisfy $|\omega'| \leqslant |\omega|, \omega' \neq \omega$, and compare $\mathscr{M}(\omega)$ with $R(\omega', \rho(\omega'))$.
(A) If $z \in R(\omega', \rho(\omega'))$ then $|z - \omega'| < \rho(\omega')/2$ while $|\omega - \omega'| \geqslant \rho(\omega')$. Hence the angle β between $z - \omega$ and $\omega' - \omega$ must satisfy $\sin \beta \leqslant \frac{1}{2}$, i.e. $0 \leqslant \beta \leqslant \pi/6$.
(B) If $z \in \mathscr{M}(\omega)$ then $z = re^{2\pi i\theta}\omega$ and $\omega' = se^{2\pi i\phi}\omega$ for some $r, s, |\theta| < r(\omega)/500$ and $-\frac{1}{2} < \phi \leqslant \frac{1}{2}$. Then

$$z - \omega = (re^{2\pi i\theta} - 1)\omega = r'e^{2\pi i\theta'}\omega,$$

$$\omega' - \omega = (se^{2\pi i\phi} - 1)\omega = s'e^{2\pi i\phi'}\omega$$

say, and the angle β is given by $2\pi|\theta' - \phi'|$. But $|\omega'| \leqslant |\omega| \Rightarrow |\omega'/\omega| \leqslant 1 \Rightarrow$ $\Rightarrow Re(\omega'/\omega - 1) \leqslant 0 \Rightarrow |\phi'| \geqslant 1/4$. So using $|\phi'| \leqslant |\theta' - \phi'| + |\theta'|$ gives

$$\beta \geqslant \frac{\pi}{2} - 2\pi|\theta'|.$$

Yet,

$$\begin{aligned} \text{Re } (re^{2\pi i\theta} - 1) &= r \cos 2\pi|\theta| - 1 \\ &\geqslant r(1 - 4|\theta|) - 1 \\ &\geqslant \frac{8}{100} r(\omega) \end{aligned}$$

and

$$\begin{aligned} \text{Im } (re^{2\pi i\theta} - 1)| &= r\,|\sin 2\pi\theta| \leqslant 2\pi r|\theta| \\ &\leqslant \frac{1}{400} r(\omega) \end{aligned}$$

by the definition of $\mathscr{M}(\omega)$. Hence $|\theta'| \leqslant |\tan \theta'| \leqslant 1/32$ and $\beta \geqslant \pi(\frac{1}{2} - \frac{1}{16})$.

From (A) and (B) we have that $\mathscr{M}(\omega)$ and $\mathscr{R}(\omega', \rho(\omega'))$ are disjoint. In particular the $\mathscr{M}(\omega)$ are pairwise disjoint.

Set $l = r(\omega)/2000$. Consider $\mathfrak{a} \in \mathscr{S}(y, \psi, l)$ with $(y, \psi) \in I(\omega)$. Then combining the conditions in $\mathscr{S}(y, \psi, l)$ and $I(\omega)$,

$$|N\mathfrak{a} - \delta^2 N\omega| \leqslant l(y + N\omega) \leqslant l(1 + \delta^2 + l)N\omega,$$
$$|\psi(\mathfrak{a} - \psi(\omega)| \leqslant 2l.$$

Choose α, where $\mathfrak{a} = (\alpha)$, with $\alpha = re^{2\pi i\theta}\omega$, $-1/8 < \theta \leqslant 1/8$. Then

$$2\pi|\theta| = \frac{2\pi}{4}|\psi(\mathfrak{a}) - \psi(\omega)| < 4l \leqslant \frac{r(\omega)}{500}$$

and

$$|r - \delta| = \frac{r^2 - \delta^2}{r + \delta} \leqslant \frac{l(1 + \delta^2 + l)}{\delta} < 4l \leqslant \frac{r(\omega)}{500}.$$

Hence $\alpha \in \mathscr{M}(\omega)$ and so $\mathscr{S}(y, \psi, l) \subseteq \mathscr{M}(\omega) \subseteq R(\omega, \rho(\omega))$.

Under the map

$$(y, \psi) \longmapsto y^{\frac{1}{2}} e^{2\pi i(\psi/4)} = (y^{\frac{1}{2}}/|\omega|)e^{2\pi i(\psi - \psi(\omega))/4}\omega$$

the image of $I(\omega)$ lies within $\mathscr{M}(\omega)$. The mapping is one-to-one, so the $I(\omega)$ are pairwise disjoint.

Let $1 \leqslant H < x$ and consider ω such that

$$\frac{\rho(\omega)}{|\omega|} > \left(\frac{H}{x}\right)^{1/2}.$$

Then we have seen that

$$\mathscr{S}\left(y, \psi, \frac{1}{2000}\left(\frac{H}{x}\right)^{1/2}\right) \subseteq \mathscr{S}\left(y, \psi, \frac{r(\omega)}{2000}\right) \subseteq \mathscr{M}(\omega) \subseteq \mathscr{R}(\omega, \rho(\omega))$$

for all $(y, \psi) \in I(\omega)$. And so $\theta(y, \psi, (H/x)^{1/2}/2000) = 0$ and

$$\iint_{I(\omega)} \left| \frac{\theta\left(y, \psi, \frac{1}{2000}\left(\frac{H}{x}\right)^{\frac{1}{2}}\right) - \left(\frac{2}{2000}\right)^2 \frac{H}{x} y}{y} \right|^2 dy\, d\psi$$

$$= \left(\frac{1}{1000}\right)^4 \left(\frac{H}{x}\right)^2 \iint_{I(\omega)} dy\, d\psi$$

$$= \left(\frac{1}{1000}\right)^4 \left(\frac{H}{x}\right)^2 4l^2 N\omega = \left(\frac{1}{1000}\right)^6 \left(\frac{H}{x}\right)^2 \rho^2(\omega).$$

Sum over $N\omega \leqslant x$ using the fact that the $I(\omega)$ are pairwise disjoint and then apply Theorem 5 (with $A = 8$, say), to obtain

$$\sum_{\substack{\rho(\omega) > (H/x)^{1/2}|\omega| \\ N\omega \leqslant x}} \rho^2(\omega) \ll \frac{x}{H} \log^2 x \tag{19}$$

valid for $H < x^{3/4}$. From (2) we see that $\rho(\omega) > (H/x)^{1/2}|\omega|$ can only possibly hold if $|\omega| < (x/H)\log^2 x$. For $H > x^{3/4}$ this is a stronger condition than $N\omega < x$ and we can again apply Theorem 5 (with $A = 5$, say) to obtain (19) with $x^{3/4} \leqslant H \leqslant x$. Integrating (19) over $1 \leqslant H \leqslant x$ gives the required result.

There are alternative ways to define ρ. For instance, given $-\frac{1}{2} \leqslant \theta < \frac{1}{2}$, $0 \leqslant \phi < \frac{1}{2}$, define

$$\rho(\omega) = \min\left\{|\omega' - \omega| : \left|\frac{1}{2\pi}\arg(\omega' - \omega) - \theta\right| \leqslant \phi, \omega' \neq \omega\right\}.$$

Then the same result (1) holds for this ρ, with just a change in the implied constant. The proof consists of defining a region similar to $\mathscr{M}(\omega)$ within the sector

$$\left\{z : \left|\frac{1}{2\pi}\arg(z - \omega) - \theta\right| \leqslant \phi,\ |z| \leqslant \rho(\omega)\right\}.$$

We can then deduce that given a direction (θ) almost all Gaussian primes have a close prime in that direction. The methods of this, and previous papers [1] and [2], say, are optimal for square-like regions $\mathscr{S}(x, \psi, l)$. But with this homogeneity in distribution we see that we may not necessarily derive better results if we could replace $\mathscr{S}(x, \psi, l)$ by more general shaped regions.

Proof of Lemma 2. To each prime $\omega, N\omega \leqslant x$, associate the region $\mathscr{M}(\omega)$ defined in the previous proof. These regions are distinct, have area $\gg \rho^2(\omega)$ and all lie within the disc, centre 0, radius $2x^{1/2}$ in $\mathbb{C}$. Hence the upper bound follows.

We demonstrate a slightly stronger lower bound by defining $\rho_0(\omega) = \min|\omega' - \omega|, \omega' \neq \omega$ and showing that $\sum_{N\omega \leqslant x} \rho_0^2(\omega) \gg x$. From [8], Satz 17,

$$|\{N\omega \leqslant x : \omega + \alpha \text{ prime }\}| < c_1 \prod_{\mathfrak{p}|(\alpha)} \left(1 + \frac{1}{N\mathfrak{p}}\right) \frac{x}{\log^2 x}$$

for some constant c_1, a simple consequence of Selbergs Sieve applied to the set $\{\xi(\xi + \alpha), \xi \in \mathbb{Q}(i)\}$. Hence

$$\sum_{N\alpha \leqslant c_2 \log x} |\{N\omega \leqslant x : \omega + \alpha \text{ prime }\}| < c_3 \frac{x}{\log x},$$

where $c_3 = 4c_1c_2\zeta_{\mathbb{Q}(i)}(2)\zeta_{\mathbb{Q}(i)}^{-1}(4)$. Choosing c_2 so that $c_3 < 1$, we find that $\rho_0^2(\omega) > c_2 \log x$ for $> (1 - c_3)x/\log x$ of the primes $N\omega \leqslant x$. Hence the lower bound follows.

For the proof of Theorem 3 we require

Lemma 8. *Let $G : [0, \infty] \to \mathbb{R}$ be a monotonically increasing function satisfying $G(x) \to \infty$ as $x \to \infty$. Then there exists $f : [0, \infty] \to \mathbb{R}$ satisfying $f(x) \to \infty$ and*

$$\frac{1}{f(x)} G\left(\frac{x}{f(x)}\right) \to \infty$$

as $x \to \infty$.

Proof. Since $G(x) \to \infty$ monotonically we can choose a sequence $\{x_n\}_{n \geq 1}$ such that $x_n \to \infty$ and $G(x) > n^2$ for all $x \geqslant x_n$. Define $f(x) = n$ for $nx_n \leqslant x < (n+1)x_{n+1}$. Then $f(x) \to \infty$ and, for $nx_n \leqslant x < (n+1)x_{n+1}$,

$$\frac{1}{f(x)} G\left(\frac{x}{f(x)}\right) = \frac{1}{n} G\left(\frac{x}{n}\right) > n$$

giving the required result.

Proof of Theorem 3. Assume the norm-form f_γ and a sequence $(\varepsilon_i)_{1 \leqslant i \leqslant r_1}, \varepsilon_i = \pm 1$ is given. It is explained in Section 5 of [2] how to construct integers $s, t \in \mathcal{O}_K$, integral ideals $\mathfrak{f}, \mathfrak{e}$, and classes $I \in I_\mathfrak{f}/P_{\mathfrak{f}\infty}$ such that if $\mathfrak{p} \in I$ then there exists $\lambda \in \mathcal{O}_K$ with $\mathfrak{p}\mathfrak{e} = (\lambda)$ and $t\lambda = s(\gamma + \mu)$ for some $\mu \in M$, where sign $(\gamma^{(i)} + \mu^{(i)}) = \varepsilon_i, 1 \leqslant i \leqslant r_1$. In particular $\mu = \alpha(\mathbf{m})$ for some $\mathbf{m} \in \mathbb{Z}^n$ and $|f_\gamma(\mathbf{m}) = |N\mathfrak{p}|$.

In Section 4 of [2] an explicit construction of ψ and the corresponding Groessencharakter λ is given. This goes back to HECKE [6]. But we find that

$$\psi : \mathcal{O}_K \to \mathbb{R}^r \times [0,1]^{r_2}$$

such that if $\alpha \in \mathcal{O}_K$, there exists a unique unit $u \in \mathcal{U}(\mathfrak{f})$ for which $\psi(u\alpha) \in [0,1]^{n-1}$. Given $\mathbf{x} \in \mathbb{R}^n$ define $\psi(\mathbf{x}) = \psi(s(\gamma + \alpha(\mathbf{x}))/t) - \psi(\mathfrak{e})$. We do not want $\psi(\mathbf{x}) \equiv \psi(\mathbf{x}')$ mod 1 for any $\mathbf{x} \neq \mathbf{x}'$ so we partition $\mathbb{R}^n$ by decomposing the image of ψ.

Consider $\mathcal{B}(N) = \{\mathbf{x} \in \mathbb{R}^n, N/2 \leqslant \|\mathbf{x}\|_2 \leqslant N\}$. From Section 6 of [2] we have that given $0 < \delta < 1$ there exists $\mathcal{R}(\delta) \subseteq \mathcal{B}(N)$ (depending on the norm-form f_γ) with vol $\{\mathcal{R}(\delta) \cap \mathcal{B}(N)\} \leqslant c\delta$ vol $\mathcal{B}(N)$ for some c independent of δ, such that if $\mathbf{x} \notin \mathcal{R}(\delta)$ then

$$\delta \|\mathbf{x}\|_2 \leqslant |\gamma^{(i)} + \alpha^{(i)}(\mathbf{x})| \ll \|\mathbf{x}\|_2. \tag{20}$$

Then, for $\mathbf{x} \in \mathcal{B}(N) \backslash \mathcal{R}(\delta)$ we have $a\delta^n N^n \leqslant |f_\gamma(\mathbf{x})| \leqslant bN^n$ for some $a, b > 0$. And, as is seen in Section 7 of [2], the image of ψ lies in $B \times [0,1]^{r_2}$, $B \subseteq \mathbb{R}^r$ a box of side length $\ll \log \delta^{-1}$. By splitting the interval for $|f_\gamma(\mathbf{x})|$ and the box B into boxes B_i of side length 1, we decompose $\mathcal{B}(N) \backslash \mathcal{R}(\delta)$ into $\ll (\log \delta^{-1})^{r+1}$ sets each contained in some

$$C = \{\mathbf{x} \in \mathbb{R}^n,\ X/2 \leqslant |f_\gamma(\mathbf{x})| < X,\ \text{sign}\ (\gamma^{(j)} + \alpha^{(j)}(\mathbf{x})) = \varepsilon_j,$$
$$1 \leqslant j \leqslant r_1, \psi(\mathbf{x}) \in B_i \times [0,1]^{r_2}\}$$

where $\delta^n N^n \ll X \ll N^n$.

Given Φ define Φ_2 and Δ by

$$\Phi_2^{(n/4)+1}(y^n) = \Phi(y/2) \quad \text{and} \quad \Delta(y) = \Phi_2^{n/4}(y).$$

We then further subdivide

$$C = \bigcup_\kappa C_\kappa = \bigcup_\kappa \{X_\kappa \leqslant |f_\gamma(\mathbf{x})| < X_\kappa + X/2\Delta,\ \mathbf{x} \in C\}$$

with $\Delta = \Delta(X)$ and $X_\kappa = X(1 + \kappa/\Delta)/2$. Note that with a as above, $G(y) = \Delta(ay^n)$ is monotonically increasing so we can apply Lemma 8 to find f satisfying $f(y) \to \infty$ and $f^{-1}(y)G(y/f(y)) = f^{-1}(y)\Delta(a(y/f(y))^n) \to \infty$. Then we choose $\delta = \delta(N) = f^{-1}(N)$ so that $\delta(N) \to 0$ and $\delta(N)\Delta(a\delta^n(N)N^n) \to \infty$ as $N \to \infty$.

Define $D_\kappa = (\mathscr{B}(N)\backslash\mathscr{R}(\delta)) \cap C_\kappa$ and $\theta(\mathbf{x}, l) = \theta(x, \psi, l)$ where $x = |f(\mathbf{x})|$ and $\psi = \psi(\mathbf{x})$. Then

$$\int_{D_\kappa} \left| \frac{\theta(\mathbf{x}, l) - (2l)^n x/h(\mathfrak{f})}{x} \right|^2 d\mathbf{x} \ll \int_{X_\kappa}^{X_\kappa + X/\Delta} \int_{\mathbf{T}^{n-1}} \left| \frac{\theta(x, \psi, l) - (2l)^n x/h(\mathfrak{f})}{x} \right|^2 dx\, d\psi$$

(the Jacobian is a constant depending on the form f_γ)

$$\ll l^n \log^2(1/l)$$

by Theorem 5. In which case

$$\left| \theta(\mathbf{x}, l) - \frac{(2l)^n x}{h(\mathfrak{f})} \right| \ll \frac{X l^{n/2} \log(1/l)}{|D_\kappa|^{1/2}} \left(\frac{|D_\kappa|}{|C_\kappa|} \Delta(X) \right)^{\frac{1}{2}}$$

excepting a set of size

$$\ll |D_\kappa| / \left(\frac{|D_\kappa|}{|C_\kappa|} \Delta(X) \right) = \frac{|C_\kappa|}{\Delta(X)}.$$

But $|C_\kappa| \asymp X/\Delta(X)$, so

$$\left| \theta(\mathbf{x}, l) - \frac{(2l)^n x}{h(\mathfrak{f})} \right| \ll X^{1/2} l^{n/2} \Delta(X) \log(1/l) \tag{21}$$

for all $\mathbf{x} \in C_\kappa$ excepting a set of size $\ll X/\Delta^2(X)$. We now choose $l(y) = y^{-1/n} \Phi_2(y) \log^{2/n} y$. The condition $\Phi(x) < x$ is sufficient to deduce that $l(y)$ is decreasing. Then, if $\mathbf{x}$ satisfies (21),

$$\begin{aligned} \theta(\mathbf{x}, l(x)) - \frac{(2l(x))^n x}{h(\mathfrak{f})} & \\ \leqslant \theta(\mathbf{x}, l(X_\kappa)) - \frac{(2l(X_\kappa))^n}{h(\mathfrak{f})} x + \frac{2^n}{h(\mathfrak{f})} & (l(X_\kappa)^n x - l(x)^n x) \\ \leqslant C \left\{ X^{1/2} l^{n/2}(X_\kappa) \Delta(X) \log(1/l) + \left(\frac{x}{X_\kappa} - 1 \right) \Phi_2^n(x) \log^2 x \right\} & \end{aligned}$$

having used the definition of l and that Φ_2 is an increasing function. But $x/X_\kappa - 1 \leqslant 1/\Delta(X)$ and so

$$\theta(\mathbf{x}, l(x)) - \frac{(2l(x))^n x}{h(\mathfrak{f})} \leqslant C \Delta^3(X) \log^2 x$$

for some $C > 0$. Similarly

$$\begin{aligned} \theta(\mathbf{x}, l(x)) - \frac{(2l(x))^n x}{h(\mathfrak{f})} & \\ \geqslant \theta(\mathbf{x}, l(X_{\kappa+1})) - \frac{(2l(X_{\kappa+1}))^n}{h(\mathfrak{f})} x + \frac{2^n}{h(\mathfrak{f})} & (l(X_{\kappa+1})^n x - l(x)^n x) \\ \geqslant - C \left\{ X^{1/2} l^{n/2}(X_{\kappa+1}) \Delta(X) \log(1/l) + \left(\frac{x}{X_{\kappa+1}} - 1 \right) \Phi_2^n(x) \log^2 x \right\} & \\ \geqslant - C \Delta^3(X) \log^2 x & \end{aligned}$$

for some $C > 0$. Hence

$$|\theta(\mathbf{x}, l(x)) - (2l(x))^n x/h(\mathfrak{f})| \ll \Delta^3(X)\log^2 x \tag{22}$$

for those $\mathbf{x} \in C_\kappa$ satisfying (21). Taking the union of the C_κ, X and B_i we have that (22) holds for all $N/2 \leqslant \|\mathbf{x}\|_2 \leqslant N$ excepting a set of size

$$\ll \delta N^n + \sum_{B_i}\sum_{X}\sum_{\kappa}\frac{X}{\Delta^2(X)} \ll \left(\delta + \frac{(\log \delta^{-1})^{r+1}}{\Delta(a\delta^n N^n)}\right)N^n$$

$$\ll \left(\delta + \frac{(\log \delta^{-1})^{r+1}}{\delta^{-1}}\frac{1}{\delta\Delta(a\delta^n N^n)}\right)N^n = o(N^n).$$

So for almost all $\mathbf{x} \in \mathscr{B}(N)$, $\theta(\mathbf{x}, l(x)) > 0$. By Section 4 of [2] there then exists $\mathbf{m}$ for which $|f_\gamma(\mathbf{m})|$ is prime and

$$\left|\sum_{i=1}^{n}(m_i - x_i)\alpha_i^{(j)}\right| \ll |\gamma^{(j)} + \alpha^{(j)}(\mathbf{x})|l(x), \quad 1 \leqslant j \leqslant n,$$

$$\ll \|\mathbf{x}\|_2 l(x) \tag{23}$$

by (20) since $\mathbf{x} \notin \mathscr{R}(\delta)$. For the same reason $\delta^n\|\mathbf{x}\|_2^n \leqslant x \ll \|\mathbf{x}\|_2^n$ and, because $(\alpha_i^{(j)})_{1\leqslant i,j\leqslant n}$ is invertible,

$$\|\mathbf{m} - \mathbf{x}\|_2 \ll \|\mathbf{x}\|_2 l(x) = \|\mathbf{x}\|_2 x^{-1/n}\Phi_2(x)\log^{2/n} x$$

$$\ll \delta^{-1}(N)\Phi_2(N^n)\log^{2/n} N$$

$$\leqslant \Delta(N^n)\Phi_2(N^n)\log^{2/n} N$$

(by definition $\delta^{-1}(N) \leqslant \Delta(N^n)$)

$$\ll \Phi(N/2)\log^{2/n} N$$

$$\ll \Phi(\|\mathbf{x}\|_2)\log^{2/n}\|\mathbf{x}\|_2,$$

as required.

Acknowledgements. I would like to thank the referee for his/her comments and in particular for the simplified proof of Lemma 8 that I have presented here.

References

[1] Coleman MD (1990) The distribution of points at which binary quadratic forms are prime. Proc London Math Soc (3) **61**: 433–456

[2] Coleman MD (1992) The distribution of points at which norm-forms are prime. J Number Theory **41**: 359–378

[3] Coleman MD (1996) Relative norms of prime ideals in small regions. Mathematika **43**: 40–62

[4] Davenport H (1980) Multiplicative Number Theory, 2nd edn. Berlin Heidelberg New York: Springer

[5] Duke W (1989) Some problems in multidimensional analytic number theory. Acta Arith **52**: 203–228

[6] Hecke E (1920) Eine neue Art von Zetafunktionen und ihre Beziehung zur Verteilung der Primzahlen I, II. Math Z **1**: 357–376; **6**: 11–51

[7] Odoni RWK (1979) The distribution of integral and prime-integral values of systems of full-norm polynomials and affine-decomposable polynomials. Mathematika **26**: 80–87

[8] RIEGER GJ (1961) Verallgemeinerung der Siebmethode von A. Selberg auf algebraische Zahlkörper. III. J Reine Angew. Math **213**: 79–90

[9] SELBERG A (1989) On the normal density of primes in small intervals and the difference between consecutive primes. Archiv for Mathematik og Naturvidenskab B. **47**: 6, 87–105 (1943); Also in: CHANDRASEKHRAN K (ed) Collected Papers, Vol. 1. pp 160–178. Berlin: Springer

[10] SELBERG A (1989) Contributions to the theory of Dirichlet's L-functions. Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo I Mat-Naturv. Klasse (1946) No. 3, 1–62; Also In: CHANDERSEKHRAN K (ed) Collected Papers, Vol. 1. pp 281–340. Berlin: Springer

MD COLEMAN
Department of Mathematics
UMIST
P.O. Box 88
Manchester M60 1QD
UK

Mh. Math. 125, 127–145 (1998)

Monatshefte für
Mathematik

Printed in Austria

Highly Saturated Packings and Reduced Coverings

By

G. Fejes Tóth, Budapest, **G. Kuperberg,** Davis, CA,
and **W. Kuperberg,** Auburn, AL

With 6 Figures

(Received 19 February 1996)

Abstract. We introduce and study certain notions which might serve as substitutes for maximum density packings and minimum density coverings. A body is a compact connected set which is the closure of its interior. A packing $\mathscr{P}$ with congruent replicas of a body K is n-saturated if no $n-1$ members of it can be replaced with n replicas of K, and it is completely saturated if it is n-saturated for each $n \geqslant 1$. Similarly, a covering $\mathscr{C}$ with congruent replicas of a body K is n-reduced if no n members of it can be replaced by $n-1$ replicas of K without uncovering a portion of the space, and it is completely reduced if it is n-reduced for each $n \geqslant 1$. We prove that every body K in d-dimensional Euclidean or hyperbolic space admits both an n-saturated packing and an n-reduced covering with replicas of K. Under some assumptions on $K \subset \mathbb{E}^d$ (somewhat weaker than convexity), we prove the existence of completely saturated packings and completely reduced coverings, but in general, the problem of existence of completely saturated packings and completely reduced coverings remains unsolved. Also, we investigate some problems related to the the densities of n-saturated packings and n-reduced coverings. Among other things, we prove that there exists an upper bound for the density of a $d+2$-reduced covering of $\mathbb{E}^d$ with congruent balls, and we produce some density bounds for the n-saturated packings and n-reduced coverings of the plane with congruent circles.

1. Introduction and Preliminaries

Two of the basic problems in the theory of packing and covering are to determine the most efficient packing and covering with replicas of (meaning sets congruent to) a given set K in some metric space. Recall that a packing is a family of sets whose interiors are mutually disjoint, and that a covering is a family of sets whose union is the whole space. By a space we mean either d-dimensional Euclidean space $\mathbb{E}^d$ or d-dimensional hyperbolic space $\mathbb{H}^d$, although the definitions that follow are sometimes more general. We shall consider packings and coverings with replicas of a nonempty compact connected set which is the closure of its interior, a *body*, for short.

The usual measure of the efficiency of an arrangement in Euclidean space is density. Roughly speaking, the density of an arrangement is the total volume of the members of the arrangement divided by the volume of the whole space. Rigorously, density can be defined by an appropriate limit [12]. The maximum

1991 Mathematics Subject Classification: 52C17
Key words: packing, covering, saturated, reduced, convex, body

density of a packing of the space with replicas of a (measurable) set K is denoted by $\delta(K)$ and is called the *packing density of* K. The minimum density of a covering with replicas of K is denoted by $\vartheta(K)$ and is called the *covering density of* K. It is known that each of the maximum and the minimum density is attained [13].

There are some disadvantages of using density for measuring the efficiency of an arrangement. In the first place, optimum density is a global notion and it does not imply the local efficiency of an arrangement. Secondly, the notion of density cannot be extended in full generality to hyperbolic geometry [1], [12]. In what follows we introduce and study certain notions which might serve as substitutes for maximum density packings and minimum density coverings.

Let K be a body and let $\mathscr{P}$ be a packing of space with replicas of K. $\mathscr{P}$ is said to be *saturated* if it cannot be augmented with any additional replica of K without overlapping with a member of $\mathscr{P}$. More generally, $\mathscr{P}$ is *n-saturated* if no $n-1$ members of it can be replaced with n replicas of K. A packing is *completely saturated* if it is n-saturated for every $n \geqslant 1$.

Note that FEJES TÓTH and HEPPES [11] define the term "n-saturated" differently, but we hope that our definition causes no confusion.

A covering $\mathscr{C}$ of E^d with replicas of K is *reduced* if no proper sub-family of $\mathscr{C}$ is a covering. Similarly, we say that $\mathscr{C}$ is *n-reduced* if no n members of it can be replaced by $n-1$ replicas of K without uncovering a portion of the space. A covering is *completely reduced* if it is n-reduced for every $n \geqslant 1$.

Conjecture. *Every body K in $\mathbb{E}^d$ (resp. in $\mathbb{H}^d$) admits a completely saturated packing and a completely reduced covering with replicas of K.*

This conjecture is supported by the following results:

Theorem 1.1. *Every convex body in $\mathbb{E}^d$ admits a completely saturated packing and a completely reduced covering of $\mathbb{E}^d$ with replicas of the body.*

Theorem 1.2. *Every body K in $\mathbb{E}^d$ (resp. in $\mathbb{H}^d$) admits both an n-saturated packing and an n-reduced covering with replicas of K.*

Section 2 presents a proof of Theorem 1.1. We note there that the theorem holds for bodies satisfying the strict nested similarity property, a condition weaker than convexity. Theorem 1.2 is proved in Section 3 (for the Euclidean case) and Section 4 (for the hyperbolic case), each as a corollary of a more general statement. The hyperbolic case involves some elements of the theory of hyperbolic manifolds, which we review for the benefit of the unfamiliar reader.

2. Complete Saturation and Reduction

In this section we give a proof of Theorem 1.1. We precede the proof with some definitions and two lemmas. Throughout the argument, K is a given convex body in $\mathbb{E}^d$, and, as before, $V(A)$ denotes the volume of A. We use the Hausdorff distance between closed sets to measure the distance between a pair of (finite)

packings or coverings, extending the Hausdorff distance function to the space of finite (unordered) collections of compact sets in the natural way.

Let c be a "center" point in the interior of K, say the center of gravity of K. Let $B(r, p)$ be the sphere of radius r centered at p. A packing with replicas of K is *completely saturated in* $B(r, p)$ if no n replicas contained in $B(r, P)$ can be replaced by $n+1$ replicas contained in $B(r, p)$ for every integer n (the replicas not contained in $B(r, p)$ are not to be moved in this process). A packing is *unsaturated in* $B(r, p)$ for short if it fails to be completely saturated in $B(r, p)$. An arbitrary packing of $\mathbb{E}^d$ can be altered within $B(r, p)$ so as to result in a packing completely saturated in the ball: delete all replicas of K contained in the ball and replace them with the maximum number of replicas of K that will fit in the ball without overlapping with each other or with any of the replicas that partially invade the ball.

A *homothetic thinning* $T_h(\mathscr{P})$ of a packing $\mathscr{P}$ with replicas of K by a factor of $h > 1$ is a new packing such that each center $c \in \mathbb{E}^d$ maps to hc, but such that the replicas of K are translated without expansion. The analogous concepts for coverings (*completely reduced in* $B(r, p)$, *unreduced in* $B(r, p)$, *homothetic thickening* by a factor of $h < 1$) are defined similarly. The proof of Theorem 1.1 is given only for packings, since the proof for coverings is the same except for one modification which is mentioned afterwards.

Lemma 2.1. *For every r and $\varepsilon > 0$, there exists a $\delta > 0$ such that if a $\mathscr{P}$ is less than δ away (in Hausdorff distance) from a packing which is unsaturated in $B(r, 0)$ then $T_{1+\varepsilon}(\mathscr{P})$ is unsaturated in $B((1+\varepsilon)r, 0)$.*

The proof is left to the reader.

Lemma 2.2. *Let $r > 0$ and $\eta > 0$. Then there exists an s_0 and a $\delta > 0$ such that for every $s > s_0$, a packing $\mathscr{P}$ of replicas of K which is densest relative to $B(2s + r, 0)$ has the following property: If p is chosen at random in $B(s, 0)$, $\mathscr{P}$ is at least δ away from unsaturated in $B(r, p)$ with probability at least $1 - \eta$.*

Proof. Informally, if ε is sufficiently small and s is sufficiently large, then if $\mathscr{P}$ is expanded by $1 + \varepsilon$, the loss of density from replicas of K sliding over the edge of $B(2sR + r, 0)$ is outweighed by the gain in finding a η proportion of $B(r, p)$'s inside that are unsaturated and re-saturating the packing in a disjoint collection of these smaller balls. Then δ can be chosen based on ε and Lemma 2.1. A more precise argument follows.

Temporarily fix $\delta > 0$ and $s > 0$, and suppose that, to the contrary, the set X of points $p \in B(s, 0)$ such that $\mathscr{P}$ is less than δ away from unsaturated in $B(r, p)$ has measure at least $\eta V(B(s, 0))$. We will arrive at a contradiction for δ sufficiently small and s sufficiently large.

Since X has measure $\eta V(B(s, 0))$, it cannot be covered by fewer than

$$\eta \frac{V(B(s, 0))}{V(B(2r, 0))} = \frac{\eta}{(2r)^d} s^d$$

balls of radius $2r$. It follows that there exists a packing $\{B(r, p_i)\}_{1 \leqslant i \leqslant k}$ of k balls of radius r entirely within $B(s + r, 0)$ such that the restriction of $\mathscr{P}$ to each ball is

less than δ away from an unsaturated packing, where

$$k > \eta \frac{V(B(s, 0))}{V(B(2r, 0))} > cs^d$$

for some constant c depending only on r and η. In other words,

$$\frac{kV(K)}{V(B(s, 0))} > \frac{cV(K)}{V(B(1, 0))}.$$

Let

$$C = \frac{cV(K)}{V(B(1, 0))} < 1,$$

let

$$\varepsilon = \min\left((1 - C)^{-\frac{1}{d}} - 1, \frac{1}{2}\right),$$

and let δ be given by Lemma 2.1.

Observe that the difference between the density of $\mathscr{P}$ and that of the homothetic thinning $T_{1+\varepsilon}(\mathscr{P})$ (both relative to $B(s, 0)$) is at most $1 - (1 - \varepsilon)^{-d}$. On the other hand, since $T_{1+\varepsilon}(\mathscr{P})$ is unsaturated in each of the balls $B((1 + \varepsilon)r, (1 + \varepsilon)p_i)$, the density of $T_{1+\varepsilon}(\mathscr{P})$ relative to $B(s, 0)$ can be increased through saturation by an amount greater than

$$k(s) V(K)/V(B(s, 0) > cV(K)/V(B(1, 0)) = C.$$

(Note that by our choice of ε, each $B((1 + \varepsilon)r, (1 + \varepsilon)p_i)$ is contained in $B(2s + r, 0)$.) By our choice of ε, we have $1 - (1 - \varepsilon)^{-d} < C$. Thus, the thinned and then re-saturated packing is denser in $B(2s + r, 0)$ than the original packing $\mathscr{P}$, which is a contradiction.

Proof of Theorem. Let $n > 0$ be an integer and for each $1 \leqslant k \leqslant n$, let δ_k and s_k be given by the second lemma with $r = k$ and $\eta = 3^{-k}$. Let s be the supremum of the s_k's, and let $\mathscr{P}$ be a packing which is densest relative to $B(2s + n, 0)$. Then for at least

$$1 - \left(\frac{1}{3} + \frac{1}{9} + \cdots + \frac{1}{3^n}\right) > 1/2$$

of $p \in B(s, 0)$, $\mathscr{P}$ is δ_k away from unsaturated in all $B(k, p)$'s simultaneously. After translation by $-p$, $\mathscr{P}$ becomes a packing $\mathscr{P}_n$ which is simultaneously δ_k away from unsaturated in each $B(k, 0)$. The sequence $\mathscr{P}_n$ has a subsequence which converges in the Hausdorff topology to a limit $\bar{\mathscr{P}}$. The packing $\bar{\mathscr{P}}$ is δ_n away from unsaturated in $B(n, 0)$ for every $n > 0$, and it is therefore completely saturated in $\mathbb{E}^d$.

Remark 1. The argument for the analogous theorem for coverings requires two minor modifications in the formulation and proof of Lemma 2.2. Firstly, instead of using a covering which extremizes density in $B(2s + r, 0)$, we use a minimum

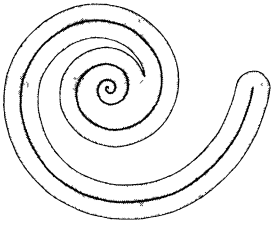


Figure 1. A body with the strict nested similarity property

cardinality arrangement of replicas of K that covers $B(2s + r, 0)$. In particular, an optimal covering of $B(2s + r, 0)$ in this sense has no replicas disjoint from $B(2s + r, 0)$. Secondly, a homothetic thickening of a covering of $B(2s + r, 0)$ by a factor of $1 - \varepsilon$ is not in general again a covering. To repair it, we identify a cube $\kappa \subset K$ and we cover the annular region $B(2s + r, 0) - B((1 - \varepsilon)(2s + r), 0)$ by non-overlapping translates of κ. The resulting gain in number of replicas due to homothetic thickening is comparable to the loss of density due to homothetic thinning.

Remark 2. The hypothesis that K is convex can be weakened somewhat without any changes in the proof. It suffices that K have the *strict nested similarity property*, which requires that, for every positive number $h < 1$, the interior of K contains a replica of hK. For example, if K is *strictly starlike*, i.e., K contains an interior point such that every ray emanating from it meets the boundary of K at a single point, then K has the strict nested similarity property. Another example of a body with the strict nested similarity property is an ε-neighborhood of a logarithmic spiral, as shown in Figure 1.

3. Lattices of Isometries

We recall a construction from the topological theory of covering spaces, to be used in this section and in the next one, which we apply to produce n-saturated packings and n-reduced coverings in various spaces. For the basic notions and facts related to that theory we refer the reader to [22, Ch. 1–2].

Let M be a locally compact, connected metric space (for our purposes it is sufficient to assume that M is a non-compact Riemannian manifold) with the metric ϱ. A group G of self-isometries of M is a *fixed-point-free, uniform lattice of isometries*, or *lattice of isometries* or *lattice* for short, if G satisfies the following two conditions:

(i) there is a number $\gamma > 0$ such that for every $x \in M$ and every $g \in G$ other than the identity, $\varrho(x, g(x)) \geqslant \gamma$;

(ii) the quotient space $B = M/G$ (whose points are the orbits $\{g(x) : g \in G\}$ of points of M under G, furnished with the quotient topology) is compact.

Every lattice G of isometries of M has the crucial property that the quotient map p from M to B, assigning to each point to M its orbit, is a covering projection (see [22, p. 88]). We shall refer to M as the covering space and to B as the base

space. Also, the covering is regular, for G acts transitively on each point-inverse. The base space is metrizable: a specific metric for B is defined by $\varrho(x, y) = \inf\{\varrho(\tilde{x}, \tilde{y}) : p(\tilde{x}) = x, p(\tilde{y}) = y\}$. Under this metric, the covering projection p is a local isometry: the restriction of p to any set of diameter smaller than γ is an isometry.

Conversely, given a regular covering $p : M \to B$, where B is a compact manifold, M is connected and endowed with the metric lifted from B, then the group of covering transformations of M is a lattice of isometries. Moreover, if M is simply connected, then the lattice of isometries of M is isomorphic to the fundamental group $\pi_1(B)$ of the base space B (see [22, Sec. 2.6]).

If M is d-dimensional Euclidean space, then a lattice of isometries which consists of translations reduces to the classical concept of a lattice of vectors, and the resulting base space is a d-dimensional torus. However, there are other lattices of isometries even in $\mathbb{E}^2$. For instance, the group of isometries of the Cartesian plane generated by a translation in the x direction and a glide-reflection in the y direction is a lattice of isometries, but the resulting base space is a Klein bottle and not a torus.

Define the *girth* of a lattice G as the infimum of the distances $\varrho(x, g(x))$ over all non-identity elements $g \in G$. If M is a simply-connected Riemannian manifold (such as $\mathbb{H}^d$), then the same number is the infimum of the lengths of all non-trivial loops in the base space, hence the name "girth". For many manifolds, including Euclidean and hyperbolic manifolds, the girth is twice the *injectivity radius* of the quotient manifold, which is defined as the largest r such that no metric ball of radius r overlaps itself.

Observe that if S is a subset of M whose diameter is smaller than the girth of G, then the image $p(S)$ is a replica of S in B. Also, the set $p^{-1}(p(S))$ is the union of a discrete collection of mutually disjoint replicas of S, namely it is the orbit of S under G. We call this discrete collection of replicas of S *a lifting* of $p(S)$ (in M).

Theorem 3.1. *Let M be a locally compact connected metric space and let K be a body in M. If for every $c > 0$, M admits a lattice of girth greater than c, then there exist an n-saturated packing of M and an n-reduced covering of M with replicas of K, for every positive integer n.*

Proof. We restrict our attention to packings only, since the case of coverings is completely analogous.

Let G be a lattice of isometries of M of girth greater than $2n + 1$ times the diameter of K, let B be the base space associated with the lattice, and let $p : M \to B$ be the covering projection. In the base space B, arrange a packing with a maximum number of bodies of the form $p(K')$, where K' is a replica of K in M. The maximum is finite because B is compact and K has a non-empty interior. Let $\mathscr{P}$ be the lifting (in M) of this packing. Obviously, $\mathscr{P}$ is a packing of M with replicas of K. We assert that $\mathscr{P}$ is n-saturated.

Suppose the contrary, and let $m \leqslant n$ be the smallest positive integer such that $\mathscr{P}$ is not m-saturated. Therefore there are m members of $\mathscr{P}$, say $K_1, K_2, \ldots, K_m$ which can be replaced by $m + 1$ other replicas of K, say $L_1, L_2, \ldots, L_{m+1}$, and m is the smallest integer with this property. By the "pigenhole principle," the set

$S = (\bigcup K_i) \cup (\bigcup L_j)$ is connected. Therefore the diameter of S is smaller than the sum of the diameters of the K_i's and the L_j's, thus smaller than the girth of G. If we replace in B the sets $p(K_i)$ $(i = 1, 2, \ldots, m)$ with the sets $p(L_j)$ $(j = 1, 2, \ldots, m+1)$, we exceed the maximum number defined above. This is a contradiction. □

The Euclidean case of Theorem 1.2 is an immediate corollary of Theorem 3.1. Also, the remaining part of Theorem 1.2 is now reduced to the problem of existence of lattices of arbitrarily large girth in d-dimensional hyperbolic space. This problem is addressed in the next section.

4. Hyperbolic Lattices of Large Girth

A lattice of isometries of hyperbolic space $\mathbb{H}^d$ will be called a (d-dimensional) *hyperbolic lattice* for short. The aim of this section is to prove the following:

Theorem 4.1. *For every c, there exists a d-dimensional hyperbolic lattice of girth greater than c.*

Although this fact and the methods used for proving it have been known for a long time, we could not find a suitable reference and we include a proof for completeness.

We begin with some algebraic preliminaries. A group G is *residually finite* if for every $g \in G$ other than the identity e, there exists a normal subgroup N of finite index which does not contain g. Equivalently, G is residually finite if for every $g \in G$, $g \neq e$, there is a homomorphism φ from G to a finite group such that $\varphi(g)$ is not the identity.

Since the intersection of two normal, finite-index subgroups of G is again a normal subgroup of finite index, we get immediately:

Proposition 4.2. *If G is residually finite, then for any finite set $F \subset G$ not containing the identity there exists a normal subgroup $N \subset G$ of finite index which does not intersect F.*

The group of non-singular $n \times n$ matrices with real coefficients is denoted by $GL(n, \mathbb{R})$ and $I(\mathbb{H}^d)$ denotes the group of isometries of $\mathbb{H}^d$. Since $\mathbb{H}^d$ can be modelled as one sheet of a two-sheeted hyperboloid in $\mathbb{R}^{d+1}$, where the isometries of $\mathbb{H}^d$ are those linear transformations of $\mathbb{R}^{d+1}$ which preserve the sheet (see [5, Sec. A2]), the group $I(\mathbb{H}^d)$ is isomorphic to a subgroup of $GL(d+1, \mathbb{R})$.

The following lemma is a direct consequence of a theorem of MAL'CEV ([16, Th. VII]). We include a version of MAL'CEV's proof.

Lemma 4.3. *Every finitely generated subgroup G of $GL(n, \mathbb{R})$ is residually finite.*

Proof. Let g be a non-identity element of G. The aim of the proof is to construct a homomorphism from G to a finite group such that the image of g is also not the identity. This is accomplished by the composition of three homomorphisms. The first one, α, sends G into the group of algebraic matrices (matrices

whose entries are algebraic numbers) of the same size as the matrices in G; the second one, β, is a map to a group of (larger) rational matrices; and the third one, γ, is a map to a group of matrices over a finite field $\mathbb{Z}/p$.

Let $g_1, g_2, \ldots, g_k$ be a set of generators for G. If a set of algebraic equations in finitely many variables has a real solution, then it has (possibly complex) algebraic solutions arbitrarily close to the real solution. The defining relations between the g_i's impose some constraints on the entries of these matrices. Since the constraints are algebraic, algebraic matrices can be found that satisfy the same relations as the g_i's do, and are arbitrarily close to them. We pick algebraic matrices for the images under α of the generators that are close enough to the original real matrices so that g's image $\alpha(g)$ is not the identity.

The coefficients of all of the $\alpha(g_i)$'s are algebraic numbers that, all together, lie in some field F which is a finite-dimensional vector space over $\mathbb{Q}$. The algebraic numbers can then themselves be understood as rational linear transformations of F. Therefore, possibly by passing to larger matrices, we can assign to each matrix $\alpha(g_i)$ a larger matrix with rational entries, and this assignment extends to a monomorphism β. Thus, $\beta\alpha(h)$ is a rational matrix assigned to h for every $h \in G$.

To define the third and last homomorphism, let p be a prime which does not divide the denominator of any $\beta\alpha(g_i)$. (The prime p therefore also does not divide the denominator of any coefficient of any $\beta\alpha(h)$.) In general, if p does not divide b, the fraction a/b is well-defined as an element of $\mathbb{Z}/p$. Therefore we can reduce all $\beta\alpha(g_i)$'s mod p to obtain a homomorphism γ of $\beta\alpha(G)$ to a group of matrices over $\mathbb{Z}/p$ if p fails to divide all denominators in all $\beta\alpha(g_i)'s$. We know that $\beta\alpha(g)$ for the originally-chosen g is not the identity matrix. Therefore if p is larger than all numerators in the matrix $\beta\alpha(g)$ as well, then $\gamma\beta\alpha(g)$ will also be distinct from the identity. □

Corollary 4.4 *Every hyperbolic lattice is residually finite.*

Proof. For a d-dimensional hyperbolic lattice G, the base space $B = \mathbb{H}^d/G$ is a smooth and closed, hence triangulable, manifold. Therefore the fundamental group $\pi_1(B)$ is finitely generated. In effect, G is isomorphic to a finitely generated group of matrices.

Proof of Theorem. The proof consists of two parts. In the first part we show the existence of a d-dimensional hyperbolic lattice, and in the second part, given a number c, we show that every hyperbolic lattice contains a sublattice G' whose girth is greater than c. For the first part, we quote directly from the introduction to [4, p. 111]:

> A Clifford-Klein form of a connected and simply connected Riemannian manifold M is a Riemannian manifold M' whose universal Riemannian covering (universal covering endowed with the metric lifted from M') is isomorphic to M. The main purpose of this note is to prove the following:

Theorem A. *A simply connected Riemannian symmetric space M always has a compact Clifford-Klein form. Any such form M' has a finite Galois covering which is proper, unless M' is isomorphic to M.*

We recall that a Riemannian manifold X is symmetric, in the sense of Cartan, if it is connected and if every point x in X is an isolated fixed point of an involutive isometry s_x.

Obviously, hyperbolic space $\mathbb{H}^d$ is a simply-connected Riemannian symmetric space, and a compact Clifford-Klein form of $\mathbb{H}^d$ produces immediately a lattice of isometries of $\mathbb{H}^d$ as the group of covering transformations.

For the second part of the proof, assume that G is a lattice of isometries of $\mathbb{H}^d$ and let $c > 0$. Let B be the quotient manifold $\mathbb{H}^d/G$ and let $p : \mathbb{H}^d \to B$ be the covering map. The girth of G is determined by the shortest non-contractible (unbased) loop in B.

If λ is a non-contractible unbased loop, λ represents a conjugacy class of elements of $\pi_1(M)$. If g is an element of this conjugacy class and N is a normal subgroup of $\pi_1(M)$ that does not contain g, then N does not contain any conjugate of g either. Since B is compact, it admits only finitely many homotopy classes with loops of length at most c, say $\{[\lambda_1], [\lambda_2], \ldots, [\lambda_k]\}$. Each $[\lambda_i]$ represents a conjugacy class of an element $g_i \in \pi_1(M)$. Let $F = \{g_1, g_2, , g_k\}$. By Corollary 4.4 and Proposition 4.2, $\pi_1(M)$ has a finite-index normal subgroup N that does not intersect F. Thus N does not contain any conjugate of any of the g_i's.

Let $q : \tilde{B} \to B$ be the covering of B corresponding to N. Since B is compact and N has finite index, B is compact as well. By the universality of the covering $p : \mathbb{H}^d \to B$, there exists a covering $\tilde{p} : \mathbb{H}^d \to \tilde{B}$, hence $\tilde{B}$ determines a d-dimensional hyperbolic lattice $\tilde{G}$. It is clear that the girth of $\tilde{G}$ is greater than c, or, in other words, the length of every non-contractible (unbased) loop in $\tilde{B}$ is greater than c, because $q_* : \pi_1(\tilde{B}) \to \pi_1(B)$ is a monomorphism, $q_*\pi_1(\tilde{B}) = N$ and q does not increase the length of any loop.

5. Dense n-reduced Coverings

We begin with an example of an arbitrarily dense 2-saturated lattice covering of $\mathbb{E}^d$ with unit balls ($d \geqslant 2$). Let $e_1, e_2, \ldots, e_d$ be an orthonormal basis for $\mathbb{E}^d$. Consider the lattice generated by the vectors $v_i = ae_i$ for $1 \leqslant i \leqslant d-1$ and

$$v_d = \left(1 + \sqrt{1 - \frac{a^2}{4}(d-1)}\right) e_d + \frac{a}{2} \sum_{i=1}^{d-1} e_i,$$

where $0 < a < \dfrac{2}{\sqrt{d-1}}$.

Clearly, the unit balls centered at the lattice points form a (simply-) reduced covering. Moreover, each ball covers pairs of point not contained in any other ball such that the distance between them approaches 2. Therefore, if two balls are deleted, then one can find four uncovered points that form the vertices of a parallelogram with two sides of lengths approaching 2. Since no such set of four points can be covered by a single unit ball, the covering is 2-reduced. However, the

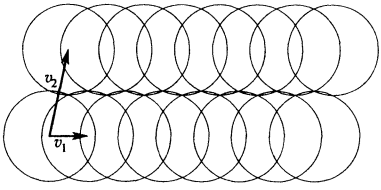


Figure 2. A high-density, 2-reduced covering by circles

density of the covering is arbitrarily large for sufficiently small a. Figure 2 illustrates this covering for $d = 2$.

An equally simple construction yields an infinitely dense 2-saturated covering of $\mathbb{E}^d$ with unit balls ($d \geqslant 2$), in fact locally infinitely dense at every point. Let P be a hyperplane in $\mathbb{E}^d$ containing the origin and let Q be a dense subset of P such that $P \backslash Q$ is dense in P as well. Let v denote the vector normal to P of length 2. The collection of unit balls centered at the points of the form $q + 2iv$ for $q \in Q$ and $p + (2j + 1)v$ for $p \in P \backslash Q$, $i, j \in \mathbb{Z}$, is a covering. Since the removal of any single ball from this collection uncovers both end points of the ball's diameter parallel to v, the covering is not only reduced, but 2-reduced as well.

The second construction generalizes to an arbitrary body by choosing the hyperplane P to be perpendicular to a diameter of the body and giving v the same length as the diameter.

Let $\Theta_n(K)$ be the supremum of the densities of all n-reduced coverings with K. For example, the above constructions show that $\Theta_1(K) = \Theta_2(K) = \infty$. However, the simple relation

$$\lim_{n \to \infty} \Theta_n(K) = \vartheta(K)$$

implies the existence of a smallest positive integer $l(K)$ such that $\Theta_n(K) < \infty$ for all $n \geqslant l(K)$. The notion of the Newton covering number of a convex body (introduced below) yields an upper bound for $l(K)$, where K is an arbitrary convex body in $\mathbb{E}^d$. In addition, an application of a theorem of Bárány yields a slightly better bound for $l(B^d)$, where B^d denotes the unit ball in $\mathbb{E}^d$.

Recall that the Hadwiger covering number $H_c(K)$ of a convex body K in $\mathbb{E}^d$ is the minimum number of translates of K whose union contains a neighborhood of K. HADWIGER [14] asks for the maximum value of $H_c(K)$ over all convex bodies K in $\mathbb{E}^d$. The problem was stated independently by others, also in the context of the equivalent problem of illumination of K, and it is conjectured that $H_c(K) \leqslant 2^d$ with equality for parallelotopes only. The conjecture is still open in every dimension $d \geqslant 3$.

Similarly, one can consider $N_c(K)$, the minimum number of replicas of K whose union contains a neighborhood of K. The quantities $N_c(K)$ and $H_c(K)$ have their dual counterparts $N(K)$ and $H(K)$, called the Newton (or kissing) number and the Hadwiger number, respectively, in the context of packings (see [12]). We call $N_c(K)$ the *Newton covering number* of K to extend the analogy. Obviously, $N_c(K) \leqslant H_c(K)$ for all K.

The following theorem establishes a relation between $l(K)$ and $N_c(K)$.

Theorem 5.1. *For every convex body K in Euclidean space,*

$$l(K) \leqslant N_c(K) + 1.$$

In the proof of this theorem, as well as in the next section, we use certain relations between the (global) density of an arrangement and its density with respect to some bounded domain. In what follows, the volume of a (measurable) set S will be denoted by $V(S)$. As usual, B^d denotes the unit ball in $\mathbb{E}^d$, so rB^d is the ball of radius r centered at the origin. Let $\mathscr{A}$ be a locally finite arrangement of uniformly bounded measurable sets, and let G be a bounded domain. The density $d(\mathscr{A} \mid \mathscr{G})$ of $\mathscr{A}$ relative to G is defined by

$$d(\mathscr{A} \mid G) = \frac{1}{V(G)} \sum_{A \in \mathscr{A}} V(A \cap G),$$

and the average density $d_{\text{av}}(\mathscr{A} \mid G)$ to $\mathscr{A}$ relative to all translates of G is defined by

$$d_{\text{av}}(\mathscr{A} \mid G) = \lim_{r \to \infty} \frac{1}{V(rB^d)} \int_{rB^d} d(\mathscr{A} \mid (G + x)),$$

provided the limit exists. Otherwise we take the lim sup or the lim inf, and it is usually clear from context which limit is meant. Also, the domain of the integral above is rB^d, just as rB^d is frequently used to define the density of $\mathscr{A}$ as a limit.

The following proposition is derived from these definitions by routine methods of real analysis, interchanging sums and limits with integrals, and applying Fubini's theorem.

Proposition 5.2. *For every locally finite arrangement $\mathscr{A}$ of uniformly bounded measurable sets, and any bounded domain G, the average density $d_{\text{av}}(\mathscr{A} \mid G)$ coincides with the density of $\mathscr{A}$.*

As a direct corollary, we get:

Proposition 5.3. *Let $\mathscr{A}$ be a locally finite arrangement of uniformly bounded measurable sets, and let G be a bounded domain. If the density of $\mathscr{A}$ is a, then there exists a translate of G such that $d(\mathscr{A} \mid G) \geqslant a$ and there exists a translate of G such that $d(\mathscr{A} \mid G)) \leqslant a$.*

Proposition 5.2 can be put in an equivalent, discrete form:

Proposition 5.4. *Let D be a locally finite set of points, and let G be a bounded domain. If the number density of D is a, then the average number of points contained in a translate of G is equal to $aV(G)$.*

Again, as a corollary, we get:

Proposition 5.5. *Let D be a locally finite set of points, and let G be a bounded domain. If the number density of D is a, then there exists a translate of G which contains at least $aV(G)$ points of D and there exists a translate of G which contains at most $aV(G)$ points of D.*

Proof of Theorem. By definition, there exists an $\varepsilon > 0$ such that some $N_c(K)$ replicas of K cover the ε-neighborhood (the outer parallel domain of radius ε) of K. Let $p \in K$, and for each member K_i of the covering, let p_i be the image of p under an isometry that takes K to K_i. Since the group of isometries of the space that fix p is compact, it can be partitioned into a finite collection of sets such that, if g and h belong to the same set, then the Hausdorff distance between $g(K)$ and $h(K)$ is smaller than $\varepsilon/2$, *i.e.*, each of $g(K)$ and $h(K)$ lies in the other's $\varepsilon/2$-neighborhood.

If the covering is sufficiently dense, then by Proposition 5.5, there exists a ball of radius $\varepsilon/2$ containing $N_c(K) + 1$ points p_i such that the Hausdorff distance between each two of the sets $K_i - p_i$ (each set K_i shifted so that p_i is moved back to the origin) is smaller than $\varepsilon/2$. We have now $N_c(K) + 1$ replicas of K, say $K_1, K_2, \ldots, K_{N_c(K)+1}$, such that K_i lies in the ε-neighborhood of K_1 for $2 \leqslant i \leqslant N_c(K)) + 1$. By the definition of the Newton covering number of K, these $N_c(K) + 1$ replicas can be replaced by $N_c(K)$ others without uncovering any points. Thus, every sufficiently dense covering with replicas of K fails to be $(N_c(K) + 1)$-reduced. □

Since $N_c(B^d) = d + 1$, the above theorem implies immediately that $l(B^d) \leqslant \leqslant d + 2$. However, using a result of BÁRÁNY [3, Th. 2] which generalizes a theorem of ERDŐS and SZEKERES [7], one can improve this inequality as follows.

Theorem 5.6. $l(B^d) \leqslant d + 1$.

Proof. Bárány's theorem states: For any $\varepsilon > 0$ and $d \geqslant 2$ there exists a constant $n(d, \varepsilon)$ such that every finite set $V \subset \mathbb{E}^d$ contains a subset $W \subset V$, $| W | \leqslant n(d, \varepsilon)$ with the property that for $v \in V \backslash W$ there are points $w_1, w_2 \in W$ with $\measuredangle(w_1 v w_2) > \pi - \varepsilon$. Given positive numbers ε and $\delta < 1$, if a covering of $\mathbb{E}^d$ with unit balls is sufficiently dense, then, by Proposition 5.5, some ball of radius δ contains at least $(d + 1)\binom{n(d,\varepsilon)}{2}$ centers of the unit balls. If we let V be the set of these centers and we apply Bárány's theorem, we obtain $d + 3$ distinct points $w_1, w_2, v_1, v_2, \ldots, v_{d+1}$ in V such that $\measuredangle(w_1 v_i w_2) > \pi - \varepsilon$ for $i = 1, 2, \ldots, d + 1$. It follows that the points $v_1, v_2, \ldots, v_{d+1}$ lie in the "double cone" C (the union of two congruent non-overlapping right cones with a common base) whose apexes are w_1 and w_2 and whose angle at each apex is 2ε.

Let $B(p)$ be the unit ball centered at the point p. Each of the $d + 1$ balls $B(v_i)$ $(1 \leqslant i \leqslant d + 1)$ is contained in the outer parallel domain P of radius 1 of C. Observe that the set $P \backslash (B(w_1) \cup B(w_2))$ is a neighborhood of the $d - 2$-dimensional unit sphere centered at the midpoint of $w_1 w_2$ which lies in the hyperplane perpendicular to $w_1 w_2$, and that his neighborhood is arbitrarily close to the sphere for sufficiently small ε and δ. Since such a neighborhood can be covered by d unit balls, it follows that a very dense covering of $\mathbb{E}^d$ with unit balls cannot be $(d + 1)$-reduced. Figure 3 shows the set $P \backslash (B(w_1) \cup B(w_2))$ in dimension 2, where $\varepsilon = \pi/6$ and $\delta = 1$ are small enough for our purpose. □

Remark. The proof does not use the fact that the given collection of balls is a covering. Define an n-reduced arrangement (not necessarily a covering) of replicas of K by the property that it is not possible to delete n members of the arrangement

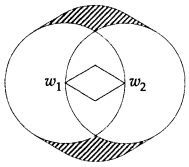


Figure 3. The set $p\backslash(B(w_1) \cup B(w_2))$

and replace them with some $n-1$ replicas of K without uncovering any point that was covered by the original arrangement. Then the above argument demonstrates that a $d+1$-reduced arrangement of unit balls cannot have arbitrarily high density, and for this generalization the bound of $d+1$ is the best possible.

So far we have considered arrangements of replicas of a given body, without restrictions on the isometries that send one of them onto another. However, we can also consider arrangements with restrictions on the allowed isometries. For example, many of the ideas and results of this section and previous sections generalize to arrangements of translates of K (which is only an appropriate restriction in the Euclidean case, of course). In particular, Theorem 1.2 becomes:

Theorem 5.7. *Every body K in $\mathbb{E}^d$ admits both an n-saturated packing and an n-reduced covering in the class of packings and coverings with translates of K.*

Theorem 5.1 also has a "translative" version with a completely analogous proof. (In the corresponding notation, we indicate the restriction to translates by the subscript "T".)

Theorem 5.8. *For every convex body K in Euclidean space,*

$$l_T(K) \leqslant H_c(K) + 1.$$

Example. While the bounds given in Theorem 5.1 and Theorem 5.6 could be far from optimal, the following example indicates that Theorem 5.8 is close to optimal in at least some cases.

Let P denote a right pyramid in the coordinate space $\mathbb{E}^d$ whose base is the unit $d-1$-dimensional cube $Q^{d-1} = \{(x_1, x_2, \ldots, x_{d-1}, 0) \in \mathbb{E}^d : |x_i| \leqslant 1/2 \text{ for } i < d\}$. Observe that $H_c(P) = 1 + H_c(Q^{d-1}) = 1 + 2^{d-1}$. Consider the set Λ of vectors of the form $(n_1, n_2, \ldots, n_{d-1}, x)$ where each n_i is an integer and $x \in \mathbb{R}$ is rational if and only if $\Sigma\, n_i$ is even. The translates of P by all vectors of Λ is a covering of $\mathbb{E}^d$ of infinite density. It is easily verified that this covering is $(H_c(P)-1)$-reduced. Therefore $l_T(P) \geqslant H_c(P)$.

The following estimate for the Hadwiger covering number of a convex body in $\mathbb{E}^d$ is due to Rogers (unpublished):

$$H_c(K) \leqslant \frac{V(K-K)}{V(K)}\,(d \log d + d \log\log d + 5d),$$

where $K-K$ is the difference body of K, consisting of points of the form $x-y$ where $x, y \in K$. The inequality follows from the result of Rogers [17] which states that each d-dimensional covex body K admits a covering of $\mathbb{E}^d$ by its translates of density at most $d \log d + d \log\log d + 5d$. If $\{K + a_i\}$ is a covering of density

guaranteed by the theorem of Rogers, then Proposition 5.5 implies that there exists a $\lambda > 1$ and a translate $(\lambda K) - K + c$ of $(\lambda K) - K$ containing at most $(d \log d + d \log\log d + 5d)\, V(K-K)/V(K)$ of the points a_i. Since $(K + a_i) \cap \lambda K + c \neq \emptyset$ if and only if $a_i \in \lambda K - K + c$, it follows that the respective translates of K cover $\lambda K + c$.

For a centrally symmetric body K in $\mathbb{E}^d$, we have $V(K-K)/V(K) = 2^d$, thus in this case, the Rogers bound for $H_c(K)$ is reasonably close to the conjectured best upper bound of 2^d. In the general case, a result of ROGERS and SHEPHARD [19] states that $V(K-K)/V(K) \leqslant \binom{2d}{d}$ for every convex body $K \subset \mathbb{E}^d$, which yields the asymptotic bound $H_c(K) \leqslant 4^{d+o(d)}$.

6. Asymptotic Density Bounds

In Section 5, we defined $\Theta_n(K)$ as the supremal density of all n-reduced coverings with replicas of K, and we mentioned the simple relation

$$\lim_{n\to\infty} \Theta_n(K) = \vartheta(K).$$

Analogously, let $\Delta_n(K)$ be the infimum of the densities of all n-saturated packings with replicas of K, and note the analogous relation

$$\lim_{n\to\infty} \Delta_n(K) = \delta(K).$$

Also, observe that $\Delta_n(K) > 0$ for every body K and every $n \geqslant 1$. Obviously, each of the two sequences $\{\Delta_n(K)\}$ and $\{\Theta_n(K)\}$ is monotonic. The following inequalities give estimates for the rate of convergence of the sequences $\{\Delta_n(K)\}$ and $\{\Theta_n(K)\}$:

$$\Delta_n(K) \geqslant \delta(K) - O(n^{-1/d}) \tag{6.1}$$

and

$$\Theta_n(K) \leqslant \vartheta(K) + O(n^{-1/d}). \tag{6.2}$$

To prove inequality (6.1), assume that K is of body of diameter 1 and volume V and let r denote the minimum radius of a ball that can intersect n non-overlapping replicas of K. Let σ_d denote the volume of the unit ball in $\mathbb{E}^d$. By the definition of the packing density of K and Proposition 5.3,

$$nV/(\sigma_d r^d) \geqslant \delta(K) - \varepsilon$$

for every $\varepsilon > 0$, hence

$$\frac{nV}{\sigma_d r^d} \geqslant \delta(K).$$

Assume now that $\mathscr{P}$ is an n-saturated packing with replicas of K, and let p denote the density of this packing. Any ball of radius $r + 2$ must contain at least n members of $\mathscr{P}$, for otherwise the members of $\mathscr{P}$ contained in the ball could be replaced by n non-overlapping replicas of K intersecting the concentric ball of radius r. Thus the total volume of the intersections of such a ball with all members

of $\mathscr{P}$ is at least nV. Using Proposition 5.3 again, we obtain

$$p \geqslant \frac{nV}{\sigma_d (r+2)^d}.$$

It follows immediately that

$$p \geqslant \delta(K) \left(\frac{r}{r+2} \right)^d.$$

By the definition of r, a ball of radius r intersects at least n non-overlapping replicas of K. Each of these replicas is contained in the concentric ball of radius $r+1$. Thus $\sigma_d (r+1)^d \geqslant nV$, and we get

$$r \geqslant \left(\frac{V}{\sigma_d} n \right)^{1/d} - 1.$$

Since the function $f(x) = (x/(x+2))^d$ is increasing, we get:

$$p \geqslant \delta(K) \left(\frac{cn^{1/d} - 1}{cn^{1/d} + 1} \right)^d,$$

where $c = (V/\sigma_d)^{1/d}$, and inequality (6.1) follows.

The proof of inequality (6.2) is analogous.

The above method can be refined as follows to yield some specific density bounds for n-saturated packings and n-reduced coverings of $\mathbb{E}^d$ with unit balls. For packings, consider a "cluster" of n non-overlapping unit balls and let G be the outer parallel domain of radius 1 of their union. Let $\mathscr{P}$ be an n-saturated packing with unit balls. Then every translate of G contains at least n centers of the members of $\mathscr{P}$. It follows, by Proposition 5.5, that the density of $\mathscr{P}$ is at least $n\sigma_d / V(G)$.

The smaller the volume of G, the greater the resulting bound, which raises the problem of arranging n non-overlapping unit balls in $\mathbb{E}^d$ so that the volume of the outer parallel domain of radius 1 of their union is minimum. A similar method can be used for coverings, and it leads to the problem of arranging n unit balls in $\mathbb{E}^d$,

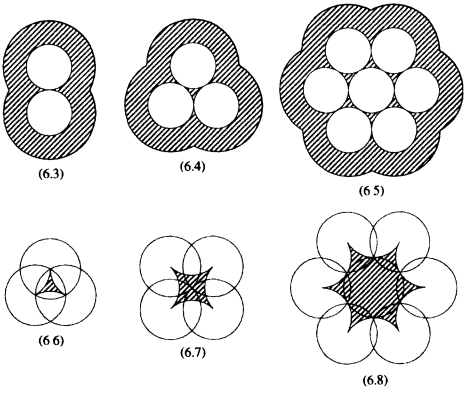


Figure 4. Economical clusters of unit circles

this time allowing overlaps, so that the volume of the inner parallel domain of radius 1 of their union is maximum. Of course, this method only works if $n \geqslant d+1$.

Except for some trivial cases, we do not know the solution to these problems even for $d=2$. However, some clusters of unit circles in $\mathbb{E}^2$, shown in Figure 4, seem reasonably economical for the method described above. Using translates of these clusters to estimate density bounds, we get the following:

$$\Delta_2(B^2) \geqslant \frac{3\pi}{3\sqrt{3}+8\pi} = 0.31075\ldots \tag{6.3}$$

$$\Delta_3(B^2) \geqslant \frac{3\pi}{4\sqrt{3}+6\pi} = 0.36561\ldots \tag{6.4}$$

$$\Delta_7(B^2) \geqslant \frac{7\pi}{12\sqrt{3}+8\pi} = 0.47892\ldots \tag{6.5}$$

$$\Theta_3(B^2) \leqslant \frac{6\pi}{2\sqrt{3}-\pi} = 58.44661\ldots \tag{6.6}$$

$$\Theta_4(B^2) \leqslant \frac{6\pi}{4-\pi} = 14.63916\ldots \tag{6.7}$$

$$\Theta_7(B^2) \leqslant \frac{7\pi}{6\sqrt{3}-2\pi} = 5.35179\ldots\ . \tag{6.8}$$

Clearly, these inequalities are far from sharp. Among good estimates for $\Delta_n(B^2)$ and $\Theta_n(B^2)$, one stands out. Clearly, any 1-saturated packing with unit balls becomes a covering if the radius of each ball is increased to 2. Since $\vartheta(B^2) = 2\pi/\sqrt{27}$ (a well-known results of KERSHNER [15]), it follows that

$$\Delta_1(B^2) = \pi/6\sqrt{3} = 0.302299\ldots$$

Also, as we mentioned before,

$$\Theta_1(B^2) = \Theta_2(B^2) = \infty.$$

Apart from these three cases, it seems difficult to determine the exact values of $\Delta_n(B^2)$ and $\Theta_n(B^2)$.

7. Remarks, Open Problems and Conjectures

In relation to the conjecture stated in the introduction, claiming the existence of completely saturated packings and completely reduced coverings, observe the following:

(i) Complete saturation implies maximum density and complete reduction implies minimum density. More precisely, the density of a completely saturated packing with replicas of a body K exists and is equal to $\delta(K)$. Similarly, the density of a completely reduced covering with replicas of K exists and is $\vartheta(K)$.

(ii) Obviously, the converse of (i) is false. But a weaker statement holds: A periodic packing with replicas of K with density $\vartheta(K)$ is completely saturated

and a periodic covering with replicas of K whose density is $\vartheta(K)$ is completely reduced.

The first observations indicates that the conjecture on existence of completely saturated packings and reduced coverings is not as obvious as it might appear. The conjecture, if true, would imply a version of GROEMER's result [13] on the existence of maximum density packings and minimum density coverings.

The second observation brings to mind the well-known problem: Given a body K, is there a periodic packing [covering] with replicas of K, whose density is $\delta(K)[\vartheta(K)]$? A positive answer to this question would imply our conjecture. However, SCHMITT [20] constructed a strictly star-shaped prototile for a monohedral tiling in $\mathbb{E}^3$ such that no tiling with its replicas is periodic, and by a slight modification of Schmitt's construction Conway produced a convex prototile with this property. For $\mathbb{E}^2$ no such example is known, but according to another result of SCHMITT [21] there is a strictly star-shaped set $K \subset \mathbb{E}^2$ whose replicas do not admit a periodic packing of density $\delta(K)$.

Generally, it seems extremely difficult to determine whether a given convex body admits a periodic packing (covering) of maximum (minimum) density. In particular, the answer is not known for the d-dimensional ball ($d \geqslant 3$). The case $d = 2$ offers some answers, since it is known (see [8, 10]) that every centrally-symmetric convex disk attains its packing density in a lattice packing. The analogous statement for converings is only a conjecture, supported by a partial result under the restriction to crossing-free coverings (see [8, 10]).

There are only a handful of cases in which sphere packings in Euclidean or hyperbolic space are known to be completely saturated. Without exception, they follow from the ROGERS [18] and the BÖRÖCZKY [2] bounds: The density of any sphere packing in d dimensions is at most the density in a regular simplex of $d+1$ kissing spheres with centers at the vertices of the simplex. If the regular simplex tiles space, there exists a corresponding periodic sphere packing that achieves the bound. The only regular simplices that tile Euclidean and hyperbolic space are:

- Equilateral triangles in $\mathbb{E}^2$.
- Triangles in $\mathbb{H}^2$ with angles of $2\pi/n$ for $n \geqslant 7$.
- Simplices in $\mathbb{H}^4$ with dihedral angles of $2\pi/5$.

The analogous bound for coverings also holds in Euclidean space (see COXETER-FEW-ROGERS [6]), although it is open for sphere coverings of hyperbolic d-space and the d-sphere for $d > 2$. (For $\mathbb{H}^2$ the bound for circle coverings is due to L. FEJES TÓTH [9]). Therefore the same simplices also produce completely reduced coverings of the same types, except perhaps in $\mathbb{H}^4$.

Clearly, the familiar densest lattice packing (covering) of $\mathbb{E}^2$ with unit circles is completely saturated (reduced), but we do not know whether there is a non-lattice, completely saturated packing (completely reduced covering) with unit circles. We do not even know whether or not the circle packing in Figure 5 is completely saturated. The arrangement of circles shown there is given by dividing lattice packing into two "half-plane" parts along a pair of adjacent rows of circles and

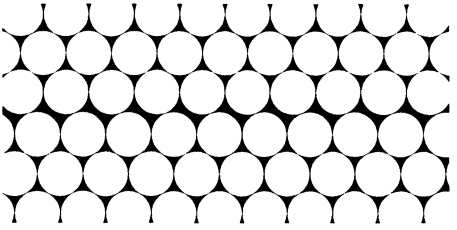


Figure 5. A sub-optimal packing which could be completely saturated

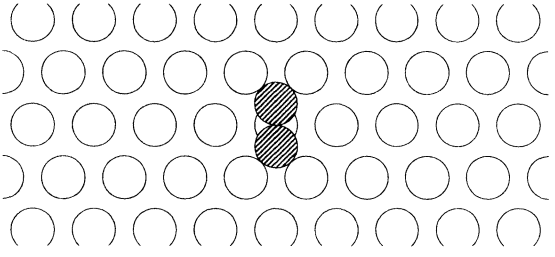


Figure 6. Possibly the least dense 2-saturated packing

then separating the parts slightly while maintaining contact between the two adjacent rows.

Although it seems difficult to determine $\Delta_n(B^2)$ and $\Theta_n(B^2)$, we conjecture that

$$\Delta_2(B^2) = \pi(3 - \sqrt{5})/\sqrt{27} = 0.461873\ldots,$$

which is the density of the packing shown in Figure 6. But we do not even have a conjecture for the other constants.

Theorem 5.1 relates the Newton covering number $N_c(K)$ for a convex body K to $l(K)$. Newton covering numbers are of interest in their own right: Among all convex bodies in d dimensions, which one has the greatest Newton covering number, and what is that number? Let P denote the right pyramid over a $d-1$-dimensional cube Q^{d-1}, as in the remark following Theorem 5.8. Is $N_c(P) = N_c(Q^{d-1}) + 1$? Does $N_c(P)$ depend on the height of the pyramid? What is the Newton covering number of the cube Q^d?

The inequality $l(B^d) \leqslant d + 1$ (Theorem 5.6) is sharp for $d = 2$ (see Figure 2), but we suspect that for $d \geqslant 3$ this is not the case. It might even turn out that $l(B^d) = 3$ for all $d \geqslant 2$. This problem can be stated more simply as follows: Given a very dense covering of $\mathbb{E}^d$ with unit balls, can one always make a new covering by replacing three balls by two?

Acknowledgements. The authors acknowledge, with gratitude, that during the preparation of this paper, the research of G. Fejes Tóth has been supported by the Hungarian Foundation for Scientific Research (OTKA), grants no. 1907 and no. 14218, and that of W. Kuperberg by the National Science Foundation, grant no. DMS-9403515.

References

[1] Böröczky K (1974) Sphere packing in spaces of constant curvature I. Mat Lapok **25**: 265–306
[2] Böröczky K (1978) Packing of spheres in spaces of constant curvature. Acta Math Acad Sci Hungar **32**: 243–261
[3] Bárány I (1987) An extension of the Erdös-Szekeres theorem on large angles. Combinatorica **7**: 161–169
[4] Borel A (1963) Compact Clifford-Klein forms of symmetric spaces. Topology **2**: 111–122
[5] Benedetti R, Petronio C (1992) Lectures on Hyperbolic Geometry. Berlin Heidelberg New York: Springer
[6] Coxeter HSM, Few L, Rogers CA (1959) Covering space with equal spheres. Mathematika **6**: 147–157
[7] Erdös P, Szekeres G (1960–61) On some extremum problems in elementary geometry. Ann Univ Sci Budapest, Eötvös Sect Math **3–4**: 53–62
[8] Fejes Tóth L (1950) Some packing and covering theorems. Acta Sci Math Szeged **12/A**: 62–67
[9] Fejes Tóth L (1964) Regular Figures. Oxford: Pergamon Press
[10] Fejes Tóth L (1972) Lagerungen in der Ebene, auf der Kugel und im Raum. 2nd edn. Berlin Heidelberg New York: Springer
[11] Fejes Tóth L, Heppes A (1980) Multi-saturated packings of circles. Studia Sci Math Hungar **15**: 303–307
[12] Fejes Tóth G, Kuperberg W (1993) Packing and covering with convex sets. In: Gruber PM, Wills JM (eds) Handbook of Convex Geometry, pp. 799–860. Amsterdam: North-Holland
[13] Groemer H (1963) Existenzsätze für Lagerungen im Euklidischen Raum. Math Z **81**: 260–278
[14] Hadwiger H (1957) Ungelöstes Problem Nr 20. Elem Math **12**: 121
[15] Kershner R (1939) The number of circles covering a set. Amer J Math **61**: 665–671
[16] Mal'cev AI (1940) On the faithful representation of infinite groups by matrices. Mat Sb **8**: 405–422
[17] Rogers CA (1957) A note on coverings. Mathematika **4**: 1–6
[18] Rogers CA (1958) The packing of equal spheres. Proc London Math Soc (3) **8**: 609–620
[19] Rogers CA, Shephard GC (1957) The difference body of a convex body. Arch Math **8**: 220–233
[20] Schmitt P (1988) An aperiodic prototile in space. Preprint
[21] Schmitt P (1991) Discs with special properties of densest packings. Discrete Comput Geom **6**: 181–190
[22] Spanier EH (1966) Algebraic Topology. Berlin Heidelberg New York: Springer

G. Fejes Tóth
Mathematical Institute
Hungarian Academy of Sciences
P.O. Box 127
H-1364 Budapest
Hungary
e-mail: gfejes@math-inst.hu

G. Kuperberg
Department of Mathematics
University of California
Davis, CA 95616-8633
U.S.A.
e-mail: greg@math.ucdavis.edu

W. Kuperberg
Department of Mathematics
Auburn University
Auburn, AL 36849-5310
U.S.A.
e-mail: kuperwl@mail.auburn.edu

Mh. Math. 125, 147–164 (1998)

Monatshefte für
Mathematik

Printed in Austria

Equidistribution and Brownian Motion on the Sierpiński Gasket*

By

Peter J. Grabner and **Robert F. Tichy**, Graz

With 2 Figures

(Received 28 February 1996)

Abstract. We introduce several concepts of discrepancy for sequences on the Sierpiński gasket. Furthermore a law of iterated logarithm for the discrepancy of trajectories of Brownian motion is proved. The main tools for this result are regularity properties of the heat kernel on the Sierpiński gasket. Some of the results can be generalized to arbitrary nested fractals in the sense of T. Lindstrøm.

1. Introduction

As a starting point we consider the Sierpiński gasket, a well known planar fractal set introduced by *W. Sierpiński* [26]. Let A_0 be a closed equilateral triangle of unit sides e_1, e_2, e_3 with vertices $P_1\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right), P_2(0,0), P_3(1,0)$. Let A_1 be the set obtained by deleting the open equilateral triangle whose vertices are the midpoints of the edges of A_0. Thus A_1 consists of three equilateral triangles with side $\frac{1}{2}$. Repeating this procedure we obtain successively $A_2, A_3, \ldots A_n$ consists of 3^n equilateral triangles of side 2^{-n}, which are called *elementary triangles of level n*. Furthermore, we denote the set of all vertices of A_n by V_n and the boundary of A_n by E_n. Thus, $F_n = (V_n, E_n)$ is defining a finite graph.

Definition 1. The set $G = \bigcap_{n=0}^{\infty} A_n$ is called the (*bounded*) *Sierpinski gasket.*

Remark 1. Any point $p \in G$ can be represented by the triple (k_1, k_2, k_3) with $k_1 + k_2 + k_3 = 2$, where

$$k_i = k_i(p) = \sum_{l=1}^{\infty} \frac{\varepsilon_l^{(i)}}{2^l}, \quad \varepsilon_l^{(i)} = 0 \quad or \ 1$$

and $\varepsilon_l^{(1)} + \varepsilon_l^{(2)} + \varepsilon_l^{(3)} = 2$ for all $l \geqslant 1$. Note that $(1 - k_i)\frac{\sqrt{3}}{2}$ is just the distance of P to the side e_i.

---

*Dedicated to Prof. Edmund Hlawka on the occasion of his 80th birthday

1991 Mathematics Subject Classification: 60B99, 11K06

Key words: diffusion processes, fractals, discrepancy, uniform distribution

The authors are supported by the Austrian Science Foundation project Nr. P10223-PHY and by the Austrian-Italian scientific cooperation program project Nr. 39

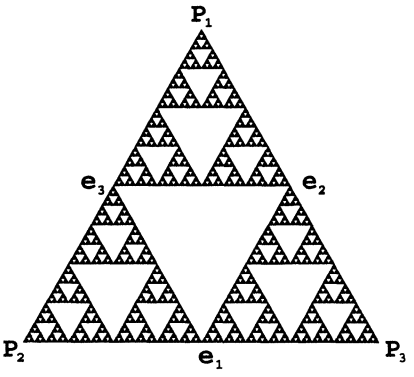


Figure 1

By standard techniques, as described in [26], it is easy to see that the Sierpiński gasket has Hausdorff dimension $\alpha = \log 3/\log 2$ and finite positive Hausdorff measure. Let μ denote the (normalized) Hausdorff measure of dimension α on G

In a series of papers, methods and results from classical potential theory in the Euclidean space were extended to the Sierpiński gasket, the Sierpiński carpet, where the whole potential theory is developed in a series of papers (cf. [1] for further references) or more generally to so called nested fractals. We want to mention here the fundamental paper of M. T. BARLOW and E. A. PERKINS [2], where a systematic theory of Brownian motion on the Sierpiński gasket is developed. Let $\tilde{F}$ be the infinite graph defined by $\tilde{F} = \bigcup_k 2^k F_k$.

Definition 2. The topological closure of

$$\bigcup_{k=0}^{\infty} 2^{-k}\tilde{F}$$

is called *the infinite Sierpiński gasket* $\tilde{G}$.

The Brownian motion is introduced as a suitable limit process of discrete random walks on the vertices of $\tilde{F}$: Let Y_k be the random walk on $\tilde{F}$ with transition probabilities $\frac{1}{4}$ from each vertex to its neighbour. In order to give a proper definition of the limiting process one has to rescale the time according to the eigenvalues of the transition matrix. Thus we consider the processes

$$X^{(n)}(t) = 2^{-n} Y_{[5^n t]} \quad \text{for } t \geqslant 0,\ n \in \mathbb{N}.$$

In, [2, Theorem 2.8] it is shown, that the processes $X^{(n)}$ converge weakly to a process X, where X is a continuous, non-constant, $\tilde{G}$-valued, strong Markov process starting at O. This process $X = X_t$ can be considered as the Brownian motion on $\tilde{G}$. The Laplacian is obtained as the infinitesimal generator of the semi-group describing this process. Among other very interesting results the authors obtain regularity properties of the heat kernel.

LINDSTRØM [20] studies Brownian motion on compact nested fractals. In the case of the finite (=compact) Sierpiński gasket G a trajectory of Brownian motion

is obtained from a trajectory on $\tilde{G}$ just by factorizing $\tilde{G}$ modulo the equivalence relation ρ, which identifies all the translates of G whose union is $\tilde{G}$. Notice that ρ identifies the points P_1, P_2 and P_3.

Another way to obtain the Brownian motion on G is to consider the limit process of random walks on the vertices of A_n, where the transition probabilities in each point is $\frac{1}{4}$ to any neighbouring point, and the transition probabilities for leaving the vertices of A_0 is $\frac{1}{2}$ (this is in direct correspondence to the equivalence relation constructed above). This actually is the approach of Lindstrøm. The Laplacian is again defined as the infinitesimal generator of this process. The main results of Lindstrøm are concerned with the asymptotic behaviour of the eigenvalues of the Laplacian.

A completely different approach is due to a Japanese school. In [18] KIGAMI considers finite difference operators, so called harmonic differences on F_n. The Laplacian then is defined as a certain limit of these operators. The key idea is to use the step by step construction of the graphs to investigate the "evolution" of the eigenvalues (cf. [25]). Kigami starts with a detailed investigation of harmonic functions on the Sierpiński gasket and its N-dimensional generalizations. FUKUSHIMA and SHIMA [14] use this approach in order to develop a precise spectral analysis. A survey on these developments can be found in the monographs [8] and [10].

In classical papers on uniformly distributed sequences and functions the distribution behaviour of the trajectories of the Brownian motion on the P-dimensional torus $\mathbb{R}^p/\mathbb{Z}^p$ were analyzed (cf. [27], where the one-dimensional case is considered). W. FLEISCHER [13] has considered the p-dimensional case, and later on in [6] this problem could be settled in the case of Brownian motion on Riemannian manifolds. In [6] a law of iterated logarithm for the discrepancy is proved by applying a general technique due to W. PHILIPP [23] are bounds for the eigenvalues of the Laplace-Beltrami-operator on the manifold.

In Section 2 we develop the basic properties of uniformly distributed sequences on the Sierpiński gasket. We define a natural metric and introduce various concepts of discrepancy and obtain inequalities comparing these discrepancies. We discuss special sequences including irregularities of distribution. Furthermore we prove explicit formulae for the Hausdorff measure of certain triangles contained in the Sierpiński gasket. For related subsets of G this measure was computed in [15] as an application of summation formulae for special q-multiplicative arithmetic functions. This is a consequence of the digital description of the gasket, which we have presented above. In section 3 we conclude by proving a law of iterated logarithm for the discrepancy of the trajectories of the Brownian motion on the compact gasket G using a method of BLÜMLINGER [5].

2. Uniform Distribution on the Gasket

2.1. The Geodesic Metric on G. G is a compact space the topology of which is induced by the following metric d. Any two points a and b in G are contained in elementary triangles of level k, $\Delta_k(a)$, $\Delta_k(b)$, respectively. Let a_k, b_k be the lower

left vertices of $\Delta_k(a)$, $\Delta_k(b)$ respectively, and observe that a_k and b_k are vertices of the finite graph F_k. We set

$$d(a, b) = \lim_{k\to\infty} 2^{-k} d_k(a_k, b_k),$$

where d_k is the minimal length of a chain connecting a_k and b_k. Obviously $d(a, b)$ is the geodesic distance of a and b, i.e. the length of the shortest continuous curve in G connecting a and b. This distance has already been used in [2].

Proposition 1. *Let a and b be two points in G given by their digital representation $a = (\varepsilon_l^{(i)})$, $b = (\delta_l^{(i)})$, $i = 1, 2, 3$, $l = 1, 2, \ldots$. Let L be the first index such that the triples $(\varepsilon_L^{(i)})$ and $(\delta_L^{(i)})$ are distinct and define the indices i and j by $\varepsilon_L^{(i)} = 0$ and $\delta_L^{(j)} = 0$. Then the distance of a and b is given by*

$$\sum_{l=L}^{\infty} 2^{-l}(\varepsilon_l^{(j)} + \delta_l^{(i)} - 1).$$

(*The formula does not depend on different representations of the same points.*)

Proof. Assume first that a and b are contained in one elementary triangle of level 1. Blowing up this triangle by a factor 2 yields $d(a,b) = \frac{1}{2} d(\tilde{a}, \tilde{b})$, where $\tilde{a}$ and $\tilde{b}$ are the homothetic images of a and b. This procedure can be continued as far as these iterated homothetic images of a and b lie in two different elementary triangles of level 1. This happens after $L - 1$ iterations. In this case we have $d(a,b) = \frac{1}{2} d(\tilde{a}, P_j) + \frac{1}{2} d(\tilde{b}, P_i)$. Thus, we only have to compute the distances of a given point to one of the points P_1, P_2, P_3. Observing that $d(p, P_m) = k_m(p)$ ($m = 1, 2, 3$, see Remark 1) and $d(\tilde{p}, P_m) = 2k_m(p) - 1$ for $k_m(p) \geqslant \frac{1}{2}$, we obtain $d(a,b) = k_j(a) + k_i(b) - 1$. Inserting the digital representations of Remark 1 we obtain the desired result. □

Remark 2. Let $p \in G$ be a point different from P_i, $i = 1, 2, 3$ and let $\varepsilon > 0$ be sufficiently small. Then the ε-ball

$$B(p, \varepsilon) = \{x \in G \mid d(x, p) < \varepsilon\}$$

consists of two congruent equilateral triangles (intersected with G) with one common vertex. Obviously, $B(P_i, \varepsilon)$ consists of one triangle. Thus the metric d induces the topology of the gasket.

Since G is a compact metric space the general theory of uniform distribution (cf. [19]) can be applied. A sequence (x_n) of points in G is called uniformly distributed (with respect to the Hausdorff measure μ) if

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_G f(x) d\mu(x) \tag{2.1}$$

holds for all continuous functions f on G. By [19], Theorem 1.2, p. 175, (x_n) is uniformly distributed if and only if (2.1) is satisfied for all functions $f = \chi_M$, where M is a Borel set with negligible boundary. In order to describe the distribution behaviour more precisely we introduce several concepts of discrepancy.

2.2. Several Notions of Discrepancy. Let $\mathscr{D}$ be some system of Borel sets A, such that the boundary of A is a null set. Then the discrepancy of a sequence (x_n) with respect to $\mathscr{D}$ is defined by

$$D_N(x_n) = D_N^{\mathscr{D}}(x_n) = \sup_{A\in\mathscr{D}} \left| \frac{1}{N} \sum_{n=1}^{N} \chi_A(x_n) - \mu(A) \right|, \tag{2.2}$$

where χ_A is the characteristic function of the set A. Of special interest are discrepancy systems $\mathscr{D}$ which are "nice" from a topological or geometric point of view.

The first system we want to consider is the system $\mathscr{B}$ of all balls $B(p, \varepsilon)$ with $p \in G$ and $\varepsilon > 0$. We will call the corresponding discrepancy *ball discrepancy.*

We next introduce the *gasket discrepancy*: Let $\mathscr{G}$ be the system of all sets which are intersections of G with triangles the sides of which are parallel to the sides of A_0 and whose vertices are elements of G and define $D_N^{\mathscr{G}}$ as in (2.2). Furthermore, we consider the *star discrepancy* D_N^S, which is defined via the discrepancy system $\mathscr{S}$ consisting of triangles of $\mathscr{G}$ that have one side in the boundary of A_0 (see Figure 2). Finally we introduce the *elementary discrepancy* D_N^{ε}. In this case the supremum in (2.2) is extended over all elementary triangles.

In the following we establish some easy relations between these four types of discrepancy. Let $y \in G$ be arbitrary and let $\Delta_1(y)$, $\Delta_2(y)$ and $\Delta_3(y)$ be three triangles as defined in Figure 2. Note that one side of Δ_i is a part of the side e_i of the equilateral triangle A_0 $(i = 1, 2, 3)$.

We introduce three discrepancy functions of the sequence (x_n):

$$D_N^{(i)}(x_n, y) = \frac{1}{N} \sum_{n=1}^{N} \chi_{\Delta_i(y)}(x_n) - \mu(\Delta_i(y)), \quad \text{for} \quad i = 1, 2, 3. \tag{2.3}$$

Furthermore, we define the corresponding discrepancies of a given sequence (x_n)

$$D_N^{(i)}(x_n) = \sup_{y\in G} \left| D_N^{(i)}(x_n, y) \right|. \tag{2.4}$$

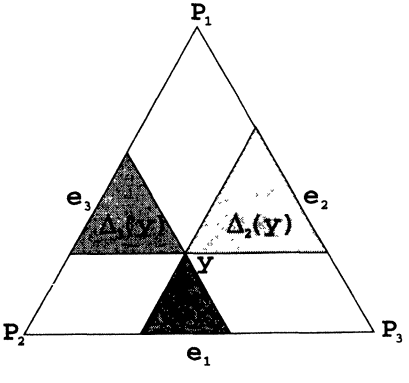


Figure 2

The measures $\mu(\Delta_i(p))$ can be computed explicitely using the digital representation of Remark 1. This is a generalization of a result given in [15].

Proposition 2. *Let p be a point in G given by its representation (k_1, k_2, k_3) as in Remark 1*

$$k_i = \sum_{l=1}^{\infty} \frac{\varepsilon_l^{(i)}}{2^l}.$$

Then the Hausdorff measure of $\Delta_1(p)$ is given by

$$\sum_{l=2}^{\infty} 3^{-l} \sum_{n=1}^{l-1} \varepsilon_n^{(2)} \varepsilon_n^{(3)} ((1+\varepsilon_{n+1}^{(2)}) \dots (1+\varepsilon_{l-1}^{(2)}) \varepsilon_l^{(2)} + (1+\varepsilon_{n+1}^{(3)}) \dots (1+\varepsilon_{l+1}^{(3)}) \varepsilon_l^{(3)}).$$

Proof. We note first that $\Delta_1(p) = \{q = (m_1, m_2, m_2) \mid m_2 \leqslant k_2, m_3 \leqslant k_3\}$. In order to compute the Hausdorff measure of this set we consider its finite approximations by elementary triangles and count their numbers. Let Φ_N be the digital function given by

$$\begin{aligned}&\Phi_N(\varepsilon_1^{(2)}, \dots, \varepsilon_N^{(2)}, \varepsilon_1^{(3)}, \dots, \varepsilon_N^{(3)}) \\ &\quad = \#\{(\delta_1^{(2)}, \dots, \delta_N^{(2)}, \delta_1^{(3)}, \dots, \delta_N^{(3)}) \mid (\delta_1^{(i)}, \dots, \delta_N^{(i)}) \leqslant (\varepsilon_1^{(i)}, \dots, \varepsilon_N^{(i)}) \\ &\quad \text{for} \quad i = 2, 3 \quad \text{and} \quad \delta_n^{(2)} + \delta_n^{(3)} > 0 \quad \text{for} \quad n \leqslant N\}.\end{aligned}$$

An easy observation shows the following recurrence relation

$$\begin{aligned}\Phi_N(\varepsilon_1^{(2)}, \dots, \varepsilon_N^{(2)}, \varepsilon_1^{(3)}, \dots, \varepsilon_N^{(3)}) &= \Phi_{N-1}(\varepsilon_2^{(2)}, \dots, \varepsilon_N^{(2)}, \varepsilon_2^{(3)}, \dots, \varepsilon_N^{(3)}) \\ &\quad + \varepsilon_1^{(2)} \varepsilon_1^{(3)} (\Phi_{N-1}(\varepsilon_2^{(2)}, \dots, \varepsilon_N^{(2)}, 1, \dots, 1) \\ &\quad + \Phi_{N-1}(1, \dots, 1, \varepsilon_2^{(3)}, \dots, \varepsilon_N^{(3)})).\end{aligned} \tag{2.15}$$

By inserting special values we get

$$\begin{aligned}\Phi_N(1, \dots, 1, \varepsilon_1, \dots, \varepsilon_N) &= (1+\varepsilon_1) \Phi_{N-1}(1, \dots, 1, \varepsilon_2, \dots, \varepsilon_N) \\ &\quad + \varepsilon_1 \Phi_{N-1}(1, \dots, 1, 1 \dots, 1).\end{aligned}$$

Inserting $\Phi_N(1, \dots, 1, 1 \dots, 1) = 3^N$ into this equation yields

$$\begin{aligned}\Phi(1, \dots, 1, \varepsilon_1, \dots, \varepsilon_N) &= (1+\varepsilon_1) \dots (1+\varepsilon_N) \\ &\quad + \sum_{k=1}^{N} (1+\varepsilon_1) \dots (1+\varepsilon_{k-1}) \varepsilon_k 3^{N-k}.\end{aligned}$$

Inserting this into (2.15) yields the following explicit formula

$$\begin{aligned}
&\Phi_N(\varepsilon_1^{(2)},\dots,\varepsilon_N^{(2)},\varepsilon_1^{(3)},\dots,\varepsilon_N^{(3)})\\
&\quad=\sum_{n=1}^{N}\varepsilon_n^{(2)}\varepsilon_n^{(3)}\Bigg((1+\varepsilon_{n+1}^{(2)})\dots(1+\varepsilon_N^{(2)})\\
&\qquad+\sum_{k=1}^{N-n}(1+\varepsilon_{n+1}^{(2)})\dots(1+\varepsilon_{n+k-1}^{(2)})\varepsilon_{n+k}^{(2)}3^{N-n-k}\\
&\qquad+(1+\varepsilon_{n+1}^{(3)})\dots(1+\varepsilon_N^{(3)})\\
&\qquad+\sum_{k=1}^{N-n}(1+\epsilon_{n+1}^{(3)})\dots(1+\varepsilon_{n+k-1}^{(3)})\varepsilon_{n+k}^{(3)}3^{N-n-k}\Bigg).
\end{aligned}$$

We use this formula and the fact that

$$\mu(\Delta_1(p))=\lim_{N\to\infty}3^{-N}\Phi_N(\varepsilon_1^{(2)},\dots,\varepsilon_N^{(2)},\varepsilon_1^{(3)},\dots,\varepsilon_N^{(3)})$$

to obtain the desired result. □

We note that any triangle with sides parallel to the sides of A_0 can be represented as set-theoretic sum or difference of at most six triangles of types $\Delta_1,\Delta_2,\Delta_3$. Thus we have

$$D_N^{\mathscr{S}}(x_n)\leqslant D_N^{\mathscr{G}}(x_n)\leqslant 6D_N^{\mathscr{S}}(x_n). \tag{2.5}$$

In order to compare the elementary and the gasket discrepancy, we have to estimate how many elementary traingles are necessary to aproximate a given gasket triangle T contained in $\mathscr{G}$. For this purpose we define P_n as the union of all elementary triangles of level n, which are contained in T. Observe now, that $P_{n+1}\backslash P_n$ consists of elementary triangles of level $n+1$ which are contained in elementary triangles of level n which intersect the boundary of T. The number of elementary triangles of level n, which intersect the boundary of T is at most $3\cdot 2^n$. Thus T can be approximated by a union of $6\cdot 2^n$ elementary triangles of level $\leqslant n$ with an error of at most $6\cdot(\frac{2}{3})^n$. From this we derive the inequality

$$D_N^{\mathscr{G}}\leqslant 6\cdot 2^m\,(D_N^{\varepsilon}+3^{-m})$$

for any m. By inserting $m=[\log_3\frac{1}{D_N^{\varepsilon}}]$ we get

$$D_N^{\varepsilon}\leqslant D_N^{\mathscr{G}}\leqslant 24(D_N^{\varepsilon})^{\frac{\alpha-1}{\alpha}}, \tag{2.6}$$

where the left inequality is obvious.

Finally, we compare the ball discrepany with the elementary discrepancy. For this purpose we observe that any elementary triangle Δ of sidelength 2^{-k} can be exhausted by balls by the following procedure: Take the midpoint p of an edge of Δ and consider the ball $B(p,2^{-k-1})$. Then $\Delta\backslash B(p,2^{-k-1})$ is an elementary triangle of sidelength 2^{-k-1} and we can iterate the procedure. Thus we need K balls to approximate an elementary triangle with accuracy $\mu(\Delta)3^{-K}$. This yields

$$D_N^{\varepsilon}\leqslant K\,D_N^{\mathscr{B}}+3^{-K}$$

for any integer $K \geqslant 0$. We set $K = [\log_3 \frac{1}{D_N^{\mathscr{B}}}]$ to obtain

$$D_N^{\varepsilon} \leqslant D_N^{\mathscr{B}} \left(\log_3 \frac{1}{D_N^{\mathscr{B}}} + 3 \right). \tag{2.7}$$

In order to obtain an inequality in the opposite direction we have to exhaust a given ball by elementary triangles. The procedure to do this is quite the same as in the proof of (2.6) and yields

$$D_N^{\mathscr{B}} \leqslant 72 \, (D_N^{\varepsilon})^{\frac{\alpha-1}{\alpha}}. \tag{2.8}$$

Remark 3. For any triangle $\Delta \in \mathscr{G}$ and arbitrary $k \in \mathbb{N}$ there exist two triangles Δ' and Δ'' with vertices in V_k such that $\Delta' \subseteq \Delta \subseteq \Delta''$ and $\mu(\Delta'' \backslash \Delta') = \mathcal{O}((\frac{2}{3})^k)$. Since V_k contains only finitely many points and $\mu(\Delta)$ is a continuous function of the vertices of Δ, compactness and uniform continuity immediately yield

$$\lim_{N \to \infty} D_N^{\mathscr{G}}(x_n) = 0$$

if and only if (x_n) is uniformly distributed in G. The inequalities (2.5)–(2.8) imply that this holds for all the notions of discrepany discussed above.

2.3. L^p-Discrepancy. We introduce the L^p-discrepancy of (x_n) for arbitrary $p \geqslant 1$:

$$L_N^{(p)} = \left(\int_0^1 \int_G \left| \frac{1}{N} \sum_{n=1}^{N} \chi_{B(y,r)}(x_n) - \mu(B(y,r)) \right|^p d\mu(y) \, dr \right)^{\frac{1}{p}}.$$

Obviously $L_N^{(p)} \leqslant D_N^{\mathscr{B}}(x_n)$. In order to prove an opposite inequality we sketch a procedure used in [28] to derive a general inequality of this type on compact metric spaces endowed with a Borel probability measure. This is a more general version of an inequality between the usual discrepancy on $[0, 1)^s$ and the corresponding L_p-discrepancy proved in [22].

Theorem 1. *Let (X, d) be a compact metric space and λ and ζ be two Borel probability measures on X, where λ satisfies the following additional conditions*

$$|\lambda(B(x, r_1)) - \lambda(B(x, r_2))| \leqslant L_1 |r_2 - r_1|^{\beta},$$
$$|\lambda(B(x_1, r)) - \lambda(B(x_2, r))| \leqslant L_2 d(x_1, x_2)^{\beta},$$
$$\lambda(B(x, r)) \geqslant L_0 r^s.$$

Then the discrepancy function $D(y, r) = \zeta(B(y, r)) - \lambda(B(y, r))$ satisfies the following inequality

$$\int_X \int_0^{\vartheta} \varphi(|D(y, r)|) \, dr \, d\lambda(y) \geqslant c \, \| D \|_{\infty}^{\frac{s+1}{\beta}} \, \varphi\left(\frac{1}{6} \, \| D \|_{\infty} \right)$$

for any increasing function φ on $[0, 1]$, where c is a positive constant only depending on X, L_0, L_2, β and s. ϑ denotes the diameter of X.

Corollary 1. *Let $L_N^{(p)}$ denote the L^p-discrepancy defined in (2.9). Then the following inequality holds*

$$L_N^{(p)} \geqslant c\,(D_N^{\mathscr{B}})^{\frac{\alpha+1}{(\alpha-1)p}+1}$$

for a suitable positive constant c depending only on p. Thus $\lim_{N\to\infty} L_N^{(p)}(x_n) = 0$ *is equivalent with the uniform distribution of the sequence* (x_n).

Sketch Proof of Theorem 1. Let $D = \|D\|_\infty$. Then for any $\varepsilon > 0$ there exists a pair $(x_0, r) \in X \times \mathbb{R}^+$ such that $|D(x_0, r)| > D - \varepsilon$. We set $\vartheta(x_0) = \sup_{y\in X} d(x_0, y)$ and show that

$$m(a) := \sup_{r\in[a,\vartheta(x_0)-a]} |D(x_0, r)| \geqslant \frac{1}{3} D - \frac{2}{3} L_1 a^\beta$$

for arbitrary $0 < a < \frac{1}{2}\vartheta(x_0)$. We choose

$$a = \min\left(\left(\frac{1}{2}\vartheta(x_0)\right)^{\frac{1}{\beta}}, \left(4L_1 + 6\frac{L_1 + L_2}{2^\beta}^{-\frac{1}{\beta}}, \frac{1}{2}\vartheta(x_0)\right) D^{\frac{1}{\beta}}\right)$$

and take an $r_0 \in [a, \vartheta(x_0) - a]$ such that

$$|D(x_0, r)| \geqslant \frac{1}{3} D - \frac{2}{3} L_1 a^\beta - \varepsilon$$

for arbitrary $\varepsilon > 0$.

For $D(x_0, r_0) > 0$ we have $B(y, r) \supseteq B(x_0, r_0)$ for every $y \in B(x_0, \frac{a}{4})$ and every $r \in [r_0 + \frac{a}{2}, r_0 + a] =: I_a$. Thus we have (using the monotonicity of $\zeta(B(y, r))$)

$$\begin{aligned}\zeta(B(y, r)) - \lambda(B(y, r)) &\geqslant \zeta(B(x_0, r_0)) - \lambda(B(y, r_0)) - L_1(r - r_0)^\beta \\ &\geqslant \zeta(B(x_0, r_0)) - \lambda(B(x_0, r_0)) - L_1(r - r_0)^\beta - L_2\left(\frac{a}{4}\right)^\beta \\ &\geqslant D(x_0, r_0) - (L_1 + L_2)\left(\frac{a}{4}\right)^\beta,\end{aligned}$$

and by the choice of x_0, r_0 and a we derive

$$|D(y, r)| \geqslant \frac{1}{6} D. \tag{2.10}$$

For $D(x_0, r_0) < 0$ we have $B(y, r) \supseteq B(x_0, r_0)$ for every $y \in B(x_0, \frac{a}{4})$ and every $r \in [r_0 - \frac{a}{2}, r_0 \frac{a}{4}] =: I_a$ Thus we have $|D(y, r)| \geqslant \frac{1}{6} D$.

Combining (2.10) and the last condition on the measure λ yields

$$\begin{aligned}\int_X \int_0^\vartheta \varphi(|D(y, r)|)\, dr\, d\lambda(y) &\geqslant \int_{B(x_0,\frac{a}{4})} \int_{I_a} \varphi\left(\frac{1}{6} D\right) dr\, d\lambda(y) \\ &\geqslant L_0\left(\frac{a}{4}\right)^{s+1} \varphi\left(\frac{1}{6} D\right),\end{aligned}$$

which (by the choice of a) gives the desired result. □

Proof of Corollary 1. In order to prove the Corollary we notice that

$$\zeta(E) = \frac{1}{N}\sum_{n=1}^{N} \chi_E(x_n)$$

is a Borel measure, and $\lambda = \mu$ satisfies the conditions of Theorem 1 with $\beta = \alpha - 1$ and $s = \alpha$ and some suitable constants L_0, L_1, L_2. □

2.4 Special Sequences and Irregularities of Distribution. Obviously we have

$$\frac{1}{N} \leqslant D_N(x_n) \leqslant 1 \tag{2.11}$$

for the four discrepancy systems under consideration in Section 2.2. We note here that in the case of the elementary discrepancy it is possible to find sequences (x_n) such that $N D_N^{\varepsilon}(x_n)$ is bounded. Such sequences can be compared with the well-known net-sequences in the unit cube. We remark here that these net-sequences have recently been used for various applications in quasi Monte Carlo methods (cf. [21]). For the other two notions of discrepancy there is the phenomenon of *irregularities of distribution.*

In the following we want to describe a gasket analogon of the well-known van der Corput sequence $\gamma = (\gamma_n)$. For this purpose we note that the digital expansion described in Remark 1 can also be given as follows: let $\delta_l \in \{0, 1, 2\}$ be the index i mod 3 such that $\varepsilon_l^{(i)} = 0$. Then every point in the gasket can be encoded (not necessarily uniquely) as an infinite triadic string (cf. [7]). For defining the sequence γ we expand every integer n in triadic expansion

$$n = \sum_{l=0}^{L} \delta_{l+1}(n) 3^l$$

and define γ_n as the point encoded by $(\delta_1, \delta_2, \ldots, \delta_L, 0^{\infty})$.

Remark 4. Note that any elementary triangle of level k corresponds to a residue class mod 3^k. Thus the elementary discrepancy is $\mathcal{O}(\frac{1}{N})$. By (2.5), (2.6) and (2.8) we immediately derive

$$D_N^{\mathscr{S}}(\gamma),\ D_N^{\mathscr{G}}(\gamma),\ D_N^{\mathscr{B}}(\gamma) = \mathcal{O}\left(\frac{1}{N^{\frac{\alpha-1}{\alpha}}}\right). \tag{2.12}$$

By a standard technique due to W. PHILIPP [23] the average rate of growth of the discrepancy of an arbitrary sequence $x_n \in G$ can be determined.

Proposition 3. *The following law of iterated logarithm holds for* $\mathscr{D} = \mathscr{G}, \mathscr{S}, \mathscr{B}$

$$\limsup_{N\to\infty} \frac{D_N^{\mathscr{D}}(x_n)\sqrt{N}}{\sqrt{2\log\log N}} = \frac{1}{2}$$

for almost all (with respect to the infinite product measure generated by μ*) sequences on G.*

Remark 5. For the elementary discrepancy a similar law of the iterated logarithm can be shown by much simpler arguments; the constant $\frac{1}{2}$ has to be replaced by $\frac{\sqrt{2}}{3}$.

The example of van der Corput sequence shows that there is a gap between the lower bound $\frac{1}{N}$ and the upper bound $\mathcal{O}\left(\frac{1}{N^{\frac{\alpha-1}{\alpha}}}\right)$. A simple application of W. SCHMIDT's theorem on irregularities of distribution [24] yields

Proposition 4. *Let x_n be a sequence in G. Then*

$$D_N^{\mathscr{S}}(x_n) \geqslant c\frac{\log N}{N}$$

holds for infinitely many N (where $c > 0$ denotes an absolute constant).

Proof. For the sequence x_n we consider the sequence $k_1(x_n) \in [0, 1]$ (cf. Remark 1). Note that for uniformly distributed $x_n, k_1(x_n)$ has the asymptotic distribution function $F(x) = \mu(\Delta_2(p))$, where p is given by $k_1(p) = x$ and $k_3(p) = 1$, see Proposition 2 and [15]. Clearly this function is continuous and strictly increasing. Applying Schmidt's lower bound to the sequence $F^{-1}(k_1(x_n))$ yields

$$D\mathscr{S}_N(x_n) \geqslant D_N^*(F^{-1}(k_1(x_n))) \geqslant \frac{1}{66\log 4}\frac{\log N}{N},$$

where D_N^* denotes the usual star-discrepancy in the unit interval. □

Remark 6. Clearly, this is a very weak bound, since we have used only very special gasket triangles to derive this inequality. It remains as an interesting open problem to improve this lower bound. Since we have no natural group structure it seems to be very hard to apply Beck's Fourier transform approach to the gasket.

Concluding this section we present a probabilistic aproach for constructing a set Γ_N of N points in G with small discrepancy.

Theorem 2. *For every positive integer $N>1$ there exists a point set Γ_N consisting of N points such that*

$$D_N^{\mathscr{S}}(\Gamma_N) \leqslant cN^{\frac{1}{2\alpha}-1}(\log N)^{\frac{1}{2}},$$

where $C > 0$ is an absolute constant.

Proof. In order to prove this theorem we use BECK's probabilistic approach [3]. We define N sets $Q_1, \ldots, Q_N$ as follows: let k be the uniquely determined integer such that $3^{k-1} < N \leqslant 3^k$ and take $Q_1, \ldots, Q_{N_1}$ as elementary triangles of level k, $Q_{N_1+1}, \ldots, Q_{N_1+N_2}$ as the union of two elementary triangles of level k and $Q_{N_1+N_2+1}, \ldots, Q_{N_1+N_2+N_3}$ as the union of three elementary triangles of level k, where $N_1 + N_2 + N_3 = N$ and $N_1 + 2N_2 + 3N_3 = 3^k$. We choose $N_3 = \max(3^k - 2N, 0)$, $N_2 = 3^k - N - 2N_3$ and $N_1 = 2N - 3^k + N_3$ (these values are all non-negative).

Let $Z_1, \dots, Z_N$ be random variables such that Z_n is uniformly distributed on Q_n (with respect to μ), $n = 1, \dots, N$. We observe that

$$\mathscr{B}_l = \left\{ B(x, r) \,|\, x \in V_l,\ r = \frac{m}{2^l},\ m = 0, \dots, 2^l \right\}$$

has the property that for any ball $B(x, r)$ there exist two balls $B', B'' \in \mathscr{B}_l$ such that $B' \subseteq B \subseteq B''$ and $\mu(B'' \setminus B') \ll \left(\frac{2}{3}\right)^l$. Furthermore $\#\mathscr{B}_l \leqslant 2 \cdot 6^l$. Now we compute the expected value of the random variables $X_n(S) = \chi_S(Z_n)$ for a ball $S \in \mathscr{B}_l$ for some l, which will be chosen later. Clearly $\mathbb{E}X_n(S) = \dfrac{\mu(S \cap Q_n)}{\mu(Q_n)}$. Thus we obtain that $X_n(S) \equiv \mathbb{E}X_n(S)$ if $S \cap Q_n = \emptyset$ or $S \cap Q_n = Q_n$. As in the proof of (2.6) we have

$$\#\{n \,|\, \emptyset \neq S \cup Q_n \neq Q_n\} \ll N^{\log_3 2},$$

where the implied constant is absolute. By [3], Lemma 8.2, we derive

$$\text{Prob}\left(\left| \sum_{n=1}^{N} (X_n(S) - \mathbb{E}X_n(S)) \right| \geqslant \gamma \right) \leqslant 2 \exp\left(-C\gamma^2 N^{-\log_3 2}\right).$$

Setting $\gamma = C' N^{\frac{1}{2}\log_3 2 - 1} \sqrt{\log N}$ and $l = \left\lfloor \dfrac{2\log 3 - \log 2}{2(\log 3 - \log 2)\log 3} \log N \right\rfloor$, a suitable choice of C' yields

$$\text{Prob}\left(\left| \frac{1}{N} \sum_{n=1}^{N} \chi_S(Z_n) - \mu(S) \right| \geqslant C' N^{\frac{1}{2}\log_3 2 - 1} \sqrt{\log N}, \quad \text{for some } S \in \mathscr{B}_l \right) \leqslant \frac{1}{2}.$$

Thus there exists a point set Γ_N satisfying the bound given in Theorem 2. □

Remark 7. The main ingredient of the proof is the approximation of the discrepancy system by a finite system. Thus an analogous theorem can be proved for $D^{\mathscr{S}}$ and $D^{\mathscr{G}}$.

2.5. Uniform Distribution of Curves. We recall here the definition of discrepancy for continuous functions $x(t)$ on the gasket:

$$D_T(x(t)) = D_T^{\mathscr{D}}(x(t)) = \sup_{A \in \mathscr{D}} \left| \frac{1}{T} \int_0^T \chi_A(x(t))dt - \mu(A) \right|. \tag{2.15}$$

If we interpret $x(t)$ as the motion of a particle on the gasket, the discrepancy can be considered as the deviation of the mean with respect to time and the spatial mean. The motion is called equidistributed if $\lim_{T\to\infty} D_T(x(t)) = 0$. The theory of uniform distribution for continuous functions was developed in a series of papers by HLAWKA and KUIPERS et al. (cf. [19]). We consider all discrepancy systems $\mathscr{D}$ introduced above. An application of a general result in [9] yields

Proposition 5. *Let $\mathscr{D}$ be one discrepancy systems considered in Section 2.2 and $x(t)$ a continuous function $\mathbb{R}_0^+ \to G$ with finite arclength $s(T)$ and $\lim_{T\to\infty} s(T) = \infty$. Then there exists a constant $c(\mathscr{D})>0$ such that*

$$D_T^{\mathscr{D}}(x(t)) \geqslant c(\mathscr{D}) \left(\frac{1}{s(T)} \right)^{\frac{\alpha}{\alpha - 1}} \quad \text{for} \quad T \geqslant T_0.$$

Remark 8. Note that the arclength of a continuous function $x(t)$ is defined by

$$s(T) = \sup \sum_{n=0}^{N-1} d(x(t_n), x(t_{n+1})),$$

where the supremum is extended over all partitions $0 = t_0 < t_1 < \cdots < t_N = T$.

Remark 9. The proof is verbally the same as the proof of Theorem 1 in [9]. We only note, that in the technical condition (2.2) in [9] it is not necessary to take balls $B(x, r)$ and $B(y, R)$ with the same center $x = y$.

Remark 10. Obviously the general inequalities (2.5–2.8) remain valid for the continuous versions of the different kinds of discrepancies.

3. A Uniform Law of Iterated Logarithm for Brownian Motion on G

Our aim is the generalization of the law of iterated logarithm (2.13) to the trajectories of Brownian motion. For Brownian motion on manifolds similar results can be found in [6] and [5].

In the introduction we have defined Brownian motion on G as a limit process of a discrete random walk. Barlow and Perkins [2] have shown for the corresponding process on the infinite gasket $\tilde{G}$ that this is a symmetric Markow process with jointly continuous transition densities $\tilde{p}(t, x, y)$ on $[0, \infty] \times \tilde{G} \times \tilde{G}$. Furthermore $t \mapsto \tilde{p}(t, x, y)$ is C^∞, and for all $t>0$, $x, y \in \tilde{G}$ the following estimate holds

$$c_1 t^{\frac{\log 3}{\log 5}} \exp\left(-c_2 d(x, y)^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}}\right) \leqslant \tilde{p}(t, x, y)$$
$$\leqslant c_3 t^{\frac{\log 3}{\log 5}} \exp\left(-c_4 d(x, y)^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}}\right)$$

with suitable positive constants c_1, c_2, c_3, c_4.

As described in the introduction, the Brownian motion on G is obtained by factorizing $\tilde{G}$ modulo a suitable equivalence relation ρ. Thus we can decompose $\tilde{G}$ as $\bigcup_k T_k G$, where the T_k are translations, which originate from the definition of ρ. Since

$$p(t, x, y) = \sum_k \tilde{p}(t, x, T_k y),$$

we have to combine the estimates (3.1) in order to give upper and lower bounds. It follows from [15] that the number of copies of G in $\tilde{G}$ whose points p have a distance $l \leqslant d(p, O) \leqslant l + 1$ is $2^{s(l)+1}$, where $s(l)$ denotes the binary sum-of-digits function. Therefore we get the bounds

$$\gamma_t = 2c_1 t^{\frac{\log 3}{\log 5}} \sum_{l=0}^{\infty} 2^{s(l)} \exp\left(-c_2 (l+1)^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}}\right)$$
$$\leqslant p(t, x, y) \leqslant \gamma_t' \qquad (3.2)$$
$$= 2c_3 t^{\frac{\log 3}{\log 5}} \sum_{l=0}^{\infty} 2^{s(l)} \exp\left(-c_4 l^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}}\right).$$

We want to prove that the functions γ_t γ_t' are bounded from above and below by positive constants for $t \geqslant 1$. In order to prove this and to estimate γ_t from below we apply partial summation to the first sum in (3.2). This yields

$$\gamma_t = 2c_1 t^{\frac{\log 3}{\log 5}} \sum_{l=1}^{\infty} \sum_{k=0}^{l-1} 2^{s(k)} \times \left(\exp\left(-c_2 l^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}} \right) - \exp\left(-c_2 (l+1)^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}} \right) \right).$$

From [16], [12] and [15] we know that

$$\frac{1}{2} N^{\log_2 3} \leqslant \sum_{n=0}^{N-1} 2^{s(n)} \leqslant N^{\log_2 3},$$

which implies

$$\begin{aligned} \gamma_t &\geqslant c_1 t^{\frac{\log 3}{\log 5}} \sum_{l=1}^{\infty} l^{\log_2 3} \times \\ &\quad \times \left(\exp\left(-c_2 l^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}} \right) - \exp\left(-c_2 (l+1)^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}} \right) \right) \\ &\geqslant c_1 t^{\frac{\log 3}{\log 5}} \sum_{l=1}^{\infty} l^{\log_2 3 - 1} \exp\left(-c_2 l^{\frac{\log 5}{\log 5 - \log 2}} t^{-\frac{\log 2}{\log 5 - \log 2}} \right). \end{aligned} \tag{3.3}$$

The last sum has been studied by Ramanujan (cf. [4]). For estimating this sum we apply the Mellin transform to the sum

$$f(u) = \sum_{n=1}^{\infty} n^{\log_2 3 - 1} \exp\left(-c_2 n^{\frac{\log 5}{\log 5 - \log 2}} u \right),$$

which yields

$$f^*(s) = \int_0^{\infty} f(u) u^{s-1} \, du = \zeta\left(\frac{\log 5}{\log 5 - \log 2} s - \log_2 3 + 1 \right) \Gamma(s).$$

By the well-known correspondence between the singularities of the transform and the asymptotic behaviour of the function we obtain $f(u) \sim c_5 u^{\frac{\log 5 (\log 2 - \log 2)}{\log 2 \log 5}}$ with some explicit constant c_5. Inserting this into (3.3) ($u = t^{-\frac{\log 5}{\log 5 - \log 2}}$) and applying the same procedure to γ_t' we derive

$$0 < c_6 \leqslant p(t, x, y) \leqslant c_7 \quad \text{for } t \geqslant 1. \tag{3.4}$$

Now we introduce a gasket analogon of the classical Wiener measure. Let $\mathscr{C}_w$ be the set of all continuous curves in G starting in a given point $w \in G$. Then for fixed $0 < t_1 < t_2 < \cdots < t_n$ and a Borel set $E \subseteq G^n$ the corresponding Wiener measure is given by

$$\mu_w(\{x \in \mathscr{C}_w \mid (x(t_1), x(t_2), \ldots, x(t_n)) \in E\}) = \int_E p(t_n - t_{n-1}, x_n, x_{n-1}) \ldots p(t_2 - t_1, x_2, x_1) p(t_1, x_1, w) \, d\mu(x_1) \ldots d\mu(x_n).$$

Proposition 6. *Let* $P(t, x, S) = \int_S p(t, x, y)d\mu(y)$ *be the transition probabilities of Brownian motion on G. Then*

$$|p(t, x, S) - \mu(S)| \leqslant Ae^{-at}, \tag{3.6}$$

$$||P(t, x, dy) - d\mu(y)|| \leqslant 2Ae^{-at}, \tag{3.7}$$

with positive constants A, a, independent of x and the measurable set S. ($\|\cdot\|$ *denotes the uniform norm with respect to x.*)

Proof. From [29], p. 197 if follows that there exists a measure ν^t such that

$$|P(nt, x, S) - \nu^t(S)| \leqslant (1 - \gamma_t)^{n-1} \tag{3.8}$$

for $t > 0$. Applying the Chapman-Kolmogorov equation we obtain

$$\int_G P(nt, y, S)P(ns, x, dy) = P(n(s + t), x, S). \tag{3.9}$$

For $n \to \infty$, $P(nt, x, S)$ converges to $\nu^t(S)$ uniformly in $x \in G$ by (3.8). Thus the integral in (3.9) converges to $\nu^t(S)$, whereas the right hand side converges to $\nu^{s+t}(S)$. Hence $\nu = \nu^t$ is independent of t.

Next we identify ν as the Hausdorff measure μ. As $p(t, x, y)$ is bounded by an absolute constant ν is absolutely continuous with respect to μ. Let f be the density of ν with respect to μ. By the above arguments f is essentially bounded. Let $q(t, x, y)$ be the transition density with respect to the measure ν, i.e. $q(t, x, y) = p(t, x, y)f(y)$. We want to show that $f \equiv 1$ and proceed indirectly, assuming that f is non-constant. Let now C be the essential supremum of f and set

$$A_\varepsilon = \{x \,|\, f(x) < C - \varepsilon\}.$$

For $\varepsilon > 0$ we have

$$\begin{aligned} &q(t, x, y_0) - q(t, y_0, x) \\ &\quad = p(t, x, y_0)(f(y_0) - f(x)) > 0 \quad \text{for} \quad x \in A_\varepsilon \quad \text{and} \quad y_0 \in A_\varepsilon^C. \end{aligned}$$

Next we choose ε so small that $\mu(A_\varepsilon) > 0$. Thus we obtain

$$\int_{A_\varepsilon^C} (q(t, x, y_0) - q(t, y_0, x))d\nu(x) \leqslant 0$$

and

$$\int_{A\varepsilon} (q(t, x, y_0) - q(t, y_0, x))d\nu(x) > 0.$$

Since $\mu(A_\varepsilon^C) \to 0$ for $\varepsilon \to 0$ we have

$$\int_G (q(t, x, y_0) - q(t, y_0, x))d\nu(x) > 0.$$

On the other hand the integral on the left hand side is 0, which is a contradiction. Thus $f \equiv 1$ and the two measures μ and ν are equal.

It follows from (3.4) and (3.8) that

$$|P(t, x, S) - \mu(S)| \leqslant e^{-at+a}$$

with $a = -\log(1 - \gamma_1)$. Setting $A = e^a$ yields (3.6). The estimate (3.7) is an immediate consequence of the Hahn decomposition theorem. □

Let I be a time interval and $\mathscr{F}_I$ the σ-algebra generated by events in I. A process is called φ-mixing if

$$|P(E_2 \mid E_1) - P(E_2)| \leqslant \varphi(t), \tag{3.10}$$

and ψ-mixing if

$$|P(E_2|E_1) - P(E_2)| \leqslant P(E_2)\psi(t) \tag{3.11}$$

for events E_1, E_2, with E_1 being $\mathscr{F}_{[0,s]}$-measurable, E_2 being $\mathscr{F}_{[s+t,\infty]}$-measurable and $\varphi(t) \to 0$, $\psi(t) \to 0$ as $t \to \infty$.

Proposition 7. *The Brownian motion on G has the ψ-mixing property with $\psi(t) = Ke^{-at}$ for $t \geqslant 1$. It satisfies the φ-mixing property for $t \geqslant 0$ with $\varphi(t) = K'e^{-at}$.*

Proof. By the Chapman-Kolmogorov equation and (3.7) we obtain

$$|p(s + t, x, y) - 1| \leqslant 2\gamma_s' Ae^{-at}.$$

Using the Markov property, a simple computation yields for $t \geqslant 1$

$$|P(E_2 \mid E_1) - P(E_2)| \leqslant P(E_2)Ke^{-at},$$

where $K = \dfrac{4\gamma'_{1/2}A}{\inf_{x,y\in G} p(1, x, y)}$. Since $|P(E_2 \mid E_1) - P(E_2)| \leqslant 1$, the second assertion follows immediately. □

Next we approach our main result, a uniform law of the iterated logarithm. It follows from a general result of W. PHILIPP [23] and is a gasket analogon of Theorem 4 in [5]. As usual we will use the notation $E(f) = \int_G f(x)d\mu(x)$.

Theorem 4. *Let θ be a positive integer and let $\mathscr{A}_\theta$ be a family of real-valued uniformly bounded measurable functions on G such that $1 \leqslant \#\mathscr{A}_\theta \leqslant e^{\theta k_1}$ with some constant k_1. Let $\mathscr{A}$ be the set of all functions on G, uniformly bounded by 1 and having the following approximation property:*

For all $f \in \mathscr{A}$ there exists a sequence of functions $h_\theta, \bar{h}_\theta \in \mathscr{A}$ such that for all positive integers L

$$\sum_{\theta=1}^{L} h_\theta \leqslant f \leqslant \sum_{\theta=1}^{L} \bar{h}_\theta,$$

$$\left\| \sum_{\theta=1}^{L} (\bar{h}_\theta - h_\theta) \right\|_1 \leqslant e^{-k_2 L},$$

where k_2 is a positive constant.

Let w be a given point in G. Then for arbitrary $\varepsilon > 0$ and for μ_w-almost all curves $x(t)$ in $\mathscr{C}_w$ there exists a $T_0 > 0$ such that

$$\left| \frac{\int_0^T f(x(t))dt - TE(f)}{\sqrt{2T \log\log T}} \right| < \sigma(f) + \varepsilon$$

for all $f \in \mathscr{A}$ and all $T > T_0$. Furthermore, for μ_w-almost all curves $x(t)$ in $\mathscr{C}_w$

$$\limsup_{T\to\infty} \left| \frac{\int_0^T f(x(t))dt - TE(f)}{\sqrt{2T \log\log T}} \right| = \sigma(f)$$

holds uniformly for all $f \in \mathscr{A}$.

Sketch of the Proof. We use the notation $X_n = \int_{n-1}^n f(x(t))dt$. Since the functions in $\mathscr{A}$ are uniformly bounded we may consider integer values for T only. Furthermore, w. l. o. g. we may assume that $\int_G f d\mu = 0$, $E(X_n(f)) = 0$ for all positive integers n. The theorem is an immediate consequence of theorems 1.3.1, 1.3.2 in [23] and Proposition 7, if we can verify the following conditions (3.13), (3.14), (3.15):

$$E\left(\sum_{n=1}^N X_n(f)\right)^2 = N\sigma^2(f) + O(N), \tag{3.13}$$

$$\sigma^2(f) = O(1), \tag{3.14}$$

$$E\left(\sum_{n=1}^N X_n(h_\theta)\right)^2 = O(Ne^{-k_1\theta}), \tag{3.15}$$

where the O-constants are absolute ones.

(3.13) and (3.14) follow from Proposition 6 and (3.12) after some standard calculations and estimates, see [5]. (3.15) is a direct consequence of the approximation property stated in the theorem. □

Remark 11. As in [5, Theorem 5] it can be shown that $0 < \sigma(f) < \infty$ for all non-zero $f \in L^\infty$.

Corollary 2. *The following law of iterated logarithm holds for the discrepancy systems $\mathscr{D} = \mathscr{G}, \mathscr{S}, \mathscr{B}$*

$$\limsup_{N\to\infty} \frac{D_N^{\mathscr{D}}(x(t))\sqrt{N}}{\sqrt{2\log\log N}} = \sigma$$

for μ_w-almost all functions $x(t) \in C_w$, where w is a given point in G and σ is some positive constant.

Proof. As in Section 2.2 the exponential approximation property can be derived for discrepancy systems. □

Acknowledgements. We are indebted to Martin Blümlinger and Wolfgang Woess for valuable discussions and for providing some recent references.

References

[1] BARLOW MT, BASS RF (1993) Coupling and Harnack inequalities for Sierpiński carpets. Bull Amer Math Soc **29**: 208–212

[2] BARLOW MT, PERKINS EA (1988) Brownian motion on the Sierpiński gasket. Probab Th Re Fields **79**: 543–623

[3] BECK J, CHEN W (1987) Irregularities of Distribution. Cambridge: Univ Press

[4] BERNDT B (1989) Ramanujan's Notebooks, Part II. Berlin Heidelberg New York: Springer

[5] BLÜMLINGER M (1989) Sample Path Properties of Diffusion Processes on Compact Manifolds. In: HLAWKA E and TICHY RE (eds.) Number-Theoretic Analysis Lect Notes Math 1452, pp. 6–19

[6] BLÜMLINGER M, DRMOTA M, TICHY RF (1989) A uniform law of the iterated logarithm for Brownian motion on compact Riemannian manifolds. Math Z **201**: 495–507

[7] CUOCO AA (1991) Visualizing the p-adic integers. Amer Math Monthly **98**: 355–364

[8] DOBRUSHIN RL, KUSUOKA S (1993) Statistical Mechanics and Fractals. Lect Notes Math 1567

[9] DRMOTA M, TICHY RF (1988) C-uniform distribution on compact metric spaces. J Math Anal Appl **129**: 284–292

[10] ELWORTHY KD, IKEDA N (1993) Asymptotic Problems in Probability Theory: Stochastic Model and Diffusions on Fractals. Pitman Res Notes Math 283

[11] FALCONER KJ (1985) The Geometry of Fractals Sets. Cambridge: Univ Press

[12] FLAJOLET P, GRABNER PJ, KRISCHENHOFER P, PRODINGER H, TICHY RF (1994) Mellin transforms and asymptotics: digital sums. Theor Comput Sci **123**: 291–314

[13] FLEISCHER W (1971) Das Wienersche Maß einer gewissen Menge von Vektorfunktionen, Mh Math **75**: 193–197

[14] FUKUSHIMA M, SHIMA T (1992) On a spectral analysis for the Sierpiński gasket. Potential Analysis **1**: 1–35

[15] GRABNER PJ (1993) Completely q-multiplicative functions: the Mellin transform approach. Acta Arith **65**: 85–96

[16] HARBORTH H (1977) Number of odd binomial coefficients. Proc Amer Math Soc **62**: 19–22

[17] HLAWKA E (1960) Über C-Gleichverteilung. Ann Math Pure Appl **49**: 311–326

[18] KIGAMI J (1989): A harmonic calculus on the Sierpinski spaces. Japan J Appl Math **6**: 259–290

[19] KUIPERS L, NEIDERREITER H (1974) Uniform Distribution of Sequences. New York: Wiley

[20] LINDSTRØM T (1990) Brownian motion on nested fractals. Memoirs Amer Math Soc **83**

[21] NIEDERREITER H (1992) Random Number Generation and Quasi-Monte Carlo Methods. SIAM Lecture Notes 63. Philadelphia: SIAM

[22] NIEDERREITER H, TICHY RF, TURNWALD G (1990) An inequality for differences of distribution functions. Arch Math **54**: 166–172

[23] PHILIPP W (1971) Mixing Sequences of Random Variables in Probabilistic Number Theory. Memoirs Amer Math Soc **114**

[24] SCHMIDT W (1972) Irregularities of distribution VII. Acta Arith **21**: 45–50

[25] SHIMA T (1991) On eigenvalue problems for the random walks on the Sierpiński pregasket, Japan J Indust Appl Math **8**: 127–141

[26] SIERPIŃSKI W (1915) Sur une courbe dont tout point est un point de ramification. CR Acad Sci Paris **160**: 302–305

[27] STACKELBERG O (1971) A uniform law of the iterated logarithm for functions C-uniformly distributed mod 1. Indiana Univ Math J **21**: 515–528

[28] TICHY RF (1991) A general inequality with applications to the discrepancy of sequences. Grazer Math Ber **313**: 65–72

[29] DOOB J (1953) Stochastic processes. New York: Wiley

P. J. GRABNER and R. F. TICHY
Institut für Mathematik
Technische Universität Graz
Steyrergasse 30
8010 Graz
Austria
e-mail: grabner@weyl.math.tu-graz.ac.at
tichy@weyl.math.tu-graz.ac.at

Mh. Math. 125, 165–172 (1998)

Monatshefte für
Mathematik

Printed in Austria

Decompositions of Factor Maps Involving Bi-closing Maps

By

Paul Trow, Memphis, TN

(Received 23 January 1996)

Abstract. We describe all possible decompositions of a finite-to-one factor map $\theta : \Sigma_A \to S$, from an irreducible shift of finite type onto a sofic shift, into two maps $\theta = \gamma\phi$, such that the range of ϕ is a shift of finite type, and γ is bi-closing. We also give necessary and sufficient conditions for θ to be almost topologically conjugate over S to a bi-closing map.

Introduction

In [1], ADLER, KITCHENS and MARCUS classified finite-to-one factor maps between shifts of finite type, up to certain equivalence relations. They also gave an example of a factor map θ, of degree d, which is impossible to decompose into two maps, $\theta = \gamma\phi$, such that γ is d-to-1 everywhere ([1, Example 2, p. 492]). In this paper, we describe all possible decompositions of a finite-to-one factor map $\theta : \Sigma_A \to S$, from a shift of finite type onto a sofic shift, into two maps $\phi : \Sigma_A \to \Sigma_B$ followed by $\gamma : \Sigma_B \to S$, such that Σ_B is a shift of finite type and γ is bi-closing. When S is a shift of finite type, the bi-closing condition is equivalent to γ being k-to-1 everywhere for some k ([9, Theorem 7.3] or [4, Prop. 1]). We show that for any such decomposition, the map γ must be conjugate over S to one of a finite collection of maps, which can be explicitly constructed (Theorem 1.5). This yields a finite procedure for determining all such decompositions, up to conjugacy (Theorem 1.8).

In general, the existence of such a decomposition $\theta = \gamma\phi$ implies that S must be almost finite type (see [4, p. 408]); conversely, if S is almost finite type, then there always exists such a decomposition, where γ is the minimal cover of S and $\deg(\gamma) = 1$ (by [4, Theorem 9 and Corollary 10]). If we require that $\deg(\gamma)$ have a specified value greater than one, then such a decomposition may not exist, as Example 2 in [1] shows. In particular, if we require that $\deg(\gamma) = \deg(\theta)$, then the existence of such a decomposition is equivalent to the statement that θ is almost topologically conjugate over S to a bi-closing map. In Theorem 1.7, we show that this is equivalent to the statement that a certain map, which can be explicitly constructed, is bi-closing. This result can be viewed as a generalization of [4, Theorem 9 and Corollary 10].

1991 Mathematics Subject Classification: 58F03
Key words: shift of finite type, sofic shift, factor map

Example 2 in [1] gives a finite-to-one factor map which is not almost topologically conjugate over its range to a bi-closing map. On the other hand, it follows from [7, Theorem 4.1] that any finite-to-one factor map is almost topologically conjugate over its range to a right closing map, and to a left closing map (see Theorem 1.2).

We briefly summarize some background material. See [11], [8], [5], [4] or [1] for more details. A *factor map* $\theta : \Sigma_A \to S$, from a shift of finite type onto a sofic shift, is a continuous, surjective map which commutes with the shift. A factor map is *finite-to-one* if every point in the range has a uniformly bounded number of preimages. If θ is finite-to-one and Σ_A is irreducible, then there exists a unique positive integer d, called the *degree* of θ and denoted $\deg(\theta)$, such that every doubly transitive point in S has exactly d preimages ([8, Chap. 9, Def. 1.2]). We say that θ is *d-to-one almost everywhere.* A factor map is *right closing* if it does not identify two distinct points which are left asymptotic under the shift map, and *left closing* if it does not identify two distinct points which are right asymptotic. A map is *bi-closing* if it is both right and left closing.

A *decomposition* of θ is a pair of maps $\phi : \Sigma_A \to \Sigma_B$, $\gamma : \Sigma_B \to S$ such that Σ_B is a shift of finite type and $\gamma\phi = \theta$. Two factor maps $\theta_1 : \Sigma_{A_1} \to S$ and $\theta_2 : \Sigma_{A_2} \to S$ are *conjugate over S* if there is a topological conjugacy $\alpha : \Sigma_{A_1} \to \Sigma_{A_2}$ such that $\theta_2\alpha = \theta_1$. (In [11], this relation was referred to simply as *conjugacy.*) The maps θ_1 and θ_2 are *almost topologically conjugate over S* if there exists an irreducible shift of finite type Σ_D and one-to-one almost everywhere factor maps ψ_i $(i = 1, 2)$ such that $\theta_1\psi_1 = \theta_2\psi_2$ (see [1, p. 487]).

The *fiber product* of the maps θ_1 and θ_2 is the shift of finite type $\Sigma_E = \{(x, y) \in \Sigma_{A_1} \times \Sigma_{A_2} | \theta_1(x) = \theta_2(y)\}$, together with the obvious projection maps $\psi_i : \Sigma_E \to \Sigma_{A_i}$, $i = 1, 2$ (see [8, Chap. 8, Def. 3.2]). The following fact is well-known.

Lemma 0.1. *Let $\theta_1 : \Sigma_{A_1} \to S$ and $\theta_2 : \Sigma_{A_2} \to S$ be factor maps from irreducible shifts of finite type onto a sofic shift. Let Σ_E be the fiber product of θ_1 and θ_2, and assume that $\Sigma_{E'}$ is an irreducible component of Σ_E, of maximal entropy in Σ_E, and that θ_2 is one-to-one almost everywhere. Then the restriction $\psi_1|_{\Sigma_{E'}}$ is a one-to-one almost everywhere factor map.*

Proof. Since θ_2 is finite-to-one, it is easy to see that ψ_1 is finite-to-one. Since $\Sigma_{E'}$ is irreducible, it follows from [6, Theorem 3.3] that $\psi_1|_{\Sigma_{E'}}$ is onto (see the proof of [1, Theorem 9]). To see that $\psi_1|_{\Sigma_{E'}}$ is one-to-one almost everywhere, suppose that $x \in \Sigma_{A_1}$ is doubly transitive. By [8, Chap. 9, Lemma 1.13], $\theta_1(x)$ is doubly transitive. If (x, y) and (x, y') are two preimages of x under $\psi_1|_{\Sigma_{E'}}$, then since $\theta_2(y) = \theta_1(x) = \theta_2(y')$, and θ_2 is one-to-one almost everywhere, we must have $y = y'$. Therefore $\psi_1|_{\Sigma_{E'}}$ is one-to-one almost everywhere. □

It follows from Lemma 0.1 that if θ_1 and θ_2 are one-to-one almost everywhere, then they are almost topologically conjugate over S. From this, using the fiber product construction, one can easily show that almost topological conjugacy over S is an equivalence relation on finite-to-one factor maps.

The following is a simple generalization of [4, Theorem 9].

Lemma 0.2. *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. Then θ is almost topologically conjugate over S to a bi-closing map if and only if there exists a shift of finite type Σ_B, a one-to-one almost everywhere factor map $\phi : \Sigma_A \to \Sigma_B$ and a bi-closing map $\gamma : \Sigma_B \to S$ such that $\theta = \gamma\phi$.*

Proof. The if direction is clear, taking ψ_1 to be the identity and $\psi_2 = \phi$. For the only if direction, suppose that θ is almost topologically conjugate over S to a bi-closing map γ. By [4, Theorem 1.1], there is an irreducible component $\Sigma_{E'}$ of the fiber product of θ and γ, of maximal entropy, such that the restrictions $\psi_1|_{\Sigma_{E'}}$ and $\psi_2|_{\Sigma_{E'}}$ are one-to-one almost everywhere. Since γ is bi-closing, it is not hard to show that $\psi_1|_{\Sigma_{E'}}$ is bi-closing. Since Σ_A is a shift of finite type, it follows from [9, Theorem 7.3] that $\psi_1|_{\Sigma_{E'}}$ is constant-to-one, and since it is one-to-one almost everywhere, it must be a conjugacy. Now just take $\phi = (\psi_2|_{\Sigma_{E'}})\,(\psi_1|_{\Sigma_{E'}})^{-1}$. □

Our main tools in this paper will be the right closing cover and left closing cover induced by a θ-congruence partition, which is a generalization of NASU's right closing cover ([10]). We next briefly summarize the construction of this cover, the details of which are given in [11].

Let $\theta : \Sigma_A \to S$ be a finite-to-one, one-block factor map, of degree d, from an irreducible shift of finite type onto a sofic shift. Assume that A is an $n \times n$ non-negative, integral matrix. Let $m = m_1 \cdots m_r$ be a magic word for θ, with magic coordinate m_s ([11, Def. 2.1]). Let $S(\theta, m)$ denote the set of d symbols which occur at coordinate s in the set of preimages of m under θ ([11, Def. 2.2]). The map θ gives rise to a group $\mathscr{G}$ of permutations on the set $S(\theta, m)$ (see [11, the remarks following Lemma 2.6]). A *θ-congruence partition* is a partition $\mathscr{P}$ of $S(\theta, m)$ which is invariant under the action of $\mathscr{G}$ ([11, Def. 3.1]). Any decomposition of θ naturally induces a θ-congruence partition ([11, Def. 2.10 and Lemma 3.2]). The *partition into singleton sets* is the partition of $S(\theta, m)$ into sets each of which contains a single element; this is always a θ-congruence partition, as is the partition containing the single set $S(\theta, m)$ (see [11, remarks following Def. 3.1]).

For a given θ-congruence partition $\mathscr{P}$ and $P \in \mathscr{P}$, and an S-word $w = mv$ beginning with m, we define a vector $l^{(w,P)} \in \mathbb{Z}^n$ by

$$(l^{(w,P)})_i = \begin{cases} 1 & \text{if there exists } ab \in \theta^{-1}(mv), \\ & \text{ending at } i, \text{ such that } a_s \in P \\ 0 & \text{otherwise} \end{cases}$$

([11, Def. 3.3]). We define a shift of finite type $\Sigma_{R(\theta,\mathscr{P})}$, whose states are vectors of the form $l^{(w,P)}$, with transitions defined by $l^{(w,P)} \to l^{(wa,P)}$, where a is a symbol for S such that wa is allowed. We label the corresponding edge in the directed graph for $\Sigma_{R(\theta,\mathscr{P})}$ by a. This labelling defines a right resolving factor map $\pi_{r(\theta,\mathscr{P})} : \Sigma_{R(\theta,\mathscr{P})} \to S$, called the *right closing cover induced by the partition $\mathscr{P}$* ([11, Def. 3.5]). There is also a left closing cover, $\pi_{l(\theta,\mathscr{P})} : \Sigma_{L(\theta,\mathscr{P})} \to S$, which is defined in a similar fashion.

We will use the following result from [11].

Theorem 0.3. ([11, Theorem 4.7]) *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift, and suppose that there is a shift of finite type Σ_B, a factor map $\phi : \Sigma_A \to \Sigma_B$ and a factor map $\gamma : \Sigma_B \to S$ such that $\gamma\phi = \theta$. Let $\mathscr{P}$ be the partition induced by $\gamma\phi$.*

(i) *If θ is left closing and γ is right closing, then γ is conjugate over S to the right closing cover $\pi_{r(\theta,\mathscr{P})}$.*

(ii) *If ϕ is right closing and γ is left closing, then γ is conjugate over S to the left closing cover $\pi_{l(\theta,\mathscr{P})}$.*

Decompositions

Lemma 1.1. *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. Let $\mathscr{P}$ be a θ-congruence partition, and suppose that $\mathscr{P}$ contains d distinct sets. Then* $\deg(\pi_{r(\theta,\mathscr{P})}) = d$. *If $\mathscr{P}$ is the partition into singleton sets, then* $\deg(\pi_{r(\theta,\mathscr{P})}) = \deg(\theta)$.

Proof. We may assume that θ is a one-block map and that m is a magic word for θ. Write $\mathscr{P} = \{P_1, \ldots P_d\}$. If mv is any S-word beginning with the magic word m, then it follows from [11, Lemma 2.3] that the d vectors of the form $l^{(mv,P_i)}$, $1 \leqslant i \leqslant d$, are pairwise distinct. Now, let w be any S-word. Since S is irreducible, there exists an S-word u such that muw is allowed. For $1 \leqslant i \leqslant d$, let w_i be the unique word in $(\pi_{r(\theta,\mathscr{P})})^{-1}(w)$ beginning at $l^{(mu,P_i)}$, and ending at $l^{(muw,P_i)}$, which exists by definition of $\pi_{r(\theta,\mathscr{P})}$ ([11, Def. 3.5]). By the previous remark, the words of the form w_i are d pairwise mutually separated preimages of w, which shows that $\deg(\pi_{r(\theta,\mathscr{P})}) \geqslant d$.

To see the reverse inequality, note that any preimage of m under $\pi_{r(\theta,\mathscr{P})}$ is a $\Sigma_{R(\theta,\mathscr{P})}$-word beginning at some state $l^{(mw,P_i)}$ and ending at $l^{(mwm,P_i)}$, where w is an S-word such that mwm is allowed. It follows from [11, Lemma 6.2] that any state of the form $l^{(mwm,P_i)}$ is equal to a state of the form $l^{(m,P_j)}$, for some j. Therefore any preimage of m ends at one of the d states $l^{(m,P_j)}$, $1 \leqslant j \leqslant d$. It follows that $\deg(\pi_{r(\theta,\mathscr{P})}) \leqslant d$ (and in fact m is a magic word for $\pi_{r(\theta,\mathscr{P})}$). Combining this with the previous inequality yields $\deg(\pi_{r(\theta,\mathscr{P})}) = d$.

If $\mathscr{P}$ is the partition into d singleton sets, then $d = \deg(\theta)$ by definition of $S(\theta, m)$ ([11, Def. 2.2]), which proves the last statement. □

The next result shows that any finite-to-one factor map is almost topologically conjugate over its range to a right closing map, and to a left closing map.

Theorem 1.2. ([7, Theorem 4.1]) *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. Let $\mathscr{P}$ be the partition into singleton sets. Then θ is almost topologically conjugate over S to the induced right closing cover $\pi_{r(\theta,\mathscr{P})}$ and to the induced left closing cover $\pi_{l(\theta,\mathscr{P})}$.*

Proof. Let $\bar{\mathscr{P}}$ denote the θ-congruence partition consisting of the single set $S(\theta, m)$ (where m is a magic word for θ), and let $\pi_{l(\theta,\mathscr{P})} : \Sigma_{L(\theta,\mathscr{P})} \to S$ be its induced left closing cover. In the proof of [7, Theorem 4.11], it is shown that there exists an irreducible component of maximal entropy, Σ_D, of the fiber product of $\pi_{r(\theta,\mathscr{P})}$

and $\pi_{l(\theta,\bar{\mathscr{P}})}$, and factor maps $p : \Sigma_D \to \Sigma_{R(\theta,\mathscr{P})}$ and $g : \Sigma_D \to \Sigma_A$ such that $\pi_{r(\theta,\mathscr{P})}p = \theta g$. (Note the change in notation: in [7], $\pi_{r(\theta,\mathscr{P})}$ is denoted π_L, $\pi_{l(\theta,\mathscr{P})}$ is denoted π_R and $p = p_L$.) The map p is the restriction to Σ_D of the natural projection map from the fiber product onto $\Sigma_{R(\theta,\mathscr{P})}$. Since $\bar{\mathscr{P}}$ contains exactly one set, it follows from Lemma 1.1 that $\deg(\pi_{l(\theta,\bar{\mathscr{P}})}) = 1$. By Lemma 0.1, we have $\deg(p) = 1$. Since $\mathscr{P}$ is the partition into singleton sets, it follows from Lemma 1.1 that $\deg(\theta) = \deg(\pi_{r(\theta,\mathscr{P})})$. Applying [11, Lemma 2.7], we obtain $\deg(\pi_{r(\theta,\mathscr{P})})\deg(p) = \deg(\pi_{r(\theta,\mathscr{P})}p) = \deg(\theta g) = \deg(\theta)\deg(g)$, and it follows that $\deg(g) = \deg(p) = 1$. Therefore θ is almost topologically conjugate over S to $\pi_{r(\theta,\mathscr{P})}$.

A similar argument shows that θ is almost topologically conjugate over S to $\pi_{l(\theta,\mathscr{P})}$. □

For right closing maps, the conclusion of Theorem 1.2 can be strengthened from almost topologically conjugate over S to conjugate over S.

Lemma 1.3. *Let $\gamma : \Sigma_B \to S$ be a right closing factor map from an irreducible shift of finite type onto a sofic shift. Let $\mathscr{P}'$ be the partition into singleton sets. Then γ is conjugate over S to the right closing cover $\pi_{r(\gamma,\mathscr{P}')}$.*

Proof. Let $\phi : \Sigma_B \to \Sigma_B$ denote the identity. Then $\mathscr{P}'$ is the congruence partition induced by $\gamma\phi$. Since $\theta = \gamma\phi$ is a decomposition of θ into a left closing factor map followed by a right closing factor map, the result now follows from Theorem 0.3 (i). □

Lemma 1.4. *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. Suppose that there exists a shift of finite type Σ_B, a factor map $\phi : \Sigma_A \to \Sigma_B$ and a right closing factor map $\gamma : \Sigma_B \to S$ such that $\gamma\phi = \theta$. Let $\mathscr{P}$ be the θ-congruence partition induced by the decomposition $\gamma\phi$. Then there exists a right closing factor map $\rho : \Sigma_{R(\theta,\mathscr{P})} \to \Sigma_B$, of degree one, such that $\gamma\rho = \pi_{r(\theta,\mathscr{P})}$.*

Proof. Suppose that there exists such a decomposition, and assume that $\deg(\gamma) = d$. Let $\mathscr{P}'$ be the γ-congruence partition into d singleton sets. It follows from the definition of the partition induced by $\gamma\phi$ ([11, Def. 2.10]) that $\mathscr{P}$ contains exactly d sets. By Lemma 1.1, we have $\deg(\pi_{r(\theta,\mathscr{P})}) = d$. By [11, Lemma 3.7], there exists a factor map $\nu_\phi : \Sigma_{R(\theta,\mathscr{P})} \to \Sigma_{R(\gamma,\mathscr{P}')}$ such that $\pi_{r(\gamma,\mathscr{P}')}\nu_\phi = \pi_{r(\theta,\mathscr{P})}$. Since γ is right closing, it follows from Lemma 1.3 that there exists a conjugacy $\delta : \Sigma_{R(\gamma,\mathscr{P}')} \to \Sigma_B$ such that $\pi_{r(\theta,\mathscr{P}')} = \gamma\delta$. Let $\rho = \delta\nu_\phi$. Clearly, ρ is right closing and $\gamma\rho = \gamma\delta\nu_\phi = \pi_{r(\gamma,\mathscr{P}')}\nu_\phi = \pi_{r(\theta,\mathscr{P})}$. Since $\deg(\gamma\rho) = \deg(\gamma)\deg(\rho)$, by [11, Lemma 2.7], and $\deg(\gamma) = d = \deg(\pi_{r(\theta,\mathscr{P})})$, if follows that $\deg(\rho) = 1$. □

The next result shows that if there exists a decomposition $\theta = \gamma\phi$, where the range of ϕ is a shift of finite type and γ is bi-closing, then the map γ is determined up to conjugacy by the θ-congruence partition induced by $\gamma\phi$. First, we simplify our notation as follows.

Notation: If $\mathscr{P}$ is a θ-congruence partition, and $\mathscr{P}'$ is the $\pi_{r(\theta,\mathscr{P})}$-congruence partition into singleton sets, then we let $\pi_{l(\pi_{r(\theta,\mathscr{P})})} = \pi_{l(\pi_{r(\theta,\mathscr{P})},\mathscr{P}')} : \Sigma_{L(\pi_{r(\theta,\mathscr{P})},\mathscr{P}')} \to S$

denote the left closing cover induced on $\pi_{r(\theta,\mathscr{P})}$ by $\mathscr{P}'$. Similarly, if $\mathscr{P}'$ is the $\pi_{l(\theta,\mathscr{P})}$-congruence partition into singleton sets, we let $\pi_{r(\pi_{l(\theta,\mathscr{P})})} = \pi_{r(\pi_{l(\theta,\mathscr{P})},\mathscr{P}')}$.

Theorem 1.5. *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. Suppose that there exists a shift of finite type Σ_B, a factor map $\phi : \Sigma_A \to \Sigma_B$ and a bi-closing factor map $\gamma : \Sigma_B \to S$ such that $\gamma\phi = \theta$. Let $\mathscr{P}$ be the θ-congruence partition induced by the decomposition $\gamma\phi$. Let $\mathscr{P}'$ be the $\pi_{r(\theta,\mathscr{P})}$-congruence partition into singleton sets. Then γ is conjugate over S to the left closing cover $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ induced by the partition $\mathscr{P}'$.*

Proof. Since γ is right closing, by Lemma 1.4 there exists a right closing factor map $\rho : \Sigma_{R(\theta,\mathscr{P})} \to \Sigma_B$, of degree one, such that $\gamma\rho = \pi_{r(\theta,\mathscr{P})}$. The decomposition $\gamma\rho$ is a decomposition of $\pi_{r(\theta,\mathscr{P})}$ into a right closing map followed by a left closing map, since γ is left closing. Since $\deg(\rho) = 1$, it follows from [11, Def. 2.10 and Lemma 2.8] that $\mathscr{P}'$ is the $\pi_{r(\theta,\mathscr{P})}$-congruence partition induced by $\gamma\rho$. The result now follows from Theorem 0.3 (ii). □

The map $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ in Theorem 1.5 is obtained by taking the right closing cover induced by $\mathscr{P}$, and then the left closing cover induced on $\pi_{r(\theta,\mathscr{P})}$ by $\mathscr{P}'$. However, the theorem also holds if we first induce on the left and then on the right; that is, γ is also conjugate over S to the map $\pi_{r(\pi_{l(\theta,\mathscr{P})})}$, obtained by first taking the left closing cover induced by $\mathscr{P}$, and then the right closing cover induced on $\pi_{l(\theta,\mathscr{P})}$ by $\mathscr{P}'$. This gives the following result.

Corollary 1.6. *Under the same hypotheses as in Theorem 1.5, the two maps $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ and $\pi_{r(\pi_{l(\theta,\mathscr{P})})}$ are conjugate over S, and both are bi-closing.*

In the next result we characterize when a finite-to-one factor map is almost topologically conjugate over its range to a bi-closing map.

Theorem 1.7. *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. Let $\mathscr{P}$ be the partition into singleton sets. The following are equivalent.*

(i) *θ is almost topologically conjugate over S to a bi-closing map.*

(ii) *There exists a shift of finite type Σ_B, a one-to-one almost everywhere factor map $\phi : \Sigma_A \to \Sigma_B$ and a bi-closing map $\gamma : \Sigma_B \to S$ such that $\theta = \gamma\phi$.*

(iii) *$\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ is bi-closing.*

(iv) *$\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ is topologically conjugate to $\pi_{r(\pi_{l(\theta,\mathscr{P})})}$.*

Proof. (i) ⇔ (ii) follows from Lemma 0.2.

(ii)⇒(iv) If (ii) holds, then since ϕ is one-to-one almost everywhere, it follows from [11, Def. 2.10] that the congruence partition induced by $\gamma\phi$ is the partition $\mathscr{P}$. By Corollary 1.6, $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ and $\pi_{r(\pi_{l(\theta,\mathscr{P})})}$ are conjugate over S and therefore topologically conjugate.

(iv)⇒(iii) Since $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ is left closing and $\pi_{r(\pi_{l(\theta,\mathscr{P})})}$ is right closing (see [11, Def. 3.5]), if they are topologically conjugate then both maps are bi-closing.

(iii)⇒(i) By Theorem 1.2, θ is almost topologically conjugate over S to its induced right closing cover, $\pi_{r(\theta,\mathscr{P})}$, and $\pi_{r(\theta,\mathscr{P})}$ is almost topologically conjugate

over S to its induced left closing cover, $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$. Since almost topologically conjugacy over S is an equivalence relation (by the remarks following Lemma 0.1), θ is almost topologically conjugate over S to $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$. So, if $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ is bi-closing, then (i) holds. □

If the conditions in Theorem 1.7 hold, then the map γ is conjugate over S to $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$, and so is uniquely determined up to conjugacy.

The conditions in Theorem 1.7 are decidable, since there is a finite procedure for deciding whether the map $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ is bi-closing. (The proof of [4, Prop. 1] gives a finite procedure for deciding whether a factor map is right closing or left closing.)

If θ is one-to-one almost everywhere, then it follows from [4, Theorem 9 and Corollary 10] that conditions (i)–(iv) are equivalent to S being almost finite type, in which case $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ is conjugate over S to the minimal cover of S.

Finally, we give a procedure for describing all possible decompositions $\theta = \gamma\phi$, where the range of ϕ is a shift of finite type and γ is biclosing.

Theorem 1.8. *Let $\theta : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. There is finite procedure for determining all decompositions of the form $\phi : \Sigma_A \to \Sigma_B$, $\gamma : \Sigma_B \to S$, where Σ_B is a shift of finite type and γ is bi-closing, up to conjugacy.*

Proof. We may assume that θ is a one-block map. For each θ-congruence partition $\mathscr{P}$, there is a finite procedure for construcing the map $\pi_{l(\pi_{r(\theta,\mathscr{P})})} : \Sigma_{L(\pi_{r(\theta,\mathscr{P})},\mathscr{P}')} \to S$, and a finite procedure for deciding whether $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ is bi-closing ([4, Prop. 1]). Now, for each map $\pi_{l(\pi_{r(\theta,\mathscr{P})})}$ which is bi-closing, there is a finite procedure for determining all factor maps $\rho : \Sigma_A \to \Sigma_{L(\pi_{r(\theta,\mathscr{P})},\mathscr{P}')}$ such that $\pi_{l(\pi_{r(\theta,\mathscr{P})})}\rho = \theta$ ([11, Corollary 5.3]). This generates a finite list of decompositions of θ, and it follows from Theorem 1.5 that any decomposition of θ as in the statement of Theorem 1.8 must be conjugate over S to one on the list. □

M. Boyle has recently given a general method of determining all possible decompositions of a factor map, up to conjugacy, and has shown that the number of these is finite ([3]).

References

[1] Adler R, Kitchens B, Marcus B (1985) Almost topological classification of finite-to-one factor maps between shifts of finite type. Ergodic Th Dynam Syst **5**: 485–500

[2] Adler R, Marcus B (1979) Topological entropy and equivalence of dynamical systems. Memoirs Amer Math Soc **219**

[3] Boyle M Factoring factor maps. J London Math Soc (to appear)

[4] Boyle M, Kitchens B, Marcus B (1985) A note on minimal covers for sofic systems. Proc Amer Math Soc **95**: 403–411

[5] Boyle M, Marcus B, Trow P (1987) Resolving maps and the dimension group for shifts of finite type. Memoirs Amer Math Soc **377**

[6] Coven E, Paul M (1974) Endomorphisms of irreducible shifts of finite type. Math Syst Theory **8**: 167–175

[7] Kitchens B, Marcus B, Trow P (1991) Eventual factor maps and compositions of closing maps. Ergodic Th Dynam Syst **11**: 85–113

[8] Lind D, Marcus B (1995) An Introduction to Symbolic Dynamics and Coding. New York: Cambridge Univ Press

[9] NASU M (1983) Constant-to-one and onto global maps of homomorphisms between strongly connected graphs. Ergodic Th Dynam Syst **3**: 387–413
[10] NASU M (1988) Topological conjugacy for sofic systems and extensions of automorphisms of finite subsystems of topological Markov shifts. In: ALEXANDER JC (ed) Dynamical Systems. Proceedings, University of Maryland, 1986–87, pp 564–607. Lect Notes Math **1342**
[11] TROW P (1995) Decompositions for finite-to-one factor maps. Israel J Math **91**: 129–155

P. TROW
Department of Mathematical Sciences
The University of Memphis
Memphis
Tennessee 38152
USA

Buchbesprechungen – Book Reviews

Gardner, M.: *The Universe in a Handkerchief.* Lewis Carroll's Mathematical Recreations, Games, Puzzles, and Word Plays. X, 159 pp. Springer, New York, 1996. Cloth DM 33,–; öS 233,–.

Lewis Carroll belongs to the most eminent recreational mathematicians ever since as can be seen already in those books he is especially renowned for: "Alice's Adventures in Wonderland" and "Through the Looking-Glass". A great many of puzzles, paradoxes and mathematical amusements, which are commonly known today, go back to him. Regarding this the bibliophil book, written by another master of recreational mathematics, gives an interesting and informative survey thus allowing to look at the peculiar way of Carroll's mathematical resp. logical thinking.

G. Kowol, Wien

Rigatelli, L. T.: *Evariste Galois 1811–1832* (Vita Mathematica, Vol. 11). 162 pp. Birkhäuser, Basel Berlin Boston, 1996. DM 35,–; öS 278,–.

Evariste Galois has been – and is – fascinating mathematicians because of his outstanding work as well as the mysterious circumstances of his death. The author having supplied some new documents sheds light on that last point. She makes evident that Galois death had nothing to do with a coquette but had been a spontaneous – lastly pointless – political sacrifice. The other occurrences of Galois' short life are just as scrupulously researched and are combined to a well-rounded biography. It includes also a separate chapter on the mathematical contributions of Galois and, in addition to the bibliography, references on classical Galois theory.

G. Kowol, Wien

Janusz, G. J.: *Algebraic Number Fields.* Second Edition (Graduate Studies in Mathematics, Vol. 7). X, 276 pp. American Mathematical Society, Providence, Rhode Island, 1996. Cloth US $ 44,–.

This is the second edition of a book originally published by Academic Press in 1973. It contains an introduction to algebraic number fields and class field theory as well. The contents have remained more or less the same; some material, like L-series, has been added, some formulations were changed and misprints were corrected. Although written in a less modern style than e.g. Neukirch's expositions, the reviewer likes the less abstract access chosen in the book under discussion.

J. Schoissengeier, Wien

Kopytov, V. M., Medvedev, N. Y.: *Right-Ordered Groups* (Siberian School of Algebra and Logic). IX, 250 pp. Consultants Bureau, London Moscow, 1996. Cloth US $ 95,–.

A group $(G, \cdot)$ is called partially right-ordered (p.r.o) if a partial order is defined on G, which is compatible with multiplication on the right. Some books on partially ordered (p.o.) groups (for which the ordering is compatible with multiplication on both sides) contained some chapters on the theory of p.r.o. groups (Kokorin–Kopytov 1974, Bigard–Keimel–Wolfenstein 1977, Botto Mura–Rhemtulla 1977). This book intends to give a complete

review of methods and results on these interesting algebraic objects. In fact, it deals with the theory of torsion-free groups seen from an order theoretical point of view. Surprisingly, a large number of theorems on p.o. groups are treated in the text, which are not used in the theory of p.r.o. groups. In contrast, some elementary facts on p.r.o. groups are not included (as the existence of partial right-orders, connections with left-compatibility, lattice orders, etc.). Unfortunately, the translations and some formulations are sometimes misleading.

H. Mitsch, Wien

Nathanson, M. B.: *Additive Number Theory.* Inverse Problems and the Geometry of Sumsets (Graduate Texts in Mathematics, Vol. 165). XIV, 291 pp. Springer, Berlin Heidelberg New York, 1996. Cloth DM 80,–; öS 568,–.

This book is devoted to the study of inverse problems in additive number theory. It is the companion volume of "Additive Number Theory: The Classical Bases" by the same author, but can be read independently. A typical question is of the following shape: Let A and B be finite subsets of an abelian group. What can be deduced about A and B from the subset A + B? A typical answer would be: If A + B is small than A and B have a certain structure, e.g. are subsets of arithmetic progressions. One of the book's highlights is an elegant proof by I. Z. Ruzsa of a deep theorem of G. A. Freiman which is of this type. Whereas several of the proofs are rather involved the prerequisites are small and it is enjoyable to read. It is therefore the reviewer's hope that it will also find its way into the hands of some interested non-experts.

C. Baxa, Wien

Leutbecher, A.: *Zahlentheorie.* Eine Einführung in die Algebra. 9 Abb., 6 Tab., 1 Faltb., XI, 354 S. Springer, Berlin Heidelberg New York, 1996. Brosch. DM 49,–; öS 350,–.

Erklärtes Ziel des Autors war es, eine Einführung in die elementare und algebraische Zahlentheorie für Leser mit besonders geringen Vorkenntnissen, wie Studierende ab dem 3. Semester oder Nichtmathematiker aus benachbarten Gebieten, zu schaffen. Um kein Wissen aus Algebra voraussetzen zu müssen, werden Resultate über Gruppen und Körper parallel bereitgestellt. Dennoch dringt das Buch viel tiefer in die algebraische Zahlentheorie ein, als man beim ersten Aufschlagen erwarten würde. So findet man beispielsweise den Dirichletschen Einheitensatz, die analytische Klassenzahlformel für die Dedekindsche Zetafunktion und den Satz von Kronecker–Weber. Der unkonventionelle Aufbau des Buches ist manchmal verblüffend – etwa, wenn auf das Kapitel "Das Hilbertsche Normenrestsymbol" die "Elemente der Gruppentheorie" folgen – kann aber als gelungen bezeichnet werden.

C. Baxa, Wien

Bundschuh, P.: *Einführung in die Zahlentheorie*, 3. Aufl. 7 Abb., XIV, 336 S. Springer, Berlin Heidelberg New York, 1996. Brosch. DM 56,–; öS 394,–.

Daß dieses schöne Lehrbuch bereits seine dritte Auflage erlebt, sollte als verdienter Erfolg betrachtet werden. Zu seinen Vorzügen zählt, neben guter Lesbarkeit, den Lernenden in viele Teilgebiete der Zahlentheorie einzuführen. Neben elementarer enthält es die Anfänge von analytischer und algebraischer Zahlentheorie und ein Kapitel über diophantische Approximation mit besonderer Betonung der Theorie der transzendenten Zahlen. In zahlreichen Bemerkungen schildert der Autor die historische Entwicklung von Fragestellungen und Lösungen. Für diese Auflage ist eine kurze Geschichte des Beweises des großen Fermats durch Andrew Wiles hinzugefügt worden.

C. Baxa, Wien

Cassels, J. W. S., Flynn, E. V.: *Prolegomena to a Middlebrow Arithmetic of Curves of Genus 2* (London Mathematical Society Lecture Note Series, Vol. 230). XIV, 218 pp. Cambridge University Press, Cambridge New York Oakleigh, 1996. Softcover US $ 37,95.

While number theory of curves of genus 0 or 1 is well investigated, the corresponding theory for curves of higher genus is still in a sad state insofar as the general theory (Jacobian, Mordell–Weil group, Falting's theorem) is not really suitable to deal with concrete examples. The credit for having investigated concrete examples of curves of genus 2 and for having linked with the general theory as far as possible goes largely to the authors of the book. Of course, one cannot expect the manipulations to be done by "pen and paper" (note that the Jacobian of such a curve is a surface in $\mathbb{P}_{15}$). The reader will find the accompanying computer programm, written in MAPLE, in an appendix. It is clear that this often subtle and witty book is required reading for those working in the field.

J. SCHOISSENGEIER, Wien

Berndt, B. C., Diamond, H. G., Hildebrand, A. J. (Eds.): *Analytic Number Theory.* Proceedings of a Conference in Honor of Heini Halberstam, Volume 1 (Progress in Mathematics, Vol. 138). XII, 449 pp. Birkhäuser, Basel Berlin Boston, 1996. Cloth DM 182,–; öS 1300,–.

Berndt, B. C., Diamond, H. G., Hildebrand, A. J. (Eds.): *Analytic Number Theory.* Proceedings of a Conference in Honor of Heini Halberstam, Volume 2 (Progress in Mathematics, Vol. 139). XII, 436 pp. Birkhäuser, Basel Berlin Boston, 1996. Cloth DM 182,–; öS 1300,–.

On May 16–20, 1995, a conference on Analytic Number Theory took place at the University of Illinois. G. Andrews, J. Bourgain, J. M. Deshouillers, H. Halberstam, D. R. Heath-Brown, H. Iwaniec, H. L. Montgomery, M. Ram Murty, C. Pomerance and R. C. Vaughan were invited plenary speakers. The two volumes of the proceedings contain 49 papers of a broad spectrum of Analytic Number Theory and two fotos of H. Halberstam, one of which portrays him in younger years.

J. SCHOISSENGEIER, Wien

Redmond, D.: *Number Theory. An Introduction.* Pure and Applied Mathematics XII, 749 pp. Dekker, New York Basel Hong Kong, 1996. Cloth US $ 175,–.

The book is divided up into ten chapters, the first three of them cover the usual elementary number theory, starting from divisibility up to the quadratic reprocity formula. Chapter four contains diophantine approximation in the one dimensional case (continued fractions, Farey series etc.) The next two chapters deal with diophantine equations of genus 0. Three chapters are dedicated to analytic number theory, including the prime number theorem, while the last one contains a brief and very elementary introduction to algebraic number theory. On the end of every chapter one finds additional exercises. The book is best suited for beginners.

J. SCHOISSENGEIER, Wien

Mollin, R. A.: *Quadratics.* XX, 387 pp. CRC Press, Boca Raton New York London, 1996. Cloth US $ 74,95.

This book presents the theory of quadratic number fields in a complete new form. Instead of presenting the usual abstract theory of orders and ideals, the author looks at the so-called infrastructure of an order, introduced by D. Shanks. This infrastructure links the theory of purely periodic continued fractions with a certain class of ideals of an order. This

point of view is one of the principal tools of the whole book and is best suited for modern computation, e.g. for computing class numbers. The book is well readable even for those who have never seen the classical theory of quadratic orders. Nevertheless it may be of some help for these students if they inform themselves roughly about this theory in another book.

J. SCHOISSENGEIER, Wien

Laurinčikas, A.: *Limit Theorems for the Riemann Zeta-Function* (Mathematics and its Applications, Vol. 352). XIII, 297 pp. Kluwer, Dordrecht Boston London, 1996. Cloth US $ 149,–.

This book deals with limit theorems for the zeta-function, Dirichlet series in various parts of the complex plane and other limit theorems with regard to the zeta-function. The volume is badly organized. This harsh criticism calls for justification. The book starts with the abstract definition of probability spaces (Definition 1.3), but assumes that the reader is familiar with abstract integration (see Definition 1.6). Some of the definitions are in fact theorems (Definition 3.2). The definition of the Cartesian product of sets is given on page 179 (!). After having already used the residue theorem on page 29, the author gives a miscarried definition of the concept of an analytic function on page 214. See also the curious definition of an entire function, and so on. I am unwilling to recommend the book to students. At best, it may have some merits for the professional mathematician, who is able to overlook strange formulations and who can guess what was meant by the author.

J. SCHOISSENGEIER, Wien

Nathanson, M. B.: *Additive Number Theory.* The Classical Bases (Graduate Texts in Mathematics 164). XIV, 342 pp. Springer, New York Berlin Heidelberg, 1996. Cloth DM 80,–; öS 569,–.

This book presents an excellent introduction to a difficult theme: additive number theory. All proofs are given in full detail and the specialists will look upon some steps as having been unnecessary. Nevertheless, the reviewer likes it just for this reason, as it is the first textbook on the theme written in this style. It contains, apart from an appendix, Waring's problem, Vinogradov's three prime number theorem and Chen's theorem. At the end of every section exercises have been added. The book is best suited for beginners. The misprints will not detract from the value of the book.

J. SCHOISSENGEIER, Wien

Glynn, J.: *Mathematik entdecken mit DERIVE – von der Algebra bis zur Differentialrechnung.* 154 S. Birkhäuser, Basel Berlin Boston, 1995. DM 42,–; öS 296,40.

Dieses gut lesbare Buch wendet sich gleichermaßen an Lehrer und Lehramtsstudenten wie auch an interessierte Schüler. Es bietet die Möglichkeit, sich auf unterhaltsame Weise im Selbststudium mit dem Computeralgebraprogramm DERIVE vertraut zu machen. Anhand gut ausgewählter Aufgabenstellungen lernt der Leser spielerisch Bedienung und Einsatzmöglichkeiten von DERIVE kennen, und wird gleichzeitig motiviert, sich mit elementaren mathematischen Fragestellungen auseinanderzusetzen.

M. KOTH, Wien

Wolfart, J.: *Einführung in die Zahlentheorie und Algebra.* X, 223 S. Vieweg, Braunschweig Wiesbaden, 1996. DM 38,–.

Diese gut lesbare, übersichtlich gegliederte Einführung mit Übungsaufgaben bietet eine ausgewogene, sinnvolle Synthese zwischen Zahlentheorie und Algebra, enthält alle

grundlegenden wichtigen Resultate bis zum quadratischen Reziprozitätsgesetz, den Gaußschen Summen und zur Galoistheorie mit der Charakterisierung von durch Radikale auflösbaren Gleichungen und den Transzendenzsatz von Lindemann–Weierstraß, aber auch neuere Resultate, z.B. über Primzahltests und Primfaktorzerlegung, die großen gelösten und ungelösten Problemkreise im Umfeld, in verständlicher, den Gesamtzusammenhang erhellender Form.

H. RINDLER, Wien

Griebel, M., Dornseifer, T., Neunhoeffer, T.: *Numerische Simulation in der Strömungsmechanik* (Eine praxisorientierte Einführung). XIII, 219 S. Vieweg, Braunschweig Wiesbaden, 1995. DM 51,–, Brosch. öS 357,–.

In übersichtlicher Weise gibt das Buch eine Einführung in die numerische Simulation, und zwar eben am Beispiel der 2D Navier–Stokes-Gleichungen. Besonderes Bemühen gilt dabei der Herausarbeitung des Umstandes, daß wissenschaftliches Rechnen ein wesentlich interdisziplinäres Unterfangen ist. Entsprechend werden Fragen spezieller Hardware, der Softwareentwicklung und der Visualisierung behandelt. Breiten Raum nehmen natürlich die Modellbildung und die Diskretisierung der Gleichungen ein. Insgesamt ein für den Praktiker sehr nützliches und inhaltsreiches einführendes Buch, zu dem noch dazu Software über das WWW abgerufen werden kann.

H. MUTHSAM, Wien

Rivasseau, V. R. (Ed.): *Constructive Physics.* Results in Field Theory, Statistical Mechanics and Condensed Matter Physics (Lecture Notes in Physics, Vol. 446). VII, 337 pp. Springer, Berlin Heidelberg New York, 1995. Cloth DM 130,–; öS 764,40.

Constructive physics, i.e. the rigorous study of particular physical models by hard mathematical methods such as expansions giving detailed information on the properties of the models observables, is now some 30 years old. This well-edited proceedings volume of the 1994 conference on "Constructive results in field theory and statistical mechanics" at Ecole Polytechnique in Palaiseau (France), collecting the main lectures and selected additional contributions, offers a rather wide spectrum of subjects, but with an emphasis on pedagogical presentation. Topics covered include basic mathematical methods, constructive field theory, solid-state physics and statistical mechanics, and constructive methods for classical partial differential equations. This book should serve well to "invite more pure mathematicians to join" constructive physics, as the editor puts it in his Preface, alluding to an additional general lecture.

T. HUDETZ, Wien

Menezes, A. J., Oorschot, P. C. van, Vanstone, S. A.: *Handbook of Applied Cryptography* (Discrete Mathematics and its Applications). XXVIII, 780 pp. CRC Press, Boca Raton New York, London, 1997. Cloth US $ 96,–.

Cryptography has emerged in the last 20 years as an important discipline providing the foundation for information security in most areas of data communication (such as in financial services industry, in the public sector, and in private electronic mail.) This handbook offers an integrated treatment of the subject and will serve as a valuable text for the novice as well as for the expert. Providing all the mathematical background it allows easy and rapid access of information including more than 200 algorithms and protocols. In spite of the fast development of that field, this volume will constitute the main reference of this subject for many years to come. Contents: Overview; Mathematical Background; Public-Key Parameters; Pseudorandom Bits; Stream and Black Ciphers; Public-Key Encryption, Hash Functions and Data Integrity; Identification and Entity Authentication;

Digital Signatures; Key Establishment Protocols; Key Management Techniques; Efficient Implementation; Patents and Standards.

H. MITSCH, Wien

Horst, R., Pardalos, P. M. (Eds.): *Handbook of Global Optimization.* 880 pp. Kluwer, Boston Dordrecht London, 1995. Cloth £ 176,50.

This book contains up to date information on the field of global optimization. It starts with an introduction and a survey on complexity results. Then a series of expository contributions by various authors describe a wide variety of stochastic and deterministic approaches as well as suboptimal algorithms for solving global optimization problems. The following subjects are covered: Concave minimization, D. C. optimization, quadratic optimization, complementarity problems, minimax problems, multiplicative programming, Lipschitz optimization, fractional programming, network problems, trajectory methods, homotopy methods, interval methods, and stochastic methods. The book is a good reference guide for everybody working in the subject.

H. SCHICHL, Wien

Kutzler, B.: *Mathematik unterrichten mit DERIVE.* Ein Leitfaden für Lehrer. 190 S. Addison-Wesley, Bonn Paris Reading, 1995. Geb. DM 39,90.

Dieses Buch richtet sich an Mathematiklehrer höherer Schulen und Lehramtsstudenten, die bereits über Grundkenntnisse im Umgang mit dem Computeralgebraprogramm DERIVE verfügen. Ein Teil des Buches nimmt auf die derzeit aktuelle DERIVE-Version 3 Bezug und bietet neben einer kurzen Vorstellung von DERIVE interessante praktische Beispiele für den Unterricht sowie diverse Tips und Tricks für das Arbeiten mit diesem Programm. Die übrigen Kapitel sind weitgehend systemunabhängig gestaltet: Hier werden die Einsatzmöglichkeiten von Computeralgebrasystemen im derzeitigen traditionellen Mathematikunterricht diskutiert, aber auch Konzepte für einen zukünftigen „computerzeitaltergemäßen" Mathematikunterricht vorgestellt.

M. KOTH, Wien

Berry, J. S., Graham, E., Watkins, A. J. P.: *Mathematik lernen mit DERIVE.* XII, 354 S. Birkhäuser, Basel Berlin Boston, 1995. DM 70,–; öS 496,40.

Dieses Werk kann Studienanfängern sowie auch interessierten Schülern der Sekundarstufe II als begleitendes Arbeitsbuch empfohlen werden. Anhand von konkreten Aufgabenstellungen wird der Leser angeregt, sich mit den Stoffgebieten Funktionen, Differential- und Integralrechnung sowie Algebra (Komplexe Zahlen, Matrizen) aktiv auseinanderzusetzen. Gleichzeitig erhält er eine praktische Einführung in das Computeralgebraprogramm DERIVE. Die vorgestellten ausgearbeiteten Beipiele sind durch eine große Zahl von Übungen ergänzt, deren Lösungen im Anhang des Buches nachgeschlagen werden können.

M. KOTH, Wien

Verleger: Springer-Verlag KG, Sachsenplatz 4–6, A-1201 Wien. — Herausgeber: Prof. Dr. Karl Sigmund, Institut für Mathematik der Universität Wien, Strudlhofgasse 4, A-1090 Wien. — Redaktion: Strudlhofgasse 4, A-1090 Wien. — Satz und Umbruch: Thomson Press (India) Ltd., New Delhi. — Offsetdruck: Druckerei Eugen Ketterl Ges.m.b.H., A-1180 Wien. Verlagsort: Wien. — Herstellungsort: Wien. — Printed in Austria.

Mh. Math. 125, 179–187 (1998)

Monatshefte für
Mathematik

Printed in Austria

Trace and Spectrum Preserving Linear Mappings in Jordan-Banach Algebras

By

Bernard Aupetit, Québec

(Received 21 February 1996; in revised form 24 September 1996)

Abstract. In the first section we define the trace on the socle of a Jordan-Banach algebra in a purely spectral way and we prove that it satisfies several identities. In particular this trace defines the Faulkner bilinear form. In the second section, using analytic tools and the properties of the trace, we prove that a spectrum preserving linear mapping from J onto J', where J and J' are semisimple Jordan-Banach algebras, is not far from being a Jordan isomorphism. It is in particular a Jordan isomorphism if J' is primitive with non-zero socle.

Introduction

The use of analytic multifunctions has given a lot of interesting new applications in the field of Jordan-Banach algebras (for general surveys see [3] and [4], Chapter 8). In particular, using this technique, the concepts of rank, spectral multiplicity, trace and determinant have been defined in the case of general Jordan-Banach algebras (see [5, 9]), extending former results obtained in the associative case [8].

In Section 1, new identities concerning the trace defined on the socle of a Jordan-Banach algebra are proved. They are used to prove that $\mathrm{Tr}(xy)$ coincides with the Faulkner form on the socle and then to characterize the annihilator of the socle as the orthogonal of the socle for the bilinear form defined by this trace.

Given two semisimple Banach algebras with identity we say that T is a spectrum preserving linear mapping from A onto B if T is surjective and we have $\mathrm{Sp}\, T(x) = \mathrm{Sp}\, x$, for all x in A. In [7] we gave the historical motivation to this notion and we proved several results:

(i) T is a continuous linear bijection from A onto B such that $T(\mathrm{Soc}\, A) = \mathrm{Soc}\, B$ and $T(kh(\mathrm{Soc}\, A)) = kh(\mathrm{Soc}\, B)$,

(ii) $(Ta^2 - (Ta)^2)x = 0$ for $a \in A$ and $x \in \mathrm{Soc}\, B$, in particular T is a Jordan isomorphism on $\mathrm{Soc}\, A$,

(iii) T is an isomorphism or an anti-isomorphism if B is a primitive Banach algebra with non-zero socle,

1991 Mathematics Subject Classification: 17C65, 46H70

Key words: Jordan-Banach algebra, spectrum preserving mapping, Jordan isomorphism, socle, trace

(iv) T is a Jordan isomorphism if B is a separable scattered Banach algebra.

Is it possible to extend some of these results to the Jordan-Banach situation? Using the results of Section 1 the answer is yes. We shall see this in Section 2.

1. The Trace

To simplify, throughout the paper we suppose that J is a semisimple Jordan-Banach algebra with identity, but in many cases these hypotheses are not really necessary. We denote by Soc J the socle of J. For its definition and its properties see [3, 5, 6, 9, 11, 12].

We recall a few definitions and properties taken from [5, 9]. If $x \in J$ we define the *rank* by

$$\operatorname{rank}(x) = \max_{y \in J} \#(\operatorname{Sp} U_y(x) \setminus \{0\}) \leqslant +\infty,$$

and then the socle coincides with the set of finite-rank elements. If $x \in \operatorname{Soc} J$ then x is algebraic, consequently its spectrum is finite. Let $\lambda \in \operatorname{Sp} x$ and Δ be a small disk centred at λ isolating λ from the remaining spectrum, then the *multiplicity* of x at λ is given by

$$m(\lambda, x) = \#\operatorname{Sp}(U_b(x) \cap \Delta),$$

where b is an arbitrary element of $B \cap E(x)$, B denoting the unit ball centred at the identity and $E(x) = \{y : \#(\operatorname{Sp} U_y(x) \setminus \{0\}) = \operatorname{rank}(x)\}$.

Then for $x \in \operatorname{Soc} J$ the *trace* is defined by

$$\operatorname{Tr}(x) = \sum_{\lambda \in \operatorname{Sp} x} \lambda m(\lambda, x).$$

This trace is additive on the socle and continuous when restricted to a set of bounded-rank elements.

Proposition 1.1. *Let Φ be an automorphism of J. Then the socle is invariant under Φ, $\operatorname{Sp}\Phi(x) = \operatorname{Sp} x$, $\operatorname{rank}\Phi(x) = \operatorname{rank}(x), m(\lambda, \Phi(x)) = m(\lambda, x)$, for every $x \in \operatorname{Soc} J$ and $\lambda \in \operatorname{Sp} x$. Consequently $\operatorname{Tr}\Phi(x) = \operatorname{Tr}(x)$, for every $x \in \operatorname{Soc} J$. If D is a derivation on J, then the socle is invariant under D and moreover if D is bounded then $\operatorname{Tr} Dx = 0$, for every $x \in \operatorname{Soc} J$.*

Proof. The image of a minimal quadratic ideal by Φ is also a minimal quadratic ideal so, by the definition of the socle, it is invariant under Φ. Because $\Phi(1) = 1$ and x invertible is equivalent to $\Phi(x)$ invertible, then $\operatorname{Sp}\Phi(x) = \operatorname{Sp} x$. From the definition we see easily that

$$\operatorname{rank}(x) = \operatorname{rank}(\Phi(x)).$$

A similar argument proves that $m(\lambda, \Phi(x)) = m(\lambda, x)$. It is also possible to use the fact that the Riesz projection p associated to λ and x satisfies $m(\lambda, x) = \operatorname{rank}(p)$ ([9], Theorem 1.6) and to notice that the Riesz projection associated to λ and $\Phi(x)$ is $\Phi(p)$. Hence, from the definition of the trace, we have $\operatorname{Tr}\Phi(x) = \operatorname{Tr}(x)$. Since the socle is regular by [11], let $a = U_a b$ in the socle. Then $Da = \{a, b, Da\} + U_a(Db)$

is in the socle, since the socle is an ideal. If D is bounded, then $e^{\lambda D}$ is an automorphism of J for $\lambda \in \mathbb{C}$, consequently $\operatorname{Tr} e^{\lambda D}x = \operatorname{Tr} x$, for $\lambda \in \mathbb{C}$ and $x \in J$. If we set

$$f(\lambda) = \begin{cases} \frac{e^{\lambda D}-1}{\lambda}x, & \text{for } \lambda \neq 0, \\ Dx, & \text{for } \lambda = 0. \end{cases} \tag{1}$$

we know that $\lambda \to \operatorname{Tr} f(\lambda)$ is entire ([9], Theorem 2.1), moreover $\operatorname{Tr} f(\lambda) = 0$ for $\lambda \neq 0$, by additivity of the trace, so $\operatorname{Tr} f(0) = \operatorname{Tr} Dx = 0$. □

Corollary 1.2. *Let $a \in$ Soc $J, b, x, y \in J$. Then we have* $\operatorname{Tr}(aL_xL_yb + bL_xL_ya) = \operatorname{Tr}(aL_yL_xb + bL_yL_xa)$. *In particular* $\operatorname{Tr}(x(ya)) = \operatorname{Tr}(y(xa))$.

Proof. We take $D = [L_x, L_y]$, which is a bounded derivation. So by the previous theorem we have $\operatorname{Tr} D(ab) = 0$. Then using additivity of the trace we get the first formula. Taking $b = 1$ we get the second one. □

Corollary 1.3. *Let $a \in$ Soc J be such that* $\operatorname{Tr}(au) = 0$, *for every $u \in$ Soc J. Then $a = 0$.*

Proof. Let $x \in J$, $z \in$ Soc J be arbitrary. By Corollary 1.2 and the hypothesis we have $\operatorname{Tr}(x^2a)z = \operatorname{Tr} z(x^2a) = \operatorname{Tr} x^2(za) = \operatorname{Tr} x^2(az) = \operatorname{Tr} a(x^2z) = 0$, because $x^2z \in$ Soc J. For similar reasons we have $\operatorname{Tr}(x(xa))z = \operatorname{Tr} z(x(xa)) = \operatorname{Tr} x(z(xa)) = \operatorname{Tr} x((xa)z) = \operatorname{Tr}(xa)(xz) = \operatorname{Tr}(xz)(xa) = \operatorname{Tr} x((xz)a) = \operatorname{Tr} x(a(xz)) = \operatorname{Tr} a(x(xz)) = 0$. By additivity of the trace, we have proved that

$$\operatorname{Tr} U_x(a)z = 0, \quad \text{for } z \in \operatorname{Soc} J, \quad x \in J. \tag{2}$$

Taking for z the powers of $U_x(a)$, which are in Soc J, we conclude that

$$\operatorname{Tr}(U_x(a))^n = 0, \quad \text{for } x \in J, \quad n = 1, 2, \ldots. \tag{3}$$

Consequently by [9], Theorem 2.2 (iii), we have $U_x(a)$ quasi-nilpotent for every $x \in A$. Then by [2], Corollary 1, we have $a \in \operatorname{Rad} J$, that is $a = 0$. □

Let $a \in$ Soc J and $E(a) = \{x \in J : \#(\operatorname{Sp} U_x(a) \setminus \{0\}) = \operatorname{rank}(a)\}$. In [9], Theorem 1.1, we proved that this open set is dense in J. We now slightly improve this result. We recall that $a \in$ Soc A is said to be a *maximal finite-rank element* if $\#(\operatorname{Sp} a \setminus \{0\}) = \operatorname{rank}(a)$. This is the case for $U_x(a)$ if $x \in E(a)$.

Theorem 1.4. *The set $F(a)$ of invertible x in $E(a)$ such that* $\operatorname{rank} U_x(a) = \operatorname{rank} U_x(a)^2 = \operatorname{rank}(a)$ *is dense in the set of invertible elements of J. Consequently every $a \in$ Soc J is the limit of a sequence (a_n) of maximal finite-rank elements such that* $\operatorname{rank}(a_n^2) = \operatorname{rank}(a_n) = \operatorname{rank}(a)$.

Proof. Suppose there exists an open set V of invertible elements containing no x of $E(a)$ such $\operatorname{rank} U_x(a) = \operatorname{rank} U_x(a)^2$. By Theorem 3.5 and Corollary 3.6 of [5] we have $\operatorname{rank} U_x(a)^2 \leqslant \operatorname{rank} U_x(a) = \operatorname{rank}(a)$, because x is invertible. Let $r = \operatorname{rank}(a)$, then we have $\operatorname{rank} U_x(a)^2 \leqslant r - 1$ on $V \cap E(a)$, which is a non-empty open set, by density of $E(a)$. But $U_x(a)^2 = U_{U_x(a)}(1) = U_xU_a(x^2)$, consequently, by the same argument, we have $\operatorname{rank} U_x(a)^2 = \operatorname{rank} U_a(x^2)$, for $x \in V \cap E(a)$. Fixing $x_0 \in V \cap E(a), x \in J$ arbitrary and taking $f(\lambda) =$

$= x_0 + \lambda(x - x_0)$, by the Scarcity Theorem for the rank ([5]), Theorem 3.4) applied to rank $U_a(f(\lambda)^2)$ we conclude that rank $U_a(x^2) \leqslant r - 1$ for every $x \in J$, because this is true for $|\lambda|$ small. Now we take y arbitrary in J, then $y - \lambda$ is a square for $|\lambda| > \rho(y)$ (the spectral radius of y) by the Holomorphic Functional Calculus, so again applying the Scarcity Theorem for the rank to rank $U_a(y - \lambda)$, we conclude that rank $U_a(y) \leqslant r - 1$, for every $y \in J$. But we know that the socle is von Neumann regular ([11], Theorem 1), so there exists $y \in J$ such that $a = U_a(y)$, consequently rank $U_a(y) = \text{rank}\,(a) \leqslant r - 1$, which is absurd. So the first part is proved. To prove the second one, we just take $a_n = U_{x_n}(a)$ where $\lim x_n = 1$ and $x_n \in F(a)$. □

Remark. Exactly as in the associative case ([8], Theorem 2.8), A . Fernández López recently proved (in a private letter) that the maximal finite-rank elements have the form $\lambda_1 p_1 + \cdots + \lambda_n p_n$, where the λ_i are the spectral values and the p_i are orthogonal rank one projections. Then, using Corollary 2.3 of [9], we conclude that $F(a)$ coincides with the set of invertible elements of $E(a)$. So these new results simplify the previous argument.

Theorem 1.5. *Let $a \in$ Soc J and $x \in J$. Then we have* $\text{Tr}\, U_a(x) = \text{Tr}(a^2 x)$ *and* $\text{Tr}\, U_x(a) = \text{Tr}\,(x^2 a)$.

Proof. First step. The first identity is a consequence of Corollary 1.2, because $\text{Tr}\,(a(xa)) = \text{Tr}\,(x(aa))$, and of the additivity of the trace.

Second step. We suppose that a is a maximal finite-rank element and that rank $(a) = \text{rank}\ (a^2)$. Let $x \in E(a^2)$, then $U_x(a^2)$ is maximal so

$$\text{rank}\, U_x(a^2) = \#(\text{Sp}\, U_x(a^2) \backslash \{0\}) = \text{rank}\,(a^2).$$

This implies that each non-zero point of the spectrum of $U_x(a^2)$ has multiplicity one. By [5], Corollary 3.7, and the hypothesis, we have rank $U_a(x^2) \leqslant \text{rank}\,(a^2)$. By the Shifting Principle of McCrimmon ([14], Proposition 3 and the following remark) we have

$$\text{Sp}\, U_x(a^2) \backslash \{0\} = \text{Sp}\, U_a(x^2) \backslash \{0\}.$$

So this implies that $U_a(x^2)$ is also maximal, hence all its non-zero spectral values have multiplicity one. By definition of the trace we conclude that $\text{Tr}\, U_x(a^2) = \text{Tr}\, U_a(x^2)$, when $x \in E(a^2)$. The density of $E(a^2)$ in J and continuity of the trace on $\mathscr{F}_{2r}$, where $r = \text{rank}\,(a)$ ([9], Theorem 2.2 (ii)) implies that the same is true for every $x \in J$. Then by the first step we have

$$\text{Tr}\, U_x(a^2) = \text{Tr}\, U_a(x^2) = \text{Tr}\,(x^2 a^2), \tag{4}$$

for every $x \in J$.

Third step. Let $a \in$ Soc J be arbitrary, $r = \text{rank}\,(a)$. By Theorem 1.4, we can apply (4) to the sequence (a_n). Again by continuity of the trace on $\mathscr{F}_r$ we conclude that (4) is true for $a \in$ Soc J and $x \in J$.

Fourth step. Because J is a Jordan-Banach algebra the socle is a von Neumann regular ideal ([11], Theorem 2), so for $a \in$ Soc J there exists $b \in$ Soc J such that

$a = U_a(b)$. Using the polarization of the product $xy = \frac{1}{2}((x+y^2 - (x-y)^2)$ it is easy to conclude that a is a finite sum of squares of elements of the socle. So by additivity of the trace and formula (4) the theorem is proved. □

Professor Ottmar Loos suggested to the author that the trace is probably related to the Faulkner form defined on the socle [15]. The next corollary proves this fact. The argument only uses the first identity of Theorem 1.5.

Corollary 1.6. *Let F be the Faulkner bilinear form defined on the socle. Then $F(y,z) = \operatorname{Tr}(yz)$, for $y, z \in \operatorname{Soc} J$.*

Proof. By the Osborn-Racine theorem the socle is linearly spanned by the ideals $I(p)$ generated by the minimal projections p and $I(p)$ is linearly spanned by the elements $y = U_{x_1} \cdots U_{x_n}(p)$ which are in $\mathcal{F}_1$, by [5], Theorem 3.5 and Corollary 3.8, and which are *reduced* in the terminology of [15], that is they satisfy $U_y(J) \subset \mathbb{C}_y$. Consequently, because F and Tr are bilinear, it is enough to suppose y reduced and $z \in J$. From the definition of F we have

$$U_y(z) = F(y,z)y. \tag{5}$$

Consequently we have

$$(U_z(y))^2 = U_{U_{z(y)}}(1) = U_z U_y(z^2) = F(y,z^2)U_z(y). \tag{6}$$

So we have $\operatorname{Sp} U_z(y) \subset \{0, F(y,z^2)\}$. If $U_z(y)$ is nilpotent then by (6) we have $F(y,z^2) = 0$ and $0 = \operatorname{Tr} U_z(y) = \operatorname{Tr}(yz^2)$ by the first identity of Theorem 1.5. If $U_z(y)$ is not nilpotent then $\operatorname{Sp} U_z(y) = \{0, F(y,z^2)\}$ and the multiplicity of the non-zero spectral value is one, so $\operatorname{Tr} U_z(y) = \operatorname{Tr}(yz^2) = F(y,z^2)$. Now using the fact that every element of the socle is a finite sum of squares we obtain that $F(y,z) = \operatorname{Tr}(yz)$ for y reduced and $z \in \operatorname{Soc} J$. So, by linearity, the result is proved. □

It is known that F is nondegenerate on the socle, that is $\operatorname{Tr}(au) = 0$, for every $u \in \operatorname{Soc} J$ and some $a \in \operatorname{Soc} J$ implies $a = 0$. This is in fact Corollary 1.3, for which we gave a new proof. We now extend this result.

Denote by $(\operatorname{Soc} J)^{\perp}$ the *orthogonal of the socle* of J, that is the set of $a \in J$ such that $\operatorname{Tr}(au) = 0$, for every $u \in \operatorname{Soc} J$.

Lemma 1.7. *Let D be a bounded derivation on J. Then $(\operatorname{Soc} J)^{\perp}$ is invariant under D. Consequently, if D takes its values in the socle, then D is zero on $(\operatorname{Soc} J)^{\perp}$.*

Proof. Let $a \in (\operatorname{Soc} J)^{\perp}$ and $x \in \operatorname{Soc} J$. Then $D(ax) = a \cdot Dx + Da \cdot x$. By Theorem 1.1 we have $\operatorname{Tr} D(ax) = 0$. By hypothesis $\operatorname{Tr}(a \cdot Dx) = 0$ because $Dx \in \operatorname{Soc} J$. Consequently $\operatorname{Tr}(Da \cdot x) = 0$ for every $x \in \operatorname{Soc} J$ which proves the first part. If D takes its values in $\operatorname{Soc} J$ then $Da \in (\operatorname{Soc} J)^{\perp} \cap (\operatorname{Soc} J)$, which is zero by Corollary 1.3. □

Corollary 1.8. *Let $a \in (\operatorname{Soc} J)^{\perp}, x \in J, y \in \operatorname{Soc} J$. Then $y(xa) = x(ya)$.*

Proof. We apply the previous lemma to the derivation $D = [L_y, L_x]$ which takes its values in the socle. □

The notion of *annihilator* was introduced by E. ZEL'MANOV (see for instance [17]). For more information on annihilators, in particular the annihilator of the socle which is an ideal, see [10]. In this particular case $\mathrm{Ann}(\mathrm{Soc}\, J) = \{a \in J : a\, \mathrm{Soc}\, J = \{0\}\}$.

At that point the author expresses his best thanks to Professor Antonio Fernández López for all the necessary information on annihilators needed to prove the next theorem.

Theorem 1.9. *$(Soc\, J)^{\perp} = Ann\,(Soc\, J)$.*

Proof. It is obvious that $\mathrm{Ann}\ (\mathrm{Soc}\ J) \subset (\mathrm{Soc}\, J)^{\perp}$. So suppose that $\mathrm{Tr}\,(au) = 0$, for every $u \in \mathrm{Soc}\, J$. Let $x \in J$ be arbitrary. We have $u(ax) = x(ua)$ by Corollary 1.8. Moreover $\mathrm{Tr}\, x(ua) = \mathrm{Tr}\, x(au) = \mathrm{Tr}(xu) = 0$, by Corollary 1.2 and the hypothesis. So this implies that $\mathrm{Tr}(ax)u = 0$, for every $u \in \mathrm{Soc}\, J$. With additivity of the trace, this implies that $(\mathrm{Soc}\ J)^{\perp}$ is an ideal. By Corollary 1.3 we know that $\mathrm{Soc}\, J \cap (\mathrm{Soc}\, J)^{\perp} = \{0\}$. If $a \in (\mathrm{Soc}\, J)^{\perp}$ then $U_a\,(\mathrm{Soc}\, J) \subset \mathrm{Soc}\, J \cap (\mathrm{Soc}\, J)^{\perp} = \{0\}$, because Soc J and $(\mathrm{Soc}\ J)^{\perp}$ are ideals. But $U_a\,(\mathrm{Soc}\, J) = \{0\}$ implies that $a \in \mathrm{Ann}\ (\mathrm{Soc}\ J)$, by a standard result of Zel'manov [17]. □

Applying the results of [10, 12], the theorem implies that $(\mathrm{Soc}\ J)^{\perp} = \{0\}$ for primitive Jordan-Banach algebras with non-zero socle.

2. Spectrum Preserving Linear Mappings

In [5], Theorem 3.13, we proved that if J and J' are two semisimple Jordan-Banach algebras with identity and if T is a spectrum preserving linear mapping from J onto J', then T is a continuous linear bijection from J onto J' such that $T1 = 1$ and $T(\mathscr{F}_n) = \mathscr{F}'_n (n = 1, 2, \ldots)$, so in particular $T\,(\mathrm{Soc}\, J) = \mathrm{Soc}\, J'$. Incidentally the fact that J' is semisimple is a consequence of the other hypotheses, using the spectral characterization of the radical [2].

We now prove that T is not far from being a Jordan isomorphism.

Theorem 2.1. *Let T be a spectrum preserving linear mapping from J onto J'. Then $Tx^2 - (Tx)^2 \in \mathrm{Ann}\ (\mathrm{Soc}\ J')$, for every $x \in J$.*

Proof. Let $x, b \in J$ be fixed. For $|\mu|\rho(x) < 1$, by the Holomorphic Functional Calculus, the element $(1 - \mu x)^{1/2}$ is well-defined and invertible. The identity

$$1 - \mu x + b = U_{(1-\mu x)^{1/2}}(1 + U_{(1-\mu)^{-1/2}} b), \tag{1}$$

proves that $1 - \mu x + b$ is non-invertible if and only if $-1 \in \mathrm{Sp}\, U_{(1-\mu x)^{-1/2}} b$. But because T is spectrum preserving and $T1 = 1$ we have $1 - \mu x + b$ non-invertible iff $1 - \mu Tx + Tb$ is non-invertible. This implies that

$$\mathrm{Sp}\, U_{(1-\mu x)^{-1/2}} b = \mathrm{Sp}\, U_{(1-\mu Tx)^{-1/2}} Tb, \tag{2}$$

for $|\mu|\rho(x) < 1$ (incidentally $\rho(x) = \rho(Tx)$). Now we suppose that $b \in \mathscr{F}_1, b \neq 0$. Then $0 \neq Tb \in \mathscr{F}_1$ and $TU_{(1-\mu x)^{-1/2}} b, U_{(1-\mu x)^{-1/2}} Tb \in \mathscr{F}_1$ by [5], Theorem 3.5 and

$T\mathscr{F}_1) = \mathscr{F}'_1$. From this and (2) we obtain

$$\operatorname{Tr} TU_{(1-\mu x)^{-1/2}}b = \operatorname{Tr} U_{(1-\mu xTx)^{-1/2}}Tb, \tag{3}$$

for $|\mu|\rho(x) < 1$, because the two spectra of the elements appearing in (3) are equal and contain at most one non-zero point of multiplicity one. Using the identity $U_{1+u}y = y + 2uy + U_u y$ and the series of $(1-a)^{-1/2}$ we get

$$\operatorname{Tr}\left(TU_{1+\frac{\mu}{2}x+\frac{3}{8}\mu^2x^2+\ldots}b\right) = \operatorname{Tr}\left(U_{1+\frac{\mu}{2}Tx+\frac{3}{8}\mu^2(Tx)^2+\ldots}Tb\right). \tag{4}$$

By identification of the coefficients of μ and μ^2 we obtain

$$\operatorname{Tr} T(xb) = \operatorname{Tr}(Tx \cdot Tb) \tag{5}$$

$$\operatorname{Tr}[3t(x^2b) + TU_x(b)] = \operatorname{Tr}[3(Tx)^2Tb + U_{Tx}(Tb)]. \tag{6}$$

It is also clear that $\operatorname{Tr} TU_x(b) = \operatorname{Tr} U_x(b)$ because T preserves spectrum and rank. Using this remark and (5) we obtain

$$3\operatorname{Tr}(Tx^2 \cdot Tb) + \operatorname{Tr} U_x(b) = 3\operatorname{Tr}(Tx)^2Tb + \operatorname{Tr} U_{Tx}(Tb). \tag{7}$$

Now applying the second identity of Theorem 1.5 to $U_x(b)$ and $U_{Tx}(Tb)$ we obtain

$$\begin{aligned} 3\operatorname{Tr}(Tx^2 \cdot Tb) + \operatorname{Tr}(x^2b) &= 4\operatorname{Tr}(Tx)^2Tb = \\ &= 3\operatorname{Tr}(Tx^2 \cdot Tb) + \operatorname{Tr} T(x^2b) = 4\operatorname{Tr}(Tx^2 \cdot Tb), \end{aligned} \tag{8}$$

again by (5) and the fact the $\operatorname{Tr} T(x^2b) = \operatorname{Tr}(x^2b)$. Finally we have proved that

$$\operatorname{Tr}((Tx)^2 - Tx^2) \cdot Tb = 0, \tag{9}$$

for every $b \in \mathscr{F}_1$. Because the trace is additive and the socle is linearly generated by rank one elements (see the beginning of the proof of Corollary 1.6) this proves, by Theorem 1.9, that $(Tx)^2 - Tx^2$ is in the annihilator of the socle of J'. □

Corollary 2.2. *With the previous hypotheses, T is a Jordan isomorphism from* Soc J *onto* Soc J'.

Proof. This is obvious because $(Tx)^2 - Tx^2 \in \operatorname{Soc} J' \cap (\operatorname{Soc} J')^{\perp} = \{0\}$, by Corollary 1.3. □

Corollary 2.3. *If T satisfies the previous hypotheses and J' is a primitive Jordan-Banach algebra with non-zero socle, then T is a Jordan isomorphism from J onto J'.*

Proof. This is immediate from Theorem 2.1 because $\operatorname{Ann}(\operatorname{Soc} J') = \{0\}$. □

In the associative case it was proved (see [7], Lemma 3.6 and the proof of Theorem 3.7) that T is a Jordan isomorphism from $kh(\operatorname{Soc} J)$ onto $kh(\operatorname{Soc} J')$. It is possible to prove the same result in the case of Jordan-Banach algebras but more work has to be done. We recall that $kh(\operatorname{Soc} J)$ is the preimage of the radical of $J/\operatorname{Soc} J$. Since the radical of a Jordan algebra is the intersection of its maximal modular quadratic ideals we have $kh(\operatorname{Soc} J) = kh(\overline{\operatorname{Soc} J})$. The results of [1] have been extended to Jordan-Banach algebras by A. Maouche ([13], Theorem 2.3.2

and Corollaries 2.3.3–2.3.5) and T. J. D. WILKINS [16]. In particular they proved that $kh(\mathrm{Soc}\, J)$ is an inessential ideal of J, this means that the set of limit points of $\mathrm{Sp}\, a$ is included in $\{0\}$ for every a in $kh(\mathrm{Soc}\, J)$ and this implies that the Riesz projections associated to the non-zero spectral values of a are in the socle. Recently T. J. D. WILKINS [16] gave the following spectral characterization of $kh(\mathrm{Soc}\, J)$, inspired from the spectral characterization of the radical given in [2]. Namely $a \in kh(\mathrm{Soc}\, J)$ if and only if $\sup_{t\in\mathbb{C}} \rho(\mathrm{Sp}(x+ta)') < +\infty$, for every $x \in J$, where $\rho((\mathrm{Sp}\, y)')$ denotes the radius of the topological derivative of the spectrum of y. With all these prerequisites we are ready to prove the following.

Corollary 2.4. *With the previous hypotheses, T is a Jordan isomorphism from $kh(\mathrm{Soc}\, J)$ onto $kh(\mathrm{Soc}\, J')$.*

Proof. a) T maps $kh(\mathrm{Soc}\, J)$ onto $kh(\mathrm{Soc}\, J')$. Let $a \in kh(\mathrm{Soc}\, J)$ then $\rho(\mathrm{Sp}(x+ta)') = \rho(\mathrm{Sp}(Tx+tTa)')$, because T is additive and preserves spectrum. So, by Wilkins's result, $Ta \in kh(\mathrm{Soc}\, J')$. We know that T is injective and surjective, so arguing with T^{-1} this step is proved.

b) T is a Jordan isomorphism on $kh(\mathrm{Soc}\, J)$. For $a \in kh(\mathrm{Soc}\, J)$, by part a) and Theorem 2.1, we have $Ta^2 - (Ta)^2 \in I = kh(\mathrm{Soc}\ J') \cap \mathrm{Ann}\,(\mathrm{Soc}\ J')$. We now prove that the ideal I is zero. First we prove that $u \in I$ implies $\rho(u) = 0$. Suppose $\rho(u) \neq 0$, then there exists $\alpha \neq 0, \alpha \in \mathrm{Sp}\, u$. The non-zero Riesz projection p associated to u and α is in the socle of J'. Moreover $p = \frac{u}{2\pi i}\int_\Gamma \frac{1}{\lambda}(\lambda - u)^{-1} d\lambda$, where Γ is a small circle centred at α. Because $u \in \mathrm{Ann}\ (\mathrm{Soc}\ J')$, which is an ideal, we also have $p \in \mathrm{Ann}\ (\mathrm{Soc}\ J')$, hence by Corollary 1.3, $p = 0$ and this is a contradiction. The ideal I consists only of quasi-nilpotent 1.3, elements, so $I \subset \mathrm{Rad}\, J' = \{0\}$. Hence the corollary is proved. □

Modifying Harte's theorem and the argument of the proof of Theorem 3.7 of [7] and using the structure theorem for separable scattered Jordan-Banach algebras ([6], Theorem 19) it is then possible to extend result (iv) of the introduction to scattered Jordan-Banach algebras.

References

[1] AUPETIT B (1986) Inessential elements in Banach algebras. Bull London Math Soc **18**: 493–497

[2] AUPETIT B (1993) Spectral characterization of the radical in Banach and Jordan-Banach algebras. AMath Proc Camb Phil Soc **114**: 31–35

[3] AUPETIT B (1994) Recent trends in the field of Jordan-Banach algebras. In: ZEMÁNEK J (ed) Functional Analysis and Operator Theory, vol 30. Polish Academy of Sciences: Banach Center Publ

[4] AUPETIT B (1994) Analytic Multifunctions and Their Applications. In: GAUTHIER PM (ed) Complex Potential Theory. Kluwer

[5] AUPETIT B (1995) Spectral characterization of the socle in Jordan-Banach algebras. Math Proc Camb Phil Soc **117**: 479–489

[6] AUPETIT B, BARIBEAU L (1989) Sur le socle dans les algèbres de Jordan-Banach. Can J Math **41**: 1090–1100

[7] AUPETIT B, MOUTON H DU T. (1994) Spectrum preserving linear mappings in Banach algebras. Studia Math **109**: 91–100

[8] AUPETIT B, MOUTON H DU T. (1996) Trace and determinant in Banach algebras. Studia Math **121**: 195–200

[9] AUPETIT B, MAOUCHE A, MOUTON H DU T. (preprint) Trace and determinant in Jordan-Banach algebras

[10] Fernández López A (1992) On annihilators in Jordan algebras. Publ Matemàtiques **36**: 569–589
[11] Fernández López A, Rodríguez Palacios A (1986) On the socle of a non-commutatives Jordan algebra. Manuscripta Math **56**: 269–278
[12] Fernández López A, Rodríguez Palacios A (1986) Primitive non-commutative Jordan algebras with non-zero socle. Proc Amer Math Soc **96**: 199–206
[13] Maouche A (1994) Théorie spectrale dans les systèmes de Jordan-Banach. Thesis, Univ Laval, Québec
[14] McCrimmon K (1971) A characterization of the radical of a Jordan algebra. J Algebra **18**: 101–111
[15] McCrimmon K (1985) Reduced elements in Jordan triple systems. J Algebra **97**: 540–564
[16] Wilkins TJD (1997) Inessential ideals in Jordan-Banach algebras. Bull London Math Soc **136**: 73–81
[17] Zel'manov E (1987) Goldie's theorem for Jordan algebras. Siberian Math J **28**: 44–52

B. Aupetit
Département de Mathématiques et de Statistique
Faculté des Sciences et de Génie
Université Laval
Québec G1K 7P4
Canada

Mh. Math. 125, 189–209 (1998)

Monatshefte für
Mathematik

Printed in Austria

Closed Curves and Geodesics with Two Self-Intersections on the Punctured Torus

By

David Crisp, Adelaide, **Susan Dziadosz**, Ann Arbor, MI, **Dennis J. Garity**, Corvallis, OR, **Thomas Insel**, Berkeley, CA, **Thomas A. Schmidt**, Corvallis, CA, and **Peter Wiles**, Madison, WI

With 12 Figures

(Received 11 March 1996; in revised form 15 January 1997)

Abstract. We classify the free homotopy classes of closed curves with minimal self intersection number two on a once punctured torus, T, up to homeomorphism. Of these, there are six primitive classes and two imprimitive. The classification leads to the topological result that, up to homeomorphism, there is a unique curve in each class realizing the minimum self intersection number. The classification yields a complete classification of geodesics on hyperbolic T which have self intersection number two. We also derive new results on the Markoff spectrum of diophantine approximation; in particular, exactly three of the imprimitive classes correspond to families of Markoff values below Hall's ray.

Introduction

We classify the free homotopy classes of loops on a once–punctured torus, T, whose self-intersection number is two. Our classification is up to homeomorphism type; that is, we identify two free homotopy classes if there is a self homeomorphism of the punctured torus taking one class to the other. There are eight such classes. This classification has topological, geometric and number theoretic consequences.

The classification leads to the topological result that each homeomorphism class of free homotopy classes of intersection number two contains a unique curve, up to homeomorphism, realizing the minimum intersection number. In addition, all but two of the classes are distinguished by the closures of the complementary domains of any curve in the class that realizes the minimum self intersection number. In particular, if two curves have self intersection number two and are not freely homotopic to curves with lower self intersection number, and if they are in the same class, then there is a homeomorphism of T taking one curve to the other.

A geometric consequence of the classification is that the classes we obtain include the classes which, for any hyperbolic metric, admit a closed geodesic of

1991 Mathematics Subject Classification 53A35, 57M50; 11J06
Key words: geodesic, punctured torus, Markoff spectrum, diophantine approximation
Research started during the Summer 1994 NSF REU Program at Oregon State University and partially supported by NSF DMS 9300281

two self-intersections. Thus, given a hyperbolic metric on T, we obtain a classification of geodesics on T with two self intersections. Up to homeomorphism of T, and thus up to automorphism of the fundamental group of T, there are six such classes.

In work which responds to C. SERIES' closing question in [15], D. CRISP and W. MORAN [6], [5], [7] classified the singly self-intersecting geodesics of a particular hyperbolic punctured torus; they then proceeded to obtain results about the Markoff spectrum of diophantine approximation. We use our classification of doubly self-intersecting geodesics on T to obtain further information about the Markoff spectrum.

The simple curves on a punctured torus are well studied: for the use of studying the generators of the fundamental group and the action of its automorphisms [11], [3]; for identifying Teichmüller space [13], [18]; and for their relationship to the Markoff spectrum of diophantine approximation [4] (for an overview), [9], [16]. That the geodesics of a hyperbolic punctured torus are rather special is well known, see [19] for a convincing theorem to this effect. That these geodesics are somehow fundamental in the study of more general surfaces is demonstrated in [2; Chapter 3].

We also point out here that classifications of the nature we obtain are actually valid for all hyperbolic metrics on the punctured torus. We mention that our method easily generalizes to the case of higher self-intersection numbers. Although dependent upon earlier works, our approach is somewhat new.

This paper is an outgrowth of an NSF REU project at Oregon State University. Crisp and Schmidt thank Bill Moran for earlier related conversations.

In Section 1, we give the topological background for this project and in Section 2, we give the geometric and number theoretic background. In Section 3, we obtain the actual classification. The case analysis involved uses standard combinatorial topological methods. The proof that the classes obtained are distinct is more interesting and is postponed until a later section.

In Section 4, we give the topological consequences of the classification, in Section 5 we give the geometric consequences, in Section 6 we prove the distinctness of the classes and in Section 7 we discuss the Markoff values associated to the geodesics for a particular hyperbolic metric. We close with some questions and suggestions for further research.

1. Topological Background

Let T be a punctured torus. The fundamental group of T, $\pi_1(T)$, is isomorphic to the free group on two letters, $F(a,b)$. We fix such an isomorphism. There is a natural bijection between free homotopy classes of closed curves on T and conjugacy classes of elements of $F(a,b)$.

A loop is said to have a single nontrivial self-intersection if the loop has a single transverse intersection, and is not freely homotopic to a simple loop. A loop is said to have two nontrivial self-intersections if it has two transverse self-intersection points, and if it is not freely homotopic to a loop with a single nontrivial self-intersection or to a simple loop. A free homotopy class of loops is

said to have intersection number two if it contains a loop with two nontrivial self-intersections.

Our initial goal is to classify those free homotopy classes with intersection number two. The classification will be as follows. We view a loop as a map from S^1 into T. Given a free homotopy class which contains a loop ℓ with two nontrivial intersections, we show that the class is completely determined by the pair $(T, \ell(S^1))$. That is, if ℓ' is another loop with two nontrivial self intersections in the class, then there is a homeomorphism h of pairs taking $(T, \ell(S^1))$ to $(T, \ell'(S^1))$. Moreover, this homeomorphism can be chosen so as to preserve orientation in the sense that there is an orientation preserving homeomorphism k of S^1 so that $h \circ \ell = \ell' \circ k$.

We also show that a free homotopy class of intersection number two is completely determined by the pattern of closures of the complements of a loop ℓ realizing the minimum intersection number.

We state a result that follows directly from the classification of surfaces. See BIRMAN and SERIES [1] for a discussion of this and other results about simple loops on surfaces.

Theorem 1.1. *The conjugacy class of a simple loop l on T is either*
(a) *the identity, and l bounds a disc,*
(b) *one of $[aba^{-1}b^{-1}]$ or $[bab^{-1}a^{-1}]$, and l bounds punctured disc, or*
(c) *$[w]$ where w is a generator of $\pi_1(T)$, and l does not separate T.*

In [6], [5] and [7], a classification of loops with a single nontrivial self-intersection is given. The methods used are to apply the above result to each of the subloops.

Theorem 1.2. (Crisp). *The conjugacy class in $\pi_1(T)$ of a loop with a single nontrivial self-intersection on T is one of*
(a) *$[(aba^{-1}b^{-1})^2]$ or $[(aba^{-1}b^{-1})^{-2}]$,*
(b) *$[g(a^2)]$ for some $g \in$ Aut $\pi_1(T)$,*
(c) *$[g(abab^{-1})]$ for some $g \in$ Aut $\pi_1(T)$, or*
(b) *$[g(aaba^{-1}b^{-1})]$ for some $g \in$ Aut $\pi_1(T)$.*
Conversely, each of these conjugacy classes contains such a loop.

2. Geometric Background

We now let T be a punctured torus with a hyperbolic metric. A free homotopy class of closed curves on T is said to be primitive if it is not a non-trivial power of some other class.

It is well known that a closed curve on T is freely homotopic to a geodesic if and only if the curve lies in a primitive free homotopy class which contains no simple curve bounding either a disc or a punctured disc. Furthermore, a geodesic has the minimal number of self-intersections amongst all curves in its free homotopy class.

J. NIELSEN [11] showed that every automorphism of $\pi_1(T)$ can be realized, up to inner automorphism, by some homeomorphism. He further showed that every

automorphism of $\pi_1(T) \cong F(a,b)$ takes $aba^{-1}b^{-1}$ to $x(aba^{-1}b^{-1})^{\pm 1}x^{-1}$ for some $x \in F(a,b)$.

Theorem 2.1. *Let T_1 and T_2 be two hyperbolic punctured tori. Let (A_1, B_1) and (A_2, B_2) be fixed generating pairs of their respective fundamental groups. Let n be a positive integer. Let $W(a,b)$ be a word in the letters a and b. If a geodesic on T_1 in the free homotopy class of $W(A_1, B_1)$ has n intersections, then any geodesic on T_2 in the free homotopy class of $W(A_2, B_2)$ also has n intersections.*

Proof. There is a homeomorphism between T_1 and T_2, call it h. The images of the free homotopy classes $[A_1]$ and $[B_1]$ are, say, $[\alpha]$ and $[\beta]$. Similarly, $[W(A_1, B_1)]$ is sent to $[W(\alpha, \beta)]$. Since h is a homeomorphism, these have the same minimal number of self-intersections.

Since h induces an isomorphism on π_1, α and β are a generating pair for $\pi_1(T_2)$. In particular, there is an automorphism of $\pi_1(T_2)$ taking (α, β) to (A_2, B_2). This automorphism sends $[W(\alpha, \beta)]$ to $[W(A_2, B_2)]$. But, the automorphism is realizable by a homeomorphism, hence the minimal number of self-intersections in these classes must be the same. Since the geodesic in each of these classes achieves this minimal number of self-intersections, we are done. □

Corollary 2.2. *A classification for a particular hyperbolic punctured torus of the automorphism classes of those free homotopy classes which contain closed geodesics which are n-times self-intersecting is in fact valid for all such punctured tori.*

Remark. We have thus shown that the results of [6], as proven in [5] and [7], are in fact valid for all hyperbolic punctured tori. Indeed, the classification is fundamentally topological. We point out that it has consequences even for arbitrary Riemannian metrics on the punctured torus: a class which for a hyperbolic metric has a unique closed geodesic with m self-intersections, will for an arbitrary Riemannian metric admit a geodesic of m self-intersections; such a class may admit geodesics with more self-intersections, but never less. Thus, a classification of the classes which for a hyperbolic metric admit a geodesic with up to m self-intersections identifies the set of classes which for an arbitrary Riemannian metric may admit geodesics of m self-intersections. □

For the classification of twice self-intersecting geodesics, we specialize to a *fixed* metric only as one means for showing that two homeomorphism classes of geodesics are distinct. The surface we then use is the quotient of the Poincaré upper half-plane by the commutator subgroup of the modular group, exactly that studied by the above mentioned authors. Indeed, we will also give number theoretic implications of our classification by interpreting them with respect to this particular metric.

2.1 Specialization to the Γ'-Metric

We consider the particular once-punctured torus **T**, the quotient of the Poincaré upper half-plane, $\mathscr{H}$, by the commutator subgroup of the modular group, Γ'. This torus has constant curvature minus one – thus, there is at most one geodesic in

each free homotopy class – and admits $\mathscr{H}$ as its universal Riemannian cover. The action of Γ on $\mathscr{H}$ is given by linear fractional transformations. We use a standard fundamental domain $\mathscr{D}$ for this action – a quadrilateral with hyperbolic line segment boundaries of vertices $-1, 0, 1, \infty$.

We take $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$ as generators of $\Gamma' \cong F(A, B)$. Thus, given a word W in A and B, by matrix multiplication we find a corresponding matrix in Γ'. A hyperbolic element of Γ' is one of trace greater than 2 in absolute value. Each such fixes an axis – a geodesic of $\mathscr{H}$, thus a semi-circle with center on $\mathbb{R} \cup \{\infty\}$. Free homotopy classes correspond to conjugacy classes in Γ'. A free homotopy class has a closed geodesic in it if and only if the corresponding conjugacy class is hyperbolic. The geodesic in such a homotopy class is then the projection of the axes fixed by the elements of the conjugacy class. Indeed, the geodesic is the 1-1 projection of those segments of the fixed axes which lie in $\mathscr{D}$, where we note that $\mathscr{D}$ has certain sides paired by the actions of A and B. Furthermore, for a reduced word in A and B representing an element of the conjugacy class, the cyclic permutations of this word determine all axes which will have geodesic segments within $\mathscr{D}$. We refer to [5] for a more detailed presentation of these standard facts.

Thus, to find the number of self-intersections of the unique geodesic in the class $[W]$, for a given (hyperbolic) word W in A and B, we find the number of intersections within $\mathscr{D}$ of the axes of those matrices which arise from the cyclic permutations of the word of W. That is, for each cyclic permutation of W we find the corresponding element $M \in \Gamma'$ as a matrix. There is a straightforward formula for the endpoints $p_1, p_2 \in \mathbb{R}$ of the axis of M. Thus, we find each axis. We then simply count the number of intersections which lie in $\mathscr{D}$ of these finitely many axes. Note that this determination of the number of self-intersections of the geodesic of W has thus been reduced to arithmetic.

Thus, by Theorem 2.1, we can determine the self-intersection number of any free homotopy class for a general once-punctured torus. In what follows we will refer to the result of such a calculation as coming from *specialization to the* Γ'*-metric.*

2.2 Geodesics Mapped to Geodesics

We will have need to decide whether a specific pair of free homotopy classes are equivalent up to the action of the outer automorphisms of the fundamental group. As we will discuss in Section 5, there are various ways to address this problem. One is to specialize to a particular hyperbolic metric and to consider the action of the Teichmüller modular group.

As discussed in say [10], the Teichmüller modular orbit of a hyperbolic punctured torus arises from a relabeling of generators of the fundamental group. That is, the action of the outer automorphisms of the fundamental group are realized by that of the Teichmüller modular group. Moreover, within an orbit, this action can be given by way of isometries of the universal Riemannian cover, $\mathscr{H}$. In particular, the action of the automorphisms can be realized by diffeomorphisms of hyperbolic punctured tori which send geodesics to geodesics.

Thus, for geodesics with respect to a fixed hyperbolic metric to be related under some outer automorphism, the complements of the geodesics must be of the same topological type. This is a true strengthening over the mere assumption that outer automorphisms be realized by homeomorphisms – under this, we merely know that a geodesic in a class is sent to some curve in the class of the second; free homotopy could allow for vastly different complements.

This brief discussion thus indicates that two free homotopy classes are not equivalent under the action of the outer automorphisms of the fundamental group if under some specialization to a hyperbolic metric the complements of their respective geodesics have differing topological type. We will use this in Section 5.

2.3 Geometry of the Markoff Spectrum

In the case of the Γ'-metric, the simple closed geodesics are known to be in 1–1 correspondence with the Markoff numbers of Diophantine approximation. Briefly, the Markoff spectrum can be considered as the set of suprema of Euclidean diameters of lifts of the geodesics of $\Gamma'\backslash\mathscr{H}$. It is known that there is a ray approaching infinity, the Hall ray, of values in this spectrum; for further details, see the discussion of [8]. A. HAAS [9] remarked that any geodesic which has a loop about the puncture will have Markoff value at least 6. But, 6 is known to be in the Hall ray of the Markoff spectrum. The lowest part of the spectrum corresponds to the simple closed geodesics. The intermediate section of the Markoff spectrum is still rather mysterious.

Another characterization of the Markoff spectrum is in terms of continued fractions. Recall that a real number can be expanded in a continued fraction of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots}},$$

where a_0 is an integer and the remaining a_i are positive integers. We use a flat notation for this continued fraction expansion: $[a_0; a_1, \ldots]$. Given a doubly infinite sequence of positive integers, $\mathscr{A} = \ldots, a_{-1}, a_0, a_1, \ldots$, one defines $\lambda_i(\mathscr{A})$ to be $[a_i; a_{i+1}, \ldots] + [0; a_{i-1}, a_{i-2}, \ldots]$. The Markoff value of $\mathscr{A}$ is the supremum of the $\lambda_i(\mathscr{A})$. The above mentioned Markoff numbers is the discrete set of values with unique limit value 3, given by certain, well-characterized, periodic doubly infinite sequences of 1, 1's and 2, 2's. For further details, we again refer to [8].

The above indicates that the Markoff values associated to the simple closed geodesics for the Γ'-metric are isolated within the Markoff spectrum. The Crisp-Moran conjecture, introduced in [6], is that all Markoff values corresponding to the single self intersecting geodesics which contain no loop about the puncture are also isolated within the Markoff spectrum. Thus, there are important aspects of the Markoff values associated to geodesics of the Γ'-metric which are known to depend only upon the homeomorphism class of the geodesics. This is what in part motivated the present work.

3. Loops with Two Self-Intersections

Our goal in this section is to classify loops on T with two nontrivial self-intersections. We consider the composition of three simple loops, as justified by the following lemma, obtained by the authors Dziadosz, Insel, and Wiles.

Lemma 3.1. *Up to free homotopy, any loop l on T with k transverse self-intersection points can be formed as the composition of $k+1$ simple loops, which intersect at only one point.*

Proof. Consider a loop l on T with k transverse self-intersections occurring at distinct points. Call a segment that part of the loop which is minimally between two (possibly nondistinct) points of intersection. We claim that there are $2k$ distinct segments connecting the intersections of l. Indeed, as four (possibly non-distinct) segments converge at each intersection and each segment connects two intersections, there are exactly $4k/2 = 2k$ total distinct segments connecting the self-intersections of l.

Now, pick some point p on T. For each self-intersection of l, we can continuously deform l so that the intersection occurs at p, in effect collapsing a segment of l. In order to bring all intersections to one point, we perform $k-1$ collapsings, leaving $2k-(k-1) = k+1$ segments. □

The lemma shows that it is sufficient to consider the compositions of three simple loops in order to determine all loops with two transverse self-intersections. In addition, the following Corollary to this Lemma limits the number of cases that need to be considered. This Corollary follows immediately from the manner in which the three simple loops were obtained in the Lemma.

Corollary 3.2. *Let l be a loop on T formed as the composition of 3 simple loops which intersect only at a single point p. Assume that l arises as in the Lemma from a loop $\tilde{l}$ with 2 transverse intersections. Then the point p has a neighborhood U in T such that $U \cap l$ is a wedge of six segments with wedge point p as in Figure 1. Moreover, this wedge has a unique "axis" c–d as in Figure 1 along which the collapsing was done to obtain l from $\tilde{l}$, and $\tilde{l}$ can be recovered from l by reversing the collapsing along the axis c–d.*

Corollary 3.3. *Let l and U be as in Corollary 3.2. Suppose that l is the composition of loops l_1, l_2 and l_3 of basepoint p of T. For $i \in \{1,2,3\}$, let i and i' label respectively the initial and final segments of $l_i \cap U$. The resulting graph, which we call a basepoint graph, is one of the graphs in Figure 2.*

Proof. We use a cyclic ordering on $\{1, 2, 3\}$. From the graph of $l \cap U$, labelling some segment as i' also determines the segment labelled $i+1$. Moreover, the axis

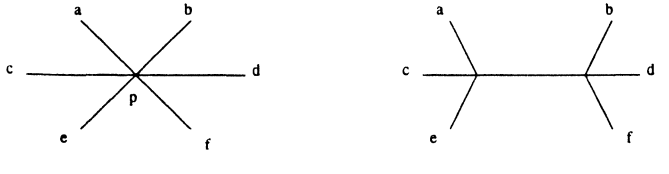


Figure 1. Basic wedge pattern

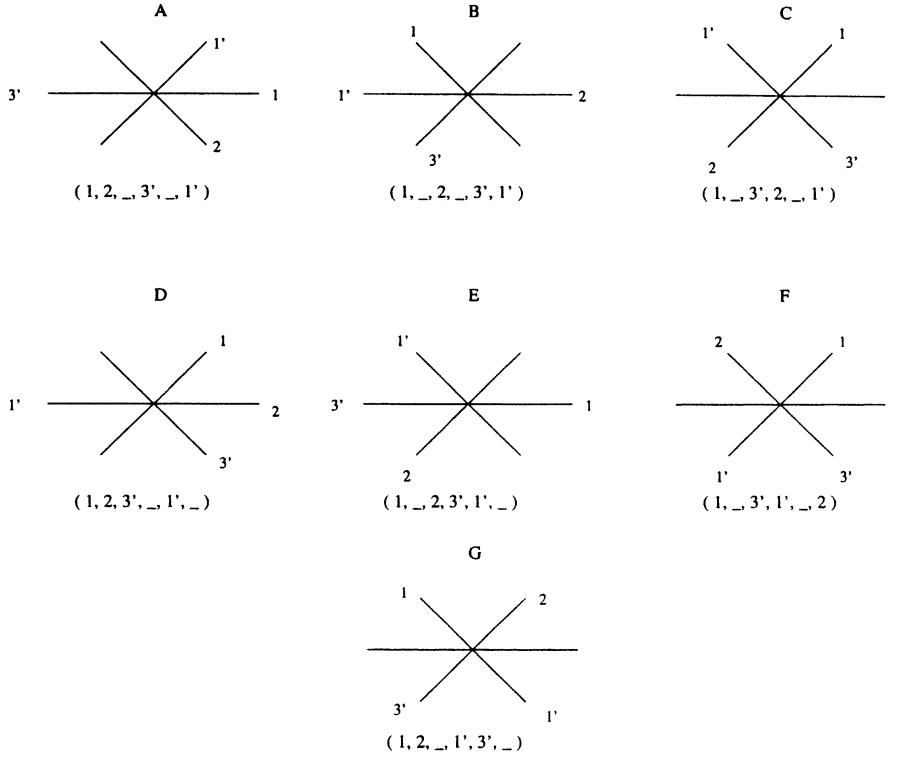


Figure 2. Possible basepoint graphs

must be labelled $(j', j+1)$ for some j. One now lists all possible basepoint graphs. We label clockwise, beginning with 1, and consider the number of segments between $1'$ and 1: $1'$ and 1 are adjacent (A, B, C); separated by one segment (D, E); by two (F, G). Note that in each of these cases there are two unlabeled segments, allowing for a choice of orientation. □

We will refer to an *analysis of basepoint segment pattern* as the process of determining which $\{A, \ldots, G\}$ could come from a given configuration of (often unoriented) loops l_i. In practice, this will allow us to determine which orientations of the l_i are possible. In general, this process will be in two steps. Since the l_i are (in general) initially known only up to orientation, we first consider j' as equivalent to j and write out the possible sequences of $j \in \{1, 2, 3\}$ coming from the pattern of the l_i. Thus, each j appears twice in each such sequence. This is followed by determining the possible ways to label one of each pair of j with a prime so as to achieve one of the basepoint graphs in $\{A, \ldots, G\}$.

We continue now by classifying the possible compositions of three simple loops that can arise from loops with two transverse intersections.

Theorem 3.4. *Let l be a loop on T formed as the composition of* 3 *simple loops which intersect only at a single point. Assume that l arises as in Lemma* 3.1 *from a*

loop $\tilde{l}$ with precisely 2 transverse intersections. Then the conjugacy class of l in $\pi_1(T)$, $[l]$, is one of

(a) $[g(aaba^{-1}b^{-1}aba^{-1}b^{-1})]$,
(b) $[g(aaaba^{-1}b^{-1})]$,
(c) $[g(aaabb)]$,
(d) $[g(abab^{-1}aba^{-1}b^{-1})]$,
(e) $[g(aaba^{-1}a^{-1}b^{-1})]$,
(f) $[g(aabab^{-1})]$,
(g) $[g(a^3)]$,
(h) $[g((aba^{-1}b^{-1})^3)]$,
(i) *the conjugacy class of a loop with one non-trivial self-intersection, or*
(j) *the conjugacy class of a simple loop*

for some $g \in \text{Aut } \pi_1(T)$.

Proof. Let $l = l_1 l_2 l_3$, where l_1, l_2 and l_3 are simple loops on T. We can move the common intersection point of the simple loops to the basepoint of $\pi_1(T)$ by an isotopy of T. Let w, w_1, w_2 and w_3 be the homotopy classes of l, l_1, l_2 and l_3 respectively. The corresponding free homotopy classes are $[w]$, $[w_1]$, $[w_2]$ and $[w_3]$.

Note that $[w] = [w_1 w_2 w_3] = [w_3 w_1 w_2] = [w_2 w_3 w_1]$ since cyclic permutations are conjugate by the obvious elements of the free homotopy group.

In the following proof, we use without explicit mention the topology of simple curves on the torus. ROLFSEN's book on knot theory contains the necessary results [12]. Often, after cutting apart the torus along a non-separating curve, we work with simple curves on the resulting punctured disc. The possibilities that arise for the various curves come from repeated application of the Schoenflies theorem.

By Theorem 1.1, each of l_1, l_2 and l_3 can be either the identity, a generator, or a loop around the puncture. Thus, the following cases arise:

Case 1. For some i, $[w_i] = [\text{Id}]$;

Case 2. Each of $[w_1]$, $[w_2]$, $[w_3] \in \{[aba^{-1}b^{-1}], [bab^{-1}a^{-1}]\}$;

Case 3. For some i, $[w_i] = [g(a)]$ and $\{[w_j] | j \neq i\} \subset \{[aba^{-1}b^{-1}], [bab^{-1}a^{-1}]\}$;

Case 4. Exactly two of $[w_1]$, $[w_2]$, $[w_3]$ are of the form $[g(a)]$ for some automorphism g of $\pi_1(T)$, and the third is in $\{[aba^{-1}b^{-1}], [bab^{-1}a^{-1}]\}$; and

Case 5. All of the homotopy classes represent generators.

Since the details involved in the above cases are similar, we provide a complete argument only for Case 1, Case 2 and a few subcases of Case 4.

Case 1: Some $[w_i] = [\text{Id}]$.

Without loss of generality, $i = 1$. In this case, $[w] = [w_2 w_3]$, the conjugacy class of either a simple loop or a loop with one non-trivial self-intersection. In fact, Corollary 3.2 shows that if we recover the loop $\tilde{l}$ with two nontrivial intersections from l, then the part of $\tilde{l}$ corresponding to l_1 is as pictured in Figure 3 where the region enclosed by the curve is a disc on T. In the first case in this figure, one intersection can be removed by a free homotopy and the loop is freely homotopic

Figure 3. Figure for Case 1

to a loop with a single transverse self intersection. In the second case in this figure, both intersections can be removed by a free homotopy and the loop is freely homotopic to a simple loop.

Case 2: $[w_1], [w_2], [w_3] \in \{[aba^{-1}b^{-1}], [bab^{-1}a^{-1}]\}$.

Each of the loops l_i is a loop bounding a punctured disc. We know that $[aba^{-1}b^{-1}]$ and $[bab^{-1}a^{-1}]$ are the free homotopy classes of simple loops around the puncture, and that they have opposite orientations. Without loss of generality, l_1 bounds a disc around the puncture that contains the punctured discs bounded by l_2 and l_3. An analysis of the basepoint segment pattern shows that only case A with pattern (1, 2, 3, 3′, 2′, 1′) or case B with pattern (1, 3, 2, 2′, 3′, 1′) from Figure 2 are possible. In each of these cases, $w_1 = w_2 = w_3$ and

$$[w] = [w_1^3] = [g((aba^{-1}b^{-1})^3)]. \tag{h}$$

For Case 4, we make use of methods from [5]. First, note that we can induce an automorphism between the w_i and words in a and b. Typically, we use $g(a) = w_j$ and $g(b) = w_k$ for $j \neq k$, but occasionally we use more complicated automorphisms in order to simplify a word. In addition, by an automorphism g, we actually mean *some* automorphism g. For example, $[g(a)]$, refers to any generator in $\pi_1(T)$.

The second method we adopt is the technique of cutting T along some loop l whose image in $\pi_1(T)$ is a generator to obtain a disk bounded by l containing the puncture and a hole also bounded by l. We draw the resulting disc as in Figure 4, with the loop l oriented clockwise around the outer boundary. Whenever possible, we choose the loop l that we cut T open along in such a way that the other loops have their base point at the left side of resulting figure as in Figure 4. Finally, we frequently introduce an additional path whose image in $\pi_1(T)$ forms a basis with the homotopy class of the loop bounding this disk.

Case 4: Two of $[w_1], [w_2], [w_3]$ are of the form $[g(a)]$ for some automorphism g of $\pi_1(T)$, and the third is in $\{[aba^{-1}b^{-1}], [[bab^{-1}a^{-1}]\}$.

Three subcases arise. The *first* is that the two homotopy classes that represent generators form a generating pair for $\pi_1(T)$. The second is that the two homotopy classes that represent generators do not form a generating pair for $\pi_1(T)$, and that these two homotopy classes are either equal or are inverses of each other. The third

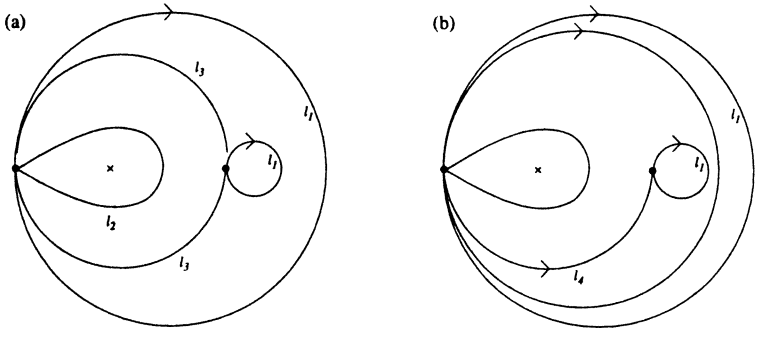


Figure 4. The torus T cut along l_1 in the first and second parts of Case 4.

is that the two homotopy classes that represent generators do not form a generating pair for $\pi_1(T)$, and that these two homotopy classes are neither equal to each other nor are inverses of each other. We provide details only for the second subcase.

Second Subcase: Assume that the two homotopy classes that represent generators do not form a generating pair for $\pi_1(T)$, and that these two homotopy classes are either equal or are inverses of each other. Without loss of generality, l_1 is one of the homotopy classes representing a generator and the torus may be cut along l_1 to obtain a figure with the other loops positioned as in Figure 4 (b). Define a path l_4, disjoint except for the base point from l_1, l_2 and l_3 such that w_1 and w_4 form a generating pair for $\pi_1(T)$. An analysis of the basepoint segment pattern shows that only case *A* with pattern (1, 2, 3, 3′, 2′, 1′) or case B with pattern (1, 3, 2, 2′, 3′, 1′) are possible. So the other loop that represents a generator is homotopic to l_1 as indicated in the figure. In the first case $w_2 = w_1$ and $w_3 = w_1 w_4 w_1^{-1} w_4^{-1}$ and in the second case, $w_3 = w_1$ and $w_3 = w_1 w_4 w_1^{-1} w_4^{-1}$.

Thus,

$$[w] = [w_1 w_2 w_3] = [w_1 w_1 w_1 w_4 w_1^{-1} w_4^{-1}] = [g(aaaba^{-1}b^{-1})], \tag{b}$$

or

$$[w] = [w_1 w_2 w_3] = [w_1 w_1 w_4 w_1^{-1} w_4^{-1} w_1] = [g(aaba^{-1}b^{-1}a)]. \tag{b}$$

A conjugation shows that the above two classes are equal.

A similar analysis of the first subcase gives

$$[w] = [w_1 w_2 w_3] = [g(ab^{-1}aba^{-1}b)] \tag{c}$$

Similarly, in the *third subcase* one finds that either

$$[w] = [w_1 w_2 w_3] = [g(abab^{-1}aba^{-1}b^{-1})], \tag{d}$$

or

$$[w] = [w_1 w_2 w_3] = [g(aba^{-1}a^{-1}b^{-1}a)]. \tag{e}$$

Case 3: A similar analysis yields

$$[w] = [w_1 w_2 w_3] = [g(aaba^{-1}b^{-1}aba^{-1}b^{-1})]. \tag{a}$$

Case 5: One finds

$$[w] = [w_1 w_2 w_3] = [g(a)], \tag{j}$$

$$[w] = [w_1 w_2 w_3] = [g(a^3)], \tag{g}$$

$$[w] = [w_1 w_2 w_3] = [g(bab^{-1}aa)], \tag{f}$$

$$[w] = [w_1 w_2 w_3] = [g(a^2)], \tag{i}$$

or

$$[w] = [w_1 w_2 w_3] = [g(ab^{-1}abb)]. \tag{c}$$

We have completed the classification of the compositions of three simple loops that arise from loops with two transverse intersections. It remains to show that the classes that we obtained are distinct. This is done after the next two sections on the discussion of topological and geometric applications of this result. We now state the following result on curves with two nontrivial intersections.

Theorem 3.5. *The conjugacy class in $\pi_1(T)$ of a loop on T with two non-trivial self-intersections is one of*

(a) $[g(aaba^{-1}b^{-1}aba^{-1}b^{-1})]$,
(b) $[g(aaaba^{-1}b^{-1})]$,
(c) $[g(aaabb)]$,
(d) $[g(abab^{-1}aba^{-1}b^{-1})]$,
(e) $[g(aaba^{-1}a^{-1}b^{-1})]$,
(f) $[g(aabab^{-1})]$,
(g) $[g(a^3)]$, or
(h) $[g((aba^{-1}b^{-1})^3)]$

for some $g \in$ Aut $\pi_1(T)$. Moreover, these classes are distinct.

Proof. We apply Theorem 3.4 and eliminate the simple loops, and loops with a single non-trivial self-intersection.

The only candidates remaining are (a)–(h) above. It is sufficient to demonstrate that each of these classes does not contain a simple loop or a loop with a single non-trivial intersection. Since each of classes (c) through (h) is primitive, specialization to the Γ' metric shows that two is the minimal self intersection number for each of these classes. Classes (a) and (b) are the non-primitive classes and a simple algebraic argument along with Theorems 1.1 and 1.2 shows that they do not contain simple loops or loops with a single non-trivial intersection. As mentioned above, the proof of distinctness is postponed until after a discussion of the geometric and topological consequences of the classification. □

4. Topological Consequences of the Classification

Theorem 4.1. *Each of the classes in Theorem* 3.5 *contains, up to homeomorphism, a unique curve of self intersection number* 2. *That is, if l and l' are closed curves of self intersection number two in one of the classes from Theorem* 3.5, *then there is a homeomorphism h of T and an orientation preserving homeomorphism k of S^1 so that $h \circ l = l' \circ k$.*

Proof. To obtain this result, it suffices to reexamine the proof of Theorem 3.4 and to recover the original curves with two self intersections by reversing the collapsing as described in Corollary 3.2. During the analysis carried out in the proof of Theorem 3.4, cyclically permuting the three curves making up the loop corresponds to applying an orientation preserving homeomorphism of S^1 prior to mapping S^1 into the punctured torus. The classes in Theorem 3.4 arise in a number of different cases during the analysis. In every instance, the different cases corresponding to the same class yield the same topological type of curve on reversing the collapsing process. The curves obtained are those in Figures 5 through 12 with one of the two possible orientations.

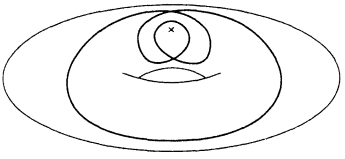


Figure 5. A loop on T in the free homotopy class $[aaba^{-1}b^{-1}aba^{-1}b^{-1}]$.

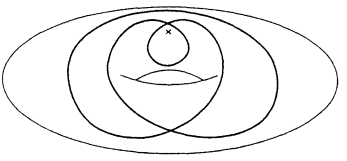


Figure 6. A loop on T in the free homotopy class $[aaaba^{-1}b^{-1}]$.

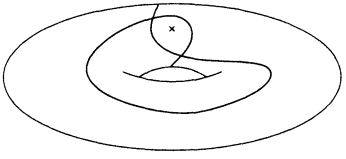


Figure 7. A loop on T in the free homotopy class $[aaabb]$.

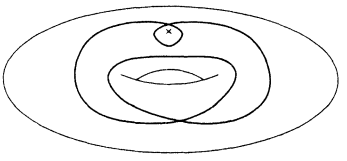


Figure 8. A loop on T in the free homotopy class $[abab^{-1}aba^{-1}b^{-1}]$.

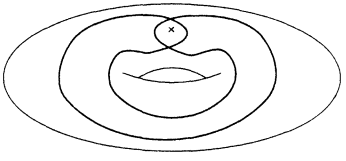


Figure 9. A loop on T in the free homotopy class $[aaba^{-1}a^{-1}b^{-1}]$.

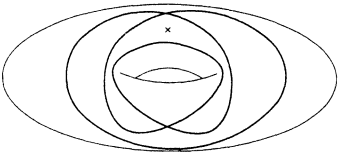


Figure 10. A loop on T in the free homotopy class $[aabab^{-1}]$.

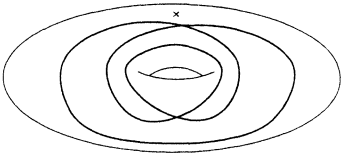


Figure 11. A loop on T in the free homotopy class $[a^3]$.

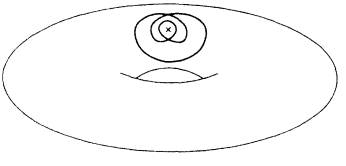


Figure 12. A loop on T in the free homotopy class $[(aba^{-1}b^{-1})^3]$.

Note that for each of the figures, there is a self homeomorphism of the punctured torus that takes the given curve to itself with orientation reversed. The result now follows.

Note that in every diagram but the third, there are two possible segments to collapse to obtain a wedge of three simple loops as in Lemma 3.1. In the third case, there are four possible segments to collapse. This leads to this case occurring in apparently different subcases of Theorem 3.4. □

Theorem 4.2. *Let l be a closed curve of self intersection number 2 on the punctured torus T. Assume that l is not freely homotopic to a curve with fewer self intersections. Then for all but two of the free homotopy classes listed in Theorem 3.5, the class of l is completely determined by closures of the complementary domains of the image of l in T. For the remaining two classes, the class of l is determined by the pattern of intersections of the closures of the complementary domains of the image of l.*

Proof. It suffices to list the components and closures of components and to note the pattern of intersections. This is done in Table 1 below.

Note that all the classes above with the exception of the fourth and sixth are distinguished by the closures of the components of the complements of the curves. The fourth and sixth class are distinguished as follows. In the fourth class, the intersection of the closures of the first and second listed components is a single point. In the sixth class, the same intersection is the entire boundary of the closed punctured disc. □

In the following table, c represents $aba^{-1}b^{-1}$.

Table 1. Topological types of twice self intersecting loops

Class	Number of Components	Components	Closure of Components
a^3	3	punctured open cylinder open disc open disc	closed cylinder pinched cylinder pinched cylinder
c^3	4	punctured open disc open disc open disc punctured open cylinder	punctured closed disc pinched cylinder pinched cylinder closed one handle
ac^2	3	punctured open disc open disc open cylinder	punctured closed disc pinched cylinder closed one handle
a^2c	3	punctured open disc open disc open cylinder	punctured closed disc pinched cylinder closed pinched one handle
a^3b^2	2	punctured open disc open disc	punctured closed disc closed one handle
$abab^{-1}c$	3	punctured open disc open disc open cylinder	punctured closed disc pinched cylinder closed pinched one handle
$aaba^{-2}b^{-1}$	3	punctured open disc open disc open cylinder	punctured closed disc punctured closed cylinder closed cylinder
$aabab^{-1}$	3	punctured open disc open disc open cylinder	punctured pinched cylinder pinched cylinder closed cylinder

5. Geometric Consequences of the Classification

We identify those classes which contain geodesics with two self-intersections. This proof parallels Theorem 3.2 in [5], but is included for completeness.

Theorem 5.1. *A closed geodesic on T has two self-intersections if and only if it is in one of the following free homotopy classes*

(a) $[g(aaba^{-1}b^{-1}aba^{-1}b^{-1})]$,
(b) $[g(aaab^{-1}b^{-1})]$,
(c) $[g(aaabb)]$,
(d) $[g(abab^{-1}aba^{-1}b^{-1})]$,
(e) $[g(aaba^{-1}a^{-1}b^{-1})]$, *or*
(f) $[g(aabab^{-1})]$

for some $g \in \mathrm{Aut}\,(F(a,b))$. *Any two geodesics in the same class are of the same topological type. Moreover, these free homotopy classes are distinct.*

Proof. Closed geodesics on T lie in primitive free homotopy classes and realize the minimum number of self-intersections for their free homotopy classes. We apply Theorem 3.5 and exclude the non-primitive classes $[g((aba^{-1}b^{-1})^3)$ and $[g(a^3)]$ Since the remaining classes are primitive, each class contains a geodesic. By the topological classification in the preceding section, there is only one topological type of geodesic in each class.

It remains to be shown that these classes are distinct. This follows from the result in the next section. □

6. Distinctness of the Classes

Theorem 6.1. *The classes obtained in Theorem 3.5 are all distinct.*

Proof. We indicate several means of proof. First, note that $[g(a^3)]$ and $[g(aba^{-1}b^{-1})^3]$, being imprimitive, are distinct from the other classes. The aforementioned result of Nielsen shows that they are also distinct from each other.

Approach A. An algorithm for deciding whether there is an automorphism of $F(a,b)$ taking a specific primitive word w_1 to a specific primitive word w_2 has been described by J.H.C. Whitehead [17]. One can apply this in our situation and prove the Theorem.

Approach B. We now use the fact that every automorphism of a punctured torus can be realized by a diffeomorphism which preserves the set of geodesics under a given metric. The background for this was briefly summarized in Section 2.2; the main implication is that two primitive free homotopy classes can be in the same homeomorphism class only if the topological type of the associated geodesics for any fixed hyperbolic metric are the same. In particular, we can specialize to the Γ'-metric and find that the information of Table 1 guarantees that the Theorem is true.

Approach C. Note that two loops on T cannot be in automorphic conjugacy classes in $\pi_1(T)$ if they are not in automorphic conjugacy classes on the usual, non-punctured, torus T'. When we remove the puncture from T, loops can be deformed as follows:

(i) $[aaba^{-1}a^{-1}b^{-1}]$ is in $[\mathrm{Id}]'$

(ii) $[aaba^{-1}b^{-1}aba^{-1}b^{-1}]$ and $[aaabb]$ are in $[g(a)]'$

(iii) $[aaaba^{-1}b^{-1}]$ and $[abab^{-1}aba^{-1}b^{-1}]$ are in $[g(a^2)]'$

(iv) $[aabab^{-1}]$ is in $[g(a^3)]'$

for some $g \in \mathrm{Aut}\,\pi_1(T)$. This enables us to distinguish all but two pairs of classes.

We show directly that there are no automorphisms taking aCC to $aaabb$, nor aaC to $abab^{-1}C$, where $C := aba^{-1}b^{-1}$. To this end, we recall two results of J. NIELSEN [11]: (1) a cyclically reduced word W in a and b is a generator of $F(a, b)$ only if all exponents of a in W are of the same sign, similarly for those of b; (2) automorphisms preserve the set of conjugates of $C^{\pm 1}$.

If in the first of these cases there were such an automorphism ϕ, then we could solve for $\phi(a) = aaabbV$, with $V = XC^{\pm 2}X^{-1}$. This right hand side cannot reduce to have either of a or b appear to the same exponent sign throughout. Since an automorphic image of a is a generator, this right hand side must not be cyclically reduced. But this implies that there is an automorphism ψ such that $\psi(a) = aabbaY\,C^{\pm 2}Y^{-1}$. Again, this cannot be reduced so as to have same sign exponents of a nor b. Indeed, proceeding cyclically, we find that there can be no such ϕ.

In the second case, an initial automorphism allows us to consider the existence of a ϕ with $\phi(a^2) = abbaCW$, with $W = X\,C^{\pm 1}X^{-1}$. But, if W is not C^{-1}, there can be no reduction so as to produce a right hand side of same sign exponents in either a or b. Hence, either $\phi(a^2) = abba$, or this right hand side is not cyclically reduced. Proceeding through all possible cyclic permutations, we find that $\phi(a^2)$ must be equal to one of the cyclic permutations of $abba$.

Now, if some $Z = w_1 U w_2$ for letters w_i and word U, then the reduced word of Z^2 is either $w_1 U w_2 w_1 U w_2$ or $w_1 U^2 w_2$. This latter case arises only if $w_1 = w_2^{-1}$. It is easily checked that for the above choices, $\phi(a^2)$ cannot have the form of a square. Hence, no such automorphism exists. □

7. Applications to the Markoff Spectrum

Recall from the discussion in Section 2.3 that A. Haas has shown that a closed geodesic for the Γ'-metric corresponds to a Markoff value greater than 6 if and only if the geodesic contains a loop about the puncture. We thus conclude the following.

Lemma 7.1. *Of all twice self-intersecting geodesics on $\Gamma'\backslash\mathscr{H}$, only those which are automorphic images of (i) A^3B^2, (ii) A^2BAB^{-1} or (iii) $A^2BA^{-2}B^{-1}$ have Markoff values less than 6.*

Our new examples of geodesics has led us to the following

Theorem 7.2. *Let W(A, B) be a cyclically reduced primitive word in A and B of positive exponents. Then the Markoff values of the images of W under the*

automorphisms of $F(A,B)$ lie between $\sqrt{5}$ and 6. Furthermore, 3 is a limit value of this set of Markoff values.

Proof. Since $W(A,B)$ is a primitive word in A and B of positive exponents, it clearly has no occurrence of the commutator of A and B within it. In particular, the corresponding geodesic does not contain a loop about the puncture. Therefore neither do any of the geodesics corresponding to automorphic images of W. But, by the remark of Haas, this implies that the corresponding Markoff values are then at most 6. It is well known that the minimum Markoff value is $\sqrt{5}$.

Let $C = B^{-1}A^{-1}B^{-1}$ and $D = B^{-1}$. Then (C,D) is a generating pair for $F(A,B)$. In the Γ'-metric, C acts as $Cz = [2; 2, z]$ and D as $Dz = [1; 1, z]$. Thus words in positive powers of C and D correspond to (doubly infinite) sequences whose entries are double 1s and double 2s.

The work of [5], [6] and [7] shows that there is a binary tree of isometries, so called Dehn twists, generated by λ and ρ; branching to the left is given by λ which replaces C by CD and fixes D, while branching to the right is given by ρ which replaces D by CD and fixes C. Now, since the automorphism group of Γ' is generated by λ and ρ, the corresponding tree of images of $W(C,D)$ is indeed infinite.

Each of these images of W has a corresponding periodic sequence of double 1s and double 2s; furthermore, as one penetrates deep into the tree, the sequences have ever longer subsets which agree with the periods corresponding to the simple geodesics. Thus, the images of W correspond to Markoff values approaching the limit value of the Markoff values of the simple geodesics. But, this unique limit value is well known to be 3. □

Corollary 7.3. *The automorphism classes of (i) A^3B^2 and (ii) A^2BAB^{-1} have Markoff values which admit 3 as a limit.*

Proof. That A^3B^2 meets the hypotheses of the Theorem is obvious. As to A^2BAB^{-1}, we note that the automorphism given by $(A,B) \to (AB,B)$ converts this to a word in positive powers of A and B. □

We now address the third class of "low height" twice self-intersecting geodesics.

Proposition 7.4. *The automorphic images of $A^2BA^{-2}B^{-1}$ have Markoff values between 3 and 6. Furthermore, 6 is the unique limit value of this set of Markoff values.*

For typographical reasons, in what follows, we use $\bar{X}$ to denote the inverse of a group element X.

Proof. The automorphism $(A,B) \to (DC,\bar{D})$, sends $A^2B\bar{A}^2\bar{B}$ to $(DC)^2(\bar{D}\bar{C})^2$. We again restrict to the tree of images corresponding to repeated applications of the automorphisms λ and ρ as above. Let α be one of the resulting composite automorphisms. That is, suppose $\alpha = x_1 \cdots x_n$ where each x_i is λ or ρ.

By cyclically permuting, we replace $(DC)^2(\bar{D}\bar{C})^2$ by $C\bar{D}\bar{C}\bar{D}\bar{C}DCD$. We will consider the effect of applying α. Using induction, it is not hard to verify that

$\alpha(CD) = CWD$ where W is a symmetrical word in positive powers of C and D. Therefore, $(\alpha(CD))^{-1} = \bar{D}\bar{W}\bar{C}$.

Claim. For all α as above, $\alpha(C\bar{D}\bar{C}\bar{D}\bar{C}DCD) = C\bar{D} \cdot \bar{C}\bar{W}\bar{D} \cdot \bar{C}D \cdot CWD$.

The proof is by induction on n, the "length" of α. It thus suffices to check applications of λ and ρ to words of the form indicated. Applying λ, we have

$$\begin{aligned}\lambda\alpha(C\bar{D}\bar{C}\bar{D}\bar{C}DCD) &= \lambda(C\bar{D} \cdot \bar{C}\bar{W}\bar{D} \cdot \bar{C}DCWD)\\ &= C\bar{D} \cdot \bar{C}\lambda(\bar{W})\bar{D}\bar{D} \cdot \bar{C}D \cdot CD\lambda(W)D.\end{aligned} \tag{7.1}$$

Since $\lambda\alpha(CD) = \lambda(CWD) = CD\lambda(W)D$, we find $W_{\lambda\alpha} = D\lambda(W)$ and we are done in this case. Similarly, applying ρ, we have

$$\begin{aligned}\rho\alpha(C\bar{D}\bar{C}\bar{D}\bar{C}DCD) &= \rho(C\bar{D} \cdot \bar{C}\bar{W}\bar{D} \cdot \bar{C}D \cdot \bar{C}WD)\\ &= C\bar{D} \cdot \bar{C}^2\rho(\bar{W}\bar{D} \cdot \bar{C}D \cdot C\rho(W)CD.\end{aligned} \tag{7.2}$$

But, $W_{\rho\alpha} = \rho(W)C$ and our claim follows.

We can now describe the transformation $\alpha(C\bar{D}\bar{C}\bar{D}\bar{C}DCD)$ as a continued fraction. We have $\alpha(C\bar{D}\bar{C}\bar{D}\bar{C}DCD) = C\bar{D} \cdot \bar{C}\bar{W}\bar{D} \cdot \bar{C}D \cdot CWD$. One easily checks that

$$\begin{aligned}C\bar{D}\bar{C}(z) &= [5; 2, -1/z];\\ \bar{D}(z) &= -1/([1; 1, -1/z]);\\ \bar{C}DC(z) &= -1/([5; 2, z]);\\ D(z) &= [1; 1, z].\end{aligned} \tag{7.3}$$

Since W is a word in positive powers of C and D, we also have

$$W(z) = [a_0; a_1, \ldots, a_n, z]$$

where $a_0; a_1, \ldots, a_n$ is a sequence of 1,1's and 2,2's. It follows that

$$\bar{W}(z) = -1/([a_n, a_{n-1}, \ldots, a_0, -1/z]).$$

Using the symmetry of W we therefore have $\bar{W}(z) = -1/([a_n, a_{n-1}, \ldots, a_0, -1/z])$.

We now compose the various continued fractions to conclude that

$$\begin{aligned}\alpha(C(\bar{D}\bar{C})^2DCD)(z) &= C\bar{D} \cdot \bar{C}\bar{W}\bar{D} \cdot \bar{C}D \cdot CWD(z)\\ &= [5; 2, a_0, a_1, \ldots, a_n, 1, 1, 5, 2, a_0, a_1, \ldots, a_n, 1, 1, z].\end{aligned} \tag{7.4}$$

The associated doubly infinite sequence of integers is $\mathscr{A} = (5, 2, a_0, a_1, \ldots, a_n, 1, 1)^\infty$. Now recall that $\alpha(CD) = CWD$ and that this word corresponds to a simple closed geodesic, whose Markoff value hence is a Markoff number – one of a discrete set of values lying between $\sqrt{5}$ and 3. Of course, $\alpha(CD)(z) = CWD(z) = [2; 2, a_0, a_1, \ldots, a_n, 1, 1, z]$, of associated sequence $\mathscr{B} = (2, 2, a_0, a_1, \ldots, a_n, 1, 1)^\infty$. It follows that the Markoff value of $\mathscr{A}$ is exactly the Markoff value of $\mathscr{B}$ plus 3. □

Remark. It is likely that each of the above two classes which admit 3 as a limit of their Markoff values actually admit this as their unique limit. However, we have included only the above proof of this nature. □

We now discuss a property of possible arithmetic interest which differentiates the two classes which admit 3 as a limit of their Markoff values. M. Sheingorn and one of the authors has noted that many of the families of isolated values in the Markoff spectrum correspond to geodesics which enjoy a special symmetry. Briefly, A. Schmidt pointed out that every hyperbolic punctured torus admits a unique Riemann surface which it double covers. The covered surface is a punctured sphere with 3 points of ramification, each of order two. (Indeed, there is an elliptic involution on the torus; dividing by the involution, we find the sphere. The expected fourth point of ramification corresponds to the puncture itself.) In the case of Γ', this punctured sphere is $\Gamma^3\backslash\mathscr{H}$, where Γ^3 is the normal subgroup generated by all of the cubes in the full modular group. We call a closed geodesic of this surface a *bouncer* if it passes through two of the elliptic fixed points of order two. SHEINGORN [16] showed that the simple closed geodesics of $\Gamma'\backslash\mathscr{H}$ project to be bouncers; [14] shows that the proper singly self-intersecting geodesics of [6] do as well; in fact, various infinite families of known isolated values correspond to bouncers [14].

Of the twice self-intersecting geodesics, we now show that only those which are images of A^3B^2 are bouncers. It is known that Γ^3 is generated by the three elliptic elements, each of order two, with no non-trivial relations amongst them:

$$T_0 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T_1 = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 2 & -5 \\ 1 & -2 \end{pmatrix}. \tag{7.5}$$

It is easily checked that $A = T_1T_2T_1T_0$ and $B = T_0T_1$. Note that since each T_i is its own inverse, each of A and B is the product of two elliptic elements (elements of trace 0) in Γ^3. This shows that the corresponding geodesics are indeed bouncers. It is fairly straightforward to check that the automorphisms of Γ' preserve the set of elements which are products of two elliptics of Γ^3. Thus, we can now write out words for each of cases (c), (e) and (f) in terms of A and B, substitute to obtain words in the T_i and check. (There are perhaps more elegant ways to perform this check, but the present manner is probably the most accessible.) For an explicit member of this type, we have

$$A^3B^2 = T_1(T_2T_1T_0T_1)T_2(T_1T_0T_1T_2)T_0T_1. \tag{7.6}$$

Thus, by conjugating by T_1, we see that the corresponding geodesic bounces from the projection of the elliptic fixed point of T_2 to that of T_0.

One easily checks the other classes, so as to conclude the

Proposition 7.5. *Of the twice self-intersecting geodesics on $\Gamma'\backslash\mathscr{H}$, only those which are automorphic images of A^3B^2 project to geodesics of $\Gamma^3\backslash\mathscr{H}$ which bounce between two elliptic fixed points of order two.*

Whereas [14] shows that each proper singly self-intersecting geodesic of [6] is forced to have its point of intersection at one of the three pre-images of the points

of ramification, it is not the case that these bouncing 2-self-intersectors have these as their points of self-intersection. A quick calculation readily shows this for the geodesic of A^3B^2.

Remark. We note that the results of even this section are quite strongly independent of hyperbolic metric. A. HAAS [9] showed that for any such metric, the simple closed geodesics have a particular "height" as their limit. Similarly, for a fixed hyperbolic metric, the property of a geodesic having a loop about the puncture can be expressed in terms of a fixed height (basically, this is given by the translation width at infinity of the corresponding group). Furthermore, the work of M. SHEINGORN [16] indicates that a word corresponding to a geodesic with respect to a fixed hyperbolic metric on a punctured torus which projects to a bouncing geodesic on the punctured sphere double covered by the torus will have this same property with respect to any hyperbolic metric. □

8. Final Comments

The techniques described in this paper can be used to obtain a classification of curves with higher self intersection number than two on the punctured torus. The analysis becomes more complicated because the number of cases increases.

It would be interesting to see if it is always the case that for a fixed n, the free homotopy classes of self-intersection number n are distinguished by their topological types (including complements).

Our results on the Markoff spectrum also lead to various questions. It would be quite interesting were it the case that one of the automorphism classes would give isolated Markoff values, as conjectured by Crisp and Moran for the proper single self intersectors.

References

[1] BIRMAN J, SERIES C (1984) An algorithm for simple curves on surfaces. J London Math Soc (2) **29**: 331–342
[2] BUSER P (1992) Geometry and Spectra of Compact Riemann Surfaces. Boston: Birkhäuser
[3] BUSER P, SEMMLER K-D (1988) The geometry and spectrum of the one holed torus. Comment Math Helv **63**: 259–274
[4] COHN H (1993) Markoff geodesics in matrix theory. In: POLLINGTON A, MORAN W (eds) Number Theory with an Emphasis on the Markoff Spectrum, pp 69–82. New York: Dekker
[5] CRISP D (1993) The Markoff Spectrum and Geodesics on the Punctured Torus. PhD Thesis Univ of Adelaide
[6] CRISP D, MORAN W (1993) Single self-intersection geodesics and the Markoff spectrum. In: POLLINGTON A, MORAN W (eds) The Markoff Spectrum, Diophantine Analysis and Analytic Number Theory, pp. 83–94. New York: Dekker
[7] CRISP D, MORAN W (1995) The Markoff spectrum and geodesics with one self-intersection on the punctured torus. Flinders Univ Preprint
[8] CUSICK TW, FLAHIVE ME (1989) The Markoff and Lagrange Spectra. Providence, RI: Amer Math Soc
[9] HAAS A (1988) Diophantine approximation on hyperbolic orbifolds. Duke Math J **56**: 531–547
[10] MASKIT B (1989) Parameters for Fuchsian groups II: topological type (1, 1). Ann Acad Sci Fennicae **14**: 365–375
[11] NIELSEN J (1918) Die Isomorphismen der allgemeinen unendlichen Gruppe mit zwei Erzeugenden. Math Ann **78**: 385–397
[12] ROLFSEN D (1990) Knots and Links. Houston, Texas: Publish or Perish

[13] SCHMIDT A (1976) The minimum of quadratic forms with respect to Fuchsian groups, I. J Reine Angew Math (Crelle) **287**: 341–368
[14] SCHMIDT TA, SHEINGORN M (1997) Markoff geometry on $\Gamma^3\backslash\mathscr{H}$. Oregon State Univ Preprint
[15] SERIES C (1985) The geometry of Markoff numbers. Math Intel **7**(3): 20–29
[16] SHEINGORN M (1985) Characterization of simple closed geodesics on Fricke surfaces. Duke Math J **52**: 535–545
[17] WHITEHEAD JHC (1936) On equivalent sets of elements in a free group. Ann Math **37**: 782–800
[18] WOLPERT S (1983) On the Kähler form of the moduli space of once punctured tori. Comment Math Helv **58**: 246–256
[19] ZIESCHANG H (1986) Minimal geodesics of a torus with a hole (Russian). Izv Acad Sci USSR **50**: 1097–1105; English transl. in: Math USSR Izv **29**: 449–457 (1987)

D. CRISP
Flinders University
Department of Mathematics
Adelaide 5001
Australia

T. INSEL
University of California
Department of Mathematics
Berkeley, California 94720
USA

S. DZIADOSZ
University of Michigan
Department of Mathematics
Ann Arbor, Michigan 48109
USA

T. A. SCHMIDT
Oregon State University
Department of Mathematics
Kidder Hall 368
Corvallis, Oregon 97331-4605
USA

D. J. GARITY
Oregon State University
Department of Mathematics
Kidder Hall 368
Corvallis, Oregon 97331-4605
USA

P. WILES
University of Wisconsin
Department of Mathematics
Madison, Wisconsin 53706
USA

Mh. Math. 125, 211–217 (1998)

Monatshefte für
Mathematik

Lower Bounds for the Discrepancy of Triples of Inversive Congruential Pseudorandom Numbers with Power of Two Modulus

By

Jürgen Eichenauer-Herrmann, Darmstadt, and **Harald Niederreiter,** Wien

(Received 10 April 1996)

Abstract. This paper deals with the inversive congruential method with power of two modulus m for generating uniform pseudorandom numbers. Statistical independence properties of the generated sequences are studied based on the distribution of triples of successive pseudorandom numbers. It is shown that there exist parameters in the inversive congruential method such that the discrepancy of the corresponding point sets in the unit cube is of an order of magnitude at least $m^{-1/3}$. The method of proof relies on a detailed analysis of certain rational exponential sums.

1. Introduction

Nonlinear congruential methods of generating uniform pseudorandom numbers in the interval [0, 1) have been studied intensively during the last years. Reviews of the development of this area can be found in the survey articles [2, 3, 5, 10, 13, 15] and in the monograph [14]. These surveys indicate that particularly attractive nonlinear congruential methods are based on multiplicative inversion in modular arithmetic. The present paper concentrates on the important case of a power of two modulus $m = 2^\omega$ with some integer $\omega \geqslant 4$. Let $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ for integers $n \geqslant 1$, and write $\mathbb{Z}_n^*$ for the set of all odd integers in $\mathbb{Z}_n$. For $y_0 \in \mathbb{Z}_m^*$ and *parameters* $a, c \in \mathbb{Z}_m^*$ with $a \equiv 1 \pmod 4$, an *inversive congruential sequence* $(y_n)_{n \geqslant 0}$ of elements of $\mathbb{Z}_m^*$ is defined by

$$y_{n+1} \equiv ac^2 y_n^{-1} + 2c \pmod m, n \geqslant 0,$$

where z^{-1} denotes the multiplicative inverse mod m of $z \in \mathbb{Z}_m^*$. A sequence $(x_n)_{n \geqslant 0}$ of *inversive congruential pseudorandom numbers* in the interval [0,1) is obtained by $x_n = y_n/m$ for $n \geqslant 0$. It follows from [1] that these sequences are purely periodic with the maximum possible period length $m/2$, i.e., $\{y_0, y_1, \ldots, y_{(m/2)-1}\} = \mathbb{Z}_m^*$. Equidistribution and statistical independence properties of the generated sequences have been studied in [4, 7, 8, 11]. Statistical independence properties

1991 Mathematics Subject Classification: 65C10, 11K45
Key words: uniform pseudorandom numbers, inversive congruential method, statistical independence, discrepancy of triples, rational exponential sums

are very important in many stochastic simulations. They can be analysed based on the discrepancy of tuples of successive pseudorandom numbers. For N arbitrary points $\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_{N-1} \in [0,1)^k$ the *discrepancy* is defined by

$$D_N(\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_{N-1}) = \sup_J |F_N(J) - V(J)|,$$

where the supremum is extended over all subintervals J of $[0,1)^k$, $F_N(J)$ is N^{-1} times the number of points among $\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_{N-1}$ falling into J, and $V(J)$ denotes the k-dimensional volume of J. In the present paper, lower bounds for the discrepancy of triples of inversive congruential pseudorandom numbers are established. The main results are presented and discussed in the third section. The second section contains several auxiliary results.

2. Auxiliary Results

For real t, the abbreviation $e(t) = e^{2\pi i t}$ is used, and $\mathbf{u} \cdot \mathbf{v}$ stands for the standard inner product of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^k$. First, a known auxiliary result is stated which follows from [12, Lemma 1; 14, Corollary 3.17].

Lemma 1. *The discrepancy of N arbitrary points $\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_{N-1} \in [0,1)^k$ satisfies*

$$D_N(\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_{N-1}) \geqslant \frac{\pi}{2N((\pi+1)^\ell - 1) \prod_{j=1}^k \max(1, |h_j|)} \left| \sum_{n=0}^{N-1} e(\mathbf{h} \cdot \mathbf{t}_n) \right|$$

for any nonzero lattice point $\mathbf{h} = (h_1, \ldots, h_k) \in \mathbb{Z}^k$, where ℓ denotes the number of nonzero coordinates of $\mathbf{h}$.

Now, for integers u, v, w a *rational exponential sum* is defined by

$$S(u,v,w;a,c;m) = \sum_{z \in \mathbb{Z}_m^*} e(c(uz + vaz^{-1} + waz(2z+a)^{-1})/m),$$

where the parameters a, c, m belong to the underlying inversive congruential generator. In the following, let $\omega = 3\nu + \mu$ with suitable integers $\nu \geqslant 1$ and $\mu \in \{0,1,2\}$.

Lemma 2. *Let u, v, w be integers with $v \equiv 2 \pmod 4$. Then*

$$S(u,v,w;a,c;m) = \sum_{x \in \mathbb{Z}_{2^\nu}^*} e(c(ux + vax^{-1} + wax(2x+a)^{-1})/m)$$
$$\cdot \sum_{y \in \mathbb{Z}_{2^{\omega-\nu}}} e\Big(c((u - vax^{-2} + wa^2(2x+a)^{-2} + 2^{2\nu+1})y$$
$$+2^\nu a(vx^{-3} - 2wa(2x+a)^{-3})y^2)/2^{\omega-\nu}\Big).$$

Proof. Straightforward calculations show that

$$
\begin{aligned}
&S(u,v,w;a,c;m)\\
&\quad=\sum_{x\in\mathbb{Z}_{2^\nu}^*}\sum_{y\in\mathbb{Z}_{2^{\omega-\nu}}} e\Big(c\Big(u(x+2^\nu y)+va(x+2^\nu y)^{-1}\\
&\qquad+wa(x+2^\nu y)(2x+a+2^{\nu+1}y)^{-1}\Big)/m\Big)\\
&\quad=\sum_{x\in\mathbb{Z}_{2^\nu}^*}\sum_{y\in\mathbb{Z}_{2^{\omega-\nu}}} e\Big(c\big(u(x+2^\nu y)+va(x^{-1}-2^\nu x^{-2}y+2^{2\nu}x^{-3}y^2-2^{3\nu}x^{-4}y^3)\\
&\qquad+wa(x+2^\nu y)((2x+a)^{-1}-2^{\nu+1}(2x+a)^{-2}y+2^{2\nu+2}(2x+a)^{-3}y^2)\big)/m\Big)\\
&\quad=\sum_{x\in\mathbb{Z}_{2^\nu}^*}\sum_{y\in\mathbb{Z}_{2^{\omega-\nu}}} e\Big(c\big(u(x+2^\nu y)+va(x^{-1}-2^\nu x^{-2}y+2^{2\nu}x^{-3}y^2)+2^{3\nu+1}y\\
&\qquad+wa(x(2x+a)^{-1}+2^\nu a(2x+a)^{-2}y-2^{2\nu+1}a(2x+a)^{-3}y^2)\big)/m\Big)\\
&\quad=\sum_{x\in\mathbb{Z}_{2^\nu}^*} e\Big(c(ux+vax^{-1}+wax(2x+a)^{-1})/m\Big)\\
&\qquad\cdot\sum_{y\in\mathbb{Z}_{2^{\omega-\nu}}} e\Big(c((u-vax^{-2}+wa^2(2x+a)^{-2}+2^{2\nu+1})y\\
&\qquad+2^\nu a(vx^{-3}-2wa(2x+a)^{-3})y^2)/2^{\omega-\nu}\Big),
\end{aligned}
$$

which is the desired result. □

For any parameters a, m in the inversive congruential method, a set $R(a,m)$ of integer triples (u,v,w) is defined by

$$
\begin{aligned}
R(a,m)=\{(u,v,w)\in\mathbb{Z}^3 \mid & u\equiv 1 \pmod 2, v\equiv -2uaz^3+2^{2\nu+1} \pmod{2^{\omega-\nu}},\\
& w\equiv -u(2z+1)^3 \pmod{2^{\omega-\nu}}, z\in\mathbb{Z}_{2^{\omega-\nu-1}}^*\},
\end{aligned}
$$

which plays a crucial role in the following analysis.

Lemma 3. *Let $(u,v,w)\in R(a,m)$ with a corresponding integer $z\in\mathbb{Z}_{2^{\omega-\nu-1}}^*$. Then*

$$
\begin{aligned}
&u-vax^{-2}+wa^2(2x+a)^{-2}+2^{2\nu+1}\\
&\qquad\equiv 2u(x-az)^2(2x^2+2a(2z+1)x+a^2z)(2x^2+ax)^{-2} \pmod{2^{\omega-\nu}}
\end{aligned}
$$

and

$$
\begin{aligned}
&2^\nu a(vx^{-3}-2wa(2x+a)^{-3})\equiv 2^{\nu+1}ua^2(x-az)\\
&\qquad\cdot((12z^2+6z+1)x^2+az(6z+1)x+a^2z^2)(2x^2+ax)^{-3} \pmod{2^{\omega-\nu}}
\end{aligned}
$$

for all $x\in\mathbb{Z}_{2^\nu}^*$.

Proof. It follows at once from $(u, v, w) \in R(a, m)$ that

$$\begin{aligned} u - vax^{-2} + wa^2(2x+a)^{-2} + 2^{2\nu+1} &\equiv u(1 + 2a^2x^{-2}z^3 - a^2(2x+a)^{-2}(2z+1)^3) \\ &\equiv u((2x+a)^2x^2 + 2a^2(2x+a)^2z^3 - a^2x^2(2z+1)^3)(2x^2+ax)^{-2} \\ &\equiv 2u(2x^4 + 2ax^3 - 3a^2z(2z+1)x^2 + 4a^3z^3x + a^4z^3)(2x^2+ax)^{-2} \\ &\equiv 2u(x-az)^2(2x^2 + 2a(2z+1)x + a^2z)(2x^2+ax)^{-2} \pmod{2^{\omega-\nu}} \end{aligned}$$

and

$$\begin{aligned} 2^\nu a(vx^{-3} - 2wa(2x+a)^{-3}) &\equiv 2^{\nu+1}ua^2(x^3(2z+1)^3 - (2x+a)^3z^3)(2x^2+ax)^{-3} \\ &\equiv 2^{\nu+1}ua^2(x-az)((12z^2+6z+1)x^2 + az(6z+1)x + a^2z^2) \\ &\quad \cdot (2x^2+ax)^{-3} \pmod{2^{\omega-\nu}} \end{aligned}$$

for $x \in \mathbb{Z}_{2^\nu}^*$. □

Lemma 4. *Let $(u, v, w) \in R(a, m)$. Then*

$$|S(u, v, w; a, c; m)| = 2^{\omega-\nu}$$

for any parameter $c \in \mathbb{Z}_m^*$.

Proof. Let $z \in \mathbb{Z}_{2^{\omega-\nu-1}}^*$ denote an integer corresponding to $(u, v, w) \in R(a, m)$. For any fixed integer $x \in \mathbb{Z}_{2^\nu}^*$, let $\xi \in \{1, 2, \ldots, \nu+1\}$ be defined by $\gcd(x - az, 2^{\nu+1}) = 2^\xi$. Since $u \equiv 1 \pmod 2$, $2x^2 + 2a(2z+1)x + a^2z \equiv 1 \pmod 2$, and $(12z^2+6z+1)x^2 + az(6z+1)x + a^2z^2 \equiv 1 \pmod 2$, it follows from Lemma 3 that

$$\gcd(u - vax^{-2} + wa^2(2x+a)^{-2} + 2^{2\nu+1}, 2^{\omega-\nu}) = 2^{\min(2\xi+1, \omega-\nu)}$$

and

$$\gcd(2^\nu a(vx^{-3} - 2wa(2x+a)^{-3}), 2^{\omega-\nu}) = 2^{\min(\nu+\xi+1, \omega-\nu)}.$$

If $\xi \leqslant \nu - 1$, then $2\xi + 1 < \nu + \xi + 1 \leqslant \omega - \nu$, and therefore [6, Lemma 6] implies that

$$\sum_{y \in \mathbb{Z}_{2^{\omega-\nu}}} e(c((u - vax^{-2} + wa^2(2x+a)^{-2} + 2^{2\nu+1})y + 2^\nu a(vx^{-3} - 2wa(2x+a)^{-3})y^2)/2^{\omega-\nu}) = 0.$$

If $\xi = \nu$, then $2\xi + 1 = \nu + \xi + 1 = 2\nu + 1 \geqslant \omega - \nu - 1$, which implies that

$$\sum_{y \in \mathbb{Z}_{2^{\omega-\nu}}} e(c((u - vax^{-2} + wa^2(2x+a)^{-2} + 2^{2\nu+1})y + 2^\nu a(vx^{-3} - 2wa(2x+a)^{-3})y^2)/2^{\omega-\nu}) = 2^{\omega-\nu};$$

if $\xi = \nu + 1$, then $2\xi + 1 > \nu + \xi + 1 \geqslant \omega - \nu$, which yields the same result. Since there exists exactly one $x \in \mathbb{Z}_{2^\nu}^*$ with $x \equiv az \pmod{2^\nu}$, i.e., $\xi \geqslant \nu$, it follows from Lemma 2 that

$$|S(u, v, w; a, c; m)| = 2^{\omega - \nu}$$

for any parameter $c \in \mathbb{Z}_m^*$. □

3. Main Results

Subsequently, triples $\mathbf{x}_n = (x_n, x_{n+1}, x_{n+2}) \in [0, 1)^3$ of inversive congruential pseudorandom numbers are considered, and the abbreviation

$$D_{m/2}^{(3)} = D_{m/2}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{(m/2)-1})$$

is used for their discrepancy. For any parameters a, m in the inversive congruential method, let

$$r(a, m) = \min \{|uvw| \in \mathbb{Z} | (u, v, w) \in R(a, m)\},$$

where the set $R(a, m)$ is defined as in the second section. In the following, it will still be assumed that $\omega = 3\nu + \mu$ with suitable integers $\nu \geqslant 1$ and $\mu \in \{0, 1, 2\}$.

Theorem. *The discrepancy $D_{m/2}^{(3)}$ of triples in the inversive congruential method satisfies*

$$D_{m/2}^{(3)} \geqslant \frac{2^{\mu/3}}{(\pi^2 + 3\pi + 3) r(a, m)} m^{-1/3}$$

for any parameters $a, c \in \mathbb{Z}_m^*$ with $a \equiv 1 \pmod 4$.

Proof. First, Lemma 1 is applied with $k = \ell = 3, N = m/2, \mathbf{t}_n = \mathbf{x}_n$ for $0 \leqslant n < m/2$, and $\mathbf{h} = (u, v, w) \in \mathbb{Z}^3$, where $(u, v, w) \in R(a, m)$ with $|uvw| = r(a, m)$. This yields

$$D_{m/2}^{(3)} \geqslant \frac{1}{(\pi^2 + 3\pi + 3) r(a, m) m} \left| \sum_{n=0}^{(m/2)-1} e(\mathbf{h} \cdot \mathbf{x}_n) \right|.$$

Now, it follows from $\mathbf{x}_n = (y_n, y_{n+1}, y_{n+2})/m, y_{n+1} \equiv ac^2 y_n^{-1} + 2c \pmod m$, and $y_{n+2} \equiv ac(acy_n^{-1} + 2)^{-1} + 2c \pmod m$ for $n \geqslant 0$ that

$$\begin{aligned} \left| \sum_{n=0}^{(m/2)-1} e(\mathbf{h} \cdot \mathbf{x}_n) \right| &= \left| \sum_{n=0}^{(m/2)-1} e((uy_n + vac^2 y_n^{-1} + wac(acy_n^{-1} + 2)^{-1})/m) \right| \\ &= \left| \sum_{y \in \mathbb{Z}_m^*} e((uy + vac^2 y^{-1} + wac(acy^{-1} + 2)^{-1})/m) \right|, \end{aligned}$$

where in the last step $\{y_0, y_1, \ldots, y_{(m/2)-1}\} = \mathbb{Z}_m^*$ has been used. Finally, the

transformation $y \equiv cz \pmod{m}$ and Lemma 4 imply that

$$\left| \sum_{n=0}^{(m/2)-1} e(\mathbf{h} \cdot \mathbf{x}_n) \right| = \left| \sum_{z \in \mathbb{Z}_m^*} e\Big((ucz + vacz^{-1} + wac(az^{-1} + 2)^{-1})/m \Big) \right|$$
$$= |S(u, v, w; a, c; m)| = 2^{\omega - \nu} = 2^{\mu/3} m^{2/3}.$$

This yields the desired result. □

Corollary. *Let ρ be an integer with $\rho \equiv 1 \pmod 4$ and $|\rho| < 2^{\omega-\nu-2}$. Then the discrepancy $D_{m/2}^{(3)}$ of triples in the inversive congruential method satisfies*

$$D_{m/2}^{(3)} \geqslant \frac{2^{\mu/3}}{2(\pi^2 + 3\pi + 3)|\rho|} m^{-1/3}$$

for any parameters $a, c \in \mathbb{Z}_m^*$ with $a \equiv \rho + 2^{2\nu} \pmod{2^{\omega-\nu-1}}$.

Proof. First, observe that the integer triple $(u, v, w) = (1, 2\rho, 1)$ belongs to the set $R(a, m)$ (with corresponding integer $z = 2^{\omega-\nu-1} - 1$). This implies that $r(a, m) \leqslant 2|\rho|$. Now, the desired result follows at once from the Theorem. □

The results above show that in the inversive congruential method with power of two modulus, there exist parameters a such that for all parameters c the discrepancy of triples is of an order of magnitude at least $m^{-1/3}$. This order of magnitude is too large compared with the asymptotic behaviour of $m/2$ true random points in $[0, 1)^k$ which have a discrepancy of about $m^{-1/2}$. Finally, it should be observed that all lower bounds remain valid for the discrepancy of k-tuples with $k \geqslant 4$, since their discrepancy is at least as large as the discrepancy of triples.

References

[1] EICHENAUER J, LEHN J, TOPUZOĞLU A (1988) A nonlinear congruential pseudorandom number generator with power of two modulus. Math Comp **51**: 757–759

[2] EICHENAUER-HERRMANN J (1992) Inversive congruential pseudorandom numbers: a tutorial. Int Statist Rev **60**: 167–176

[3] EICHENAUER-HERRMANN J (1995) Pseudorandom number generation by nonlinear methods. Int Statist Rev **63**: 247–255

[4] EICHENAUER-HERRMANN J (1996) Equidistribution properties of inversive congruential pseudorandom numbers with power of two modulus. Metrika **44**: 199–205

[5] EICHENAUER-HERRMANN J, EMMERICH F (1995) A review of compound methods for pseudorandom number generation. In: [9], pp 5–14

[6] EICHENAUER-HERRMANN J, NIEDERREITER H (1991) On the discrepancy of quadratic congruential pseudorandom numbers. J Comput Appl Math **34**: 243–249

[7] EICHENAUER-HERRMANN J, NIEDERREITER H (1992) Lower bounds for the discrepancy of inversive congruential pseudorandom numbers with power of two modulus. Math Comp **58**: 775–779

[8] EICHENAUER-HERRMANN J, NIEDERREITER H (1993) Kloosterman-type sums and the discrepancy of nonoverlapping pairs of inversive congruential pseudorandom numbers. Acta Arith **65**: 185–194

[9] HELLEKALEK P, LARCHER G, ZINTERHOF P (eds) (1995) Proceedings of the 1st Salzburg Mini-symposium on Pseudorandom Number Generation and Quasi-Monte Carlo Methods. Vienna: Austrian Center for Parallel Computation

[10] L'ECUYER P (1994) Uniform random number generation. Ann Operations Res **53**: 77–120

[11] NIEDERREITER H (1989) The serial test for congruential pseudorandom numbers generated by inversions. Math Comp **52**: 135–144

[12] NIEDERREITER H (1990) Lower bounds for the discrepancy of inversive congruential pseudorandom numbers. Math Comp **55**: 277–287
[13] NIEDERREITER H (1991) Recent trends in random number and random vector generation. Ann Operations Res **31**: 323–345
[14] NIEDERREITER H (1992) Random Number Generation and Quasi-Monte Carlo Methods. Philadelphia: SIAM
[15] NIEDERREITER H (1995) New developments in uniform pseudorandom number and vector generation. In: NIEDERREITER H, SHIUE PJ-S (eds) Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. Lecture Notes in Statistics **106**, 87–120. New York: Springer

J. EICHENAUER-HERRMANN
Fachbereich Mathematik
Technische Hochschule
Schloßgartenstraße 7
D-64289 Darmstadt
Germany

H. NIEDERREITER
Institut für Informationsverarbeitung
Österr. Akademie der Wissenschaften
Sonnenfelsgasse 19
A-1010 Wien
Austria

Mh. Math. 125, 219–227 (1998)

Monatshefte für
Mathematik

Printed in Austria

Inequalities Involving Integrals of Polar-Conjugate Concave Functions

By

M. Meyer, Marne-la-Vallée, and **S. Reisner,** Haifa

(Received 9 April 1996; in revised form 9 September 1996)

Abstract. An inequality of K. Mahler, together with its case of equality, due to M. Meyer, are extended to integrals of powers of polar-conjugate concave functions. An application to estimation of the volume-product of certain convex bodies is given.

Introduction

In [2] (1939), K. MAHLER established a lower bound to the product of the area of a plane convex figure by the area of its polar convex figure. The case of equality in Mahler's inequality was established much later by M. MEYER [3]. Using a symmetrization argument, it is shown in [3] that Mahler's inequality is equivalent to

Theorem A (MAHLER, MEYER). *Let f be a concave non-negative function defined on $[-a,b], a,b>0$, and for $s\in[-\frac{1}{a},\frac{1}{b}]$ define*

$$f^*(s)=\inf_{t\in[-a,b]}\frac{1-st}{f(t)}.$$

Then

$$\int_{-a}^{b}f(t)dt\int_{-1/a}^{1/b}f^*(s)ds\geqslant\frac{27}{16}. \qquad (*)$$

Equality holds if and only if f is affine and either $f(-a)=0$ and $\frac{a}{b}=2$ or $f(b)=0$ and $\frac{b}{a}=2$.

The main result of this paper (the 'Theorem') is an extension of Theorem A. In this extension the inequality $(*)$ is replaced by an analogous inequality, in which f and f^* appear in a power n, where n is any positive integer, instead on $n=1$ in $(*)$.

The motivation to deal with such an extension, which is of interest for its own sake, was its connection to the well-known conjecture concerning the so called 'inverse Santaló inequality' (see the remark preceding Corollary 1). A certain contribution to the knowledge concerning this conjecture is given in Corollary 1, using the main result.

1991 Mathematics Subject Classification: 52A20, 52A40
Key words: convex bodies, inverse Santaló, concave functions

Proposition 1. *Let f be a concave non-negative function on $[0,a]$ such that $f(0)=1$, and suppose that for some $v\geqslant 0$, $f(t)\leqslant 1+vt$ for every $t\in[0,a]$. Then*

$$F(p)=\left(1+(p+1)v\int_0^a f^p(t)dt\right)^{1/(p+1)}$$

is a decreasing function of $p\in[0,+\infty[$.

Proof. We may assume that $a=1$. Setting $g(t)=\mu(\{f>t\})$ (μ is the Lebesgue measure), we have $g(0)=1$ and g is concave non-negative and decreasing on $[0,M]$, where $M=\max\{f(t);\ t\in[0,1]\}\leqslant 1+v$, and satisfies:

$$g(t)\leqslant g(1)-\frac{t-1}{v} \tag{1}$$

(this last condition corresponds to $f(t)\leqslant 1+vt$). We have, for $p>0$,

$$F(p)=\left(1+(p+1)pv\int_0^M g(t)t^{p-1}dt\right)^{1/(p+1)}.$$

By (1) and by continuity, there exists $c\in[0,1]$ such that the function h which is defined by: $h(t)=1-(1-c)t$ for $t\in[0,1]$ and $h(t)=(c-(t-1)/v)_+$ for $t\in[1,+\infty[$, is concave on its support and satisfies;

$$\int_0^M g(t)t^{p-1}dt=\int_0^\infty h(t)t^{p-1}dt. \tag{2}$$

It follows that $g\geqslant h$ on some interval $[0,b]$ and $g\leqslant h$ on $[b,\infty[$. For $q\geqslant 0$, let

$$\phi_q(u)=\int_u^{+\infty}(g(t)-h(t))t^q dt.$$

Observe that $\phi_q(0)=\phi_q(+\infty)=0$, and that for some $m\geqslant 0$, $\phi_q'(u)=(h(u)-g(u))u^q$ is negative on $[0,m]$ and positive on $[m,+\infty[$, hence ϕ_q is negative on $[0,+\infty[$. Therefore, for $n\geqslant p$ we get:

$$\begin{aligned}\int_0^{+\infty}g(t)t^{n-1}dt&=(n-p)\int_0^{+\infty}g(t)t^{p-1}\left(\int_0^t u^{n-p-1}du\right)dt\\&=(n-p)\int_0^{+\infty}\left(\int_u^{+\infty}g(t)t^{p-1}dt\right)u^{n-p-1}du\\&\leqslant(n-p)\int_0^{+\infty}h(t)t^{p-1}\left(\int_0^t u^{n-p-1}du\right)dt\\&=\int_0^{+\infty}h(t)t^{n-1}dt.\end{aligned} \tag{3}$$

Now, for any $r\geqslant 1$, we have

$$\int_0^\infty h(t)t^{r-1}dt=\frac{(vc+1)^{r+1}-(vc+1)+v}{vr(r+1)}.$$

Thus, (2) and (3) give

$$F(p) = ((vc+1)^{p+1} - vc + v)^{1/(p+1)} \quad \text{and} \quad F(n) \leqslant ((vc+1)^{n+1} - vc + v)^{1/(n+1)}.$$

Setting $C = vc + 1 \geqslant 1$ and $d = v - vc \geqslant 0$, the inequality $F(n) \leqslant F(p)$ follows from the fact that the function

$$x \to (C^x + d)^{1/x} = \exp\left(\frac{\ln(C^x + d)}{x}\right)$$

is decreasing on $[1, +\infty[$. This is clear, because setting $C^x = y$, the function

$$y \to \frac{\ln(y+d)}{\ln(y)}$$

is decreasing on $[1, +\infty[$. □

In [1], H. ALZER proved an integral inequality for concave functions. The following Proposition 2, though formally not a special case of this inequality, can be proved using a similar proof. For the sake of completeness we bring here its proof, which is adapted from [1].

Proposition 2. *Let f be a concave non-negative function on $[a, b]$, then we have*

$$(n+1)\int_a^b f^n(t)dt \geqslant nf(x)\int_a^b f^{n-1}(t)dt + (x-a)f^n(a)$$

for every $x \in [a, b]$ and every $n \geqslant 1$.

Proof. By uniform approximation we may assume that f is differentiable. By concavity we have for all $x, t \in [a, b]$

$$f(x) \leqslant f(t) + (x-t)f'(t).$$

Multiply both sides of the last inequality by $f(t)^{n-1}$ and integrate. This gives

$$f(x)\int_a^b f^{n-1}(t)dt \leqslant \int_a^b f^n(t)dt + \frac{1}{n}\int (x-t)(f^n)'(t)dt,$$

integration by parts gives

$$\int_a^b (x-t)(f^n)'(t)dt = (x-b)f^n(t) - (x-a)f^n(a) + \int_a^b f^n(t)dt$$

$$\leqslant -(x-a)f^n(a) + \int_a^b f^n(t)dt.$$

The last two inequalities taken together, prove the required inequality. □

Let f be a concave non-negative function defined on $[-a, b]$, $a, b > 0$ and let $u \in [-a, b]$ be such that $\max_{t \in [-a,b]} f(t) = f(u) = M$. For $s \in [-\frac{1}{a}, \frac{1}{b}]$ define

$$f^*(s) = \inf_{t \in [-a,b]} \frac{1 - st}{f(t)}.$$

Then it is easy to see that

$$f^*(s) = \inf_{t\in[u,b]} \frac{1-st}{f(t)}, \quad \text{for every } s \in \left[0, \frac{1}{b}\right] \tag{4}$$

$$f^*(s) = \inf_{t\in[-a,u]} \frac{1-st}{f(t)}, \quad \text{for every } s \in \left[-\frac{1}{a}, 0\right].$$

Moreover, f^* is concave on $[-\frac{1}{a}, \frac{1}{b}]$ and satisfies:

$$f^*(0) = \frac{1}{M} \quad \text{and} \quad f^*(s) \leqslant \frac{1-us}{M}, \quad \text{for every } s \in \left[-\frac{1}{a}, \frac{1}{b}\right].$$

The duality between the concave functions f and f^* is in fact an expression of a geometric polar-duality: Let C be a convex figure in $\mathbb{R}^2$ which contains the origin in its interior and is symmetric about the x-axis. Then C may be represented as

$$C = \{(x,y);\ x \in [-a,b],\ |y| \leqslant f(x)\}$$

where f is as above. In this representation, the polar (about the origin) convex figure C^* turns out to be

$$C^* = \{(x,y);\ x \in [\tfrac{1}{a}, \tfrac{1}{b}],\ |y| \leqslant f^*(x)\}$$

and f^* is defined as above.

Thus, the following theorem is, in a sense, a "power-n" extension of Mahler's inequality for the area-product of polar-conjugate convex figures (cf. [2], [3]).

Theorem. *Let f be a concave non-negative function defined on $[-a,b]$, $a,b>0$ and for $s \in [-\frac{1}{a}, \frac{1}{b}]$ define*

$$f^*(s) = \inf_{t\in[-a,b]} \frac{1-st}{f(t)}.$$

Then, for every integer $n \geqslant 0$

$$\left(\int_{-a}^{b} (f(t))^n dt\right)\left(\int_{-1/a}^{1/b} (f^*(s))^n ds\right) \geqslant \frac{(n+2)^{n+2}}{(n+1)^{n+3}}.$$

Equality holds if and only if f is affine and either $f(-a)=0$ and $\frac{a}{b} = n+1$ or $f(b)=0$ and $\frac{b}{a} = n+1$.

Proof. Let u be as above, we may assume that $M = f(u) = \max_{t\in[-a,b]} f(t) = 1$. For $n \geqslant 0$, we define the numbers $a_n, \alpha_n, b_n, \beta_n$ by

$$a_n = (n+1)\int_u^b f(t)^n dt, \quad \alpha_n = (n+1)\int_{-a}^{u} f(t)^n dt,$$

$$b_n = (n+1)\int_0^{1/b} f^*(s)^n ds, \quad \beta_n = (n+1)\int_{-1/a}^{0} f^*(s)^n ds$$

(not that $\beta_0 = \frac{1}{a}$ and $b_0 = \frac{1}{b}$).

By Proposition 2 we have for $n \geqslant 1$:

$$a_n \geqslant f(t)a_{n-1} + (t-u), \quad \text{for every } t \in [u, b] \tag{5}$$

$$\alpha_n \geqslant f(t)\alpha_{n-1} + (u-t), \quad \text{for every } t \in [-a, u] \tag{6}$$

$$b_n \geqslant f^*(s)b_{n-1} + s, \quad \text{for every } s \in [0, \tfrac{1}{b}] \tag{7}$$

$$\beta_n \geqslant f^*(s)\beta_{n-1} - s, \quad \text{for every } s \in [-\tfrac{1}{a}, 0]. \tag{8}$$

It follows that

$$\frac{a_{n-1}f(t)}{a_n + u} + \frac{t}{a_n + u} \leqslant 1, \quad \text{for every } t \in [u, b]$$

$$\frac{a_{n-1}f(t)}{\alpha_n - u} + \frac{-t}{\alpha_n - u} \leqslant 1, \quad \text{for every } t \in [-a, u]$$

which gives, by the definition of f^*,

$$f^*\left(\frac{1}{a_n + u}\right) \geqslant \frac{a_{n-1}}{a_n + u} \quad \text{and} \quad f^*\left(\frac{-1}{\alpha_n - u}\right) \geqslant \frac{\alpha_{n-1}}{\alpha_n - u}.$$

Using the inequalities (7) and (8), we get for every $n \geqslant 1$

$$b_n(a_n + u) \geqslant a_{n-1}b_{n-1} + 1 \quad \text{and} \quad \beta_n(\alpha_n - u) \geqslant \alpha_{n-1}\beta_{n-1} + 1 \tag{9}$$

hence

$$a_n \geqslant \frac{a_{n-1}b_{n-1} + 1}{b_n} - u \quad \text{and} \quad \alpha_n \geqslant \frac{\alpha_{n-1}\beta_{n-1} + 1}{\beta_n} + u.$$

Adding these last inequalities together, we get

$$(a_n + \alpha_n)(b_n + \beta_n) \geqslant (b_n + \beta_n)\left(\frac{1}{b_n}(a_{n-1}b_{n-1} + 1) + \frac{1}{\beta_n}(\alpha_{n-1}\beta_{n-1} + 1)\right).$$

For $n \geqslant 2$, the inequalities (9), applied to $n-1$ instead of n, become;

$$a_{n-1}b_{n-1} \geqslant a_{n-2}b_{n-2} - ub_{n-1} \quad \text{and} \quad \alpha_{n-1}\beta_{n-1} \geqslant \alpha_{n-2}\beta_{n-2} + u\beta_{n-1}$$

and, by induction, we get

$$a_{n-1}b_{n-1} \geqslant a_0b_0 + n - 1 - u(b_1 + \cdots + b_{n-1})$$

and

$$\alpha_{n-1}\beta_{n-1} \geqslant \alpha_0\beta_0 + n - 1 + u(\beta_1 + \cdots + \beta_{n-1}).$$

It follows that

$$(a_n + \alpha_n)(b_n + \beta_n) \geqslant (b_n + \beta_n)\left[\frac{1}{b_n}\left(a_0b_0 + n - u\sum_{p=1}^{n-1} b_p\right) + \frac{1}{\beta_n}\left(\alpha_0\beta_0 + n + u\sum_{p=1}^{n-1}\beta_p\right)\right].$$

But since $a_0 = b - u,\ b_0 = \frac{1}{b},\ \alpha_0 = a + u,\ \beta_0 = \frac{1}{a}$, we have $a_0 b_0 = 1 - ub_0$ and $\alpha_0\beta_0 = 1 + u\beta_0$, so that

$$(a_n + \alpha_n)(b_n + \beta_n) \geqslant (b_n + \beta_n)\left[\frac{1}{b_n}\left(n + 1 - u\sum_{p=0}^{n-1} b_p\right) + \frac{1}{\beta_n}\left(n + 1 + u\sum_{p=0}^{n-1}\beta_p\right)\right],$$

or, setting $t_n = \frac{\beta_n}{b_n}$,

$$\begin{aligned}(a_n + \alpha_n)(b_n + \beta_n) &\geqslant \frac{t_n + 1}{t_n}\left[t_n\left(n + 1 - u\sum_{p=0}^{n-1} b_p\right) + n + 1 + u\sum_{p=0}^{n-1}\beta_p\right] \\ &= \frac{t_n + 1}{t_n}\left[(n+1)(t_n+1) + u\left(\sum_{p=0}^{n-1}\beta_p - t_n\sum_{p=0}^{n-1} b_p\right)\right].\end{aligned} \tag{10}$$

We may assume that $u \geqslant 0$ (the other case is treated analogously). Assuming this we claim that

$$t_n(n+1) + u\left(\sum_{p=0}^{n-1}\beta_p - t_n\sum_{p=0}^{n-1} b_p\right) \geqslant \frac{1}{b_n}\sum_{p=0}^{n}\beta_p. \tag{11}$$

In fact, for $0 \leqslant p \leqslant n$ we have $\beta_p \leqslant \beta_n$ and $b_p \leqslant b_n$. Also,

$$b_n \leqslant \frac{1}{u}\left[1 - \left(1 - \frac{u}{b}\right)^{n+1}\right], \tag{12}$$

hence, $ub_n \leqslant 1$ with equality only if $u = b$. We therefore have

$$(1 - ub_n)\sum_{p=0}^{n}\beta_p \leqslant (n+1)\beta_n(1 - ub_n) \leqslant \beta_n\left(n + 1 - u\sum_{p=1}^{n} b_p\right).$$

Hence

$$\begin{aligned}\sum_{p=0}^{n}\beta_p &\leqslant ub_n\sum_{p=0}^{n}\beta_p + \beta_n\left(n + 1 - u\sum_{p=0}^{n-1} b_p - ub_n\right) \\ &= ub_n\sum_{p=0}^{n-1}\beta_p - u\beta_n\sum_{p=0}^{n-1} b_p + \beta_n(n+1).\end{aligned} \tag{13}$$

Dividing the inequality (13) by b_n, we get (11) and

$$(a_n + \alpha_n)(b_n + \beta_n) \geqslant \frac{1 + t_n}{t_n}\left[n + 1 + \frac{1}{b_n}\sum_{p=0}^{n}\beta_p\right] = \frac{1 + t_n}{t_n}\sum_{p=0}^{n}\left(1 + \frac{\beta_p}{b_n}\right). \tag{14}$$

As $u \leqslant \frac{1}{b_n}$ we have (since we assume $M = 1$) for $t \leqslant 0$:

$$f^*(t) \leqslant 1 - ut \leqslant 1 - \frac{1}{b_n}\, t.$$

Hence, by Proposition 1,

$$\left(1 + \frac{\beta_p}{b_n}\right)^{1/p+1} \geqslant \left(1 + \frac{\beta_n}{b_n}\right)^{1/n+1} = (1 + t_n)^{1/n+1} \tag{15}$$

and the right-hand side of (14) is not less than

$$\frac{1 + t_n}{t_n} \sum_{p=0}^{n} (1 + t_n)^{p+1/n+1} = \frac{(1 + t_n)^{n+2/n+1}}{(1 + t_n)^{1/n+1} - 1}. \tag{16}$$

Substituting $s = (1 + t_n)^{1/(n+1)}$ this is

$$\frac{s^{n+2}}{s - 1} \geqslant \frac{(n + 2)^{n+2}}{(n + 1)^{n+1}} \quad (\text{for } s > 1), \tag{17}$$

thus the inequality is proved.

The case of equality: It is easy to check that if f satisfies the conditions at the end of the Theorem, equality holds. Suppose, on the other hand that we have equality. We may assume, as before, that $u \geqslant 0$. Since we have used (12), it follows now that $u = b = b_p^{-1}$ for all $0 \leqslant p \leqslant n$. This means, in particular, that f is increasing on $[-a, b]$. So, by (4), f^* satisfies

$$f^* = 1 - bt \text{ on } [0, \tfrac{1}{b}] \quad \text{and} \quad f^* \leqslant 1 - bt \text{ on } [-\tfrac{1}{a}, 0]. \tag{18}$$

Equality in (15) means, in particular, that

$$\left(1 + \frac{\beta_p}{b_n}\right)^{1/p+1} = (1 + b\beta_p)^{1/p+1}$$

is constant for $0 \leqslant p \leqslant n$. Using this together with (18) we get

$$\begin{aligned} 1 + \frac{b}{a} = 1 + b\beta_0 = (1 + b\beta_1)^{\frac{1}{2}} \\ = \left(1 + 2b \int_{-\frac{1}{a}}^{0} f^*(t)\, dt\right)^{\frac{1}{2}} \\ \leqslant \left(1 + 2b \int_{-\frac{1}{a}}^{0} (1 - bt)\, dt\right)^{\frac{1}{2}} = 1 + \frac{b}{a}. \end{aligned}$$

Hence, $f^* = 1 - bt$ in the whole interval $[-\frac{1}{a}, \frac{1}{b}]$ and f is affine as well, with $f(-a) = 0$. Equality in (17) is obtained only for

$$t_n = \left(\frac{n + 2}{n + 1}\right)^{n+1} - 1. \tag{19}$$

Equating the value of t_n in (19) with the value of $t_n = b\beta_n$ computed with $f^*(t) = 1 - bt$ we get $\frac{b}{a} = \frac{1}{n+1}$. □

It is clear that no finite upper bound which is independent of f exists for the product of integrals of the Theorem. However, by an appropriate translation we have the following observation. Here $f_x(t) = f(t-x)$ and for the sake of brevity we denote $(f_x)^*$ by f_x^*. We denote the volume of the Euclidean unit ball in $\mathbb{R}^n$ by χ_n.

Proposition 3. *Let f satisfy the conditions of the Theorem. Then there exists a point $x_0 \in]-b, a[$ such that*

$$\left(\int_{-a+x_0}^{b+x_0} (f_{x_0}(t))^n \, dt\right)\left(\int_{\frac{1}{-a+x_0}}^{\frac{1}{b+x_0}} (f_{x_0}^*(s))^n \, ds\right) \leqslant \frac{\chi_{n+1}^2}{\chi_n^2}, \tag{20}$$

with equality if and only if the graph of f is the upper half of an ellipse (with an axis on the line $\{y = 0\}$).

The proof of Proposition 3 is outlined right after the proof of the coming Corollary 1. We only remark here that the point x_0 is defined by the property that

$$\int_{\frac{1}{-a+x_0}}^{\frac{1}{b+x_0}} (f_{x_0}^*(s))^n s \, ds = 0$$

and also that the first integral on the left hand side of (20) does not depend on x_0.

We also make the remark that, combining the Theorem and Proposition 3, we get the estimate:

$$h(n) \leqslant (n+1)\left(\int_{-a+x_0}^{b+x_0} (f_{x_0}(t))^n \, dt\right)\left(\int_{\frac{1}{-a+x_0}}^{\frac{1}{b+x_0}} (f_{x_0}^*(s))^n \, ds\right) \leqslant g(n),$$

where $h(n)$ tends to e and $g(n)$ tends to 2π, as n tends to infinity.

Let K be a convex body (i.e. a compact convex set with non-empty interior) in $\mathbb{R}^n$, containing the origin. Let K^* denote its polar body with respect to the origin. If K is any convex body in $\mathbb{R}^n$ and $z \in \operatorname{int} K$, we denote $K^z = (K - z)^*$.

It is well known (cf. [6] and also [7]) that there exists a unique point z_0 in int K, called the *Santaló point of K*, such that

$$\operatorname{vol}_n(K^{z_0}) = \min_{z \in \operatorname{int} K} \operatorname{vol}_n(K^z).$$

With this z_0 we denote

$$P(K) = \operatorname{vol}_n(K)\operatorname{vol}_n(K^{z_0}).$$

P(K) is an affine-invariant parameter associated with K. A famous open problem is whether $P(K)$ is minimal for (and only for) $K = \Sigma_n$, the n-dimensional simplex. For a restricted family of convex bodies ('bodies of revolution' in a certain sense) we have the following result:

Corollary 1. *Let K be a convex body in $\mathbb{R}^n$ and $u \in S^{n-1}$. For $t \in \mathbb{R}$ let $K(t) = \{x \in K;\ \langle x, u\rangle = t\}$, and $I = \{t \in \mathbb{R}; \operatorname{vol}_{n-1}(K(t)) \neq 0\}$. Assume that there is a fixed convex body $L \subset \{x \in \mathbb{R}^n; \langle x, u\rangle = 0\}$ and $v \in \mathbb{R}^n$ with $\langle u, v\rangle = 1$, such*

that 0 is the Santaló point of L and, for all $t \in I$, $K(t) - tv = f(t)L$, *for some* $f(t) > 0$. *Then*

$$P(K) \geqslant \frac{P(\Sigma_n)}{P(\Sigma_{n-1})} P(L),$$

where Σ_n *is the n-dimensional simplex.*

Proof. It is clear that the Santaló point of K lies on the line $\{tv; t \in \mathbb{R}\}$ and we may assume that it is the origin, so that $0 \in I$. It is routine to check now that K^* is constructed in the same way as K, replacing L by L^* (in $\mathbb{R}^{n-1}$) and f by f^*. We get:

$$P(K) = \left(\int_I f(t)^{n-1}\, dt \right) \left(\int_{I^*} f^*(t)^{n-1}\, dt \right) P(L)$$

and use the Theorem. □

Proof of Proposition 3. Given f we construct in $\mathbb{R}^{n+1}$ a convex body K similar to the one in Corollary 1, with $K(t)$ congruent to $f(t)B_2^n$ (B_2^n being the Euclidean unit ball $\mathbb{R}^n$). We use the Blaschke-Santaló inequality, together with the case of equality (cf. [6] [5] and [4]). □

For a convex body K in $\mathbb{R}^n$, and $u \in S^{n-1}$, a *Schwarz rounding* of K about the direction u is any translate of the convex (by Brunn-Minkowski) body K for which for all $t \in \mathbb{R}$, $\tilde{K}(t)$, if it is not empty or a single point, is an $(n-1)$-dimensional Euclidean ball and $\mathrm{vol}_{n-1}(\tilde{K}(t)) = \mathrm{vol}_{n-1}(K(t))$. A special case of Corollary 1 is:

Corollary 2. *Let K be a convex body in* $\mathbb{R}^n$. *For any Schwarz rounding* $\tilde{K}$ *of K we have*

$$P(\tilde{K}) \geqslant \frac{(n+1)^{n+1}}{n^{n+2}} \chi_{n-1}^2.$$

References

[1] Alzer H (1992) On an integral inequality for concave functions. Acta Sci Math **56**: 79–82

[2] Mahler K (1939) Ein Minimalproblem für konvexe Polygone. Mathematica (Zutphen) **B7**: 118–127

[3] Meyer M (1991) Convex bodies with minimal volume product in $\mathbb{R}^2$. Mh Math **112**: 297–301

[4] Meyer M, Pajor A (1990) On the Blaschke-Santaló inequality, Arch Math **55**: 82–93

[5] Petty CM (1985) Affine isoperimetric problems. In: Goodman JE, Lutwak E, Malkevitch J, Pollack R (eds) Discrete Geometry and Convexity. Ann New York Acad Sci **440**: 113–127

[6] Santaló LA (1949) Un invariante afin para los cuerpos convexos del espacio de n dimensiones. Portugaliae Math **8**: 155–161

[7] Schneider R (1993) Convex Bodies: The Brunn-Minkowski Theory. Cambridge: University Press

M. Meyer
Equipe d'Analyse et de
Mathématiques Appliquées
Université de Marne-la-Vallée
2, rue de la Butte Verte
F-93166 Noisy-le-Grand Cedex
France
e-mail: meyer@math.univ-mlv.fr

S. Reisner
Department of Mathematics and
School of Education – Oramin
University of Haifa
Haifa, 31905
Israel
and
Department of Mathematics and Computer Science
University of Denver
Denver, CO 80208
USA
e-mail: reisner@mathcs2.haifa.ac.il

Mh. Math. 125, 229–239 (1998)

Monatshefte für
Mathematik

Printed in Austria

Radicals Coinciding with the Von Neumann Regular Radical on Artinian Rings

By

R. Mlitz*, Wien, **A. D. Sands****, Dundee, and **R. Wiegandt***, Budapest

(Received 11 April 1996; in final form 20 May 1997)

Abstract. We study radicals which coincide on artinian rings with Jacobson semisimple rings or equivalently with von Neumann regular rings. Exact lower and upper bounds for strong coincidence are given. For weak coincidence the exact lower bound is that for strong coincidence. We determine the smallest homomorphically closed class which contains all radicals coinciding in the weak sense with the von Neumann regular radical on artinian rings, but we do not know even the existence of the upper bound for weak coincidence. If a radical γ coincides with the von Neumann regular radical on artinian rings in the strong sense, then $\gamma(A)$ is a direct summand in A for every artinian ring A.

1. Introduction

DIVINSKY [1] localized the position of radicals which are nilpotent on artinian rings, in other words, which coincide with the Jacobson radical and with the Baer (prime) radical β on artinian rings in the strong and in the weak sense. In this note we shall deal with the dual problem, and determine all radicals γ which coincide on artinian rings with the Jacobson semisimple rings. As is well-known, the Jacobson semisimple artinian rings are exactly the von Neumann regular artinian rings, and the maximal von Neumann regular ideal $\mathcal{Q}(A)$, called the von Neumann regular radical, of an artinian ring A is again a Jacobson semisimple artinian ring, and so a direct summand in A. Thus from a structure theoretic point of view it is of interest to determine the radicals γ for which $\gamma(A) = \mathcal{Q}(A)$ for artinian rings A, that is, those radicals which coincide with $\mathcal{Q}$ on $\mathcal{A}$ in the strong sense. The main features are:

i) exact lower and upper bounds for strong coincidence with $\mathcal{Q}$ on $\mathcal{A}$ are given,

ii) for weak coincidence of a radical γ with $\mathcal{Q}$ on $\mathcal{A}$, that is, for $\gamma \cap \mathcal{A} = \mathcal{Q} \cap \mathcal{A}$ the exact lower bound is that for strong coincidence; we give

1991 Mathematics Subject Classification: 16N80

Key words: subidempotent and hypoidempotent radical, lower radical, hereditarily idempotent ring, artinian ring.

* Research carried out within the Austro-Hungarian Bilateral Intergovernmental Cooperation Program A-31. Research partially supported by Hungarian National Foundation for Scientific Research Grant No. T4265

** The second author gratefully acknowledges the support of the Carnegie Trust for Universities of Scotland

the smallest homomorphically closed class which contains all such radicals γ but we do not know even the existence of the exact upper bound of such radicals.

iii) neither the largest subidempotent radical $\mathscr{H}$ nor the largest hypoidempotent radical $\mathscr{I}$ are exact upper bounds for the coincidence with $\mathscr{Q}$ on $\mathscr{A}$; $\mathscr{H}$ is too small, $\mathscr{I}$ is too large,

iv) the exact lower bound does not contain the radical class of all strongly regular rings,

v) if a radical γ coincides with $\mathscr{Q}$ on $\mathscr{A}$ in the strong sense, then $\gamma(A)$ is a direct summand in A for every artinian ring A, and $\gamma(A)$ is a finite direct sum of matrix rings over division rings.

In the sequel, for the sake of convenience, we shall determine the radicals γ which coincide with $\mathscr{H}$ on artinian rings, and we shall see that they are exactly those which coincide with the von Neumann regular radical $\mathscr{Q}$ on artinian rings (cf. Corollary 9 and Theorem 15).

2. Preliminaries

In this note only associative rings and Kurosh-Amitsur radicals will be considered.

We remind the reader that a radical class γ is said to be *hypernilpotent* (*hypoidempotent*), if γ contains all nilpotent rings, (consists of idempotent rings, respectively). A hereditary hypernilpotent radical (hereditary hypoidempotent radical) is called a *supernilpotent radical* (*a subidempotent radical*). As usual, $\mathscr{S}$ will denote the semisimple operator designating the class

$$\mathscr{S}\gamma = \{\text{all rings } A \mid \gamma(A) = 0\}$$

to each radical γ. The letter $\mathscr{L}$ will stand for the lower radical operator assigning to every homomorphically closed class χ of rings the smallest radical class $\mathscr{L}\chi$ which contains χ.

We say that *two radical classes γ and δ coincide on a class $\mathscr{C}$ of rings in the weak sense*, if $\gamma \cap \mathscr{C} = \delta \cap \mathscr{C}$, and that *$\gamma$ and δ coincide on $\mathscr{C}$ in the strong sense*, if $\gamma(A) = \delta(A)$ for every ring $A \in \mathscr{C}$. Coincidence in the strong sense implies coincidence in the weak sense, but not conversely (see Lemma 28 and Example 6 in [2]). However, if the class $\mathscr{C}$ is hereditary, then coincidence in the strong and weak sense are the same. Strong coincidence of radicals means weak coincidence of the corresponding semisimple classes, as seen from the following

Proposition 1. *Let $\mathscr{C}$ be a homomorphically closed class of rings. For two radicals γ and δ the following conditions are equivalent*:

(i) *γ and δ coincide on $\mathscr{C}$ in the strong sense*,

(ii) $\mathscr{S}\gamma \cap \mathscr{C} = \mathscr{S}\delta \cap \mathscr{C}$.

Proof. (i) $\Rightarrow$ (ii) Suppose that $A \in \mathscr{S}\gamma \cap \mathscr{C}$. Then by (i) we have $0 = \gamma(A) = \delta(A)$. Hence $A \in \mathscr{S}\delta \cap \mathscr{C}$ holds, proving the containment $\mathscr{S}\gamma \cap \mathscr{C} \subseteq \mathscr{S}\delta \cap \mathscr{C}$. The opposite inclusion can be proved similarly.

(ii) $\Rightarrow$ (i) Let $A \in \mathscr{C}$. Then we have

$$\delta(A)/(\delta(A) \cap \gamma(A)) \cong (\delta(A) + \gamma(A))/\gamma(A) \lhd A/\gamma(A) \in \mathscr{S}\gamma \cap \mathscr{C} = \mathscr{S}\delta \cap \mathscr{C},$$

which implies

$$\delta(A)/(\delta(A) \cap \gamma(A)) = 0,$$

that is, $\gamma(A) \subseteq \delta(A)$. Analogous reasoning yields the opposite inclusion.

Remark. The assumption that $\mathscr{C}$ is homomorphically closed, was used only at the implication (ii) $\Rightarrow$ (i).

In the sequel we denote the class of all artinian rings by $\mathscr{A}$. The Baer (prime) radical will be denoted by β, $\mathscr{L}$ and $\mathscr{H}$ will stand for the radical classes of all von Neumann regular rings and of all hereditarily idempotent rings, respectively. More on radical theory can be found in the books [2], [3] and [5].

3. Exact Lower Bound

We start with

Proposition 2. *A hereditarily idempotent artinian ring is isomorphic to a finite direct sum of matrix rings over division rings, and conversely. In symbols:* $\mathscr{H} \cap \mathscr{A} = \mathscr{S}\beta \cap \mathscr{A}$. *The class* $\mathscr{S}\beta \cap \mathscr{A}$ *is hereditary, and hence so is* $\mathscr{H} \cap \mathscr{A}$.

Proof. The Baer radical of an artinian ring is nilpotent. Hence a hereditarily idempotent artinian ring has to be Baer semisimple, and so the Wedderburn-Artin Structure Theorems infer the assertions.

Next, we determine the exact lower bound of radicals coinciding with $\mathscr{H}$ on artinian rings in the weak sense.

Proposition 3. *The lower radical* $\mathscr{L}\mu$ *of the class* μ *of all matrix rings over division rings coincides with* $\mathscr{H}$ *on* $\mathscr{A}$ *in the weak sense.*

Proof. $\mu \subseteq \mathscr{H}$ implies $\mathscr{L}\mu \subseteq \mathscr{H}$ and also $\mathscr{L}\mu \cap \mathscr{A} \subseteq \mathscr{H} \cap \mathscr{A}$. If $A \in \mathscr{H} \cap \mathscr{A}$, then by Proposition 2 it follows that $A \in \mathscr{L}\mu \cap \mathscr{A}$. Thus $\mathscr{L}\mu \cap \mathscr{A} = \mathscr{H} \cap \mathscr{A}$ has been established.

Proposition 4. *If* $\gamma \cap \mathscr{A} = \mathscr{H} \cap \mathscr{A}$ *for a radical* γ, *then* $\mathscr{L}\mu \subseteq \gamma$.

Proof. We have

$$\mu \subseteq \mathscr{H} \cap \mathscr{A} = \gamma \cap \mathscr{A} \subseteq \gamma,$$

whence $\mathscr{L}\mu \subseteq \gamma$.

Proposition 5. *The lower radical* $\mathscr{L}\mu$ *of the class* μ *of all matrix rings over division rings is the unique smallest radical which coincides with* $\mathscr{H}$ *on* $\mathscr{A}$ *in the weak sense.*

Proof. Trivial by Propositions 3 and 4.

Let us recall that $\mathscr{H}$ is the largest homomorphically closed subclass in the class of all semiprime rings, and hence *a radical class* γ *consists of semiprime rings if and only if* $\gamma \subseteq \mathscr{H}$.

Proposition 6. *If γ is a radical satisfying $\gamma(A) \in \mathscr{H}$ for all $A \in \mathscr{A}$, then $\gamma(A)$ is a Jacobson semisimple artinian ring and hence a direct summand in A for each $A \in \mathscr{A}$.*

Proof. Denoting the Jacobson radical of $A \in \mathscr{A}$ by $J(A)$, we have

$$\gamma(A) \cap J(A) \subseteq \mathscr{H}(A) \cap J(A) = 0,$$

because the intersections are both idempotent and nilpotent. Hence we get that

$$\gamma(A) \cong \gamma(A)/(\gamma(A) \cap J(A)) \cong (\gamma(A) + J(A))/J(A) \lhd A/J(A),$$

and $A/J(A)$ is a Jacobson semisimple artinian ring. Consequently also $\gamma(A)$ is Jacobson semisimple and artinian, whence $\gamma(A)$ has an identity, and so $\gamma(A)$ is a direct summand of A.

Proposition 7. *The lower radical $\mathscr{L}\mu$ is the unique smallest radical which coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense.*

Proof. Since $\mathscr{L}\mu \subseteq \mathscr{H}$, the relation $\mathscr{L}\mu(A) \subseteq \mathscr{H}(A)$ is true for every ring A. Let A be an arbitrary artinian ring. Then applying Proposition 6 for the radical $\gamma = \mathscr{H}$, we get that $\mathscr{H}(A)$ is an artinian ring, and so by Proposition 2 it follows that $\mathscr{H}(A)$ is contained in the class $\mathscr{L}\mu$, whence $\mathscr{H}(A) \subseteq \mathscr{L}\mu(A)$. Thus $\mathscr{L}\mu$ coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense. Since by Proposition 5 $\mathscr{L}\mu$ is the smallest radical coinciding with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense, by the above statement $\mathscr{L}\mu$ must be also the smallest radical which coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense.

Theorem 8. *Let γ be any radical such that $\mathscr{L}\mu \subseteq \gamma \subseteq \mathscr{H}$. Then γ coincides with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense as well as in the strong sense.*

Proof. The statements are immediate consequences of Propositions 5 and 7.

Corollary 9. *The von Neumann regular radical $\mathscr{Q}$ coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense.*

Proof. Since $\mathscr{L}\mu \subsetneqq \mathscr{Q} \subsetneqq \mathscr{H}$, the assertion follows from Theorem 8.

Example 10. The class $\mathscr{K}$ of *all strongly regular rings*, (that is, $a \in a^2A$ for all $a \in A$), is a subclass of $\mathscr{Q}$ and a subidempotent radical. Since a strongly regular simple ring is a division ring, it follows $\mathscr{L}\mu \not\subseteq \mathscr{K}$.

Of course, a radical γ satisfying $\mathscr{L}\mu \subseteq \gamma \subseteq \mathscr{H}$ need not be subidempotent, as seen from the following

Example 11. Let V be a (countably) infinite dimensional vector space over a prime field F, and let T denote the ring of all finite valued linear transformations of V. As is well-known, T is a simple ring without unity, (cf. [2] Example 11). The ring A built on the cartesian product $T \times F$ with componentwise addition and with the multiplication

$$(t, a)(s, b) = (ts + tb + as, ab), \quad \forall t, s \in T;\ a, b \in F$$

has obviously the following properties:

i) A contains a unity element, namely (0,1);
ii) $T \cong \{(t,0), t \in T\} \lhd A$ and $A/T \cong F$;
iii) the ideals of A are 0, T and A.

Let us consider now the class $\mu \cup \{A\}$ which is clearly homomorphically closed. As one readily sees, the lower radical $\mathscr{L}(\mu \cup \{A\})$ contains A but does not contain T. Moreover, $\mathscr{L}\mu \subsetneqq \mathscr{L}(\mu \cup \{A\}) \subsetneqq \mathscr{H}$, *and* $\mathscr{L}(\mu \cup \{A\})$ *is not a hereditary, hence not a subidempotent radical.*

4. Upper Bounds

The radical class $\mathscr{I}$ of all idempotent rings is an upper bound of those radicals which coincide with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense, although $\mathscr{I}$ does not coincide with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense.

Proposition 12. *If a radical γ coincides with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense, then γ is a hypoidempotent radical.*

Proof. Suppose that γ is not hypoidempotent. Then there exists a ring $A \in \gamma$ which is not idempotent, and hence $B = A/A^2 \in \gamma$ and $B \neq 0$. The ring B is a subdirect sum of subdirectly irreducible zero-rings which latter are zero-rings $Z(p^k)$ for some $k = 1, 2, \ldots, \infty$, on cyclic (or quasi-cyclic) groups. Thus $Z(p^k) \in \gamma \cap \mathscr{A}$ for at least one k, though $Z(p^k)$ is not idempotent. This shows that $\gamma \cap \mathscr{A} \subsetneqq \mathscr{H} \cap \mathscr{A}$, a contradiction.

Note that in the proof we needed only the fact that the class γ is homomorphically closed.

Proposition 13. *The class $\mathscr{I}$ of all idempotent rings does not coincide with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense.*

Proof. The ring of integers modulo 4 is an example for an artinian ring which is in $\mathscr{I}$ but not in $\mathscr{H}$. This proves $\mathscr{H} \cap \mathscr{A} \subsetneqq \mathscr{I} \cap \mathscr{A}$.

Propositions 12 and 13 infer

Corollary 14. *If a radical γ coincides with $\mathscr{H}$ on $\mathscr{A}$ in the weak (or strong) sense, then $\gamma \subsetneqq \mathscr{I}$.*

In what follows, we shall position the radicals which coincide with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense, and give the exact upper bound for strong coincidence. For this purpose, let us consider the class

$$\mathscr{B} = \{A \in \mathscr{A} \mid 0 \neq B \lhd A \text{ implies } B \notin \mathscr{H}\} = \mathscr{A} \cap \mathscr{S}\mathscr{H}$$

which by Corollary 9 and Proposition 1 coincides with

$$\mathscr{B} = \{A \in \mathscr{A} \mid 0 \neq B \lhd A \text{ implies } B \notin \mathscr{Q}\} = \mathscr{A} \cap \mathscr{S}\mathscr{Q}.$$

Theorem 15. *The class*

$$\mathscr{U}\mathscr{B} = \{\text{all rings } A \mid A \text{ has no nonzero homomorphic image in } \mathscr{B}\}$$

is the largest homomorphically closed class of rings for which $\mathscr{U}\mathscr{B} \cap \mathscr{A} = \mathscr{H} \cap \mathscr{A}$.

Proof. By definition we have

$$\mathcal{U}\mathcal{B} \cap \mathcal{A} = \mathcal{U}(\mathcal{A} \cap \mathcal{S}\mathcal{H}) \cap \mathcal{A} = \mathcal{U}\mathcal{S}\mathcal{H} \cap \mathcal{A} = \mathcal{H} \cap \mathcal{A}.$$

Let γ be any homomorphically closed class such that $\gamma \cap \mathcal{A} = \mathcal{H} \cap \mathcal{A}$. If $A \in \gamma$ and $A/I \in \gamma \cap \mathcal{B}$, then

$$A/I \in \gamma \cap \mathcal{B} = \gamma \cap \mathcal{A} \cap \mathcal{S}\mathcal{H} = \mathcal{H} \cap \mathcal{A} \cap \mathcal{S}\mathcal{H} = \{0\},$$

that is, $A \in \mathcal{U}\mathcal{B}$. Hence $\gamma \subseteq \mathcal{U}\mathcal{B}$.

An immediate consequence is

Corollary 16. *A radical γ coincides with $\mathcal{H}$ on $\mathcal{A}$ in the weak sense if and only if $\mu \subseteq \gamma \subseteq \mathcal{U}\mathcal{B}$.*

Remark. The class $\mathcal{U}\mathcal{B}$ is certainly homomorphically closed, but we do not know as whether $\mathcal{U}\mathcal{B}$ is a radical class. Moreover, we do not know the existence of largest radical(s) in the interval $\mu \subseteq \gamma \subseteq \mathcal{U}\mathcal{B}$.

Problem 1. Does there exist an exact upper bound for radicals which coincide with $\mathcal{H}$ on $\mathcal{A}$ in the weak sense?

Proposition 17. *A radical γ coincides with $\mathcal{H}$ on $\mathcal{A}$ in the strong sense if and only if $\mathcal{B} \subseteq \mathcal{S}\gamma \subseteq \mathcal{S}\mathcal{L}\mu$.*

Proof. In the case of strong coincidence Proposition 1 yields

$$\mathcal{B} = \mathcal{S}\mathcal{H} \cap \mathcal{A} = \mathcal{S}\gamma \cap \mathcal{A} \subseteq \mathcal{S}\gamma.$$

Further, by Proposition 7 we have $\mathcal{L}\mu \subseteq \gamma$ which implies $\mathcal{S}\gamma \subseteq \mathcal{S}\mathcal{L}\mu$.

Conversely, $\mathcal{B} \subseteq \mathcal{S}\gamma \subseteq \mathcal{S}\mathcal{L}\mu$ implies

$$\mathcal{S}\mathcal{H} \cap \mathcal{A} \subseteq \mathcal{S}\gamma \cap \mathcal{A} \subseteq \mathcal{S}\mathcal{L}\mu \cap \mathcal{A}.$$

Hence again by Propositions 1 and 7 γ coincides with $\mathcal{H}$ on $\mathcal{A}$ in the strong sense.

Let $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ denote the hereditary closures of the classes $\mathcal{A}$ and $\mathcal{B}$ respectively. The following assertion provides some useful information.

Proposition 18. $\bar{\mathcal{B}} = \bar{\mathcal{A}} \cap \mathcal{S}\mathcal{H}$.

Proof. The containment $\bar{\mathcal{B}} \subseteq \bar{\mathcal{A}} \cap \mathcal{S}\mathcal{H}$ is trivially true. So, let A be any ring from $\bar{\mathcal{A}} \cap \mathcal{S}\mathcal{H}$. Then A is an accessible subring in a ring $C \in A$. Moreover, we have that $\mathcal{H}(C) \cap A \in \mathcal{H} \cap \mathcal{S}\mathcal{H} = \{0\}$. Hence $A \cong (A + \mathcal{H}(C))/\mathcal{H}(C)$ and the latter is an accessible subring in $C/\mathcal{H}(C) \in \mathcal{A} \cap \mathcal{S}\mathcal{H} = \mathcal{B}$. Thus $A \in \bar{\mathcal{B}}$ holds.

Theorem 19. *The upper radical*

$$\vartheta = \mathcal{U}\bar{\mathcal{B}} = \{\text{all rings } A \mid A \text{ has no nonzero homomorphic image in } \bar{\mathcal{B}}\}$$

is the unique largest radical which coincides with $\mathcal{H}$ on $\mathcal{A}$ in the strong sense.

Proof. By Proposition 17, if a radical γ coincides with $\mathcal{H}$ on $\mathcal{A}$ in the strong sense, then its semisimple class $\mathcal{S}\gamma$ must contain $\mathcal{B}$ and thus, in view of the heredity of semisimple classes of associative rings, also $\bar{\mathcal{B}}$. Therefore the radical

class γ has to be contained in $\vartheta = \mathscr{U}\bar{\mathscr{B}}$ which itself is a radical class since $\bar{\mathscr{B}}$ is hereditary.

We show that the radical ϑ coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense. In view of Proposition 1 we use semisimple classes. Since $\bar{\mathscr{B}} \subseteq \mathscr{S}\mathscr{H}$, also $\mathscr{S}\vartheta = \mathscr{S}\mathscr{U}\bar{\mathscr{B}} \subseteq \mathscr{S}\mathscr{H}$ holds, implying $\mathscr{S}\vartheta \cap \mathscr{A} \subseteq \mathscr{S}\mathscr{H} \cap \mathscr{A}$. Hence from

$$\mathscr{S}\mathscr{H} \cap \mathscr{A} = \mathscr{B} \subseteq \bar{\mathscr{B}} \subseteq \mathscr{S}\mathscr{U}\bar{\mathscr{B}} = \mathscr{S}\vartheta$$

we conclude that $\mathscr{S}\vartheta \cap \mathscr{A} = \mathscr{S}\mathscr{H} \cap \mathscr{A}$.

Next, we characterise the radical ϑ via the class

$$\mathscr{B}' = \{\text{all rings } A \mid \exists C \in \mathscr{A} \cap \mathscr{S}\mathscr{H} \text{ such that } A \lhd C\}.$$

Corollary 20. $\vartheta = \mathscr{U}\mathscr{B}'$.

Proof. $\mathscr{B} \subseteq \mathscr{B}' \subseteq \bar{\mathscr{B}}$ implies $\mathscr{U}\mathscr{B} \subseteq \mathscr{U}\mathscr{B}' \subseteq \mathscr{U}\bar{\mathscr{B}}$. As observed after Proposition 12, by Theorem 15 the class $\mathscr{U}\mathscr{B}$ consists of idempotent rings, and so does $\mathscr{U}\mathscr{B}'$. Assume that $\mathscr{U}\mathscr{B} \neq \mathscr{U}\mathscr{B}'$. Then there exists a ring A in $\mathscr{U}\mathscr{B}' \setminus \mathscr{U}\bar{\mathscr{B}}$. Hence A has a nonzero factor ring A/K in $\bar{\mathscr{B}}$ which is idempotent. $A/K \in \bar{\mathscr{B}}$ means that A/K is an accessible subring of some ring $C \in \mathscr{A} \cap \mathscr{S}\mathscr{H}$ (cf. Proposition 18). Since A/K is an idempotent ring. it is an ideal of C, and thus belongs to $\mathscr{B}'$, contradicting $A \in \mathscr{U}\mathscr{B}'$.

Corollary 21. *A radical γ coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense if and only if*

$$\mathscr{L}\mu \subseteq \gamma \subseteq \vartheta.$$

If $\mathscr{L}\mu \subseteq \gamma \subseteq \vartheta$, then $\gamma(A)$ is a direct summand in every artinian ring A and $\gamma(A)$ is a finite direct sum of matrix rings over division rings.

Proof. The first part is a direct consequence of Proposition 7 and Theorem 19. The second part follows from Proposition 6.

Problem 2. Is the equality $\vartheta = \mathscr{U}\mathscr{B}$ true? Or, does there exist a radical γ such that

i) $\gamma \cap \mathscr{A} = \vartheta \cap \mathscr{A} = \mathscr{H} \cap \mathscr{A}$,

ii) $\gamma(A) \neq \vartheta(A) = \mathscr{H}(A)$ for some $A \in \mathscr{A}$?

If such a radical γ exists, then necessarily $\gamma \not\subseteq \vartheta$ and by Proposition 12 γ is hypoidempotent. If such a radical γ does not exist, then coincidence with $\mathscr{H}$ on A in the weak and in the strong sense are the same and ϑ is the solution for Problem 1.

Remark. The torsion radical $\tau = \{A|$ the additive group of A is a torsion group$\}$ has the property that $\tau(A)$ is a direct summand in every artinian ring A but τ does not coincide with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense.

We are going to give an example showing that the radical ϑ is larger than $\mathscr{H}$. For this end we need some preparations. Let V be a countably infinite-dimensional vector space, and let T denote the ring of all finite-valued linear transformations of

V. Writing linear transformations from the left, T is not a left artinian ring (cf. [2] Example 11). On the cartesian product $V \times T$ we define addition componentwise and multiplication by the rule

$$(a,t)(b,s) = (tb, ts) \qquad \forall a, b, \in V; t, s \in T.$$

One easily verifies that in this way we have got a ring S.

Lemma 22. *The ring S has the following properties:*
(i) $(V,0) = \{(a,0) \mid a \in V\} \lhd S$, $(V,0)^2 = 0$, *and* $S/(V,0) \cong T$;
(ii) *The ideals of S are* 0, $(V,0)$, S, *and so S is a subdirectly irreducible ring with heart $H = (V,0)$;*
(iii) *The nonzero factor rings of S are S, T and they are not (left) artinian.*
(iv) *If S is an accessible subring in an artinian ring A, then $S \lhd A$ and $H \lhd A$. The same is true for T;*
(v) *If L is a left ideal of S such that $H \subseteq L$, then L is a left ideal of A and $AL = L$.*

Proof. (i) Obvious.
(ii) Let $0 \neq (a,t) \in I \lhd S$. First, suppose that $t \neq 0$. Then there exists a vector $b \in V$ with $tb \neq 0$. Since $t \in T$, necessarily $\ker t \neq 0$, and so we may choose a linear transformation $s \in T$ such that

$$0 \neq s(\ker t) \subseteq \ker t.$$

Now we have

$$0 \neq (tb, 0) = (tb, ts) = (a,t)(b,s) \in I,$$

and therefore we may cofine ourselves to the case $(a,0) \in I$.

Next, let us consider a basis $\{a = e_1, \ldots, e_\alpha, \ldots\}$ for V, and define linear transformations $q_\alpha \in T$ for each α by

$$q_\alpha e_1 = e_\alpha \quad \text{and} \quad q_\alpha e_\varepsilon = 0 \qquad \text{for every index } \varepsilon \neq 1.$$

Then

$$(e_\alpha, 0) = (q_\alpha a, 0) = (0, q_\alpha)(a,0) \in I$$

holds for each index α, whence $(V,0) \subseteq I$ follows. Since $S/(V,0) \cong T$ is a simple ring, (ii) has been established.
(iii) Trivial.
(iv) The rings S and T are obviously idempotent for $TV = V$. Hence by induction it follows that $S \lhd A$. Since $S/H \cong T$ is a prime ring, $H \lhd S \lhd A$ implies $H \lhd A$.
(v) For any element $(0,l) \in L$, lV is a finite dimensional subspace with a basis $\{e_1, \ldots, e_n\}$. Let us expand this basis to a basis $\{e_1, \ldots, e_n, e_{n+1}, \ldots\}$ of V, and take the linear transformation $t \in T$ defined by

$$te_i = e_i \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad te_j = 0 \quad \text{for} \quad j > n.$$

Then for each basis vector e_k we have $le_k = tle_k$, whence $l = tl$. Furthermore,

$$(0,T)(V,0) = (V,0) = H$$

holds implying $SH = H$. Since $H \subseteq L$, for every element $(a, b) \in L$ we have

$$(a, l) = (a, 0) + (0, l) = (a, 0) + (0, tl) \in H + SL \subseteq SL.$$

Thus it follows that $SL \subseteq L \subseteq SL$, and so

$$L \subseteq AL = A(SL) = (AS)L \subseteq SL \subseteq L$$

holds.

Proposition 23. *The class $\mathscr{H}$ is properly contained in the class $\vartheta = \mathscr{U}\bar{\mathscr{B}}$.*

Proof. The containment $\mathscr{H} \subseteq \mathscr{U}\bar{\mathscr{B}}$ is obvious by Corollary 21. We claim that the ring S of Lemma 22 is in $\mathscr{U}\bar{\mathscr{B}}$ but not in $\mathscr{H}$. Since $H \lhd S$ and $H^2 = 0$ by Lemma 22 (i), S is certainly not in $\mathscr{H}$. We have to show that S has no nonzero homomorphic image in $\bar{\mathscr{B}}$, which means by Lemma 22 (iii) that $S \notin \bar{\mathscr{B}}$ and $T \notin \bar{\mathscr{B}}$. The latter is trivially true for T is an idempotent simple ring, hence in $\mathscr{H}$.

Assume that $S \in \bar{\mathscr{A}}$. Then S is an accessible subring in an artinian ring A and by Lemma 22 (iv) we have that $S \lhd A$. Since S and T are not (left) artinian rings, there is a strictly descending infinite chain

$$L_1 \supset L_2 \supset \cdots$$

of left ideals of S such that $H \subseteq \bigcap_{n=1}^{\infty} L_n$. By Lemma 22 (v) all these left ideals L_n are left ideals also in A, a contradiction. Thus $S \notin \bar{\mathscr{A}}$ and consequently $S \notin \bar{\mathscr{B}}$.

Let us mention that to prove Proposition 18 also the Zassenhaus algebra (Example 3 in [2]) could have been used. Also we show that radicals coinciding with $\mathscr{H}$ on $\mathscr{A}$ in any sense cannot be obtained as upper radicals of supernilpotent radicals as seen from

Proposition 24. *Let ϱ be a supernilpotent radical. Its upper radical*

$$\mathscr{U}\varrho = \{A \mid A \text{ cannot be mapped homomorphically onto a nonzero ring in } \varrho\}$$

is a hypoidempotent radical such that $\mathscr{U}\varrho \cap \mathscr{A} \neq \mathscr{H} \cap \mathscr{A}$. In particular, if $\varrho \cap \mathscr{A} = \beta \cap \mathscr{A}$, then $\mathscr{U}\varrho \cap \mathscr{A} = \mathscr{I} \cap \mathscr{A}$.

Proof. Let $\mathscr{N}$ denote the class of all nilpotent rings. As is well-known

$$\mathscr{N} \subsetneqq \beta \subseteq \varrho \quad \text{and} \quad \mathscr{U}\varrho \subseteq \mathscr{U}\beta \subseteq \mathscr{U}\mathscr{N} = \mathscr{I},$$

showing that $\mathscr{U}\varrho$ is hypoidempotent.

Assume that $\varrho \cap \mathscr{A} \neq \beta \cap \mathscr{A}$. Then by $\beta \subseteq \varrho$ it follows that $\beta \cap \mathscr{A} \subsetneqq \varrho \cap \mathscr{A}$. Then there exists a matrix ring M over a division ring such that $M \in \varrho$, that is, $M \in \mathscr{A} \setminus \mathscr{U}\varrho$. Thus $\mathscr{U}\varrho$ does not coincide with $\mathscr{H}$ on $\mathscr{A}$ in the weak sense.

Suppose that $\varrho \cap \mathscr{A} = \beta \cap \mathscr{A}$. Then we have

$$\begin{aligned}\mathscr{I} \cap \mathscr{A} &= \mathscr{U}\mathscr{N} \cap \mathscr{A} = \mathscr{U}(\mathscr{N} \cap \mathscr{A}) \cap \mathscr{A} = \mathscr{U}(\beta \cap \mathscr{A}) \cap \mathscr{A} \\ &= \mathscr{U}(\varrho \cap \mathscr{A}) \cap \mathscr{A} = \mathscr{U}\varrho \cap \mathscr{A}.\end{aligned}$$

Further, by Proposition 13 it follows also $\mathscr{U}\varrho \cap \mathscr{A} \neq \mathscr{H} \cap \mathscr{A}$.

5. Characterisations through Lower Radicals

Our next aim is to characterize the radicals γ which coincide with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense as lower radicals.

Theorem 25. *For a radical γ the following assertions are equivalent:*

(i) *γ is hypoidempotent and coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense;*

(ii) *$\mathscr{L}\mu \subseteq \gamma$ and γ is the lower radical $\mathscr{L}\mathscr{C}$ of some homomorphically closed class $\mathscr{C}$ contained in the class*

$$\mathscr{M} = \mathscr{I}\backslash\{I \in \mathscr{S}\mathscr{H} | \exists A \in \mathscr{A} \text{ such that } 0 \neq I \lhd A\};$$

(iii) *$\mathscr{L}\mu \subseteq \gamma$ and γ is the lower radical $\mathscr{L}\mathscr{C}$ of some homomorphically closed class $\mathscr{C}$ contained in the class*

$$\mathscr{M}' = \mathscr{I}\backslash\{K \neq 0 | \exists A \in \mathscr{A} \cap \mathscr{S}\mathscr{H} \text{ such that } K \lhd A\}.$$

In addition $\gamma \subseteq \mathscr{M} \subseteq \mathscr{M}'$ holds.

Proof. (i) $\Rightarrow$ (ii) We choose $\mathscr{C} = \gamma$. In view of Proposition 7 we have $\mathscr{L}\mu \subseteq \gamma \subset \mathscr{I}$. Suppose that $\gamma \not\subseteq \mathscr{M}$, that is, γ contains a ring $I \in \mathscr{S}\mathscr{H}$ with $0 \neq I \lhd A \in \mathscr{A}$. $I \in \gamma$ implies $I \subseteq \gamma(A) = \mathscr{H}(A)$ by $A \in \mathscr{A}$ and (i), hence the hereditariness of $\mathscr{H}$ implies $I \in \mathscr{H}$, contradicting $0 \neq I \in \mathscr{S}\mathscr{H}$.

(ii) $\Rightarrow$ (iii) By the hereditariness of $\mathscr{S}\mathscr{H}$, $I \lhd A \in \mathscr{A} \cap \mathscr{S}\mathscr{H}$ infers $I \in \mathscr{S}\mathscr{H}$. Hence $\mathscr{M} \subseteq \mathscr{M}'$ and $\mathscr{C} \subseteq \mathscr{M}$ implies $\mathscr{C} \subseteq \mathscr{M}'$.

(iii) $\Rightarrow$ (i) Suppose that there is an artinian ring A such that $\gamma(A) \neq \mathscr{H}(A)$. Then we distinguish two cases.

If $\gamma(A) \subset \mathscr{H}(A)$, then the radical class $\gamma \cap \mathscr{H}$ is properly contained in $\mathscr{H}$. Hence by $\mathscr{L}\mu \subseteq \gamma$ Proposition 5 is applicable to the radical $\gamma \cap \mathscr{H}$ yielding that $\gamma \cap \mathscr{H} \cap \mathscr{A} = \mathscr{H} \cap \mathscr{A}$, that is, $\gamma \cap \mathscr{H}$ coincides with $\mathscr{H}$ on $\mathscr{A}$. This, however, contradicts

$$(\gamma \cap \mathscr{H})(A) = \gamma(A) \subsetneqq \mathscr{H}(A).$$

The other case is $\gamma(A) \not\subseteq \mathscr{H}(A)$. Now we have

$$0 \neq K = \gamma(A)/(\gamma(A) \cap \mathscr{H}(A)) \cong (\gamma(A) + \mathscr{H}(A))/\mathscr{H}(A) \lhd A/\mathscr{H}(A) \in \mathscr{A} \cap \mathscr{S}\mathscr{H}.$$

Since $\mathscr{C} \subseteq \mathscr{I}$, a result of HOFFMAN and LEAVITT [4] (cf. [5] Corollary 12.6) tells us that

$$\gamma = \{A | \text{ every nonzero homomorphic image of } A \text{ contains a nonzero ideal in } \mathscr{C}\}.$$

Thus, by $\gamma(A) \in \gamma$ the ring K contains a nonzero ideal $I \in \mathscr{C}$. Let $\bar{I}$ be the ideal of $A/\mathscr{H}(A)$ generated by I. Since $\mathscr{C}$ consists of idempotent rings, the Andrunakievich Lemma tells us that

$$I = I^3 \subseteq \bar{I}^3 \subseteq I.$$

Hence I is itself an ideal of $A/\mathscr{H}(A) \in \mathscr{A} \cap \mathscr{S}\mathscr{H}$, which contradicts the assumption $\mathscr{C} \subseteq \mathscr{M}'$.

Theorem 20 infers the following characterization of the largest radical $\vartheta = \mathscr{U}\bar{\mathscr{B}}$ which coincides with $\mathscr{H}$ on $\mathscr{A}$ in the strong sense.

Corollary 26. *$\vartheta = \mathcal{U}\bar{\mathcal{B}}$ is the largest homomorphically closed subclass in the class $\mathcal{M}$, or equivalently, in the class $\mathcal{M}'$.*

Proof. For the largest homomorphically closed subclass

$$\mathcal{C}' = \{A \in \mathcal{M} \mid A/I \in \mathcal{M} \quad \text{for all} \quad I \lhd A\}$$

of $\mathcal{M}$ and for any radical γ coinciding with $\mathcal{H}$ on $\mathcal{A}$ in the strong sense we have

$$\mathcal{L}\mu \subseteq \gamma \subseteq \mathcal{C}' \subseteq \mathcal{L}\mathcal{C}' \subseteq \mathcal{M}$$

by Theorem 25, and $\mathcal{L}\mathcal{C}'$ is contained in $\mathcal{C}'$ since $\mathcal{L}\mathcal{C}'$ is homomorphically closed. Thus $\mathcal{C}' = \mathcal{L}\mathcal{C}' = \vartheta$.

Acknowledgement. The authors thank the referee for valuable suggestions and improvements.

References

[1] Divinsky N (1961) General radicals that coincide with the classical radical on rings with D. C. C. Can J Math **13**: 639–644
[2] Divinsky N (1965) Rings and Radicals. London: Allen & Unwin
[3] Gardner BJ (1988) Radical Theory. New York: Longman
[4] Hoffman AE, Leavitt WG (1968) A note on the termination of the lower radical construction. J London Math Soc **43**: 617–618
[5] Wiegandt R (1974) Radical and semisimple classes of rings. Queen's Papers in Pure & Appl Math **37**. Kingston, Ontario: Univ Queens

R. Mlitz
Institut für Angewandte Mathematik
TU Wien
Wiedner Hauptstr. 8–10
A-1040 Wien
Österreich
e-mail: mlitz@umbriel.tuwien.ac.at

D. Sands
Department of Mathematics
University of Dundee
Dundee, DD1 4HN
United Kingdom
e-mail: adsands@mcs. dund.ac.uk

R. Wiegandt
Mathematical Institute HAS
P.O. Box 127
H-1364 Budapest
Hungary
e-mail: wiegandt@math-inst.hu

Mh. Math. 125, 241–253 (1998)

Monatshefte für
Mathematik

Printed in Austria

On the Metric Theory of the Nearest Integer Continued Fraction Expansion

By

R. Nair, Liverpool

(Received 13 December 1994; in final form 20 November 1996)

Abstract. Suppose k_n denotes either $\phi(n)$ or $\phi(p_n)(n = 1, 2, \ldots)$ where the polynomial ϕ maps the natural numbers to themselves and p_k denotes the k^{th} rational prime. Also let $\left(\frac{r_n}{q_n}\right)_{n=1}^{\infty}$ denote the sequence of convergents to a real number x and let $(c_n(x))_{n=1}^{\infty}$ be the corresponding sequence of partial quotients for the nearest integer continued fraction expansion. Define the sequence of approximation constants $(\theta_n(x))_{n=1}^{\infty}$ by

$$\theta_n(x) = q_n^2\left|x - \frac{r_n}{q_n}\right|. \qquad (n = 1, 2, \ldots)$$

In this paper we study the behaviour of the sequences $(\theta_{k_n}(x))_{n=1}^{\infty}$ and $(c_{k_n}(x))_{n=1}^{\infty}$ for almost all x with respect to the Lebesgue measure. In the special case where $k_n = n$ $(n = 1, 2, \ldots)$ these results are known and due to H. Jager, G. J. Rieger and others.

1. Introduction

In this paper we refine some results about a relative of the regular continued fraction expansion called the nearest continued fraction expansion proved by G. J. Rieger [11] and C. Kraaikamp [6]. We first introduce the notion of a semi-regular continued fraction expansion, of which both the regular continued fraction expansion and the nearest integer continued fraction expansion are examples. For a real number x we write

$$x = c_0 + \cfrac{\varepsilon_1}{c_1 + \cfrac{\varepsilon_2}{c_2 + \cfrac{\varepsilon_3}{c_3 + \cfrac{\varepsilon_4}{c_4 \cdots}}}},$$

where $(c_i)_{n=1}^{\infty}$ is a sequence of integers and $\varepsilon_i \in \{-1, 1\}$. The numbers c_i $(i = 1, 2, \ldots)$ are called the partial quotients of the expansion and for each natural number n the truncates

$$\frac{P_n}{Q_n} = [c_0; \varepsilon_1 c_1, \ldots, \varepsilon_n c_n]$$

1991 Mathematics Subject Classification: 11K50
Key words: nearest integer continued fraction, metric number theory, natural extensions

are called the convergents of the expansion. The expansion is called semi-regular if: (i) c_n is a natural number, for positive n; (ii) $\varepsilon_{n+1} + c_{n+1} \geqslant 1$ and (iii) $\varepsilon_{n+1} + c_{n+1} \geqslant 2$ for infinitely many n if the expansion is itself infinite. Central to the class of semi-regular continued fraction expansions is the regular continued fraction expansion which is also the most familiar and obtained when c_n is a natural number for natural numbers n and ε_i takes the value one for all i. Here and henceforth for a real number y let $[y]$ denote the greatest integer less than y and let $\langle y \rangle$ denote its fractional part, that is $y - [y]$. Notice that for the regular continued fraction expansion $c_0 = [x]$. It is thus convenient and no real restriction to assume x is in $[0, 1)$. If this is done we define the Gauss map

$$Tx = \left\langle \frac{1}{x} \right\rangle, \; x \neq 0; \quad T0 = 0.$$

on $[0, 1)$. We see that $c_i(x) = c_1(T^{i-1}x)$ $(i = 1, 2, \ldots)$. The nearest integer continued fraction expansion is defined by an analogous procedure. Here we set $c_0(x)$ to be the integer nearest to x and we thus also implicity define $\varepsilon_1(x)$. Given this information set Ω to be $\left(-\frac{1}{2}, \frac{1}{2}\right) \backslash \mathbf{Q}$ and define the map $S : \Omega \to \Omega$ by

$$Sx = \frac{\varepsilon_1}{x} - \left[|x|^{-1} + \frac{1}{2} \right].$$

In Section 2 we present some results from ergodic theory we need to carry out our investigation. In Section 3 we state and prove some results concerning the average behaviour of the convergents of nearest continued fraction expansion. These results extend certain earlier work of G. J. RIEGER [11]. The sequence $\left(\frac{P_n}{Q_n}\right)_{n=1}^{\infty}$ in the case of the nearest integer continued fraction expansion is a subsequence of the corresponding sequence for the regular continued fraction expansion [10]. Recall the inequality

$$\left| x - \frac{P_n}{Q_n} \right| \leqslant \frac{1}{Q_n^2},$$

which is classical, well known and due to L. J. DIRICHLET [3] for regular continued fractions and hence also for the nearest integer continued fractions, their convergents being a subsequence of those of the regular continued fraction expansion. Clearly if for each natural number n we set

$$\theta_n(x) = Q_n^2 \left| x - \frac{P_n}{Q_n} \right|$$

for the nearest integer continued fraction expansion, then for each x the sequence of numbers $(\theta_n(x))_{n=1}^{\infty}$ lies in the interval $[0, g]$ where $g = \frac{1}{2}(\sqrt{5} - 1)$ [10]. In Section 4, extending work of H. JAGER [5] and C. KRAAIKAMP [6] we state and prove results concerning the distribution of the sequence $(\theta_n(x))_{n=1}^{\infty}$ for almost all x with respect to the Lebesgue measure. In Section 5 further corollaries of the main results of Section 4 are described. In Section 6 the methods of Sections 4 and 5 are used to study other sequences attached to the nearest integer continued fraction expansion.

2. Basic Ergodic Theory

Here and throughout the rest of the paper by a dynamical system (X, β, μ, T) we mean a set X, together with a σ-algebra β of subsets of X, a probability measure μ on the measurable space (X, β) and a measurable self map T of X that is also measure preserving. By this we mean that if given an element A of β if we set $T^{-1}A = \{x \in X : Tx \in A\}$ then $\mu(A) = \mu(T^{-1}A)$. We say a dynamical system is ergodic if $T^{-1}A = A$ for some A in β means that $\mu(A)$ is either zero or one in value. We say the dynamical system (X, β, μ, T) is weakly mixing (among other equivalent formulations [13]) if for each pair of sets A and B in β we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\mu(T^{-n}A \cap B) - \mu(A)\mu(B)| = 0.$$

To be weakly mixing is a strictly stronger condition than ergodicity. A piece of terminology that is becoming increasingly standard is to call a sequence $\mathbf{k} = (k_n)_{n=1}^{\infty}$ of non-negative integers L^p good universal if given any dynamical system (X, β, μ, T) and any function f in $L^p(X, \beta, \mu)$ it is true that

$$l_f(x) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^{k_n}x),$$

exists almost everywhere with respect to the measure μ. The following theorem is proved in [9].

Theorem 2.1. *Suppose the sequence* $\mathbf{k} = (k_n)_{n=1}^{\infty}$ *of non-negative integers is such that for each irrational numbers* α *the sequence* $(\langle k_n\alpha\rangle)_{n=1}^{\infty}$ *is uniformly distributed modulo one and that for a particular* p *greater or equal to one that* $\mathbf{k} = (k_n)_{n=1}^{\infty}$ *is* L^p *good universal. Then if the dynamical system* (X, β, μ, T) *is weakly mixing,* $l_f(x) = \int_X f(t)d\mu(t)$ *almost everywhere with respect to* μ.

If k_n denotes either $\phi(n)$ or $\phi(p_n)$, where ϕ denotes any non-constant polynomial mapping the natural numbers to themselves and p_n denotes the n^{th} rational prime, then $\mathbf{k}$ is L^p good universal for any p greater than one. See [1] and [8], respectively, for proofs. The fact that for each irrational number α the sequence $(\langle k_n\alpha\rangle)_{n=1}^{\infty}$ is uniformly distributed modulo one in both instances are well known classical results. See [12] and [14], respectively. Other sequences are known by the author to satisfy the both hypothesis but these results have yet to appear in print. Henceforth, for reasons of brevity, we will call a sequence $\mathbf{k} = (k_n)_{n=1}^{\infty}$ p-good if it satisfies the hypothesis of Theorem 2.1 and good when it is p-good in the special case $p = \infty$. The weakly mixing property of the dynamical system (Ω, β, η, S) discussed in Section 3 and that of its natural extension discussed in Section 4 are fundamental to this investigation in that they allow one to use Theorem 2.1 and in consequence the powerful new subsequence pointwise ergodic theorems described above. The methods used in this paper also apply to other weakly mixing continued fraction expansions like the nearest integer continued fraction expansion such as those contained in the class of maximal S-expansions [7] for which the corresponding integrals arising in the use of Theorem 2.1 can be evaluated.

3. Average Behaviour of Convergents

Suppose η denotes the measure on $(-\frac{1}{2},\frac{1}{2})$ given for arbitrary Lebesgue measurable sets $E \subset (-\frac{1}{2},\frac{1}{2})$ by

$$\eta(E) = \frac{1}{\log G}\int_E \rho(t)dt,$$

where if $G = \frac{1}{2}(\sqrt{5}+1)$

$$\rho(t) = \begin{cases} (G+t)^{-1} & \text{if } t \geqslant 0 \\ (G+1+t)^{-1} & \text{if } t < 0 \end{cases}$$

It was shown in [11] that η is preserved by the map S and that the dynamical system (Ω, β, η, S), where β denotes the σ-algebra of Lebesgue measurable sets, is weakly mixing and in fact weakly Bernoulli. This fact via Theorem 2.1 has a number of arithmetic consequences. Our first theorem is the following

Theorem 3.1. *Suppose we are given two numbers a and ε with $a \geqslant 2$, $\varepsilon \in \{-1, 1\}$ and such that $a + \varepsilon \geqslant 2$. Suppose also that $\mathbf{k} = (k_n)_{n=1}^{\infty}$ is good then the relative frequency of elements of $\mathbf{k} = (k_n)_{n=1}^{\infty}$ such that $c_{k_n}(x) = c$ and $\varepsilon_{k_n}(x) = \varepsilon$ is*

$$\frac{1}{\log G}\log\frac{(4c+\varepsilon-3+2\sqrt{5})(4c+\varepsilon-5+2\sqrt{5})}{(4c+\varepsilon-7+2\sqrt{5})(4c+\varepsilon-1+2\sqrt{5})},$$

almost everywhere with respect to the Lebesgue measure.

Proof. Note that x is in

$$B^+_{c,\varepsilon} = \left[\frac{4}{4c+1}, \frac{4}{4c-1}\right]$$

if and only if

$$\varepsilon_1(x) = \min(1, c_1(x)) = \min(\varepsilon_2(x), c) = \varepsilon,$$

and that x is in

$$B^-_{c,\varepsilon} = \left[\frac{-4}{4c+1}, \frac{-4}{4c-1}\right]$$

if and only if

$$\varepsilon_1(x) = \min(-1, c_1(x)) = \min(\varepsilon_2(x), c) = \varepsilon.$$

Applying Theorem 2.1 to the characteristic function of the set $B^+_{c,\varepsilon} \cup B^-_{c,\varepsilon}$ we see that the required relative frequency is

$$\frac{1}{\log G}\int_{B^+_{c,\varepsilon}\cup B^-_{c,\varepsilon}} \rho(t)dt = \frac{2}{\log G}\log\frac{(4c+\varepsilon-3+2\sqrt{5})(4c+\varepsilon-5+2\sqrt{5})}{(4c+\varepsilon-7+2\sqrt{5})(4c+\varepsilon-1+2\sqrt{5})},$$

as required. □

We have the following general result concerning the average behaviour of the convergents. See [9] for an analogous result about the convergents of regular continued fraction expansion.

Theorem 3.2. *Suppose the function F with domain the non-negative real numbers and range the real numbers is continuous and increasing. Suppose also that F in $L^p(\eta)$ for some fixed p in $[1,\infty]$ or F is not in $L^1(\eta)$. For each natural number n and arbitrary non-negative real numbers $a_1,\ldots,a_n$ we define*

$$M_{F,n}(a_1,\ldots,a_n) = F^{-1}\left[\frac{1}{n}\sum_{j=1}^{n} F(a_j)\right].$$

Then if $\mathbf{k} = (k_n)_{n=1}^{\infty}$ is p-good

$$\lim_{n\to\infty} M_{F,n}(c_{k_1}(x),\ldots,c_{k_n}(x)) = F^{-1}\left[\frac{1}{\log G}\int_{-\frac{1}{2}}^{\frac{1}{2}} F(c_1(t))d\rho(t)\right],$$

almost everywhere with respect to Lebesgue measure.

Proof. If F in $L^p(\eta)$ the result follows immediately from Theorem 2.1. If however F is not in $L^1(\eta)$, set

$$f_M(x) = \begin{cases} F(c_1(x)) & \text{if } F(c_1(x)) \leqslant M \\ M & \text{if } F(c_1(x)) > M. \end{cases}$$

This means that for each $M \geqslant F(1)$ and almost all x with respect to the Lebesgue measure we have

$$\limsup_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} F(c_{k_n}(x)) \geqslant \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f_M(S^{k_n}(x))$$
$$= \frac{1}{\log G}\int_{-\frac{1}{2}}^{\frac{1}{2}} f_M(t)d\rho(t),$$

which tends to infinity with M, as required. □

Finally in this section we note the following result.

Theorem 3.3. *If $\mathbf{k} = (k_n)_{n=1}^{\infty}$ is good then*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \varepsilon_{k_n}(x) = \frac{1}{\log G}\log\frac{G^3}{4},$$

almost everywhere with respect to the Lebesgue measure.

Proof. Note that $\varepsilon_{n-1}(Sx) = \varepsilon_n(x)$ and application of Theorem 2.1 gives the result. □

4. Statistical Properties of the Sequence $(\theta_n(x))_{n=1}^{\infty}$

In this section we prove a result relating to the distribution of the sequence of pairs of numbers $(\theta_{n-1}(x),\theta_n(x))_{n=1}^{\infty}$ for almost all x with respect to the Lebesgue

measure. Let Γ_1 be the interior of the quadrilateral; with vertices $(0,0)$, $(g,0)$, $(2g^3,g^2)$ and $(0,\frac{1}{2})$ where $g=\frac{1}{2}(\sqrt{5}-1)$. Also let Γ_{-1} be the interior of the quadrilateral with vertices $(0,0)$, $(g^2,0)$, $(2g^3,g)$ and $(0,\frac{1}{2})$. Further put $\Gamma=\Gamma_1\cup\Gamma_{-1}$. In [6] it is shown that for all irrational x the sequence is distributed over Γ. Further where $G=\frac{1}{2}(\sqrt{5}+1)$ set

$$f_1(\alpha,\beta)=\frac{1}{\log G}\frac{1}{\sqrt{1-4\alpha\beta}} \quad \text{if } 1-4\alpha\beta>0$$

and

$$f_{-1}(\alpha,\beta)=\frac{1}{\log G}\frac{1}{\sqrt{1+4\alpha\beta}} \quad \text{if } 1+4\alpha\beta>0.$$

Further set

$$f(\alpha,\beta)=\begin{cases} f_1(\alpha,\beta) & \text{if } (\alpha,\beta)\in\Gamma_1\backslash\Gamma_{-1};\\ f_1(\alpha,\beta)+f_{-1}(\alpha,\beta) & \text{if } (\alpha,\beta)\in\Gamma_1\cap\Gamma_{-1};\\ f_{-1}(\alpha,\beta) & \text{if } (\alpha,\beta)\in\Gamma_{-1}\backslash\Gamma_1;\end{cases}$$

Then we have the following theorem which is the main result of this section.

Theorem 4.1. *Suppose A is a Borel subset of the set Γ then if $\mathbf{k}=(k_n)_{n=1}^{\infty}$ is good we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\chi_A(\theta_{k_n-1}(x),\theta_{k_n}(x))=\int_A f(\alpha,\beta)d\alpha\,d\beta,$$

almost everywhere with respect to the Lebesgue measure.

A variant of Theorem 4.1 in the special case where $k_n=n$ $(n=1,2,\ldots)$ appears in [6]. To prove Theorem 4.1 we need to use the ergodic properties of the map S associated to the nearest integer continued fraction expansion or more accurately its natural extension. In particular we need the following theorem to be found in [4]. The proof of Theorem 4.1 appearing here is essentially the same as that of Theorem 5 in [6]. For the details of the natural extension see [2].

Theorem 4.2. *Let*

$$D=\{x\in\Omega: x\geqslant 0\}\times[0,1-g)\cup\{x\in\Omega: x\geqslant 0\}\times(0,g).$$

Let β denote the σ-algebra of Borel sets in D and let μ be the measure defined on the measurable space (D,β) with Radon Nikodym derivative $(\log G)^{-1}(1+xy)^2$ relative to two dimensional Lebesque measure on D. Also define the map $N: D\to D$ by

$$N(x,y)=(Sx,(c_1+\varepsilon_1 y)^{-1}).$$

Then the dynamical system (D,β,μ,N) is weakly mixing.

If x is irrational then $\frac{Q_{n-1}}{Q_n}<g$ and if $c_n\geqslant 3$ then $\frac{Q_{n-1}}{Q_n}<g^2$. See [10] for a proof of this result due to A. Hurwitz. We readily see that $(S^n x,\frac{Q_{n-1}}{Q_n})$ is in D.

To prove Theorem 4.1 we need to prove the following theorem.

Theorem 4.3. *Suppose A is a Borel subset of the set D then if $\mathbf{k} = (k_n)_{n=1}^{\infty}$ is good we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi_A\left(S^{k_n}x, \frac{Q_{k_n-1}}{Q_{k_n}}\right) = \frac{1}{\log G}\int_A \frac{dxdy}{(1+xy)^2},$$

almost everywhere with respect to Lebesgue measure.

Proof. Note that $\varepsilon_n(x) = \varepsilon_1(S^{n-1}x)$ and so from the procedure for defining the nearest integer continued fraction expansion we observe the following recurrence relations:

$$P_{-1} = 1; \quad P_0 = c_0; \quad P_n = c_nP_{n-1} + \varepsilon_nP_{n-2} \ (n = 1, 2, \ldots) \tag{4.1}$$

and

$$Q_{-1} = 0; \quad Q_0 = 1; \quad Q_n = c_nQ_{n-1} + \varepsilon_nQ_{n-2} \ (n = 1, 2, \ldots). \tag{4.2}$$

The analoguous recurrence relation for the regular continued fraction expansion is well known and proved similarly. Using (4.1) and (4.2) and the definition of the map S we readily see that for each natural number n

$$N^n(x, 0) = \left(S^nx, \frac{Q_{n-1}}{Q_n}\right). \tag{4.3}$$

We also easily check that for any y such that $(x, y) \in D$

$$\lim_{n\to\infty} (N^n(x, 0) - N^n(x, y)) = 0.$$

Let B be the set of x for which the conclusion of Theorem 4.3 fails and let

$$C = \{x \in B : x \leqslant 0\} \times [0, 1 - g) \cup \{x \in B : x \geqslant 0\} \times [0, g).$$

Then for almost all (x, y) in C the sequence $N^{k_n}(x, y)$ is not distributed with respect to the measure having Radon-Nikodym derivative $(\log G)^{-1}(1 + xy)^2$ relative to two dimensional Lebesgue measure on D. This is in contradiction to Theorem 2.1 unless the measure of B is zero as required. □

We now complete the proof of Theorem 4.1. First we note by induction that

$$x - \frac{P_n}{Q_n} = \frac{(-1)^n\varepsilon_1(x)\cdots\varepsilon_n(x)S^nx}{Q_n(Q_n + Q_{n-1}S^nx)} \qquad (n = 1, 2, \ldots)$$

and from the fact $\varepsilon_{n+1}S^nx > 0$ we see that

$$\theta_n(x) = \frac{\varepsilon_{n+1}S^nx}{1 + \frac{Q_{n-1}}{Q_n}S^nx} \qquad (n = 1, 2, \ldots).$$

Also because

$$\frac{1}{\varepsilon_{n+1}S^nx} = c_{n+1} + S^{n+1}$$

and

$$\frac{Q_{n+1}}{Q_n} = c_{n+1} + \varepsilon_{n+1}\frac{Q_{n-1}}{Q_n},$$

we have

$$\theta_n(x) = \frac{Q_n}{Q_{n-1}} \frac{1}{1 + \frac{Q_n}{Q_{n+1}} S^{n-1}x}.$$

Let

$$(\alpha, \beta) = F(a, b) = \left(\frac{b}{1+ab}, \frac{a}{1+ab} \right)$$

for $ab \neq 1$. Then F has derivative

$$F'(a,b) = \begin{pmatrix} \frac{-b^2}{(1+ab)^2} & \frac{1}{(1+ab)^2} \\ \frac{1}{(1+ab)^2} & \frac{-a^2}{(1+ab)^2} \end{pmatrix}$$

and Jacobian $J = -(1-ab)(1+ab)^{-3}$. Then as a consequence of Theorem 4.1 for almost all x with respect to Lebesgue measure $(F(S^{k_n x}, \frac{Q_{k_n-1}}{Q_{k_n}}))_{n=1}^{\infty}$ which is just the sequence $(\theta_{k_n}(x), \varepsilon_{k_n+1}\theta_{k_n+1}(x))_{n=1}^{\infty}$ is distributed over $\Gamma_1 \cup \Gamma_1^*$, where Γ_1^* is the reflection of Γ_1 in the α axis, with density $\frac{1}{|J| \log G} \frac{1}{(1+xy)^2}$. Thus from the definition of F we have

$$\frac{1}{|J| \log G} \frac{1}{(1+ab)^2} + \frac{1}{|J| \log G} \frac{1}{(1-4\alpha\beta)^{\frac{1}{2}}}$$

and the other details follow analoguously and so Theorem 4.1 is proved. □

5. Corollaries of Theorem 4.1

Apply Theorem 4.1 with

$$A = \{(x, y) \in D : x < a \text{ and } y < b\}.$$

We have

Corollary 5.1. *Let*

$$G(z) = \begin{cases} \frac{z}{\log G} & \text{if } z \in [0, 1-g]; \\ \frac{1}{\log G}(z - \frac{z}{1-g} + \log \frac{z}{1-g} + 1) & \text{if } z \in [1-g, g]. \end{cases}$$

Then if $\mathbf{k} = (k_n)_{n=1}^{\infty}$ *is good we have*

$$\lim_{N \to \infty} \frac{1}{N} |\{1 \leqslant n \leqslant N : \theta_{k_n}(x) \leqslant z\}| = G(z)$$

for almost all x with respect to the Lebesgue measure.

To derive our next corollary we need the following Lemma [7].

Lemma 5.2. *Let* $\Delta(a)$ *be the interior of the triangle in the* (α, β) *plane with vertices* (0,0), $(0, a)$ *and* $(a, 0)$ *with* $0 < a \leqslant 5g - 2$. *Set*

$$H_1(a) = \frac{1}{\log G} \int\int_{\Delta(a) \cap \Gamma_1} f(x, y) dx dy = \int_0^a h_1(t) dt$$

and

$$H_{-1}(a) = \frac{1}{\log G}\int\int_{\Delta(a)\cap\Gamma_{-1}} f(x,y)dxdy = \int_0^a h_{-1}(t)dt.$$

Then for a in $[0, 5g-2]$

$$h_1(a) = \begin{cases} \frac{1}{2\log G}\log\frac{1+a}{1-a} & \text{if } a \in [0,\frac{1}{2}] \\ \frac{1}{2\log G}\log 3 & \text{if } a \in [\frac{1}{2}, g] \\ \frac{1}{2\log G}\left(\log\frac{1+(3g-1)}{1-(3g-1)} - \log\frac{1+a}{1-a}\right) & \text{if } a \in (g, 3g-1] \\ 0 & \text{if } a \in (3g-1, 5g-2] \end{cases}$$

and

$$h_{-1}(a) = \begin{cases} \frac{1}{\log G}\arctan a & \text{if } a \in [0, g^2] \\ \frac{1}{\log G}\arctan g^2 & \text{if } a \in (g^2, \frac{1}{2}] \\ \frac{1}{\log G}(\arctan(5g-2) - \arctan a) & \text{if } a \in (\frac{1}{2}, 5g-2] \end{cases}$$

We have the following result.

Corollary 5.3. *Let* $h = h_1 + h_{-1}$. *Then if* $\mathbf{k} = (k_n)_{n=1}^{\infty}$ *is good*

$$\lim_{N\to\infty}\frac{1}{N}|\{1 \leqslant n \leqslant N : \theta_{k_n-1}(x) + \theta_{k_n}(x) < a\}| = \int_0^a h(t)dt,$$

almost everywhere with respect to the Lebesgue measure.

Proof. If $\Delta(a)$ is as in the statement of Lemma 5.2 we have

$$\lim_{N\to\infty}\frac{1}{N}|\{1 \leqslant n \leqslant N : \theta_{k_n-1}(x) + \theta_{k_n}(x) < a : \varepsilon_n = 1\}|$$
$$= \frac{1}{\log G}\int\int_{\Delta(a)\cap\Gamma_1} f(x,y)dxdy,$$

and

$$\lim_{N\to\infty}\frac{1}{N}|\{1 \leqslant n \leqslant N : \theta_{k_n-1}(x) + \theta_{k_n}(x) < a : \varepsilon_n = -1\}|$$
$$= \frac{1}{\log G}\int\int_{\Delta(a)\cap\Gamma_{-1}} f(x,y)dxdy.$$

Thus Corollary 5.3 follows from Lemma 5.2. □

In [6] it is shown that for irrational x, $-g < \theta_{n-1}(x) - \theta_n(x) < g$ for all positive integers n. We have the following result proved in the same way to Corollary 5.3.

Corollary 5.4. *Let* $k = k_1 + k_{-1}$

$$k_1(a) = \begin{cases} \frac{1}{2\log G}(\log 3 + \log\frac{1+a}{1-a}) & \text{if } a \in [-\frac{1}{2}, 3g-2] \\ \frac{1}{2\log G}\frac{1}{2}\log 5 & \text{if } a \in (3g-2, 0] \\ \frac{1}{2\log G}(\frac{1}{2}\log 5 - \log\frac{1+a}{1-a}) & \text{if } a \in (g, 3g-1] \\ 0 & \text{if } a \in (g^2, g] \end{cases}$$

and

$$k_{-1}(a) = \begin{cases} \frac{1}{\log G}(\arctan \frac{1}{2} + \arctan a) & \text{if } a \in [-\frac{1}{2}, 0] \\ \frac{1}{\log G} \arctan \frac{1}{2} & \text{if } a \in (0, 5g-3] \\ \frac{1}{\log G}(\arctan g - \arctan a) & \text{if } a \in (5g-3, g]. \end{cases}$$

Then if $\mathbf{k} = (k_n)_{n=1}^{\infty}$ *is good*

$$\lim_{N\to\infty} \frac{1}{N} |\{1 \leqslant n \leqslant N : \theta_{k_n-1}(x) - \theta_{k_n}(x) < a\}| = \int_0^a k(t)dt,$$

almost everywhere with respect to the Lebesgue measure.

6. Other Sequences Attached to the Nearest Integer Continued Fraction Expansion

Theorem 2.1 has a number of other consequences for the nearest integer continued fraction expansion which we now describe. Let

$$L(z) = \begin{cases} \frac{1}{\log G}(\log(1+\frac{z}{2}) - \log(1-\frac{z}{2})) & \text{if } z \in [0, g^2] \\ \frac{1}{\log G}(\log(1+\frac{z}{2}) - \log(1-\frac{g^2}{2})) & \text{if } z \in (g^2, g] \end{cases}$$

Note that L is monotonically increasing, continuous and such that $L(0) = 0$ with $L(g) = 1$. We have the following theorem

Theorem 6.1. *If* $\mathbf{k} = (k_n)_{n=1}^{\infty}$ *is good then*

$$\lim_{N\to\infty} \frac{1}{N} \left| \left\{ 1 \leqslant n \leqslant N : \frac{Q_{k_n-1}}{Q_{k_n}} \leqslant z \right\} \right| = L(z),$$

almost everywhere with respect to the Lebesgue measure.

Proof. Let

$$A(z) = \{(x, y) : (x, y) \in D; y \leqslant z\}.$$

We note that

$$\frac{Q_{n-1}}{Q_n} = [0; c_n, \varepsilon_n c_{n-1}, \ldots, \varepsilon_2 c_1].$$

As a consequence of this observation and (4.3) we see that for $n > n_0(\varepsilon)$ and any $(x, y) \in D$ if

$$N^n(x, y) \in A(z - \varepsilon)$$

Then

$$N^n(x, y) \in A(z)$$

which is equivalent to $\frac{Q_{n-1}}{Q_n} \leqslant z$ and which in turn implies

$$N^n(x, y) \in A(z + \varepsilon).$$

Hence, for almost all x with respect to the Lebesgue measure. Using the argument used to prove Theorem 4.3 we see that

$$\lim_{N\to\infty}\frac{1}{N}\left|\left\{1\leqslant n\leqslant N:\frac{Q_{k_n-1}}{Q_{k_n}}\leqslant z\right\}\right|=\mu(A(z)).$$

Computation reveals that $\mu(A(z))=L(z)$, as required. □

Taking first moments and evaluating $\int_0^g z dL(z)$ we readily have the following theorem.

Theorem 6.2. *If* $\mathbf{k}=(k_n)_{n=1}^{\infty}$ *is good then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{Q_{k_n-1}}{Q_{k_n}}=\frac{1}{\log G}(\sqrt{5}-2+4\log 2+2\log(\sqrt{5}-2))=0.25225\cdots,$$

almost everywhere with respect to the Lebesgue measure.

For $z\in[0,\frac{g}{2}]$ let

$$M_1(z)=\frac{1}{\log G}\log(1+z)-\frac{z}{1+z}\log(2Gz))$$

and for $z\in[0,\frac{g^2}{2}]$ let

$$M_2(z)=\frac{1}{\log G}\left(-\log(1-z)-\frac{z}{1-z}\log(2Gz)\right).$$

Next define $M:[0,\frac{g}{2}]\to[0,1]$ that is continuous and monotonically increasing by

$$M(z)=\begin{cases}M_1(z)+M_2(z) & \text{if } z\in[0,\frac{g^2}{2}]\\ M_1(z)+M_2(\frac{g^2}{2}) & \text{if } z\in(\frac{g^2}{2},\frac{g}{2}].\end{cases}$$

Further let

$$R_n=\frac{|x-\frac{P_n}{Q_n}|}{|x-\frac{P_{n-1}}{Q_{n-1}}|}\qquad(n=1,2,\ldots).$$

In [10] it is shown that $0\leqslant R_n\leqslant\frac{g}{2}$ for irrational x. We have the following theorem

Theorem 6.3. *If* $\mathbf{k}=(k_n)_{n=1}^{\infty}$ *is good then*

$$\lim_{N\to\infty}\frac{1}{N}|\{1\leqslant n\leqslant N:R_{k_n}\leqslant z\}|=M(z),$$

almost everywhere with respect to the Lebesgue measure.

Proof. Let

$$B_+(z)=\{(x,y)\in D:x\geqslant 0;xy\leqslant z\}$$

and

$$B_-(z)=\{(x,y)\in D:x<0;|x|y\leqslant z\}.$$

Then using the fact that

$$R_n = \frac{Q_{n-1}}{Q_n}|N^n x|$$

and the argument used in the proof of Theorem 6.1 we have

$$\lim_{N\to\infty}\frac{1}{N}|\{1\leqslant n\leqslant N : R_{k_n}\leqslant z\}| = \mu(B_+(z)) + \mu(B_-(z)).$$

Computation verifies that if $z \in [0, \frac{g}{2}]$

$$\mu(B_+(z)) = M_1(z)$$

and that if $z \in [0, \frac{g^2}{2}]$

$$\mu(B_-(z)) = M_2(z),$$

completing the proof of Theorem 6.3. □

Let $\mathrm{Li}_2(x)$ denote the dilogaritm defined for instance in [11]. Again taking first moments we have the following result

Theorem 6.4. *If* $\mathbf{k} = (k_n)_{n=1}^{\infty}$ *is good then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} R_{k_n} = -3 - \log 2 + \frac{1}{\log G}\left(\frac{\pi^2}{12} + \log 4 - 2\,\mathrm{Li}_2\left(\frac{g^2}{2}\right)\right) = 0.06158\cdots,$$

almost everywhere with respect to the Lebesgue measure.

Acknowledgement. I thank the referee for his detailed and constructive comments which have improved the paper substantially.

References

[1] BOURGAIN J (1989) Pointwise ergodic theorems for arithmetic sets. Publ. I. H. E. S. **69**: 5–45
[2] CORNFELD IP, FORMIN SV, SINAI YG (1982) Ergodic Theory. Berlin Heidelberg New York: Springer
[3] HARDY, GH, WRIGHT E (1979) An Introduction to Number Theory. Oxford: Univ Press
[4] ITO, S, NAKADA H, TANAKA S (1981) On the invariant measure for the transformation associated with some real continued fractions. Keio Engineering Reports **30**: 61–69
[5] JAGER H (1986) The distribution of certain sequences connected with the continued fraction. Indag Math **48**: 61–69
[6] KRAAIKAMP C (1987) The distribution of some sequences connected with the nearest integer continued fraction expansion. Indag Math **49**: 177–191
[7] KRAAIKAMP C (1993) Maximal S-expansions are Bernoulli shifts. Bull Soc Math France **121**: 117–131
[8] NAIR R (1993) On Polynomials in primes and J. Bourgains's circle method approach to ergodic theorems II. Studia Math **105**: 207–233
[9] NAIR R (1994) On the metrical theory of continued fractions. Proc Amer Math Soc **120**: 1041–1046
[10] PERRON O (1935) Die Lehre von den Kettenbrüchen, Band 1, 3. Aufl. Stuttgart: Teubner
[11] RIEGER GJ (1979) Mischung und Ergodizität bei Kettenbrüchen nach nächsten Ganzen. J reine angew Math **310**: 171–181
[12] RHIN G (1975) Repartition modulo 1 de $f(p_n)$ quand f est une series entire. Lect Notes Math **475**: 176–244

[13] Walters P (1982) An Introduction to Ergodic Theory, pp 313–361
[14] Weyl H (1976) Über die Gleichverteilung von Zahlen mod. Eins Math Ann **77**: 313–361

R. Nair
Dept. of Pure Mathematics
University of Liverpool
P.O. Box 147
Liverpool L69 3BX
U.K.

Mh. Math. 125, 255–261 (1998)

Monatshefte für
Mathematik

Printed in Austria

Generalized Solution to a Semilinear Hyperbolic System with a Non-Lipshitz Nonlinearity

By

Marko Nedeljkov and **Stevan Pilipović**, Novi Sad

(Received 16 January 1996; in revised form 2 September 1996)

Abstract. *Let*

$$(\partial_t + \Lambda(x,t)\partial_x)y(x,t) = F(x,t,y(x,t)), y(x,0) = A(x) \tag{1}$$

be a semilinear hyperbolic system, where Λ is a real diagonal matrix and a mapping $y \to F(x,t,y)$ is in $\mathcal{O}_M(\mathbb{C}^n)$ with uniform bounds for $(x,t) \in K \subset\subset \mathbb{R}^2$. OBERGUGGENBERGER [6] has constructed a generalized solution to (1) when A is an arbitrary generalized function and F has a bounded gradient with respect to y for $(x,t) \in K \subset\subset \mathbb{R}^2$. The above system, in the case when the gradient of the nonlinear term F with respect to y is not bounded, is the subject of this paper. F is substituted by $F_{h(\varepsilon)}$ which has a bounded gradient with respect to y for every fixed (ϕ, ε) and converges pointwise to F as $\varepsilon \to 0$. A generalized solution to

$$(\partial_t + \Lambda(x,t)\partial_x)y(x,t) = F_{h(\varepsilon)}(x,t,y(x,t)), y(x,0) = A(x) \tag{2}$$

is obtained. It is compared to a continuous solution to (1) (if it exists) and the coherence between them is proved.

1. Introduction

System (1) has been solved in Colombeau's algebra $\mathcal{G}^n$ under the following assumptions ([6], [7]): Λ is a real $n \times n$ diagonal matrix of smooth functions, $A \in \mathcal{G}^n$, each component of $y \mapsto F(x,t,y)$ is in $\mathcal{O}_M$ with uniform bounds for $(x,t) \in K$, and each component of F has a globally bounded gradient with respect to y for $(x,t) \in K$, where K is an arbitrary compact subset of $\mathbb{R}^2$.

We consider system (1) in the case when the gradient with respect to y of F is not bounded. This nonlinear term is substituted by a suitable $F_{h(\varepsilon)}$ such that all the components of it have the gradient with respect to y bounded by $C \log \varepsilon^{-1}$. In Proposition 1 we construct a solution to system (2) by using the classical existence theorem for smooth functions (cf. [6]).

Our procedure is similar to the one described in [4] but without the use of regularized derivatives and with explicit form of the nonlinear term $F_{h(\varepsilon)}$. Although (1) may be uniquely solved by the truncation method and by using the regularized derivatives as in [4], the analysis of classical solutions to (1) and the ones obtained by this method has not been performed in that paper. This was the motivation for our paper.

1991 Mathematics Subject Classification: 35A05, 35L60, 46F10
Key words: semilinear hyperbolic system, generalized functions

We investigate the relations of classical solutions to (1) and the corresponding generalized solutions to (2). Such analysis for linear symmetric and homogeneous quasilinear systems with the regularized derivatives has been done in [3].

Let K_0 be a compact set in $\mathbb{R}$. K_T denotes the set bounded by $K_0 \times \{t = 0\}$, lines $t = \pm T$, and by the characteristics starting from the end points of K_0.

It is known from [6] that an L^1_{loc}-solution to a system of integral equations

$$g^j(x,t) = a^j(\gamma^j(x,t,0)) + \int_0^t F^j(\gamma^j(x,t,\tau),\tau,g(\gamma^j(x,t,\tau),\tau))d\tau,$$

where γ^j is the j-th characteristic curve for (1), $j = 1,\dots,n$, is a distributional solutions to (1). We shall suppose that all γ^j globally exist. So, the n-tuple of continuous functions $(g^1,\dots,g^n)$ is a continuous solution to (1) in K_T if it satisfies the system of integral equations in this domain.

In Proposition 2 we prove that a continuous local solution to (1) is associated with the generalized solution to the regularized system (2) on a slightly smaller interval. Also, if (1) is well posed globally (this is the case if, for example, F is globally bounded), then the generalized solution to (2) is associated with the classical one in the domain K_T for every $T > 0$.

2. Basic Notions

We slightly change the notation of [2] and [6]. We denote by $\mathcal{A}_0(\mathbb{R})$ a subset of $C_0^\infty(\mathbb{R})$ defined as follows: There exists a fixed constant $C_0 > 0$ such that for every $\phi \in \mathcal{A}_0(\mathbb{R})$

$$\sup\{|\phi(x)|, |\phi'(x)|, x \in \mathbb{R}\} \leqslant C_0 \tag{3}$$

and

$$\operatorname{diam}(\operatorname{supp}(\phi)) = 1, \quad \int_{\mathbb{R}} \phi(x)dx = 1.$$

Then $\mathcal{A}_q(\mathbb{R}), q \in \mathbb{N}$ is defined as the subset of $\mathcal{A}_0(\mathbb{R})$ whose elements satisfy the additional assumption

$$\int_{\mathbb{R}} x^\alpha \phi(x)dx = 0, \alpha \in \mathbb{N}_0, 1 \leqslant \alpha \leqslant q.$$

Put

$$\mathcal{A}_q(\mathbb{R}^n) = \{\phi \in \mathcal{D}(\mathbb{R})^n, \phi(x_1,\dots,x_n) = \bar{\phi}(x_1)\cdot\dots\cdot\bar{\phi}(x_n), \bar{\phi} \in \mathcal{A}_q(\mathbb{R})\}.$$

It is clear that $\mathcal{A}_0(\mathbb{R}^n) \supsetneq \mathcal{A}_1(\mathbb{R}^n) \supsetneq \dots$ As usual, ϕ_ε denotes the function $\varepsilon^{-n}\phi(\cdot/\varepsilon)$.

Further, let $\mathcal{E} = \mathcal{E}(\mathbb{R}^n)$ denote the space of functions $(\phi,\varepsilon,x) \to F_{\phi,\varepsilon}(x) \in \mathbb{C}, (\phi,\varepsilon,x) \in \mathcal{A}_0 \times (0,1) \times \mathbb{R}^n$ which are smooth on $\mathbb{R}^n$ for every fixed (ϕ,ε). $\mathbb{C}_M$ is the space of $A_{\phi,\varepsilon} : \mathcal{A}_0 \times (0,1) \to \mathbb{C}$ for which there exists $N \in \mathbb{N}_0$ such that for every $\phi \in \mathcal{A}_N$ there exist $C > 0$ and $\eta > 0$ such that $|A_{\phi,\varepsilon}| \leqslant C\varepsilon^{-N}, \varepsilon < \eta$. $\mathcal{E}_M = \mathcal{E}_M(\mathbb{R}^n)$ is the space of $G_{\phi,\varepsilon} \in \mathcal{E}$ such that for every compact set K and every $\beta \in \mathbb{N}_0^n$ there exists $N \in \mathbb{N}_0$ such that for every $\phi \in \mathcal{A}_N$ there exist $C > 0$ and $\eta > 0$ such that $|\partial^\beta G_{\phi,\varepsilon}(x)| \leqslant C\varepsilon^{-N}, x \in K, \varepsilon < \eta$. Let γ be the family

of real-valued increasing sequences which tend to infinity. $\mathcal{N} = \mathcal{N}(\mathbb{R}^n)$ is the space of $G \in \mathcal{E}_M$ such that for every $\beta \in \mathbb{N}_0^n$ and every compact set K there exist $N \in \mathbb{N}_0$ and $g \in \gamma$ such that for every $\phi \in \mathcal{A}_q$, $q \geqslant N$, there exist $C > 0$ and $\eta > 0$ such that $|\partial^\beta G_{\phi,\varepsilon}(x)| \leqslant C\varepsilon^{g(q)-N}, x \in K, \varepsilon < \eta$. $\mathbb{C}_0$ is the space of $A \in \mathbb{C}_M$ such that there exist $g \in \gamma$ and $N \in \mathbb{N}_0$ such that for every $\phi \in \mathcal{A}_q, q \geqslant N$, there exist $C > 0$ and $\eta > 0$ such that $|A_{\phi,\varepsilon}| \leqslant C\varepsilon^{g(q)-N}, \varepsilon < \eta$.

The spaces of Colombeau's generalized complex numbers and generalized functions are defined by $\bar{\mathbb{C}} = \mathbb{C}_M/\mathbb{C}_0$ and $\mathcal{G} = \mathcal{G}(\mathbb{R}^n) = \mathcal{E}_M/\mathcal{N}$, respectively. We denote by G or $[G_{\phi,\varepsilon}]$ the equivalence class of $G_{\phi,\varepsilon}$. It is said that a generalized constant $A \in \bar{\mathbb{C}}$ is associated to a constant $a \in \mathbb{C} (A \approx a)$ if it has a representative $A_{\phi,\varepsilon}$ such that there exists $N \in \mathbb{N}$ such that $\lim_{\varepsilon \to 0} A_{\phi,\varepsilon} = a$, for $\phi \in \mathcal{A}_N$. Generalized functions G and H are associated in $\mathcal{G}(\Omega) (G \approx H)$ if for every ψ in $C_0^\infty(\Omega), \langle G - H, \psi \rangle \approx 0$.

We introduce a stronger concept of association. Elements $G, H \in \mathcal{G}(\Omega)$ are L^∞-associated on a compact set $K \subset \Omega (G \overset{L^\infty(K)}{\approx} H)$ if $\|G_{\phi,\varepsilon} - H_{\phi,\varepsilon}\|_{L^\infty(K)} \approx 0$. They are L^∞-associated on Ω if they are L^∞-associated on every compact set $K \subset \Omega$. Obviously, the L^∞-association implies the ordinary one.

A function $f : \mathbb{C}^p \to \mathbb{C}$ is called slowly increasing at infinity if for some $C > 0$ and $r \in \mathbb{N}_0$

$$|f(z_1, \ldots, z_p)| < C(1 + |z_1|^2 + \cdots + |z_p|^2)^{r/2}, \quad z = (z_1, \ldots, z_p) \in \mathbb{C}^p.$$

$\mathcal{O}_M(\mathbb{C}^p)$ denotes the subspace of functions f of $C^\infty(\mathbb{C}^p)$ for which $f^{(\alpha)}$ is slowly increasing at infinity for every $\alpha \in \mathbb{N}_0^{2p}$, where we identify $f(z_1, \ldots, z_p)$ with $f(x_1, y_1 \cdots, x_p, y_p)$ and $f^{(\alpha)}$ denotes $\dfrac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial y_1^{\alpha_2} \cdots \partial x_p^{\alpha_{2p-1}} \partial y_p^{\alpha_{2p}}}$.

In the sequel we shall use the notation $F(x, t, u + iv) = F(x, t, u, v), x, t \in \mathbb{R}$, $y = u + iv \in \mathbb{C}^n$.

3. Regularized Systems

In the sequel Λ denotes a real $n \times n$ diagonal matrix of smooth functions. We fix a decreasing function $h : (0, 1) \to (0, \infty)$ such that

$$h(\varepsilon) = \mathcal{O}((\log \varepsilon^{-1})^{1/2}), \quad h(\varepsilon) \to \infty \quad \text{as } \varepsilon \to 0. \tag{4}$$

Denote by B_r the cube $|x| \leqslant r, |t| \leqslant r, |y| \leqslant r$, where $y = (u_1, v_1, \ldots, u_n, v_n)$. Let ε_i be a decreasing sequence of positive numbers such that $h(\varepsilon_{i+1}) = i, i \in \mathbb{N}$.

This implies that $h(\varepsilon) \geqslant i - 1$ if $\varepsilon < \varepsilon_i$. Let

$$\begin{aligned} Si = B_i \cap \{(x, t, u, v), |F(x, t, u, v)| \leqslant i - 1\} \\ \cap \{(x, t, u, v), |\nabla_{u,v} F(x, t, u, v)| \leqslant i - 1\}, \quad i \in \mathbb{N}. \end{aligned}$$

Let κ_i be the characteristic function of $S_i, i \in \mathbb{N}$. Put

$$\begin{aligned} \kappa_{h(\varepsilon)} &= (\kappa_i * \phi_{1/h(\varepsilon)}), \quad \varepsilon \in [\varepsilon_{i+1}, \varepsilon_i), \quad i \in \mathbb{N}, \\ F^k_{h(\varepsilon)} &= F^k \kappa_{h(\varepsilon)}, \quad \varepsilon \in (0, \varepsilon_1), \quad k \in \{1, \ldots, n\}. \end{aligned}$$

Then there exists a constant $C = C(C_0) > 0$ which does not depend on $\phi \in \mathscr{A}_0$ such that

$$\begin{aligned} &\|F_{h(\varepsilon)}\|_{L^\infty(\mathbb{R}^{2+2n})} \leqslant Ch(\varepsilon), \\ &\|\nabla_{u,v} F_{h(\varepsilon)}\|_{L^\infty(\mathbb{R}^{2+2n})} \leqslant Ch(\varepsilon)^2, \quad \varepsilon \in (0, \varepsilon_1). \end{aligned} \tag{5}$$

Definition 1. $G = (G_1, \dots, G_n) \in (\mathscr{G}(\mathbb{R}^2))^n$ is a solution to (2) if any of its representatives $G_{\phi,\varepsilon}$ satisfies the system

$$\begin{aligned} (\partial_t + \Lambda(x,t)\partial_x) G_{\phi,\varepsilon}(x,t) &= F_{h(\varepsilon)}(x,t,G_{\phi,\varepsilon}(x,t)) + d_{1,\phi,\varepsilon}(x,t), \\ G_{\phi,\varepsilon}(x,0) &= A_{\phi,\varepsilon}(x) + d_{2,\phi,\varepsilon}(x), \end{aligned} \tag{2'}$$

where $A_{\phi,\varepsilon} \in (\mathscr{E}_M(\mathbb{R}))^n$ is a representative of A, for some $d_{2,\phi,\varepsilon} \in (\mathscr{N}(\mathbb{R}))^n$, and $d_{1,\phi,\varepsilon} \in (\mathscr{N}(\mathbb{R}^2))^n$. We call (2) and (2′) the h-regularized system.

We need the following well-known result:

Lemma 1. (cf. [6]) *Let* $v = (v_1, \dots, v_n) \in (C^\infty(\mathbb{R}^2))^n$ *be a solution to*

$$\begin{aligned} (\partial_t + \Lambda(x,t)\partial_x) v(x,t) &= f(x,t)v(x,t) + g(x,t), \\ v(x,0) &= b(x), \end{aligned}$$

where f *is a smooth* $n \times n$ *matrix, and* g *and* b *are smooth vectors. Then*

$$\sup_{(x,t)\in K_T} |v(x,t)| \leqslant (\sup_{x\in K_0} |b(x)| + T \sup_{(x,t)\in K_T} |g(x,t)|) \exp(nT \sup_{(x,t)\in K_T} |f(x,t)|),$$

where

$$\sup_{(x,t)\in K_T} |f(x,t)| = \|f\|_{L^\infty(K_T)} = \sup_{\substack{(x,t)\in K_T \\ i=1,\dots,n}} |f_i(x,t)|.$$

Proposition 1. *Assume that every component of the mapping* $y \to F(x,t,y)$ *belongs to* $\mathscr{O}_M(\mathbb{C}^n)$ *and has uniform bounds for* $(x,t) \in K \subset\subset \mathbb{R}^2$. *Then the h-regularized system* (2) *has a unique solution in* $(\mathscr{G}(\mathbb{R}^2))^n$ *whenever the initial data is in* $(\mathscr{G}(\mathbb{R}))^n$.

Proof. We shall identify $\phi \in \mathscr{A}_0(\mathbb{R})$ and $\phi \otimes \phi \in \mathscr{A}_0(\mathbb{R}^2)$. Well-known theory (cf. [6]) implies that for given $\phi \in \mathscr{A}_0$ and $\varepsilon > 0$ there exists a smooth solution $G_{\phi,\varepsilon}$ of (2′) where $A_{\phi,\varepsilon}$ is a representative of A. Obviously, $G_{\phi,\varepsilon} \in (\mathscr{E}(\mathbb{R}^2))^n$. The existence will be proved when we show that $G_{\phi,\varepsilon}$ is in $(\mathscr{E}_M(\mathbb{R}^2))^n$. The procedure is the same as in [6]. Let us write the system (2) in the following form

$$\begin{aligned} (\partial_t + \Lambda\partial_x) G_{\phi,\varepsilon}(x,t) &= \left(\int_0^1 \nabla_y F(x,t,\sigma G_{\phi,\varepsilon}(x,t)) d\sigma \right) G_{\phi,\varepsilon}(x,t) + F(x,t,0), \\ G_{\phi,\varepsilon}(x,0) &= A_{\phi,\varepsilon}(x). \end{aligned}$$

We will prove the necessary estimates for $G_{\phi,\varepsilon}$ in the domain K_T, because each compact set of $\mathbb{R}^2$ lies in some K_T. Note $A_{\phi,\varepsilon} \in (\mathscr{E}_M(\mathbb{R}))^n$ implies that there exists $N \in \mathbb{N}$ such that for every $\phi \in \mathscr{A}_N$ there exist $C > 0$ and $\eta > 0$ such that

$$\sup_{x\in K_0} |(A_{1,\phi,\varepsilon}(x), \dots, A_{n,\phi,\varepsilon}(x))| \leqslant C\varepsilon^{-N}, \quad \varepsilon < \eta.$$

Since $K_T \subset B_i$ for large enough $i \in \mathbb{N}$, (5) and Lemma 1 imply that there exist $q > 0$ and $\eta > 0$ such that

$$|G_{\phi,\varepsilon}(x,t)| \leqslant (C\varepsilon^{-N} + T \sup_{(x,t)\in K_T} |F_{h(\varepsilon)}(x,t,0)|) \exp (nT \sup_{\substack{(x,t)\in K_T \\ y\in\mathbb{C}^n}} |\nabla_y F_{h(\varepsilon)}(x,t,y)|)$$

$$\leqslant (C\varepsilon^{-N} + C_2 T)\varepsilon^{-nTq} \leqslant C_1 \varepsilon^{-N_1}, \quad \varepsilon < \eta.$$

Let us estimate $\partial_x G_{\phi,\varepsilon}$. Note that $[\partial_t + \Lambda\partial_x, \partial_x] = -(\partial_x\Lambda)\partial_x$. Then the following equality holds

$$\begin{aligned}(\partial_t + \Lambda\partial_x)(\partial_x G_{\phi,\varepsilon}) &= \partial_x F_{h(\varepsilon)}(x,t,G_{\phi,\varepsilon}) + [\partial_t + \Lambda\partial_x, \partial_x] G_{\phi,\varepsilon} \\ &= (\nabla_G F_{h(\varepsilon)}(x,t,G_{\phi,\varepsilon}) - (\partial_x\Lambda(x,t))\partial_x G_{\phi,\varepsilon} + (\partial_x F_{h(\varepsilon)})(x,t,G_{\phi,\varepsilon}), \\ \partial_x G_{\phi,\varepsilon}(x,0) &= \partial_x A_{\phi,\varepsilon}(x).\end{aligned}$$

This means that $\partial_x G_{\phi,\varepsilon}$ satisfies the equation in Lemma 1. By using this argument we obtain estimates for all derivatives of $G_{\phi,\varepsilon}$ with respect to x. Estimates for t, x- and t-derivatives can be obtained by successive differentiation. This proves that $G_{\phi,\varepsilon} \in (\mathscr{E}_M(\mathbb{R}^2))^n$.

Let us prove the uniqueness. Let G and V be solutions to (2). This implies

$$\begin{aligned}(\partial_t + \Lambda\partial_x)(G_{\phi,\varepsilon} - V_{\phi,\varepsilon})(x,t) = &\left(\int_0^1 \nabla_y F_{h(\varepsilon)}(x,t,\sigma G_{\phi,\varepsilon}(x,t) + (1-\sigma)V_{\phi,\varepsilon}(x,t))d\sigma \right) \\ &\times (G_{\phi,\varepsilon}(x,t) - V_{\phi,\varepsilon}(x,t)) + D_{2,\phi,\varepsilon} \\ (G_{\phi,\varepsilon} - V_{\phi,\varepsilon})(x,0) = &\, D_{1,\phi,\varepsilon},\end{aligned}$$

for some $D_{1,\phi,\varepsilon} \in (\mathscr{N}(\mathbb{R}))^n$ and $D_{2,\phi,\varepsilon} \in (\mathscr{N}(\mathbb{R}^2))^n$. We have to prove that $G_{\phi,\varepsilon} - V_{\phi,\varepsilon}$ is in $(\mathscr{N}(\mathbb{R}^2))^n$. This can be done in the same way as in the first part of the proof.

Proposition 2. *Let the initial data* $(a_1, \ldots, a_n)$ *in* (1) *belong to* $(C(\mathbb{R}))^n$.

(a) *The solution* G_h *to the regularized system* (2) *is* L^∞*-associated with the continuous local solution* g *to* (1) *in* K_{T_0}, *for some* $T_0 > 0$.

(b) *Assume that* (1) *is globally well posed. Then the solution* G_h *to* (2) *is* L^∞*-associated with the continuous solution* g *to* (1) *on each* K_T.

Remark. If for every compact set $K \subset \mathbb{R}^2$ there exists $C > 0$ such that

$$\sup_{\substack{(x,t)\in K \\ y\in\mathbb{C}^n}} |F(x,t,y)| \leqslant C$$

or if for every compact set $K \subset \mathbb{R}^2$ there exists $C > 0$ such that

$$\sup_{\substack{(x,t)\in K \\ y\in\mathbb{C}^n}} |\nabla_y F(x,t,y)| \leqslant C,$$

then system (1) is globally well posed.

Proof. (a) Since a is bounded in any compact set K_0, there exists $C > 0$ such that

$$\|a * \phi_\varepsilon(x)\|_{L^\infty(K_0)} = \|(a_1 * \phi_\varepsilon(x), \ldots, a_n * \phi_\varepsilon(x))\|_{L^\infty(K_0)} \leqslant C, x \in K_0, \varepsilon \in (0,1).$$

Therefore, there is some $T_0 > 0$ such that we may apply the fixed point theorem as in [4, Proposition 13.1 a)] and obtain a $(C^\infty)^n$-family of solutions $\{G_{h,\phi,\varepsilon}(x,t), \phi \in \mathscr{A}_0, \varepsilon \in (0,1)\}$ defined on K_{T_0} which are uniformly bounded with respect to ε. This implies that for $\varepsilon < \varepsilon_0$,

$$F_h^k(x,t,G_{h,\phi,\varepsilon}(x,t)) = F^k(x,t,G_{h,\phi,\varepsilon}(x,t)), (x,t) \in K_{T_0}, k \in \{1,\dots,n\}. \tag{6}$$

Gronwall's inequality implies that for some $C_3 > 0$

$$|G_{h,\phi,\varepsilon}(x,t) - g(x,t)| \leqslant C_3 \|a*\phi_\varepsilon - a\|_{L^\infty(K_0)} \to 0, \quad \text{as } \varepsilon \to 0.$$

This finishes the proof of (a).

(b) Let g be the continuous solutions to

$$(\partial_t + \Lambda(x,t)\partial_x)g = F(x,t,g), g(x,0) = a(x) \quad \text{on } K_T$$

and $G_{\phi,\varepsilon} \in (C^\infty(\mathbb{R}^2))^n$ be the solution to

$$(\partial_t + \Lambda(x,t)\partial_x)G_{\phi,\varepsilon} = F(x,t,G_{\phi,\varepsilon}), \quad G_{\phi,\varepsilon}(x,0) = a*\phi_\varepsilon(x)$$

on K_T. Since (1) is well posed, we have $\lim_{\varepsilon\to 0} G_{\phi,\varepsilon} = g$, i.e. for every $\delta > 0$ and every $K_T \subset\subset \mathbb{R}^2$ there exists $\eta_1 > 0$ such that

$$\|G_{\phi,\varepsilon}(x,t) - g(x,t)\|_{L^\infty(K_T)} < \delta \quad \text{for } \varepsilon < \eta_1,$$

which implies that there exists $C_g > 0$ such that

$$\|G_{\phi,\varepsilon}(x,t)\|_{L^\infty(K_T)} \leqslant C_g, \varepsilon < \eta_1.$$

Let $\varepsilon < \min\{\varepsilon_{i_0}, \eta_1\}$. Then $\{(x,t,y) | (x,t) \in K_T, |y| \leqslant C_g\} \subset B_{i_0}$, and the properties of $F^k_{h(\varepsilon)}$ imply

$$F(x,t,G_{\phi,\varepsilon}(x,t)) = F_{h(\varepsilon)}(x,t,G_{\phi,\varepsilon}(x,t)),$$

for $\varepsilon < \min\{\eta_{i_0}, \delta_1\}$ and $(x,t) \in K_T \subset B_{i_0}$. Therefore, $G_{\phi,\varepsilon}$ is a solution to (2′) in K_T. By Proposition 1, the solution to (2′) is unique. Hence we have $G_{h,\phi,\varepsilon} = G_{\phi,\varepsilon}$ in K_T up to an element of $\mathscr{N}$.

Remarks. (a) As long as the smooth solution g to (1) with smooth initial data exists, it is a solution to (2) for every h, because it does not depend on ε and there exists $i \in \mathbb{N}$ such that

$$\sup_{(x,t)\in K \subset \pi(B_i)} \{|g(x,t))|\} < h(\varepsilon), \varepsilon < \varepsilon_i,$$

where π denotes the projection $\mathbb{R}^{2+2n} \to \mathbb{R}^2$.

(b) If there is a solution to (1) which is a generalized function of bounded type, then this is the solution to (2) for every h of logarithmic growth, since $F = F_{h(\varepsilon)}$ in this case.

(c) The supposition that F is a smooth function can be relaxed in the following way. If $F \in C^1$ with polynomial growth in $\pm\infty$, then, if we regularize it by

$$F_{h(\varepsilon)} = (F*\phi_\varepsilon)\kappa_{h(\varepsilon)},$$

Propositions 1 and 2 a) hold if

$$h(\varepsilon) = \mathcal{O}((\log \varepsilon^{-1})^{r/2}), \varepsilon \to 0, \quad \text{for some } r < 1.$$

Proposition 2 b) holds with the additional assumptions that imply the boundeness of the family $\{G_{h,\phi,\varepsilon}, \varepsilon \in (0, \varepsilon_0)\}$ on compact subsets of $\mathbb{R}^2$.

Acknowledgement. The authors would like to express their gratitude to M. Oberguggenberger for his valuable remarks concerning the paper.

References

[1] Biagioni HA (1990) A Nonlinear Theory of Generalized Functions. Berlin Heidelberg New York: Springer
[2] Colombeau JF (1985) Elementary Introduction in New Generalized Functions. Amsterdam North Holland
[3] Colombeau JF, Heibig A, Oberguggenberger M (1996) Generalized solutions to partial differential equations of evolution type. Acta Applicandae Mathematicae **45**: 115–142
[4] Colombeau JF, Heibig A (1994) Generalized solutions to Cauchy problems. Mh Math **117**: 33–49
[5] Illner R, Reed MC (1981) The decay of solutions of the Carleman model. Math Meth Appl Sci **3**: 121–127
[6] Oberguggenberger M (1992) Multiplication of Distributions and Applications to partial Differential Equations. Pitman Res Not Math **259**. Essex: Longman
[7] Oberguggenberger M (1987) Generalized solutions to semilinear hyperbolic systems. Mh Math **103**: 133–144
[8] Oberguggenberger M (1991) The Carleman system with positive measures as initial data – generalized solutions. Transport Theory Statist Physics **20**: 177–197

M. Nedeljkov and S. Pilipović
University of Novi Sad
Faculty of Science
Institute of Mathematics
Trg D. Obradovića 4
Novi Sad
Yugoslavia

Buchbesprechungen – Book Reviews

Fischer, H., Riedmüller, B., Schäffler, S. (Eds.): *Applied Mathematics and Parallel Computing.* Festschrift for Klaus Ritter. 39 Figs., VIII, 376 pp. Physica, Heidelberg, 1996. Cloth DM 128,–.

This Festschrift for Klaus Ritter honours him on the occasion of his sixtieth birthday. Klaus Ritter has, over the years, done research in a rather unconventional combination of areas. These include data analysis, among others the proper representation of high dimensional data; various aspects of automatic differentiation; questions in hydrology; optimization and optimal control; and various technical questions. In tackling these problems, computer software and even hardware has been developed by Ritter. The articles of his friends and colleagues belong more or less to those areas of research even in that several of them discuss extensively parallel aspects of the resulting algorithms and even the implementation on parallel machines, in particular on transputer networks.

H. MUTHSAM, Wien

Rade, L., Westergren, B.: *Springers Mathematische Formeln.* Taschenbuch für Ingenieure, Naturwissenschaftler, Wirtschaftswissenschaftler. 543 S. Springer, Berlin Heidelberg New York, 1996. Brosch. DM 48,–; öS 350,40.

Dies ist eine nützliche und umfangreiche Formelsammlung, die folgende Gebiet umfaßt: Grundlagen und diskrete Mathematik, Algebra und Lineare Algebra, Geometrie und Trigonometrie, elementare Funktionen, Folgen und Reihen, ein- und mehrdimensionale Differentialrechnung, Integrale, gewöhnliche Differentialgleichungen, Vektoranalysis, Spezielle Funktionen, Transformationen, komplexe Analysis, Optimierung, numerische Mathematik, Wahrscheinlichkeitstheorie und Statistik. Außerdem erleichtert ein umfangreicher Index das Auffinden von Formeln.

R. BÜRGER, Wien

Dunham, W.: *Mathematik von a bis z.* Eine alphabetische Tour durch vier Jahrtausende. 360 S. Birkhäuser, Basel Berlin Boston, 1996. Geb. DM 68,–.

Dieses vorzüglich organisierte, abwechslungsreiche, in vielen Richtungen höchst anregende, mit Liebe geschriebene, informative Buch eines sehr kompetenten Autors über grundlegende mathematische Themen und Persönlichkeiten, mit vielen köstlichen Anekdoten, wird auch den der Mathematik eher skeptisch gegenüberstehenden Leser erfreuen und neugierig machen. (Im für einen Angloamerikaner nicht selbstverständlichen Kapitel „Eine Lanze für Leibniz" hätte ich auf wichtige außermathematische Leistungen dieses Universalgelehrten, z. B. für die Geschichtsforschung, hingewiesen, im Gegensatz zu Newtons „Leistungen". Ein Kuriosum: Als Gegenbeispiel zu einer Vermutung von Polignac wird die Zahl 127 angegeben, das kleinste Gegenbeispiel ist aber 905; Polignac selbst gibt nur das nächstgrößere Gegenbeipiel 959 an [C.R. Paris 1849, p. 738–739] für eine Zahl, die keine Primzahl ist und nicht als Summe einer Potenz von 2 und einer Primzahl geschrieben werden kann. Diese Gegenbeispiele eignen sich gut zur Konstruktion kniffliger Denksportaufgaben.)

H. RINDLER, Wien

Casti, J.: *Die großen Fünf.* Mathematische Theorien, die unser Jahrhundert prägten. 217 S. Birkhäuser, Basel Berlin Boston, 1996. Geb. DM 68,–.

Der Autor behandelt hier für einen breiten Leserkreis fünf in vieler Hinsicht bedeutende mathematische Theoreme dieses Jahrhunderts: den Minimax-Satz, den Brouwerschen Fix punkt satz, das Morse-Theorem, den Anhalte-Satz und die Simplexmethode, ihr mathematisches Umfeld und ihre Bezüge zu anderen Wissensgebieten. Ausgehend von allgemein verständlichen interessanten Fragestellungen, vermittelt der Autor grundlegende Ideen schwieriger mathematischer Sachverhalte in oft unterhaltsamer Weise, manchmal etwas ungenau oder verkürzend, aufgelockert durch vorzügliche bis kühne, aber immer anregende Analogien.

H. RINDLER, Wien

Rautenberg, W.: *Einführung in die Mathematische Logik.* Ein Lehrbuch mit Berücksichtigung der Logikprogrammierung (Vieweg Lehrbuch Mathematik). XII, 250 S. Vieweg, Braunschweig Wiesbaden, 1996. Brosch. DM 39,50; öS 288,–.

Einführend wird außer den Grundbegriffen der Aussagenlogik und der Prädikatenlogik der Gödelsche Unvollständigkeitssatz eingehend behandelt. Darüber hinaus gibt es weiterführende Kapitel, in denen „Grundlagen der Logikprogrammierung", „Elemente der Modelltheorie" vorgestellt und ein Ausblick auf die Rekursionstheorie gegeben werden.

K. AUINGER, Wien

Drake, F. R., Singh, D.: *Intermediate Set Theory.* X, 234 pp. Wiley, Chichester New York Brisbane, 1996, £ 16,99.

"Intermediate" in the title is best explained by means of the list of contents: 1. History; 2. First-order logic; 3. Axioms of ZF; 4. Cardinals; 5. Ordered sets; 6. The number system; 7. Axiom of Choice; 8. Constructible sets and forcing; 9. Further topics. Approximately half of the text offers material which usually is not covered by introductions to set theory. It takes the reader to a level which allows to follow the researches of the last thirty years or so. In particular, proofs of the basic facts about Gödel's constructible sets and Cohen's method of forcing are included. This justifies the title of this book, which will be of great help for everybody interested in the logical foundations of mathematics.

H. MITSCH, Wien

Booth, D., Ziegler, R. (Eds.): *Finsler Set Theory: Platonism and Circularity.* Translation of Paul Finsler's Papers on Set Theory with Introductory Comments. IX, 278 pp. Birkhäuser, Basel Berlin Boston, 1996. Cloth DM 98,–; öS 716,–.

It is little known that the Swiss geometer Paul Finsler concerned himself intensively with the foundations of mathematics. His unusual approach in founding set theory had been objected to by R. Baer and his argument had been endorsed in the following willingly enough, for after all formalism clashed with Platonism. Although Baer's argument was erroneous from the Platonic view, Finsler's theory, probably the nearest to Cantor's conception, was washed away by the flood of formal set theory. Today interest in the former is growing since it comprises much more sets as usual – including non-founded ones – and combinatorial aspects come to light. Finsler's papers on set theory had been published already in 1975, but are out-of-print for a long time past. The most essential of them are presented here in English translation and arranged in three parts: philosophical, foundational, combinatorial. In contrast to the previous edition the editors have added extremely helpful commentaries which give today's reader, who is grown up in formalist views, an understanding of Finsler's Platonism and way of argumentation and by the way

make this "absolute" set theory accessible to anyone who is interested in the foundations of mathematics.

G. KOWOL, Wien

Glodstern, M., Judah, H.: *The Incompleteness Phenomenon. A New Course in Mathematical Logic.* XIII, 247 pp. A. K. Peters, Wellesley, Massachusetts, 1995. US $ 49,95.

This book is a well written and well organized course in mathematical logic, which leads from the basics (mathematical languages and logical systems) over Gödel's completeness theorem (provability=validity) and model theory to Gödel's incompleteness theorem (and recursion). "We think that the material of this book should be part of the basic background of every student in any discipline which employs deductive and formal reasoning as a part of its methodology. This definitely includes a large part of the social sciences." This is a very reasonable position, and the authors do their best not to scare away any potential reader, but since mathematical logic is (extremely) abstract by nature, there is not too much hope ...

P. SCHMITT, Wien

Waldschmidt, M., Moussa, P., Luck, J. M., Itzykson, C. (Eds.): *From Number Theory to Physics.* 93 Figs., X, 690 pp. Springer, Berlin Heidelberg New York, 1995. Cloth DM 138,–.

The fourteen contributions in this book are excellent self-contained courses on specific themes which cover a large part of number theory and reveal connections to physics. The editors gained prominent mathematicians and physicists for these articles. The headlines of the chapters read as follows. "An Introduction to Zeta Functions" (P. Cartier), "Introduction to Compact Riemann Surfaces, Jacobians and Abelian Varieties" (J.-B. Bost), "Elliptic Curves" (H. Cohen), "Introduction to Modular Forms" (D. Zagier), "Decorated Elliptic Curves: Modular Aspects" (R. Gergondey), "Galois Theory, Algebraic Number Theory, and Zeta Functions" (H. M. Stark), "Galois Theory for Coverings and Riemann Surfaces" (E. Reyssat), "Differential Galois Theory" (F. Beukers), "p-adic Numbers and Ultrametricity" (G. Christol), "Introduction to Lattice Geometry" (M. Senechal), "A Short Introduction to Quasicrystallography" (A. Katz), "Gap Labelling Theorems for Schrödinger Operators" (J. Bellissard), "Circle Maps: Irrationality Winding" (P. Cvitanović), "An Introduction to Small Divisors Problems" (J.-C. Yoccoz). It is a second and corrected version of a book originally published in 1992. It contains a photograph of C. Itzykson who died on May 22, 1995.

J. SCHOISSENGEIER, Wien

Kuznetsov, Y. A.: *Elements of Applied Bifurcation Theory* (Applied Mathematical Sciences, Vol. 112). 232 Figs., XV, 515 pp. Springer, Berlin Heidelberg New York, 1995. Cloth DM 98,–; öS 764,40.

The first five chapters cover the standard material: codimension one bifurcations in differential equations (fold and Andronov–Hopf) and difference equations (fold, flip and Neimark–Sacker), and center manifold reduction. Here the author derives an extremely concise formula for the direction of the Hopf bifurcation in any dimension. What I missed, is a discussion of the transcritical bifurcation. Chapters 6 to 10, which occupy two thirds of the book collect much advanced material: homoclinic bifurcations, with a detailed description of the work of Shil'nikov; some "exotic" bifurcations; 100 pages on the five types of codimension two bifurcations of vector fields; a similar chapter for maps (with a discussion of the open problems). For some of these more complex topics, the author

makes clear the relevant geometric ideas behind the proofs, rather than giving every detail. This approach makes the book readable for a wide audience. Each chapter puts much emphasis on how to do concrete bifurcation calculations, with practical illustrations through classical examples from the applied sciences. Although the book concentrates on the finite dimensional case, some detailed examples of reaction–diffusion equations demonstrate its applicability beyond that. The final chapter is devoted to the numerical analysis of bifurcations and continuation methods and ends with a brief survey of available software packages. (The author is well-known for his program LOCBIF). The book comes with many exercises, detailed bibliographical notes and a huge list of references. – This is an extremely valuable book, both for the theoretical worker, who gets a detailed account of the work of the Russian schools and a balanced description of the state of the art, and even more for the practitioner, who is taught a rich toolbox of applied bifurcation analysis.

J. HOFBAUER, Wien

Hagen, R., Roch, S., Silbermann, B.: *Spectral Theory of Approximation Methods for Convolution Equations* (Operator Theory, Advances and Applications, Vol. 74). 27 Figs., XII, 373 pp. Birkhäuser, Boston, 1995. Cloth DM 198,–; öS 1466,–.

The basic question throughout this book is the question of different kinds of invertibility for certain operators. Besides usual inversion problems constructive aspects, such as the convergence of solutions to approximate problems (e.g. using spline approximation, Galerkin methods, etc.) have gained interest recently. As a unifying idea the authors propose to reformulate related concepts in the frame-work of (classical) invertibility in new Banach algebras (mostly consisting of suitable sequences of operators). The book covers in a concise way a very interesting chapter of applied functional analysis, holding promises for interesting developments.

H. G. FEICHTINGER, Wien

Wan, F. Y. M.: *Introduction to the Calculus of Variations and its Applications.* XVIII, 638 pp. Chapman & Hall, New York Albany Bonn, 1995. Cloth £ 49,95.

The calculus of variations and control theory are the classical tools for analyzing and solving optimization problems involving variable functions. Solutions are typically smooth or piece-wise smooth; weak and strong solutions are discussed in this book. Three quarters of the book cover problems, theory and solution methods involving vector-valued functions of a single variable, first without then with constraints. The remainder is about functions in higher dimensions, covering Plateau's problem and problems from elasticity theory, plate theory, and fluid mechanics. For each of the 18 chapters, about a dozen exercises are provided.

A. NEUMAIER, Wien

Morgan, F.: *Riemannian Geometry. A Beginner's Guide.* 119 pp. Jones and Bartlett, Boston London, 1995. US $ 30,–.

This pleasant little book offers a quick tour to Riemannian Geometry. It focuses on curvature (its central concept) and is based on surfaces in n-dimensional space (rather than on abstract manifolds), presented in such a way that the more sophisticated intrinsic formulas are natural consequences. According to its character, it emphasizes the main ideas (and visual imagination). Thus, proofs (when given at all) concentrate on the crucial steps and leave out the (often messy) details. The power and usefulness of the theory is demonstrated by examples from hyperbolic geometry, general relativity (orbit of Mercury), global differential geometry (Gauss–Bonnet theorem, geodesics), and energy-minimizing curves (isoperimetric problem and crystals). The exposition is also well suited to be read before or parallel to a more comprehensive course or textbook.

P. SCHMITT, Wien

Miles, R. E.: *Symmetric Bends. How to Join Two Lengths of Cord* (Series on Knots and Everything, Vol. 8). XII, 163 pp. World Scientific, Singapore New Jersey London, 1995. £ 23,–.

This monograph "relates to everyday concrete things – knots which join two lengths of cord". It is (probably) the first effort to deal with this subject (more precisely: with a certain subtopic, called *symmetric bends*) from a mathematical point of view (i.e., as a special type of knots and links in the framework of knot theory). But since it is written in a leisurely style (which requires only few prerequisistes from mathematics) and includes discussions of practical aspects, as well, it is also a piece of recreational mathematics, and a book for (non-mathematical) knot enthusiasts, in general.

P. SCHMITT, Wien

Shampine, L. F.: *Numerical Solution of Ordinary Differential Equations.* X, 484 pp. Chapman & Hall, New York London, 1994. Cloth £ 49,95.

A practical book about solving ordinary differential equations (ODEs) numerically. There are very few theorems (the first one appears on p. 216) but many formulas, figures, examples and exercises explaining and illustrating what goes on and what can go wrong with problem formulations, solution methods, and software packages. The reader thus gets a vivid impression of the role that order, memory, stability, stiffness, large and variable step sizes, error control, Jacobian information and other things play in the design and use of modern ODE solvers. The book thus complements nicely more formal texts on the numerical analysis of ODEs.

A. NEUMAIER, Wien

Strehmel, K., Weiner, R.: *Numerik gewöhnlicher Differentialgleichungen.* 462 S. Teubner, Stuttgart, 1995. Brosch. DM 49,80.

Dieses Lehrbuch über die Methoden zur numerischen Behandlung gewöhnlicher Differentialgleichungen bespricht in drei Teilen die Behandlung von Anfangswertproblemen für nichtsteife und steife Differentialgleichungen, sowie für Algebro-Differentialgleichungen und retardierte Differentialgleichungen. Es wurde entworfen als Bindeglied zwischen der Einführung in numerische Methoden zur Lösung von Differentialgleichungen, wie sie in Lehrbüchern zur numerischen Mathematik geboten wird, und anspruchsvollen Monographien für Spezialisten. Theorie und Algorithmen der wesentlichen Verfahren sind gut dargestellt. Daher ist dieses Buch sehr brauchbar als Lehrbuch oder als Begleitbuch zu einer Vorlesung.

H. SCHICHL, Wien

Gander, W., Hrebicek, J. (Eds.): *Solving Problems in Scientific Computing Using Maple and MATLAB*, 2nd Edn. 315 pp. Springer, Berlin Heidelberg, 1995. Softcover DM 73,–; öS 530,40,–.

Treating a variety of realistic scientific problems this volume provides a well written account of the capacity of modern symbolic and numeric software packages. Topics covered in this volume include trajectories of a spinning tennis ball, orbits in the planar three-body problem, internal field in semiconductors, the radar problem, symbolic computation of explicit Runge–Kutta formulas and many more. In this second edition, two new chapters (transient response of a two-phase half-wave rectifier, circuits in power electronics) have been added. The philosophy behind this book, namely that solving nontrivial examples is the best way of acquiring in-depth knowledge of the subject is particularly valid in the field of scientific computing.

M. KUNZINGER, Wien

Gaylord, R. J., Kamin, S. N., Wellin, P. R.: *Einführung in die Programmierung mit Mathematica.* XIX, 310 S. Birkhäuser, Basel Berlin Boston, 1995. Brosch. DM 70,–; öS 496,–.

Das Buch wendet sich an leicht fortgeschrittene Matheamtica-Benutzer, die versuchen wollen, nicht nur isoliert die vorgegebenen Befehle zu verwenden, sondern auch durch selbstdefinierte Funktionen zu ergänzen oder in Prozeduren und Paketen zusammenzufassen. Dazu werden die Datenstrukturen und die von Mathematica ermöglichten (vielseitigen) Programmiertechniken erläutert (damit kann das Buch auch als allgemeine erste Einführung in Programmierung verwendet werden). Die Themenbehandlung erfolgt in anschaulicher Form an Hand einer großen Zahl von Beispielen aus den verschiedensten Bereichen der Mathematik: numerische, kombinatorische Aufgaben bis hin zur Graphikprogrammierung.

V. Losert, Wien

Casey, J., Crochet, M. J. (Eds.): *Theoretical, Experimental, and Numerical Contributions to the Mechanics of Fluids and Solids.* ZAMP, Zeitschrift für angewandte Mathematik und Physik; A collection of papers in honor of Paul M. Naghdi (Special issue, Vol. 46). 847 pp. Birkhäuser, Basel Berlin Boston, 1995. Cloth DM 460,–; öS 3.260,40,–.

A considerable number of authors have contributed to the celebration volume of P. M. Naghdi's 70th brithday; soon afterwards, Naghdi died. Naghdi contributed to many aspects of continuum mechanics, the celebration papers do so as well. While a few papers treat with numerical aspects or methods, the larger part is in the classical analytical style of mechanics.

H. Muthsam, Wien

Adam, A., Bellomo, N.: *A Survey of Models for Tumor-Immune System Dynamics* (Modelling and Simulation in Science, Engineering and Technology). XIV, 344 pp. Birkhäuser, Basel Berlin Boston, 1997. Cloth DM 144,–; öS 1008,–.

Dieses Buch ist eine Sammlung von sieben Übersichtsartikeln, welche Modelle für Tumordynamik und die Wechselwirkung zwischen Tumoren und Immunsystem beschreiben. Dabei wird auch viel Hintergrundinformation geliefert (über Immunologie ebenso wie über Methoden der angewandten Mathematik), so daß dieses Buch zweifellos eine wichtige Bereicherung der Literatur auf diesem rasch wachsenden Gebiet darstellt.

K. Sigmund, Wien

Verleger: Springer-Verlag KG, Sachsenplatz 4–6, A-1201 Wien. — Herausgeber: Prof. Dr. Karl Sigmund, Institut für Mathematik der Universität Wien, Strudlhofgasse 4, A-1090 Wien. — Redaktion: Strudlhofgasse 4, A-1090 Wien. — Satz und Umbruch: Thomson Press (India) Ltd., New Delhi. — Offsetdruck: Druckerei Eugen Ketterl Ges.m.b.H., A-1180 Wien. Verlagsort: Wien. — Herstellungsort: Wien. — Printed in Austria.

Mh. Math. 125, 269–278 (1998)

Monatshefte für
Mathematik

Printed in Austria

Binomial-Coefficient Multiples of Irrationals

By

Terrence M. Adams and **Karl E. Petersen**, Chapel Hill, NC

(Received 2 August 1996; in revised form 24 February 1997)

Abstract. Denote by x a random infinite path in the graph of Pascal's triangle (left and right turns are selected independently with fixed probabilities) and by $d_n(x)$ the binomial coefficient at the n'th level along the path x. Then for a dense G_δ set of θ in the unit interval, $\{d_n(x)\theta\}$ is almost surely dense but not uniformly distributed modulo 1.

1. Introduction

The *Pascal graph* is the directed infinite planar graph with vertices (n, k), for $n = 0, 1, \ldots$ and $k = 0, \ldots, n$ and two edges coming out of each vertex (n, k), one to $(n + 1, k)$ and one to $(n + 1, k + 1)$. Labeling edges of the first kind by 0 and of the second kind by 1 produces a natural correspondence between infinite sequences $x \in \Omega = \{0, 1\}^{\mathbb{N}}$ and infinite paths in the Pascal graph which start at the root vertex (0, 0). We will denote by $d_n(x)$ the binomial coefficient $C(n, k_n(x))$ found at the n'th vertex of x, if the Pascal graph is superimposed on the Pascal triangle. The *Pascal adic transformation* on the space X of infinite paths (see [9, 11, 12, 6]) corresponds to the map $T : \Omega \to \Omega$ given by $T(1^p 0^q 01 \ldots) = 0^q 1^p 10 \ldots$ for $p, q \geqslant 0$. VERSHIK [9] noted that the invariant ergodic measures for this map are exactly the Bernoulli measures $\mu_\alpha = \mathscr{B}(\alpha, 1 - \alpha)$ on Ω and conjectured [10] that they are weakly mixing. It was noted in [6] that if, for a fixed Bernoulli measure on Ω, λ is an eigenvalue of T, then $\lambda^{d_n(x)} \to 1$ for a.e. x. The question of whether or not there exist such λ, and its variants concerning the distribution of the points $\lambda^{d_n(x)}$ on the unit circle for typical x or indeed for all x, some of them also mentioned in [6], lead to the study of the distribution modulo 1 of binomial-coefficient multiples of irrationals; answering many of these questions seems to demand deeper understanding of the divisibility properties of binomial coefficients than we have at present.

While we are not yet able to answer the question of weak mixing for the Pascal adic transformation, we do have some progress on related questions. First, we note that if x is a path in the Pascal graph which tracks a line of a fixed slope α, then the proportion of j's, $0 \leqslant j \leqslant n$, for which $d_j(x)$ is divisible by a fixed prime q tends to 1

1991 Mathematics Subject Classification: 28D05, 28D99

Key words: Uniform distribution modulo 1, Pascal adic transformation, measure preserving transformation, eigenvalue, weak mixing

as $n \to \infty$. Using this, we construct an uncountable set of λ on the unit circle such that for a.e. path x in the Pascal graph (with respect to a fixed Bernoulli measure) the points $\lambda^{d_n(x)}$ are not uniformly distributed on the circle, since asymptotically too large a fraction of them are near 1. Thus these points might be candidates for eigenvalues of T – but we construct such λ for which the $\lambda^{d_n(x)}$ are dense. We also list several further observations, questions, and conjectures about the distribution of these points; perhaps the strongest conjecture (also mentioned in [6]) is the following: If $\lambda \in \mathbb{C}$ and there exists a path x in the Pascal graph for which $\lambda^{d_n(x)} \to 1$, then $\lambda = 1$.

The second author gratefully acknowledges the support of the Erwin Schrödinger Institute, Vienna, where part of this research was conducted.

2. Intersections of Lines with Sierpinski's Gasket

Let E denote the triangle with vertices (0, 0), (1, 0), and (0, 1). In this section we construct Sierpinski's gasket as a subset of E. Then we prove that every straight line path (with slope strictly greater than 0) through Sierpinski's gasket intersects the gasket in a set with one-dimensional Lebesgue measure 0. (Since Sierpinski's gasket has two-dimensional Lebesgue measure 0, Fubini's theorem guarantees that almost every line with a specified slope intersects Sierpinski's gasket with one-dimensional Lebesgue measure 0, but this is not sufficient for our purpose.) We give two lemmas which set up the general proof. Also we apply this to the generalized Sierpinski gasket defined at the end of this section.

Given a natural number n, let E_n^1 denote the interior of the right triangle with vertices $((2^{n-1}-1)/2^{n-1}, 1/2^n), ((2^n-1)/2^n, 0)$ and $((2^n-1)/2^n, 1/2^n)$ and let E_n^2 be the interior of the triangle with vertices $(1/2^n, (2^{n-1}-1)/2^{n-1})$, $(0, (2^n-1)/2^n)$ and $(1/2^n, (2^n-1)/2^n)$. We view the collection of E_n^i's contained in the larger triangle E with vertices (0, 0), (1, 0) and (0, 1). The next two lemmas concern the intersection of the E_n^i's with a straight line of slope $\gamma > 0$ (where γ corresponds to the invariant Bernoulli measure μ_α, with $\alpha = 1/(1+\gamma)$ for the Pascal adic transformation).

Lemma 2.3. *Given $\alpha > 0$ there exists $\eta = \eta(\alpha) > 0$ such that: for any straight line L of slope α which intersects the interior of E, there exist n and i such that*

$$\frac{\mu(L \cap E_n^i)}{\mu(L \cap E)} \geqslant \eta.$$

Before we prove Lemma 2.3 we state and prove the preliminary Lemma 2.2.

Lemma 2.2. *Given $\gamma > 0$ there exist positive real numbers $\varepsilon = \varepsilon(\gamma)$ and $\eta = \eta(\gamma)$ such that for all $b \in [-\varepsilon, (1-\varepsilon)/2]$ and $L_b = \{(x, \gamma(x-b)) : x \in \mathbb{R}\}$ we have*

$$\frac{\mu(L_b \cap E_1)}{\mu(L_b \cap E)} \geqslant \eta.$$

Proof. First we find the intersection of L_b with the hypotenuse of E_1. Solving $y = \gamma(x-b)$ and $y = 1/2 - x$ simultaneously, we obtain $x = (1/2 + \gamma b)/(1+\gamma)$

and $y = \gamma(1/2 - b)/(1 + \gamma)$. Thus if we choose $\varepsilon < 1/(2\gamma)$ our point of intersection will have positive x and y coordinates. Hence the function $f(b) = \mu(L_b \cap E_1)/\mu(L_b \cap E)$ is positive and continuous on the closed interval $[-\varepsilon, (1 - \varepsilon)/2]$ and therefore achieves a positive minimum value η. □

Proof of Lemma 2.3. Any line L which intersects the interior of E intersects either the line segment joining (0, 0) to (1, 0) or the line segment joining (0, 0) to (0, 1). Without loss of generality let us consider line segments L intersecting the segment joining (0, 0) to (1, 0). If we let $I = \{(x, 0) : 0 \leqslant x < 1\}$, then $I \cap L \neq \emptyset$. In this case we may focus on the sets $E_n = E_n^1$.

Choose $\varepsilon > 0$ as in Lemma 2.2. (In particular $\varepsilon < 1/(2\gamma)$ will work.) We may cover I with subintervals $I_n = [(2^{n-1} - 1)/2^{n-1} - \varepsilon/2^{n-1}, (2^n - 1)/2^n - \varepsilon/2^n)]$; so we have $I = \bigcup_{n=1}^{\infty} I_n$. For each $b \in I_n$ define $L_b = \{(x, \gamma(x - b)) : x \in \mathbb{R}\}$ and $f_n : I_n \to [0, 1]$ as $f_n(b) = \mu(L_b \cap E_n)/\mu(L_b \cap E)$. The self-similarity properties of the triangles E_n imply that each $f_n : I_n \to [0, 1]$ is continuous, and they all have the same image. Therefore by Lemma 2.2 there exists a single real number $\eta > 0$ such that

$$f_n(b) = \frac{\mu(L_b \cap E_n)}{\mu(L_b \cap E)} \geqslant \eta$$

for all positive integers n and all $b \in I_n$. □

Now we construct Sierpinski's gasket as a closed nowhere dense subset of E. We call triangles in the plane *lower triangles* if we can label the vertices $(a_1, b_1), (a_2, b_2)$ and (a_3, b_3) so that the right angle is at (a_2, b_2) and $a_2 = \min\{a_1, a_3\}$ and $b_2 = \min\{b_1, b_3\}$. *Upper triangles* have right angle at (a_2, b_2) with $a_2 = \max\{a_1, a_3\}$ and $b_2 = \max\{b_1, b_3\}$. Note that given a lower triangle R there is a unique upper triangle (inscribed in R) whose vertices are the midpoints of the sides of R. This upper triangle is denoted $\mathcal{U}(R)$; let $\mathcal{L}(R) = \{R_1, R_2, R_3\}$ be the collection of lower triangles remaining when we extract $\mathcal{U}(R)$ from R. Also, given a collection $\mathcal{C}$ of lower triangles let $\mathcal{U}(\mathcal{C}) = \{\mathcal{U}(R) : R \in \mathcal{C}\}$ and let

$$\mathcal{L}(\mathcal{C}) = \bigcup_{R \in \mathcal{C}} \mathcal{L}(R).$$

We take the triangles in $\mathcal{L}(R)$ to be closed.

The following proposition uses Lemma 2.3 to prove that $\mu(L \cap G) = 0$ for any line L with slope $\gamma > 0$. First note that Lemmas 2.2 and 2.3 can be extended to any lower triangle playing the role of the initial triangle E. Also note that $E_n^i \in \mathcal{U}(\mathcal{L}^{n-1}(E))$ for positive integers n and $i = 1, 2$. This implies that each $R \in \mathcal{L}^n(E)$ is disjoint from the interior of E_n^i. Let

$$G_n = \bigcup_{R \in \mathcal{L}^n(E)} R.$$

Then $G = \bigcap_{n=1}^{\infty} G_n$ is *Sierpinski's gasket.*

Proposition 2.4. *If* $\gamma > 0$ *and* $L = \{(x, \gamma x) : x \in \mathbb{R}\}$, *then* $\mu(G \cap L) = \lim_{n \to \infty} \mu(G_n \cap L) = 0$.

Proof. Choose $\eta = \eta(\gamma) > 0$ as in Lemma 2.3. We construct inductively a sequence n_j of natural numbers such that for all positive integers j we have

$$\frac{\mu(L \cap G_{n_j})}{\mu(L \cap E)} \leqslant (1-\eta)^j.$$

For the primary case Lemma 2.3 ensures that there exists $E_{n_1}^{i_1} \in \mathscr{U}(\mathscr{L}^{n_1-1}(E))$ such that $\mu(L \cap E_{n_1}^{i_1})/\mu(L \cap E) \geqslant \eta$. (Actually $n_1 = 1$.) Hence $\mu(L \cap G_{n_1})/\mu(L \cap E) \leqslant \leqslant 1 - \eta$.

For the general case suppose that n_k satisfies $\mu(L \cap G_{n_k})/\mu(L \cap E) \leqslant (1-\eta)^k$. Now $\mathscr{L}^{n_k}(E) = \{R_1, \ldots, R_p\}$ is composed of a finite number of closed triangles. Let $\mathscr{M} = \{m : 1 \leqslant m \leqslant p, R_m \cap L \neq \emptyset\}$. Thus $\mu(R_m \cap L) = 0$ for $m \notin \mathscr{M}$. For each $m \in \mathscr{M}$, by Lemma 2.3 there exists a positive integer $e(m)$ and $E(m) \in \mathscr{U}(\mathscr{L}^{e(m)-1}(R_m))$ such that $\mu(L \cap E(m))/\mu(L \cap R_m) \geqslant \eta$. Let $e = \max_{m \in \mathscr{M}}\{e(m)\}$, and let $n_{k+1} = n_k + e$. Therefore we have

$$\begin{aligned}\mu(L \cap G_{n_{k+1}}) &\leqslant (1-\eta)\mu(L \cap G_{n_k}) \\ &\leqslant (1-\eta)^{k+1}.\end{aligned}$$

□

Now we define the *generalized Sierpinski gasket* and give the analogous proposition which may be proved by the same method. Let E be the closed triangle described above. Given a positive integer q, the lines $y = p/q, x = p/q$ and $y = p/q - x$ partition E into q^2 triangles with $q(q-1)/2$ upper triangles and $q(q+1)/2$ lower triangles. Let $\mathscr{U}_q(E)$ be the collection of upper triangles and $\mathscr{L}_q(E)$ be the collection of lower triangles. (Take the members of $\mathscr{U}_q(E)$ to be open and the members of $\mathscr{L}_q(E)$ to be closed). Define $\mathscr{U}_q(\mathscr{C})$ and $\mathscr{L}_q(\mathscr{C})$ analogously for any collection $\mathscr{C}$ of lower triangles. We obtain

$$G_q^n = \bigcup_{R \in \mathscr{L}_q^n(E)} R, \qquad G_q = \bigcap_{n=1}^{\infty} G_q^n.$$

Proposition 2.5. *If $\gamma > 0$ and $L = \{(x, \gamma x) : x \in \mathbb{R}\}$, then for all positive integers q*

$$\mu(G_q \cap L) = \lim_{n \to \infty} \mu(G_q^n \cap L) = 0.$$

Proof. We explain how to choose sets $E_n^i(q)$ analogous to the E_n^i above. Then the analogues of the previous lemmas and proposition follow in the same manner as before.

Given a lower triangle R let $\mathscr{L}_q(R)$ be the collection of two lower triangles: the top left, lower triangle from $\mathscr{L}_q(R)$ and the bottom right, lower triangle from $\mathscr{L}_q(R)$. Similarly we define $\mathscr{L}_q^n(\mathscr{C})$. The sets $E_n^i(q)$ are chosen in $\mathscr{U}_q(\mathscr{L}_q^{n-1}(E))$. □

3. Binomial Coefficients Modulo a Prime Along a Random Path in Pascal's Triangle

If Pascal's triangle is reduced modulo a prime q, a well-known self-similar pattern (which can be produced by a cellular automaton) results; this is a

consequence of KUMMER's Carry Theorem [4] and the resulting formula of LUCAS [5]. The parts of the triangle that correspond to the upper triangles (which form $(G_q)^c$) (the 'voids') consist of regions in which the binomial coefficients are divisible by q. The following theorem says that since a line of slope γ spends most of its time outside of each G_q^n, a μ_α-typical path x, which eventually approaches a line of slope $\gamma = (1-\alpha)/\alpha$ in Pascal's triangle, spends most of its time on vertices which carry binomial coefficients divisible by q.

Theorem 3.1. *If q is prime and $0 < \alpha < 1$, then for μ_α-almost all x we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} |e^{2\pi i d_j(x)/q} - 1| = 0.$$

Proof. Let $\varepsilon > 0$. Using Proposition 2.5, choose N so large that if $n \geqslant N$ then $\mu(G_q^n \cap L_\alpha) < \varepsilon$. Notice that since the complement of G_q^N is a union of finitely many triangles, if we move L_α just a small amount we cannot decrease the Lebesgue measure of its intersection with $(G_q^N)^c$ by very much. Thus we may choose $\delta > 0$ and then a large-enough natural number M such that if the part within our unit triangle E of the band of width δ about the line L_α is cut into M equally-spaced chunks by lines parallel to the hypotenuse of E, and if one point is chosen from each of those chunks, then the proportion of those points which are in G_q^n in still less than 2ε.

Let $S_k(x)$ denote the number of 1's along the path x (regarded as a sequence in $\Omega = \{0,1\}^{\mathbb{Z}}$). By the Ergodic Theorem, for μ_α-almost every x there is $K = K(x)$ such that

$$|S_k(x) - k\alpha| < k\delta \quad \text{for all } k \geqslant K.$$

Choose a time $M > K/\delta$, and consider Pascal's triangle down to that level, including subtriangles of rank up to N. When this part of Pascal's triangle is scaled down to lie over our unit triangle E, we see the subtriangles that form $(G_q^N)^c$, and the scaled-down path x lies entirely inside the band of width δ about L_γ, the line of slope $\gamma = (1-\alpha)/\alpha$ in the Sierpinski gasket. We have arranged that the proportion of vertices of the scaled-down path which are in $(G_q^N)^c$ is at least $1 - 2\varepsilon$, and hence the proportion of vertices of the path x at which the binomial coefficients $d_j(x)$ are divisible by q is at least $1 - 2\varepsilon$. □

4. Main Result: A Construction of Special Irrationals

In this section we use Theorem 3.1 to construct an uncountable dense set (in fact a G_δ) $\Lambda \subset [0,1)$ such that for each $\theta \in \Lambda$ and for μ_α-almost every $x \in X$ we have that the sequence $\{d_j(x)\theta\}$ is not uniformly distributed modulo 1. In particular we obtain a result similar to Theorem 3.1, but with the limit replaced by lim inf.

Theorem 4.1. *There exist a dense G_δ set $\Lambda \subset [0,1)$ and a set of full μ_α-measure $Y \subset X$ so that for each $\theta \in \Lambda$ and $x \in Y$ we have*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} |e^{2\pi i d_j(x)\theta} - 1| = 0.$$

Proof. Let $\{\varepsilon_n\}$ be a sequence of positive real numbers satisfying $\sum_{n=1}^{\infty} \varepsilon_n < \infty$, and let $\{q_n\}$ be a sequence of primes increasing to ∞. We will produce sequences R_n of natural numbers and $\delta_n > 0$ so that if

$$Y_n = \left\{ x \in X : \frac{1}{R_n} \sum_{j=0}^{R_n - 1} |e^{2\pi i d_j(x) p/q_n} - 1| < \frac{1}{n} \text{ for } p = 0, 1, \ldots, q_n - 1 \right\},$$

$$\Lambda_n = \{\theta \in [0,1) : \text{there exists } p = 0, 1, \ldots, q_n - 1 \text{ with } |\theta - \tfrac{p}{q_n}| < \delta_n\},$$

$$\Lambda = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \Lambda_n, \quad \text{and} \quad Y = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} Y_n,$$

then

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} |e^{2\pi i d_j(x)\theta} - 1| = 0$$

for all $\theta \in \Lambda$ and all $x \in Y$.

For each n, by Theorem 3.1 we may choose R_n so that $\mu_\alpha(Y_n) > 1 - \varepsilon_n$. Then, since

$$\left\{ \frac{1}{R_n} \sum_{j=0}^{R_n - 1} |e^{2\pi i d_j(x)\theta} - 1| : x \in Y_n \right\}$$

is a finite collection of continuous functions of θ, we may choose $\delta_n > 0$ so that

$$\frac{1}{R_n} \sum_{j=0}^{R_n - 1} |e^{2\pi i d_j(x)\theta} - 1| < \frac{2}{n}$$

for all $x \in Y_n$ and all $\theta \in \Lambda_n$. Then Λ is a dense G_δ with the usual topology, and $\mu_\alpha(Y) = 1$ because $\sum_{n=1}^{\infty} \varepsilon_n$ converges.

To verify the outcome of the theorem, first choose $\theta \in \Lambda$ and $x \in Y$. Then there exists a sequence $n_m \to \infty$ such that $\theta \in \Lambda_{n_m}$ for all positive integers m. Also there exists k such that $x \in Y_n$ for $n \geqslant k$. Hence for $n_m \geqslant k$ we have

$$\frac{1}{R_{n_m}} \sum_{j=0}^{R_{n_m} - 1} |e^{2\pi i d_j(x)\theta} - 1| < \frac{2}{n_m}. \qquad \square$$

5. Density Without Uniform Distribution

In the previous section we constructed a dense G_δ set $\Lambda \subset [0,1)$ such that for each $\theta \in \Lambda$, $\{d_j(x)\theta\}$ is not uniformly distributed modulo 1 for μ_α-almost every $x \in X$. Those $\theta \in \Lambda$, which are irrational, remain candidates for eigenvalues of the Pascal adic transformation. However in this section we will show that if the sequence $\{\delta_n\}$ converges to zero sufficiently fast, then $\{d_j(x)\theta\}$ is dense modulo 1 for each $\theta \in \Lambda$ and for μ_α-almost every $x \in X$. This excludes these θ as eigenvalues for the Pascal adic.

We begin by considering Pascal's triangle modulo a prime q. Recall that for $n \in \mathbb{N}$ and $1 \leqslant k \leqslant q^n - 1$, we have $C(q^n, k) \cong_q 0$. Hence $C(q^n - 1, k) \cong_q (-1)^k$ for $0 \leqslant k \leqslant q^n - 1$, which gives a 'blocking line' on the triangle. It is this 'blocking line' that yields total ergodicity of the Pascal adic. In Lemma 5.1 we note that among the binomial coefficients in the row numbered $q^n - 2$ one can find all the congruence classes modulo q, and in fact they appear in a regular way. This allows us to show that along a random path in the triangle a hit of congruence class r at level q^m and of congruence class p at level q^n are approximately independent if m and n are far apart, and therefore with probability 1 no congruence class modulo q can be avoided. Consequently, if θ is very well approximated by rationals p_n/q_n, then $\{d_j(x)\theta\}$ must be dense modulo 1.

Lemma 5.1. *Let q be prime and n a natural number. Then for $k = 0, \ldots, q^n - 2$ we have the following formula:*

$$C(q^n - 2, k) \cong_q (-1)^k (k + 1). \tag{1}$$

Moreover, for natural numbers k and p satisfying $0 \leqslant p \leqslant q - 1$ and $0 \leqslant k \leqslant q^n - 2q - 1$ the set

$$\{i : k \leqslant i \leqslant k + 2q - 1, C(q^n - 2, i) \cong_q p\} \tag{2}$$

has exactly two elements.

Proof. First we derive the formula inductively. The primary case in trivial: $C(q^n - 2, 0) = 1 = (-1)^0(1)$. Assume the formula holds for $k = l - 1$. Thus for $k = l$ we have

$$\begin{aligned} C(q^n - 2, l) &= C(q^n - 1, l) - C(q^n - 2, l - 1) \\ &\cong_q (-1)^l - (-1)^{l-1} l \\ &= (-1)^l [1 + l]. \end{aligned}$$

Now for the second part of the lemma, we note that our formula gives the following:

$$C(q^n - 2, i + 2) \cong_q \begin{cases} C(q^n - 2, i) + 2 & \text{if } i \text{ even} \\ C(q^n - 2, i) - 2 & \text{if } i \text{ odd.} \end{cases}$$

Hence if $q = 2$ we obtain $C(q^n - 2, k) \cong_q 1$ for k even and $C(q^n - 2, k) \cong_q 0$ for k odd. For $q \neq 2$ we have that each of the maps $j \mapsto C(q^n - 2, k + 2j) \bmod q$ and $j \mapsto C(q^n - 2, k + 1 + 2j) \bmod q$ gives a bijection of $\{0, \ldots, q - 1\}$. □

Lemma 5.2. *Suppose that for each $n \in \mathbb{N}$ we have a unimodal distribution f_n on the set $\{0, \ldots, q^n - 2\}$. For each p satisfying $0 \leqslant p \leqslant q - 1$, let*

$$M_p = \{m : 0 \leqslant m \leqslant q^n - 2, C(q^n - 2, m) \cong_q p\}.$$

If $\lim_{n \to \infty} \max\{f_n(m) : 0 \leqslant m \leqslant q^n - 2\} = 0$, then

$$\lim_{n \to \infty} \sum_{m \in M_p} f_n(m) = \frac{1}{q}.$$

Proof. We will show that

$$\lim_{n\to\infty}\left(\sum_{m\in M_p} f_n(m) - \sum_{m\in M_r} f_n(m)\right) = 0$$

for all p and r. Without loss of generality, assume that

$$\sum_{m\in M_p} f_n(m) \geqslant \sum_{m\in M_r} f_n(m).$$

Partition $\{0,\dots,q^n-2\}$ into subintervals of $2q$ consecutive numbers with one remaining subinterval of at most $2q$ consecutive numbers. Discard the subinterval which contains the peak of the distribution f_n, as well as its two adjacent subintervals, from the set M_p. Call the remaining set M_p^*. Now for each $m^* \in M_p^*$ there exists $m \in M_r$ in the next interval of length $2q$ towards the peak of f_n such that $f_n(m) \geqslant f_n(m^*)$. Therefore

$$\sum_{m\in M_r} f_n(m) \geqslant \sum_{m\in M_p^*} f_n(m) \geqslant \sum_{m\in M_p} f_n(m) - 6\max\{f_n(m) : 0 \leqslant m \leqslant q^n-2\}. \qquad \square$$

Let $F_n(q,p)$ be the set of paths which pass through a vertex (q^n-2,k) satisfying $C(q^n-2,k) \cong_q p$. Lemma 5.2 implies that the conditional probability of the set $F_n(q,p)$, given that the path passes through a fixed vertex, converges to $1/q$ as $n\to\infty$. Therefore for each $m \in \mathbb{N}$

$$\lim_{n\to\infty} \mu_\alpha(F_n(q,p)\cap F_m(q,r)) = \frac{1}{q}\mu_\alpha(F_m(q,r)). \tag{3}$$

A standard Hilbert space argument of A. RÉNYI [8] implies that $F_n(q,p), n = 1,2,\dots$ is a mixing sequence of sets. In particular we have Lemma 5.3, which says $F_n(q,p)$ 'sweeps out'. Finally we prove Theorem 5.4 using an approximation technique similar to that used in the previous section.

Lemma 5.3. *For $0 \leqslant p < q$ with q prime and $0 < \alpha < 1$.*

$$\mu_\alpha\left(\bigcup_{n=1}^{\infty} F_n(q,p)\right) = 1.$$

Theorem 5.4. *There exist a dense G_δ set $\Lambda \subset [0,1)$ and a set of full μ_α-measure $Y \subset X$ so that for each $\theta \in \Lambda$ and $x \in Y$ the set $\{e^{2\pi i d_j(x)\theta} : j \in \mathbb{N}\}$ is dense (but not uniformly distributed) in S^1.*

Proof. Let $\{\varepsilon_n\}$ be a sequence of positive real numbers satisfying $\sum_{n=1}^{\infty}\varepsilon_n < \infty$, and let $\{q_n\}$ be a sequence of primes increasing to ∞. We will produce sequences R_n of natural numbers and $\delta_n > 0$ so that if

$$Y_n = \bigcap_{p=0}^{q_n-1}\bigcup_{j=0}^{R_n} F_j(q_n,p),$$

$$\Lambda_n = \{\theta \in [0,1) : \text{there exists } p_n = 0, 1, \ldots, q_n - 1 \text{ with } |\theta - \tfrac{p_n}{q_n}| < \delta_n\},$$

$$\Lambda = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} \Lambda_n, \quad \text{and} \quad Y = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} Y_n,$$

then $\{e^{2\pi i d_j(x)\theta} : j \in \mathbb{N}\}$ is dense but not uniformly distributed for all $\theta \in \Lambda$ and all $x \in Y$.

For each n, by Lemma 5.3 we may choose R_n so that $\mu_\alpha(Y_n) > 1 - \varepsilon_n$. Then choose

$$\delta_n < \frac{1}{nC(q_n^{R_n} - 2, (q_n^{R_n} - 2)/2)}.$$

As before, if $\theta \in \Lambda$ and $x \in Y$, then we can find arbitrarily large n such that $\theta \in \Lambda_n$ and $x \in Y_n$. Now θ is very well approximated by a rational p_n/q_n, and as p runs through the congruence classes modulo q_n, the points pp_n/q_n are $1/q_n$-dense modulo 1. Further, for each congruence class p modulo q_n there is $j = 1, \ldots, R_n$ such that at level $s = q_n^j$ the path x has its binomial coefficient $d_s(x)$ hit that congruence class. Since δ_n has been chosen so small that all the points $d_s(x)\theta$ under consideration are very close to the points $d_s(x)p_n/q_n$, and the latter are $1/q_n$-dense, we are done. □

6. Questions and Conjectures

1. *Conjecture* [10, 6]: For each Bernoulli measure μ_α, the Pascal adic transformation T is weakly mixing. This would follow if one could show that $\lambda^{d_n(x)} \to 1$ for a.e. x with respect to μ_α implies $\lambda = 1$.

2. *Conjecture*: If there is a path x such that $\lambda^{d_n(x)} \to 1$, then $\lambda = 1$.

3. Does there exist any λ in the unit circle such that $\{\lambda^{d_n(x)}\}$ is uniformly distributed in the circle *for every* x (except for the two paths down the edges)? For such a λ, the skew-product transformation

$$S(z_1, z_2, z_3, \ldots) = (\lambda z_1, z_1 z_2, z_2 z_3, \ldots)$$

on the infinite torus, known to be uniquely ergodic by results of Weyl [13, 14], Furstenberg [1], Hahn [3], and Postnikov [7], would have the very strong property that we would see a uniformly distributed sequence $\{(S^j z)_k\}$ not only when we looked in a fixed coordinate k at the orbit of a point z, but also when we allowed our view to shift one place to the right from time to time: $\{(S^j z)_{k_j}\}$ would be uniformly distributed for each z and each choice of $\{k_j\} \subset \mathbb{N}$ with $k_{j+1} - k_j \in \{0, 1\}$ for each j. (The λ that we construct above are at another extreme from this property.)

4. From another theorem of Weyl and Tonelli's Theorem it follows that for almost every λ the sequence $\{\lambda^{d_n(x)}\}$ is uniformly distributed for a.e. x, with respect to each Bernoulli measure μ_α. For which λ does this hold? Similarly, what paths x have the property that this sequence is uniformly distributed for each λ that is not a root of unity? (By Weyl's Theorem, this is the case for each path x that is eventually diagonal.)

5. Studies like these on divisibility of binomial coefficients by primes suggest questions on simultaneous divisibility by several primes. For example, thinking about the central path in Pascal's triangle and divisibility by 2 and 3 leads to the following *Conjecture*: The only solutions in nonnegative integers r and distinct $s_1, \ldots, s_m$ of an equation

$$2^r = 3^{s_1} + \cdots + 3^{s_m}$$

are $1 = 1, 4 = 1 + 3$, and $256 = 1 + 3 + 9 + 243$. We thank Charles Giffen for pointing out that this conjecture was already made by Erdös–see [2]. More generally, if for a prime p we define an integer Cantor set $H(p)$ to consist of all those expansions base p with coefficients in the interval $[0, p/2]$ (or subject to some other restriction), is $H(p_1) \cap \cdots \cap H(p_n)$ typically finite?

References

[1] Furstenberg H (1961) Strict ergodicity and transformations of the torus. Amer J Math **83**: 573–601
[2] Guy RK (1994) Unsolved Problems in Number Theory, 2nd edn, p 88. Berlin Heidelberg New York: Springer
[3] Hahn FJ (1963) On affine transformations of compact abelian groups. Amer J Math **85**: 428–446
[4] Kummer EE (1852) Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen. J für Math **44**: 115–116
[5] Lucas E (1878) Théorie des fonctions numériques simplement périodiques. Amer J Math **1**: 184–240
[6] Petersen K, Schmidt K (1997) Symmetric Gibbs measures. Trans Amer Math Soc **349**: 2775–2811
[7] Postnikov AG (1966) Ergodic problems in the theory of congruences and of Diophantine approximations. Proc Steklov Inst Math **82**: 3–112
[8] Rényi A (1958) On mixing sequences of sets. Acad Sci Hung **9**: 215–228
[9] Vershik AM (1974) Description of invariant measures for actions of some infinite groups. Dokl Akad Nauk SSSR **218**: 749–752; (1974) Soviet Math Dokl **15**: 1396–1400
[10] Vershik AM (private communication, 1991)
[11] Vershik AM (1981) Uniform algebraic approximation of shift and multiplication operators. Dokl Akad Nauk SSSR **259**: 526–529; Soviet Math Dokl **24**: 97–100
[12] Vershik AM, Livshitz AN (1992) Adic models of ergodic transformations, spectral theory, substitutions, and related topics. Adv Soviet Math **9**: 185–204
[13] Weyl H (1914) Über ein Problem aus dem Gebiete der diophantischen Approximationen. Nachr Ges Wiss Göttingen 234–244
[14] Weyl H (1916) Über die Gleichverteilung von Zahlen mod. Eins. Math Ann **77**: 313–352

T. A. Adams
Department of Mathematics
CB 3250 Phillips Hall
University of North Carolina
Chapel Hill, NC 27599-3250
USA
e-mail: tadams@math.ohio-state.edu

K. E. Petersen
Department of Mathematics
CB 3250 Phillips Hall
University of North Carolina
Chapel Hill, NC 27599-3250
USA
e-mail: petersen@math.unc.edu

Current address:
Department of Mathematics
Ohio State University
Columbus, OH 43210,
USA

Mh. Math. 125, 279–292 (1998)

Monatshefte für
Mathematik

Printed in Austria

Quadratic Equations over Finite Fields and Class Numbers of Real Quadratic Fields

By

Takashi Agoh and **Toshiaki Shoji,** Tokyo

(Received 28 May 1996; in revised form 27 January 1997)

Abstract. Let p be an odd prime and $\mathbb{F}_p$ the finite field with p elements. In the present paper we shall investigate the number of points of certain quadratic hypersurfaces in the vector space $\mathbb{F}_p^n$ and derive explicit formulas for them. In addition, we shall show that the class number of the real quadratic field $\mathbb{Q}(\sqrt{p})$ (where $p \equiv 1 \pmod 4$) over the field $\mathbb{Q}$ of rational numbers can be expressed by means of these formulas.

1. Introduction

1.1. Let $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ be the set of natural numbers, the ring of integers, the field of rational numbers, and the field of real numbers, respectively. For an odd prime p, let $\mathbb{F}_p$ be the finite field with p elements and $V = \mathbb{F}_p^n$ the vector space over $\mathbb{F}_p$. We denote the point $(x_1, x_2, \cdots, x_n) \in V$ by the single letter x.

For $\alpha \in \mathbb{F}_p$, consider the following quadratic surface S_α in V and its open subset S_α^0:

$$S_\alpha = \{x \in V \mid \sum_{i=1}^{n} x_i^2 = \alpha\},$$

$$S_\alpha^0 = \{x \in S_\alpha \mid x_i \neq 0, x_i \neq \pm x_j (i \neq j, 1 \leqslant i, j \leqslant n)\}.$$

If we denote by $\mathscr{A}$ the set of hyperplanes in V given by the equations

$$x_i = 0,\ x_i = x_j,\ x_i = -x_j \quad (i \neq j, 1 \leqslant i, j \leqslant n), \tag{1.1.1}$$

then the set S_α^0 can be expressed as $S_\alpha^0 = S_\alpha - \bigcup_{H \in \mathscr{A}} (H \cap S_\alpha)$. Let $L = L(\mathscr{A})$ be the lattice associated with $\mathscr{A}$, i.e., L is the set consisting of various intersections of hyperplanes in $\mathscr{A}$. If $\mathscr{A}_0$ is the set of hyperplanes in a real vector space $\mathbb{R}^n$ given by (1.1.1), then the lattice $L(\mathscr{A})$ is isomorphic to the lattice $L(\mathscr{A}_0)$ whenever $p \neq 2$.

Let W be the subgroup of the orthogonal group $O_n(\mathbb{R})$ of degree n generated by reflections with respect to $H \in \mathscr{A}_0$. Then W is isomorphic to the Weyl group of type B_n.

1991 Mathematics Subject Classification: 05A19, 05E15, 11E04, 11R11, 11R29, 20B30
Key words: Quadratic forms over finite fields, Weyl groups, hyperplane complements, partitions, combinatorial identities, class numbers, real quadratic fields

Now the Möbius function $\mu : L \times L \to \mathbb{Z}$ is defined recursively by the following conditions: $\mu(X,X) = 1, \sum_{X \leqslant Z \leqslant Y} \mu(X,Z) = 0$ if $X < Y$ and $\mu(X,Y) = 0$ otherwise. Here for each $X, Y \in L$, we write $X < Y$ if the corresponding subspaces $X, Y \subseteq V$ satisfy $Y \subset X$. Then by the Möbius inversion formula we obtain

$$|S_\alpha^0| = \sum_{X \in L} \mu(V,X)|S_\alpha \cap X|, \tag{1.1.2}$$

where $|M|$ means the number of elements of a set M.

1.2. Let m be a square-free integer with $m \neq -1, -3$ and d be the discriminant of the quadratic field $\mathbb{Q}(\sqrt{m})$. If $L(s,\chi)$ is the L-function attached to the Dirichlet character χ of conductor $|d|$, then the class number h of $\mathbb{Q}(\sqrt{m})$ can be given by

$$h = \begin{cases} \dfrac{\sqrt{d}}{2\log\varepsilon} L(1,\chi) & \text{for } m > 0, \\ \dfrac{\sqrt{|d|}}{\pi} L(1,\chi) & \text{for } m < 0, \end{cases}$$

where ε is the fundamental unit of $\mathbb{Q}(\sqrt{m})$.

By using this formula we can deduce that if $m = p$ is an odd prime with $p \equiv 1 \pmod 4$, then

$$\varepsilon^{2h} = \prod_{k=1}^{p-1} (1 - \zeta^k)^{-\chi(k)}, \tag{1.2.1}$$

where ζ is a primitive p-th root of unity and χ is the Legendre symbol in this case.

1.3. In the present paper we shall investigate the number of points of the quadratic surface S_α and its open part S_α^0, and derive the explicit formulas for them by using (1.1.2) together with some combinatorial results due to Orlik and Solomon. These formulas are described in terms of elements of the Weyl group W of type B_n. In addition, using (1.2.1) we shall show that the class number of $\mathbb{Q}(\sqrt{p})$ for $p \equiv 1 \pmod 4$ can be explicitly expressed by means of these formulas.

We note that the class number formulas of $\mathbb{Q}(\sqrt{p})$ deduced in Section 5 may be not so efficient as an algorithm for real computing of h compared with, e.g., the well-known continued fraction method. However, we emphasize that our aim is to derive explicit formulas for h by means of the number of points of the quadratic surface S_α and its open subset S_α^0.

2. Some Calculations on Quadratic Surfaces

2.1. For a positive integer t, we consider a t-tuple $\lambda = (\lambda_1, \ldots, \lambda_t) \in \mathbb{N}^t$. Let $\mathbb{F}_q$ be a finite field consisting of q elements, with q odd. For each $\alpha \in \mathbb{F}_q$, we define a quadratic surface $S_\alpha(\lambda) \subset \mathbb{F}_q^t$ as follows:

$$S_\alpha(\lambda) = \{x \in \mathbb{F}_q^t \mid \sum_{i=1}^t \lambda_i x_i^2 = \alpha\}. \tag{2.1.1}$$

Let $\mathbb{F}_q^*$ be the multiplicative group of $\mathbb{F}_q$ and let $\theta : \mathbb{F}_p^* \to \{\pm 1\}$ be the Legendre symbol defined by $\theta(x) = x^{(q-1)/2} (x \in \mathbb{F}_q^*)$. As usual, we extend θ to the function on $\mathbb{F}_q$ by setting $\theta(0) = 0$. Also, we define $\theta(\lambda)$ for $\lambda \in \mathbb{N}^t$ by $\theta(\lambda) = \prod_{i=1}^t \theta(\lambda_i)$.

Now the following proposition describes the cardinality of $S_\alpha(\lambda)$.

Proposition 2.2. *Let $S_\alpha(\lambda)$ be as above, and let $\kappa = \theta(-1) = (-1)^{(q-1)/2}$. Then*

$$|S_\alpha(\lambda)| = \begin{cases} q^{t-1} + \theta(\alpha)\theta(\lambda)(\kappa q)^{(t-1)/2} & \text{if } t \text{ is odd,} \\ q^{t-1} - \dfrac{1}{q}\theta(\lambda)(\kappa q)^{t/2} & \text{if } t \text{ is even and } \alpha \neq 0, \\ q^{t-1} + \dfrac{q-1}{q}\theta(\lambda)(\kappa q)^{t/2} & \text{if } t \text{ is even and } \alpha = 0. \end{cases}$$

This result is already known (see LIDL and NIEDERREITER [4], Theorems 6.26 and 6.27). Therefore we omit the proof, and just indicate below the outline of the proof for the sake of completeness.

Let Q be a quadratic form on a vector space V over $\mathbb{F}_q$, and for a fixed $\alpha \in \mathbb{F}_q$ define a quadratic space $X_\alpha(Q)$ by

$$X_\alpha(Q) = \{x \in V | Q(x) = \alpha\}.$$

We compute the cardinality of $X_\alpha(Q)$. First consider the special case where $Q(x) = x_1^2 - \delta x_2^2$ with $\delta \in \mathbb{F}_q^*$, which case is easily treated. Next consider the case where Q is of the canonical form with respect to the Witt decomposition. The cardinality of $X_\alpha(Q)$ in this case can be computed by making use of the natural filtration of $X_\alpha(Q)$ together with the above computation. Then the general case follows from the fact that two quadratic forms Q and Q' are equivalent if and only if $d_Q \equiv d_{Q'} \bmod (\mathbb{F}_q^*)^2$, which is also equivalent to the condition that $\theta(d_Q) = \theta(d_{Q'})$ (here, d_Q denotes the discriminant of Q).

3. The Explicit Formulas for $|S_\alpha^0|$

3.1. In this section we shall give an explicit formula for the cardinality of the set S_α^0 given in Section 1. As in Section 1, we return to the setting of the prime field $\mathbb{F}_p$. But note that the results in this section hold without change if $\mathbb{F}_p$ is replaced by a finite field $\mathbb{F}_q$. Before starting the computation, we summarize some properties of the Weyl group W. So, let W be the Weyl group of type B_n, realized as the reflection subgroup of $O_n(\mathbb{R})$ as given in Section 1. We fix the standard basis $e_1, e_2, \ldots, e_n$ of $\mathbb{R}^n$. Then each $w \in W$ gives a permutation of the basis (with signature), $w : e_i \to \pm e_{\sigma(i)}$. The set of conjugacy classes of W is in one to one correspondence with the set $\mathscr{P}_n$ of all the pairs of partitions $\nu = (\lambda; \mu)$ of n, where

$$\begin{aligned} &\lambda : \lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_t > 0, \\ &\mu : \mu_1 \geqq \mu_2 \geqq \cdots \geqq \mu_s > 0 \end{aligned} \tag{3.1.1}$$

for positive integers λ_i, μ_j such that $\sum \lambda_i + \sum \mu_j = n$. The correspondence is given as follows: for each sequence I of m distinct numbers $i_1, i_2, \ldots, i_m$ such that

$1 \leqslant i_k \leqslant n$, we consider a cyclic permutation (with signature) $w_I : e_{i_1} \to \pm e_{i_2} \to \to \cdots \to \pm e_{i_m} \to \pm e_{i_1}$. Then $w_I \in W$, and we say that w_I is a positive cycle (resp. a negative cycle) of length m if $w_I^m(e_{i_1}) = e_{i_1}$ (resp. $w_I^m(e_{i_1}) = -e_{i_1}$). Now any element w of W can be expressed as a product of mutually commuting positive cycles of length λ_i and negative cycles of length μ_j such that $\sum \lambda_i + \sum \mu_j = n$, and the corresponding $\nu = (\lambda; \mu) \in \mathscr{P}_n$ (after rearranging so that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots$, $\mu_1 \geqslant \mu_2 \geqslant \cdots$) gives the required bijection. For a given $\nu \in \mathscr{P}_n$ as in (3.1.1), we denote by w_ν (a representative of) the corresponding conjugacy class in W. Then it is easy to see from the above description that

$$\det(w_\nu) = \prod_{i=1}^{t}(-1)^{\lambda_i+1} \cdot \prod_{j=1}^{s}(-1)^{\mu_j} = (-1)^{n+t}. \tag{3.1.2}$$

For a given $w_\nu \in W$, let c_ν be the number of elements in the conjugacy class containing w_ν. Then c_ν is explicitly given as follows; we denote by $\lambda = (\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_t) = (1^{m_1}, 2^{m_2}, \ldots, i^{m_i}, \ldots)$, i.e., $m_i = \#\{j \in \mathbb{N} \mid \lambda_j = j\}$, and similarly, $\mu = (1^{n_1}, 2^{n_2}, \ldots)$.

Then we have

$$c_\nu = 2^n n! \left(\prod_{i=1}^{t} 2\lambda_i \cdot \prod_{i \geqslant 1} m_i!\right)^{-1} \left(\prod_{j=1}^{s} 2\mu_j \cdot \prod_{j \geqslant 1} n_j!\right)^{-1}. \tag{3.1.3}$$

3.2. For a given $\nu \in \mathscr{P}_n$, we choose a specific representative w_ν as follows: for each $i\,(1 \leqslant i \leqslant t)$ and $j\,(1 \leqslant j \leqslant s)$, w_ν maps $e_{k_i+1} \to e_{k_i+2} \to \cdots \to e_{k_i+\lambda_i} \to e_{k_i+1}$, and $e_{l_j+1} \to e_{l_j+2} \to \cdots \to e_{l_j+\mu_j} \to -e_{l_j+1}$, where $k_i = \lambda_1 + \lambda_2 + \cdots + \lambda_{i-1}$ and $l_j = \sum_{i=1}^{t} \lambda_i + (\mu_1 + \cdots + \mu_{j-1})$. Then the fixed point subspace $\mathrm{Ker}(w_\nu - 1)$ of w_ν in $\mathbb{F}_p^n$ can be expressed as follows:

$$\begin{aligned}\mathrm{Ker}(w_\nu - 1) = \{x \in V \mid\, & x_{k_i+1} = x_{k_i+2} = \cdots = x_{k_i+\lambda_i} (1 \leqslant i \leqslant t), \\ & x_{l_1+1} = x_{l_1+2} = \cdots = x_n = 0\}.\end{aligned} \tag{3.2.1}$$

In particular, we see that $\dim \mathrm{Ker}(w_\nu - 1) = t$.

3.3. For each $w \in W$, let $n(w)$ be the smallest number m such that w is expressed as a product of m reflections of W. Then it is known that

$$n(w) = \mathrm{rank}(w - 1) = \dim V - \dim \mathrm{Ker}(w - 1)$$

(cf. [7]). Hence, by 3.2, $n(w) = n - t$ for $w = w_\nu \in W$. Now the following formula is a special case of a formula due to L. SOLOMON ([8]), which is also valid for general Weyl groups.

Let q be an indeterminate. Assume that W is the Weyl group of type B_n. Then we have

$$\sum_{w \in W} q^{n(w)} = \prod_{i=1}^{n}(1 + (2i - 1)q). \tag{3.3.1}$$

For each $\nu = (\lambda; \mu) \in \mathscr{P}_n$, we denote by $t = t(\nu)$ the number of λ_i appearing in the partition $\lambda = (\lambda_i)$ (see (3.1.1)). Then if we note that $n(w) = n - t(\nu)$

for $w = w_\nu$, the following formula is easily deduced from (3.3.1):

$$\sum_{\nu \in \mathscr{P}_n} (-1)^{n+t(\nu)} c_\nu q^{t(\nu)} = \prod_{i=1}^{n} (q - (2i-1)). \tag{3.3.2}$$

It follows from (3.3.2) that

$$\sum_{\substack{\nu \in \mathscr{P}_n \\ t(\nu)=0}} c_\nu = \prod_{i=1}^{n} (2i-1) = \frac{(2n)!}{2^n n!}. \tag{3.3.3}$$

In fact, if we denote by Z the left-hand side of (3.3.3), then (3.3.2) implies that

$$\sum_{\substack{\nu \in \mathscr{P}_n \\ t(\nu)>0}} c_\nu q^{t(\nu)-1} = \frac{1}{q} \left\{ \prod_{i=1}^{n} (q - (2i-1)) - (-1)^n Z \right\}.$$

The left-hand side of this equality is a polynomial in q. Hence, by considering the coefficient of q^{-1} in the right hand side, we obtain (3.3.3).

3.4. Returning to the original setting in Section 1, we shall give an explicit formula for $|S_\alpha^0|$. We want to compute it by making use of (1.1.2). For this, we need to evaluate the values $\mu(V, X)$ for each $X \in L$, which will be given by the following result of Orlik and Solomon.

Proposition 3.5 (Orlik-Solomon [7]). *For $X \in L$, let*

$$F_X = \{w \in W | \mathrm{Ker}(w-1) = X\}.$$

Then we have

$$\mu(V, X) = \sum_{w \in F_x} \det(w).$$

3.6. Using Proposition 3.5, we rewrite (1.1.2) as

$$\begin{aligned} |S_\alpha^0| &= \sum_{X \in L} \left(\sum_{w \in F_X} \det(w) \right) |S_\alpha \cap X| \\ &= \sum_{w \in W} \det(w) |S_\alpha \cap \mathrm{Ker}(w-1)|. \end{aligned} \tag{3.6.1}$$

Note that each term in the sum of the last formula depends only on the conjugacy class of W. Hence, by the discussion in 3.1 the sum is taken over all the pairs of partitions $\nu = (\lambda; \mu) \in \mathscr{P}_n$. Then using (3.1.2), we see that

$$|S_\alpha^0| = \sum_{\nu \in \mathscr{P}_n} (-1)^{n+t(\nu)} c_\nu |S_\alpha \cap \mathrm{Ker}(w_\nu - 1)|, \tag{3.6.2}$$

where c_ν is as in (3.1.3). Here we may assume that w_ν is as in 3.2. Then (3.2.1) implies that the set $S_\alpha \cap \mathrm{Ker}(w_\nu - 1)$ can be identified with the set $S_\alpha(\lambda)$ given in (2.1.1) if $t = t(\nu) > 0$. In this case $|S_\alpha(\lambda)|$ is described in Proposition 2.2. If

$t(\nu)=0$, then $S_\alpha(\lambda)=\{0\}$ (resp. $S_\alpha(\lambda)=\emptyset$) in the case where $\alpha=0$ (resp. $\alpha\neq 0$), and so $|S_\alpha(\lambda)|=1$ (resp. $|S_\alpha(\lambda)|=0$), respectively.

Now, by making use of Proposition 2.2, we can express (3.6.2) in a more explicit form. Let

$$\begin{aligned} A &= (-1)^n \sum_{\substack{\nu\in\mathscr{P}_n \\ t(\nu)>0}} c_\nu (-p)^{t(\nu)}, \\ B &= (-1)^n \sum_{\substack{\nu\in\mathscr{P}_n \\ t(\nu)\text{odd}}} c_\nu \theta(\lambda)(\kappa p)^{(t(\nu)-1)/2}, \\ C &= (-1)^n \sum_{\substack{\nu\in\mathscr{P}_n \\ t(\nu)>0\text{ even}}} c_\nu \theta(\lambda)(\kappa p)^{t(\nu)/2}, \\ D &= (-1)^n \sum_{\substack{\nu\in\mathscr{P}_n \\ t(\nu)=0}} c_\nu. \end{aligned} \tag{3.6.3}$$

Noting that the restriction of $\sum_{i=1}^n x_i^2$ to $\mathrm{Ker}(w_\nu-1)$ is $\lambda_1 x_{\lambda_1}^2+\lambda_2 x_{\lambda_1+\lambda_2}^2+\cdots+\lambda_t X_{\lambda_1+\cdots+\lambda_t}^2$, we see that

$$|S_\alpha^0| = \begin{cases} \dfrac{1}{p}A-\theta(\alpha)B-\dfrac{1}{p}C & \text{if } \alpha\neq 0, \\ \dfrac{1}{p}A+\dfrac{p-1}{p}C+D & \text{if } \alpha=0. \end{cases} \tag{3.6.4}$$

Note, by (3.3.2) and (3.3.3), that A and D can be written as

$$\begin{aligned} A &= \prod_{i=1}^n (p-(2i-1)) - (-1)^n \prod_{i=1}^n (2i-1), \\ D &= (-1)^n \frac{(2n)!}{2^n n!} \end{aligned}$$

and B, C are independent of α. If we put $A_0=\prod_{i=1}^n (p-(2i-1))$, we have $A=A_0-D$.

Moreover, by setting $E=C+D$ and modifying the formula (3.6.4), we get the following theorem:

Theorem 3.7. *Let* $\kappa=(-1)^{(p-1)/2}$ *and* $\theta:\mathbb{F}_p^*\to\{\pm 1\}$ *be the Legendre symbol. Use the same notations as in* 3.1, 3.2 *and* 3.3 *set*

$$\begin{aligned} A_0 &= \prod_{i=1}^n (p-(2i-1)), \\ B &= (-1)^n \sum_{\substack{\nu\in\mathscr{P}_n \\ t(\nu)\text{ odd}}} c_\nu \theta(\lambda)(\kappa p)^{(t(\nu)-1)/2}, \\ E &= C+D = (-1)^n \sum_{\substack{\nu\in\mathscr{P}_n \\ t(\nu)\geqslant 0\text{ even}}} c_\nu \theta(\lambda)(\kappa p)^{t(\nu)/2}. \end{aligned}$$

Then we have

$$|S_\alpha^0| = \begin{cases} \dfrac{1}{p}A_0 - \theta(\alpha)B - \dfrac{1}{p}E & \text{if } \alpha \neq 0, \\ \dfrac{1}{p}A_0 + \dfrac{p-1}{p}E & \text{if } \alpha = 0. \end{cases}$$

4. The Identification with a Result of Maohua Le

4.1. In the case where $\alpha = 0$, the number of solutions $|S_\alpha^0|$ was computed by M.-H. Le in [5] by a different method. In this section, we give the identification of his result with ours. First, we explain his result. Let $\Delta = \sqrt{\kappa p}$, hence $d = \Delta^2$ is the discriminant of $\mathbb{Q}(\Delta)$ as mentioned in Section 1, and let

$$s_k = \frac{1}{2}(-1 + \theta(k)\Delta), 1 \leqslant k \leqslant \frac{p-1}{2}. \tag{4.1.1}$$

We set, for each integer $n \geqslant 1$,

$$\sigma_n = \sum_{\substack{n_i \geqslant 0, i=1,2,\\ n_1+2n_2+\cdots=n}} (-1)^{n_2+n_4+\cdots} \prod_{i=1}^{n} \frac{s_i^{n_i}}{i^{n_i} n_i!}, \tag{4.1.2}$$

and define integers A_n, B_n by the condition that $\sigma_n = (A_n + B_n\Delta)/2$. (To prove his theorem, Le defines σ_n by a recursive formula involving s_k. However in the last remark of his paper, Le also gives an explicit formula for σ_n which is nothing but (4.1.2)).

Let

$$N_n = \frac{1}{p}\binom{(p-1)/2}{n} + \frac{p-1}{2p}A_n. \tag{4.1.3}$$

Then Le's result is stated as follows.

Proposition 4.2 (Maohua Le [5]). *Assume that* $\alpha = 0$. *Then*

$$N_n = \frac{1}{2^n n!}|S_\alpha^0|.$$

Proof. We shall deduce the proposition from our result. It follows from Theorem 3.7 that $|S_\alpha^0|$ can be written as

$$|S_\alpha^0| = \frac{1}{p}A_0 + \frac{p-1}{p}E$$

with A_0, E as in Theorem 3.7. It is easy to see that

$$\frac{1}{2^n n!}A_0 = \binom{(p-1)/2}{n}.$$

So we have only to show that

$$\frac{1}{2^n n!}E = \frac{1}{2}A_n. \tag{4.2.1}$$

We now show (4.2.1). Using (4.1.1), we have

$$\prod_{i=1}^{n} s_i^{n_i} = \prod_{i=1}^{n}\left\{\frac{1}{2^{n_i}}\sum_{0\leqslant m_i\leqslant n_i}(-1)^{n_i-m_i}(\theta(i)\Delta)^{m_i}\binom{n_i}{m_i}\right\}$$
$$= \sum_{\substack{m_1,m_2,\cdots\\ 0\leqslant m_i\leqslant n_i}}(-1)^{\sum n_i-\sum m_i}\Delta^{\sum m_i}\prod_{i=1}^{n}\left\{\frac{1}{2^{n_i}}\binom{n_i}{m_i}\theta(i)^{m_i}\right\}.$$

In the last expression, only the terms such that $\sum m_i$ is even give the contribution for A_n. Now, let $\mathscr{P}_n^0$ be the set of partitions $\xi = (1^{n_1}, 2^{n_2}, \cdots)$ of n, i.e., $n_i \geqslant 0$ are integers satisfying $n_1 + 2n_2 + \cdots = n$. More generally, for a sequence $\lambda = (1^{m_1}, 2^{m_2}, \cdots)$ with integers $m_i \geqslant 0$, we denote $\lambda \leqslant \xi$ if $m_i \leqslant n_i$ for all $i \geqslant 1$. If such λ is written as $\lambda = (\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_t)$, we put $t = t(\lambda)$. Under this notation, (4.1.2) implies that

$$\frac{1}{2}A_n = \sum_{\xi\in\mathscr{P}_n^0}(-1)^{n_2+n_4\cdots}\sum_{\substack{\lambda\leqslant\xi\\ t(\lambda)\text{ even}}}(-1)^{t(\xi)}\theta(\lambda)\Delta^{t(\lambda)}\prod_{i=1}^{n}\binom{n_i}{m_i}\frac{1}{(2i)^{n_i}n_i!}, \tag{4.2.2}$$

where we used the notations $\xi = (1^{n_1}, 2^{n_2}, \cdots)$ and $\lambda = (1^{m_1}, 2^{m_2}, \cdots)$ in the sum.

Now, since $\Delta^2 = \kappa p$, and since $t(\lambda)$ is even, we see that $\Delta^{t(\lambda)} = (\kappa p)^{t(\lambda)/2}$. Moreover, for the above $\xi = (1^{n_1}, 2^{n_2}\cdots)$ and $\lambda = (1^{m_1}, 2^{m_2}, \cdots)$ with $\lambda \leqslant \xi$, if we put $\mu = (1^{n_1-m_1}, 2^{n_2-m_2}, \cdots)$, then $\nu = (\lambda; \mu)$ gives rise to an element in $\mathscr{P}_n$ (see 3.1). The corresponding c_ν is described by the formula (3.1.3). In this case, we obtain

$$c_\nu = 2^n n!\prod_{i=1}^{n}\binom{n_i}{m_i}\frac{1}{(2i)^{n_i}n_i!}.$$

On the other hand, we note that

$$(-1)^{n_2+n_4+\cdots}\cdot(-1)^{t(\xi)} = (-1)^n.$$

In fact, since $n = \sum_i in_i \equiv \sum_{i\,\text{odd}} n_i \pmod 2$, we see that

$$t(\xi) + (n_2 + n_4 + \cdots) = \sum_i n_i + (n_2 + n_4 + \cdots)$$
$$\equiv \sum_{i\text{ odd}} n_i$$
$$\equiv n \pmod 2.$$

Combining these results, it is easy to see that the right hand side of (4.2.2) coincides with $(2^n n!)^{-1}E$. This shows (4.2.1) and so proves the proposition. □

5. The Class Number of $\mathbb{Q}(\sqrt{p})$

5.1. In this section we deal with the class number h of the real quadratic field $\mathbb{Q}(\sqrt{p})$. In the following, we assume that p is an odd prime with $p \equiv 1 \pmod 4$.

Recall the formula (1.2.1) from Section 1 and write it as

$$\varepsilon^{2h} = \frac{\prod_s (1-\zeta^s)}{\prod_r (1-\zeta^r)}, \tag{5.1.1}$$

where r, s run over the quadratic residues and nonresidues of p with $1 \leqslant r, s \leqslant p-1$, respectively.

For simplicity, set $\Pi_n = \prod_n (1-\zeta^n)$ for $n = r, s$. Then it is easily seen that

Lemma 5.2. *There exist integers a, b and c such that*

$$\Pi_r = a + b\sum_r \zeta^r + c\sum_s \zeta^s \quad \text{and} \quad \Pi_s = a + b\sum_s \zeta^s + c\sum_r \zeta^r,$$

where the sums $\sum_r, \sum_s$ are taken over all integers r, s in the interval $(0, p)$ such that $\theta(r) = 1, \theta(s) = -1$, respectively.

Here, we note that the above integers a, b and c satisfy

$$a + \frac{p-1}{2}(b+c) = 0 \tag{5.2.1}$$

and

$$(b-c)^2 + 4 = p(b+c)^2. \tag{5.2.2}$$

Since $\Pi_r \cdot \Pi_s = p, \Pi_r + \Pi_s = 2a - (b+c) = -(b+c)p > 0$ and $\Pi_r - \Pi_s = (b-c)\sqrt{p} < 0$, it follows from (5.1.1) that

$$\begin{aligned}\varepsilon^h = \frac{\Pi_s}{\sqrt{p}} &= \frac{1}{2}\left\{\frac{\Pi_s - \Pi_r}{\sqrt{p}} + \frac{\Pi_r + \Pi_s}{p}\sqrt{p}\right\} \\ &= \frac{1}{2}\{-b + c - (b+c)\sqrt{p}\}.\end{aligned}$$

Further, using (5.2.1) and (5.2.2) we may rewrite this formula as

$$\varepsilon^h = \frac{1}{p-1}\left\{\sqrt{pa^2 - (p-1)^2} + a\sqrt{p}\right\}.$$

Consequently, we can state the following theorem (cf. [1]):

Theorem 5.3. *If $p \equiv 1 \pmod 4$, then*

$$\begin{aligned}h &= \frac{\log\left\{\sqrt{pa^2 - (p-1)^2} + a\sqrt{p}\right\} - \log(p-1)}{\log \varepsilon} \\ &= \frac{\log\{-b + c - (b+c)\sqrt{p}\} - \log 2}{\log \varepsilon}.\end{aligned}$$

5.4. Here, the problem how to determine the integers a, b and c arises. In view of Lemma 5.2, the integer $a = a(p)$ may be written as

$$a = 1 + \sum_{n=1}^{(p-1)/2} (-1)^n N_n,$$

where N_n is the number of points of

$$\left\{ x \in V \mid \sum_{i=1}^{n} x_i^2 = 0 (1 \leqslant x_1 < x_2 < \cdots < x_n \leqslant (p-1)/2) \right\},$$

hence $N_n = (2^n n!)^{-1} |S_\alpha^0|$ for $\alpha = 0$, as mentioned in Proposition 4.2. Similarly, we get

$$b = \sum_{n=1}^{(p-1)/2} (-1)^n N_n' \quad \text{and} \quad c = \sum_{n=1}^{(p-1)/2} (-1)^n N_n'',$$

where $N_n' = (2^n n!)^{-1} |S_\alpha^0|$ for $\theta(\alpha) = 1$ and $N_n'' = (2^n n!)^{-1} |S_\alpha^0|$ for $\theta(\alpha) = -1$.

As stated in Section 4, it was shown by Le that N_n can be determined recursively by means of (4.1.3).

We now apply the formulas for $|S_\alpha^0|$ mentioned in Theorem 3.7 for evaluating the integers a, b, c and present the following explicit expressions of them.

Proposition 5.5. *Let $B = B(n)$ and $E = E(n)$ be as in Theorem* 3.7. *Then*

$$a = \frac{p-1}{p} \left\{ 1 + \sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} E(n) \right\},$$

$$b = -\frac{1}{p} \left\{ 1 + \sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} (pB(n) + E(n)) \right\},$$

$$c = -\frac{1}{p} \left\{ 1 + \sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} (-pB(n) + E(n)) \right\}.$$

Proof. The proofs of the above formulas are almost the same, so we shall give here the proof only for the integer a. By making use of the formula for $|S_\alpha^0|$ in Theorem 3.7 we deduce

$$\begin{aligned} a &= 1 + \frac{1}{p} \sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} \{A_0(n) + (p-1)E(n)\} \\ &= 1 + \frac{1}{p} \sum_{n=1}^{(p-1)/2} \left\{ (-1)^n \binom{(p-1)/2}{n} + (p-1) \frac{(-1)^n}{2^n n!} E(n) \right\} \\ &= \frac{p-1}{p} \left\{ 1 + \sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} E(n) \right\}, \end{aligned}$$

which proves the first formula as desired. □

According to Theorem 5.3 and Proposition 5.5, the formula for the class number h of $\mathbb{Q}(\sqrt{p})$ for $p \equiv 1 \pmod 4$ can be explicitly derived.

5.6. Next, we would like to treat the case where $p \equiv 5 \pmod 8$. This case is a special case of $p \equiv 1 \pmod 4$, however it leads to other results of interest. For instance, we can deduce a different form of expression of h in this case.

Set $\Pi'_n = \prod_n (1+\zeta^n)$ for $n = r, s$. Similarly to Lemma 5.2, we get the following lemma.

Lemma 5.7. *There exist integers $\bar{a}, \bar{b}$ and $\bar{c}$ such that*

$$\Pi'_n = \bar{a} + \bar{b}\sum_r \zeta^r + \bar{c}\sum_s \zeta^s \quad \text{and} \quad \Pi'_s = \bar{a} + \bar{b}\sum_s \zeta^s + \bar{c}\sum_r \zeta^r,$$

where the sums $\sum_r, \sum_s$ are taken over all integers r, s in the interval $(0, p)$ such that $\theta(r) = 1, \theta(s) = -1$, respectively.

Here, we easily see that

$$\bar{a} + \frac{p-1}{2}(\bar{b} + \bar{c}) = 2^{(p-1)/2} \tag{5.7.1}$$

and

$$\left\{2^{(p+1)/2} - (\bar{b} + \bar{c})p\right\}^2 - 4 = p(\bar{b} - \bar{c})^2. \tag{5.7.2}$$

Since $\theta(2) = (-1)^{(p^2-1)/8} = -1$ for $p \equiv 5 (\mathrm{mod}\, 8)$, $\Pi'_r + \Pi'_s = 2\bar{a} - (\bar{b} + \bar{c}) = 2^{(p+1)/2} - (\bar{b} + \bar{c})p > 0$ and $\Pi'_r - \Pi'_s = (\bar{b} - \bar{c})\sqrt{p} > 0$, we get from (5.7.1) and (5.7.2)

$$\begin{aligned}
\varepsilon^{2h} = \Pi'_r &= \frac{1}{2}\left\{\Pi'_r + \Pi'_s + \frac{\Pi'_r - \Pi'_s}{\sqrt{p}}\sqrt{p}\right\} \\
&= \frac{1}{2}\left\{2^{(p+1)/2} - (\bar{b} + \bar{c})p + (\bar{b} - \bar{c})\sqrt{p}\right\} \\
&= \frac{1}{p-1}\left\{p\bar{a} - 2^{(p-1)/2} + \sqrt{(p\bar{a} - 2^{(p-1)/2})^2 - (p-1)^2}\right\}.
\end{aligned}$$

Using this formula we obtain immediately the analogous result to Theorem 5.3 as follows (cf. [1]):

Theorem 5.8. *If $p \equiv 5 \pmod 8$, then*

$$\begin{aligned}
h &= \frac{\log\left\{p\bar{a} - 2^{(p-1)/2} + \sqrt{(p\bar{a} - 2^{(p-1)/2})^2 - (p-1)^2}\right\} - \log(p-1)}{2\log\varepsilon} \\
&= \frac{\log\left\{2^{(p+1)/2} - (\bar{b} + \bar{c})p + (\bar{b} - \bar{c}\sqrt{p}\right\} - \log 2}{2\log\varepsilon}.
\end{aligned}$$

We see from Lemma 5.7 that the integers $\bar{a}, \bar{b}, \bar{c}$ can be expressed as follows:

$$\bar{a} = 1 + \sum_{n=1}^{(p-1)/2} N_n, \quad \bar{b} = \sum_{n=1}^{(p-1)/2} N'_n, \quad \bar{c} = \sum_{n=1}^{(p-1)/2} N''_n.$$

In the next proposition we shall derive explicit expressions of these numbers.

Proposition 5.9. *Let $B = B(n)$ and $E = E(n)$ be as in Theorem 3.7. Then*

$$\bar{a} = 1 + \frac{1}{p}\left(2^{(p-1)/2} - 1\right) + \frac{p-1}{p} \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} E(n),$$

$$\bar{b} = \frac{1}{p}\left(2^{(p-1)/2} - 1\right) - \frac{1}{p} \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} (pB(n) + E(n)),$$

$$\bar{c} = \frac{1}{p}\left(2^{(p-1)/2} - 1\right) - \frac{1}{p} \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} (-pB(n) + E(n)).$$

Proof. Applying the formula for $|S^0_\alpha|$ (where $\alpha = 0$) in Theorem 3.7, we easily deduce

$$\begin{aligned}
\bar{a} &= 1 + \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} |S^0_\alpha| \\
&= 1 + \frac{1}{p} \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} \{A_0(n) + (p-1)E(n)\} \\
&= 1 + \frac{1}{p} \left\{ \sum_{n=1}^{(p-1)/2} \binom{(p-1)/2}{n} + (p-1) \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} E(n) \right\} \\
&= 1 + \frac{1}{p} \left\{ 2^{(p-1)/2} - 1 + (p-1) \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} E(n) \right\},
\end{aligned}$$

as indicated. The other assertions for $\bar{b}$ and $\bar{c}$ are also immediate from Theorem 3.7. □

By Theorem 5.8 and Proposition 5.9 we can give the explicit formula of the class number h of $\mathbb{Q}(\sqrt{p})$ in the case where $p \equiv 5 \pmod 8$.

6. The Ankeny-Artin-Chowla Conjecture

6.1. Let $\varepsilon = (T + U\sqrt{p})/2 > 1$ be the fundamental unit of $\mathbb{Q}(\sqrt{p})$ for $p \equiv 1 \pmod 4$, hence T and U are the least positive solutions of Pell's equation $X^2 + 4 = pY^2$.

In their paper [2], ANKENY, ARTIN and CHOWLA proved an important property between the class number and the fundamental unit of $\mathbb{Q}(\sqrt{p})$ as follows:

$$\frac{hU}{T} \equiv B_{(p-1)/2} (\mathrm{mod}\ p), \tag{6.1.1}$$

where B_m is the m-th Bernoulli number in the "*even suffix*" notation, thus $B_0 = 1, B_1 = -1/2, B_2 = 1/6, B_3 = 0$, and so on.

The Ankeny-Artin-Chowla conjecture enunciated in [2, 3] asserts that $U \not\equiv 0 \pmod{p}$, which is equivalent to $B_{(p-1)/2} \not\equiv 0 \pmod{p}$ by (6.1.1), because $h < p$, more precisely $h \leqslant [\sqrt{p}/2]$ (see [6], Theorem-(a)), where $[x]$ means the greatest integer $\leqslant x$. In spite of various efforts by many authors, this conjecture is still unsettled. It may be not as simple as one would like. Actually, this is an extremely deep and difficult problem in connection with the distribution of quadratic residues and nonresidues modulo p.

If we write $\varepsilon^k = (T_k + U_k\sqrt{p})/2$ for $k \geqslant 1$, then it is clear that the above conjecture is also equivalent to $U_k \not\equiv 0 \pmod{p}$ for each $k \geqslant 1$, hence to $p \nmid a$ and $p \nmid b + c$ in consideration of the case $k = h$.

Incidentally, we note that (6.1.1) yields the congruence

$$\frac{a}{\sqrt{pa^2 - (p-1)^2}} = \frac{b+c}{b-c} \equiv B_{(p-1)/2} (\mathrm{mod}\ p).$$

By means of Proposition 5.5 we may describe the following statement.

Proposition 6.2. *Let* $p \equiv 1 \pmod{4}$. *Then the condition* $U \equiv 0 \pmod{p}$ *is equivalent to*

$$\sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} \sum_{\substack{\nu \in \mathfrak{P}_n \\ t(\nu)=2}} c_\nu \theta(\lambda) \equiv -1 (\mathrm{mod}\ p).$$

Proof. Let $C = C(n), D = D(n)$ be as in (3.6.3) and put $E = E(n) = D(n) + C(n)$. Since

$$D(n) = (-1)^n \frac{(2n)!}{2^n n!} = \frac{(-1)^n n!}{2^n} \binom{2n}{n}$$

and

$$\sum_{n=0}^{m} \frac{1}{2^{2n}} \binom{2n}{n} = \frac{2m+1}{2^{2m}} \binom{2m}{m},$$

we have

$$\sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} D(n) = \sum_{n=1}^{(p-1)/2} \frac{1}{2^{2n}} \binom{2n}{n} = \frac{p}{2^{p-1}} \binom{p-1}{(p-1)/2} - 1.$$

Noting that $\binom{p-1}{k} \equiv (-1)^k \pmod{p}$, it follows from Proposition 5.5 that

$$\begin{aligned} a &= \frac{p-1}{p}\left\{1 + \sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} E(n)\right\} \\ &= \frac{p-1}{p} + \frac{p-1}{p}\left\{\sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} D(n) + \sum_{n=1}^{(p-1)/2} \frac{(-1)^n}{2^n n!} C(n))\right\} \\ &= \frac{p-1}{2^{p-1}}\binom{p-1}{(p-1)/2} + \frac{p-1}{p}\sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} \sum_{\substack{\nu \in \mathscr{P}_n \\ t(\nu)>0 \text{ even}}} c_\nu \theta(\lambda)(\kappa p)^{t(\nu)/2}, \\ &\equiv -\left(1 + \sum_{n=1}^{(p-1)/2} \frac{1}{2^n n!} \sum_{\substack{\nu \in \mathscr{P}_n \\ t(\nu)=2}} c_\nu \theta(\lambda)\right) \pmod{p}. \end{aligned}$$

Therefore, the condition $p|a$ equivalent to $U \equiv 0 \pmod{p}$ implies the congruence indicated in the statement, and the reverse is also true. □

References

[1] AGOH T (1989) A note on unit and class number of real quadratic fields. Acta Math Sinica (NS) **5**: 281–288

[2] ANKENY NC, ARTIN E, CHOWLA S (1952) The class number of real quadratic fields. Ann Math **56**: 479–493

[3] ANKENY NC, CHOWLA S (1962) A further note on the class number of real quadratic fields. Acta Arith **7**: 271–272

[4] LIDL R, NIEDERREITER H (1983) Finite Fields. Reading, MA: Addison-Wesley

[5] LE M-H (1993) The number of solutions of a certain quadratic congruence related to the class number of $\mathbb{Q}(\sqrt{p})$. Proc Amer Math Soc **117**: 1–3

[6] LE M-H (1994) Upper bounds for class number of real quadratic fields. Acta Arith **68**: 141–144

[7] ORLIK P, SOLOMON L (1980) Unitary reflection groups and cohomology. Invent Math **59**: 77–94

[8] SOLOMON L (1963) Invariants of finite reflection groups. Nagoya Math J **22**: 57–64

T. AGOH and T. SHOJI
Department of Mathematics
Science University of Tokyo
Noda, Chiba 278
Japan
e-mail: agoh@ma.noda.sut.ac.jp
shoji@ma.noda.sut.ac.jp

Mh. Math. 125, 293–308 (1998)

Monatshefte für
Mathematik

Printed in Austria

On a Conjecture of Siegel

By

Enrico Bombieri, Princeton, NJ, and **Julia Mueller**, Bronx, NY

(Received 12 May 1997; in revised form 13 August 1997)

Abstract. Let k be an algebraically closed field of characteristic 0 and let $f(x,y) \in k[t][x,y]$ be a polynomial in two variables with coefficients in $k[t]$. One is interested in solving the equation $f(x,y) = 0$ with polynomials $x, y \in k[t]$. Two solutions (x,y), (x',y') are *proportional* if x'/x and y'/y are non-zero constants in k and a solution (x,y) is *primitive* if the polynomials x and y have no common root. The main result of this paper is that for a certain class of polynomials f, which includes Thue equations with sufficiently lacunary exponents, the number of non-proportional, primitive solutions is bounded solely in terms of the number of monomials $a_i(t)x^{\alpha_i}y^{\beta_i}$ appearing in the polynomial $f(x,y)$. This verifies the analogue of a conjecture of Siegel for this class of polynomials. The proof is an application of the *abc*-theorem in function fields to certain determinantal varieties arising from the elimination of the coefficients of the polynomial $f(x,y)$, together with an inductive argument on the number r of monomials in $f(x,y)$.

1. Introduction

In his celebrated paper [7, §7] SIEGEL considers the problem of obtaining bounds for the number of solutions of the diophantine equation $f(x,y) = 0$, as a function of the coefficients of f, assuming that this number is finite. Then he says "Man kann nun vermuten, daß sich sogar eine Schranke finden läßt, die nur von der Anzahl der Koeffizienten abhängt; doch dürfte dies recht schwer zu beweisen sein."

In this form, the conjecture is not quite true, since the number of integral points on a nonsingular plane curve can be arbitrarily high; as remarked by Mahler, it suffices to take an elliptic curve of positive rank, take any finite set of rational points on this curve and then make a change of coordinates $(x',y') = (Nx, Ny)$ with N the least common multiple of the denominators of the coordinates of these rational points, to obtain a curve with integral points.

On the other hand, for most points (x',y') obtained by means of this procedure one sees that x' and y' cannot be coprime, thus Mahler's remark does not apply to the case of *primitive* solutions, in which the coordinates x and y are restricted to be coprime.

Hence let us consider only the case of primitive solutions. Even then, some additional condition is needed, as shown by the equation $xy - a = 0$ with a a squarefree integer with s prime factors. The number of solutions is twice the

1991 Mathematics Subject Classification: 11D99
Key words: Diophantine equations

number of divisors of a, namely 2^{s+1}. Similar examples can be constructed with the equation $x^p y^q - a = 0$.

Another obvious case of failure is an equation $f(x) = 0$, considered as an equation in (x, y), since in this case y can be chosen arbitrarily.

The example of an equation $ax^m + by^n = c$ in a number field shows that we should consider as equivalent two solutions (x_1, y_1) and (x_2, y_2) for which $x_1^m = x_2^m$ and $y_1^n = y_2^n$. Thus, as in a preliminary study of binomial and trinomial equations in [4], [5] and [1], we say that two solutions (x, y) and (x', y') are *proportional* if both x/x' and y/y' are units.

Even after excluding these obvious cases of failure, we note that Siegel's conjecture in the case of a polynomial of degree 3 already lies very deep. Let $r_3(N)$ be the number of distinct representations of N as a sum of two cubes of coprime integers (Ramanujan's famous taxicab number[1] $1729 = 1^3 + 12^3 = 9^3 + 10^3$ is the smallest integer admitting two essentially distinct decompositions into two integer cubes). One can show that the rank of the elliptic curve $x^3 + y^3 = N$ is at least $c \log r_3(N)$, therefore if $r_3(N)$ were unbounded one would have elliptic curves over $\mathbb{Q}$ with arbitrarily large rank, solving the notoriously difficult problem of the boundedness of the rank of elliptic curves over $\mathbb{Q}$.

One can formulate several variants and generalizations of Siegel's conjecture, for example replacing $\mathbb{Z}$ by the ring of S-integers of a number field and asking how the number of primitive solutions depends, if at all[2], on S and the degree and discriminant of the field. Also, it is possible to formulate several higher dimensional generalizations of this conjecture, all related in some way to Lang's well-known conjectures on diophantine equations. These variants may include for example substituting geometric conditions in place of primitivity, and one may consider as well rational solutions instead of just integral solutions.

It is an interesting problem to formulate general conditions under which Siegel's conjecture is valid, not only over number fields but also in the easier case of function fields. In this paper, we study a special case of Siegel's conjecture, in the ring $R = k[t]$ of polynomials in one variable over an algebraically closed field k of characteristic 0. In this case, proportionality of solutions means that x/x' and y/y' are constants in k.

We consider in this paper a special class of equations, namely

$$\sum_{i=1}^{r} a_i x^{\alpha_{1i}} y^{\alpha_{2i}} = c \tag{1.1}$$

with $r \geqslant 2$, where we make the following assumptions:

(*A*) $a_i \neq 0$ *for* $i = 1, \ldots, r$;

(*B*) *for* $v = 1, 2$ *the exponents* $\alpha_{vi}, i = 1, \ldots, r$ *are K-lacunary*:

$$|\alpha_{vi} - \alpha_{vj}| \geqslant K \qquad \text{for } i \neq j;$$

(*C*) $\alpha_{11} > \alpha_{12} > \cdots > \alpha_{1r}$ *and* $\alpha_{21} < \alpha_{22} < \cdots < \alpha_{2r}$;

[1] G. H. Hardy, *Ramanujan*, Cambridge Univ. Press 1940, p 12

[2] Quite possibly it is independent of such data

and ask for primitive, non-proportional solutions $(x, y) \in k[t]^2$ to (1.1). Note that we do not ask for $c \neq 0$ in (1.1), except in the otherwise trivial case $r = 2$.

Hypothesis (C) is perhaps the simplest hereditary property (i.e. a property which remains valid after restriction to subsets) which eliminates unwanted cases, for example an equation $f(x) = 0$ considered as an equation in (x, y). Although restrictive, it is satisfied by the important class of Thue equations.

We shall prove the following result. Let $N(r, K)$ be the maximum number of primitive, non-proportional solutions of an equation of type (1.1) satisfying conditions (A), (B), (C).

Theorem. *The number of non-proportional, primitive solutions of equation (1.1) satisfying (A), (B) and (C) with* $K = (2r+1)^{(r+1)(r+2)}$ *is at most* $(2r+1)^{r-1}$.

As in [4], [5], [1] our arguments depend on the $abc \cdots z$-theorem of Stothers, Mason, Brownawell and Masser, and Voloch, in function fields, see [2]. Most of our considerations will go through in the arithmetic case, once the appropriate $abc \cdots z$-conjecture becomes available. We note here that a modified form of Siegel's conjecture over $\mathbb{Z}$ has been proved by J. MUELLER and W. M. SCHMIDT [6] in the case of Thue equations, by a completely different method. However it is not clear how to extend the arguments of [6] to S-integers of number fields or to substantially wider classes of equations than Thue equations. We consider our result to indicate that Siegel's conjecture is true, also over number fields, in substantial generality notwithstanding the exceptions mentioned before.

2. The Determinantal Map

We begin by considering the more general situation of a variety in n-dimensional affine space $\mathbb{A}^n$ over the ring $R = k[t]$. We denote by $\mathbf{x} = (x_1, \ldots, x_n)$ the standard coordinates in $\mathbb{A}^n$.

Let us abbreviate $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and consider a finite set $\mathscr{A}$ of exponent vectors $\boldsymbol{\alpha}$. We say that a subvariety $X \subset \mathbb{A}^n$ is an $\mathscr{A}$-variety if it can be defined by polynomial equations of type

$$\sum_{\boldsymbol{\alpha} \in \mathscr{A}} a_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}} = 0 \tag{2.1}$$

with coefficients $a_{\boldsymbol{\alpha}} \in R$.

Let $m = |\mathscr{A}|$ be the number of monomials $\mathbf{x}^{\boldsymbol{\alpha}}$ which may be involved in the equation. We consider the affine space

$$\mathbb{A}^{mn} = \underbrace{\mathbb{A}^n \times \mathbb{A}^n \times \cdots \times \mathbb{A}^n}_{m \text{ times}}$$

with coordinates $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$, and denote by $W(\mathscr{A})$ the subvariety of $\mathbb{A}^{mn}$ defined by the equation

$$\det \begin{pmatrix} \mathbf{x}_1^{\boldsymbol{\alpha}} \\ \mathbf{x}_2^{\boldsymbol{\alpha}} \\ \cdot \\ \cdot \\ \mathbf{x}_m^{\boldsymbol{\alpha}} \end{pmatrix}_{\boldsymbol{\alpha} \in \mathscr{A}} = 0 \tag{2.2}$$

with columns indexed by the m vectors $\boldsymbol{\alpha}$.

The following simple but basic fact is fundamental for us.

Proposition 1. *Let X be an $\mathscr{A}$-variety, $m = |\mathscr{A}|$ and let X^m be the cartesian product of X with itself m times. Then X^m is a subvariety of $W(\mathscr{A})$.*

Proof. For the proof, we need to show that if $\mathbf{x}_i \in X$ then the rows of the matrix (2.2) are linearly dependent. Since X is an $\mathscr{A}$-variety, the points $\mathbf{x}_i$ satisfy a nontrivial equation of type (2.1). The equations

$$\sum_{\boldsymbol{\alpha}\in\mathscr{A}} a_{\boldsymbol{\alpha}}\mathbf{x}_i^{\boldsymbol{\alpha}} = 0$$

for $i = 1, \dots, m$ can be viewed as a linear system of m equations in the m unknowns $\boldsymbol{\alpha}$, admitting a non-zero solution. Thus the m rows $(\mathbf{x}_i^{\boldsymbol{\alpha}})$ of the matrix are linearly dependent, completing the argument.

Now the main point is that any ordered set of m integral points on the $\mathscr{A}$-variety X determines an integral point on $W(\mathscr{A})$; the same of course occurs if we deal with rational points. Unlike projective spaces, the varieties $W(\mathscr{A})$ often are of general type and, as predicted by Lang's conjectures, one does not expect them to have too many integral or rational points. Thus the strategy of our arguments consists in studying the distribution of integral or rational points on the variety $W(\mathscr{A})$ and deduce corresponding results for $\mathscr{A}$-varieties.

In this paper, rather than relying on unproven general conjectures, we shall systematically exploit the general *abc*-inequality in function fields to obtain in the end unconditional theorems.

It would be quite interesting to determine the structure of the special locus (in the sense of LANG [3, Ch.I, §3]) of the varieties $W(\mathscr{A})$. However this appears to be a difficult problem requiring deep tools from algebraic geometry, and we have nothing to say about it except the obvious.

3. Application of the *abc*-Inequality

In this section, to each place v of $k(t)$ we associate an additive valuation $v(\ )$ normalized so that the product formula

$$\sum_{v} v(x) = 0$$

holds for every $x \in k(t)$, $x \neq 0$.

Let $\mathbf{X}$ be an ordered m-tuple of solutions of (2.1). We write $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$. By Proposition 1, we have

$$\det(\mathbf{X}^{\boldsymbol{\alpha}}) = \det\begin{pmatrix} \mathbf{x}_1^{\boldsymbol{\alpha}} \\ \mathbf{x}_2^{\boldsymbol{\alpha}} \\ \cdot \\ \cdot \\ \mathbf{x}_m^{\boldsymbol{\alpha}} \end{pmatrix}_{\boldsymbol{\alpha}\in\mathscr{A}} = 0.$$

We expand the determinant and obtain

$$\sum_{\sigma\in S(m)} \varepsilon(\sigma) \prod_{j=1}^{m} \mathbf{x}_{\sigma(j)}^{\boldsymbol{\alpha}_j} = 0, \tag{3.1}$$

where σ runs over the symmetric group $S(m)$ of permutations of $\{1,2,\dots,m\}$, $\varepsilon(\sigma)$ is the parity of σ and α_j denotes an ordering of $\mathscr{A}$, which we fix once for all. We also abbreviate

$$m_\sigma = \varepsilon(\sigma)\prod_{j=1}^{m}\mathbf{x}_{\sigma(j)}^{\alpha_j}.$$

A subset $\mathscr{B}\subseteq S(m)$ is called a block (with respect to the given m-tuple of solutions $\mathbf{X}$) if

$$\sum_{\sigma\in\mathscr{B}} m_\sigma = 0;$$

a block is called irreducible if it does not contain any proper subblock.

By (3.1), we know that $S(m)$ itself is a block, although it need not be irreducible. Let

$$\sum_{\sigma\in S(m)} m_\sigma = \sum_{\mathscr{B}}\left(\sum_{\sigma\in\mathscr{B}} m_\sigma\right)$$

be a decomposition of the sum according to irreducible subblocks. This decomposition need not be unique, so in what follows we will always consider not just $\mathbf{X}$ but rather a pair consisting of $\mathbf{X}$ together with an associated block decomposition.

Given a place v of $k(t)$, an m-tuple of solutions $\mathbf{X}$ and a block decomposition $\{\mathscr{B}\}$ into irreducible blocks, we define

$$I_v(\mathscr{B}) = \max_{\sigma\in\mathscr{B}} v(m_\sigma) - \min_{\sigma\in\mathscr{B}} v(m_\sigma).$$

Definition. Let $\mathbf{X}$ be an ordered m-tuple of solutions of (2.1) together with an associated block decomposition $\{\mathscr{B}\}$. The place v is said to be:

K-good for $\mathbf{X}$, if there is $\mathscr{B}$ such that $I_v(\mathscr{B})\geqslant K$;

neutral for $\mathbf{X}$, if $I_v(\mathscr{B}) = 0$ for every block $\mathscr{B}$ associated to $\mathbf{X}$;

K-bad for $\mathbf{X}$, if v is neither K-good nor neutral for $\mathbf{X}$.

Let also $\mathscr{X}$ be a set of solutions of (2.1). Then v is said to be:

K-good for $\mathscr{X}$, if it is K-good for some $\mathbf{X}\in\mathscr{X}^m$;

neutral for $\mathscr{X}$, if it is neutral for every $\mathbf{X}\in\mathscr{X}^m$;

K-bad for $\mathscr{X}$, if v is K-bad or neutral for every $\mathbf{X}\in\mathscr{X}^m$ and moreover is K-bad for at least one such $\mathbf{X}$.

We also say that v is *strongly neutral for* $\mathscr{X}$ if for every $\mathbf{X}\in\mathscr{X}^m$ we have that $v(m_\sigma)$ is independent of σ.

Now we recall the *abc*-inequality of Brownawell, Masser and Voloch, which we state in the following form.

***abc*-inequality.** *Let $a_i, i = 1,2,\dots,n$ be elements $a_i\in k(t)$ and suppose that*

$$\sum_{i=1}^{n} a_i = 0$$

and that no proper subsum of $\sum a_i$ *vanishes. Then we have*

$$-\sum_v \min_i v(a_i) \leqslant \frac{1}{2}(n-1)(n-2)|S|$$

where $\sum_v$ *runs over all places of* $k(t)$ *and where* S *is the set of places* v *for which* $\max_i v(a_i) > \min_i v(a_i)$.

We bring this in a more convenient form for our purposes by noting that the product formula yields

$$-\sum_v \min_i v(a_i) = \sum_v \left(v(a_j) - \min_i v(a_i) \right)$$

for every j, whence

$$\begin{aligned} -n\sum_v \min_i v(a_i) &= \sum_v \sum_{i=1}^{n} \left(v(a_j) - \min_i v(a_i) \right) \\ &\geqslant \sum_v \left(\max_i v(a_i) - \min_i v(a_i) \right). \end{aligned}$$

A fortiori it follows from the *abc*-inequality that

Corollary. *We have*

$$\begin{aligned} &\sum_v \left(\max_i v(a_i) - \min_i v(a_i) \right) \\ &\leqslant \frac{1}{2} n(n-1)(n-2) \cdot |v : \max_i v(a_i) > \min_i v(a_i)|. \end{aligned} \tag{3.2}$$

Lemma 1. *Let* $\mathscr{X}$ *be a set of solutions of (2.1) and let* $K > \frac{1}{2}(m!)^3 \cdot |\mathscr{X}|^m$. *Suppose that not every* v *is neutral for* $\mathscr{X}$. *Then there is at least one* K*-bad* v *for* $\mathscr{X}$.

Proof. In what follows, $\mathscr{B} = \mathscr{B}(\mathbf{X})$ will refer to irreducible blocks associated to the chosen decomposition of $\det(\mathbf{X}^{\alpha})$ into irreducible blocks. We assume that there is at least one K-good place for $\mathscr{X}$ and no K-bad places for $\mathscr{X}$, and derive a contradiction at the end.

We apply (3.2) to each sum

$$\sum_{\sigma \in \mathscr{B}} m_\sigma = 0$$

getting

$$\begin{aligned} \sum_{\mathbf{X}} \sum_{\mathscr{B}} \sum_v I_v(\mathscr{B}) &\leqslant \sum_{\mathbf{X}} \sum_{\mathscr{B}} \frac{1}{2} |\mathscr{B}|^3 \cdot |v : I_v(\mathscr{B}) > 0| \\ &\leqslant \frac{1}{2} \left(\sum_{\mathbf{X}} \sum_{\mathscr{B}} |\mathscr{B}|^3 \right) \cdot |v : I_v(\mathscr{B}) > 0 \text{ for some } \mathbf{X} \text{ and } \mathscr{B}| \\ &\leqslant \frac{1}{2} \left(\sum_{\mathbf{X}} \left(\sum_{\mathscr{B}} |\mathscr{B}| \right)^3 \right) \cdot |v : I_v(\mathscr{B}) > 0 \text{ for some } \mathbf{X} \text{ and } \mathscr{B}| \\ &= \frac{1}{2} (m!)^3 \cdot |\mathscr{X}|^m \cdot |v : I_v(\mathscr{B}) > 0 \text{ for some } \mathbf{X} \text{ and } \mathscr{B}|. \end{aligned}$$

On the other hand, the left-hand side of this chain of inequalities is bounded below by $K \cdot |v : v \text{ is } K\text{-good}|$. Also, by our definitions, we have

$$|v : I_v(\mathscr{B}) > 0 \text{ for some } \mathbf{X} \text{ and } \mathscr{B}| = |v : v \text{ is } K\text{-good}| + |v : v \text{ is } K\text{-bad}|.$$

Now if there were no K-bad places we would deduce

$$K \cdot |v : v \text{ is } K\text{-good}| \leqslant \frac{1}{2}(m!)^3 \cdot |\mathscr{X}|^m \cdot |v : v \text{ is } K\text{-good}|.$$

By assumption, there is at least one K-good place so we may divide both sides of this inequality by $|v : v \text{ is } K\text{-good}|$ (which is not zero), getting $K \leqslant \frac{1}{2}(m!)^3 |\mathscr{X}|^m$. This is a contradiction.

4. Auxiliary Results

We consider here a special class of equations, namely

$$\sum_{i=1}^{r} a_i x^{\alpha_{1i}} y^{\alpha_{2i}} = c \tag{1.1}$$

where $r \geqslant 2$ and:

(A) $a_i \neq 0$ for $i = 1, \ldots, r$;

(B) for $v = 1, 2$ the exponents $\alpha_{vi}, i = 1, \ldots, r$ are K-lacunary:

$$|\alpha_{vi} - \alpha_{vj}| \geqslant K \quad \text{for } i \neq j;$$

(C) $\alpha_{11} > \alpha_{12} > \cdots > \alpha_{1r}$ and $\alpha_{21} < \alpha_{22} < \cdots < \alpha_{2r}$.

Assumption (C) will be used only in the proof of Lemma 4 below.

Note that we do not assume $c \neq 0$, thus the number m of monomials in (1.1) is

$$m = \begin{cases} r+1 & \text{if } c \neq 0 \\ r & \text{if } c = 0. \end{cases}$$

If $m = r + 1$ then $\boldsymbol{\alpha}_m = (0, 0)$.

In what follows, we write $\mathbf{x}^{\boldsymbol{\alpha}_i}$ for the monomial $x^{\alpha_{1i}} y^{\alpha_{2i}}$. Let $\mathscr{X}$ be a set of primitive solution to (1.1) and fix a place v. The assignment $(x, y) \mapsto (v(x), v(y))$ gives us a map $\phi : \mathscr{X} \to \mathbb{Z}^2$. It is clear that

$$\phi(\mathscr{X}) \subset (0,0) \cup (\mathbb{N}, 0) \cup (0, \mathbb{N}). \tag{4.1}$$

Lemma 2. *Suppose* $|\phi(\mathscr{X}) \cap (\mathbb{N}, 0)| \geqslant m$. *Then either* v *is* K*-good for* $\mathscr{X}$ *or every* $\mathbf{x} \in \phi^{-1}((\mathbb{N}, 0))$ *satisfies an equation*

$$\sum_{i=1}^{r-1} b_i x^{\beta_{1i}} y^{\beta_{2i}} = d$$

with $b_i \neq 0$ *and where the exponents* $\boldsymbol{\beta}_i$ *are a subset of the exponents* $\boldsymbol{\alpha}_j, j = 1, \ldots, r$. *A corresponding statement holds if* $|\phi(\mathscr{X}) \cap (0, \mathbb{N})| \geqslant m$.

Proof. There is no loss in generality in assuming that $\alpha_{11} > \alpha_{12} > \cdots > \alpha_{1r} \geqslant \alpha_{1m}$. Now let $\mathbf{X}$ be any set of m elements $\mathbf{x}_i \in \mathscr{X}$ such that $v(x_1) > v(x_2) > \cdots > v(x_m) > 0$.

We start again from the basic equation for $\mathbf{X} \in \mathscr{X}^m$, namely

$$\det(\mathbf{X}^{\boldsymbol{\alpha}}) = \det \begin{pmatrix} \mathbf{x}_1^{\alpha_1} & \mathbf{x}_1^{\alpha_2} & \cdots & \mathbf{x}_1^{\alpha_m} \\ \mathbf{x}_2^{\alpha_1} & \mathbf{x}_2^{\alpha_2} & \cdots & \mathbf{x}_2^{\alpha_m} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \mathbf{x}_m^{\alpha_1} & \mathbf{x}_m^{\alpha_2} & \cdots & \mathbf{x}_m^{\alpha_m} \end{pmatrix} = \sum_{\mathscr{B}} \sum_{\sigma \in \mathscr{B}} m_\sigma = 0 \tag{4.2}$$

where σ runs over permutations of $\{1, \dots, m\}$. Let us denote by a dot the scalar product. Since $\mathbf{x}_i$ is primitive and $v(x_i) > 0$ by construction, we must have $v(y_i) = 0$ for every i. It follows that

$$\begin{aligned} v(m_\sigma) &= \sum_{i=1}^{m} \boldsymbol{\alpha}_i \cdot v(\mathbf{x}_{\sigma(i)}) \\ &= \sum_{i=1}^{m} \left(\alpha_{1i} v(x_{\sigma(i)}) + \alpha_{2i} v(y_{\sigma(i)})\right) \\ &= \sum_{i=1}^{m} \alpha_{1i} v(x_{\sigma(i)}). \end{aligned} \tag{4.3}$$

Let *id* denote the identity permutation. We verify that (4.3) implies that $v(m_\sigma)$ takes its maximum value precisely when $\sigma = id$ if $m = r$ or $m = r+1$ and $\alpha_{1r} > 0$, while if $m = r+1$ and $\alpha_{1r} = 0$ we have the two possibilities $\sigma = id$ and σ equal to the transposition on $(r, r+1)$. Also, if σ is not one of the extremal permutations for which $v(m_\sigma)$ is a maximum, then K-lacunarity implies

$$v(m_{id}) - v(m_\sigma) \geqslant K. \tag{4.4}$$

To see this, suppose that there is a pair (i,j) with $i < j$ and $\sigma(i) > \sigma(j)$. Let τ be the new permutation such that $\tau(i) = \sigma(j), \tau(j) = \sigma(i)$ and $\tau(l) = \sigma(l)$ if $l \neq i, j$. Then we see that

$$v(m_\tau) - v(m_\sigma) = (\alpha_{1i} - \alpha_{1j})(v_{\sigma(j)} - v_{\sigma(i)}),$$

and this quantity is always non-negative and is at least K unless $\alpha_{1i} - \alpha_{1j} = 0$, in which case $m = r+1, \alpha_{1r} = 0$ and (i,j) is either $(r, r+1)$ or $(r+1, r)$. This shows that if $v(m_\sigma)$ is a maximum then σ is increasing on $\{1, 2, \dots, r-1\}$, hence σ is either the identity or the transposition on $(r, r+1)$, in which case we also have $m = r+1$ and $\alpha_{1r} = 0$; in every other case (4.4) holds, as asserted.

If $\sigma = id$ is the only extremal permutation then *id* belongs to some irreducible block, which must contain at least another permutation. By (4.4), v is K-good for $\mathbf{X}$, which is the first alternative of Lemma 2.

Henceforth suppose that $m = r+1$ and $\alpha_{1r} = 0$, so that we have two extremal permutations, namely *id* and the transposition on $(r, r+1)$. The sum of the two monomials for extremal σ is

$$\mathbf{x}_1^{\alpha_1} \cdots \mathbf{x}_{r-1}^{\alpha_{r-1}} (y_r^{\alpha_{2r}} - y_{r+1}^{\alpha_{2r}}). \tag{4.5}$$

If (4.5) does not vanish we conclude as before that v is good for $\mathbf{X}$, so it remains for consideration the case in which (4.5) vanishes.

Suppose (4.5) vanishes, hence $y_r^{\alpha_{2r}} = y_{r+1}^{\alpha_{2r}}$. We claim that either v is K-good for $\mathbf{X}$ or

$$y_1^{\alpha_{2r}} = y_2^{\alpha_{2r}} = \cdots = y_{r+1}^{\alpha_{2r}}. \tag{4.6}$$

In order to verify this, suppose we have shown

$$y_{s+1}^{\alpha_{2r}} = \cdots = y_{r+1}^{\alpha_{2r}}$$

but $y_s^{\alpha_{2r}} \neq y_{r+1}^{\alpha_{2r}}$.

If we remove $y_{r+1}^{\alpha_{2r}}$ times the last column from the penultimate column of the matrix in (4.2) we obtain the equation

$$\det \begin{pmatrix} \mathbf{x}_1^{\alpha_1} & \cdots & y_1^{\alpha_{2r}} & - & y_{r+1}^{\alpha_{2r}} & 1 \\ \mathbf{x}_2^{\alpha_1} & \cdots & y_2^{\alpha_{2r}} & - & y_{r+1}^{\alpha_{2r}} & 1 \\ \cdot & \cdots & & \cdot & & \cdot \\ \mathbf{x}_s^{\alpha_1} & \cdots & y_s^{\alpha_{2r}} & - & y_{r+1}^{\alpha_{2r}} & 1 \\ \mathbf{x}_{s+1}^{\alpha_1} & \cdots & & 0 & & 1 \\ \mathbf{x}_{s+2}^{\alpha_1} & \cdots & & 0 & & 1 \\ \cdot & \cdots & & \cdot & & \cdot \\ \mathbf{x}_{r+1}^{\alpha_1} & \cdots & & 0 & & 1 \end{pmatrix} = 0. \tag{4.7}$$

If we make a Laplace expansion of (4.7) according to the penultimate column of the matrix we see that

$$\sum_{j=1}^{s} (-1)^{r+j} M_j \cdot (y_j^{\alpha_{2r}} - y_{r+1}^{\alpha_{2r}}) = 0 \tag{4.8}$$

where

$$M_j = \det \begin{pmatrix} \mathbf{x}_1^{\alpha_1} & \cdots & \mathbf{x}_1^{\alpha_{r-1}} & 1 \\ \mathbf{x}_2^{\alpha_1} & \cdots & \mathbf{x}_2^{\alpha_{r-1}} & 1 \\ \cdot & \cdots & \cdot & \cdot \\ \mathbf{x}_{j-1}^{\alpha_1} & \cdots & \mathbf{x}_{j-1}^{\alpha_{r-1}} & 1 \\ \mathbf{x}_{j+1}^{\alpha_1} & \cdots & \mathbf{x}_{j+1}^{\alpha_{r-1}} & 1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ \mathbf{x}_{r+1}^{\alpha_1} & \cdots & \mathbf{x}_{r+1}^{\alpha_{r-1}} & 1 \end{pmatrix}$$

Now, by computing the order of monomials, we see that in the determinant expansion of M_j there is a unique monomial with maximum order, namely the monomial $m(j)$ coming from the principal diagonal with order

$$v(m(j)) = \sum_{i=1}^{j-1} \alpha_{1i} v(x_i) + \sum_{i=j+1}^{r} \alpha_{1,i-1} v(x_i);$$

all other monomials in the expansion for M_j have smaller order by at least K. Since

$$v(m(j+1)) - v(m(j)) = \alpha_{ij}(v(x_j) - v(x_{j+1})) \geqslant K$$

we conclude that in the expansion (4.8) the maximum order is reached precisely for $m(s) \cdot (y_s^{\alpha_{2r}} - y_{r+1}^{\alpha_{2r}})$ and all other terms have smaller order by at least K. Since $y_s^{\alpha_{2r}} \neq y_{r+1}^{\alpha_{2r}}$ by hypothesis, the two terms $m(s)y_s^{\alpha_{2r}} - m(s)y_{r+1}^{\alpha_{2r}}$ cannot determine a block and any extension to a block gives a K-good block for $\mathbf{X}$, completing the proof of our claim.

Finally suppose that v is not K-good for $\mathscr{X}$. We have shown that the only remaining alternative is that $y_1^{\alpha_{2r}} = \cdots = y_{r+1}^{\alpha_{2r}}$ whenever $\mathbf{X}$ is an $(r+1)$-tuple of solutions in $\mathscr{X}$ such that $v(x_1) > v(x_2) > \cdots > v(x_{r+1}) > 0$. This implies that $y^{\alpha_{2r}}$ is constant on the subset of $\mathscr{X}$ defined by

$$\mathscr{X}' = \phi^{-1}((\mathbb{N}, 0)).$$

In this case, any $\mathbf{x} \in \mathscr{X}'$ verifies

$$\sum_{i=1}^{r-1} a_i x^{\alpha_{1i}} y^{\alpha_{2i}} = d$$

with the new constant d given by $d = c - a_r y^{\alpha_{2r}}$.

This is the second alternative of Lemma 2, and the result follows.

Lemma 3. *Let A, B be two distinct points of type either $(A,B) = ((a,0),(b,0))$ or $(A,B) = ((0,a),(0,b))$ and suppose $|\phi^{-1}(A)| \geqslant m-1$ and $|\phi^{-1}(B)| \geqslant 1$. Then either v is K-good for $\mathscr{X}$, or all elements of $\phi^{-1}(A)$ satisfy an equation*

$$\sum_{i=1}^{s} b_i x^{\beta_{1i}} y^{\beta_{2i}} = d$$

with $s < r, b_i \neq 0$ and where the exponents $\boldsymbol{\beta}_i, i = 1, \ldots, s$ are a subset of the exponents $\boldsymbol{\alpha}_j, j = 1, \ldots, r$.

Proof. It suffices to consider the case $(A,B) = ((a,0),(b,0))$. Suppose first $b > a \geqslant 0$. As in the proof of Lemma 2, there is no loss of generality in assuming $\alpha_{11} > \alpha_{12} > \cdots > \alpha_{1r}$.

We pick any $m-1$ solutions with $v(x_i) = a, i = 1, \ldots, m-1$ and choose for $\mathbf{x}_m$ a solution for which $v(x_m) = b$. We have $v(y_i) = 0$ for $i = 1, \ldots, m$. Again (4.3) and (4.4) hold but now

$$v(m_\sigma) = \sum_{i=1}^{m} \alpha_{1i} v(x_{\sigma(i)}) = \left(\sum_{i=1}^{m} \alpha_{1i} \right) \cdot a + \alpha_{1\sigma^{-1}(m)} \cdot (b-a).$$

Since $b > a$, this is a maximum precisely when σ is a permutation such that $\alpha_{1\sigma^{-1}(m)}$ is a maximum, that is $\sigma(1) = m$; the order of any other monomial m_σ with $\sigma(1) \neq m$ is smaller by at least K.

The sum of all monomials m_σ achieving maximum order is

$$\sum_{v(m_\sigma)=\max} m_\sigma = (-1)^{m+1}\mathbf{x}_m^{\alpha_1} \det \begin{pmatrix} \mathbf{x}_1^{\alpha_2} & \cdots & \mathbf{x}_1^{\alpha_m} \\ \mathbf{x}_2^{\alpha_2} & \cdots & \mathbf{x}_2^{\alpha_m} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \mathbf{x}_{m-1}^{\alpha_2} & \cdots & \mathbf{x}_{m-1}^{\alpha_m} \end{pmatrix}. \tag{4.9}$$

Now we have two cases.

Case I: For some choice of an m-tuple in $\phi^{-1}(A)$, the sum in (4.9) *does not determine a block.* In this case there must be an irreducible block containing a permutation from the sum in (4.9) and another permutation not appearing in (4.9). Since the order of this element is smaller than the maximum order by at least K, we infer that v is K-good for $\mathbf{X}$ and hence for $\mathscr{X}$.

Case II: The sum in (4.9) *determines a block for every m-tuple in $\phi^{-1}(A)$.* In this case (4.9) shows that

$$\det \begin{pmatrix} \mathbf{x}_1^{\alpha_2} & \cdots & \mathbf{x}_1^{\alpha_m} \\ \mathbf{x}_2^{\alpha_2} & \cdots & \mathbf{x}_2^{\alpha_m} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \mathbf{x}_{m-1}^{\alpha_2} & \cdots & \mathbf{x}_{m-1}^{\alpha_m} \end{pmatrix} = 0 \tag{4.10}$$

for every choice of $\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}$ in $\phi^{-1}(A)$. Let s be the maximum rank of such a matrix, occurring for some $m-1$-tuple $\mathbf{X}^*$; then (4.10) shows that $s \leqslant m-2$. Without loss of generality, we may assume that the first s rows of the corresponding matrix in (4.10) are linearly independent.

Consider now the $(s+1) \times (m-1)$ matrix

$$M = \begin{pmatrix} (\mathbf{x}_1^*)^{\alpha_2} & \cdots & (\mathbf{x}_1^*)^{\alpha_m} \\ (\mathbf{x}_2^*)^{\alpha_2} & \cdots & (\mathbf{x}_2^*)^{\alpha_m} \\ \cdot & \cdots & \cdot \\ (\mathbf{x}_s^*)^{\alpha_2} & \cdots & (\mathbf{x}_s^*)^{\alpha_m} \\ \mathbf{x}^{\alpha_2} & \cdots & \mathbf{x}^{\alpha_m} \end{pmatrix}.$$

Then M has rank s whenever $\mathbf{x} \in \phi^{-1}(A)$. Equating to 0 the determinant of an $(s+1) \times (s+1)$ maximal minor of M yields

$$\sum_{i \in I} b_{i,I}\mathbf{x}^{\alpha_i} = 0$$

where I is a subset of cardinality $s+1 \leqslant m-1$ of $\{2, \ldots, m\}$. The coefficients $b_{i,I}$ are determined by $\mathbf{x}_1^*, \ldots, \mathbf{x}_s^*$. Now at least one of these equations is not identically 0, and the last alternative of Lemma 3 follows.

If instead $0 \leqslant b < a$, we consider the monomials m_σ of minimum order rather than those of maximal order, with exactly the same discussion as before.

Lemma 4. *Let A, B be two distinct points of type $(A,B) = ((a,0),(0,b))$ with $a,b > 0$ and $|\phi^{-1}(A)| \geqslant m-1$ and $|\phi^{-1}(B)| \geqslant 1$. Then either v is K-good for $\mathscr{X}$, or all elements of $\phi^{-1}(A)$ satisfy an equation*

$$\sum_{i=1}^{s} b_i x^{\beta_{1i}} y^{\beta_{2i}} = d$$

with $s < r, b_i \neq 0$ and where the exponents $\boldsymbol{\beta}_i, i = 1,\dots,s$ are a subset of the exponents $\boldsymbol{\alpha}_j, j = 1,\dots,r$.

A corresponding statement holds if $|\phi^{-1}(A)| \geqslant 1$ and $|\phi^{-1}(B)| \geqslant m-1$.

Proof. The argument starts in the same way as in the proof of Lemma 3, selecting any $m-1$ elements in $\phi^{-1}(A)$ and one element in $\phi^{-1}(B)$.
We have

$$\det(\mathbf{X}^{\boldsymbol{\alpha}}) = \sum_{j=1}^{m} (-1)^{m+j} M_j \mathbf{x}_m^{\boldsymbol{\alpha}_j}$$

where M_j is the cofactor of $\mathbf{x}_m^{\boldsymbol{\alpha}_j}$:

$$M_j = \det \begin{pmatrix} \mathbf{x}_1^{\boldsymbol{\alpha}_1} & \cdots & \mathbf{x}_1^{\boldsymbol{\alpha}_{j-1}} & \mathbf{x}_1^{\boldsymbol{\alpha}_{j+1}} & \cdots & \mathbf{x}_1^{\boldsymbol{\alpha}_m} \\ \mathbf{x}_2^{\boldsymbol{\alpha}_1} & \cdots & \mathbf{x}_2^{\boldsymbol{\alpha}_{j-1}} & \mathbf{x}_2^{\boldsymbol{\alpha}_{j+1}} & \cdots & \mathbf{x}_2^{\boldsymbol{\alpha}_m} \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \mathbf{x}_{m-1}^{\boldsymbol{\alpha}_1} & \cdots & \mathbf{x}_{m-1}^{\boldsymbol{\alpha}_{j-1}} & \mathbf{x}_{m-1}^{\boldsymbol{\alpha}_{j+1}} & \cdots & \mathbf{x}_{m-1}^{\boldsymbol{\alpha}_m} \end{pmatrix}.$$

The sum of all monomials m_σ with $\sigma(j) = m$ is $(-1)^{m+j} M_j \mathbf{x}_m^{\boldsymbol{\alpha}_j}$. The order of monomials with $\sigma(j) = m$ depends only on j and is

$$v_j = \left(\sum_{i=1}^{m} \alpha_{1i} \right) \cdot a + (\alpha_{2j} b - \alpha_{1j} a).$$

If v is K-good for $\mathbf{X}$, we have the first alternative of the lemma. If not, for any two permutations σ, τ belonging to a same irreducible block we must have, writing $\sigma^{-1}(m) = i$ and $\tau^{-1}(m) = j$:

$$|v(m_\tau) - v(m_\sigma)| = |v_j - v_i| < K. \tag{4.11}$$

Now if $1 \leqslant i < j \leqslant r$ we have

$$v_j - v_i = (\alpha_{2j} - \alpha_{2i}) b + (\alpha_{1i} - \alpha_{1j}) a \geqslant 2K \tag{4.12}$$

as one sees using assumption (C). In particular, since $r \geqslant 2$ the triangle inequality gives

$$2K \leqslant |v_1 - v_r| \leqslant |v_1 - v_m| + |v_m - v_r|,$$

therefore either $|v_r - v_m| \geqslant K$ or $|v_1 - v_m| \geqslant K$.

We claim that if $|v_1 - v_m| \geqslant K$ then $M_1 = 0$, while in the alternative case $|v_r - v_m| \geqslant K$ (which implies $m = r+1$) we have $M_r = 0$. Suppose for example

$|v_1 - v_m| \geqslant K$. Let σ be a permutation with $\sigma(1) = m$ and let τ be another permutation in an irreducible block containing σ. Let j be defined by $\tau(j) = m$. Then (4.11) and (4.12) show that $j \neq 2, \ldots, r$, while (4.11) and $|v_1 - v_m| \geqslant K$ show that $j \neq m$. Thus we must have $\tau(1) = m$, with the conclusion that the set of permutations with $\sigma(1) = m$ forms a block. This proves that $M_1 \mathbf{x}_m^{\mathbf{a}_1} = 0$.

We have shown that if $|v_1 - v_m| \geqslant K$ then $M_1 = 0$ for any $(m-1)$-tuple $\mathbf{X}$ of elements of $\phi^{-1}(A)$, and the second alternative of Lemma 4 follows as in Case II in the proof of Lemma 3.

If $|v_r - v_m| \geqslant K$, we argue in the same way, this time using $M_r = 0$.

5. Proof of Theorem

We estimate the number of solutions of an equation (1.1) as follows.

Let $r \geqslant 2$, let $N(r, K)$ be the maximum number of non-proportional, primitive solutions to equation (1.1) and set

$$N^*(r, K) = \max_{2 \leqslant s \leqslant r} N(s, K).$$

Note that

$$N^*(r, K) \geqslant r$$

because we can prescribe arbitrarily r values for $\mathbf{x}$ and solve (1.1) considered as a linear homogeneous system with unknowns the $r+1$ coefficients a_i and c.

We assume that equation (1.1) has at least $N = N_r$ non-proportional primitive solutions, where N_r is sufficiently large as a function of r and $K = K_r$ is also sufficiently large, and obtain a contradiction at the end. This will show that $N(r, K) < N_r$. Our choices for N_r and K_r will be

$$N_r - 1 = (2r+1)(N_{r-1} - 1), \quad K_r = \frac{1}{2}(n+1)!^3 N_r^{r+1} + 1. \tag{5.1}$$

The initial value N_2 needs to be determined, and we shall see that we can take $N_2 = 6$. By induction, we may assume that

$$N(s, K_s) < N_s \quad \text{for} \quad s = 2, \ldots, r-1$$

and we verify by induction that, with K_r given by (5.1), we have $N(r, K_r) < N_r$.

The proof is by contradiction.

Lemma 5. *Suppose $r \geqslant 3, K \geqslant K_{r-1}$ and let $\mathscr{X}$ be a set of solutions of* (1.1) *of cardinality $|\mathscr{X}| = N_r$. Let v be a place of $k(t)$. If v is not K-good for $\mathscr{X}$ then v is strongly neutral for $\mathscr{X}$.*

Corollary. *Each v is either K-good or strongly neutral for $\mathscr{X}$, and in particular there are no K-bad paces for $\mathscr{X}$.*

Proof. Let $\phi(\mathbf{x}) = (v(x), v(y))$ be the map considered in the preceding section. If $|\phi(\mathscr{X})| = 1$ then $v(x), v(y)$ are independent of $\mathbf{x}$ so it becomes obvious that v is strongly neutral. Hence suppose that $|\phi(\mathscr{X})| \geqslant 2$. There are two distinct points A, B such that $|\phi^{-1}(A)| \geqslant |\phi^{-1}(B)| \geqslant 1$.

Case I: $(A,B)=((a,0),(b,0))$ or $(A,B)=((0,a),(0,b))$. We apply Lemma 3 to this situation. If the hypothesis of Lemma 3 is not satisfied we must have $|\phi^{-1}(A)|\leqslant r-1\leqslant N^*(r-1,K)$. Otherwise, Lemma 3 applies and since we assume that v is not K-good we deduce that every $\mathbf{x}\in\phi^{-1}(A)$ satisfies an equation

$$\sum_{i=1}^{s} b_i\mathbf{x}^{\boldsymbol{\beta}_i}=d$$

involving not more than $s<r$ non-constant monomials, whose exponents still satisfy (A), (B) and (C) with the same value of K. If $s\geqslant 2$, then by definition of $N^*(r,K)$ this equation has not more than $N(s,K)\leqslant N^*(r-1,K)$ solutions. If instead $s=1$ then $\mathbf{x}^{\boldsymbol{\beta}_1}$ (which is a monomial $\mathbf{x}^{\boldsymbol{\alpha}_i}$ for a suitable i) is independent of $\mathbf{x}$ for $\mathbf{x}\in\phi^{-1}(A)$. Thus we can move the term $a_i\mathbf{x}^{\boldsymbol{\alpha}_i}$ to the right-hand side of (1.1), obtaining a new equation with $r-1$ non-constant monomials, satisfied by every element in $\phi^{-1}(A)$. Since we assume $r\geqslant 3$, we must have again $|\phi^{-1}(A)|\leqslant N^*(r-1,K)$.

We conclude that in Case I we have $|\phi^{-1}(A)|\leqslant N^*(r-1,K)$.

Case II: $(A,B)=((a,0),(0,b))$ *or* $(A,B)=((0,a),(b,0))$ *with positive* a,b. This time we apply Lemma 4 and the same argument as in Case I shows that $|\phi^{-1}(A)|\leqslant N^*(r-1,K)$. Thus for every A we have

$$|\phi^{-1}(A)|\leqslant N^*(r-1,K). \tag{5.2}$$

Now we apply Lemma 2 to this situation. If $|\phi(\mathscr{X})\cap(\mathbb{N},0)|\geqslant r+1$, the hypothesis of Lemma 2 is satisfied. Since we assume that v is not K-good, the second alternative of Lemma 2 applies and we get $|\phi^{-1}((\mathbb{N},0))|\leqslant N^*(r-1,K)$. If instead $|\phi(\mathscr{X})\cap(\mathbb{N},0)|\leqslant r$, by (5.2) we get $|\phi^{-1}((\mathbb{N},0))|\leqslant rN^*(r-1,K)$. Thus in any case we have

$$|\phi^{-1}((\mathbb{N},0))|\leqslant rN^*(r-1,K).$$

The same bound holds for $|\phi^{-1}((0,\mathbb{N}))|$, and (5.2) shows that $|\phi^{-1}((0,0))|\leqslant N^*(r-1,K)$. It follows by (4.1) that

$$N=|\mathscr{X}|=|\phi^{-1}((0,0))|+|\phi^{-1}((\mathbb{N},0))|+|\phi^{-1}((0,\mathbb{N}))|\leqslant(2r+1)N^*(r-1,K).$$

Finally, we note that

$$(2r+1)N^*(r-1,K)\leqslant(2r+1)N^*(r-1,K_{r-1})\leqslant(2r+1)(N_{r-1}-1)=N_r-1,$$

completing the proof.

Suppose now that

$$K>\frac{1}{2}((r+1)!)^3N_r^{r+1}. \tag{5.3}$$

Then, since no v is K-bad for $\mathscr{X}$, Lemma 1 shows that v cannot be K-good for $\mathscr{X}$. By Lemma 5, Corollary, we infer that v must be strongly neutral for $\mathscr{X}$.

Lemma 6. *If every v is strongly neutral for $\mathscr{X}$ then all elements in $\mathscr{X}$ are proportional.*

Proof. If v is strongly neutral, $v(m_\sigma)$ is constant for every σ, thus $v(m_\sigma/m_\tau) = 0$ for every σ and τ. Hence if every v is strongly neutral we have $v(m_\sigma/m_\tau) = 0$ for every v, thus m_σ/m_τ has height 0 and therefore must lie in k^*. Choose σ to be the identity and τ to be the transposition in (1, 2). Then

$$m_\sigma/m_\tau = -(\mathbf{x}_1/\mathbf{x}_2)^{\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2} \in k^*$$

therefore

$$(x_1/x_2)^{\alpha_{1i}-\alpha_{1j}}(y_i/y_2)^{\alpha_{2i}-\alpha_{2j}} \in k^* \tag{5.4}$$

for any choice of i,j. If we take two such relations and eliminate either x_1/x_2 or y_1/y_2 we find

$$(x_1/x_2)^D \in k^*, \quad (y_1/y_2)^D \in k^* \tag{5.5}$$

with

$$D = \det\begin{pmatrix} \alpha_{1i} - \alpha_{1j} & \alpha_{2i} - \alpha_{2j} \\ \alpha_{1p} - \alpha_{1q} & \alpha_{2q} - \alpha_{2q} \end{pmatrix}, \tag{5.6}$$

for any choice of i,j,p,q.

If one of the determinants (5.6) is not 0 then (5.5) shows that $\mathbf{x}_1$ and $\mathbf{x}_2$ are proportional.

Suppose instead every determinant (5.6) vanishes. Let $h = \mathrm{GCD}(\alpha_{11} - \alpha_{12}, \alpha_{22} - \alpha_{21})$ and write $ha = \alpha_{11} - \alpha_{12}$ and $hb = \alpha_{22} - \alpha_{21}$ with a and b coprime positive integers. The vanishing of the determinant with $(i,j,p,q) = (1,2,1,q)$ shows that

$$ha \cdot (\alpha_{2q} - \alpha_{21}) = hb \cdot (\alpha_{11} - \alpha_{1q})$$

therefore, since a and b are coprime, we have

$$\alpha_{2q} - \alpha_{21} = t_q b, \quad \alpha_{11} - \alpha_{1q} = t_q a \tag{5.7}$$

for certain positive integers t_q. We substitute (5.7) into (5.4) (with $(i,j) = (1,q)$) obtaining $y_2^b/x_2^a = \lambda \cdot (y_1^b/x_1^a)$, where $\lambda \in k^*$. Since x_i and y_i are coprime, we infer that $\mathbf{x}_1$ and $\mathbf{x}_2$ are proportional, completing the proof.

Since all solutions in $\mathscr{X}$ were non-proportional by hypothesis, Lemma 6 leads to a contradiction, hence our initial assumption that we could choose a set of N_r primitive non-proportional solutions cannot hold. It follows that if $r \geqslant 3$ we have

$$N^*(r, K_r) \leqslant N_r - 1$$

with K_r and N_r determined inductively by (5.1) and $N_2 = N^*(2, K_2) + 1$, provided K_2 verifies (5.1).

It remains to bound $N(2, K_2)$. Now (1.1) becomes $a_1 x^{\alpha_{11}} y^{\alpha_{21}} + a_2 x^{\alpha_{12}} y^{\alpha_{22}} = c$. If we had $c = 0$ then we obtain that $\mathbf{x}^{\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2}$ is independent of $\mathbf{x}$, therefore setting $a = \alpha_{11} - \alpha_{12}$ and $b = \alpha_{22} - \alpha_{21}$ with a,b positive integers we obtain $y_2^b/x_2^a = y_1^b/x_1^a$ for any two solutions $\mathbf{x}_1$ and $\mathbf{x}_2$. As noted before, this implies proportionality of $\mathbf{x}_1$ and $\mathbf{x}_2$ because x_i and y_i are coprime. Thus the case $c = 0$ does not occur.

Hence suppose $c \neq 0$, so that $m = 3$. We claim that in this case the last alternative in any one of Lemmas 2, 3, or 4 does not occur. Otherwise we would obtain two solutions $\mathbf{x}_1$ and $\mathbf{x}_2$ such that for example $\mathbf{x}_1^{\alpha_1} = \mathbf{x}_2^{\alpha_1}$, which using (1.1) would imply $\mathbf{x}_1^{\alpha_2} = \mathbf{x}_2^{\alpha_2}$, and $\mathbf{x}_1^{\alpha_1 - \alpha_2} = \mathbf{x}_2^{\alpha_1 - \alpha_2}$, with the same contradiction as before.

Now the same argument in the proof of Lemma 5 yields $N(2, K_2) \leqslant 5$ provided $K_2 > \frac{1}{2} 6^6$. The result follows by taking $N_2 = 6$.

Acknowledgement. During the period this research was done the second author was partially supported by NSA Grant MDA 904-94-H-2004.

References

[1] Bombieri E, Mueller J (1995) Trinomial equations in function fields. Astérisque **228**: 19–40
[2] Brownawell EWD, Masser DW (1986) Vanishing sums in function fields. Math Proc Cambridge Phil Soc **100**: 427–434
[3] Lang S (1991) Number Theory III. Berlin Heidelberg New York: Springer
[4] Mueller J (1990) Binomial Thue's equation over function fields. Composition Math **73**: 189–197
[5] Mueller J (1990) On binomial equations over function fields and a conjecture of Siegel. In: Berndt BC, Diamond HG, Halberstam H, Hildebrand A (eds) Analytic Number Theory, Conference Proceedings in honor of Paul T. Bateman, pp 383–393. Basel: Birkhäuser
[6] Mueller J, Schmidt WM (1988) Thue's equation and a conjecture of Siegel. Acta Math **160**: 207–247
[7] Siegel CL (1929) Über einige Anwendungen diophantischer Approximationen. Abh Preuß Akad Wissenschaften, Phys-math Klasse Nr. 1. Also in: Gesammelte Abh, pp 209–274 Berlin Heidelberg New York: Springer 1966

E. Bombieri
School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540
USA

J. Mueller
Department of Mathematics
Fordham University
Bronx, NY 10458
USA

Mh. Math. 125, 309–314 (1998)

Monatshefte für
Mathematik

Printed in Austria

New Examples of Weakly Symmetric Spaces

By

J. C. González-Dávila*, La Laguna, and **L. Vanhecke,** Leuven

(Received 8 March 1996; in revised form 2 October 1996)

Abstract. We provide some new examples of weakly symmetric spaces inside the class of complete, simply connected Riemannian manifolds equipped with a complete unit Killing vector field such that the reflections with respect to its flow lines are global isometries.

1. Introduction

Weakly symmetric spaces have been introduced by A. SELBERG [16] who proved that for such homogeneous Riemannian manifolds the algebra of all isometry-invariant differential operators is commutative, that is, they are commutative spaces. But only a few non-symmetric examples were known.

The study of the geometry of weakly symmetric spaces has been started in [2] and [6] where also a series of examples is given (see also [5], [13]). This work has been continued in [1] where it is proved that any maximal geodesic in a weakly symmetric space is an orbit of a one-parameter group of isometries of that space. Further examples and new characterizations have recently been given in [3], [4], [7], [11], [12], [15], [18], [21]. In [14] examples of commutative spaces which are not weakly symmetric are provided. So, the class of weakly symmetric spaces is a proper subclass of that of the commutative spaces and this provides an answer to a question stated by SELBERG in [16].

The main purpose of this note is to provide some new examples. They belong to the class of Riemannian manifolds equipped with a complete unit Killing vector field such that the reflections with respect to the flow lines are global isometries. Several aspects of the global and local geometry of such spaces, including classification theorems, have been treated in [8], [9].

In section 2 we give some preliminaries and we provide the examples in section 3.

2. Preliminaries

Let (M, g) be a connected, n-dimensional, smooth Riemannian manifold with $n \geqslant 2$. Let ∇ denote its Levi Civita connection and R the corresponding

1991 Mathematics Subjects Classification: 53C25, 53C30

Key words: Weakly symmetric spaces, killing transversally symmetric spaces, normal flow space forms

* Supported by the "Consejería de Educación del Gobierno de Canarias"

Riemannian curvature tensor with the sign convention

$$R_{UV} = \nabla_{[U,V]} - [\nabla_U, \nabla_V]$$

for all $U, V \in \mathfrak{X}(M)$, the Lie algebra of smooth vector fields on M.

Instead of giving the original definition of [16] for a weakly symmetric space we use the following appealing geometrical characterization, derived in [6]: *(M, g) is weakly symmetric if and only if for any two points p, q in M there exists an isometry interchanging them.* This criterion is equivalent to the following one [2], which justifies the name *ray symmetric* spaces [17] for them: *(M, g) is weakly symmetric if and only if for each $m \in M$ and each non-vanishing $v \in T_mM$ there exists an isometry $s = s(m, v)$ of (M, g) such that s is a non-trivial involution on the maximal geodesic γ_v through m and tangent to v, fixing m.*

For our purposes we will need

Proposition 2.1. [4] *Let (M, g) be a simply connected, complete Riemannian manifold. Then (M, g) is weakly symmetric if and only if there exists a metric connection $\overline{\nabla}$ on (M, g) with $\overline{\nabla}$-parallel torsion $\bar{S}$ and curvature tensor $\bar{R}$, and such that at some point $o \in M$ there exists a linear isometry $\varphi_v : T_oM \to T_oM$ for each $v \in T_oM$ which preserves $\bar{R}$ and $\bar{S}$ and for which $\varphi_v v = -v$.*

Next, we summarize some facts about the Killing-transversally symmetric spaces and refer to [8], [9] for more information. So, let (M, g) be a Riemannian manifold equipped with an *isometric* flow [19] $\mathfrak{F}_\xi$ generated by a unit Killing vector field ξ. Then the flow lines are geodesics and the geodesics which are orthogonal to ξ at one of their points are orthogonal to it at all of their points. These last geodesics are called *transversal* or *horizontal* geodesics and vectors orthogonal to ξ are called *horizontal* vectors.

Now, put $HU = -\nabla_U \xi$ and $h(U, V) = g(HU, V)$ for all $U, V \in \mathfrak{X}(M)$. Since ξ is a Killing vector field, it follows that h is skew-symmetric and moreover, $h = -d\eta$ where η is the metric dual one-form of ξ. η is a contact form if and only if H has maximal rank $n - 1$, in which case $n = \dim M$ is odd. This motivates

Definition 2.1. An isometric flow $\mathfrak{F}_\xi$ on (M, g) is called a *contact flow* if η is a contact form.

Among the isometric flows, the following ones take a particular interesting place [8].

Definition 2.2. A *normal flow* $\mathfrak{F}_\xi$ on (M, g) is an isometric flow such that $R(X, Y, X, \xi) = 0$ for all horizontal vectors X, Y.

In this case, we have the following identities for R:

$$R_{UV}\xi = \eta(V)H^2U - \eta(U)H^2V, \tag{2.1}$$

$$R_{U\xi}V = g(HU, HV)\xi + \eta(V)H^2U, \tag{2.2}$$

$U, V \in \mathfrak{X}(M)$.

Now we turn to the definition of the special case we want to consider.

Definition 2.3. Let (M, g) be a Riemannian manifold and ξ a non-vanishing complete Killing vector field on it. Then $(M, g, \mathfrak{F}_\xi)$ is said to be a (*globally*) *Killing-transversally symmetric space* (briefly, a KTS-*space*) if and only if for each $m \in M$ there exists a (unique) global isometry $s_m : M \to M$ such that

$$(s_m)_*(m) = (-I + 2\eta \otimes \xi)(m)$$

on T_mM.

Note that this means that the reflections with respect to the flow lines are global isometries. Moreover, it follows that ξ is a unit vector field and $\mathfrak{F}_\xi$ is normal.

KTS-spaces are homogeneous manifolds [9]. ξ is a regular vector field and the orbit space $M' = M/\xi$ admits a unique structure of differentiable manifold such that the natural projection $\pi : M \to M'$ is a submersion. Moreover, with the metric g' defined by

$$g(X, Y) = g'(\pi_* X, \pi_* Y) \circ \pi$$

for horizontal X, Y, π becomes a Riemannian submersion. Further, (M', g') is a symmetric space [9] and π interwines the reflections s_m of M with geodesic symmetries $s_{m'=\pi(m)}$ of (M', g').

Now we consider a particular important class of KTS-spaces.

Definition 2.4. A *contact* KTS-*space* is a KTS-space such that η is a contact form.

Contact KTS-spaces are always irreducible [8]. Moreover, a reducible simply connected KTS-space is a Riemannian product of a Riemannian symmetric space and a contact KTS-space [9].

As concerns the base space (M', g') of a contact KTS-space we have

Proposition 2.2. [9] *The base space (M', g') of a simply connected contact KTS-space $(M, g, \mathfrak{F}_\xi)$ is a (simply connected) Hermitian symmetric space. Moreover, we have:*

(i) *if $M' = M'_0 \times M'_1 \times \cdots \times M'_r$ is its de Rham decomposition and $\mathscr{H}_i, i = 0, 1, \ldots, r$, are the smooth distributions on M obtained by the horizontal lifts of the tangent vectors to M'_i, then, for each $m \in M$, $\mathscr{H}(m) = \mathscr{H}_0(m) \oplus \mathscr{H}_1(m) \oplus \cdots \oplus \mathscr{H}_r(m)$ is an H-invariant orthogonal decomposition of the horizontal subspace $\mathscr{H}(m)$;*

(ii) *each sectional curvature $K(H_j, \xi), j = 1, \ldots, r$, is a positive constant c_j^2;*

(iii) *the* (1,1)*-tensor field*

$$J = J_0 \times \frac{1}{c_1} H'_1 \times \cdots \times \frac{1}{c_r} H'_r,$$

is a Hermitian structure on (M', g'), where J_0 denotes the canonical almost complex structure on $M'_0 = \mathbb{C}^p = E^{2p}(x_1, \ldots, x_{2p})$ and $H'_j = H' \circ p_j, j = 1, \ldots, r$ where $p_j : M' \to M'_j$ denotes the projection of M' onto M'_j;

(iv) $H_0' = H' \circ p_0$ *on* $E^{2p}(x_1, \ldots, x_{2p})$ *is given by*

$$\left(\begin{array}{ccc|ccc} & & & -\mu_1 & & \\ & 0 & & & \ddots & \\ & & & & & -\mu_p \\ \hline \mu_1 & & & & & \\ & \ddots & & & 0 & \\ & & \mu_p & & & \end{array}\right)$$

for certain positive real numbers $\mu_1, \ldots, \mu_p$.

Finally, we note that, since $\mathfrak{F}_\xi$ is normal, the Riemannian curvature tensors R, R' of g, g', respectively, are related by

$$\begin{aligned}(R'_{X'Y'}Z')^* = {} & R_{X'^*Y'^*}Z'^* - g(HY'^*, Z'^*)HX'^* \\ & + g(HX'^*, Z'^*)HY'^* + 2g(HX'^*, Y'^*)HZ'^*\end{aligned} \tag{2.3}$$

for X', Y', $Z' \in \mathfrak{X}(M')$ and where X'^* denotes the horizontal lift of X'.

3. Weakly Symmetric Contact KTS-Spaces

Now we shall provide some new examples of weakly symmetric spaces. We have:

Theorem 3.1. *The simply connected contact* KTS*-space fibering over products of complex space forms are weakly symmetric spaces.*

We refer to [9] for the construction of such spaces. Further, before giving the proof, we note the following. In [10] we introduced the notion of a *normal flow space form*: An (M, g) equipped with a normal contact flow $\mathfrak{F}_\xi$ is said to be a normal flow space form if the H-sectional curvature is pointwise constant, that is, the sectional curvature of a two-plane $\{X, HX\}$ for a horizontal $X \in T_mM$ is independent of X for each $m \in M$. If in addition the sectional curvature of two-plane $\{\xi, X\}$ for X horizontal is globally constant, then it turns out that (M, g) is homothetic to a Sasakian space form [10]. So, normal flow space forms are generalizations of manifolds (M, g) which are homothetic to Sasakian space forms.

The simply connected, complete normal flow space forms with globally constant H-sectional curvature or equivalently, which are KTS-spaces, have been classified completely in [10]. It turns out that they always fiber over products of complex space forms but are not always homothetic to Sasakian space forms. So, we have:

Corrollary 3.1. *Complete, simply connected normal flow space forms with globally constant H-sectional curvature are weakly symmetric spaces.*

The special case of (M, g) which are homothetic to Sasakian space forms has been considered in [6]. Now we give the

Proof of the Theorem. We shall prove the result by means of Proposition 2.1. Therefore, we consider the *canonical connection* of the isometric flow $\mathfrak{F}_\xi$, that is, the metric connection $\overline{\nabla} = \nabla - T$, where T is the tensor field of type (1, 2) given by

$$T_U V = g(HU, V)\xi + \eta(U)HV - \eta(V)HU$$

for all tangent U, V [8]. $\overline{\nabla}$ is metric since T_U is skew-symmetric. Moreover, $T_U U = 0$ and so its torsion $\bar{S}$ is given by $\bar{S} = -2T$. Since $\overline{\nabla} T = 0$, the corresponding curvature tensor $\bar{R}$ is determined by

$$\bar{R}_{UV} = R_{UV} + [T_U, T_V] - 2T_{T_U V}. \tag{3.2}$$

Finally, $\overline{\nabla}\bar{R} = \overline{\nabla}\bar{S} = 0$. (Note that this shows that the simply connected KTS-spaces are naturally reductive homogeneous spaces. T provides a naturally reductive homogeneous structure [20].)

Now, let M' be the base space of the fibering. For simplicity we suppose that $M' = M'_0 \times M'_1 \times M'_2$ is its de Rham decomposition, where $M'_0 = \mathbb{C}$ and M'_i, $i = 1, 2$, are non-flat complex space forms and where, following Proposition 2.2, we suppose that $H' = \mu J_0 \times c_1 1 \times c_2 J_2$. (The general case may be treated in a completely similar way.)

Next, let $o \in M$ and $a \in T_o M = \mathscr{H}_0(o) \oplus \mathscr{H}_1(o) \oplus \mathscr{H}_2(o) \oplus \mathbb{R}\xi$. Then we have $a = \sum_{i=0}^{2} a_i + \eta(a)\xi$, where $a_i \in \mathscr{H}_i(o)$. Now we suppose that the a_i are non-zero; the other cases may be treated in a similar way. First, let $\eta(a)$ be non-zero. Put

$$\begin{cases} e_1 = \|a_0\|^{-1} a_0, & e_2 = \|H_{a_0}\|^{-1} H_{a_0}, \\ u_1 = \|a_1\|^{-1} a_1, & u_2 = \|H_{a_1}\|^{-1} H_{a_1}, \\ v_1 = \|a_2\|^{-1} a_2, & v_2 = \|H_{a_2}\|^{-1} H_{a_2}, \end{cases}$$

and extend $(e_1, e_2, u_1, u_2, v_1, v_2)$ to an orthonormal basis $\{e_1, e_2, u_{2k-1}, u_{2k}, v_{2l-1}, v_{2l}, \xi;\ 1 \leqslant k \leqslant \dim M'_1,\ 1 \leqslant l \leqslant \dim M'_2\}$ such that

$$\begin{cases} He_1 = \mu e_2, & He_2 = -\mu e_1, \\ Hu_{2k-1} = c_1 u_{2k}, & Hu_{2k} = -c_1 u_{2k-1}, \\ Hv_{2l-1} = c_2 v_{2l}, & Hv_{2l} = -c_2 v_{2l-1}, \\ H\xi = 0. \end{cases}$$

Define a linear isometry φ of $T_o M$ by

$$\begin{cases} \varphi e_1 = -e_1, & \varphi e_2 = e_2, \\ \varphi u_{2k-1} = -u_{2k-1}, & \varphi u_{2k} = u_{2k}, \\ \varphi v_{2l-1} = -v_{2l-1}, & \varphi v_{2l} = v_{2l}, \\ \varphi\xi = -\xi. \end{cases}$$

It is easy to see that $H \circ \varphi = -\varphi \circ H$ and hence from (3.1) we see that φ preserves T and hence the torsion $\bar{S}$. Then it follows from (2.1), (2.2) and (3.2) that φ preserves $\bar{R}$ if and only if

$$\varphi R_{XY} Z = R_{\varphi X \varphi Y} \varphi Z,$$

for all horizontal X, Y, Z. It is now easy to check that this holds by using (2.3) and the well-known expression for the curvature tensor of a complex space form,

taking into account that φ preserves the spaces $\mathscr{H}_j$, $j = 1, 2$. So, Proposition 2.1 yields the existence of an isometry inducing an involution on the geodesic γ_a which fixes o.

Finally, consider the case $\eta(a) = 0$, that is, a is orthogonal to ξ. In this case we do not need Proposition 2.1 because the existence of an isometry inducing an involution on γ_a which fixes o is clear because the reflection with respect to the flow line through o provides such an isometry.

This completes the proof of the required result. □

References

[1] Berndt J, Kowalski O, Vanhecke L (1997) Geodesics in weakly symmetric spaces. Ann Global Anal Geom **15**: 153–156

[2] Berndt J, Prüfer F, Vanhecke L (1995) Symmetric-like Riemannian manifolds and geodesic symmetries. Proc Roy Soc Edinburgh Sect A **125**: 265–282

[3] Berndt J, Ricci F, Vanhecke L (1995) Weakly symmetric groups of Heisenberg type. Differential Geom Appl (to appear)

[4] Berndt J, Tondeur Ph, Vanhecke L, Examples of weakly symmetric spaces in contact geometry. Boll Un Mat Ital (to appear)

[5] Berndt J, Tricerri F, Vanhecke L (1995) Generalized Heisenberg groups and Damek-Ricci harmonic spaces. Lect Notes Math **1598**: Berlin Heidelberg New York: Springer

[6] Berndt J, Vanhecke L (1996) Geometry of weakly symmetric spaces. J Math Soc Japan **48**: 745–760

[7] Berndt J, Vanhecke L (1996) Commutativity and weak symmetry on Riemannian manifolds (preprint)

[8] González-Dávila JC, González-Dávila MC, Vanhecke L (1995) Reflections and isometric flows. Kyungpook Math J **35**: 113–144

[9] González-Dávila JC, González-Dávila MC, Vanhecke L (1996) Classification of Killing-transversally symmetric spaces. Tsukuba J Math **20**: 321–347

[10] González-Dávila JC, González-Dávila MC, Vanhecke L (1996) Normal flow space forms and their classification. Publ Math Debrecen **48**: 151–173

[11] González-Dávila, Vanhecke L (1996) A new class of weakly symmetric spaces (preprint)

[12] Kowalski O, Marinosci RA (1997) Weakly symmetric spaces in dimension five. J Geom **58**: 128–131

[13] Kowalski O, Prüfer F, Vanhecke L (1996) D'Atri spaces. In: Gindikin S (ed) Topics in Geometry: Honoring the Memory of J D'atri, pp 241–284. Boston: Birkhäuser

[14] Lauret J (1996) Commutative spaces which are not weakly symmetric (preprint)

[15] Nagai S (1995) Weakly symmetric spaces in complex and quaternionic space forms. Arch Math (Basel) **65**: 342–351

[16] Selberg A (1956) Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. J Indian Math Soc (NS) **20**: 47–87

[17] Szabó ZI (1993) Spectral theory for operator families on Riemannian manifolds. In: Greene R, Yau ST (eds) (1990) Differential Geometry. Proc Summer Research Institute on Differential Geometry. Univ of California, Los Angeles, vol 54, part 3

[18] Tamaru H (1996) Isotropy representations of weakly symmetric spaces (preprint)

[19] Tondeur Ph (1988) Foliations on Riemannian manifolds. Berlin Heidelberg New York: Springer

[20] Tricerri F, Vanhecke L (1983) Homogeneous Structures on Riemannian Manifolds. London Math Soc Lec Note Ser **83**: Cambridge: Univ Press

[21] Ziller W (1996) Weakly symmetric spaces. In: Gindikin S (ed) Topics in Geometry: Honoring the Memory of J D'Atri, pp 355–368. Boston: Birkhäuser

J. C. González-Dávila
Departamento de Matemática Fumdamental
Sección de Geometría y Topología
Universidad de La Laguna
La Laguna
Spain

L. Vanhecke
Department of Mathematics
Katholieke Universiteit Leuven
Celestijnenlaan 200 B
B-3001 Leuven
Belgium

Mh. Math. 125, 315–326 (1998)

Monatshefte für
Mathematik

Printed in Austria

Full Ideals of Polynomial Rings

By

Alexander Kreuzer*, München, and **Carl J. Maxson**, College Station, TX

(Received 9 July 1996; in revised form 11 February 1997)

Abstract. An ideal I of the ring $K[x_1, \ldots, x_n]$ of polynomials over a field K in n indeterminates is a full ideal if I is closed under substitution, $f \in I$, $g_1, \ldots, g_n \in K[x_1, \ldots, x_n]$ imply $f(g_1, \ldots, g_n) \in I$. In this paper we continue the investigation of full ideals of $K[x_1, \ldots, x_n]$. In particular we determine several classes of full ideals of $K[x, y]$ (K a finite field) and investigate properties of these classes.

1. Introduction

Let R be a commutative ring with identity 1 and $(R[x], +, \cdot)$ the ring of polynomials in a single indeterminate over R. If "$\circ$" denotes substitution or composition of these polynomials, then $(R[x], +, \cdot, \circ)$ is a composition ring or a 1-place tri-operational algebra (1-TOA) of the school of KARL MENGER [3]. When considering rings of polynomials $R[x_1, \ldots, x_n]$ in more than one indeterminate there is no longer the binary operation of composition but there is an $(n+1)$-ary operation of substitution, $(f, g_1, \ldots, g_n) \mapsto f(g_1, \ldots, g_n)$ where $f, g_1, \ldots, g_n \in R[x_1, \ldots, x_n]$. If we denote this operation by "$*$" then we have $(R[x_1, \ldots, x_n], +, \cdot, *)$, an n-place tri-operational algebra (n-TOA) of MENGER [3]. We refer the reader to the book by LAUSCH and NÖBAUER [2], and the references given there for details of Menger's work and subsequent work by Nöbauer on tri-operational algebras.

When investigating these algebras one needs to know about homomorphic images or quotient structures and so one is led to the concept of a full ideal. Menger and his school studied ideals of 1-TOA and completely determined the full ideals of $(R[x], +, \cdot, \circ)$ when R is a field. Full ideals of $R[x_1, \ldots, x_n]$ were first investigated by Nöbauer and later by STUEBEN [5] and MLITZ [4]. These authors find that a full ideal F of $(R[x_1, \ldots, x_n], +, \cdot, *)$ can be bounded above and below (via set inclusion) using the concept of an enclosing ideal developed by Nöbauer. However, in general, no explicit description of full ideals of $R[x_1, \ldots, x_n]$ is known.

In this paper we find that there are several rather diverse classes of full ideals even in the case $K[x, y]$ where K is a finite field. We focus on this case and determine several chains of full ideals. We show that $K[x, y]$ has a unique maximal full ideal and determine a generating set. We associate a minimal polynomial to

1991 Mathematics Subject Classification: 13B25, 13F20, 16W99, 16Y30
Key words: Full ideals, polynomial rings
* The first author gratefully acknowledges support from the *Deutsche Forschungsgemeinschaft*

each full ideal but this does not determine the ideal. We find that there may be several "hidden" polynomials in some full ideals.

In the next section we give background material and basic results. We also introduce the concept of a minimal polynomial for a full ideal.

2. Full Ideals in $R[x_1, \ldots, x_n]$

Throughout this section we let R denote a commutative ring with identity, 1, and we let $R[x_1, \ldots, x_n]$ denote the ring of polynomials in n indeterminates over R. For $f \in R[x_1, \ldots, x_n]$, $f = \Sigma_{a_{i_1 \ldots i_n}} x_1^{i_1} \ldots x_n^{i_n}$, let $\hat{f} : R^n \to R$ denote the evaluation function $(a_1, \ldots, a_n) \mapsto \Sigma_{a_{i_1 \ldots i_n}} a_1^{i_1} \ldots a_n^{i_n} \in R$. Moreover, for $f = \Sigma_{a_{i_1 \ldots i_n}} x_1^{i_1} \ldots x_n^{i_n} \in R[x_1, \ldots, x_n], f \neq 0$, define $\deg f := \max \{i_1 + \cdots + i_n \mid a_{i_1 \ldots i_n} x_1^{i_1} \ldots x_n^{i_n}$ appears as a nonzero summand of $f\}$. As usual, if $Y \subseteq R[x_1, \ldots, x_n]$ we let $\langle Y \rangle$ denote the (ring) ideal of $R[x_1, \cdots, x_n]$ generated by Y. Of course if $Y = \{f_1, \ldots, f_t\}$, $\langle Y \rangle = R[x_1, \ldots, x_n] f_1 + \cdots + R[x_1, \ldots, x_n] f_t$.

For $f, g_1, \ldots, g_n \in R[x_1, \ldots, x_n]$ we have the $(n+1)$-ary operation, $*$, given by $*(f, g_1, \ldots, g_n) = f(g_1, \ldots, g_n)$, i.e., the substitution of g_i for x_i in f. Straightforward computations show that for $f, h, g_1, \ldots, g_n \in R[x_1, \ldots, x_n]$, $(f+h)(g_1, \ldots, g_n) = f(g_1, \ldots, g_n) + h(g_1, \ldots, g_n)$ and $(f \cdot h)(g_1, \ldots, g_n) = f(g_1, \ldots, g_n) \cdot h(g_1, \ldots, g_n)$. We have the $(n+1)$-tri-operational algebra $(R[x_1, \ldots, x_n], +, \cdot, *)$ which we denote by $R[X]$.

A ring ideal I of $R[x_1, \ldots, x_n]$ is a *full ideal* of $R[X]$ if $R[X]/I$ is also an $(n+1)$-TOA under the natural coset operations. Thus full ideals of $R[X]$ correspond to homomorphic images of $R[X]$. The following result presents an elementwise characterization of full ideals.

Theorem 2.1 ([2], 3.5.3) ([4], Lemma 2). *An ideal I of $R[x_1, \ldots, x_n]$ is a full ideal of $R[X]$ if and only if for each $f \in I$, for all $g_1, \ldots, g_n \in R[x_1, \ldots, x_n]$, $f(g_1, \ldots, g_n) \in I$.*

We remark that full ideals arise in algebraic geometry. For suppose $V \subseteq K^n$ is an affine variety for a field K and $I(V) = \{f \in K[x_1, \cdots, x_n] | \hat{f}(V) = \{0\}\}$ is the associated ideal of V. An endomorphism φ of the coordinate ring $K[V] \cong K[x_1, \ldots, x_n]/I(V)$ of V is determined by an n-tuple $(g_1, \ldots, g_n)$ in $(K[x_1, \ldots, x_n])^n$ such that $f(g_1, \ldots, g_n) \in I(V)$ for each $f \in I(V)$ ([1, p. 35]). Thus every $(g_1, \ldots, g_n)$ determines an endomorphism if and only if $I(V)$ is a full ideal of $K[X]$.

If we let $\mathscr{F}$ denote the collection of all full ideals of $R[X]$ then for $F_1, F_2 \in \mathscr{F}$, $F_1 + F_2$, $F_1 \cap F_2$, and $F_1 \cdot F_2$ are in $\mathscr{F}$ ([2], 3.5.51 and 3.6.11). Thus $(\mathscr{F}, +, \cap)$ is a bounded lattice and $(\mathscr{F}, +, \cdot)$ a semiring of full ideals.

Let I be an ideal of R. Then $F_I := \{f \in R[X] | \hat{f}(a_1, \ldots, a_n) \in I$ for all $(a_1, \ldots, a_n) \in R^n\}$ is a full ideal of $R[X]$ ([2], 3.5.35; [4], Lemma 4). We note that $I \subseteq F_I$ where we consider the elements of I as polynomials of degree zero in $R[X]$. The next three results were obtained by Mlitz.

Lemma 2.2 [4]. *If I is a maximal ideal of R then F_I is a maximal full ideal of $R[X]$.*

Lemma 2.3 [4]. *If F is a maximal full ideal of $R[X]$ then there exists a maximal ideal, I, of R such that $F = F_I$.*

The above two lemmas are combined to obtain the following characterization of maximal full ideals in $R[X]$.

Theorem 2.4 ([4], Theorem 5). *There is a bijection between the maximal ideals of R and the maximal full ideals of $R[X]$.*

We next introduce the concept of minimal polynomial of a nonzero full ideal. We say a nonzero polynomial f in a nonzero full ideal F of $R[X]$ is a *minimal polynomial of F* if

i) $f \in R[x]$ for some $x \in \{x_1, \ldots, x_n\}$, i.e., f is a polynomial in a single indeterminate;

ii) f is monic, i.e., the leading coefficient of f is 1; and

iii) $\deg f = \min \{\deg g \mid g \in F \cap R[x], g \neq 0\}$.

We note that for every $y \in \{x_1, \ldots, x_n\}, f(y) \in F$ so the above definition does not depend on the choice of $x \in \{x_1, \ldots, x_n\}$.

Lemma 2.5. *There is at most one minimal polynomial f for a nonzero full ideal F of $R[X]$.*

Proof. If f and g satisfy the above definition then $\deg (f - g) < \deg f$ and $f - g \in F \cap R[x]$. Thus $f - g = 0$. □

Therefore, when a full ideal has a minimal polynomial we denote the unique minimal polynomial by $f_{\min}$. We will show in (2.10) for polynomials over fields that minimal polynomials always exist. It should be noted that in general when $f_{\min}$ exists in a full ideal F, $f_{\min}$ need not be the polynomial of F with lowest degree. We will see in Section 4, Theorem 4.10 that there can exist polynomials $h \in F$ with $0 < \deg h < \deg f_{\min}$.

Lemma 2.6. *Let F be a full ideal of $R[X]$. Then for any $x \in \{x_1, \ldots, x_n\}$ $I := F \cap R[x]$ is a full ideal of $(R[x], +, \cdot, \circ)$. Further, a polynomial $f \in I$ is the minimal polynomial of I if and only if f is the minimal polynomial of F.*

Proof. Clearly, I is an ideal of $R[x]$ and if $f \in I$, $g \in R[x]$, $f \circ g = f(g) \in I$. If $f_{\min} \in F$, then $f_{\min} \in F \cap R[x] = I$ and since $f_{\min}$ is uniquely determined the result follows. □

For $x \in \{x_1, \ldots, x_n\}$, define $\psi : R[X] \to R[x]$, $f(x_1, \ldots, x_n) \mapsto f(x) := f(x, \ldots, x)$. ψ is an epimorphism of $(R[x_1, \ldots, x_n], +, \cdot)$ onto $(R[x], +, \cdot)$.

Lemma 2.7. *For full ideals F_1, F_2 of $R[X]$, $\psi(F_1) = F_1 \cap R[x]$ and $\psi(F_1 \cap F_2) = \psi(F_1) \cap \psi(F_2)$.*

Proof. If $f \in F_1 \cap R[x]$, $\psi(f) = f$ so $F_1 \cap R[x] \subseteq \psi(F_1)$. If F is a full ideal, for $f \in F$, $\psi(f) = f(x, \ldots, x) \in F \cap R[x]$, hence $\psi(F) \subseteq F \cap R[x]$. Thus, $\psi(F_1 \cap F_2) = F_1 \cap F_2 \cap R[x] = \psi(F_1) \cap \psi(F_2)$. □

Corollary 2.8. *A polynomial f is the minimal polynomial of a full ideal $F = F_1 \cap F_2 \cap \ldots \cap F_t$ if and only if f is the minimal polynomial of $\psi(F) = \psi(F_1) \cap \psi(F_2) \cap \ldots \cap \psi(F_t)$.*

Proof. Apply the previous two lemmas. □

Theorem 2.9. *For every nonzero full ideal F of $R[X]$, $\psi(F) \neq \{0\}$ where R is an integral domain.*

Proof. We use induction on the number k of indeterminates in a nonzero polynomial $f \in F$. If $k = 1$, there is a nonzero polynomial f in a single indeterminate, hence $\psi(f) = f \in \psi(F)$. Suppose now that whenever F contains a nonzero polynomial with at most $k-1$ indeterminates then F, and therefore $\psi(F)$ has a nonzero polynomial in a single indeterminate. Let f be a nonzero polynomial in F and k indeterminates. Now f can be considered as a polynomial in a single indeterminate, say x_1, over the integral domain $R[x_2, \ldots, x_n]$. There is a positive integer l such that x_2^l is not a root of f, i.e., $\hat{f}(x_2^l, x_2, \ldots, x_k) \neq 0$. We now have a nonzero polynomial in F with at most $k-1$ indeterminates and so by the induction hypothesis F has a nonzero polynomial in a single indeterminate $x \in \{x_1, \ldots, x_n\}$, i.e., $\psi(F) \neq \{0\}$. □

Corollary 2.10. *If R is a field, every nonzero full ideal F of $R[X]$ has a minimal polynomial.*

Proof. By Theorem 2.9, $\psi(F)$ is a nonzero ideal in $R[x]$ and thus has a nonzero polynomial with minimal degree. Since R is a field, there is a unique monic polynomial $f_{\min}$ in $\psi(F)$ with minimal degree. By Lemma 2.6, $f_{\min}$ is the minimal polynomial of F. □

3. Full Ideals of $K[x, y]$

In this section and for the remainder of the paper we restrict our coefficient ring to be a field, K. For an infinite field, $K[X]$ has only the trivial full ideals; this result is:

Theorem 3.1 ([2], 3, 7.11). *If K is a infinite field and F is a full ideal of $N := (K[x_1, \ldots, x_n], +, \cdot, *)$ then $F = \{0\}$ or $F = N$.*

As a consequence of this theorem we henceforth take K to be a finite field. Let p be an arbitrary but fixed prime and q a power of p. We let $K = \mathrm{GF}(q)$, the finite field of characteristic p and order q. For ease of exposition we restrict ourselves to two indeterminates x, y. We recall from the previous section that there is a unique maximal full ideal $F_{\max} = \{f \in K[x, y] \mid \hat{f}(a_1, a_2) = 0 \text{ for all } a_1, a_2 \in K\}$. Further we note that

$$k(x) = x^q - x$$

is the unique monic polynomial of least degree in $K[x]$ with $\hat{k}(a) = 0$ for each $a \in K$.

For $e \in \mathbb{N}$ let $m_e(x) = x^{q^e} - x$.

In $K = \mathrm{GF}(q)$, for every prime power $r = p^j$, $j \in \mathbb{N}$, we have $k^r(x) = (x^q - x)^r = x^{qr} - x^r$ since $\binom{r}{i} = 0$ in $\mathrm{GF}(q)$, $i = 1, 2, \ldots, r-1$. Therefore

$$m_e(x) = \sum_{i=0}^{e-1} k^{q^i}(x)$$

since $k^{q^{e-1}}(x) + k^{q^{e-2}}(x) + \cdots + k(x) = x^{q^e} - x^{q^{e-1}} + x^{q^{e-1}} - x^{q^{e-2}} + \cdots + x^q - x = x^{q^e} - x$. In particular, since q is a p-power,

$$m_e(f+g) = m_e(f) + m_e(g) \quad \text{for} \quad f, g \in K[x, y].$$

Lemma 3.2. *Let $a, b \in K$, $i, j \in \mathbb{N}$ and $f, g \in K[x, y]$.*
(1) $m_e(ax + by) = am_e(x) + bm_e(y)$;
(2) $m_e(x^i) = x^{i-1}(x^{(q^e-1)(i-1)} + x^{(q^e-1)(i-2)} + \cdots + 1)m_e(x) \in \langle m_e(x)\rangle$.
(3) $m_e(x^i y^j) = y^{jq^e} m_e(x^i) + x^i m_e(y^j) \in \langle m_e(x), m_e(y)\rangle$.

Proof. (1) $m_e(ax + by) = m_e(ax) + m_e(by) = (ax)^{q^e} - ax + (by)^{q^e} - ay = a(x^{q^e} - x) + b(y^{q^e} - y) = am_e(x) + bm_e(y)$ since $a^q = a$ for all $a \in K$, hence $a^{q^e} = a$.

(2) $m_e(x^i) = x^{iq^e} - x^i = x^i(x^{i(q^e-1)} - 1) = x^i(x^{q^e-1} - 1)(x^{(i-1)(q^e-1)} + x^{(i-2)(q^e-1)} + \cdots + 1)$.

(3) $m_e(x^i y^j) = x^{iq^e} y^{jq^e} - x^i y^j = x^{iq^e} y^{jq^e} - x^i y^{yq^e} + x^i y^{jq^e} - x^i y^j = y^{jq^e}(x^{iq^e} - x^i) + x^i(y^{jq^e} - y^j) = y^{jq^e} m_e(x^i) + x^i m_e(y^j) \in \langle m_e(x), m_e(y)\rangle$. □

As a consequence of the above lemma, for $f = \Sigma c_{ij} x^i y^j \in K[x, y]$, $m_e(f) = \Sigma c_{ij} m_e(x^i y^i) \in \langle m_e(x), m_e(y)\rangle$. We have established the next result.

Theorem 3.3. *For every $e \in \mathbb{N}$, $f \in K[x, y]$, $m_e(f) \in \langle m_e(x), m_e(y)\rangle$. That is, $\langle m_e(x), m_e(y)\rangle$ is a full ideal of $K[x, y]$.*

Lemma 3.4. *For $e, n \in \mathbb{N}$, $0 \leqslant i \leqslant n$ and $f_1, f_2 \in K[x, y]$, there exist $g_1, g_2, h_1, h_2 \in k[x, y]$ with*

$$m_e^{n-i}(f_1) m_e^i(f_2) = \sum_{r=0}^{n-i} \sum_{s=0}^{i} \binom{n-i}{r} \binom{i}{s} g_1^r g_2^{n-i-r} h_1^s h_2^{i-s} m_e^{s+r}(x) m_e^{n-s-r}(y).$$

Proof. From Theorem 3.3, there exist $g_1, g_2, h_1, h_2 \in K[x, y]$ such that $m_e(f_1) = g_1 m_e(x) + g_2 m_e(y)$ and $m_e(f_2) = h_1 m_e(x) + h_2 m_e(y)$. Therefore $m_e^{n-i}(f_1) m_e^i(f_2) = (g_1 m_e(x) + g_2 m_e(y))^{n-i}(h_1 m_e(x) + h_2 m_e(y))^i$ and the result now follows by computation. □

We fix some further notation. For $e, s, e_1, \ldots, e_n, n, n_1, \ldots, n_s \in \mathbb{N}$, let

$$F(e, n) := \langle m_e^{n-i}(x) m_e^i(y) \mid i = 0, 1, 2, \ldots, n\rangle, \quad \text{and}$$
$$F(e_1, n_1, e_2, n_2, \ldots, e_s n_s) = F(e_1, n_1) \cap \cdots \cap F(e_s, n_s).$$

Theorem 3.5. *In the notation above, $F(e, n)$ is a full ideal of $K[x, y]$ with minimal polynomial $f_{\min} = m_e^n(x)$ and $F(e, n+1) \subseteq F(e, n) \subseteq F(1, n)$. Further $F(e_1, n_1, e_2, n_2, \ldots, e_s, n_s)$ is a full ideal with minimal polynomial $f_{\min} = \operatorname{lcm}\{m_{e_i}^{n_i}(x) \mid i = 1, 2, \ldots, s\}$.*

Proof. To show $F(e, n)$ is a full ideal we must show that $m_e^{n-i}(f_1) m_e^i(f_2)$ is in $F(e, n)$ for any $f_1, f_2 \in K[x, y]$. But this is the content of the previous lemma. Clearly $m_e^n(x) \in F(e, n) \cap K[x]$. It is also clear that $m_e^n(x)$ is the minimal polynomial of $F(e, n) \cap K[x]$. Hence from Lemma 2.7, $m_e^n(x)$ is the minimal polynomial of $F(e, n)$. Moreover, since the generators of $F(e, n+1)$ are multiples of the generators of $F(e, n)$ we have $F(e, n+1) \subseteq F(e, n)$. From the discussion just

prior to Lemma 3.2, $m_e(x) = k(x)\left(1 + \sum_{i=1}^{e-1} k^{q^i-1}(x)\right)$ so $m_e^i(x)$ is a multiple of $k^i(x)$. Thus $F(e, n) \subseteq F(1, n)$. We know $F(e_1, n_1, \ldots, e_s, n_s)$ is a full ideal of $K[x, y]$. Also, by Lemma 2.8, $f_{\min}$ is the minimal polynomial of $F(e_1, n_1, \ldots, e_s, n_s)$ if and only if $f_{\min}$ is the minimal polynomial of $\psi(F(e_1, n_1, \ldots, e_s, n_s) = \psi(F(e_1, n_1)) \cap \cdots \cap \psi(F(e_s, n_s)) = (F(e_1, n_1) \cap K[x]) \cap \cdots \cap (F(e_s, n_s) \cap K[x]) = K[x] \cdot m_{e_1}^{n_1}(x) \cap \cdots \cap K[x] m_{e_s}^{n_s}(x)$. Thus $f_{\min} = \operatorname{lcm}\{m_{e_1}^{n_1}(x), \ldots, m_{e_s}^{n_s}(x)\}$. □

From ([2], 3.7.21), every nontrivial full ideal V of $K[x]$ can be uniquely represented as $V = \langle x^{q^{e_1}} - x\rangle^{a_1} \cap \cdots \cap \langle x^{q^{e_s}} - x\rangle^{a_s}$ with $s > 0$, $e_1 > e_2 > \cdots > e_s$, $a_i > 0$, $i = 1, 2, \ldots, s$ and $a_j > a_i$ if e_j is a proper divisor of e_i. Therefore, from the above theorem, every polynomial f which appears as a minimal polynomial of a full ideal of $(K[x], +, \cdot, \circ)$ also appears as a minimal polynomial of a full ideal of $K[x, y]$.

We also remark that, for the minimal polynomial $f_{\min}$ of $F(e, n)$, the full ideal $F(e, n)$ is neither the minimal nor maximal full ideal with $f_{\min}$ as minimal polynomial. We consider this later in this section and in Section 4.

For computational purpose we give the next lemma.

Lemma 3.6. *Let $n, i, r, s \in \mathbb{N} \cup \{0\}$ with $0 \leqslant i \leqslant n$, $0 \leqslant r \leqslant n - i$ and $0 \leqslant s \leqslant i$. If $p \nmid \binom{n}{i}$ and $p \mid \binom{n}{r+s}$ then $p \mid \binom{i}{s}\binom{n-i}{r}$ where p is a prime integer.*

Proof. We compute

$$\begin{aligned}\binom{i}{s}\binom{n-i}{r}\binom{n}{i} &= \frac{i!}{s!(i-s)!}\,\frac{(n-i)!}{r!(n-i-r)!}\,\frac{n!}{i!(n-i)!}\\ &= \frac{n!}{s!r!(i-s)!(n-r-s-(i-s))!}\,\frac{(r+s)!}{(r+s)!}\,\frac{(n-r-s)!}{(n-r-s)!}\\ &= \binom{r+s}{r}\binom{n}{r+s}\binom{n-r-s}{i-s}.\end{aligned}$$

Since $p \mid \binom{n}{r+s}$ we get $p \mid \binom{i}{s}\binom{n-i}{r}\binom{n}{i}$ and since $p \nmid \binom{n}{i}$ we get $p \mid \binom{i}{s}\binom{n-i}{r}$ as desired. □

We introduce now another collection of full ideals of $K[x, y]$. For $e, n \in \mathbb{N}$, define $E(e, n) := \langle \mu_i m_e^{n-i}(x) m_e^i(y) \mid i = 0, 1, 2, \ldots, n\rangle$ with $\mu_i = 1$ if $p \nmid \binom{n}{i}$ and $\mu_i = 0$ if $p \mid \binom{n}{i}$.

Theorem 3.7. *For $e, n \in \mathbb{N}$, $E(e, n)$ is a full ideal with minimal polynomial $f_{\min} = m_e^n(x)$. Further, $E(e, n+1) \subseteq E(e, n) \subseteq F(e, n)$.*

Proof. Let $i \in \{0, 1, \ldots, n\}$ with $\mu_i = 1$, i.e., $p \nmid \binom{n}{i}$. We show for f_1,

$f_2 \in K[x, y]$, $m_e^{n-i}(f_1)m_e^i(f_2) \in E(e, n)$. By Lemma 3.4 we have only to show $\mu_{s+r} = 1$ whenever $\binom{n-i}{r}\binom{i}{s}$ is not zero in GF(q), i.e., whenever $p \nmid \binom{n-i}{r}\binom{i}{s}$. From Lemma 3.6, since $p \nmid \binom{n}{i}$, then $p \nmid \binom{i}{s}\binom{n-i}{r}$ implies $p \nmid \binom{n}{s+r}$. Hence $\mu_{r+s} = 1$ as required. It is clear that $E(e, n+1) \subseteq \subseteq E(e, n) \subseteq F(e, n)$. □

There are in general full ideals between $E(e, n)$ and $F(e, n)$. For example, let $\mu : \{0, 1, \ldots, n\} \to \{0, 1\}$ be a function, $i \mapsto \mu_i$, with the following two properties

$$\mu_i = 1, \quad \text{for all } i \text{ with } p \nmid \binom{n}{i} \quad \text{and}$$

$$\mu_{s+r} = 1, \quad \text{for all } s \in \{0, 1, \ldots, i\}, r \in \{0, 1, \ldots, n-i\} \quad \text{with}$$

$$\mu_i = 1 \quad \text{and } p \nmid \binom{i}{s}\binom{n-i}{r}.$$

Note that μ_i is symmetric in that $\mu_i = 1$ if and only if $\mu_{n-i} = 1$ for if $\mu_i = 1$ then for $r = n - i$, $s = 0$, $p \nmid \binom{i}{0}\binom{n-i}{n-i}$ so $\mu_{n-i} = 1$.

Lemma 3.8. *For every such function defined above,* $F_\mu := \langle \mu_i m_e^{n-i}(x)m_e^i(y) \mid i \in \{0, 1, \ldots, n\}\rangle$ *is full ideal of* $K[x, y]$ *with* $E(e, n) \subseteq \subseteq F_\mu \subseteq F(e, n)$.

Proof. To prove that F_μ is a full ideal we take $\mu_i = 1$ and let $f_1, f_2 \in K[x, y]$. Then from the second property in the definition of μ and Lemma 3.4 we see that $m_e^{n-i}(f_1)m_e^i(f_2)$ is in F. Hence F is a full ideal. Clearly $E(e, n) \subseteq F_\mu \subseteq F(e, n)$. □

For $p = 2$ and $n = 8$ we give the following examples for functions μ and associated full ideals F_μ.

$$E(e, 8) = \langle m_e^8(x), m_e^8(y)\rangle \subsetneqq \langle m_e^8(x), m_e^4(x)m_e^4(y), m_e^8(y)\rangle$$
$$\subsetneqq \langle m_e^8(x), m_e^6(x)m_e^2(y), m_e^4(x)m_e^4(y), m_e^2(x)m_e^6(y), m_e^8(y)\rangle \subsetneqq F(e, 8).$$

Also, there are in general full ideals between $F(e, n+1)$ and $F(e, n)$. For example, $E(e, n) + F(e, n+1) \supsetneqq F(e, n+1)$ and if $E(e, n) \neq F(e, n)$, $E(e, n) + \\ + F(e, n+1) \subsetneqq F(e, n)$. For $e = 1$ and $q = 3$ we have

$$F(1, 4) \subsetneqq \langle k^3(x),\ k^2(x)k^2(y),\ k^2(y)\rangle \subsetneqq F(1, 3).$$

We know that $F_{\{0\}}$ is the unique maximal full ideal of $K[x, y]$. We show $F_{\{0\}} = F(1, 1) = \langle m_1(x), m_1(y)\rangle$ which is of course $\langle k(x), k(y)\rangle$. This was also done in [2]. We include our proof since it is somewhat different.

Theorem 3.9 ([2], 3.8.2). $F_{\{0\}} = F(1, 1)$.

Proof. It is clear that $F(1, 1) \subseteq F_{\{0\}}$. To show the reverse inclusion, let $f(x, y) = \sum_{j=1}^{r} y^j f_j(x) + f_0(x) \in F_{\{0\}}$ where $f_j(x) \in K[x]$. Since $\hat{f}(a, 0) = 0$ for all $a \in K$, we have $\hat{f}_0(a) = 0$ for all $a \in K$. Thus f_0 is a multiple of $k(x)$ and so $f \in F_{\{0\}}$ if and only if $f - f_0 \in F_{\{0\}}$. For $s \in \mathbb{N}$, $r \leqslant (s+1)(q-1)$ we write $f - f_0$ as

$$f - f_0 = \sum_{t=1}^{q-1} \sum_{i=0}^{s} y^{i(q-1)+t} f_{i(q-1)+t}(x)$$

where we set $f_j(x) = 0$ for $j > r$. Let $m := q - 1$ and consider for every $t = 1, 2, \ldots, q-1$ the partial sum

$$\begin{aligned} \sum_{i=0}^{s} y^{im+t} f_{im+t(x)} &= y^t \left(\sum_{i=0}^{s} f_{im+t}(x) \right) + \sum_{i=1}^{s} (y^{im+t} - y^t) f_{im+t}(x) \\ &= y^t h_t(x) + g_t(x, y). \end{aligned}$$

Since $y^{im+t} - y^t = y^t(y^{im} - 1) = y^t(y^m - 1)(y^{(i-1)m} + y^{(i-2)m} + \cdots + 1) = y^{t-1}(y^{m+1} - y)(y^{(i-1)m} + \cdots + 1)$ and since $y^{m+1} - y = y^q - y = k(y)$ we have $g_t(x, y)$ a multiple of $k(y)$ and so $g_t(x, y) \in F(1, 1)$ for every t.

Now for $a, b \in K$, $\widehat{(f - f_0)}(a, b) = 0 = \sum_{t=1}^{q-1} b^t \hat{h}_t(a)$ since $\hat{g}_t(a, b) = 0$. If $c \in K$ is an element of order $q - 1$ then the $q - 1$ linear equations $\sum_{t=1}^{q-1} b^t \hat{h}_t(a) = 0$ for $b = c^j$, $j = 1, 2, \ldots, q-1$ have the unique solution $\hat{h}_t(a) = 0$ for every $a \in K$ since the coefficient matrix is a Vandermonde matrix. Hence for each $t \in \{1, 2, \ldots, q-1\}$, $h_t(x)$ is a multiple of $k(x)$ and so $f - f_0 \in F(1, 1)$, consequently $f \in F(1, 1)$ as was to be shown. □

4. Lower Bounds and Hidden Polynomials

In this section we associate a nonnegative integer to each polynomial in $K[x, y]$. This provides a useful tool for our work with full ideals. We use this tool to discover "hidden" polynomials in some full ideals. For the definition of hidden polynomials see the discussion following Corollary 4.9. We first collect some computational results. When we use the word "summand" of a polynomial $f \in K[x, y]$ we assume the unique representation $f = \sum_{i,j} \alpha_{ij} x^i y^j$ and we mean the term $\alpha_{ij} x^i y^j$ for some $(i, j) \in \mathbb{N}^2$.

Lemma 4.1. *Let $e, i, j, n \in \mathbb{N}$, $a, b \in K$.*

1) *The summand of $m_e^{n-i}(x) m_e^i(y)$ with lowest degree is $x^{n-i} y^i$.*

2) *For polynomials $f_i \in K[x, y]$, $i = 0, 1, \ldots, n$ and for $f = \sum_{i=0}^{n} f_i m_e^{n-i}(x) m_e^i(y) \in F(e, n) \setminus \{0\}$, every summand $f_i m_e^{n-i}(x) m_e^i(y)$ of f has degree at least n.*

3) $m_e^j(x + a) m_e^i(y + b) = m_e^j(x) m_e^i(y)$.

4) *Let* $f \in K[x, y] \backslash F(1, 1)$. *Then there exists* $c \in K \backslash \{0\}$ *and* $g \in F(1, 1)$ *with* $f = c + g$ *or there are* $a_1, a_2, b_1, b_2 \in K$ *with* $\hat{f}(a_1, b_1) \neq \hat{f}(a_1 + a_2, b_1 + b_2)$.

5) *Let* $f \in K[x, y]$ *and* $a_1, a_2, b_1, b_2 \in K$ *with* $\hat{f}(a_1, a_2) \neq \hat{f}(a_1 + a_2, b_1 + b_2)$. *Then* $h(x, y) := f(x - a_2, y - b_2) - f(x, y) \in K[x, y] \backslash F(1, 1)$ *and* $\deg h < \deg f$.

Proof. 1) Straightforward calculation since $m_e(x) = x^{q^e} - x$.

2) For $f_i \neq 0$, every summand of $f_i m_e^{n-i}(x) m_e^i(y)$ has degree at least $n - i + i = n$.

3) $m_e(x + a) = (x + a)^{q^e} - (x + a) = x^{q^e} + a^{q^e} - x - a = m_e(x)$.

4) Since $f \notin F(1, 1)$, there exist $a_1, b_1 \in K$ with $\hat{f}(a_1, b_1) = c \in K \backslash \{0\}$. If $\hat{f}(a, b) = c$ for all $a, b \in K$, then $g(x, y) := f(x, y) - c \in F(1, 1)$. If not, then there exists $a_2, b_2 \in K$ with $\hat{f}(a_1, b_1) \neq \hat{f}(a_1 + a_2, b_1 + b_2)$.

5) Since $\hat{h}(a_1 + a_2, b_1 + b_2) \neq 0$, $h \notin F(1, 1)$. Clearly all summands of f with highest degree are the same for $f(x - a_2, y - b_2)$ and thus do not appear in h. Hence $\deg h < \deg f$. □

Theorem 4.2. *Let* F *be a full ideal of* $K[x, y]$ *and for* $d, e \in \mathbb{N}$ *let* $f_i \in K[x, y]$, $i = 0, 1, \ldots, d$ *with* $f_s \notin F(1, 1)$ *for at least one* $s \in \{0, 1, \ldots, d\}$. *If*

$$f = \sum_{i=0}^{d} f_i m_e^{d-i}(x) m_e^i(y) \in F,$$

then there exists $h \in F$ *with* $h = m_e^{d-s}(x) m_e^s(y) = \sum_{i=0}^{d} h_i m_e^{d-i}(x) m_e^i(y)$ *with* $h_i \in K[x, y]$ *and* $h_s \in F(1, 1)$.

In particular h contains $x^{d-s} y^s$ as a summand and $h \notin F(e, d + 1)$.

Proof. We prove the existence of h by induction on the degree d_s of f_s. If $\deg f_s = 0, f_s \in K \backslash \{0\}$ and the result follows by $h = f_s^{-1} \cdot f$ with $h_s = 0$. Suppose the result is true for all f_s with $\deg f_s < d_s$ and now take f with $\deg f_s = d_s$. We will write "d" for "d_s".

If $f_s = c + g$ for some $c \in K \backslash \{0\}$ and $g \in F(1, 1)$, then the result follows by $h = c^{-1} \cdot f$ with $h_s = c^{-1} g \in F(1, 1)$.

On the other hand, if there exist $a_1, a_2, b_1, b_2 \in K$ with $f_s(a_1, b_1) \neq \neq f_s(a_1 + a_2, b_1 + b_2)$ (cf. Lemma 4.1 (4)), then since F is a full ideal, instead of f we may consider $g(x, y) := f(x - a_2, y - b_2) - f(x, y) \in F$ and then by Lemma 4.1 (3), (5), $f_s m_e^{d-s}(x) m_e^s(y)$ will become $g_s m_e^{d-s}(x) m_e^s(y)$ where $g_s(x, y) = = f_s(x - a_2, y - b_2) - f_s(x, y) \in K[x, y] \backslash F(1, 1)$ and $\deg g_s < \deg f_s$. Thus by the induction hypothesis $h \in F$ exists. By Lemma 4.1 (1), $m_e^{d-s}(x) m_e^s(y)$ contains a summand $x^{d-s} y^s$ while $f_i m_e^{d-i}(x) m_e^i(y)$ for $i \neq s$ does not contain $x^{d-s} y^s$.

For $h_s \neq 0$ there are $h_1, h_2 \in K[x, y]$ with $h_s = h_1 k(x) + h_2 k(y)$ for suitable h_1, $h_2 \in K[x, y]$ and every summand of h_s has degree at least 1; by Lemma 4.1 (1) every summand of $h_s m_e^{d-s}(x) m_e^s(y)$ has degree at least $d + 1$. Hence for $h_s = 0$, h contains the summand $x^{d-s} y^s$ of $m_e^{d-s}(x) m_e^s(y)$. By Lemma 4.1 (2), $h \notin F(e, d + 1)$. □

Recall that $F(1, n) = \langle k^{n-i}(x) k^i(y) \mid i = 0, 1, \ldots, n \rangle$ is a full ideal of $K[x, y]$. For simplicity of notation, in the sequel we denote $F(1, n)$ by F_n.

Lemma 4.3. *For every $d \in \mathbb{N}$ and $f \in F(1, 1) = F_1$, $f \neq 0$, with $\deg f = d$ we have $f \notin F_{d+1}$.*

Proof. Let $n \in \mathbb{N}$ and $g \in F_n \backslash \{0\}$, say $g = \sum_{i=0}^{n} f_i k^{n-i}(x) k^i(y)$. From 4.1, every summand of g has degree at least n, i.e., $\deg g \geqslant n$. Thus if $\deg f = d$, $f \notin F_{d+1}$. □

Corollary 4.4. $\bigcap_{n \in \mathbb{N}} F_n = \{0\}$.

We know $F_{n+1} \subseteq F_n$. Moreover from the definitions we have $F_{n+1} \neq F_n$. Thus from the above corollary, for every $f \in F_1 \backslash \{0\}$, there exists a unique integer $d \in \mathbb{N}$ with $f \in F_d \backslash F_{d+1}$. We call this integer d the *lower bound of* f and denote it by $d = \mathrm{lb}(f)$. If $f \in K[x, y] \backslash F_1$ or if $f = 0$ we say $\mathrm{lb}(f) = 0$.

Lemma 4.5. *For $f \in F_1$ with $d = \mathrm{lb(f)}$, there exists $f_i \in K[x, y]$, $i = 0, 1, \ldots, d$* with $f_s \notin F_1$, for at least one $s \in \{0, 1, \ldots, d\}$ and $f = \sum_{i=0}^{d} f_i k^{d-i}(x) k^i(y)$.

Proof. By definition $f \in F_d \backslash F_{d+1}$ and so there are polynomials $f_i \in K[x, y]$ with $f = \sum_{i=1}^{d} f_i k^{d-i}(x) k^i(y)$. If $f_i \in F_1$ for all $i \in \{0, 1, \ldots, d\}$ then $f_i = h_{1i} k(x) + h_{2i} k(y)$, $h_{1i}, h_{2i} \in K[x, y]$. But then $f \in F_{d+1}$, a contradiction. □

Theorem 4.6. *Let $f \in F_1$ and $d \in \mathbb{N}$. Then $\mathrm{lb}(f) = d$ if and only if there exist $f_i \in K[x, y]$ with $f_s \notin F_1$ for at least one $s \in \{0, 1, \ldots, d\}$ and* $f = \sum_{i=0}^{d} f_i k^{d-i}(x) k^i(y)$.

Proof. If $\mathrm{lb}(f) = d$, apply the previous lemma. For the converse we take $f = \sum_{i=0}^{d} f_i k^{d-i}(x) k^i(y)$ and note that $f \in F_d$. Suppose further that $f \in F_{d+1}$. Since $m_1(x) = k(x)$, from Theorem 4.2 there exists an $h \in F_{d+1}$ containing $x^{d-s} y^s$ as a summand. But from Lemma 4.1, every summand of $h \in F_{d+1}$ has degree at least $d + 1$, a contradiction. Therefore $f \notin F_{d+1}$ and $\mathrm{lb}(f) = d$. □

Corollary 4.7. *For $e, n, e_1, \ldots, e_s, n_1, \ldots, n_s \in \mathbb{N}$, $\mathrm{lb}\,(m_e^n(x)) = n$ and* $\mathrm{lb}(m_{e_1}^{n_1}(x) \ldots m_{e_s}^{n_s}(x)) = n_1 + n_2 + \cdots + n_s$.

Proof. From the discussion prior to Lemma 3.2 we see that $m_e(x) = k(x) \times \times \left(1 + \sum_{i=1}^{e-1} k^{q^i - 1}(x)\right)$ so $m_e^n(x) = k^n(x) g(x)$ with $g(x) = \left(1 + \sum_{i=1}^{e-1} k^{q^i - 1}(x)\right)^n$ and $g \notin F_1$. From the previous theorem, $\mathrm{lb}(m_e^n(x)) = n$. Also $m_{e_1}^{n_1}(x) \ldots m_{e_s}^{n_s}(x) = = k^m(x) g(x)$ with $m = n_1 + n_2 + \cdots + n_s$ and $g \notin F_1$, hence the result. □

Corollary 4.8. *If $f, g \in K[x, y] \backslash \{0\}$ then $\mathrm{lb}(fg) \geqslant \mathrm{lb}(f) + \mathrm{lb}(g)$.*

Proof. Let $\mathrm{lb}(f) = d$, $\mathrm{lb}(g) = n$ with $f = \sum_{i=0}^{d} f_i k^{d-i}(x) k^i(y)$ and $g = \sum_{i=0}^{n} g_i k^{n-i}(x) k^i(y)$. Therefore $fg = \sum_{i=0}^{d+n} h_i k^{d+n-i}(x) k^i(y)$ with $h_i \in K[x, y]$.

Since $f,\ g \in K[x, y]\backslash\{0\}$, $fg \neq 0$ and so $fg \in F_{n+d}$. But this means that $\mathrm{lb}(fg) \geqslant n + d$. □

For a full ideal F of $K[x, y]$ we define the *lower bound of* F, $\mathrm{lb}(F)$, by $\mathrm{lb}(F) := \min\{\mathrm{lb}(f) | f \in F\backslash\{0\}\}$.

Corollary 4.9. *For* $e,\ n \in \mathbb{N}$, $\mathrm{lb}(F(e, n)) = \mathrm{lb}(E(e, n)) = n$.

Proof. We know $E(e, n) \subseteq F(e, n) \subseteq F(1, n) = F_n$, hence $\mathrm{lb}(E(e, n)) \geqslant \mathrm{lb}(F(e, n)) \geqslant \mathrm{lb}(F_n)$. By definition $\mathrm{lb}(F_n) = n$ and since $m_e^n(x) \in E(e, n)$, $n \geqslant \mathrm{lb}(E(e, n))$. □

For a full ideal F with minimal polynomial $f_{\min}$ we have $\mathrm{lb}(F) \leqslant \mathrm{lb}(f_{\min})$. We will see in the remainder of this section that $\mathrm{lb}(F) \lneqq \mathrm{lb}(f_{\min})$ is possible.

Again let F be a full ideal. We say a polynomial $f \in F$ is a *hidden polynomial of* F and F *is a full ideal with hidden polynomials* if $\mathrm{lb}(f) \lneqq \mathrm{lb}(f_{\min})$ where, as usual, $f_{\min}$ is the minimal polynomial of F.

Theorem 4.10. *For every* $e,\ n \in N$, $F(e, n)$ *and* $E(e, n)$ *are full ideals without hidden polynomials.*

Proof. From Corollary 4.9, $\mathrm{lb}(F(e, n)) = \mathrm{lb}(E(e, n)) = n = \mathrm{lb}(m_e^n(x))$. Since $m_e^n(x)$ is the minimal polynomial of $F(e, n)$, (cf. Theorems 3.5 and 3.7), and $E(e, n)$ there are no hidden polynomials in these full ideals. □

On the other hand, full ideals with hidden polynomials do exist.

Theorem 4.11. *Let* $h_q := k^q(x)k(y) - k(x)k^q(y)$. *Then* $\overline{F_{q+2}} := \langle F_{q+2} \cup \{h_q\}\rangle$ *is a full ideal with hidden polynomial* h_q.

Proof. Since F_{q+2} is a full ideal, to show that $\overline{F_{q+2}}$ is a full ideal we must show that for $f_1,\ f_2 \in K[x, y]$, $h_q(f_1, f_2) \in \overline{F_{q+2}}$. Since $k(x),\ k(y) \in F(1, 1)$, $k(f_1),\ k(f_2) \in F(1, 1)$, say $k(f_1) = g_1k(x) + g_2k(y)$ and $k(f_2) = h_1k(x) + h_2k(y)$ where $g_1, g_2, h_1, h_2 \in K[x, y]$. Thus $k^q(f_1)k(f_2) = g_1^q h_1 k^{q+1}(x) + g_1^q h_2 k^q(x)k(y) + g_2^q h_1 k(x)k^q(y) + g_2^q h_2 k^{q+1}(y)$ and so $h_q(f_1, f_2) = (g_1^q h_1 - g_1 h_1^q)k^{q+1}(x) + (g_1^q h_2 - g_2 h_1^q)k^q(x)k(y) + (g_2^q h_1 - g_1 h_2^q)k(x)k^q(y) + (g_2^q h_2 - g_2 h_2^q)k^{q+1}(y)$.

Now for any $a,\ b \in K$, if $\hat{g}_1(a, b) =: c$ and $\hat{h}_1(a, b) =: d$ then $\widehat{(g_1^q h_2 - g_1 h_1^q)}(a, b) = c^q d - c d^q = cd - cd = 0$. Thus $g_1^q h_1 - g_1 h_1^q \in F(1, 1) = \langle k(x), k(y)\rangle$, hence $(g_1^q h_1 - g_1 h_1^q)k^{q+1}(x) \in \langle k^{q+2}(x),\ k^{q+1}(x)k(y)\rangle \subseteq F_{q+2} \subseteq \overline{F_{q+2}}$. In the same manner, $(g_2^q h_2 - g_2 h_2^q)k^{q+1}(y) \in \overline{F_{q+2}}$.

Let $g := (g_1^q h_2 - g_2 h_1^q)k^q(x)k(y) + (g_2^q h_1 - g_1 h_2^q)k(x)k^q(y)$ and note that $g = (g_1^q h_2 - g_2 h_1^q)h_q + [(g_1^q h_2 - g_1 h_2^q) + (g_2^q h_1 - g_2 h_1^q)]k(x)k^q(y)$. As above one finds that $g_1^q h_2 - g_1 h_2^q$ and $g_2^q h_1 - g_2 h_1^q$ are in $F(1, 1)$, consequently $g \in \langle h_q, k^2(x)k^q(y),\ k(x)k^{q+1}(y)\rangle \subseteq \overline{F_{q+2}}$. Hence $\overline{F_{q+2}}$ is a full ideal.

It remains to prove that h_q is a hidden polynomial in F. We have $\psi(h_q) = k^{q+1}(x) - k^{q+1}(x) = 0$ where ψ is the ring epimorphism defined just prior to Lemma 2.7. Then, using Lemma 2.7 and the fact that ψ is a ring epimorphism we have $\overline{F_{q+2}} \cap R[x] = \psi(\overline{F_{q+2}}) = \langle \psi(F_{q+2}) \cup \psi(h_q)\rangle = \langle \psi(F_{q+2})\rangle = \psi(F_{q+2})$. We know $m_e^{q+2}(x)$ is the minimal polynomial of F_{q+2} and so by Lemma 2.7, is the minimal polynomial of $\psi(F_{q+2}) = \psi(\overline{F_{q+2}})$. Again, by Lemma 2.7,

$m_e^{q+2}(x)$ is the minimal polynomial of $\overline{F_{q+2}}$. Thus $\mathrm{lb}(h_q) = q+1 < q+2 = \mathrm{lb}(m_e^{q+2}(x))$, hence h_q is a hidden polynomial of F. □

Corollary 4.12. *Let F be a full ideal without hidden polynomials and with $\mathrm{lb}(F) = n$. Then $F \subseteq F_n$.*

Proof. By definition, for each nonzero $f \in F$, $\mathrm{lb}(f) \geqslant n$, hence $f \in F_n$. □

Corollary 4.13. *Let F be a full ideal with minimal polynomial $f_{\min}$ such that $\mathrm{lb}(f_{\min}) = n$. Then $F \backslash F_n$ is a set of hidden polynomials of F.*

Proof. For $f \in F \backslash F_n$, $f \in F(1, 1) \backslash F_n$, hence $\mathrm{lb}(f) < n = \mathrm{lb}(f_{\min})$. □

In the above theorem, the hidden polynomials h_q had the property that $\psi(h_q) = 0$. As our next example shows this need not always be the case.

Example 4.14. Let $g_q := x^q k^q(x)k(y) - xk(x)k^q(y) = k(x)k^q(x)k(y) + xh_q \in \overline{F_{q+2}}$. As in Theorem 4.11 $\mathrm{lb}(g_q) = q+1 < q+2$, so g_q is a hidden polynomial of $\overline{F_{q+2}}$. Moreover, $\psi(g_q) = (x^q - x)k^{q+1}(x) = k^{q+2}x \neq 0$ and $\mathrm{lb}(\psi(g_q)) = q+2 > \mathrm{lb}(g_q)$.

We conclude with a result which shows that the above situations are the only ones that can occur for hidden polynomials.

Theorem 4.15. *Let f be a hidden polynomial of a full ideal F and as above let ψ: $K[x, y] \to K[x]$ be the ring epimorphism, $g \mapsto g(x, x)$. Then $\psi(f) = 0$ or $\mathrm{lb}(\psi(f)) \ngeqslant \mathrm{lb}(f)$.*

Proof. We note that $\psi(f) \in F \cap K[x]$, hence $\psi(f)$ is a multiple of $f_{\min} \in F \cap K[x]$, i.e., $\psi(f) = h \cdot f_{\min}$ with $h \in K[x]$. From Corollary 4.8, $\mathrm{lb}(\psi(f)) \geqslant \mathrm{lb}(f_{\min}) \ngeqslant \mathrm{lb}(f)$ when $\psi(f) \neq 0$. □

References

[1] EISENBUD D (1995) Commutative Algebra with a View Towards Algebraic Geometry. New York: Springer
[2] LAUSCH H, NÖBAUER W (1973) Algebra of Polynomials. Amsterdam: North Holland
[3] MENGER K (1944) Tri-operational algebras. I Rep Math Col **5**: 3–10 Notre Dame
[4] MLITZ R (1971) Ein Radikal für universale Algebren und seine Anwendung auf Polynomringe mit Komposition. Mh Math **75**: 144–152
[5] STUEBEN E (1965) Ideals in two-place tri-operational algebras. Mh Math **69**: 177–182

A. KREUZER
Mathematisches Institut
Technische Universität München
80290 München
Germany
e-mail: kreuzer@mathematik.tu-muenchen.de

C. J. MAXSON
Department of Mathematics
Texas A & M University
College Station, TX 77843-3368
USA
e-mail: cjmaxson@math.tamu.edu

Mh. Math. 125, 327–342 (1998)

Monatshefte für
Mathematik

Printed in Austria

Lifting Covers of Sofic Shifts

By

Paul Trow, Memphis, TN

(Received 23 April 1996; in final form 20 November 1996)

Abstract. We define a class of factor maps between sofic shifts, called lifting maps, which generalize the closing maps. We show that an irreducible sofic shift S has only finitely many S-conjugacy classes of minimal left (or right) lifting covers. The number of these classes is a computable conjugacy invariant of S. Furthermore, every left lifting cover factors through a minimal left lifting cover.

0. Introduction

Sofic shifts are an important class of symbolic dynamical systems. A *cover* of a sofic shift S is a factor map $\phi : \Sigma_A \to S$, where Σ_A is a shift of finite type. In this paper we define a class of covers, called *lifting covers*, which generalize the closing covers, and show that there are only finitely many S-conjugacy classes of minimal lifting covers of a given sofic shift S. The number of such classes is a computable conjugacy invariant of S. We point out that our use of the term "minimal", as defined in the next paragraph, differs from that of [2], but agrees with that of [10].

Given covers $\phi : \Sigma_A \to S$ and $\gamma : \Sigma_B \to S$, from irreducible shifts of finite type onto a sofic shift S, we say that ϕ *factors through* γ if there exists a factor map $\beta : \Sigma_A \to \Sigma_B$ such that $\phi = \gamma\beta$. If β is a topological conjugacy, we say that ϕ is *conjugate over* S to γ. This defines an equivalence relation on covers of S. We will refer to an equivalence class as an *S-conjugacy class*. A cover ϕ is *minimal* if whenever ϕ factors through a cover γ by a map β, it follows that β must be a conjugacy. For example, the left and right Fischer covers are minimal, and in fact the only minimal closing covers, up to conjugacy over S. It follows from [1, Corollary 2.7] that every finite-to-one cover with irreducible domain factors through a minimal cover.

S. Williams has shown that if S is not an almost finite type (AFT) shift, then there are infinitely many finite-to-one covers of S no pair of which factor through a common cover (see [11]). Consequently, S has infinitely many S-conjugacy classes of minimal finite-to-one covers.

We say that a class $\mathscr{C}$ of covers of a sofic shift is *closed under lower factors* if whenever $\phi \in \mathscr{C}$ and $\phi = \gamma\beta$, where γ is a cover, we must have $\gamma \in \mathscr{C}$. For

1991 Mathematics Subject Classification: 58F03
Key words: Shift of finite type, sofic shift, factor map

example, the closing covers have this property (by [3, Prop. 4.10 and 4.11]). If $\mathscr{C}$ is closed under lower factors, then every element of $\mathscr{C}$ factors through a minimal element of $\mathscr{C}$. If, in addition, there are only finitely many S-conjugacy classes of minimal covers in $\mathscr{C}$, then there is a finite set of minimal covers such that any cover in $\mathscr{C}$ factors through one in the finite set. It is natural to ask whether there are other classes of covers (besides the closing covers) which have these properties.

In Definition 2.5, we introduce such a class, the lifting maps. We show that there are only finitely many S-conjugacy classes of minimal left lifting covers of a given sofic shift S (Theorem 2.17). We call the number of such classes the *left lifting number* of S. It is easy to see that this number is a conjugacy invariant of S (see the remark following Proposition 2.8). By Corollary 2.18, there is a finite procedure for constructing a complete set of representatives of these classes, and so the left lifting number is computable. Since left lifting covers are closed under lower factors (Remark 2.7 (iii)), there is a finite set of minimal covers such that every left lifting cover of S factors through one in the finite set (Corollary 2.19). Similar statements hold for right lifting covers. In addition, we obtain a useful necessary condition for the existence of lifting factor maps between sofic shifts (Remark 2.20).

We also define a special type of one-block left lifting map, called a left determining map (Definition 2.9), and show that every left lifting map is conjugate over S to a left determining map (Corollary 2.14). This is a generalization of the fact that every left closing map is conjugate over S to a left resolving map.

Another motivation for introducing lifting maps is the fact that closing maps are not sufficient to describe all factor maps. For example, KITCHENS has shown that there are factor maps which cannot be written as a composition of closing maps ([5] or [6, p. 95]). In fact, if S is a strictly sofic shift, then it follows from [3, Prop. 4.12] that there cannot exist any factor map from S onto a shift of finite type which is a composition of closing maps. In view of this, it seems necessary to search for larger classes of maps, such as lifting maps, in order to try to understand general factor maps.

We thank the referee for suggesting a number of improvements to the paper.

1. Background

We briefly summarize some background material. An excellent source for further reference is the book [7] by LIND and MARCUS. Let A be a non-negative integral square matrix, which defines a directed graph $G(A)$, having A_{ij} edges from vertex i to vertex j. The set of vertices, or *states*, of $G(A)$ is denoted $\mathscr{S}_A$. The matrix A defines a *shift of finite type* Σ_A (see [3, Section 2]).

A *factor map* between symbolic dynamical systems is a continuous, surjective map which commutes with the shift. By the Curtis-Hedlund-Lyndon Theorem [7, Theorem 6.2.9], every factor map $\phi : S_1 \to S_2$ has the form $\phi(x)_i = \bar{\phi}(x_{x-m} \cdots x_{i+a})$, for some non-negative integers m and a, where $\bar{\phi} : \mathscr{A}_1^{m+a+1} \to \mathscr{A}_2$ is a finite block map and $\mathscr{A}_1, \mathscr{A}_2$ are the alphabets of S_1, S_2, respectively. (Usually we will let ϕ denote the map on finite blocks, as well as the map on infinite sequences.) If $m = a = 0$, then ϕ is a *one-block map*, in which case

ϕ defines a labelling of the edges of $G(A)$, where an edge e is labelled $\phi(e)$. A factor map which is injective is a *conjugacy*. A factor map is *finite-to-one* if every point in the range has a uniformly bounded number of preimages. It is *right closing* if it does not identify two distinct points which are left asymptotic under the shift map, and *left closing* if it does not identify two distinct points which are right asymptotic. Left or right closing maps are finite-to-one, by [7, Prop. 8.1.1].

If Σ_A is a shift of finite type and $\phi : \Sigma_A \to S$ is a factor map, then S is a *sofic shift*. The map ϕ is called a *cover* of S. The set of finite blocks of symbols which occur in some point of S is denoted $\mathscr{B}(S)$. Every irreducible sofic shift has a canonical right closing cover $\pi^+ : \Sigma_R \to S$, called the *right Fischer cover*, which has the following minimality property:

Theorem 1.1. ([4]) or [2, Prop. 4]) *Let $\phi : \Sigma_B \to S$ be a right closing factor map from an irreducible shift of finite type onto a sofic shift. Then there exists a unique right closing factor map $\gamma : \Sigma_B \to \Sigma_R$ such that $\pi^+\gamma = \phi$.*

The states of Σ_R are *follower sets* of the form $\mathscr{F}_S(w) = \{v \in \mathscr{B}(S) : wv \in \mathscr{B}(S)\}$, where $w \in \mathscr{B}(S)$ is a magic word. For each symbol a such that $wa \in \mathscr{B}(S)$, there is an edge from $\mathscr{F}_S(w)$ to $\mathscr{F}_S(wa)$ labelled a. There is also a canonical left closing cover $\pi^- : \Sigma_L \to S$ called the *left Fischer cover*, whose states are *predecessor sets* of the form $\mathscr{P}_S(v) = \{w \in \mathscr{B}(S) : wv \in \mathscr{B}(S)\}$, which has a similar minimality property with respect to left closing factor maps.

If $\phi : \Sigma_A \to S$ is a one-block factor map and $i \in \mathscr{S}_A$, we define the *predecessor set of i* to be $\mathscr{P}_\phi(i) = \{w \in \mathscr{B}(S)$ such that there exists $p \in \phi^{-1}(w)$ ending at $i\}$. The *follower set of i* is defined similarly (see [7, Definition 3.3.7]).

Following [7, Definition 3.3.7], we say that a labelled graph $G(A)$, corresponding to a one-block factor map $\phi : \Sigma_A \to S$, is *predecessor-separated* if for any two states $i \neq j$ for Σ_A, we have $\mathscr{P}_\phi(i) \neq \mathscr{P}_\phi(j)$. Similarly to the construction in [7], one can form the *predecessor merged graph* $G(B)$ as follows: two states $i, j \in \mathscr{S}_A$ are *equivalent* if $\mathscr{P}_\phi(i) = \mathscr{P}_\phi(j)$. The states of $G(B)$ are equivalence classes of states of $G(A)$. If $\mathscr{I}$ and $\mathscr{T}$ are equivalence classes, there is an edge f in $G(B)$, labelled a, from $\mathscr{I}$ to $\mathscr{T}$ if there exists $i \in \mathscr{I}$ and $j \in \mathscr{T}$ and an edge e from i to j labelled a. Similarly to the proof of [7, Lemma 3.3.8], it can be shown that the labelling of $G(B)$ defines a factor map $\gamma : \Sigma_B \to S$ which is predecessor-separated. Clearly, the map $\beta : \Sigma_A \to \Sigma_B$ which takes $e \to f$ is a continuous, shift-commuting map, and $\gamma\beta = \phi$. We have the following result, which is well known in symbolic dynamics and automata theory.

Proposition 1.2 *Let $\phi : \Sigma_A \to S$ be a finite-to-one, one-block factor map, where Σ_A is an irreducible shift of finite type. Then there exists a shift of finite type Σ_B, a one-block factor map $\gamma : \Sigma_B \to S$ which is predecessor separated, and a left resolving map $\beta : \Sigma_A \to \Sigma_B$ such that $\gamma\beta = \phi$.*

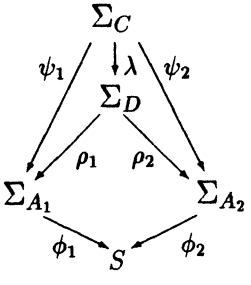
Proof. By the previous remarks, all that remains to prove is that β is left resolving (from which it follows that ϕ is surjective, by [7, Prop. 8.2.2]). Suppose that $\beta(e_1) = \beta(e_2)$, where e_1 and e_2 have the same terminal state. Let i_1 and i_2 be the initial states of e_1 and e_2, respectively. By definition of β, we have $\phi(e_1) = \phi(e_2) = a$, and $\mathscr{P}_\phi(i_1) = \mathscr{P}_\phi(i_2)$. Choose a magic word $m = m_1 \cdots m_n$

for ϕ, with magic coordinate m_k (see [8, Definition 2.1]). Choose $p = p_1 \dots p_n \in \phi^{-1}(m)$. Since Σ_A is irreducible, there are A-words q_1 and r such that $pq_1e_1rp \in \mathscr{B}(\Sigma_A)$. Let $\phi(pq_1e_1rp) = msatm$. Since pq_1 ends at i_1, we have $ms \in \mathscr{P}_\phi(i_1)$. Since $\mathscr{P}_\phi(i_1) = \mathscr{P}_\phi(i_2)$, we have $ms \in \mathscr{P}_\phi(i_2)$, and so there exists a path $p'q_2$ in $G(A)$, ending at i_2, such that $\phi(p'q_2) = ms$. Since e_1 and e_2 have the same terminal state, $p'q_2e_2rp \in \mathscr{B}(\Sigma_A)$. Since $p'q_2e_2rp \in \phi^{-1}(msatm)$, it follows from [8, Lemma 2.3] (see also [6, Lemma 2.4 (2)]) that $p_k = p'_k$. Then we must have $i_1 = i_2$ and $e_1 = e_2$, for otherwise ϕ would have a diamond, contradicting the assumption that ϕ is finite-to-one, by [6, Theorem 1.1]. Therefore β is left resolving. □

We next recall the definition of the fiber product of two factor maps ([8, Def. 8.3.2]). If $\phi_1 : X_1 \to S$ and $\phi_2 : X_2 \to S$ are factor maps between shift spaces, the *fiber product of* ϕ_1 *and* ϕ_2 is the triple (W, ρ_1, ρ_2), where $W = \{(x, y) \in X_1 \times X_2 : \phi_1(x) = \phi_2(y)\}$, and $\rho_1 : W \to X_1, \rho_2 : W \to X_2$ are the projection maps defined by $\rho_1(x, y) = x$ and $\rho_2(x, y) = y$.

Clearly $\phi_1\rho_1 = \phi_2\rho_2$. If X_1 and X_2 are shifts of finite type, then it is not hard to show that W is a shift of finite type. The following lemma is easily established.

Lemma 1.3. *Let $\phi_1 : \Sigma_{A_1} \to S, \phi_2 : \Sigma_{A_2} \to S, \psi_1 : \Sigma_C \to \Sigma_{A_1}$ and $\psi_2 : \Sigma_C \to \Sigma_{A_2}$ be finite-to-one factor maps, where $\Sigma_{A_1}, \Sigma_{A_2}$ and Σ_C are irreducible shifts of finite type, S a sofic shift, and $\phi_1\psi_1 = \phi_2\psi_2$. Then there exists an irreducible component of maximal entropy, Σ_D, of the fiber product of ϕ_1 and ϕ_2, and a factor map $\lambda : \Sigma_C \to \Sigma_D$ which makes the following diagram commute (where ρ_1 and ρ_2 are the restrictions of the projection maps to Σ_D):*



Furthermore, ρ_1 *and* ρ_2 *are surjective.*

Proof. The map λ is defined by $\lambda(x) = (\psi_1(x), \psi_2(x))$. Clearly, the diagram commutes and λ is finite-to-one. It follows that $\Sigma_D = \lambda(\Sigma_C)$ must be an irreducible component of maximal entropy of the fiber product, and therefore an irreducible shift of finite type. Since ψ_1 and ψ_2 are surjective, ρ_1 and ρ_2 are surjective. □

Recall that a cover ϕ is *minimal* if whenever ϕ factors through a cover γ by a map β, it follows that β must be a conjugacy.

Lemma 1.4. *Let $\alpha : S \to S'$ be a conjugacy. The mapping $\gamma \to \alpha\gamma$ induces a bijection between S-conjugacy classes of minimial covers of S and S'-conjugacy classes of minimal covers of S'.*

Proof. For any two covers γ_1 and γ_2 of S, it is easy to check that γ_1 and γ_2 are conjugate over S if and only if $\alpha\gamma_1$ and $\alpha\gamma_2$ are conjugate over S'. So the mapping induces a bijection between all S-conjugacy classes and S'-conjugacy classes. To complete the proof, we need only show that if $\gamma : \Sigma_M \to S$ is a minimal cover of S, then $\alpha\gamma$ is a minimal cover of S'. Suppose that $\alpha\gamma$ factors through another cover $\gamma' : \Sigma_{M'} \to S$. Then there is a factor map $\beta : \Sigma_M \to \Sigma_{M'}$ such that $\gamma'\beta = \alpha\gamma$. Since $\alpha^{-1}\gamma' : \Sigma_{M'} \to S$ is a cover of S and $\alpha^{-1}\gamma'\beta = \gamma$, it follows that β is a conjugacy. □

We next recall the right and left closing covers induced by a ϕ-congruence partition, the details of which are given in [8]. Let $\phi : \Sigma_A \to S$ be a finite-to-one, one-block factor map, from an irreducible shift of finite type onto a sofic shift. Let $m = m_1 \ldots m_r$ be a magic word for ϕ, with magic coordinate m_s ([8, Def. 2.1]). Let $S(\phi, m)$ denote the set of symbols which occur at coordinate s in the set of preimages of m under ϕ ([8, Def. 2.2]). The map ϕ gives rise to a group $\mathscr{G}$ of permutations on the set $S(\phi, m)$ (see [T1, the remarks following Lemma 2.6]). A *ϕ-congruence partition* is a partition $\mathscr{P}$ of $S(\phi, m)$ which is invariant under the action of $\mathscr{G}$ [8, Def. 3.1].

There are two congruence partitions of particular importance. The *partition into singleton sets* is the partition $\mathscr{P}_d$ of $S(\phi, m)$ into d sets, each of which contains a single element (where the degree of ϕ is d). The *trivial partition* is the partition $\mathscr{P}_1$ containing the single set $S(\phi, m)$.

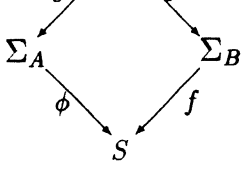
For $P \in \mathscr{P}$, and an S-word w beginning with m, we can define a vector $l_\phi^{(w,P)} \in \mathbb{Z}^n$ (where A is $n \times n$). For details, see [8, Def. 3.3]. If $\mathscr{P} = \mathscr{P}_1$ is the trivial partition, then there is just one set $P = S(\phi, m) \in \mathscr{P}_1$, in which case we simplify notation and write $l_\phi^{(w,P)} = l_\phi^w$. Similarly, one defines vectors $r_\phi^{(w,P)}$.

For any congruence partition $\mathscr{P}$, we can define a labelled graph, whose states are vectors of the form $l^{(w,P)}$ (see [8, Def. 3.5]). This labelling defines a right resolving factor map $\pi_{r(\theta,\mathscr{P})} : \Sigma_{R(\theta,\mathscr{P})} \to S$, called the *right closing cover induced by the partition $\mathscr{P}$*. There is also an induced left closing cover $\pi_{l(\phi,\mathscr{P})} : \Sigma_{L(\phi,\mathscr{P})} \to S$.

2. Lifting Maps

In this section we define lifting maps. While most of our results are stated for left lifting maps, similar results hold for right lifting maps.

Lemma 2.1. *Let $\phi : \Sigma_A \to S$ be a factor map from a shift of finite type onto a sofic shift. Suppose that there exist irreducible shifts of finite type Σ_B and Σ_D, a factor map $g : \Sigma_D \to \Sigma_A$, a right closing factor map $f : \Sigma_B \to S$ and a left closing factor map $p : \Sigma_D \to \Sigma_B$, such that the following diagram commutes:*



Let $\theta = fp = \phi g$. Let $\bar{\mathscr{P}}$ be the θ-congruence partition induced by fp [8, Def. 2.10], *and suppose that there exists a ϕ-congruence partition $\mathscr{P}$ such that $\Lambda_g^{-1}(\mathscr{P}) = \bar{\mathscr{P}}$ (where Λ_g is the map induced by g; see* [8, *Lemma* 2.9]). *Then there exists a factor map $\nu : \Sigma_B \to \Sigma_{R(\phi,\mathscr{P})}$ such that $\pi_{r(\phi,\mathscr{P})}\nu = f$.*

Proof. By [8, Theorem 4.7], f is conjugate over S to $\pi_{r(\theta,\bar{\mathscr{P}})}$. Let $\alpha : \Sigma_B \to \Sigma_{R(\theta,\bar{\mathscr{P}})}$ be a conjugacy such that $\pi_{r(\theta,\bar{\mathscr{P}})}\alpha = f$. By [8, Theorem 3.7], since $\Lambda_g^{-1}(\mathscr{P}) = \bar{\mathscr{P}}$, there exists a factor map $\bar{\nu} = \nu_g : \Sigma_{R(\theta,\bar{\mathscr{P}})} \to \Sigma_{R(\phi,\mathscr{P})}$ such that $\pi_{r(\phi,\mathscr{P})}\bar{\nu} = \pi_{r(\theta,\bar{\mathscr{P}})}$. Now let $\nu = \bar{\nu}\alpha$. □

There are a couple of cases of special interest.

Corollary 2.2. *Assume the same hypotheses as in Lemma* 2.1 *(but without assuming the existence of the partition $\mathscr{P}$).*

(i) *If f has degree one, then there exists $\nu : \Sigma_B \to \Sigma_{R(\phi,\mathscr{P}_1)}$ such that $\pi_{r(\phi,\mathscr{P}_1)}\nu = f$.*

(ii) *If ρ and g have degree one, then there exists $\nu : \Sigma_B \to \Sigma_{R(\phi,\mathscr{P}_d)}$ such that $\pi_{r(\phi,\mathscr{P}_d)}\nu = f$.*

Proof. If f has degree one, then the θ-congruence partition induced by fp is the trivial partition $\bar{\mathscr{P}}_1$, and so $\bar{\mathscr{P}}_1 = \Lambda_g^{-1}(\mathscr{P}_1)$, where $\mathscr{P}_1$ is the trivial ϕ-congruence partition. Now apply Lemma 2.1.

If g and p have degree one, then the θ-congruence partition induced by fp is the partition $\bar{\mathscr{P}}_d$ into singleton sets, and so $\bar{\mathscr{P}}_d = \Lambda_g^{-1}(\mathscr{P}_d)$, where $\mathscr{P}_d$ is the ϕ-congruence partition into singleton sets. Now apply Lemma 2.1. □

Note that in Corollary 2.2 it is not necessary to assume the existence of a partition $\mathscr{P}$ such that $\Lambda_g^{-1}(\mathscr{P}) = \bar{\mathscr{P}}$, since this holds automatically by the degree conditions.

Theorem 2.3. [6, Theorem 4.1]. *Let $\phi : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. There exists an irreducible shift of finite type Σ_D, a factor map $g : \Sigma_D \to \Sigma_A$ of degree one, and a left closing factor map $p : \Sigma_D \to \Sigma_{R(\phi,\mathscr{P}_1)}$ such that $\phi g = \pi_{r(\phi,\mathscr{P}_1)}p$. If ϕ is a one-block map, then p can be taken to be left resolving.*

Proof. By re-coding, we may assume that ϕ is a one-block map. An argument very similar to the proof of [6, Theorem 4.1] shows that there exists an irreducible component of maximal entropy, Σ_D, of the fiber product of the maps $\pi_{l(\phi,\mathscr{P}_d)}$ and $\pi_{r(\phi,\mathscr{P}_1)}$, a left resolving map $p : \Sigma_D \to \Sigma_{R(\phi,\mathscr{P}_1)}$ and a factor map $g : \Sigma_D \to \Sigma_A$ such that $\pi_{r(\phi,\mathscr{P}_1)}p = \phi g$. That g has degree one follows from the proof of [T2, Theorem 2.4]. (In [6, Theorem 4.1], it is shown that there exists a right closing factor map $p : \Sigma_D \to \Sigma_{L(\phi,\mathscr{P}_1)}$, and a factor map $g : \Sigma_D \to \Sigma_A$ such that $\pi_{l(\phi,\mathscr{P}_1)}p = \phi g$. Reversing the roles of left and right resolving covers yields Theorem 2.3.) □

Proposition 2.4 *Let $\phi : \Sigma_A \to S$ be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. Let $\gamma : \Sigma_{R(\phi,\mathscr{P}_1)} \to \Sigma_R$ be the right closing map as in Theorem* 1.1, *so that $\pi^+\gamma = \pi_{r(\phi,\mathscr{P}_1)}$. The following are equivalent:*

(i) *There exists an irreducible shift of finite type* Σ_D, *a factor map* $g : \Sigma_D \to \Sigma_A$ *and a left closing factor map* $p : \Sigma_D \to \Sigma_R$ *such that* $\phi g = \pi^+ p$.

(ii) *The map* $\gamma : \Sigma_{R(\phi,\mathscr{P}_1)} \to \Sigma_R$ *is a conjugacy.*

Proof. (i) $\Rightarrow$ (ii). Since π^+ has degree one, by Corollary 2.2 (i) there exists a factor map $\nu : \Sigma_R \to \Sigma_{R(\phi,\mathscr{P}_1)}$ such that $\pi_{r(\phi,\mathscr{P}_1)}\nu = \pi^+$. By Theorem 1.1, we have $\pi^+\gamma = \pi_{r(\phi,\mathscr{P}_1)}$. So $\pi^+\nu\gamma = \pi^+$. Since π^+ has degree one, it follows that $\nu\gamma$ is the identity on Σ_R, so that γ is a conjugacy.

(ii) $\Rightarrow$ (i). By Theorem 2.3, there exists a shift of finite type Σ_D, a factor map $g : \Sigma_D \to \Sigma_A$ and a left closing factor map $\bar{p} : \Sigma_D \to \Sigma_{R(\phi,\mathscr{P}_1)}$ such that $\phi g = \pi_{r(\phi,\mathscr{P}_1)}\bar{p}$. Let $p = \gamma\bar{p} : \Sigma_D \to \Sigma_R$. Clearly, $\phi g = \pi_{r(\phi,\mathscr{P}_1)}\bar{p} = \pi^+\gamma\bar{p} = \pi^+ p$. Since γ is a conjugacy, it is left closing, so that p is left closing. Therefore (i) holds. □

In view of the previous result, we make the following definition.

Definition 2.5. A factor map $\phi : S_1 \to S_2$ between irreducible sofic shifts is *left lifting* if there exists an irreducible shift of finite type Σ_D, a right closing factor map $\psi_1 : \Sigma_D \to S_1$ and a left closing factor map $\psi_2 : \Sigma_D \to \Sigma_R$ such that $\phi\psi_1 = \pi^+\psi_2$ (where $\pi^+ : \Sigma_R \to S_2$ is the right Fischer cover of S_2). The map ϕ is *right lifting* if there is a left closing factor map $\psi_1 : \Sigma_D \to S_1$ and a right closing factor map $\Sigma_D \to \Sigma_L$ such that $\phi\psi_1 = \pi^-\psi_2$ (where π^- is the left Fischer cover of S_2). A map which is both left and right lifting is *bi-lifting.*

Since left closing and right closing maps are finite-to-one, it is clear that left or right lifting maps are finite-to-one. For maps whose domain is an irreducible shift of finite type, Definition 2.5 is clarified by the following result.

Proposition 2.6. *Let* $\phi : \Sigma_A \to S$ *be a finite-to-one factor map from an irreducible shift of finite type onto a sofic shift. The following are equivalent:*

(i) *There exists an irreducible component of maximal entropy,* Σ_D, *of the fiber product of* ϕ *and* π^+ *such that the restriction of* ρ_1 *to* Σ_D *is a right closing factor map and the restriction of* ρ_2 *to* Σ_D *is a left closing factor map.*

(ii) ϕ *is left lifting.*

(iii) *There exists an irreducible shift of finite type* Σ_C, *a factor map* $\psi_1 : \Sigma_C \to \Sigma_A$ (*not necessarily right closing*), *and a left closing factor map* $\psi_2 : \Sigma_C \to \Sigma_R$ *such that* $\phi\psi_1 = \pi^+\psi_2$.

Proof. (i) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (iii) are obvious.

(iii) $\Rightarrow$ (i) By Lemma 1.3, there exists an irreducible component of maximal entropy, Σ_D, of the fiber product of ϕ and π^+ which makes the diagram in Lemma 1.3 commute. Since ψ_2 is left closing and Σ_D is a shift of finite type, it follows from [3, Prop. 4.10 and 4.11] that ρ_2 is left closing. By [7, Prop. 8.3.3], ρ_1 is right closing. □

Remarks 2.7. (i) It follows trivially from the definition that π^+ is left lifting and π^- is right lifting.

(ii) If $\phi : S_1 \to S_2$ is left closing, S_1, S_2 irreducible, then ϕ is left lifting. For let W be an irreducible component of maximal entropy of the fiber product of ϕ and π^+. Since ϕ is left closing, then by [7, Prop. 8.3.3], ρ_1 is right closing and ρ_2 is left

closing. Since Σ_R is a shift of finite type, it follows from [3, Prop. 4.12] that W is a shift of finite type. Therefore ϕ is left lifting. Similarly, right closing implies right lifting. It follows that π^- is left lifting, and so by (i), it is bi-lifting. Similarly, π^+ is bi-lifting.

However, left closing almost everywhere does not imply left lifting. For let $\phi : S \to \Sigma_B$ be a left closing almost everywhere factor map, where S is an irreducible non-AFT sofic shift. By [3, Prop. 4.12], ϕ cannot be left closing. Also, the right Fischer cover of Σ_B is the identity. Suppose that $\psi_1 : \Sigma_D \to S$ and $\psi_2 : \Sigma_D \to \Sigma_B$ are factor maps, where Σ_D is an irreducible shift of finite type, ψ_1 is right closing and $\phi\psi_1 = \pi^+\psi_2 = \psi_2$. Since ψ_1 factors through the right Fischer cover of S, which is not left closing (since S in not AFT), it follows that ψ_1 is not left closing. Therefore $\psi_2 = \phi\psi_1$ is not left closing, and so ϕ cannot be left lifting.

(iii) It follows easily from Proposition 2.6 (iii) that the collection of left lifting covers of S is closed under lower factors.

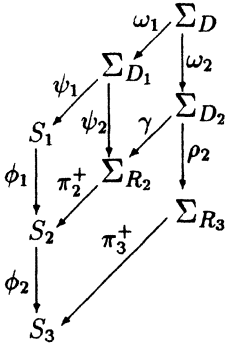
(iv) If $\phi : S_1 \to S_2$ is left lifting and S_2 is AFT, then by [2, Corollary 11], π^+ is left closing. Therefore $\pi^+\psi_2$ is left closing and it follows from [3, Prop. 4.11] that ϕ is left closing almost everywhere. If in addition S_1 is a shift of finite type, then by [3, Prop. 4.10], ϕ is left closing.

(v) If Σ_A is an irreducible shift of finite type, then by Proposition 2.4, $\phi : \Sigma_A \to S$ is left lifting if and only if the induced right closing cover $\pi_{r(\phi,\mathscr{P}_1)}$ is conjugate to the right Fischer cover of S.

We next show that compositions of left lifting maps are left lifting.

Proposition 2.8. *Let $\phi : S_1 \to S_2$ and $\phi_2 : S_2 \to S_3$ be factor maps between irreducible sofic shifts, and let $\phi = \phi_2\phi_1$. If ϕ_1 and ϕ_2 are left lifting then ϕ is left lifting.*

Proof. Let $\pi_2^+ : \Sigma_{R_2} \to S_3$ and $\pi_3^+ : \Sigma_{R_3} \to S_3$ denote the right Fischer covers of S_2 and S_3, respectively. Since ϕ_2 is left lifting, there exists an irreducible shift of finite type Σ_{D_2}, a right closing factor map $\rho_1 : \Sigma_{D_2} \to S_2$, and a left closing map $\rho_2 : \Sigma_{D_2} \to \Sigma_{R_3}$, such that $\phi_2\rho_1 = \pi_3^+\rho_2$. By Theorem 1.1, there exists a right closing factor map $\gamma : \Sigma_{D_2} \to \Sigma_{R_2}$ such that $\rho_1 = \pi_2^+\gamma$. Also, since ϕ_1 is left lifting, there exists an irreducible shift of finite type Σ_{D_1}, a right closing factor map $\psi_1 : \Sigma_{D_1} \to S_1$ and a left closing factor map $\psi_2 : \Sigma_{D_1} \to \Sigma_{R_2}$ such that $\phi_1\psi_1 = \pi_2^+\psi_2$. Now, left Σ_D be an irreducible component of maximal entropy of the fiber product of ψ_2 and γ, together with projection maps $\omega_1 : \Sigma_D \to \Sigma_{D_1}$ and $\omega_2 : \Sigma_D \to \Sigma_{D_2}$. We have the commutative diagram:



Since ψ_2 is left closing, it follows that ω_2 is left closing, and therefore $\rho_2\omega_2$ is left closing. Since γ is right closing, it follows that ω_1 is right closing, and therefore $\psi_1\omega_1$ is right closing. Therefore $\phi_2\phi_1$ is left lifting. □

Remark. If $\gamma : \Sigma_M \to S$ is a left lifting cover and $\alpha : S \to S'$ is a conjugacy, then α is left closing, and therefore left lifting, by Remark 2.7 (ii). By Proposition 2.8, $\alpha\gamma$ is left lifting. It now follows from Lemma 1.4 that the mapping $\gamma \to \alpha\gamma$ induces a bijection between S-conjugacy classes of minimial left lifting covers of S and S'-conjugacy classes of minimal left lifting covers of S'. It follows that the number of these classes, the *left lifting number* of S, is a conjugacy invariant. In Theorem 2.17, we show that this number is finite.

The converse of Proposition 2.8 does not hold: the fact that $\phi_2\phi_1$ is left lifting does not imply that either ϕ_1 or ϕ_2 is left lifting. To see that ϕ_2 need not be left lifting, take the map $\phi_2 = \phi$ described in the second paragraph of Remark 2.7 (ii). Let $\phi_1 : \Sigma_L \to S$ be the left Fischer cover of S. Then since ϕ_2 is left closing almost everywhere and ϕ_1 is left closing and has finite type domain, it follows from [3, Prop. 4.10 and 4.11] that $\phi_2\phi_1$ is left closing, and therefore left lifting, by Remark 2.7 (ii). But ϕ_2 is not left lifting.

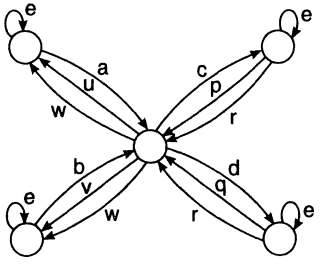
To see that ϕ_1 need not be left lifting, simply take an irreducible component of maximal entropy, Σ_D, of the fiber product of the left and right Fischer covers, π^- and π^+, of a non-AFT sofic shift S. Let $\psi_1 : \Sigma_D \to \Sigma_L$ and $\psi_2 : \Sigma_D \to \Sigma_R$ be the projection maps, and let $\theta = \pi^-\psi_1 = \pi^+\psi_2$. Then θ is left lifting. But ψ_1, which is right closing, cannot be left lifting; for if it were, then it would be left closing, by Remark 2.7 (iv), since Σ_L is a shift of finite type. But this would imply that θ is left closing, and therefore π^+ is left closing, which would imply that S is AFT by [2, Corollary 11].

We next give a simple condition which implies that a map is left lifting.

Definition 2.9. A one-block factor map $\phi : \Sigma_A \to S$ from an irreducible shift of finite type onto a sofic shift is *left determining* if for all magic words w, v, $\mathscr{F}_S(w) = \mathscr{F}_S(v)$ implies that $l^w_\phi = l^v_\phi$. Similarly, ϕ is *right determining* if $\mathscr{P}_S(w) = \mathscr{P}_S(v)$ implies that $r^w_\phi = r^v_\phi$. (The converses to these statements are true for any map ϕ.)

Since the map γ of Theorem 1.1 is defined on states by $l^w_\phi \to \mathscr{F}_S(w)$, if ϕ is left determining, then γ is actually a labelled graph isomorphism, and therefore a conjugacy. It follows by Proposition 2.4 that a left determining map is left lifting. It is not hard to show that every left resolving map is left determining.

Example 2.10.



The labelling of the directed graph above defines a one-block factor map from Σ_A onto a sofic shift S, where A is the transition matrix. It is easy to check that ϕ is both left and right determining, and therefore left and right lifting, but ϕ is neither left nor right closing. One can check that $\chi_A = (x^2 - x - 8)(x-1)^3$. It can be shown that the left Fischer cover of S has transition matrix

$$L = \begin{bmatrix} 0 & 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

while the right Fischer cover has transition matrix $R = L^T$, and that $\chi_R = \chi_L = (x^2 - x - 8)(x-1)^4$. We show that there cannot exist an integral eventually positive matrix C [3, p. 18] such that Σ_A eventually factors onto C by left closing maps and C eventually factors onto S by right closing maps (or vice versa). For if there were, then by [2, Prop. 4], C must eventually factor onto Σ_R (or Σ_L, if C eventually factors onto S by left closing maps). It would then follow from [6, Corollary 4.13] that χ_R would divide χ_A. Since this is false, no such C exists. One can contrast this example with [6, Theorem 4.12], which says that if the range of ϕ is a shift of finite type, then there always exists a matrix C with the property described above.

Next, we give a characterization of left determining maps, which is an analog of Proposition 2.4.

Proposition 2.11. *Suppose that $\phi : \Sigma_A \to S$ is a finite-to-one, one-block factor map from an irreducible shift of finite type onto a sofic shift. The following are equivalent:*

(i) *There exists an irreducible shift of finite type Σ_D, a one-block factor map $\psi_1 : \Sigma_D \to \Sigma_A$ and a left resolving factor map $\psi_2 : \Sigma_D \to \Sigma_R$ such that $\phi\psi_1 = \pi^+\psi_2$.*

(ii) *ϕ is left determining.*

Proof. (i) $\Rightarrow$ (ii) Suppose that w and v are magic words, with $\mathscr{F}_S(w) = \mathscr{F}_S(v)$. Let i denote the state $\mathscr{F}_S(w) = \mathscr{F}_S(v)$ of Σ_R. Since there is a path in $G(R)$ labelled w ending at i and a path labelled v ending at i, and π^+ is right resolving and has degree one, it is easy to see that $l^w_{\pi^+} = l^v_{\pi^+} = e_i$, where e_i is a standard basis vector. Let $\theta = \pi^+\psi_2$. Now, since ψ_2 is left resolving, it follows that there exists a path in $G(D)$ labelled w ending at $k \in \mathscr{S}_D$ if and only if $k \in (\delta\psi_2)^{-1}(i)$ (where $\delta\psi_2$ is the induced map on states – see [7, Def. 2.2.2]). This holds if and only if there exists a path labelled v ending at k. It follows that $l^w_\theta = l^v_\theta$. Now, by [8, Lemma 3.6], we have $l^w_\phi = l^v_\phi$. Therefore ϕ is left determining.

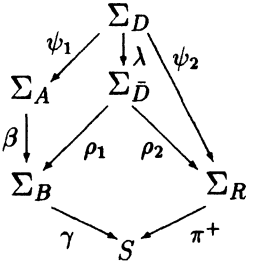
(ii) $\Rightarrow$ (i) If ϕ is left determining, then by the remarks following Definition 2.9, the map $\gamma : \Sigma_{R(\phi,\mathscr{P}_1)} \to \Sigma_R$ is a labelled graph isomorphism, and therefore left

resolving. It follows from Theorem 2.3 that there exists an irreducible shift of finite type Σ_D, a factor map $\psi_1 : \Sigma_D \to \Sigma_A$ and a left resolving map $\bar{\psi}_2 : \Sigma_D \to \Sigma_{R(\phi,\mathscr{P}_1)}$, such that $\phi\psi_1 = \pi_{r(\phi,\mathscr{P}_1)}\bar{\psi}_2$. Then $\psi_2 = \gamma\bar{\psi}_2$ is left resolving, and clearly $\phi\psi_1 = \pi^+\psi_2$. □

The following lemma is an analog for left determining maps of the fact that left lifting maps are closed under lower factors.

Lemma 2.12. *Suppose that $\phi : \Sigma_A \to S$ is a left determining factor map, Σ_A irreducible, and suppose that there exists a shift of finite type Σ_B and one-block factor maps $\beta : \Sigma_A \to \Sigma_B$ and $\gamma : \Sigma_B \to S$ such that $\gamma\beta = \phi$. Then γ is left determining.*

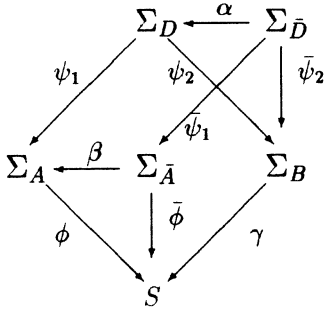
Proof. Since ϕ is left determining, it follows from Proposition 2.11 that there exists an irreducible shift of finite type Σ_D, a one-block factor map $\psi_1 : \Sigma_D \to \Sigma_A$ and a left resolving factor map $\psi_2 : \Sigma_D \to \Sigma_R$ such that $\phi\psi_1 = \pi^+\psi_2$. By Lemma 1.3, there exists an irreducible component $\Sigma_{\bar{D}}$, of the fiber product of γ and π^+, and factor maps $\lambda : \Sigma_D \to \Sigma_{\bar{D}}, \rho_1 : \Sigma_{\bar{D}} \to \Sigma_B$ and $\rho_2 : \Sigma_{\bar{D}} \to \Sigma_R$, such that the following diagram commutes:



Since γ and π^+ are one-block maps, it follows that ρ_1 and ρ_2 are one-block maps. Since ψ_1, ψ_2 and β are one-block maps, it follows that λ is a one-block map. Since ψ_2 is left resolving, it is easy to show that ρ_2 is left resolving. Therefore γ is left determining, by Proposition 2.11. □

The next result is the essential step in showing that left lifting maps can be recoded to left determining maps.

Proposition 2.13. *Suppose that $\phi : \Sigma_A \to S$ and $\gamma : \Sigma_B \to S$ are finite-to-one, one-block factor maps, where Σ_A and Σ_B are irreducible shifts of finite type and S is a sofic shift. Let Σ_D be an irreducible component of the fiber product of ϕ and γ, with projection maps $\psi_1 : \Sigma_D \to \Sigma_A$ and $\psi_2 : \Sigma_D \to \Sigma_B$, and assume that ψ_2 is right closing. Then there exist shifts of finite type $\Sigma_{\bar{D}}$ and $\Sigma_{\bar{A}}$, conjugacies $\alpha : \Sigma_{\bar{D}} \to \Sigma_D$ and $\beta : \Sigma_{\bar{A}} \to \Sigma_A$, a right resolving factor map $\bar{\psi}_2 : \Sigma_{\bar{D}} \to \Sigma_B$ and one-block factor maps $\bar{\psi}_1 : \Sigma_{\bar{D}} \to \Sigma_{\bar{A}}$ and $\bar{\phi} : \Sigma_{\bar{A}} \to S$, which make the following diagram commute:*



A similar statement holds with right closing and right resolving replaced by left closing and left resolving, respectively.

Proof. The existence of $\Sigma_{\bar{D}}$ and maps $\alpha : \Sigma_{\bar{D}} \to \Sigma_D$ and $\bar{\psi}_2 : \Sigma_{\bar{D}} \to \Sigma_B$ such that $\psi_2\alpha = \bar{\alpha}_2$ follows from [7, Prop. 5.1.11]. We recall the construction in the proof of that proposition, taking into account the fact that Σ_D is a component of the fiber product of ϕ and γ. States of $\Sigma_{\bar{D}}$ are of the form $((I,J), w_1w_2 \ldots w_d)$, where (I,J) is a state of Σ_D (so that $I \in \mathscr{S}_A$ and $J \in \mathscr{S}_B$, d is the delay of ψ_2 – see [7, Def. 5.1.4]) and $w_1w_2 \ldots w_d$ is a B-word which is the ψ_2-label of some path in $G(D)$ beginning at (I,J). By definition of the fiber product, such a path is of the form $(f_1,w_1)(f_2,w_2)\ldots(f_d,w_d)$, where $f_1f_2\ldots f_d$ is a path in $G(A)$, beginning at I, and $w_1w_2\ldots w_d$ is a path in $G(B)$, beginning at J, and $\phi(f_i) = \gamma(w_i)$, for $1 \leqslant i \leqslant d$. If $w_1w_2\ldots w_da$ is a B-word which is the ψ_2 label of a path in $G(D)$ beginning at (I,J), then among all such paths the initial edge (f_1,w_1), is unique, by the delay condition. If (f_1,w_1) ends at state (K,L), then we endow $G(\bar{D})$ with an edge $((I,J), w_1w_2\ldots w_da)$ from $((I,J), w_1w_2\ldots w_d)$ to $((K,L), w_2\ldots w_da)$. We define a one-block map $\theta : \Sigma_{\bar{D}} \to \Sigma_D$ by $\theta((I,J), w_1w_2\ldots w_da) = (f_1,w_1)$, which is easily seen to be a conjugacy. We define a right resolving map $\bar{\psi}_2 : \Sigma_{\bar{D}} :\to \Sigma_B$ by $\bar{\psi}_2((I,J), w_1w_2\ldots w_da) = a$. If $\alpha = \theta\sigma^d$, then $\psi_2\alpha = \bar{\psi}_2$.

We now mimic this construction for the map ϕ. We construct a graph $G(\bar{A})$ whose states are of the form $(I, y_1y_2\ldots y_d)$, where $I \in \mathscr{S}_A$ and $y_1y_2\ldots y_d$ is an S-word which is the ϕ-label of a path in $G(A)$ beginning at state I. Suppose that $y_1y_2\ldots y_dc$ is an S-word which is the ϕ-label of a path in $G(A)$ beginning at I. If f is the initial edge of such a path, and f ends at state K, then we endow $G(\bar{A})$ with an edge F from $(I, y_1y_2\ldots y_d)$ to $(K, y_2\ldots y_dc)$. We note that in this case, the initial edge f is not unique in general, since ϕ is not necessarily right closing. But given such an edge F, there is a unique edge f in $G(A)$ from I to K labelled y_1 by ϕ, since ϕ is finite-to-one (for otherwise ϕ would have a diamond – see [6, Theorem 1.1]). We define a one-block map $\lambda : \Sigma_{\bar{A}} \to \Sigma_A$ by $\lambda(F) = f$. As in the proof of [7, Prop. 5.1.11], λ is a conjugacy. We define a one-block map $\bar{\phi} : \Sigma_{\bar{A}} \to S$ by $\bar{\phi}(F) = c$. This is easily seen to be a factor map; however it is not necessarily right resolving, since a state $(I, y_1y_2\ldots y_d)$ and a label a do not uniquely determine the edge F. If $\beta = \lambda\sigma^d$, then $\phi\beta = \bar{\phi}$.

Finally, we define a one-block factor map $\bar{\psi}_1 : \Sigma_{\bar{D}} \to \Sigma_{\bar{A}}$ as follows. Suppose that $((I,J), w_1w_2\ldots w_da)$ is an edge in $G(\bar{D})$. By definition of $G(\bar{D})$, there exists a

path $(f_1,w_1)(f_2,w_2)\dots(f_d,w_d)(b,a)$ in $G(D)$, where $\phi(f_i)=\gamma(w_i)=y_i$, for $1\leqslant i\leqslant d$, and $\phi(b)=\gamma(b)=\gamma(a)=c$, and the initial edge (f_1,w_1) is unique among all such paths. Suppose that f_1 ends at state K. Then $f_1f_2\dots f_2b$ is a path in $G(A)$ labelled $y_1y_2\dots y_dc$ by ϕ, so there is a unique edge F in $G(\bar{A})$ from $(I,y_1y_2\dots y_d)$ to $(K,y_2\dots y_dc)$ which is labelled f by λ. We define $\bar{\psi}_1((I,J),w_1w_2\dots w_da)=F$.

We now show that the diagram above commutes. Clearly, $\lambda\bar{\psi}_1((I,J),w_1w_2\dots w_da)=\lambda(F)=f_1=\psi_1(f_1,w_1)=\psi_1\theta((I,J),w_1w_2\dots w_da)$, and therefore $\beta\bar{\psi}_1=\psi_1\alpha$, since $\bar{\psi}_1$ commutes with σ^d. Also $\bar{\phi}\bar{\psi}_1((I,J),w_1w_2\dots w_da)=\bar{\phi}_1(F)=c=\gamma(a)=\gamma\bar{\psi}_2((I,J),w_1w_2\dots w_da)$, so that $\bar{\phi}\bar{\psi}_1=\gamma\bar{\psi}_2$. Finally, we show that $\phi\beta=\bar{\phi}$. Suppose that $F_1\dots F_{d+1}$ is a path in $G(\bar{A})$, where F_j is an edge from $(I^{(j)},y_jy_{j+1}\dots y_{d+j-1})$ to $(I^{(j+1)},y_{j+1}\dots y_{d+j})$, for $1\leqslant j\leqslant d+1$. By definition, $\phi\lambda(F_{d+1})=y_{d+1}=\bar{\phi}(F_1)$. It follows that $\phi\beta=\phi\lambda\sigma^d=\bar{\phi}$. □

If $\phi:\Sigma_A\to S$ and $\bar{\phi}:\Sigma_{\bar{A}}\to S$ are factor maps, we say that ϕ is *conjugate over* S to $\bar{\phi}$ if there exists a conjugacy β such that $\phi\beta=\bar{\phi}$. The following result is a generalization of the fact that every left closing map is conjugate over S to a left resolving map ([5]).

Corollary 2.14. *Every left lifting map is conjugate over S to a left determining map.*

Proof. Let ϕ be left lifting. By [7, Prop. 1.5.12], any factor map is conjugate over S to a one-block map, so we may assume that ϕ is a one-block map. By Proposition 2.6, there is an irreducible component of maximal entropy, Σ_D, of the fiber product of ϕ and π^+ such that if ψ_1 and ψ_2 are the restrictions of the projection maps to Σ_D, then ψ_2 is left closing. By construction, ψ_1 and ψ_2 are one-block maps, since ϕ and π^+ are. Now it follows from Proposition 2.11 and the left closing version of Proposition 2.13, by taking $\gamma=\pi^+$, that ϕ is conjugate over S to a map $\bar{\phi}$ which is left determining. □

Lemma 2.15. *Let $\phi:\Sigma_A\to S$ be a left determining factor map, Σ_A irreducible. Suppose that S has n distinct follower sets. Then there are at most 2^n sets of the form $\mathscr{P}_\phi(i)$, with $i\in\mathscr{S}_A$.*

Proof. Since ϕ is left determining, by Proposition 2.11 there exist an irreducible shift of finite type Σ_D, a one-block factor map $\psi_1:\Sigma_D\to\Sigma_A$ and a left resolving factor map $\psi_2:\Sigma_D\to\Sigma_R$ such that $\phi\psi_1=\pi^+\psi_2=\theta$. Since ψ_2 is left resolving, it is easy to see that if $\delta\psi_2(i)=j$, then $\mathscr{P}_\theta(i)=\mathscr{P}_{\pi^+}(j)$ (where $\delta\psi_2$ is the induced map on states – see [7, Def. 2.2.2]). Therefore the predecessor sets of states of Σ_D are the same as those of Σ_R. Since S has n states, corresponding to the distinct follower sets of S, there are at most n distinct sets of the form $\mathscr{P}_\theta(i), i\in\mathscr{S}_D$. Now, if $k\in\mathscr{S}_A$, then since ψ_1 is a one-block map, it is easy to check that

$$\mathscr{P}_\phi(k)=\bigcup_{i\in(\delta\psi_1)^{-1}(k)}\mathscr{P}_\theta(i).$$

Therefore the predecessor sets of states of Σ_A are unions of predecessor sets of states of Σ_D, and the number of these is at most 2^n. □

Theorem 2.16. *Let $\phi : \Sigma_A \to S$ be a left lifting factor map from an irreducible shift of finite type onto a sofic shift. Suppose that S has n distinct follower sets. Then there exists a shift of finite type Σ_B with at most 2^n states, a left closing map $\beta : \Sigma_A \to \Sigma_B$ and a left determining factor map $\gamma : \Sigma_B \to S$ such that $\gamma\beta = \phi$.*

Proof. By Corollary 2.14, there exists a left determining map $\bar{\phi} : \Sigma_{\bar{A}} \to S$ and a conjugacy $\alpha : \Sigma_A \to \Sigma_{\bar{A}}$ such that $\bar{\phi}\alpha = \phi$. By Proposition 1.2, there exists a shift of finite type Σ_B, a one-block factor map $\gamma : \Sigma_B \to S$ which is predecessor-separated, and a left resolving map $\bar{\beta} : \Sigma_{\bar{A}} \to \Sigma_B$ such that $\gamma\bar{\beta} = \bar{\phi}$. By Lemma 2.12, γ is left determining. By Lemma 2.15, there are at most 2^n sets of the form $\mathscr{P}_\gamma(i)$, with $i \in \mathscr{S}_B$. Since Σ_B is predecessor-separated, it has at most 2^n states. Now, let $\beta = \bar{\beta}\alpha$, so that β is left closing and $\gamma\beta = \phi$. □

Recall that the left lifting number of S is the number of S-conjugacy classes of minimal left lifting covers of S. The following shows that this number is finite.

Theorem 2.17. *An irreducible sofic shift S has finitely many minimal left lifting covers, up to conjugacy over S.*

Proof. Assume that S has n follower sets. Let $\mathscr{C}$ denote the collection of left determining one-block covers $\gamma : \Sigma_M \to S$, such that Σ_M is an irreducible shift of finite type with at most 2^n states. Observe that for any cover in $\mathscr{C}$, the entries of M are bounded by the size of the alphabet of S, for otherwise ϕ would have a diamond, contradicting the fact that it is finite-to-one, by [6, Theorem 1.1]. So there are only finitely many possibilities for the matrix M, and for each of these there only finitely many one-block maps $\Sigma_M \to S$. Therefore $\mathscr{C}$ is finite. Now, if $\gamma' : \Sigma_{M'} \to S$ is any minimal left lifting cover, then by Theorem 2.16, there exists a left determining cover $\gamma : \Sigma_M \to S$ in $\mathscr{C}$ and a factor map $\beta : \Sigma_{M'} \to \Sigma_M$ such that $\gamma\beta = \gamma'$. Since γ' is minimal, β must be a conjugacy, so that γ' is conjugate over S to γ. □

The next result shows that the left lifting number is computable.

Corollary 2.18. *There is a finite procedure for constructing a finite set $\mathscr{M}_L$ of minimal left determining covers of S which are pairwise not conjugate over S and such that any minimal left lifting cover of S is conjugate over S to an element of $\mathscr{M}_L$.*

Proof. By the proof of Theorem 2.17, every minimal left lifting cover of S is conjugate over S to an element of $\mathscr{C}$. By the proof of [1, Corollary 2.7], there is a finite procedure for deciding whether an element of $\mathscr{C}$ is minimal. Furthermore, by the proof of [8, Prop. 5.1], two covers of a sofic shift S are conjugate over S if and only if there is an irreducible component of maximal entropy, Σ_C, of their fiber product such that the projection maps restricted to Σ_C are conjugacies. By [8, Remark 5.2 (i) and the proof of Corollary 5.3], there is a finite procedure for deciding whether this holds. Proceeding inductively, we may repeatedly discard from $\mathscr{C}$ covers which are either not minimal or which are conjugate over S to a cover still in the set, until we obtain a set $\mathscr{M}_L$ with the required property. □

It follows from Corollary 2.18 and the remark following Proposition 2.8 that the left lifting number is a computable conjugacy invariant of S.

Corollary 2.19. *Let S be an irreducible sofic shift. There exists a finite collection of covers of S such that every left lifting cover of S factors through one in the collection.*

Proof. It follows from [1, Corollary 2.7] that any finite-to-one cover $\phi : \Sigma_A \to S$ factors through a minimal cover γ. If ϕ is left lifting, then γ must be also, by Remark 2.7 (iii). So any set of representatives of the S-conjugacy classes of minimal left lifting covers, which must be finite by Theorem 2.17, has the desired property. □

Remark 2.20. If $\phi : S_1 \to S_2$ is a left lifting factor map between irreducible sofic shifts, then for any minimal left lifting cover $\gamma_1 : \Sigma_{M_1} \to S_1$, the map $\phi\gamma_1$ is left lifting. It follows that $\phi\gamma_1$ factors through a minimal left lifting cover of S_2. So, a necessary condition for the existence of the map ϕ is that for every minimal left lifting cover γ_1 of S_1, there exists a minimal left lifting cover $\gamma_2 : \Sigma_{M_2} \to S_2$ such that Σ_{M_1} factors onto Σ_{M_2}. Even in the case in which ϕ is left closing, this provides a stronger necessary condition for the existence of ϕ than would be given just by the left Fischer covers.

3. Bi-lifting Maps

Lemma 3.1. *Let $\phi : \Sigma_A \to S$ be a factor map from an irreducible shift of finite type onto a sofic shift, which has degree one and is bi-lifting. Then there exists an irreducible component Σ_D of the fiber product of π^- and π^+ and a factor map $g : \Sigma_D \to \Sigma_A$ of degree one such that $\phi g = \rho_1\pi^- = \rho_2\pi^+$.*

Proof. It follows from [6, Theorem 4.1], since ϕ has degree one, that there exists an irreducible component $\Sigma_{\bar{D}}$ of the fiber product of $\pi_{l(\phi,\mathscr{P}_1)}$ and $\pi_{r(\phi,\mathscr{P}_1)}$ and a factor map $\bar{g} : \Sigma_{\bar{D}} \to S$ of degree one such that $\phi\bar{g} = \pi^-\bar{\rho}_1 = \pi_{l(\phi,\mathscr{P}_1)}\bar{\rho}_1 = \pi_{r(\phi,\mathscr{P}_1)}\bar{\rho}_2$, where $\bar{\rho}_1$ and $\bar{\rho}_2$ are the projection maps. Let $\bar{\theta} = \phi\bar{g}$. Since ϕ is bi-lifting, it follows from Proposition 2.4 that $\pi_{l(\phi,\mathscr{P}_1)}$ is conjugate to π^- and $\pi_{r(\phi,\mathscr{P}_1)}$ is conjugate to π^+. From this it follows that there is an irreducible component Σ_D of the fiber product of π^- and π^+ such that if $\theta = \rho_1\pi^- = \rho_2\pi^+$, then θ is conjugate over S to $\bar{\theta}$. This induces a map $g : \Sigma_D \to \Sigma_A$ with the desired property. □

Finally, we characterize bi-lifting maps of degree one.

Theorem 3.2. *Let $\phi : S_1 \to S_2$ be a factor map between irreducible sofic shifts. Then ϕ has degree one and is bi-lifting if and only if there exists an irreducible component Σ_D of the fiber product of π_2^- and π_2^+, the left and right Fischer covers of S_2, respectively, and a factor map $g : \Sigma_D \to S_1$ of degree one such that $\phi g = \rho_1\pi_2^- = \rho_2\pi_2^+$.*

Proof. Suppose that ϕ has degree one and is bi-lifting. Let π_1^- be the left Fischer cover of S_1. Then by Remark 2.7 (ii), π_1^- is bi-lifting, and it follows from Proposition 2.8 that $\gamma = \phi\pi_1^-$ is bi-lifting. Since π_1^- is one-to-one almost

everywhere, so is γ. Now, by Lemma 3.1, there exists an irreducible component Σ_D of the fiber product of π_2^- and π_2^+, and a factor map $h : \Sigma_D \to \Sigma_L$ of degree one such that $\phi h = \rho_1 \pi_2^- = \rho_2 \pi_2^+$. Now, simply let $g = \pi_1^- h$. Then g has degree one and $\phi g = \gamma h = \rho_1 \pi_2^- = \rho_2 \pi_2^+$.

Conversely, suppose that there exists an irreducible component of the fiber product Σ_D of π_2^- and π_2^+ and a factor map $g : \Sigma_D \to S_1$ such that $\phi g = \rho_1 \pi^- = = \rho_2 \pi^+$. By [7, Prop. 8.3.3], ρ_1 is right closing and ρ_2 is left closing. It follows that ϕ is bi-lifting. □

We conclude with an open question: does there exist a minimal left lifting cover which is not bi-lifting?

References

[1] Boyle M Factoring factor maps. J London Math Soc (to appear)
[2] Boyle M, Kitchens B, Marcus B (1985) A note on minimal covers for sofic systems. Proc Amer Math Soc **95**: 403–411
[3] Boyle M, Marcus B, Trow P (1987) Resolving maps and the dimension group for shifts of finite type. Memoirs Amer Math Soc **377**
[4] Fischer R (1975) Graphs and symbolic dynamics. Information Theory, vol 16. Proc Colloq Math Soc Janos Bolyai
[5] Kitchens B (1981) Continuity properties of factor maps in ergodic theory. PhD Thesis, Chapel Hill: Univ North Carolina
[6] Kitchens B, Marcus B, Trow P (1991) Eventual factor maps and compositions of closing maps. Ergodic Th Dyn Sys **11**: 85–113
[7] Lind D, Marcus B (1995) An Introduction to Symbolic Dynamics and Coding. New York: Cambridge Univ Press
[8] Trow P (1995) Decompositions for finite-to-one factor maps. Israel J Math **91**: 129–155
[9] Trow P Decompositions of factor maps involving bi-closing maps. Math **125**: 165–172
[10] Williams S (1986) A sofic system with infinitely many minimal covers. Proc Amer Math Soc **98**: 503–506
[11] Williams S (1988) Covers of non-almost-finite-type systems. Proc Amer Math Soc **104**: 245–252

P. Trow
Department of Mathematical Sciences
The University of Memphis
Memphis, TN 38152
USA

Mh. Math. 125, 343–350 (1998)

Monatshefte für
Mathematik

Printed in Austria

On Existence of Positive Periodic Solutions

By

Klaudiusz Wójcik*, Kraków

(Received 22 May 1996; in revised form 11 November 1996)

Abstract. We present the geometric method for detecting periodic solutions of time periodic nonautonomous differential equations in interior of convex subset of euclidean space. The method is based on the Lefschetz fixed point theorem and the topological principle of Ważewski. Two applications to the existence of positive periodic solutions are considered.

1. Introduction

In applications of ordinary differential equations to population dynamics, $\mathbb{R}^n \times \mathbb{R}_+$ or $\mathbb{R}^n_+$, where $\mathbb{R}_+ = [0, +\infty)$, appear in a natural way phase spaces ([14], [3]). In such a context, an interesting problem which arises is the existence of positive periodic solutions, i.e. solutions which are contained in the interior of these sets. It was investigated by many authors ([1], [2], [4], [5], [7], [9], [10], [11], [13], [16]) and sufficient conditions for the existence of positive periodic solutions have been given for a wide variety of models. For more details and extensive bibliographies concerning the problem, we refer the reader to [16], [9], [11], and [7]. Our Theorem 2 is a consequence of results in [15], based on the Lefschetz Fixed Point Theorem and the Topological Principle of Ważewski. We give some topological conditions for a T-periodic local process φ on cl Ω, $\Omega \subset \mathbb{R}^n$ an open, convex set which guarantee the existence of periodic trajectories contained in Ω for all t. Note that our approach works without dissipative assumption, which is essential in many existence theorems based on topological degree theory. In contrast to the method based on the Brouwer Fixed Point Theorem, we do not need a contractible invariant set for the Poincaré map associated with the considered equation. In particular, our method can be applied to detecting positive periodic solutions which are not asymptotically stable. One can try to use our Theorem 2 to the equations for which there is a repelling rest point on the boundary of Ω. Two applications to the existence of positive periodic solutions are considered. In the first example, we apply Theorem 2 to some Lotka-Volterra type system in $\mathbb{R}^2_+$. In the second example we study a class of periodic nonautonomous differential equations on $\mathbb{R} \times \mathbb{R}_+$. We have chosen the simple equations in order to illustrate the method. However, we believe that our approach can be also applied to

1991 Mathematics Subject Classifications: 34C25, 47H10
Key words: Positive periodic solutions, Lotka-Volterra equations
* Research supported by the KBN grant 2 P03A 040 10

equations causing serious problems in determination of their dynamics. Completely different methods for similar problems are presented in [5], [6], [9], but our Theorem 3 is not a direct special case of those results. In [13] a sufficient condition was given for the existence of a positive periodic solution in the periodic competition system. Later, ALVAREZ and LAZER (see [1]) used the topological degree to prove that, under the assumptions of Gopalsamy, the periodic solution is unique and globally asymptotically stable.

2. Periodic Isolating Segments

Assume that X is a topological space and $\varphi : D \to X$ is a continuous mapping, $D \subset \mathbb{R} \times X \times \mathbb{R}$ is an open set. We will denote by $\varphi_{(\sigma,t)}$ the function $\varphi(\sigma, \cdot, t)$. φ is called a local process if the following conditions are satisfied

(1) $\forall \sigma \in \mathbb{R}, x \in X : \{t \in \mathbb{R} : (\sigma, x, t) \in D\}$ is an interval.

(2) $\forall \sigma \in \mathbb{R} : \varphi_{(\sigma,0)} = \mathrm{id}_X$

(3) $\forall \sigma \in \mathbb{R} : \varphi_{(\sigma,s+t)} = \varphi_{(\sigma+s,t)} o \varphi_{(\sigma,s)}$,

If $D = \mathbb{R} \times X \times \mathbb{R}$, we call φ a (global) process. For $(\sigma, x) \in \mathbb{R} \times X$ the set

$$\{(\sigma + t, \varphi_{(\sigma,t)}(x) \in \mathbb{R} \times X : (\sigma, x, t) \in D\}$$

is called the trajectory of (σ, x) in φ. If T is a positive number such that

(4) $\forall \sigma, t \in \mathbb{R}$ we have $\varphi_{(\sigma+T,T)} = \varphi_{(\sigma,t)}$,

we call φ a T-periodic local process. In this paper φ denotes always a T-periodic process. A local process φ on X determines a local flow Φ on $\mathbb{R} \times X$ by the formula

$$\Phi_t(\sigma, x) = (\sigma + t, \varphi_{(\sigma,t)}(x)).$$

Remark 1. The differential equation

$$\dot{x} = f(x, t), \qquad (*)$$

such that f is regular enough to guarantee the uniqueness for the solutions of the Cauchy problems associated to $(*)$, generates a local process as follows. For $x(t_0, x_0; \cdot)$ the solution of $(*)$ such that $x(t_0, x_0; t_0) = x_0$, we put

$$\varphi_{(t_0,\tau)}(x_0) = x(t_0, x_0; t_0 + \tau).$$

If f is T-periodic with respect to t, then φ is a T-periodic local process and in order to determine all T-periodic solutions of the equation $(*)$ it suffices to look for fixed points of $\varphi_{(0,T)}$ (called the Poincaré map).

Now we introduce the notion of periodic isolating segment. To this aim we use the following notation: by $\pi_1 : [0, T] \times X \to [0, T]$ and $\pi_2 : [o, T] \times X \to X$ we denote the projections and for a subset $Z \subset \mathbb{R} \times X$ and $t \in \mathbb{R}$ we put

$$Z_t = \{x \in X : (t, x) \in Z\}.$$

Let (W, W^-) be a pair of subsets of $[0, T] \times X$. We call W a periodic isolating segment over [0,T] (for the equation $(*)$) and W^- the exit set of W if:

(i) W and W^- are compact ENR's, $W_0 = W_T$ and $W_0^- = W_T^-$,

(ii) there exists a homeomorphism

$$h : [0, T] \times (W_0, W_0^-) \to (W, W^-)$$

such that $\pi_1 = \pi_1 \circ h$,

(iii) for every $\sigma \in [0, T)$ and $x \in \partial W_\sigma$ there exists a $\delta > 0$ such that for every $t \in (0, \delta)$ either $\varphi_{(\sigma,t)}(x) \notin W_{\sigma+t}$ or $\partial_{(\sigma,t)}(x) \in \operatorname{int} W_{\sigma+t}$.

(iv) $W^- \cap ([0, T) \times X) = \{(\sigma, x) \in W : \sigma < T, \exists \delta > 0 \, \forall t \in (0, \delta) : \varphi_{(\sigma,t)}(x) \notin W_{\sigma+t}\}$.

The above definition is a slight modification of the notion of a periodic isolating block in [15]. Notice that a T-periodic isolating block (in the sense of [15]) can be easily obtained by glueing together translated copies of a periodic isolating segment over $[0, T]$.

Through remainder of this section we assume that W is a periodic isolating segment over $[0, T]$. We present some notions which relate to W. Put $S^1 = \mathbb{R}/T\mathbb{Z}$ and by $[t]$ denote the equivalence class of $t \in \mathbb{R}$ in S^1. By T-periodicity of φ, the local flow Φ on $\mathbb{R} \times X$ induces the local flow $\tilde{\Phi}$ with $S^1 \times X$ as the phase space. By the condition (i), the set

$$\tilde{W} = \{([t], x) \in S^1 \times X : x \in W_t, t \in [0, T]\}$$

is an isolating block in the usual sense in the theory of isolated invariant sets (see [8]). The exit set $\tilde{W}^-$ of that isolating block is equal to $\{([t], x) : x \in W_t^-, t \in [0, T]\}$. Define a map

$$\tau_W : W_0 \ni x \to \sup\{t \geqslant 0 : \forall s \in [0, t] : \tilde{\Phi}_s([0]), x) \in \tilde{W}\} \in [0, \infty].$$

τ_W is continuous (by the argument in a proof of Ważewski's Theorem, [8]). The set

$$I_W = \bigcap_{n=-\infty}^{\infty} \{x \in W_0 : \forall t \in [0, T] : \varphi_{(0,nT+t)}(x) \in W_t\}$$

is an isolated invariant set for the Poincaré map $\varphi_{(0,T)}$, the $\tilde{\Phi}$-trajectories of the points in $\{[0]\} \times I_W$ form the maximal invariant set inside $\tilde{W}$, denoted by inv $\tilde{W}$.

Define a homeomorphism

$$\tilde{h} : (W_0, W_0^-) \to (W_T, W_T^-) = (W_0, W_0^-)$$

by $\tilde{h}(x) = \pi_2(h(T, \pi_2 h^{-1}(0, x)))$ for $x \in W_0$. Geometrically, $\tilde{h}$ moves a point $x \in W_0$ to $W_T = W_0$ along the arc $h([0, T] \times \{h^{-1}(0, x)\})$. A different choice of the homomorphism h in (ii) leads to a map which is homotopic to $\tilde{h}$ (compare [15]), hence the automorphism

$$\mu_W = \tilde{h}_* : H(W_0, W_0^-) \to H(W_0, W_0^-),$$

induced by $\tilde{h}$ in singular homology, is an invariant of the block W. Recall that its Lefschetz number is defined as

$$\operatorname{Lef}(\mu_W) = \sum_{n=0}^{\infty} (-1)^n \operatorname{tr} \tilde{h}_{*n}.$$

In particular, if $\mu_W = \operatorname{id}_{H(W_0, W_0^-)}$ then Lef (μ_W) is equal to the Euler characteristic $\chi(W_0, W_0^-)$. In the sequel we will use the following theorem which, up to slightly different notation, was proved in [15]:

Theorem 1. (1) *If W is a periodic isolating segment over $[0,T]$ then the set*

$$F_W = \{x \in X : \varphi_{(0,T)}(x) = x, \forall t \in [0,T] : \varphi_{(0,t)}(x) \in W_t\}$$

is compact and open in the set of fixed points of $\varphi_{(0,T)}$ and the fixed point indese of $\varphi_{(0,T)}$ in F_W is given by

$$\operatorname{ind}(\varphi_{(0,T)}, F_W) = \operatorname{Lef}(\mu_W),$$

(2) *Let U, W be periodic isolating segments over $[0,T]$ and $U \subset W$. If*

$$\operatorname{Lef}(\mu_U) \neq \operatorname{Lef}(\mu_W)$$

then there is a fixed point $x \in W_0$ for $\varphi_{(o,T)}$ such that $\varphi_{(0,t)}(x) \in W_t$ for every $t \in [0,T]$ and $\varphi_{(0,t_0)}(x) \notin U_{t_0}$ for some $t_0 \in [0,T]$.

3. Main Results

Assume that $\Omega \subset \mathbb{R}^n$ is an open, convex set and φ is a T-periodic local process on $\operatorname{cl}\Omega$. Note that $S^1 \times \partial\Omega$ is automatically invariant for the local flow $\tilde{\Phi}$ generated by the process φ.

Definition 1. Let $U \subset [0,T] \times \operatorname{cl}\Omega$ be a periodic isolating segment over $[0,T]$. We say that U is a repelling type iff

(a) $\operatorname{inv}\tilde{U} \subset S^1 \times \partial\Omega$,

(b) $U^- = \partial U$, where ∂U denotes the boundary U in $[0, T] \times \operatorname{cl}\Omega$.

Our main result is the following

Theorem 2. *Assume that $U \subset W \subset [0,T] \times \operatorname{cl}\Omega$ are T-periodic segments over $[0,T]$. If*

(I) *U is of repelling type and* $\operatorname{inv}\tilde{W} \cap (S^1 \times \partial\Omega) = \operatorname{inv}\tilde{U}$,

(II) $\operatorname{Lef}(\mu_W) \neq 0$

then there is a fixed point $x \in W_0$ for $\varphi_{(0,T)}$ and $\varphi_{(0,t)}(x) \in \Omega$ for all $t \in \mathbb{R}$.

Proof. By (2) in Theorem 1 it suffices to prove that $\operatorname{Lef}(\mu_U) = 0$. It follows by

Lemma 1. *If U is of repelling type then $H(U_0, U_0^-) = 0$.*

Proof. Let

$$S = \{x \in U_0 : ([0], x) \in \operatorname{inv}\tilde{U}\} \subset \partial\Omega.$$

We show that U_0^- is a strong deformation retract of $U_0 \backslash S$. For $s \in [0,T]$ we put $h_s = \pi_2 h(s, \cdot) : (U_0, U_0^-) \to)(U_s, U_s^-)$ and define $H : (U_0 \backslash S) \times [0,T] \to U_0 \backslash S$ by

$$H(x,t) = h_0 \circ h^{-1}_{[t\tau_U(x)]}(\varphi_{(0,t\tau_U(x))}(x)).$$

This is well defined because $\tau_U(x) < +\infty$ for $x \in U_0 \backslash S$ and $\Omega_{(0,t\tau_U(x))}(x) \in U^-_{[t\tau_U(x))]}$. It is easy to check that H is a strong deformation retraction. So we have

$$H(U_0, U_0^-) \cong H(U_0, U_0 \backslash S) \cong H(\operatorname{cl}\Omega, \operatorname{cl}\Omega \backslash S).$$

The last isomorphism follows by the excision property of singular homology because S is compact and contained in int U_0. By assumption, $\operatorname{cl}\Omega$ is convex and

$S \subset \partial\Omega$, so any point $p \in \Omega$ is a strong deformation retract both $\mathrm{cl}\Omega$ and $\mathrm{cl}\Omega \backslash S$ (by radial deformation), hence $H(U_0, U_0^-) = 0$.

We apply Theorem 2 to obtain existence of positive periodic solutions for some time-dependent nonautonomuous differential equations. We put $\mathbb{R}_+ = [0, +\infty)$. Consider the equation in $\mathbb{R}_+^2$

$$\begin{cases} \dot{x} = x(a - by) \\ \dot{y} = y(c - dx), \end{cases} \tag{1}$$

where $a, b, c, d : \mathbb{R} \to \mathbb{R}$ are positive, T-periodic functions and regular enough to guarantee the uniqueness for the solutions of the Cauchy problem associated to (1). We prove the following

Theorem 3. *Equation (1) has T-periodic solutions such that* $(x(t), y(t)) \in \in (0, +\infty) \times (0, +\infty)$ *for all* $t \in \mathbb{R}$.

By f we denote the vector-field in the extended phase space $X = \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}$ generated by the right-hand side of the equation (1), i.e.

$$f(x, y, t) = (x(a(t) - b(t)y), y(c(t) - d(t)x), 1).$$

Our proof of Theorem 3 consists in the construction of two periodic segments U and W over $[0, T]$ satisfying the assumptions of Theorem 2 with $\Omega = = (0, +\infty) \times (0, +\infty)$. To this end, we will introduce several auxiliary functions and sets. Let $r, R, M > 0$. Put

$$\Sigma_r^1(x, y, t) = \frac{x}{r} - 1,$$

$$\Sigma_R^2(x, y, t) = \frac{y}{R} - 1,$$

$$\Sigma_M^3(x, y, t) = xy - M.$$

For a T-periodic continuous function $h : \mathbb{R} \to \mathbb{R}$ we put $h_1 = \inf$ h and $h_2 = \sup$ h.

Lemma 2. (a) *If* $x = r, y < \dfrac{a_1}{b_2}$ *then*

$$f(x, y, t) \nabla \Sigma_r^1(x, y, t) > 0.$$

(b) *If* $y = R, x < \dfrac{c_1}{d_2}$ *then*

$$f(x, y, t) \nabla \Sigma_R^2(x, y, t) > 0.$$

(c) *If* $xy = M, M = \dfrac{L_2}{d_1 b_1}$ *for some* $L > a_2 + c_2$ *then*

$$f(x, y, t) \nabla \Sigma_M^3(x, y, t) < 0.$$

Proof. A direct calculation shows that

$$f(x, y, t) \nabla \Sigma_r^1(x, y, t) = \frac{x}{r}(a(t) - b(t)y), \tag{2}$$

$$f(x, y, t) \nabla \Sigma_R^2(x, y, t) = \frac{y}{R}(c(t) - d(t)x), \tag{3}$$

$$f(x, y, t) \nabla \Sigma_M^3(x, y, t) = xy(a(t) - b(t)y + c(t) - d(t)x), \tag{4}$$

so (a) are easy to verify. In order to prove (c), it suffices to show that for all t

$$a(t) - b(t)y + c(t) - d(t)x < 0. \tag{5}$$

Because $y = \dfrac{L^2}{d_1 b_1 x}$ and $x^2 d_1 - xL + \dfrac{L^2}{d_1} > 0$, so

$$xd_1 + yb_1 > L.$$

We have

$$a(t) + c(t) \leqslant a_2 + c_2 < L < xd_1 + yb_1 \leqslant xd(t) + yb(t),$$

hence (5) holds.

Proof of Theorem 3. We put

$$W = \{(x,y,t) \in \mathbb{R}^2_+ \times [0,T] : \Sigma^1_{r_1}(x,y,t) \leqslant 0, \Sigma^2_{R_1}(x,y,t) \leqslant 0, \Sigma^3_M(x,y,t) \leqslant 0\},$$

where $r_1 = 2b_2 M/a_1, R_1 = 2d_2 M/c_1, M = L^2/d_1 b_1$ for some $L > a_2 + c_2$. For $r < \min(r_1, a_1/b_2), R < \min(R_1, c_1/d_2)$ we define

$$U + \{(x,y,t) \in \mathbb{R}^2_+ \times [0,T] : \Sigma^1_r(x,y,t) \leqslant 0, \Sigma^2_R(x,y,t) \leqslant 0\}.$$

It follows by (a) and (b) in Lemma 2 that U is an isolating segment over $[0,T]$, $U^- = \partial U$, and U is of repelling type. It is easy to check by Lemma 2 that

$$W^- = \{(x,y,t) \in W : \Sigma^1_{r_1}(x,y,t) = 0, \Sigma_{R_1}(x,y,t) = 0\}.$$

One can check that

$$\mathrm{Lef}(\mu_W) = \mathrm{Lef}\,(\mathrm{id}_{H(W_0, W_0^-)}) = \chi(W_0, W_0^-) = -1.$$

Moreover, by (a) and (b) in Lemma 2

$$\mathrm{inv}\tilde{W} \cap (S^1 \times \partial\Omega) = \mathrm{inv}\tilde{U} = \{(0,0)\},$$

hence Theorem 2 implies the result.

Assume that $a : \mathbb{R} \to (0, +\infty)$ is continuous T-periodic, $b : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ (or $b : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_-$) and $c : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ are continuous, T-periodic in the first variable, and the equation in $\mathbb{R} \times \mathbb{R}_+$

$$\begin{cases} \dot{x} = a(t)x + b(t,y) \\ \dot{y} = yc(t,y) \end{cases} \tag{6}$$

has the uniqueness property for the solutions of the associated Cauchy problem.

Theorem 4. *If there are* $0 < r < R$ *such that*
(A) $c(t,y) > 0, \forall t \in [0,T], 0 \leqslant y \leqslant r$,
(B) $c(t,y) < 0, \forall t \in [0,T], y = R$
then equation (6) *has T-periodic solutions such that* $(x(t), y(t)) \in \mathbb{R} \times (0, +\infty)$.

We prove Theorem 4 under the assumption that $b : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ because the case $b : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_-$ is similar.

Remark 2. One can take $c(t,y) = d(t) - e(t)y$, where d, e $: \mathbb{R} \to \mathbb{R}$ are T-periodic, positive maps. Then Theorem 4 applies with $r < \dfrac{d_1}{e_2}$ and $R > \dfrac{d_2}{e_1}$.

Put

$$S_p^1(x,y,t) = \frac{x^2}{p^2} - 1,$$
$$S_w^2(x,y,t) = \frac{y}{w} - 1.$$

Lemma 3. (1) *If* $x^2 = p^2, y \leqslant R$ *and* $p > M/a_1$ *where* $M = \sup_{[0,T]\times[0,R]} b$, *then*

$$f(x,y,t)\nabla S_p^1(x,y,t) > 0.$$

(2) *If* $y = r$, *then* $f(x,y,t)\nabla S_r^2 > 0$.
(3) *If* $y = R$, *then* $f(x,y,t)\nabla S_R^2 < 0$.

Proof. We prove only (1) because (2) and (3) are easy to check by

$$f(x,y,t)\nabla S_p^2 = \frac{y}{p}c(t,y).$$

We have

$$f(x,y,t)\nabla S_p^1 = \frac{2x}{p^2}(a(t)x + b(t,y)).$$

If $x = p$, then

$$\frac{2}{p}(a(t) + b(t,y)) > 0.$$

Suppose that $x = -p$. Then, by $p > M/a_1$, we obtain

$$f(x,y,t)\nabla S_p^1 = \frac{-2}{p}(a(t)(-p) + b(t,y)) > 0.$$

Proof of Theorem 4. The isolating segments U and W are defined by

$$U = \{(x,y,t) \in \mathbb{R} \times \mathbb{R}_+ \times [0,T] : S_p^1 \leqslant 0, S_r^2 \leqslant 0\},$$
$$W = \{(x,y,t) \in \mathbb{R} \times \mathbb{R}_+ \times [0,T] : S_p^1 \leqslant 0, S_R^2 \leqslant 0\},$$

for some $p > M/a_1$. It follows by Lemma 3 that $U^- = \partial U$ and U is of repelling type. Moreover,

$$\tilde{W} \cap (S^1 \times \mathbb{R} \times \{0\}) = \tilde{U}$$

and

$$\mathrm{Lef}(\mu_W) = \chi(W_0, W_0^-) = -1,$$

so we can apply Theorem 2.

Remark 3. Note that the same result can be obtained also via a more direct method based on the Brouwer fixed point theorem.

References

[1] Alvarez C, Lazer A (1980) An application of topological degree to the periodic competing species problem. J Austral Math Soc Ser B **28**: 202–219

[2] Butler GJ, Freedman HI (1981) Periodic solutions of a predator-prey system with periodic coefficients. Math Biosci **55**: 27–38
[3] Butler GJ, Waltman P (1986) Persistence in dynamical systems. J Diff Equations **63**: 255–263
[4] Capietto A, Zanolin F (1988) An existence theorem for periodic solutions in convex sets with applications. Results in Math **14**: 10–29
[5] Cushing JM (1977) Periodic time-dependent predator-prey systems. SIAM J Appl Math **32**: 82–95
[6] Cushing JM (1980) Two species competition in a periodic environment. J Math Biol **10**: 385–400
[7] De Mottoni P, Schiaffino A (1981) Competition systems with periodic coefficients: a geometric approach. J Math Biol **11**: 319–335
[8] Conley CC (1978) Isolated invariant set and the Morse index. CBMS Regional Conf Ser no 38. Providence, RI: Amer Math Soc
[9] Ding T, Huang H, Zanolin F (1995) A priori bounds and periodic solutions for class of planar systems with applications to Lotka-Volterra equations. Discrete and Continuous Dynamical Systems **1**: 103–117
[10] Ding T, Zanolin F (1996) Periodic solutions and subharmonic solutions for a class of planar systems of Lotka-Volterra type. In: Lakshmikantham V (ed) Proc First World Congress of Nonlinear Analysis 92, pp 395–406. Berlin: de Gruyter
[11] Fernandes MLC (1990) Uniform repellers for processes with applications to periodic differential systems. J Diff Equations **86**: 141–157
[12] Fernandes MLC, Zanolin F (1990) Repelling conditions for boundary sets using Liapunov-like function, II. Persistence and periodic solutions. J Diff Equations **86**: 33–58
[13] Gopalsamy K (1982) Exchange of equilibria in two species Lotka-Volterra system. J Austral Math Soc Ser B **24**: 160–170
[14] Hofbauer J, Sigmund K (1988) The Theory of Evolution and Dynamical Systems. Cambridge: Univ Press
[15] Srzednicki R (1994) Periodic and bounded solutions in block for time-periodic nonautonomous ordinary differential equations. Nonlinear Anal Theory Meth Appl **22**: 707–737
[16] Zanolin F (1992) Permanence and positive periodic solutions for Kolmogorow competing species systems. Results in Math **21**: 224–250

K. Wójcik
Institute of Mathematics
Jagiellonian University
ul. Reymonta 4
30-059 Kraków,
Poland
e-mail: wojcik@im.uj.edu.pl

Buchbesprechungen – Book Reviews

Maor, E.: *Die Zahl e – Geschichte und Geschichten* XII, 213 S. Birkhäuser, Basel, 1996. DM 70,–; öS 351,–.

Dieser Spaziergang durch die Geschichte der Analysis hat die Eulersche Zahl *e* als (lockeren) roten Faden, der nicht allzu genau genommen wird. Angereichert mit zahlreichen Geschichten (und auch Anekdoten), beginnt er bei John Napier (1614), führt zurück zu Archimedes, weiter bis zur Funktionentheorie des neunzehnten Jahrhunderts und endet mit Ausblicken ins zwanzigste Jahrhundert (Historisch interessierte) Mathematiker werden das Buch mit Vergnügen lesen, es wendet sich aber auch an eine breite Leserschaft – wenn man (wie man sollte) Mathematik als Teil der Allgemeinbildung (und der Kultur) betrachtet, sogar an alle, jedenfalls aber an Lehrer, Studenten und interessierte Schüler.

P. Schmitt, Wien

Machover, M.: *Set Theory, Logic and Their Limitations.* IX, 288 pp. Cambridge University Press, Cambridge, 1996. £ 40,–.

This is a book on set theory and logic leading up to a chapter on "Limitative results". A rigorous axiomatic presentation of Zermelo–Fraenkel set theory is given demonstrating how basic concepts of mathematics have been reduced to set theory. This is followed by a presentation of propositional and first-order logic. Concepts and results of recursion theory are explained in intuitive terms. Finally, the results of Skolem, Tarski, Church and Gödel are proved. Thus, this text will be well-suited for beginning students in mathematical logic and also for mathematicians, who will accept that axiomatic set theory is unable to characterize some of their most basic notions.

H. Mitsch, Wien

Buechler, St.: *Essential Stability Theory.* Perspectives in Mathematical Logic. XIV, 355 pp. Springer, Berlin Heidelberg New York, 1996. Cloth DM 170,–; öS 1226,–.

Stabilitätstheorie begann in den 60er Jahren mit Morleys Kategorizitätstheorem. Später wurde sie vor allem durch Shelahs Arbeiten geprägt. In den letzten 15 Jahren entwickelte sie sich in eine völlig andere Richtung: Geometrische Stabilitätstheorie, stabile Gruppen (Zil'ber, Poizat, Hrushovski), schuf also engere Verbindungen zur Mathematik, v.a. zur Geometrie und Algebra. Diese letzten Enwicklungen werden schwerpunktmäßig behandelt und erweitern somit den Inhalt der Bücher von Baldwin und Lascar. In den ersten beiden Kapiteln werden die relevanten Voraussetzungen aus der klassischen Modelltheorie kurz wiederholt. Es folgen die Grundlagen der Stabilitätstheorie (Morleys Kategorizitätstheorem, Baldwin–Lachlan Theorem). Weiters: Geometrische Stabilitätstheorie in überabzählbaren kategorischen Theorien, stabile Theorien (allgemein), superstabile Theorien, einige spezielle fortgeschrittene Themen. Das Buch bewegt sich auf hohem Niveau, enthält Übungen, kurze historische Bemerkungen, leider kein Symbolverzeichnis, 4 Seiten Literatur (mehr findet man im Buch von Baldwin).

P. Telec, Wien

Schumacher, C.: *Chapter Zero: Fundamental Notions of Abstract Mathematics* (Addison-Wesley Higher Mathematics). XIII, 162 pp. Addison-Wesley, Reading Menlo Park New York, 1996. US $ 47,50.

This textbook is intended to help students bridge the gap between mathematics at school and at the university (between the sophomore and the undergraduate level) and to provide an 'introduction to proofs'. Topics are the (canonical) 'foundations', i.e., (naïve) set theory, orderings, and basic number theory. Guided by the conviction that mathematics has to be learnt by *doing* mathematics, the course is written as a sequence of problems with interspersed informal discussions (and, in general, will require guidance by a teacher).

P. SCHMITT, Wien

Lowen, R.: *Fuzzy Set Theory.* Basic Concepts, Techniques and Bibliography. XIV, 408 pp. Kluwer, Dordrecht Boston London, 1996. US $ 199,–.

This book is intended as a reference (on basic notions) for readers who are mainly interested in applications of fuzzy set theory. Special features are 94 figures which illustrate various properties of fuzzy sets (e.g., t-(co) norms, logical operators) and a comprehensive bibliography (160 pages). It is a useful companion (sort of a dictionary) to more specialized texts – if one is prepared to pay its price.

P. SCHMITT, Wien

Just, W., Weese, M.: *Discovering Modern Set Theory. I. The Basics* (Graduate Studies in Mathematics, Vol. 8). XVII, 210 pp. American Mathematical Society, Providence, Rhode Island, 1996. US $ 36,–.

This is an introduction to (axiomatic) set theory (within mathematical logic), but intended for a general mathematical audience: 'Our most important criterion for inclusion of an item [i.e., concept, theorem, or proof technique] was frequency of use outside of pure set theory.' The main topics of the first volume are the axioms (Zermelo–Fraenkel), the axiom of choice, ordinal and cardinal numbers. The careful exposition, written in a lively and very readable style which addresses the reader rather directly, provides (by explanations, comments, and remarks) much information and motivation. Recommended.

P. SCHMITT, Wien

Yap, H. P.: *Total Colourings of Graphs* (Lecture Notes in Mathematics, Vol. 1623). VI, 131 pp. Springer, Berlin Heidelberg New York, 1996. DM 36,–.

A total colouring of a graph is a colouring of its vertices and edges such that both incident vertices and edges, and adjacent vertices as well as adjacent edges bear different colours. These lecture notes provide an (up-to-date) exposition (introduction to and reference of) this subject.

P. SCHMITT, Wien

Jungnickel, D. (Ed.): *Designs and Finite Geometries.* 254 pp. Kluwer, Boston Dordrecht London, 1996. US $ 120,–.

This collection consists of seventeen research papers, mainly on design theory and on finite geometries. It is dedicated to Hanfried Lenz on the occasion of the 80th anniversary of his birthday. Dieter Jungnickel and Günter Pickert have contributed a brief review of the mathematical work of Lenz (including a list of his publications). Remark: The volume is the book edition of the special issue of the journal *Designs, Codes and Cryptography* **8**/1–2, 1996.

P. SCHMITT, Wien

Krabs W.: *Mathematische Modellierung.* Eine Einführung in die Problematik. 12 Abb., 144 S. Teubner, Stuttgart, 1997. DM 24,80.

Neben der Beherrschung einer breiten Palette von Methoden gehört die Modellbildung zweifellos zu den wichtigsten Aufgaben eines angewandten Mathematikers. Obwohl

diesbezügliche Erfahrungen am besten selbst gemacht werden, ist die vorbildhafte Beschreibung anhand von interessanten Beispielen, wie sie in dem vorliegenden Band dargeboten wird, eine große Hilfe (aus den Bereichen Spieltheorie, Biologie, Medizin etc.). Sinnvollerweise werden im einleitenden Kapitel allgemeine Bewertungskriterien für konkrete mathematische Modelle diskutiert.

H. G. FEICHTINGER, Wien

Hoschek, J., Kalis, P. (Eds.): *Advanced Course on FAIRSHAPE.* 64 Figs., 288 pp. B. G. Teubner, Stuttgart, 1996. Softcover DM 65,–; öS 475,–.

This book is one of the results of an EU-based research project in the frame-work of the TMR-programme, i.e. a collection of articles presented at an advanced course on "Automatic Fairing and Shape-Preserving Methodologies in CAD/CAM" near Kaiserslautern in 1996. Depending on the particular situation the act of "fairing" a curve or surface is meant to improve some of its (mostly differential geometric) properties. Usually it is applied to some initial solution of the problem (say some spline function) in order to minimize an appropriately chosen cost function. The volume contains both theoretical contributions, examples taken from applications and benchmarks.

H. G. FEICHTINGER, Wien

Washington, L. C.: *Introduction to Cyclotomic Fields.* 2nd Edn. (Graduate Texts in Mathematics, Vol. 83). XIV, 487 pp. Springer, New York Berlin Heidelberg, 1996. Cloth DM 94,–.

The second edition of this well-known book has remained essentially unaltered with respect to its first edition. Two further chapters have been added. Chapter 15 contains a proof of Thaine, Kolyvagin, and Rubin of the Main Conjecture for the p-th cyclotomic field. The last chapter contains among other things a simple proof of Sinnott that Iwasawa's μ-invariant vanishes for abelian number fields. As for the rest I refer to my review of the first edition in Monatshefte für Mathematik **96** (1983).

J. SCHOISSENGEIER, Wien

Iserles, A. (Ed.): *Acta Numerica 1995.* 491 pp. Cambridge University Press, Cambridge, 1995. Cloth £ 35,–.

Iserles, A. (Ed.): *Acta Numerica 1996.* Volume 5. 395 pp. Cambridge University Press, Cambridge, 1996. Cloth £ 36,–.

In this series, a new book is published every year containing nice surveys of methods and progress in particular fields of numerical analysis, written by well-known specialists in these areas. – Acta Numerica 1995 contains seven articles on sequential quadratic programming, Pade approximation, error control in ordinary and partial differential equations, control problems and free boundary value problems for partial differential equations, particle methods for the Boltzmann equation, and news on the qd-algorithm for the tridiagonal eigen-value problem. – Acta Numerica 1996 contains seven articles on hierarchical bases, orthogonal polynomials, automatic grid generation, optimization problems involving eigenvalues, crystal microstructure computation, linear ordinary differential equations, and level set methods for surface propagation.

A. NEUMAIER, Wien

Lorenzini, D.: *An Invitation to Arithmetic Geometry* (Graduate Studies in Mathematics, Vol. 9). XV, 397 pp. American Mathematical Society, Providence, Rhode Island, 1996. Cloth US $ 59,–.

In order to come straight to the point: this book represents an excellent introduction to Algebraic Number Theory and to Algebraic Curves as well by viewing both theories as part of Commutative Algebra. In Chapter V the finiteness of the class number for number fields and for function fields is proved. Chapter VI is dedicated to projective curves, while the

seventh deals with curves over the non-algebraically closed fields. In Chapter VIII the rationality of the zeta function of a nonsingular curve over finite fields and its functional equation is proved. The proof of Riemann–Roch's Theorem is postponed to Chapter IX. The next one contains the proof of the Riemann Hypothesis for curves over finite fields, while the last one is devoted to various topics. The reviewer likes this book as all proof are given in full detail and its concept is well thought-out.

J. SCHOISSENGEIER, Wien

Goss, D.: *Basic Structures of Function Field Arithmetic* (Ergebnisse der der Mathematik und ihrer Grenzgebiete, 3. Folge, Vol. 35). XIII, 422 pp. Springer, Berlin Heidelberg New York, 1996. Cloth DM 178,–.

In generalizing the classical exponential function to the case of characteristic p, Carlitz, investigating the arithmetic of curves over finite fields, introduced the so-called Carlitz module. This concept was further generalized by Drinfeld in 1974, who aimed at reciprocity laws (Langlands philosophy). This book is dedicated to the arithmetic of function fields over finite fields, the characteristic polynomials of the Frobenius morphism at a prime, their Euler product, which yields to an L-function in characteristic p and the many open problems connected with it. It is written on a high level and the reader should be at least familiar with classical class field theory.

J. SCHOISSENGEIER, Wien

Morandi, P.: *Field and Galois Theory* (Graduate Texts in Mathematics, 167). 18 Figs., XI, 281 pp. Springer, New York Berlin Heidelberg, 1996. Cloth DM 70,–; öS 496,–.

This is a honest introduction to the classical stock of field and Galois theory with scarcely any surprise – most of the material can be found in books on algebra, e.g. in Hungerford's. The few highlights worth to mention are small chapters on Kummer extensions, infinite algebraic extensions, algebraic varieties and function fields, derivations and differentials. 50 pages of Appendices on algebraic and topological fundamentals make the book rather self-contained.

G. KOWOL, Wien

Andradas, C., Bröcker, L., Ruiz, J. M.: *Constructible Sets in Real Geometry* (Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge Vol. 33). IX, 270 pp. Springer, Berlin Heidelberg New York, 1996. Cloth DM 160,–; öS 1153,–.

This book aims at the presentation of the different kinds of constructible sets in real geometry like semialgebraic sets, semianalytic sets (resp. set germs) from a unifying point of view. It makes necessary the study of a lot of abstract settings as real spectra of rings, spaces of ordering resp. signs, excellent rings etc. Since most of the results concerning that topic have been derived in the last decade – a good many even are proved here for the first time – and have never been presented in a book, the prerequisites for its understanding are manifold. But the interested reader is rewarded with a profound view of this lively field of research.

G. KOWOL, Wien

King, R. B.: *Beyond the Quartic Equation.* VIII, 149 pp. Birkhäuser, Basel Berlin Boston, 1996. Cloth DM 81,–; öS 570,–.

This book focusses on the solution of the general quintic, given some 120 years ago. In particular it presents the method of L. Kiepert (1878) which is adapted for use on a microcomputer. Introductory chapters more or less without proofs on group theory and symmetry, Galois theory, elliptic functions, solution by radicals, as well as an outline of the solution of algebraic equations of higher degree allow the non-specialist to glance at the background of the theory.

G. KOWOL, Wien

Tamme, G.: *Introduction to Étale Cohomology.* IX, 190 pp. Springer, Berlin Heidelberg New York, 1994. DM 85,–; öS 608,40,–.

The applications of Étale Cohomology in modern Algebraic Geometry and Number Theory are manifold. This book is an introduction to this cohomology theory. You will find therein all preliminaries from topology, the theory of sheaves, their spectral sequences and category theory. There are some examples in the book and a special chapter on the Étale Cohomology of curves. Some theorems (like the comparison theorem with classical cohomology) are stated without proofs. Be aware if you are not motivated for Étale Cohomology: You will not find in this book what Étale Cohomology is really good for.

M. NEUWIRTHER, Wien

Shafarevich, I. R. (Ed.): *Algebraic Geometry II.* Cohomology of Algebraic Varieties. Algebraic Surfaces (Encyclopedia of Mathematics, Vol. 55). 262 pp. Springer, Berlin Heidelberg New York, 1996. Cloth DM 140,–; öS 1080,–.

This volume, a translation from the russian edition 1989, contains two articles. The first one is 'Cohomology of Algebraic Varieties', by V. I. Danilov, which gives a wide view of the cohomology of the underlying topological spaces as well as of the more algebraic versions of cohomology (like etale cohomology) which were developed in the wake of the Weil conjectures. The second article is 'Algebraic Surfaces', by V. A. Iskovskikh and I. R. Shafarevich. It paints a cohesive picture of the theory of algebraic surfaces, explains its problems, and describes its main methods. As usual in this series the reader has to go for detailed proofs elsewhere.

P. MICHOR, Wien

Du, D.-Z., Pardalos, P. M.: *Minimax and Applications* (Nonconvex Optimization and its Applications, Vol. 5). XIV, 292 pp. Kluwer, Dordrecht Boston London, 1995. Cloth US $ 149,–.

First encountered in game theory and later in duality theory in linear and convex programming, minimax theory has found many more applications in optimization, for example in two-stage or bilevel programming, network problems, repellant sampling, resource allocation. Minimax problems are also at the heart of many combinatorial optimization problems such as the multiple counterfeit coin problem, various optimal triangulation problems, clustering problems, etc. This collection of papers in this book provides a representative picture of recent research in minimax theory, algorithmic approaches for finding solutions of minimax problems, and (mainly) applications.

A. NEUMAIER, Wien

Kollár, J.: *Rational Curves on Algebraic Varieties* (Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 32). 320 pp. Springer, Berlin Heidelberg New York, 1996. Cloth DM 160,–; öS 1153,–.

This book provides an introduction to the structure theory of higher dimensional ($\geqslant 3$) varieties by studying the geometry of rational curves in them. The first (technical) chapter is devotes to deformation theory of curves on varieties, i.e., the theory of Hilbert schemes and Chow varieties in the relative setting. Then follow curves on varieties, the cone theorem and minimal models, rationally connected varieties, and Fano varieties.

P. MICHOR, Wien

Lozansky, E., Rousseau, C.: *Winning Solutions* (Problem Books in Mathematics). X, 244 pp. Springer, Berlin Heidelberg New York, 1996. Softcover DM 58,–; öS 423,40.

Dieses Buch richtet sich an begabte und interessierte Mathematikstudenten und hat das erstrebenswerte Ziel, einen Beitrag zu leisten, die Lücke zwischen dem üblichen Mathematikunterricht und den Programmen der internationalen Mathematikolympiaden zu schließen. Vorwiegend werden Probleme aus der Zahlentheorie, Algebra und

Kombinatorik im Form von Aufgaben, Beispielen, Hinweisen und Lösungen behandelt, wobei erfreulicherweise auch ein tieferes mathematisches Verständnis gefördert wird.

H. RINDLER, Wien

Sahai, H., Khurshid, A.: *Statistics in Epidemiology.* Methods, Techniques, and Applications. 321 pp. CRC Press, Boca Raton New York London, 1996. Cloth $ 59,95.

This book provides a collection of statistical methods that are important for data analysis in epidemiology. It is written for epidemiologists and requires only basic knowledge in statistics. The book also contains a section of 50 exercises and fourteen appendices, that treat some of the mathematical foundations of the methods presented in the main text.

R. BÜRGER, Wien

Yap, T. K., Frieder, O., Martino, R. L.: *High Performance Computational Methods for Biological Sequence Analysis.* 15 Tabs., 11 Figs., XIII, 211 pp. Kluwer, Boston London Dordrecht, 1996. Cloth US $ 93,50.

The sheer quantity of data gained within the human genome project or the sequencing of DNA of other organisms calls for the use of high performance computing for any data storage, retrieval and analysis. This book is a practical guide to a variety of aspects concerning sequence analysis. Fortunately (for the interested mathematicians) it provides an understandable exposition of the biological and technical background concerning sequencing etc. Various sequencing analysis algorithms (concerning alignment under various circumstances etc.) are described. There are intimate connections to the computer architecture which are discussed amply. A detailed chapter explains how the databases may be accessed via Internet. – This is clearly not a mathematical book in the narrow sense but one which allows every interested mathematician (physicist, ...) to get basic acquaintance with its subject.

H. MUTHSAM, Wien

Verleger: Springer-Verlag KG, Sachsenplatz 4–6, A-1201 Wien. — Herausgeber: Prof. Dr. Karl Sigmund, Institut für Mathematik der Universität Wien, Strudlhofgasse 4, A-1090 Wien. — Redaktion: Strudlhofgasse 4, A-1090 Wien. — Satz und Umbruch: Thomson Press (India) Ltd., New Delhi. — Offsetdruck: Druckerei Eugen Ketterl Ges.m.b.H., A-1180 Wien. Verlagsort: Wien. — Herstellungsort: Wien. — Printed in Austria.