SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Mathematische Zeitschrift
Verlag: Springer
Jahr: 1991
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN266833020_0206
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN266833020_0206

Übergeordnetes Werk

Werk Id: PPN266833020
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN266833020

Terms and Conditions

The Goettingen State and University Library provides access to digitized documents strictly for noncommercial educational, research and private purposes and makes no warranty with regard to their use for other purposes. Some of our collections are protected by copyright. Publication and/or broadcast in any form (including electronic) requires prior written permission from the Goettingen State- and University Library.
Each copy of any part of this document must contain there Terms and Conditions. With the usage of the library's online system to access or download a digitized document you accept the Terms and Conditions.
Reproductions of material on the web site may not be made for or donated to other repositories, nor may be further reproduced without written permission from the Goettingen State- and University Library.

For reproduction requests and permissions, please contact us. If citing materials, please give proper attribution of the source.

Contact

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

Mathematische Zeitschrift

Band 206 · 1991

Herausgegeben von / Edited by
E. Becker, Dortmund
G. Carlsson, Princeton
K. Diederich, Wuppertal
G. Faltings, Princeton
H. Heyer, Tübingen
J. Jost, Bochum
H. Knörrer, Zürich
M. Struwe, Zürich
E. Zehnder, Zürich

Wissenschaftlicher Beirat / Advisory Board
W. P. Barth, Erlangen · K. H. Hofmann, Darmstadt
J. C. Jantzen, Eugene · W. Klingenberg, Bonn
H. H. Schaefer, Tübingen · L. Smith, Göttingen
W. v. Wahl, Bayreuth

Springer International

"Mathematische Zeitschrift" was founded in 1918 and edited by *L. Lichtenstein* in cooperation with *K. Knopp, E. Schmidt* and *I. Schur;* after Lichtenstein's death, 1933, it was edited by *K. Knopp.* The Editorial Committee was increased to include *E. Kamke* and *F.K. Schmidt* in 1933, *R. Nevanlinna* in 1936 and *H. Wielandt* in 1950. The latter became Managing Editor in 1952. He was succeeded in 1973 by *H. Salzmann,* 1977 by *H. Heyer* and 1984 by *W.P. Barth* and *W. v. Wahl.* As of 1990 *E. Becker* is the Managing Editor.

Die **„Mathematische Zeitschrift"** wurde im Jahre 1918 von *L. Lichtenstein* unter der Mitwirkung von *K. Knopp, E. Schmidt* und *I. Schur* gegründet und herausgegeben. Nach dem Tode Lichtensteins übernahm *K. Knopp* 1933 die Herausgabe. Die Schriftleitung ergänzte sich 1933 durch *E. Kamke* und *F.K. Schmidt,* 1936 durch *R. Nevanlinna* und 1950 durch *H. Wielandt,* der 1952 die Herausgabe übernahm. 1973 ging die Federführung an *H. Salzmann,* 1977 an *H. Heyer* und 1984 an *W.P. Barth* und *W. v. Wahl* über. Seit 1990 ist *E. Becker* geschäftsführender Herausgeber.

Manuscripts may be sent to / Manuskripte nehmen entgegen:

E. Becker, FB Mathematik, Universität Dortmund, Postfach 50 05 00, W-4600 Dortmund 50, Federal Republic of Germany

G. Carlsson, Department of Mathematics, University of Princeton, Fine Hall – Washington Road, Princeton, NJ 08544, USA

K. Diederich, Gesamthochschule, Fachbereich 7, Mathematik, Gaußstr. 20, W-5600 Wuppertal 1, Federal Republic of Germany

G. Faltings, Department of Mathematics, University of Princeton, Fine Hall – Washington Road, Princeton, NJ 08544, USA

H. Heyer, Mathematisches Institut der Universität, Auf der Morgenstelle 10, W-7400 Tübingen, Federal Republic of Germany

J. Jost, Mathematisches Institut der Universität, Postfach 10 21 48, W-4630 Bochum, Federal Republic of Germany

H. Knörrer, Departement Mathematik, ETH-Zentrum, Rämistr. 101, CH-8092 Zürich, Switzerland

M. Struwe, Departement Mathematik, ETH-Zentrum, Rämistr. 101, CH-8092 Zürich, Switzerland

E. Zehnder, Departement Mathematik, ETH-Zentrum, Rämistr. 101, CH-8092 Zürich, Switzerland

Copyright

Submission of a manuscript implies: that the work described has not been published before (except in the form of an abstract or as part of a published lecture, review, or thesis); that it is not under consideration for publication elsewhere; that its publication has been approved by all coauthors, if any, as well as by the responsible authorities at the institute where the work has been carried out; that, if and when the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the publisher (society); and that the manuscript will not be published elsewhere in any language without the consent of the copyright holders.

All articles published in this journal are protected by copyright, which covers the exclusive rights to reproduce and distribute the article (e.g., as offprints), as well as all translation rights. No material published in this journal may be reproduced photographically or stored on microfilm, in electronic data bases, video disks, etc., without first obtaining written permission from the publisher.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if not specifically identified, does not imply that these names are not protected by the relevant laws and regulations.

While the advice and information in this journal is believed to be true and accurate at the date of its going to press, neither the authors, the editors, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Special regulations for photocopies in the USA: Photocopies may be made for personal or in-house use beyond the limitations stipulated under Section 107 or 108 of U.S. Copyright Law, provided a fee is paid. This fee is US $0.20 per page, or a minimum of US $1.00 if an article contains fewer than five pages. All fees should be paid to the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, USA, stating the ISSN 0025-5874, the volume, and the first and last page numbers of each article copied. The copyright owner's consent does not include copying for general distribution, promotion, new works, or resale. In these cases, specific written permission must first be obtained from the publisher.

Printers: Universitätsdruckerei H. Stürtz AG, W-8700 Würzburg, Federal Republic of Germany
© Springer-Verlag, Berlin Heidelberg 1991
Springer-Verlag GmbH & Co. KG, W-1000 Berlin 33, Federal of Republic of Germany
Printed in Germany

NIEDERSÄCHS. STAATS U UNIV.- BIBLIOTHEK GÖTTINGEN

Inhalt/Contents

Amar, E.: On the extension of C.R. functions 89
Appell, J., Nazarov, V.I., Zabrejko, P.P.: Composing infinitely differentiable functions 659

Baum, H.: An upper bound for the first eigenvalue of the Dirac operator on compact spin manifolds 409
Bendersky, M.: Applications of the Ochanine Genus 443
Bessenrodt, C.: The Auslander-Reiten quiver of a modular group algebra revisited 25
Bieri, R., Hillman, J.A.: Subnormal subgroups of 3-dimensional Poincaré duality groups 67
Boas, H.P., Straube, E.J.: Sobolev estimates for the $\bar{\partial}$-Neumann operator on domains in C^n admitting a defining function that is plurisubharmonic on the boundary 81
Braun, R.: On a geometric property of the normal bundle of surfaces in $\mathbb{P}_4$ 535
Brown, K.A.: The Artin algebras associated with differential operators on singular affine curves 423
Bunce, L.J., Maitland Wright, J.D.: Complex borel structure in *JC*-algebras 225

Chen, Z.: The existence of algebraic surfaces with preassigned Chern numbers 241
Cheng, Q.-m.: Complete space-like submanifolds in a de Sitter space with parallel mean curvature vector 333

Dancer, E.N.: On the number of positive solutions of some weakly nonlinear equations on annular regions 551
Dupain, Y.: Extension a la dimension n d'un théorème de Ortel et Schneider 71

Enguehard, M.: Isométries parfaites entre blocs de groupes linéaires ou unitaires 1

Farnsteiner, R.: Cohomology groups of reduced enveloping algebras 103
Farnsteiner, R., Strade, H.: Shapiro's lemma and its consequences in the cohomology theory of modular Lie algebras 153
Fitzgerald, R.W.: Picard groups of Witt rings 303

Gamboa, J.M., Ruiz, J.M.: On rings of semialgebraic functions 527
Gellhaus, C.: Äquivariante Kompaktifizierung des $\mathbb{C}^n$ 211
Green, E.L., Zimmermann-Huisgen, B.: Finitistic dimension of artinian rings with vanishing radical cube 505
Grunewald, F., Mennicke, J., Vaserstein, L.: On symplectic groups over polynomial rings 35
Guangxin, Z.: On preordered fields related to Hilbert's 17th problem 145

Hillman, J.A. → siehe Bieri, R.
Hora, A.: Quasi-invariant measures on commutative Lie groups of infinite product type 169

Ito, H.: Action-angle coordinates at singularities for analytic integrable systems 363

Jong, A.J. de, Steenbrink, J.H.M.: Picard numbers of surfaces in 3-dimensional weighted projective spaces 341
Joshua, R.: Riemann Roch for equivariant $\mathfrak{D}$-modules-I 131
Joshua, R.: The trace formula for equivariant $\mathfrak{D}$-modules and perverse sheaves 293

Kellerhals, R.: On Schläfli's reduction formula 193
Knus, M.-A., Parimala, R., Sridharan, R.: Pfaffians, central simple algebras and similitudes 589

Lempert, L.: A precise result on the boundary regularity of biholomorphic mappings (Erratum) 501
Li, A.-M. → siehe Vrancken, L.
Loos, O.: Finiteness conditions in Jordan pairs 577

Maitland Wright, J.D. → siehe Bunce, L.J.
Mennicke, J. → siehe Grunewald, F.
Morita, J., Rehmann, U.: Symplectic K_2 of Laurent polynomials, associated Kac-Moody groups and Witt rings 57

Nakamura, H.: On galois automorphisms of the fundamental group of the projective line minus three points 617
Nakao, M.: Remarks on the existence and uniqueness of global decaying solutions of the nonlinear dissipative wave equations 265
Nazarov, V.I. → siehe Appell, J.

Ochanine, S.: Elliptic genera, modular forms over KO_*, and the Brown-Kervaire invariant 277
Ohsawa, T.: Cheeger-Goreski-MacPherson's conjecture for the varieties with isolated singularities 219
Otsu, Y.: On manifolds of positive Ricci curvature with large diameter 255

Palmer, B.: Index and stability of harmonic Gauss maps 563
Parimala, R. → siehe Knus, M.-A.
Petz, D.: On certain properties of the relative entropy of states of operator algebras 351
Pinney, K.R.: Line bundle convexity of pseudoconvex domains in complex manifolds 605

Rabinowitz, P.H., Tanaka, K.: Some results on connecting orbits for a class of Hamiltonian systems 473
Ramm, A.G., Sjöstrand, J.: An inverse problem of the wave equation 119
Rehmann, U. → siehe Morita, J.
Ruiz, J.M. → siehe Gamboa, J.M.

Saal, L., Urciuolo, M.: Singular integrals supported on the image of an analytic function 233
Schmüdgen, K.: Non-commutative moment problems 623
Sekiguchi, T., Suwa, N.: A note on extensions of algebraic and formal groups I 567
Simon, U. → siehe Vrancken, L.
Sjöstrand, J. → siehe Ramm, A.G.
Skora, R.K.: Knot and link projections in 3-manifolds 345
Sridharan, R. → siehe Knus, M.-A.
Steenbrink, J.H.M. → siehe Jong, A.J. de
Strade, H. → siehe Farnsteiner, R.
Straube, E.J. → siehe Boas, H.P.
Suwa, N. → siehe Sekiguchi, T.

Tanaka, K. → siehe Rabinowitz, P.H.

Urciuolo, M. → siehe Saal, L.

Vaserstein, L. → siehe Grunewald, F.
Vrancken, L., Li, A.-M., Simon, U.: Affine spheres with constant affine sectional curvature 651

Wolter, T.H.: Einstein Metrics on solvable groups 457

Ye, R.: On the existence of area-minimizing surfaces with free boundary 321

Zabrejko, P.P. → siehe Appell, J.
Zhou, X.: On the decomposition map of Grothendieck groups 533
Zimmermann-Huisgen, B. → siehe Green, E.L.

Covered by *Zentralblatt für Mathematik* and *Current Mathematical Publication*

A 20519

Mathematische Zeitschrift

Band 206 Heft 1 1991

2 Qel 1-473 54

Herausgegeben von / Edited by
E. Becker, Dortmund
G. Carlsson, Princeton
K. Diederich, Wuppertal
G. Faltings, Princeton
H. Heyer, Tübingen
J. Jost, Bochum
H. Knörrer, Zürich
M. Struve, Zürich
E. Zehnder, Zürich

Wissenschaftlicher Beirat / Advisory Board

W.P. Barth, Erlangen · K.H. Hofmann, Darmstadt
J.C. Jantzen, Eugene · W. Klingenberg, Bonn
H.H. Schaefer, Tübingen · L. Smith, Göttingen
W. v. Wahl, Bayreuth

8 Z Vol. 527
Z 4.156

Springer International

209 Math. Z. ISSN 0025-5874 MAZEAX 206 (1) 1-168 (1991) January 1991
Printed on acid-free paper

Staats- u. Universitätsbibliothek
Eing. 07. Feb. 1991
— Lesesaal —

K.82

Mathematische Zeitschrift

Die **„Mathematische Zeitschrift"** wurde im Jahre 1918 von *L. Lichtenstein* unter der Mitwirkung von *K. Knopp, E. Schmidt* und *I. Schur* gegründet und herausgegeben. Nach dem Tode Lichtensteins übernahm *K. Knopp* 1933 die Herausgabe. Die Schriftleitung ergänzte sich 1933 durch *E. Kamke* und *F.K. Schmidt,* 1936 durch *R. Nevanlinna* und 1950 durch *H. Wielandt,* der 1952 die Herausgabe übernahm. 1973 ging die Federführung an *H. Salzmann,* 1977 an *H. Heyer* und 1984 an *W.P. Barth* und *W. v. Wahl* über. Seit 1990 ist *E. Becker* geschäftsführender Herausgeber.

"Mathematische Zeitschrift" was founded in 1918 and edited by *L. Lichtenstein* in cooperation with *K. Knopp, E. Schmidt* and *I. Schur;* after Lichtenstein's death, 1933, it was edited by *K. Knopp.* The Editorial Committee was increased to include *E. Kamke* and *F.K. Schmidt* in 1933, *R. Nevanlinna* in 1936 and *H. Wielandt* in 1950. The latter became Managing Editor in 1952. He was succeeded in 1973 by *H. Salzmann,* 1977 by *H. Heyer* and 1984 by *W.P. Barth* and *W. v. Wahl.* As of 1990 *E. Becker* is the Managing Editor.

Die **„Mathematische Zeitschrift"** dient der Pflege der reinen und angewandten Mathematik, doch werden auch Beiträge aus den Gebieten der theoretischen Physik und Astronomie Aufnahme finden, soweit sie mathematisch von Interesse sind. Besprechungen, Aufgaben u. dgl. werden nicht zugelassen.

„Mathematische Zeitschrift" is devoted to pure and applied mathematics; papers on theoretical physics and astronomy may be accepted if they present interesting mathematical results. Reviews, problems etc. will not be published.

Von jeder Arbeit werden 100 Sonderdrucke kostenlos zur Verfügung gestellt, weitere können zum Selbstkostenpreis bezogen werden.

100 offprints of each paper are provided free of charge; additional copies may be ordered at cost price.

Hinweis für Autoren

Die Autoren werden gebeten, ihre Manuskripte gemäß den „Instructions to Authors" für den Satz vorzubereiten. Ungenügend vorbereitete Manuskripte müßten von der Redaktion zurückgesandt werden, was die Publikation verzögern würde. Um die Kommunikation zwischen Autor, Herausgeber und Verlag zu erleichtern und zu beschleunigen, sollte der Autor seine Telex- oder Fax-Verbindung auf der Titelseite des Manuskriptes angeben.

Notice to Authors

Authors are asked to prepare their manuscripts according to the "Instructions to Authors". Otherwise the manuscripts may have to be returned to the authors, which will cause a delay in publication. To facilitate communication between the authors, editors and publisher, the author should furnish a telex or fax number on the title page of the manuscript.

Geschäftsführender Herausgeber/Managing Editor

E. Becker FB Mathematik, Universität Dortmund, Postfach 50 05 00, W-4600 Dortmund 50, Federal Republic of Germany

Manuskripte nehmen außerdem entgegen/Manuscripts may also be sent to:

G. Carlsson Department of Mathematics, University of Princeton, Fine Hall – Washington Road, Princeton, NJ 08544, USA

K. Diederich Gesamthochschule, Fachbereich 7, Mathematik, Gaußstr. 20, W-5600 Wuppertal 1, Federal Republic of Germany

G. Faltings Department of Mathematics, University of Princeton, Fine Hall – Washington Road, Princeton, NJ 08544, USA

H. Heyer Mathematisches Institut der Universität, Auf der Morgenstelle 10, W-7400 Tübingen, Federal Republic of Germany

J. Jost Mathematisches Institut der Universität, Postfach 10 21 48, W-4630 Bochum, Federal Republic of Germany

H. Knörrer Mathematisches Seminar, ETH-Zentrum, Rämistr. 101, CH-8092 Zürich, Switzerland

M. Struve Mathematisches Seminar, ETH-Zentrum, Rämistr. 101, CH-8092 Zürich, Switzerland

E. Zehnder Mathematisches Seminar, ETH-Zentrum, Rämistr. 101, CH-8092 Zürich, Switzerland

Math. Z. 206, 1–24 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Isométries parfaites entre blocs de groupes linéaires ou unitaires

Michel Enguehard
D.M.I., Ecole Normale Supérieure, 45 rue d'Ulm, F-75005 Paris, France

Received April 7, 1989; in final form October 18, 1989

Soit l un nombre premier. Récemment M. Broué a introduit la notion d'*isométrie parfaite* entre l-blocs de groupes finis, ou plutôt entre les groupes de Grothendieck de ces blocs, suggérant que certaines similitudes entre blocs pouvaient résulter d'équivalences entre catégories dérivées (voir [3], la définition d'une isométrie parfaite est rappelée en 3.2).

J'ai montré dernièrement, en m'appuyant sur la formule de Murnaghan-Nakayama et un résultat de type combinatoire – rappelé en 1.4 –, qu'il existe une isométrie parfaite entre deux blocs de groupes symétriques dès que leurs poids sont égaux (voir [5]). Je démontre ici une assertion analogue sur les groupes linéaires et unitaires (1.13; le groupe de Grothendieck d'un groupe fini ou d'un bloc X sur un corps de caractéristique nulle Ω convenablement choisi est noté $\mathfrak{R}(X)$, notations en 1.1):

Théorème. *Soient G et $\tilde{G}$ deux groupes linéaires (ou deux groupes unitaires) sur un même corps fini de caractéristique différente de l. Soient B et $\tilde{B}$ des l-blocs de G et $\tilde{G}$ respectivement. Si les poids de B et de $\tilde{B}$ sont égaux, il existe une isométrie parfaite de $\mathfrak{R}(B)$ sur $\mathfrak{R}(\tilde{B})$.*

Les caractères irréductibles ordinaires des groupes finis linéaires et unitaires et leur répartition en l-blocs, (l autre que la caractéristique) ont été décrits par J.A. Green, G. Lusztig, B. Srinivasan, P. Fong, M. Broué (voir [8, 10, 7, 2]). L'invariant «poids» d'un bloc de groupe linéaire ou unitaire (redéfini ici en 1.11) a déjà été introduit par G.O. Michler et J.B. Olsson dans [11], où est en particulier vérifiée la conjecture d'Alperin sur les caractères de hauteur nul. Que deux blocs de groupes linéaires ou de groupes unitaires aient le même nombre d'irréductibles modulaires et le même nombre d'irréductibles ordinaires si leurs poids sont égaux est donc un fait déjà établi. Mais l'existence d'une isométrie parfaite dit plus. Une isométrie parfaite entre deux l-blocs de groupes finis induit un isomorphisme entre les centres de leurs algèbres sur un anneau de valuation convenable ainsi qu'une bijection préservant la hauteur entre leurs ensembles de caractères irréductibles ordinaires. En outre, les deux blocs ont même nombre de caractères modulaires, les invariants de similitude des matrices de décomposition (donc aussi ceux des matrices de Cartan) sont égaux (voir [3]).

Le théorème 1.13 et sa démonstration on fait l'objet d'une conférence au colloque de Bad Honnef «Darstellungstheorie endlicher Gruppen und endlichdimensionaler Algebren», le 30 juin 1988.

Disposant d'ores et déjà d'un théorème combinatoire analogue à 1.4 et portant sur les symboles, je prévois une extension du théorème 1.13 aux groupes «de Lie finis» de type **B**, **C** ou **D**.

Cet article comprend trois chapitres. Dans le capitre 1 sont rappelés certains faits concernant les groupes symétriques et la géométrie des crochets d'une partition sous-jacente à la formule de Murnaghan-Nakayama (1.2), puis les groupes linéaires et unitaires (1.6 à 1.12). Les groupes de Grothendieck des groupes symétriques peuvent être présentées en somme directe orthogonale: $\mathfrak{R}\mathfrak{S} := \oplus_n \mathfrak{R}(\mathfrak{S}_n)$ munie d'une structure d'algèbre de Hopf sur $\mathbb{Z}$ graduée (voir A.V. Zelevinsky, [13]). J'ai jugé utile de définir (en 1.3), pour tout d entier naturel, une loi de comodule gradué de $\mathfrak{R}\mathfrak{S}$ sur elle-même. L'existence de ces lois se déduit d'un théorème de Farahat [6]. Le théorème combinatoire 1.4, tiré de [5], qui met en relation par une bijection assortie d'une application signe, les ensembles de partitions associés à deux «e-cœurs» (i.e. partitions sans crochet de longueur e), est suffisant pour traiter le cas des blocs de poids égaux de groupes symétriques dans [5] (prendre $e=l$). Si e divise d, l'application décrite en 1.4 conserve la structure de comodule définie en 1.3 (corollaire 1.5, dont on verra l'usage au chapitre 2).

Soit $\varepsilon \in \{-1, 1\}$. Je note $\mathrm{GL}(n, \varepsilon q)$ un groupe linéaire en dimension n sur le corps $\mathbb{F}_q$ de cardinal q (si $\varepsilon = 1$) ou unitaire en dimension n relativement à l'automorphisme d'ordre 2 de $(\mathbb{F}_{q^2}/\mathbb{F}_q)$ (si $\varepsilon = -1$). En ce qui concerne la classification des éléments et des irréductibles d'un tel groupe et les calculs de caractères liés à l'autodualité (1.6 à 1.10), je me suis inspiré de l'article de F. Digne et J. Michel, [4]. La classification «à la Jordan» en «séries de Lusztig» des irréductibles des groupes finis du type de Lie, par le biais des applications dites de Lusztig R_L^G (voir [9]) est un outil essentiel pour répartir les irréductibles selon les l-blocs; ces blocs sont décrits en 1.11, où je peux enfin donner une définition du poids. En 1.12, j'ai prolongé la description de l'algèbre de Hopf $\mathfrak{R}\mathfrak{G} := \oplus_n \mathfrak{R}(\mathrm{GL}(n, \varepsilon q))$ tirée de l'article [4] déjà cité en donnant sa décomposition selon les l-blocs, en fait selon des e-cœurs de partitions.

On voit facilement, grâce à des considérations générales sur les blocs des groupes finis du type de Lie dues à M. Broué et J. Michel, qu'il suffit de démontrer le théorème 1.13 dans le seul cas où les blocs considérés contiennent des caractères unipotents; dans ce cas le poids se réduit à un entier naturel. Cette démonstration, par induction sur le poids, est l'objet des chapitres 2 et 3. Si le poids est nul, les groupes de défaut sont triviaux et les blocs projectifs. Un invariant important est l'ordre e de εq modulo l. Le bloc B est défini par son poids et un e-cœur. Si $e=1$, les groupes G et $\tilde{G}$ sont isomorphes et B et $\tilde{B}$ sont leurs blocs principaux. Si $e \neq 1$, les centres de G et $\tilde{G}$ sont d'ordres premiers à l, ce qui permet d'utiliser l'induction sur les centralisateurs de l-éléments. Dans tous les cas les groupes de défaut sont isomorphes, et on en déduit une bijection naturelle, compatible à la fusion, entre les éléments de Brauer (x, b_x) et $(\tilde{x}, b_{\tilde{x}})$ des blocs B et $\tilde{B}$ (il se trouve que pour tout $x \in G_l$, il existe au plus un seul bloc b_x de $C_G(x)$ en correspondance de Brauer avec B). Soit *tilde* cette bijection. Elle établit également une bijection entre l'ensemble des séries de Lusztig de caractères de G rencontrant B et l'ensemble des séries de Lusztig de $\tilde{G}$ rencontrant $\tilde{B}$. L'isométrie Δ: $\mathfrak{R}(B) \to \mathfrak{R}(\tilde{B})$ peut donc être définie série par

série, ce qui permet de se ramener aux caractères unipotents du bloc B, paramétrés par des partitions de e-cœur donné. Une isométrie entre groupes de Grothendieck n'étant qu'une bijection entre irréductibles assortie d'une application *signe*, le théorème 1.4 permet de définir Δ, ainsi que, pour tout (x, b_x), Δ_x: $\mathfrak{R}(b_x) \to \mathfrak{R}(b_{\tilde{x}})$. Les applications de Lusztig restreintes aux unipotents sont calculables, et ce calcul fait intervenir les lois de comodules définies en 1.3 (Prop. 2.2). On en déduit, par 1.5 et la formule de Mackey sur les applications de Lusztig la commutativité du diagramme (Prop. 2.4)

$$\begin{array}{ccc} \mathfrak{R}(B) & \xrightarrow{{}^*R^G_{C_G(x)}} & \mathfrak{R}(b_x) \\ \downarrow{\scriptstyle \Delta} & & \downarrow{\scriptstyle \Delta_x} \\ \mathfrak{R}(\tilde{B}) & \xrightarrow{{}^*R^{\tilde{G}}_{C_{\tilde{G}}(\tilde{x})}} & \mathfrak{R}(b_{\tilde{x}}) \end{array}$$

où ${}^*R^Y_X$ est l'adjointe de R^Y_X. Si les applications ${}^*R^G_{C_G(x)}$ sont connues, les applications de décomposition généralisées d^x_G le sont aussi; on obtient alors des relations de commutativité

$$\begin{array}{ccc} \mathfrak{L} \otimes \mathfrak{R}(B) & \xrightarrow{d^x_G} & \mathfrak{L} \otimes \mathfrak{R}(b_x) \\ \downarrow{\scriptstyle 1 \otimes \Delta} & & \downarrow{\scriptstyle 1 \otimes \Delta_x} \\ \mathfrak{L} \otimes \mathfrak{R}(\tilde{B}) & \xrightarrow{d^{\tilde{x}}_{\tilde{G}}} & \mathfrak{L} \otimes \mathfrak{R}(b_{\tilde{x}}) \end{array}$$

qui impliquent que les isométries Δ_x ainsi que $\Delta = \Delta_1$ transforment les combinaisons sur $\mathbb{Z}$ de caractères projectifs de b_x en combinaisons de projectifs de $b_{\tilde{x}}$ (Prop. 3.1). Que l'isométrie Δ soit parfaite s'en déduit par une démonstration tout à fait générale et très semblable à celle de [5] (voir M. Broué: «Isométries parfaites, catégories dérivées», actes du colloque de Luminy, mai 1988, à paraître in Astérisque).

1 Notations et faits

1.1 Quelques notations

Soit l un nombre premier.

Soit $\mathfrak{O}$ un anneau de valuation discrète, complet, de caractéristique nulle, et dont le corps résiduel $\mathfrak{F}$ est de caractéristique l. Soit $\mathfrak{L}$ le corps des fractions de $\mathfrak{O}$; nous supposerons $\mathfrak{L}$ assez gros pour tous les groupes considérés.

Si X est un groupe fini, le groupe de Grothendieck de la catégorie des $\mathfrak{L}X$-modules de type fini est noté $\mathfrak{R}(X)$; $\mathfrak{R}(X)$ est considéré comme un groupe abélien de fonctions centrales sur X, admettant une base formée des caractères

des représentations absolument irréductibles de X. L'ensemble de ces caractères est noté $\mathrm{Irr}(X)$.

Le groupe de Grothendieck de la catégorie des $\mathfrak{O}X$-modules projectifs de type fini est noté $\mathfrak{P}_{\mathfrak{O}}(X)$ et $e_X: \mathfrak{P}_{\mathfrak{O}}(X) \to \mathfrak{R}(X)$ est l'injection canonique. L'image de e_X est le sous-groupe des éléments de $\mathfrak{R}(X)$ de valeur nulle sur tout élément de X dont l'ordre est divisible par l.

Enfin le groupe de Grothendieck de X sur $\mathfrak{F}$, identifié au groupe des fonctions centrales sur X engendré par les caractères de Brauer des représentations de X sur $\mathfrak{F}$ est noté $\mathfrak{R}_{\mathfrak{F}}(X)$, et $d_X: \mathfrak{R}(X) \to \mathfrak{R}_{\mathfrak{F}}(X)$ est l'application de décomposition. Les applications e_X et d_X sont adjointes l'une de l'autre.

Le groupe de Grothendieck sur $\mathfrak{L}$ d'un l-bloc B de X (ou bloc de $\mathfrak{O}X$) est noté $\mathfrak{R}(B)$, et on peut le considérer comme un sous-groupe de $\mathfrak{R}(X)$ admettant pour base l'ensemble des irréductibles de $\mathfrak{L}B$. On définit de même $\mathfrak{P}_{\mathfrak{O}}(B)$, $\mathfrak{R}_{\mathfrak{F}}(B)$, $e_B: \mathfrak{P}_{\mathfrak{O}}(B) \to \mathfrak{R}(B)$ et $d_B: \mathfrak{R}(B) \to \mathfrak{R}_{\mathfrak{F}}(B)$.

1.2 Groupes symétriques et la géométrie des crochets de partitions

L'ensemble des partitions (resp. des partitions d'un entier naturel k) est noté $\mathscr{P}$ (resp. $\mathscr{P}_k$). La «somme» de deux partitions α et β est notée $\alpha \cup \beta$ (si $\alpha \in \mathscr{P}_a$ et $\beta \in \mathscr{P}_b$, $\alpha \cup \beta \in \mathscr{P}_{a+b}$). Si $\beta \in \mathscr{P}$ et d est un entier naturel non nul, notons $d \cdot \beta$ la partition dont les composantes sont les produits par d des composantes de β.

Si X est un groupe symétrique $\mathfrak{S}_n$, la base naturelle de $\mathfrak{R}(X)$ est comme à l'accoutumée indexée sur $\mathscr{P}_n$. Si $\lambda \in \mathscr{P}_n$, soit $[\lambda]$ le caractère correspondant de $\mathfrak{S}_n$. Le module graduè

$$\mathfrak{R}\mathfrak{S} := \oplus_{n \in \mathbb{N}} \mathfrak{R}(\mathfrak{G}_n)$$

peut être muni d'une structure d'algèbre de Hopf sur $\mathbb{Z}$ via l'induction depuis les – et la restriction aux – sous-groupes de Young (voir [13]). Si λ et β sont deux partitions du même entier n, la valeur de $[\lambda]$ sur un élément de $\mathfrak{S}_n$ dont la décomposition en cycles est de type β sera simplement notée $[\lambda](\beta)$; soit alors z_β l'ordre du centralisateur dans $\mathfrak{S}_n$ d'un tel élément et ϕ_β la fonction caractéristique de sa classe de conjugaison, $[\lambda](\beta)$ est le produit scalaire $\langle [\lambda], z_\beta \phi_\beta \rangle_{\mathfrak{S}_n}$ dans $\mathfrak{R}(\mathfrak{S}_n)$.

Si $\lambda \in \mathscr{P}_n$, $a \in \mathbb{N}$ et $I \subset \mathbb{N}$, notons $\mathscr{C}_a(\lambda)$ (resp. $\mathscr{C}_I(\lambda)$) l'ensemble des crochets de λ de longueur a (resp. dont la longueur appartient à I). Si $\gamma \in \mathscr{C}_n(\lambda)$, l'ablation de γ à λ fournit une partition de $(n-a)$, qui est notée $\gamma * \lambda$. La signature de γ (voir [5], 2.2 et 2.3; en termes classiques, elle vaut $(-1)^{l(\gamma)}$, où $l(\gamma)$ est la «longueur de jambe» de γ) est notée $v(\gamma)$. Ainsi la formule de Murnaghan-Nakayama s'écrit, si $(x, y) \in \mathfrak{S}_m \times \mathfrak{S}_d$, $m+d=n$ et y est un cycle de longueur d,

$$\text{(M-N)} \qquad [\lambda](xy) = \sum_{\gamma \in \mathscr{C}_d(\lambda)} v(\gamma)[\gamma * \lambda](x).$$

Soit alors $e \in \mathbb{N}^\times$. Soit l'équivalence la plus fine dans l'ensemble des partitions telle que

$$\text{«si } \gamma \in \mathscr{C}_{e\mathbb{N}}(\lambda), \lambda \text{ et } \gamma * \lambda \text{ sont équivalents.»}$$

Soit $\mathscr{B}$ une classe selon cette équivalence. Nous dirons que $\mathscr{B}$ est un «e-bloc de partitions». Il existe alors un unique $\kappa\in\mathscr{B}$ tel que $\mathscr{C}_e(\kappa)=\emptyset$; la partition κ est dite «cœur de $\mathscr{B}$». Supposons que κ soit une partition de l'entier k. On a alors

$$\mathscr{B}\cap\mathscr{P}_k=\{\kappa\},\qquad \mathscr{B}\cap\mathscr{P}_m\neq\emptyset \text{ si et seulement si } m-k\in e\mathbb{N}.$$

Dans ces conditions posons $\mathscr{B}=\mathscr{B}(\kappa)$ et $\mathscr{B}\cap\mathscr{P}_{k+eN}=\mathscr{B}_N(\kappa)$ (Ces notations ne font pas allusion à l'entier e qui est fixé.) Un tel e-bloc de partitions définit un sous-module gradué de $\mathfrak{R}\mathfrak{S}$, à savoir

$$\mathfrak{R}(\kappa):=\oplus_{N\in\mathbb{N}}\mathfrak{R}_N(\kappa),\qquad \mathfrak{R}_N(\kappa):=\sum_{\lambda\in\mathscr{B}_N(\kappa)}\mathbb{Z}[\lambda].$$

(Si e est premier, $\mathfrak{R}_N(\kappa)$ est le groupe de Grothendieck d'un e-bloc – au sens des représentations e-modulaires – du groupe symétrique $\mathfrak{S}_{k+eN}$.)

1.3 Des structures de comodules sur $\mathfrak{R}\mathfrak{S}$

Considérons maintenant un entier non nul d. Le $\mathbb{Z}$-module sous-jacent à $\mathfrak{R}\mathfrak{S}$ peut être muni d'une structure de comodule à droite sur l'algèbre $\mathfrak{R}\mathfrak{S}$, par l'application $c^{(d)}$ définie comme suit

$$c^{(d)}\colon \mathfrak{R}\mathfrak{S}\to\mathfrak{R}\mathfrak{S}\otimes\mathfrak{R}\mathfrak{S},\qquad c^{(d)}=\oplus_{(n,b)}c^{(d)}_{n,b}\qquad (db\leqq n)$$
$$c^{(d)}_{n,b}\colon \mathfrak{R}(\mathfrak{S}_n)\to\mathfrak{R}(\mathfrak{S}_{n-db})\otimes\mathfrak{R}(\mathfrak{S}_b)$$
$$c^{(d)}_{n,b}(\chi)(\mu\times\beta)=\chi(\mu\cup d\cdot\beta)\qquad (\chi\in\mathfrak{R}(\mathfrak{S}_n),\ \mu\in\mathscr{P}_{n-db},\ \beta\in\mathscr{P}_b).$$

L'application $c^{(1)}$ coïncide avec le coproduit de la cogèbre $\mathfrak{R}\mathfrak{S}$. Si, sous les hypothèses précédentes, $\chi=[\lambda]$, où $\lambda\in\mathscr{P}_n$, la formule (M-N) itérée permet de calculer $[\lambda](\mu\cup d\cdot\beta)$ comme function de μ pour β fixé (dans $\mathfrak{R}(\mathfrak{S}_{n-db})$). Que la fonction centrale ainsi définie soit un caractère généralisé de $\mathfrak{S}_{n-db}\times\mathfrak{S}_b$ a été démontré par Farahat (voir [6], p. 312). Si d est un multiple de e et κ un e-cœur, comme la formule (M-N) ne fait intervenir que des crochets de longueur multiples de d, donc aussi multiples de e, $\mathfrak{R}(\kappa)$ est un sous-comodule de $\mathfrak{R}\mathfrak{S}$ pour la structure précédente, par les restrictions $c^{(d)}_{N,b}(\kappa)$ à $\mathfrak{R}_N(\kappa)$ des applications $c^{(d)}_{k+eN,b}$

$$c^{(d)}(\kappa)\colon \mathfrak{R}(\kappa)\to\mathfrak{R}(\kappa)\otimes\mathfrak{R}\mathfrak{S},\qquad c^{(d)}(\kappa)=\oplus_{(N,b)}c^{(d)}_{N,b}(\kappa),\qquad (db\leqq eN)$$
$$c^{(d)}_{N,b}(\kappa)\colon \mathfrak{R}_N(\kappa)\to\mathfrak{R}_{N-(d/e)b}(\kappa)\otimes\mathfrak{R}(\mathfrak{S}_b).$$

La composition des applications $c^{(d)}$ est commutative. En effet, si $\mu\in\mathscr{P}_m$, $\beta_j\in\mathscr{P}_{b_j}$, et $d_j\in\mathbb{N}$ $(j=1,2)$, alors $(\mu\cup d_1\cdot\beta_1)\cup d_2\cdot\beta_2=(\mu\cup d_2\cdot\beta_2)\cup d_1\cdot\beta_1$; il en résulte que si $n=m+d_1b_1+d_2b_2$,

$$c^{(d_1)}_{n-d_2b_2,b_1}\circ c^{(d_2)}_{n,b_2}=c^{(d_2)}_{n-d_1b_1,b_2}\circ c^{(d_1)}_{n,b_1}.$$

Plus généralement, pour tout $m\in\mathbb{N}$ et pour toute famille finie de couples d'entiers non nul $\{(b(f),\ d(f))\}_f$, et si $n=\sum_f d(f)\,b(f)+m$, est définie par composition une application

$$c^{\{d(f)\}_f}_{n,\{b(f)\}_f}\colon \mathfrak{R}(\mathfrak{S}_n)\to\mathfrak{R}(\mathfrak{S}_m)\otimes\mathfrak{R}(\times_f\mathfrak{S}_{b(f)}).$$

Par restriction on obtient, si $N=N_0+\sum_f (d(f)/e)\,b(f)$,

$$c^{\{d(f)\}_f}_{N,\{b(f)\}_f}(\kappa)\colon\ \mathfrak{R}_N(\kappa)\to\mathfrak{R}_{N_0}(\kappa)\otimes\mathfrak{R}(\times_f \mathfrak{S}_{b(f)}).$$

La théorie des e-quotients et l'étude des signatures permet de démontrer le théorème volage mais fonctoriel que voici ([5], 2.17).

1.4 Théorème. D'un cœur à l'autre. *Soient deux e-cœurs $\kappa\in\mathscr{P}_k$ et $\mu\in\mathscr{P}_m$ et $\mathscr{B}(\kappa)$, $\mathscr{B}(\mu)$ les e-blocs de partitions qu'ils définissent. Il existe des applications*

$$\Psi\colon\ \mathscr{B}(\kappa)\to\mathscr{B}(\mu),\quad \textit{bijective},$$
$$\textit{pour tout } \lambda\in\mathscr{B}(\kappa),\ \Phi_\lambda\colon\ \mathscr{C}_{e\mathbf{N}}(\lambda)\to\mathscr{C}_{e\mathbf{N}}(\Psi(\lambda)),\quad \textit{bijectives},$$
$$\alpha\colon\mathscr{B}(\kappa)\to\{-1,1\},$$

telles que, si $\lambda\in\mathscr{B}_N(\kappa)$ et si $\gamma\in\mathscr{C}_{e\mathbf{N}}(\lambda)$,

a) $\Psi(\lambda)\in\mathscr{B}_N(\mu)$, Φ_λ *conserve la longueur des crochets, et* $\Phi_\lambda(\gamma)*\Psi(\lambda)=\Psi(\gamma*\lambda)$.
b) $v(\Phi_\lambda(\gamma))=\alpha(\lambda)\,\alpha(\gamma*\lambda)\,v(\gamma)$.

(De ce théorème, utilisé avec $e=l$, et de la formule de Murnaghan-Nakayama, on déduit dans [5] l'existence d'isométries parfaites entre les groupes de Grothendieck de blocs de groupes symétriques dont les poids N sont égaux.)

1.5 Corollaire. *(Hypothèses et notations du théorème) L'application*

$$I\colon\ \mathfrak{R}(\kappa)\to\mathfrak{R}(\mu)$$

définie par

$$I([\lambda])=\alpha(\lambda)[\Psi(\lambda)],\qquad(\lambda\in\mathscr{B}(\kappa))$$

est une isométrie de $\mathbb{Z}$-modules gradués respectant pour tout d multiple de e les structures de comodules sur $\mathfrak{R}\mathfrak{S}$ définies par $c^{(d)}(\kappa)$ et $c^{(d)}(\mu)$ respectivement.

Par composition on obtiendra pour toute famille finie $\{b(f),\,d(f)\}_f$ comme plus haut un diagramme commutatif

$$\begin{array}{ccc}
\mathfrak{R}_N(\kappa) & \xrightarrow{c(\kappa)} & \mathfrak{R}_{N_0}(\kappa)\otimes\mathfrak{R}(\times_f \mathfrak{S}_{b(f)}) \\
\Big\downarrow{\scriptstyle I_N} & & \Big\downarrow{\scriptstyle I_{N_0}\otimes 1} \\
\mathfrak{R}_N(\mu) & \xrightarrow{c(\mu)} & \mathfrak{R}_{N_0}(\mu)\otimes\mathfrak{R}(\times_f \mathfrak{S}_{b(f)})
\end{array}$$

($N=N_0+\sum_f (d(f)/e)\,b(f)$ et $c^{\{d(f)\}_f}_{N,\{b(f)\}_f}(\kappa)$ est abrégé en $c(\kappa)$.)

1.6 Les groupes de type GL/εq

Pour les propriétés rappelées ci-dessous ainsi que dans 1.7 à 1.12, nos références sont [12, 10, 4] en ce qui concerne les irréductibles et la structure de $\mathfrak{RG}$ (défini en 1.12) et [7, 2] à propos des blocs.

On considère les groupes linéaires en toute dimension $n \in \mathbb{N}$ sur un corps fini $\mathbb{F}_q$, de cardinal q premier à l, et les groupes unitaires relativement à l'extension quadratique $\mathbb{F}_{q^2}/\mathbb{F}_q$; les premiers sont notés GL(n, q), les seconds GL$(n, -q)$ (conventionnellement, les groupes en dimension nulle sont d'ordre 1). Les entiers l et $\varepsilon q \in \{q, -q\}$ étant fixés, soit e l'ordre de εq modulo l.

Un tel groupe $G = \mathrm{GL}(n, \varepsilon q)$ est considéré comme groupe des points rationnels sur $\mathbb{F}_q$ d'un groupe algébrique $\mathbf{G}$, groupe linéaire en même dimension sur la clôture algébrique $\mathbf{F}$ de $\mathbb{F}_q$. La convention (lettre grasse $\mathbf{X} \mapsto$ lettre maigre X) sera systématiquement utilisée – sauf ambiguïté sur εq – pour passer d'un groupe algébrique sur $\mathbf{F}$ défini sur $\mathbb{F}_q$ au groupe des points rationnels sur $\mathbb{F}_q$.

En particulier on aura à considérer les centralisateurs des éléments semi-simples de G. A cet effet soit, si r ou $-r$ est une puissance de q, $\mathscr{F}(r)$ – noté $\mathscr{F}$ si $r = \varepsilon q$ – l'ensemble des orbites selon l'application $(x \mapsto x^r)$ dans $\mathbf{F}^\times$; tout $f \in \mathscr{F}$ détermine un polynôme unitaire, élément de $\mathbf{F}[X]$, de degré $|f|$, dont l'ensemble des racines (simples) est précisément f. Une classe de semi-simples de G est caractérisée par son polynôme caractéristique dans la représentation standard, et ce polynôme décompose en produit d'éléments de $\mathscr{F}$. Les classes semi-simples (s) de G sont paramétrées par les applications

$$m_s\colon \mathscr{F} \to \mathbb{N} \quad \text{telle que} \sum_{f \in \mathscr{F}} m_s(f)|f| = n.$$

Si s est un tel semi-simple, son centralisateur $\mathbf{L}$ dans $\mathbf{G}$ est «régulier», c'est-à-dire sous-groupe de Levi rationnel d'un parabolique de $\mathbf{G}$, et à la décomposition du polynôme caractéristique correspond évidemment une décomposition de l'espace de représentation en somme directe stable par s, d'où un isomorphisme

$$\pi_s\colon L = C_G(s) \to \times_f \mathrm{GL}(m_s(f), (\varepsilon q)^{|f|}).$$

Cet isomorphisme n'est pas unique, la projection sur une composante non triviale en f est définie modulo un groupe cyclique Γ_f d'ordre $|f|$, qui opère comme un groupe de Galois si $|f|$ est impair ou si $\varepsilon = 1$ et est engendré par un automorphisme non intérieur de GL, produit de l'automorphisme de graphe par un automorphisme galoisien, sinon. Notons que deux isomorphismes π_s et π'_s sont transformés l'un de l'autre par un élément de $N_G(C_G(s))$. Ainsi toute opération de $\Gamma := \prod_f \Gamma_f$ sur les décompositions de $C_G(s)$ est produite par un élément de G.

L'opération de Γ sur les diverses paramétrisations des classes semi-simples de $C_G(s)$ se retrouve ainsi:

Pour tout entier non nul d, l'inclusion définit une application naturelle surjective

$$i_d\colon \mathscr{F}((\varepsilon q)^d) \to \mathscr{F}(\varepsilon q).$$

Dans ces hypothèses on a, pour tout $f \in \mathscr{F}(\varepsilon q)$ et $g \in i_d^{-1}(f)$, $|f| = |g|.\mathrm{pgcd}\{d, |f|\}$. D'autre part l'application $(x \mapsto x^{\varepsilon q})$ de $\mathbf{F}^\times$ en lui-même induit une bijection

de $\mathscr{F}((\varepsilon q)^d)$. Cette bijection engendre un groupe cyclique d'ordre d opérant sur $\mathscr{F}((\varepsilon q)^d)$, soit $\Gamma((\varepsilon q)^d/\varepsilon q)$ et $\mathscr{F}$ peut s'identifier au quotient de $\mathscr{F}((\varepsilon q)^d)$ par $\Gamma((\varepsilon q)^d/\varepsilon q)$. Le groupe Γ_f est naturellement isomorphe à $\Gamma((\varepsilon q)^{|f|}/\varepsilon q)$ et Γ opère ainsi sur $\prod_{f\in\mathscr{F}(\varepsilon q)} \mathscr{F}((\varepsilon q)^{|f|})$.

Considérons par exemple une classe de conjugaison (t) dans $C_G(s)$ d'éléments semi-simples. L'isomorphisme π_s étant fixé, une telle classe est paramétrée par une fonction de deux variables

$$m_{s,t}\colon \prod_{f\in\mathscr{F}(\varepsilon q)} \mathscr{F}((\varepsilon q)^{|f|}) \to \mathbb{N}$$

$$\text{telle que } \forall f\in\mathscr{F}(\varepsilon q),\ m_s(f)=\sum_{g\in\mathscr{F}((\varepsilon q)^{|f|})} m_{s,t}(f,g)|g|.$$

Par exemple, il existe $g\in\mathscr{F}(\varepsilon q)^{|f|})$ tel que $i_{|f|}(g)=f$ et $m_{s,s}(f,g)=m_s(f)$. La classe de conjugaison de G qui contient (t) est alors paramétrée par l'application m_t définie par

$$|f'|\,m_t(f')=\sum_{f\in\mathscr{F}(\varepsilon q)}\ \sum_{\{g|i_{|f|}(g)=f'\}} m_{s,t}(f,g)|f||g| \qquad (f'\in\mathscr{F}(\varepsilon q))$$

(l'application m_t ne dépend pas de π_s). Du fait de l'action des normalisateurs dans G de $C_G(s)$ et $C_G(t)$, $m_{s,t}$ ne détermine pas $m_{t,s}$. Cependant on a

$$m'(f,f'):=\sum_{\{g|i_{|f|}(g)=f'\}} m_{s,t}(f,g)=\sum_{\{g'|i_{|f'|}(g')=f\}} m_{t,s}(f',g'), \qquad (\forall(f,f')\in\mathscr{F}(\varepsilon q)^2).$$

Les classes de $C_G(t)$ paramétrées par les applications $m_{t,s}$ qui définissent par la formule précédente une même fonction m' fusionnent dans G. Si $s=t$, $m'(f,f)=m_s(f)$.

Inversement la classe semi-simple (t) de G rencontre $C_G(s)$ si, et seulement si, il existe une appplication $m'\colon \mathscr{F}(\varepsilon q)^2 \to \mathbb{N}$ telle que

$$m_s(f)=\sum_{f'} m'(f,f')|f'|(\operatorname{pgcd}\{|f|,|f'|\})^{-1},$$

$$m_t(f')=\sum_{f} m'(f,f')|f|(\operatorname{pgcd}\{|f'|,|f|\})^{-1},$$

f et f' parcourant $\mathscr{F}(\varepsilon q)$.

Soit sous ces hypothèses X le l-sous-groupe abélien de G engendré par s et t. Le groupe $C_G(X)=C_G(s)\cap C_G(t)$ peut être décrit de diverses façons comme sous-groupe de $C_G(x)$ ou de $C_G(t)$. Il existe des isomorphismes (non uniques)

$$\pi_{s,t}\colon C_G(X)\to \times_{f\in\mathscr{F}} \times_{g\in\mathscr{F}((\varepsilon q)^{|f|})} \mathrm{GL}(m_{s,t}(f,g),(\varepsilon q)^{|f||g|}),$$

$$\pi_{t,s}\colon C_G(X)\to \times_{f\in\mathscr{F}} \times_{g\in\mathscr{F}((\varepsilon q)^{|f|})} \mathrm{GL}(m_{t,s}(f,g),(\varepsilon q)^{|f||g|}).$$

Mais chaque composant de la forme $\mathrm{GL}(m,(\varepsilon q)^d)$ $(m\neq 0)$ de $C_G(X)$ correspond à une certaine représentation irréductible sur $\mathbb{F}_q$ (ou à deux représentations duales sur $\mathbb{F}_{q^2}$ si $\varepsilon=-1$) du groupe X, représentation présente avec la multiplicité m dans l'opération de X sur l'espace standard de G. La décomposition en produit direct de $C_G(X)$ est intrinsèque. Il existe donc pour tout couple (f,g) tel que $f\in\mathscr{F}$, $g\in\mathscr{F}((\varepsilon q)^{|f|})$ et $m_{s,t}(f,g)\neq 0$, un unique couple (f',g') $(f'\in\mathscr{F}$, $g'\in\mathscr{F}((\varepsilon q)^{|f'|})$

tel que $m_{t,s}(f', g') = m_{s,t}(f, g)$, et un composant direct de $C_G(X)$ qui est envoyé isomorphiquement par $\pi_{s,t}$ sur le terme $\mathrm{GL}(m_{s,t}(f, g), (\varepsilon q)^{|f||g|})$ et par $\pi_{t,s}$ sur le terme $\mathrm{GL}(m_{t,s}(f', g'), (\varepsilon q)^{|f'||g'|})$ (on a alors nécessairement $i_{|f'|}(g') = f$, $i_{|f|}(g) = f'$ et $|f'||g'| = |f||g| = \mathrm{ppcm}(|f|, |f'|)$).

La décomposition en l-blocs des algèbres des centralisateurs d'éléments semi-simples de G fera intervenir pour tout $f \in \mathscr{F}$ l'ordre modulo l de $(\varepsilon q)^{|f|}$, qui sera noté $e(f)$; on a donc $e = e(\{1\})$. La distinction nécessaire entre l- et l'-composante des éléments semi-simples conduit à noter $\mathscr{F}_l$ (resp. $\mathscr{F}_{l'}$) l'ensemble des $f \in \mathscr{F}$ dont les éléments sont d'ordre une puissance de l (resp. premier à l). Si s est un l-élément, il sera commode d'isoler la composante de $C_G(s)$ relative à $f = \{1\}$ (opérant sur l'espace des points fixes de s) des autres; si $f \neq \{1\}$ et $m_s(f) \neq 0$, la composante relative à f de $C_G(s)$ contient un élément central d'ordre l, donc $f \in (\mathscr{F}_l - \{1\})$ et e divise $|f|$ (ce qui signifie aussi $e(f) = 1$). Nous posons alors $C_G(s) = \mathrm{GL}(m_s(1), \varepsilon q) \times C_G^{(1)}(s)$ (où $m_s(1)$ est $m_s(\{1\})$).

La décomposition en l- et l'-composante d'une racine de l'unité induit une application $\mathscr{F} \to \mathscr{F}_l \times \mathscr{F}_{l'}$; notons-la $(f \mapsto (f_l, f_{l'}))$. Il est clair qu'alors $|f| = \mathrm{ppcm}\{|f_l|, |f_{l'}|\}$. Posons $d = |f_{l'}|$. Si $g \in \mathscr{F}((\varepsilon q)^d)$ et $i_d(g) = f_l$, on a aussi $|f_l| = |g| \ \mathrm{pgcd}\{d, |f_l|\}$. De ces deux égalités il résulte $|f| = |g||f_{l'}|$. Or, si $g \neq \{1\}$, $e(f)$ divise $|g|$ (de même que e divise $|h|$ si $h \in \mathscr{F}_l$ et $h \neq \{1\}$). On voit ainsi que, si $f \neq f_{l'}$, $e(f_{l'})|f_{l'}|$ divise $|f|$.

En fin de compte, en vue d'une démonstration par induction, il nous faudra considérer tout groupe de la forme $\times_j \mathrm{GL}(n_j, (\varepsilon q)^{d(j)})$, j parcourant un ensemble fini; un tel groupe sera dit «de type GL/εq». Les notions et faits présentés dans les paragraphes 1.7 à 1.12 concernent un groupe $G = \mathrm{GL}(n, \varepsilon q)$, mais seront utilisés si besoin est dans leur généralisation évidente aux groupes de type GL/εq.

1.7 Irréductibles de GL$(n, \varepsilon q)$: *les séries*

Selon les travaux de P. Deligne et G. Lusztig relatifs aux représentations des groupes finis du type de Lie (voir [9]), l'ensemble des irréductibles de G est partitionné en «séries». Cette décomposition fait appel à un groupe algébrique $\mathbf{G}^*$ en dualité avec $\mathbf{G}$ et muni d'un Frobenius F^* lié à F; on s'est donné au préable un isomorphisme ι: $\mathbf{F}^\times \to (\mathbb{Q}/\mathbb{Z})_{p'}$ et un plongement de $(\mathbb{Q}/\mathbb{Z})_{p'}$ dans $\bar{\mathbb{Q}}_l^\times$ (ou dans $\mathbb{C}^\times$).

Soit $\mathbf{T}$ un tore maximal rationnel de $\mathbf{G}$. Deligne et Lusztig ont défini un morphisme de groupes

$$R_T^G\colon\ \mathfrak{R}(T) \to \mathfrak{R}(G)$$

(une notation plus correcte ferait référence aux groupes algébriques $\mathbf{G}$ et $\mathbf{T}$). Soit $\theta \in \mathrm{Irr}(T)$. Au couple $(\mathbf{T}, \theta)$ – ou plus exactement à l'orbite de ce couple selon G – correspond par dualité une orbite selon G^* de couples $(\mathbf{T}^*, t^*)$, où $\mathbf{T}^*$ est un tore maximal rationnel de $\mathbf{G}^*$ et $t^* \in T^*$. Soit $\mathscr{E}(G, (t^*))$ l'ensemble des irréductibles de G qui interviennent dans l'un au moins des $R_T^G(\theta)$ ($\mathbf{T}^*$ parcourant l'ensemble des tores de $\mathbf{G}^*$ contenant t^*). On a alors une partition

$$\mathrm{Irr}(G) = \bigcup_{(t^*)} \mathscr{E}(G, (t^*))$$

(t^*) parcourant l'ensemble des classes de conjugaison semi-simples de G^*.

1.8 *Autodualité*

Il se trouve que **G**, étant de type **A**, est isomorphe à **G***, si bien que la décomposition précédente de $\mathrm{Irr}(G)$ peut aussi être indexée sur l'ensemble des classes de conjugaison semi-simples de G. Précisons ce fait en suivant [4].

En passant des groupes de racines de l'unité à leurs groupes de caractères, on peut définir un ensemble $\mathscr{F}^*$ dual de $\mathscr{F}$, et l'isomorphisme ι détermine une bijection $j\colon \mathscr{F} \to \mathscr{F}^*$. Ainsi $\mathscr{F}$ est en dualité avec lui-même par l'application $\langle , \rangle\colon \mathscr{F} \times \mathscr{F} \to (\mathbb{Q}/\mathbb{Z})_{l'}$ définie par si $(f, g) \in \mathscr{F}^2$, soit $m = \mathrm{ppcm}\{|f|, |g|\}$ $\langle f, g\rangle = (q^m - 1)^2\, \iota(x)\, \iota(y)/(q-1)$ $(x \in f,\ y \in g)$.

Si **T** est un tore maximal rationnel de **G**, $(\mathbf{T}, F)$ est isomorphe à un produit direct $\prod_{i \in I} (\mathbf{T}_{\lambda_i}, F_{\lambda_i})$ où $(\mathbf{T}_m, F_m)$ est la descente des scalaires de $\mathbb{F}_{q^m}$ à $\mathbb{F}_q$ du groupe multiplicatif. Dans ce produit $\{\lambda_i\}_{i \in I}$ est une partition de n qui caractérise l'orbite selon G de **T**. Si $s \in T$, puisque **T** centralise s, la composante en $i \in I$ de T est contenue dans l'une des composantes du centralisateur de s dans G, ce qui définit une application

$$\phi_{T,s}\colon I \to \mathscr{F}, \qquad |\phi_{T,s}(i)| \text{ divise } \lambda_i \quad \text{et} \quad m_s(f)|f| = \sum_{\{i \mid \phi_{T,s}(i) = f\}} \lambda_i \qquad (\forall f \in \mathscr{F}).$$

On a pour tout $f \in \mathscr{F}$ une partition $\{\lambda_i/|f|\}_{\phi(i) = f}$ de $m_s(f)$ et l'application de $\mathscr{F}$ dans $\mathscr{P}$ ainsi définie caractérise l'orbite selon G du couple $(\mathbf{T}, s)$.

Considérons maintenant deux couples $(\mathbf{T}^*, t^*)$ et $(\mathbf{T}, \theta)$ en dualité ($t^* \in T^*$ et $\theta \in \mathrm{Irr}(T)$). Les partitions qui caractérisent les orbites de **T** selon G et de $\mathbf{T}^*$ selon G^* comme les ensembles I qui indexent les décompositions de $(\mathbf{T}, F)$ et $(\mathbf{T}^*, F^*)$ peuvent être identifiés. Il existe $t \in T$ tel que $m_t = m_{t^*} \circ j$ et $j \circ \phi_{T,t} = \phi_{T^*,t^*}$. Si $s \in T$, $\theta(s)$ n'est autre que l'image dans $\mathbb{Q}_l^\times$ (ou dans $\mathbb{C}^\times$) de

$$\sum_{i \in I} \frac{\lambda_i}{\mathrm{ppcm}\{|\phi_{T,s}(i)|, |\phi_{T,t}(i)|\}} \langle \phi_{T,s}(i), \phi_{T,t}(i)\rangle.$$

Soit $m'\colon \mathscr{F}^2 \to \mathbb{N}$ la fonction définie en 1.6 relativement à t et s. Le groupe $C_G(x) \cap C_G(t)$, isomorphe à $\times_{f,g} \mathrm{GL}(m_{s,t}(f, g), (\varepsilon q)^{|f|\,|g|})$ contient un tore du type $\{(|f|\,|f'|/\mathrm{pcgd}\{|f|, |f'|\})^{m'(f,f')}\}_{(f,f')}$ et $\theta(s)$ est l'image dans $\mathbb{Q}^\times$ de

$$\sum_{(f,f') \in \mathscr{F}^2} m'(f,f') \langle f, f'\rangle.$$

Un isomorphisme de **G** sur son dual ne détermine pas une unique dualité entre un tore maximal **T** de **G** et son image dans le dual de **G** mais l'orbite de θ sous Γ est déterminée.

La série $\mathscr{E}(G, (t^*))$, définie par m_{t^*} comme par $m_t = m_{t^*} \circ j$, sera aussi notée $\mathscr{E}(G, (t))$. La partition de $\mathrm{Irr}(G)$ en série induit une décomposition du groupe de Grothendieck en somme directe

$$\mathfrak{R}(G) = \oplus_{(t)} \mathfrak{R}^{(t)}(G).$$

1.9 Classification de Lusztig

Soient θ, t et t^* comme en 1.8 et $\mathbf{L}^*$ et $\mathbf{L}$ les centralisateurs de t^* et t dans $\mathbf{G}^*$ et $\mathbf{G}$ respectivement. A t^* central et rationnel dans $\mathbf{L}^*$ correspond par dualité un caractère de degré 1 de L, soit $\overline{\theta}$, dont la restriction à Test θ.

G. Lusztig a défini un morphisme de groupes abéliens généralisant R_T^G, le tore maximal $\mathbf{T}$ étant remplacé par un sous-groupe régulier, comme l'est $\mathbf{L}$; alors

$$R_L^G\colon\ \mathfrak{R}(L)\to\mathfrak{R}(G)$$

est telle qu'il existe $\eta\in\{-1,1\}$ *tel que l'application*

$$(\zeta\mapsto\eta R_L^G(\overline{\theta}\zeta))\qquad(\zeta\in\mathscr{E}(L,(1))$$

soit bijective sur $\mathscr{E}(G,(t))$.

Une dualité entre $\mathbf{G}$ et $\mathbf{G}^*$ et un isomorphisme entre ces deux groupes ne déterminent pas une unique dualité entre le centralisateur dans $\mathbf{G}$ d'un élément semi-simple t de G et le centralisateur dans le dual de $\mathbf{G}$ de l'image de t. Dans la formule ci-dessus, t ne détermine pas θ, mais $R_L^G(\overline{\theta}\zeta)$ est bien défini.

L'application adjointe de R_L^G sera notée ${}^*R_L^G$.

Plus généralement, si $\mathbf{M}$ est un sous-groupe régulier de $\mathbf{G}$ contenant $\mathbf{L}$, l'application R_M^G induit à un signe prés une bijection de $\mathscr{E}(M,(t))$ sur $\mathscr{E}(G,(t))$.

1.10 Les caractères unipotents

A toute partition λ de n correpond une classe de tores $(\mathbf{T})$ selon G, d'où un élément $R_\lambda^G:=R_T^G(1)$ de $\mathfrak{R}(G)$. Un caractère irréductible de G est dit unipotent si, et seulement si, il intervient dans l'un des R_λ^G $(\lambda\in\mathscr{P}_n)$.

Soient $\mu\in\mathscr{P}_n$ *et*

$$v([\mu]):=\sum_{\lambda\in\mathscr{P}_n}z_\lambda^{-1}[\mu](\lambda)\,R_\lambda^G.$$

Alors $v([\mu])\in\mathfrak{R}(G)$, $v([\mu])$ *est de norme* 1, *et quand* μ *parcourt* $\mathscr{P}_n$ *on obtient ainsi au signe près tous les irréductibles unipotents de* G.

1.11 Les l-blocs

La répartition des irréductibles de G en blocs pour le nombre premier l est décrite par les trois (et demie) assertions suivantes

1.11.1 Soient $\chi\in\mathscr{E}(G,(t))$, $\xi\in\mathscr{E}(G,(s))$. *Si* χ *et* ξ *sont des irréductibles du même l-bloc* $(t_{l'})=(s_{l'})$.

Il s'agit là d'un théorème général sur les groupes réductifs finis, dû à Michel Broué et Jean Michel (voir [1], 2.2), que nous énonçons en tenant compte de l'autodualité. L'ensemble des l-blocs de G décompose donc en «séries» indexées sur l'ensemble des classes de conjugaison d'éléments semi-simples et d'ordre premier à l de G.

1.11.2 Soit $t \in G_{l'}$ et soit **L** *le centralisateur de t dans* **G**. *L'application de Lusztig R_L^G induit une isométrie parfaite entre la somme des blocs de L de la série (t), et la somme des blocs de G de la série (t).*

Là encore c'est un théorème général pour les groupes réductifs finis sous l'hypothèse «le centralisateur de t^* dans $\mathbf{G}^*$ est un sous-groupe régulier de $\mathbf{G}^*$», dû à M. Broué (voir [3], 2.2). L'isométrie induite par l'application de Lusztig y est d'ailleurs composée avec une équivalence de Morita évidente entre blocs. Reprenons la description de l'autodualité faite en 1.8. Soit $\bar{\theta}$ un caractère de degré 1 de L, associé à t par autodualité; t et $\bar{\theta}$ sont de même ordre, premier à l. La multiplication par $\bar{\theta}$ induit une équivalence de Morita entre la somme des blocs de L de la série (1) et la somme des blocs de L de la série (t). On obtient

1.11.2.bis L'application $(\zeta \mapsto R_L^G(\bar{\theta}\zeta))$ induit une isométrie parfaite entre la somme des blocs de L de la série (1) *et la somme des blocs de G de la série (t).*

Il reste à décrire l'ensemble des irréductibles d'un bloc de la série (1). Considérons donc $s \in G_l$ et $t \in G_l$ et leurs centralisateurs $L = C_G(s) \cong \mathrm{GL}(m_s(1), \varepsilon q) \times C_G^{(1)}(s)$ et $M = C_G(t) \cong \mathrm{GL}(m_t(1), \varepsilon q) \times C_G^{(1)}(t)$, décomposés comme convenu en 1.6. Soient $\bar{\theta}$ et $\bar{\sigma}$ des caractères de degré 1 de M et L déterminés par m_t et m_s respectivement. Soient enfin $\lambda \in \mathscr{P}_{m_s(1)}$, $\mu \in \mathscr{P}_{m_t(1)}$, ζ (resp. ξ) un irréductible unipotent de $C_G^{(1)}(s)$ (resp. $C_G^{(1)}(t)$). Alors $R_L^G(\bar{\sigma}(v([\lambda]) \otimes \zeta))$ et $R_M^G(\bar{\theta}(v([\mu]) \otimes \xi))$ sont à un signe près des irréductibles de G.

1.11.3 Sous les hypothèses qui précèdent, $R_M^G(\bar{\theta}(v(\mu) \otimes \xi))$ et $R_L^G(\bar{\sigma}(v(\lambda) \otimes \zeta))$ sont «dans le même l-bloc» si et seulement si les partitions λ et μ sont dans le même e-bloc de partitions.

Ainsi un e-cœur $\kappa \in \mathscr{P}_k$ et un entier naturel N déterminent un l-bloc $B(\kappa, N)$ de $\mathrm{GL}(k+eN, \varepsilon q)$ de la série (1) (i.e. contenant des caractères unipotents), et tous ces blocs s'obtiennent ainsi. Le **poids** du bloc est N. Notons que si $e=1$ (c'est-à-dire si l divise $(\varepsilon q - 1)$), κ est vide, $k=0$ et $B(\emptyset, N)$ est le bloc principal de $\mathrm{GL}(\varepsilon q, N)$. Si $N=0$, B est de défaut nul, simple et projectif.

Plus généralement, un l-bloc d'un groupe linéaire en εq est déterminé par la donnée d'une famille de couples $(\kappa(f), N(f))$ où $\kappa(f) \in \mathscr{P}_{k(f)}$ est un $e(f)$-cœur et $N(f) \in \mathbb{N}$, f parcourant un sous-ensemble fini $\mathscr{I}$ de $\mathscr{F}_{l'}$. Soient alors $m(f) = k(f) + e(f)\,N(f)$ et supposons $n = \sum_{f \in \mathscr{I}} |f|\, m(f)$. L'application m prolongée par 0 hors de $\mathscr{I}$ détermine une classe (t) de l'-éléments semi-simples de $G = \mathrm{GL}(n, \varepsilon q)$. Pour chaque $f \in \mathscr{I}$, le couple $(\kappa(f), N(f))$ détermine un l-bloc $b(\kappa(f), N(f))$ de la série (1) de la composante en f de $C_G(t)$ et le produit $b := \otimes_f b(\kappa(f), N(f))$ est un l-bloc de $C_G(t)$ de la série (1) lequel détermine à son tour un l-bloc $B((\kappa(f), N(f))_f)$ de G de la série (t) selon 1.11.2.bis. En outre l'application de Lusztig $R_{C_G(t)}^G$ composée avec la multiplication par $\bar{\theta}$ induit une isométrie parfaite de $\mathfrak{R}(b)$ sur $\mathfrak{R}(B)$. Tous les l-blocs de G s'obtiennent ainsi.

Le **poids** du bloc est l'application $N: \mathscr{F} \to \mathbb{N}$ (on peut prolonger κ par $\emptyset$ et N et k par 0 hors de $\mathscr{I}$).

1.12 Une structure d'algèbre de Hopf

F. Digne et J. Michel [4] donnent des faits présentés en 1.7–1.10 une description élégante. Considérons le $\mathbb{Z}$-module métrique gradué

$$\mathbb{R}\mathbb{G} = \mathfrak{R}\mathbb{G}(\varepsilon q) := \oplus_{n \in \mathbf{N}} \mathbb{R}(\mathrm{GL}(n, \varepsilon q)).$$

On munit $\mathfrak{RG}$ d'une structure d'algèbre de Hopf autoadjointe sur $\mathbb{Z}$, le produit étant défini composante par composante par les applications $R_L^G: \mathfrak{R}(L) \to \mathfrak{R}(G)$, où L est sous-groupe régulier de type diagonal (de la forme $=\mathrm{GL}(r, \varepsilon q) \times \mathrm{GL}(s, \varepsilon q)$) de $G=\mathrm{GL}(r+s, \varepsilon q)$.

Soit $f \in \mathscr{F}$, supposons $n=|f|$ et soit $(\mathbf{T}, \theta)$ ($\mathbf{T}$ tore maximal de $\mathbf{G}$, $\theta \in \mathrm{Irr}(T)$) tel que l'image de ϕ_{T^*,θ^*} se réduise à $\{j(f)\}$ (voir 1.8, alors T est un tore de Coxeter, cyclique). La sous-algèbre $\mathfrak{RG}^{(f)}$ de $\mathfrak{RG}$ engendrée par $R_T^G(\theta)$ est isomorphe à $\mathfrak{RS}$, les degrés étant multipliés par $|f|$. Cet isomorphisme n'est pas unique, car la multiplication par la signature est un automorphisme non trivial de $\mathfrak{RS}$, mais il peut être fixé conventionnellement.

Par exemple, pour $f=\{1\}$, le sous-module de base l'ensemble des irréductibles unipotents est une sous-algèbre graduée de $\mathfrak{RG}$, notée ici $\mathfrak{RG}^{(1)}$, isomorphe à $\mathfrak{RS}$, par un isomorphisme qui envoie $[\lambda]$ sur $v([\lambda])$ pour tout $\lambda \in \mathscr{P}$. Cet isomorphisme comme ses restrictions à des sous-modules de $\mathfrak{RS}$ sera évidemment noté v. Il est caractérisé ainsi (notations de 1.2 et 1.10):

$$\langle v(\chi), R_\beta^G \rangle_G = \chi(\beta) \quad \chi \in \mathfrak{R}(\mathfrak{S}_n), \beta \in \mathscr{P}_n).$$

Enfin $\mathfrak{RG}$ est isomorphe au produit tensoriel sur $f \in \mathscr{F}$ des sous-algèbres $\mathfrak{RG}^{(f)}$.

Un e-cœur $\kappa \in \mathscr{P}_k$ détermine un sous-module de $\mathfrak{RG}$, somme directe sur $N \in \mathbb{N}$ des groupes de Grothendieck des blocs $B(\kappa, N)$ (voir 1.11). Soit $\mathfrak{RG}^{(1)}(\kappa)$ l'image par v dans $\mathfrak{RG}^{(1)}$ de $\mathfrak{R}(\kappa)$ (ce dernier défini en 1.2). On a clairement une décomposition

$$\oplus_{N \in \mathbb{N}} \mathfrak{R}(B(\kappa, N)) = \mathfrak{RG}^{(1)}(\kappa) \otimes (\otimes_f \mathfrak{RG}^{(f)})$$

f parcourant $(\mathscr{F}_l - \{1\})$. La composante en degré $n=k+eN$ est

$$\text{(D)} \qquad \mathfrak{R}(B(\kappa, N)) = \oplus_m (\mathfrak{RG}^{(1)}_{(m(1)-k)/e}(\kappa) \otimes (\otimes_f \mathfrak{RG}^{(f)}_{m(f)}))$$

où f parcourt $\mathscr{F}_l - \{1\}$, l'indice $m(f)$ indique un degré relatif (à multiplier par $|f|$ pour obtenir le degré dans $\mathfrak{RG}$) et m parcourt l'ensemble des applications de $\mathscr{F}_l$ dans $\mathbb{N}$ telles que $\sum_{f \in \mathscr{F}_l} m(f)|f| = n$ et $(m(1) \geqq k)$. Chacune de ces applications m détermine une classe de l-éléments semi-simples (t) de G tels que $m_t = m$, et le facteur direct d'indice m dans la décomposition précédente de $\mathfrak{R}(B(\kappa, N))$ admet une base formée d'éléments de la série $\mathscr{E}(G, (t))$: c'est $\mathfrak{R}^{(t)}(B(\kappa, N)) := \mathfrak{R}(B(\kappa, N)) \cap \mathfrak{R}^{(t)}(G)$. Pour $t=1$, l'application v se restreint en

$$v: \ \mathfrak{R}_N(\kappa) \to \mathfrak{R}^{(1)}(B(\kappa, N))$$

et est fixe par tout automorphisme de G. Si $e=1$, $\kappa=\emptyset$ et $\mathfrak{R}_N(\kappa)=\mathfrak{R}(\mathfrak{S}_N)$. Si $N=0$, on a $\mathfrak{R}_0(\kappa)=\mathbb{Z}[\kappa]$ et $\mathfrak{R}(\mathfrak{B}(\kappa, 0))=\mathfrak{R}^{(1)}(B(\kappa, 0))=\mathbb{Z}v([\kappa])$.

On a des descriptions analogues pour les blocs d'une série quelconque. Soit, comme en 1.11, $\mathscr{I}$ un sous-ensemble fini de $\mathscr{F}_{l'}$, et $\{\kappa(f)\}_{f \in \mathscr{I}}$ une famille de $e(f)$-cœurs, où $\kappa(f) \in \mathscr{P}_{k(f)}$. Cette fois on a

$$\oplus_{(N(f)) \in \mathbb{N}^{\mathscr{I}}} \mathfrak{R}(B((\kappa(f), N(f))_f)) = \otimes_{f \in \mathscr{I}} (\oplus_{N \in \mathbb{N}} \mathfrak{R}(B(\kappa(f), N)))$$

$$\oplus_{N \in \mathbb{N}} \mathfrak{R}(B(\kappa(f), N)) = \mathfrak{RG}^{(f)}(\kappa(f)) \otimes (\otimes_g \mathfrak{RG}^{(g)})$$

Ici $\mathfrak{R}\mathfrak{G}^{(f)}(\kappa(f))$ est l'image de $\mathfrak{R}(\kappa(f))$ dans $\mathfrak{R}\mathfrak{G}^{(f)}$ et, dans le produit tensoriel de droite, g parcourt l'ensemble des éléments de $\mathscr{F}-\{f\}$ dont la l'-composante est f.

En degré $n=\sum_{f\in\mathscr{I}}(k(f)+e(f)N(f))|f|$, on obtient la décomposition

$$\mathfrak{R}(B((\kappa(f),N(f))_f))=\oplus_{(m_f)_f}(\otimes_{f\in\mathscr{I}}(\mathfrak{R}\mathfrak{G}^{(f)}_{(m_f(f)-k(f))/e(f)}(\kappa(f))\otimes(\otimes_g\mathfrak{R}\mathfrak{G}^{(g)}_{m_f(g)}))).$$

Dans cette dernière égalité f parcourt $\mathscr{I}$; pour chaque f, g parcourt le même ensemble que dans la formule précédente, $\{m_f\}_f$ parcourt l'ensemble des familles d'applications de $\{g\in\mathscr{F}\mid g_{l'}=f\}$ dans $\mathbb{N}$ telles que $(m_f(f)-k(f))$ soit multiple positif de $e(f)$ et $\sum_g m_f(g)|g|/(e(f)|f|)=(N(f)-m_f(f))/e(f)$ (on a vu en 1.6 que $e(f)|f|$ divise $|g|$). Là encore chaque famille $\{m_f\}_f$ d'applications détermine une classe (t) d'éléments semi-simples de G caractérisée par $m_t=\sum_{f\in\mathscr{I}}m_f$ (en prolongeant par 0 les applications m_f), et le facteur direct correspondant dans le groupe de Grothendieck du bloc n'est autre que $\mathfrak{R}(B((\kappa(f),N(f))_f))\cap\mathfrak{R}^{(t)}(G)$.

1.13 Théorème. *Soient G et $\tilde{G}$ deux groupes linéaires (ou deux groupes unitaires) sur le corps $\mathbb{F}_q$. Soient B et $\tilde{B}$ des l-blocs de G et $\tilde{G}$ respectivement. Si les poids de B et de $\tilde{B}$ sont égaux, il existe une isométrie parfaite de $\mathfrak{R}(B)$ sur $\mathfrak{R}(\tilde{B})$.*

Grâce au théorème de Broué déjà cité en 1.11.2.bis, il suffit de considérer des blocs de la série (1) pour des groupes de type $\mathrm{GL}/\varepsilon q$, donc finalement deux blocs $B(\kappa, N)$ et $B(\tilde{\kappa}, N)$ de la série (1) de poids N égaux. Les deux paragraphes suivants sont consacrés à la démonstration du théorème 1.13 dans cette dernière hypothèse. Notons que si $e=1$, le théorème est vide car $G=\tilde{G}$ et $B=\tilde{B}$. Nous supposerons donc $e\neq 1$, ce qui signifie que G n'admet aucun élément central d'ordre l. Enfin si $N=0$ les blocs B et $\tilde{B}$ sont tous deux de défauts nuls, et l'existence d'une isométrie parfaite est évidente. Le théorème 1.13 sera donc démontré par récurrence sur N pour tous les groupes de type $\mathrm{GL}/\varepsilon q$.

2 Isométries locales

Dans ce paragraphe comme dans le suivant, sont fixés εq, donc aussi e, κ et $\tilde{\kappa}$ qui sont deux e-cœurs, partitions de k et $\tilde{k}$ respectivement, $N\in\mathbb{N}$, $n=k+eN$, $\tilde{n}=\tilde{k}+eN$, $G=\mathrm{GL}(n,\varepsilon q)$, $\tilde{G}=\mathrm{GL}(\tilde{n},\varepsilon q)$ et on considère $B:=B(\kappa, N)$ et $\tilde{B}:=B(\tilde{\kappa}, N)$ qui sont des l-blocs de la série (1) de G et $\tilde{G}$ respectivement (définitions et notations dans 1.2, 1.6 et 1.11.3).

2.1 Les éléments de Brauer

Soit $x\in G_l$, dont la classe est caractérisée par $m_x\colon\mathscr{F}_l\to\mathbb{N}$. Si $m_x(1)\geqq k$, comme e divise $n-m_x(1)$, posons $N_x=(m_x(1)-k)/e$. Alors $C_G(x)$ admet le l-bloc

$$b_x:=B(\kappa,N_x)\otimes B_0(C_G^{(1)}(x)).$$

($B_0(X)$ désigne le bloc principal de X.) On sait (voir [2]; l'assertion se déduit d'ailleurs de la Prop. 2.2) que les éléments de Brauer de B sont les couples (x, b_x) comme ci-dessus. Des hypothèses faites sur B et $\tilde{B}$ et de la classification des classes semi-simples il résulte immédiatement le fait suivant

L'application $((x)\mapsto(\tilde{x}))$ *définie par*

$$m_{\tilde{x}}(1)-\tilde{k}=m_x(1)-k \quad et \quad m_{\tilde{x}}(f)=m_x(f) \quad si\ f\neq\{1\}$$

induit une bijection de l'ensemble des classes de l-éléments semi-simples de G admettant un élément de Brauer de B sur l'ensemble analogue relatif à $\tilde{G}$ *et* $\tilde{B}$.

2.2 Proposition. *Soient* $x\in G_l$, $m=m_x\colon \mathscr{F}\to\mathbb{N}$ *caractérisant la classe de x dans G et* $N_x=(m(1)-k)/e$. *Soient* $\mathfrak{R}^{(1)}(B)$ *et* $\mathfrak{R}^{(1)}(b_x)$ *les composantes unipotentes des groupes de Grothendieck des blocs B et* b_x. *Soit un isomorphisme* $\pi_x\colon C_G(x)\to \mathrm{GL}(m(1),\varepsilon q)\times(\times_f \mathrm{GL}(m(f),(\varepsilon q)^{|f|}))$ *(f parcourt* $\{f\in(\mathscr{F}-\{1\})\mid m(f)\neq 0\}$*) définissant (voir* 1.12*)*

$$v_x\colon\ \mathfrak{R}_{N_x}(\kappa)\otimes\mathfrak{R}(\times_f \mathfrak{S}_{m(f)})\to\mathfrak{R}^{(1)}(b_x).$$

Le diagramme suivant, où $c=c^{\{|f|\}_f}_{N,\{m(f)\}_f}(\kappa)$ *(notations de* 1.3*),*

$$\begin{array}{ccc}
\mathfrak{R}_N(\kappa) & \xrightarrow{\ v\ } & \mathfrak{R}^{(1)}(B) \\
\Big\downarrow{\scriptstyle c} & & \Big\downarrow{\scriptstyle {}^*R^G_{C_G(x)}} \\
\mathfrak{R}_{N_x}(\kappa)\otimes\mathfrak{R}(\times_f \mathfrak{S}_{m(f)}) & \xrightarrow{\ v_x\ } & \mathfrak{R}^{(1)}(b_x)
\end{array}$$

est commutatif.

Comme il l'est montré dans [7], 2, les applications ${}^*R^G_L$ restreintes aux unipotents se calculent à l'aide de la formule (M-N) rappelée en 1.2. Cette propriété étant incluse dans l'existence de l'isomorphisme $v\colon \mathfrak{RS}\to\mathfrak{RG}^{(1)}$ je la redémontre ici.

Posons $L=C_G(x)$. Les applications de Lusztig se restreignent aux groupes engendrés par les caractères unipotents. Si $\chi\in\mathfrak{R}(\mathfrak{S}_n)$, ${}^*R^G_L(v(\chi))$ est caractérisé par les produits scalaires $\langle {}^*R^G_L(v(\chi)),\ R^L_T(1)\rangle_L$, quand $\mathbf{T}$ parcourt l'ensemble des tores rationnels maximaux de $\mathbf{L}$. Soit $\mathscr{I}=\{f\in\mathscr{F}\mid m(f)\neq 0\}$. En considérant la composante en f de L comme un groupe de points rationnels d'un groupe linéaire en dimension $m(f)$ défini sur $\mathbb{F}_{q^{|f|}}$, on obtient bien l'isomorphisme v_x. Soit un tore T de L tel que $\pi_x(T)$ soit de type $\{\beta_f\}_f\in\{\mathscr{P}_{m(f)}\}_f$; T est de type $(\cup_f|f|\cdot\beta_f)\in\mathscr{P}_n$ dans G. D'une part, selon la définition des applications v et v_x on a

$$\langle v(\chi), R^G_T(1)\rangle_G=\chi(\cup_f|f|\cdot\beta_f)$$

et, si $\chi'=\otimes_f\chi'_f\in\otimes_f\mathfrak{R}(\mathfrak{S}_{m(f)})$,

$$\langle v_x(\chi'), R^L_T(1)\rangle_L=\prod_{f\in\mathscr{I}}\chi'_f(\beta_f)=\chi'(\cup_{f\in\mathscr{I}}\beta_f).$$

D'autre part, $R^G_T=R^G_L\circ R^L_T$ et donc $\langle v(\chi), R^G_T(1)\rangle_G=\langle {}^*R^G_L(v(\chi)), R^L_T(1)\rangle_L$.

Il en résulte que ${}^*R_L^G(v(\chi))=v'(\chi')$, où $\chi'(\cup_f \beta_f)=\chi(\cup_f |f|\cdot\beta_f)$ pour toute famille $\{\beta_f\}_f \in \{\mathscr{P}_{m(f)}\}_f$. Dans l'hypothèse où $x\in G_l$, si $f\neq\{1\}$, e divise $|f|$ et on a donc selon les définitions de 1.3,

$$ {}^*R_L^G \circ v = v' \circ c_{n,\{m(f)\}_f}^{\{|f|\}_f}(f\in(\mathscr{I}-\{1\})). $$

Que le diagramme de la proposition soit commutatif s'en déduit par restriction. □

La Prop. 2.2 montre que si $m_x(1)<k$, alors ${}^*R_{C_G(x)}^G(\mathfrak{R}^{(1)}(B))=\{0\}$. Nous l'utiliserons dans chaque centralisateur de l-élément de G, composante à composante, y compris lorsque $\kappa=\emptyset$ et $e=1$.

2.3 *Isométries: leur définition*

Soit I: $\mathfrak{R}(\kappa)\to\mathfrak{R}(\tilde{\kappa})$ une isométrie comme en 1.5 et notons I_m la restriction à $\mathfrak{R}_m(\kappa)$ de I pour tout $m\in\mathbb{N}$. Comme nous l'avons démontré dans [5], I n'est pas en général unique, on connaît une orbite selon $(\mathbb{Z}/e\mathbb{Z}, +)$ de telles isométries.

La bijection $(x)\mapsto(\tilde{x})$ définie en 2.1 s'interprète aussi comme une bijection entre les ensembles parcourus par les indices $m=m_x$ et $\tilde{m}=m_{\tilde{x}}$ dans les décompositions parallèles des groupes de Grothendieck des blocs (voir 1.12)

$$ \mathfrak{R}(B)=\oplus_{(t)}\mathfrak{R}^{(t)}(B), \qquad \mathfrak{R}(\tilde{B})=\oplus_{(\tilde{t})}(\mathfrak{R}^{(\tilde{t})}(\tilde{B}). $$

Définissons une isométrie Δ entre groupes de Grothendieck des blocs, terme à terme par transport à travers les applications v des isométries du paragraphe 1.5. On pose donc

$$ \Delta=\oplus_{(t)}\Delta^{(t)}\colon\ \mathfrak{R}(B)\to\mathfrak{R}(\tilde{B}), \qquad \Delta^{(t)}\colon\ \mathfrak{R}^{(t)}(B)\to\mathfrak{R}^{(\tilde{t})}(\tilde{B}) $$

où (t) parcourt le même ensemble que (x) dans 2.1.

Soit G_0 et $\tilde{G}_0$ des sous-groupes de G et $\tilde{G}$ «naturellement isomorphes» à $\mathrm{GL}(n-k, \varepsilon q)$ (i.e. issus de plongements compatibles avec les isomorphismes d'autodualité de $\mathrm{GL}(n-k, \mathbb{F})$ dans $\mathrm{GL}(n, \mathbb{F})$ et $\mathrm{GL}(\tilde{n}, \mathbb{F})$). Toute classe d'éléments semi-simples (t) de G telle que $m_t(1)\geqq k$ rencontre G_0 en une classe et une seule dont le paramétrage m ne diffère de m_t que par sa valeur en $\{1\}$. On a une assertion *tilde* analogue et l'isomorphisme de G_0 sur $\tilde{G}_0$ via $\mathrm{GL}(n-k, \varepsilon q)$, qui sera bien entendu noté $x\mapsto\tilde{x}$, induit en particulier la bijection décrite en 2.1: on a encore $m_{\tilde{x}}=\tilde{m}_x$.

Pour toute classe (t) dans G_l telle que $m_t(1)\geqq k$ soit un représentant $t\in G_0$, d'où $\tilde{t}\in\tilde{G}_l\cap\tilde{G}_0$. Soit $\theta\in\mathrm{Irr}(C_G(t))$ (resp. $\tilde{\theta}\in\mathrm{Irr}(C_{\tilde{G}}(\tilde{t}))$ «définis» par autodualité comme $\tilde{\theta}$ en 1.8 et 1.11. On a

$$ \mathfrak{R}^{(t)}(B)=R_{C_G(t)}^G(\theta\mathfrak{R}^{(1)}(b_t)), \qquad \mathfrak{R}^{(\tilde{t})}(\tilde{B})=R_{C_{\tilde{G}}(\tilde{t})}^{\tilde{G}}(\tilde{\theta}\mathfrak{R}^{(1)}(b_{\tilde{t}})). $$

Soient des isomorphismes π_t, $\pi_{\tilde{t}}$ comme en 1.6 induisant un isomorphisme de $C_G^{(1)}(t)$ sur $C_{\tilde{G}}^{(1)}(\tilde{t})$. On a $C_G^{(1)}(t)\subset G_0$ et $C_{\tilde{G}}^{(1)}(\tilde{t})\subset\tilde{G}_0$, et on peut supposer que $\pi_t(x)=\pi_{\tilde{t}}(\tilde{x})$ quel que soit $x\in C_{G_0}(t)$. Posons $m=m_t$, $\tilde{m}=m_{\tilde{t}}$, $N_t=(m(1)-k)/e$ $=(\tilde{m}(1)-\tilde{k})/e$. Soit v_t et $v_{\tilde{t}}$ des isomorphismes comme en 2.2.

Si $t=1$, on a $v_1=v$ et $\Delta^{(1)}$ est définie par la relation

$$\tilde{v}\circ I_N=\Delta^{(1)}\circ v.$$

Si $t\neq 1$, posons

$$\sigma^{(t)}\colon\ \mathfrak{R}_{N_t}(\kappa)\otimes\mathfrak{R}(\times_f \mathfrak{S}_{m(f)})\to\mathfrak{R}^{(t)}(B),\qquad \sigma^{(t)}(\chi)=R^G_{C_G(t)}(\theta v_t(\chi))$$
$$\tilde{\sigma}^{(\tilde{t})}\colon\ \mathfrak{R}_{N_t}(\tilde{\kappa})\otimes\mathfrak{R}(\times_f \mathfrak{S}_{m(f)})\to\mathfrak{R}^{(\tilde{t})}(\tilde{B}),\qquad \tilde{\sigma}^{(\tilde{t})}(\chi)=R^{\tilde{G}}_{C_G(\tilde{t})}(\tilde{\theta}v_{\tilde{t}}(\chi)).$$

L'isométrie $\Delta^{(t)}$ est définie par la commutativité du diagramme

$$\begin{array}{ccc}
\mathfrak{R}_{N_t}(\kappa)\otimes\mathfrak{R}(\times_f \mathfrak{S}_{m(f)}) & \xrightarrow{\sigma^{(t)}} & \mathfrak{R}^{(t)}(B)\\
\Big\downarrow{\scriptstyle I_{N_t}\otimes 1} & & \Big\downarrow{\scriptstyle \Delta^{(t)}}\\
\mathfrak{R}_{N_t}(\tilde{\kappa})\otimes\mathfrak{R}(\times_f \mathfrak{S}_{m(f)}) & \xrightarrow{\tilde{\sigma}^{(\tilde{t})}} & \mathfrak{R}^{(\tilde{t})}(\tilde{B}).
\end{array}$$

Remarquons que $\sigma^{(t)}$ est indépendant du choix de π_t, si bien que $\Delta^{(t)}$ est déterminé indépendamment du choix de t dans sa classe et du choix de π_t et $\pi_{\tilde{t}}$.

Notre démonstration de 1.13 étant inductive, nous supposerons que sont déterminées selon la même procédure et par un choix d'isomorphismes π_x compatibles avec l'application *tilde*: $C_{G_0}(x)\to C_{\tilde{G}_0}(\tilde{x})$ des isométries $\Delta_x\colon \mathfrak{R}(b_x)\to\mathfrak{R}(b_{\tilde{x}})$ pour tout $x\in G_l\cap G_0$ tel que $m_x(1)\geqq k$. Il est clair que $\pi_{\tilde{x}}$ et π_x définissent un isomorphisme de $C_G^{(1)}(x)$ sur $C_{\tilde{G}}^{(1)}(\tilde{x})$, et que cet isomorphisme induit la restriction de Δ_x depuis $\mathfrak{R}(B_0(C_G^{(1)}(x)))$ sur $\mathfrak{R}(B_0(C_{\tilde{G}}^{(1)}(\tilde{x})))$.

2.4 Proposition. *Soit $x\in G_l\cap G_0$. Le diagramme suivant*

$$\begin{array}{ccc}
\mathfrak{R}(B) & \xrightarrow{{}^*R^G_{C_G(x)}} & \mathfrak{R}(b_x)\\
\Big\downarrow{\scriptstyle \Delta} & & \Big\downarrow{\scriptstyle \Delta_x}\\
\mathfrak{R}(\tilde{B}) & \xrightarrow{{}^*R^{\tilde{G}}_{C_{\tilde{G}}(\tilde{x})}} & \mathfrak{R}(b_{\tilde{x}})
\end{array}$$

est commutatif.

La relation à démontrer équivaut par dualité à $\Delta\circ R^G_{C_G(x)}=R^{\tilde{G}}_{C_G(\tilde{x})}\circ\Delta_x$, et on voit qu'elle est vraie indépendamment du choix des isomorphismes π_x et $\pi_{\tilde{x}}$.

L'application Δ décomposant en $\oplus_{(t)}\Delta^{(t)}$, il suffit de démontrer la commutativité du diagramme

$$\begin{array}{ccc}
\mathfrak{R}^{(t)}(B) & \xrightarrow{{}^*R^G_{C_G(x)}} & \mathfrak{R}(b_x)\\
\Big\downarrow{\scriptstyle \Delta^{(t)}} & & \Big\downarrow{\scriptstyle \Delta_x}\\
\mathfrak{R}^{(\tilde{t})}(\tilde{B}) & \xrightarrow{{}^*R^{\tilde{G}}_{C_{\tilde{G}}(\tilde{x})}} & \mathfrak{R}(b_{\tilde{x}})
\end{array}$$

Si $t=1$, elle résulte de 1.5, 2.2 et de la définition des isométries. On a en effet

$$\begin{aligned} {}^*R^{\tilde{G}}_{C_{\tilde{G}}(\tilde{x})} \circ \Delta^{(1)} \circ v &= {}^*R^{\tilde{G}}_{C_{\tilde{G}}(\tilde{x})} \circ \tilde{v} \circ I_N \\ &= v_{\tilde{x}} \circ \tilde{c} \circ I_N \\ &= v_{\tilde{x}} \circ (I_{N_x} \otimes 1) \circ c \\ &= (\Delta_x)^{(1)} \circ v_x \circ c \\ &= (\Delta_x)^{(1)} \circ {}^*R^{G}_{C_G(x)} \circ v \end{aligned}$$

(on a utilisé la définition de $\Delta^{(1)}$, puis 2.2 avec $\tilde{c}=c^{\{|f|\}_{f\neq 1}}_{N,\{m_{\tilde{x}}(f)\}_{f\neq 1}}$, puis 1.5 avec $c=c^{\{|f|\}_{f\neq 1}}_{N,\{m_x(f)\}_{f\neq 1}}=\tilde{c}$, puis la définition de $(\Delta_x)^{(1)}$, puis 2.2).

Si $t\neq 1$, soit $m=m_t$ l'application de $\mathscr{F}_l$ dans $\mathbb{N}$ caractérisant la classe de t. Posons comme plus haut $N_t=(m(1)-k)/e=(\tilde{m}(1)-\tilde{k})/e$ et soient θ (resp. $\tilde{\theta}$) un caractère de degré 1 de $C_G(t)$ défini par m (resp. de $C_{\tilde{G}}(\tilde{t})$ défini par $\tilde{m}$) (voir 1.8 et 1.11). On a donc $\mathfrak{R}^{(t)}(B)=R^G_{C_G(t)}(\theta\mathfrak{R}^{(1)}(b_t))$. Soient $\chi\in\mathfrak{R}_N(\kappa)\otimes\mathfrak{R}(\times_{f\neq 1}\mathfrak{R}(\mathfrak{S}_{m(f)}))$ et $\zeta=v_t(\chi)\in\mathfrak{R}^{(1)}(b_t)$. Par définition de $\Delta^{(t)}$, on a $\Delta(R^G_{C_G(t)}(\theta\zeta))=R^{\tilde{G}}_{C_{\tilde{G}}(\tilde{t})}(\tilde{\theta}v_{\tilde{t}}(\tilde{\chi}))$ où $\tilde{\chi}=(I_{N_t}\otimes 1)(\chi)$.

La formule de Mackey pour les applications R^G_L s'écrit

$${}^*R^G_{C_G(x)} \circ R^G_{C_G(t)} = \sum_{(y)\,|\,y^g=t} R^{C_G(x)}_{C_G(x)\cap C_G(y)} \circ {}^*R^{C_G(y)}_{C_G(x)\cap C_G(y)} \circ \mathrm{Ad}(g)$$

où (y) parcourt l'ensemble de classes de conjugaison de $C_G(x)$ contenues dans (t), g est alors l'un quelconque des éléments de G tels que $y^g=t$ et $\mathrm{Ad}(g)$ désigne le transport par automorphism intérieur de G selon g. L'ensemble des fonctions $m_1\colon \times_f \mathscr{F}((\varepsilon q)^{|f|})\to\mathbb{N}$ paramétrant ces classes de $C_G(x)$ est déterminé par m et m_x, car x et y doivent se centraliser (voir 1.6). Considérons l'un des termes de la somme à droite du signe $=$, en fixant $m_{x,y}$, noté m_1.

Le paramétrage des unipotents est stable par automorphismes, et g induit un isomorphisme entre les groupes $C_G(t)$ et $C_G(y)$ de telle sorte que $\mathrm{Ad}(g)(\theta\zeta)=\theta' v_y(\chi)$, où θ' est défini par m. Soit θ_1 un caractère de degré 1 de $(C_G(y)\cap C_G(x))$ défini par $m_{y,x}$; on a ainsi

$$\begin{aligned} {}^*R^{C_G(y)}_{C_G(x)\cap C_G(y)}(\mathrm{Ad}(g)(\theta v_t(\chi))) &= \theta_1\, {}^*R^{C_G(y)}_{C_G(x)\cap C_G(y)}(v_y(\chi)), \\ {}^*R^{C_{\tilde{G}}(\tilde{y})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(\mathrm{Ad}(\tilde{g})(\tilde{\theta}v_{\tilde{t}}(\tilde{\chi}))) &= \tilde{\theta}_1\, {}^*R^{C_{\tilde{G}}(\tilde{y})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(v_{\tilde{y}}(\tilde{\chi})). \end{aligned}$$

Selon la proposition précédente appliquée à $C_G(y)$ et b_y et à $C_{\tilde{G}}(\tilde{y})$ et $b_{\tilde{y}}$, ces termes sont non nuls seulement si $m_1(1,1)\geqq k$ et $\tilde{m}_1(1,1)\geqq\tilde{k}$ (car $m_{x,y}(1,1)=m_{y,x}(1,1)$ est la dimension de l'espace des points fixes de $\langle x,y\rangle$ dans la représentation standard de G). On peut supposer que $\tilde{x}$, $\tilde{y}$, $\pi_{\tilde{x}}$, $\pi_{\tilde{y}}$, $\pi_{\tilde{x},\tilde{y}}$, $\pi_{\tilde{y},\tilde{x}}$ (notations de 1.6) sont définis depuis leurs analogues sans *tilde* via l'isomorphisme de G_0 sur $\tilde{G}_0$. On a donc $m_{\tilde{x},\tilde{y}}=\tilde{m}_1$ et, en posant $m_2=m_{y,x}$, $m_{\tilde{y},\tilde{x}}=\tilde{m}_2$; les divers paramétrages sur les groupes $C_G(x)$, $C_G(y)$, $C_G(x)\cap C_G(y)$ se transportent sur $C_{\tilde{G}}(\tilde{x})$, $C_{\tilde{G}}(\tilde{y})$ et $C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})$. En outre θ_1 est caractérisé par sa restriction au groupe $(G_0\cap C_G(x)\cap C_G(y))$.

Il est clair que sous les restrictions $m_{x,y}(1,1)\geqq k$ et $m_{\tilde{x},\tilde{y}}(1,1)\geqq\tilde{k}$, les classes (y) intervenant dans la formule de Mackey envisagée correspondent bijectivement par l'application *tilde* aux classes $(\tilde{y})$ intervenant dans la formule analogue avec $(\tilde{G},\tilde{x},\tilde{m},\tilde{y})$ à la place de (G,x,m,y) (voir 2.1 appliqué à $C_G(x)$ et $C_{\tilde{G}}(\tilde{x})$). En outre le caractère $R^{C_G(x)}_{C_G(x)\cap C_G(y)}(\theta_1 *R^{C_G(y)}_{C_G(x)\cap C_G(y)}(v_y(\chi)))$ appartient à la composante en (y) de $\mathfrak{R}(b_x)$, son image par Δ_x est donc donnée par $(\Delta_x)^{(y)}$. Il nous suffit de démontrer que, pour y et $\theta_1\in\mathrm{Irr}(C_G(x)\cap C_G(y))$ fixés, d'où $\tilde{\theta}_1\in\mathrm{Irr}(C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y}))$, on a

$$\begin{aligned}&((\Delta_x)^{(y)}\circ R^{C_G(x)}_{C_G(x)\cap C_G(y)})(\theta_1 *R^{C_G(y)}_{C_G(x)\cap C_G(y)}(v_y(\chi)))\\&\quad=R^{C_{\tilde{G}}(\tilde{x})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(\tilde{\theta}_1 *R^{C_{\tilde{G}}(\tilde{y})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(v_{\tilde{y}}(\tilde{\chi}))).\end{aligned}$$

Or la Prop. 2.2 est applicable aux groupes $C_G(y)$ et $C_{\tilde{G}}(\tilde{y})$. Soit $b_{y,x}$ le bloc de $C_G(y)\cap C_G(x)$ en correspondance de Brauer avec b_y, bloc de $C_G(y)$, donc avec B et avec b_x, bloc de $C_G(x)$. Posons $N_{y,x}=(m_2(1,1)-k)/e$. Soient de même $b_{\tilde{y},\tilde{x}}$, $N_{\tilde{y},\tilde{x}}$ et notons que $N_{\tilde{y},\tilde{x}}=N_{y,x}=N_{x,y}=N_{\tilde{x},\tilde{y}}$. L'application $(*R^{C_G(y)}_{C_G(x)\cap C_G(y)}\circ v_y)$ se calcule grâce à un isomorphisme

$$v_{y,x}\colon\ \mathfrak{R}_{N_{x,y}}(\kappa)\otimes\mathfrak{R}(\times_{(f,g)\neq(1,1)}\mathfrak{S}_{m_2(f,g)})\to\mathfrak{R}^{(1)}(b_{y,x})$$

(f parcourant $\mathscr{F}_l$ et g parcourant $\mathscr{F}((\varepsilon q)^{|f|})$). On a $*R^{C_G(y)}_{C_G(x)\cap C_G(y)}\circ v_y=v_{y,x}\circ c$, où $c=\times_f c_f$ et $c_f=c^{\{|g|\}_g}_{m(f),\{m_2(f,g)\}_g}$, (f,g) comme ci-dessus. Il résulte de 1.5 que

$$(I_{N_{y,x}}\otimes 1)(c(\chi))=c((I_{N_y}\otimes 1)(\chi))=c(\tilde{\chi}).$$

Soient η et $\tilde{\eta}$ tels que $v_{y,x}(c(\chi))=v_{x,y}(\eta)$ et $v_{\tilde{y},\tilde{x}}(c(\tilde{\chi}))=v_{\tilde{x},\tilde{y}}(\tilde{\eta})$. On passe de $c(\chi)$ à η et de $c(\tilde{\chi})$ à $\tilde{\eta}$ par des changements de paramétrages identiques (cf. 1.6). Donc par 1.5 on a aussi $(I_{N_{x,y}}\otimes 1)(\eta)=\tilde{\eta}$. On a donc

$$\begin{aligned}R^{C_G(x)}_{C_G(x)\cap C_G(y)}(\theta_1 v_{y,x}(c(\chi)))&=R^{C_G(x)}_{C_G(x)\cap C_G(y)}(\theta_1 v_{x,y}(\eta))\\&=(\sigma_x)^{(y)}(\eta)\end{aligned}$$

par définition de $(\sigma_x)^{(y)}$.

Il en résulte

$$\begin{aligned}(\Delta_x)^{(y)}(R^{C_G(x)}_{C_G(x)\cap C_G(y)}(\theta_1 *R^{C_G(y)}_{C_G(x)\cap C_G(y)}(v_y(\chi))))&=(\sigma_{\tilde{x}})^{(\tilde{y})}((I_{N_{xyv}}\otimes 1)(\eta))\\&=R^{C_{\tilde{G}}(\tilde{x})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(\tilde{\theta}_1(v_{\tilde{y},\tilde{x}}(\tilde{\eta}))\\&=R^{C_{\tilde{G}}(\tilde{x})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(\tilde{\theta}_1(v_{\tilde{y},\tilde{x}}(c(\tilde{\chi})))\\&=R^{C_{\tilde{G}}(\tilde{x})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(\tilde{\theta}_1 *R^{C_{\tilde{G}}(\tilde{y})}_{C_{\tilde{G}}(\tilde{x})\cap C_{\tilde{G}}(\tilde{y})}(v_{\tilde{y}}(\tilde{\chi}))).\end{aligned}$$

La première égalité résulte de la définition de $(\Delta_x)^{(y)}$, la seconde de la définition de $(\sigma_{\tilde{x}})^{(\tilde{y})}$, la troisième des considérations précédentes. Enfin la dernière résulte de 2.2 appliquée à $C_{\tilde{G}}(\tilde{y})$, compte-tenu de $m_{\tilde{y},\tilde{x}}=\tilde{m}_2$. □

De la démonstration de 2.4 retenons une précision sur le comportement des isométries relativement aux fonctions de Lusztig et à la décomposition en séries:

2.5. *Soit* $\zeta \in \mathfrak{R}(B)$ *et*

$$*R^G_{C_{\tilde{G}}(x)}(\zeta) = \sum_{(y)} \xi^{(y)}$$

la décomposition selon les séries de $*R^G_{C_G(x)}(\zeta)$ *(donc* $\xi^{(y)} \in \mathfrak{R}^{(y)}(b_x)$ *et* (y) *parcourt l'ensemble des l-classes de* $C_G(x)$*). Alors*

$$*R^{\tilde{G}}_{C_G(\tilde{x})}(\Delta(\zeta)) = \sum_{(y)} (\Delta_x)^{(y)}(\xi^{(y)}).$$

3 Démonstration du théorème

La démonstration du théorème 1.13 est basée sur le bon comportement des isométries Δ et Δ_x introduites en 2.3 vis-à-vis des applications de décomposition généralisées. Rappelons la définition de l'application de décomposition généralisée d^x_G: si ζ est une fonction centrale sur G à valeurs dans $\mathfrak{L}$ et $x \in G_l$, on pose

$$d^x_G(\zeta)(y) = \zeta(xy) \qquad (y \in C_G(x)_{l'})$$

et on prolonge éventuellement $d^x_G(\zeta)$ par zéro dans $C_G(x) - C_G(x)_{l'}$, obtenant une fonction centrale sur $C_G(x)$. Selon le second théorème fondamental de Brauer, si ζ s'écrit sur un l-bloc B de G, $d^x_G(\zeta)$ décompose sur les caractères de Brauer des blocs de $C_G(x)$ en correspondance avec B. Sous les hypothèses du théorème 1.13 et avec les notations de 2.1, d^x_G se restreint à $\mathfrak{L} \otimes \mathfrak{R}(B)$ en prenant valeur dans $\mathfrak{L} \otimes \mathfrak{R}(b_x)$.

3.1 Proposition. *Notations et hypothèses de* 2.3 *et* 2.4.

a) *Le diagramme suivant est commutatif*

$$\begin{array}{ccc} \mathfrak{L} \otimes \mathfrak{R}(B) & \xrightarrow{d^x_G} & \mathfrak{L} \otimes \mathfrak{R}(b_x) \\ \downarrow{\scriptstyle 1 \otimes \Delta} & & \downarrow{\scriptstyle 1 \otimes \Delta_x} \\ \mathfrak{L} \otimes \mathfrak{R}(\tilde{B}) & \xrightarrow{d^{\tilde{x}}_{\tilde{G}}} & \mathfrak{L} \otimes \mathfrak{R}(b_{\tilde{x}}) \end{array}$$

b) *On a* $\Delta(e_B(\mathfrak{P}_{\mathfrak{D}}(B))) = e_{\tilde{B}}(\mathfrak{P}_{\mathfrak{D}}(\tilde{B}))$.

Remarquons que l'assertion a) pour l'ensemble des $x \in (G_l - \{1\})$ implique b) et que b) implique a) pour $x = 1$. En effet, si $\zeta \in \mathfrak{R}(B)$, ζ est l'image par e_G

d'un projectif si, et seulement si, $d_G^x(\zeta)=0$ pour tout $x\in(G_l-\{1\})$. Si $m_x(1)<k$, il n'existe aucun élément de Brauer en x de B, donc $d_G^x(\mathfrak{R}(B))=\{0\}$. Il suffit de considérer les $x\in G_0$ et alors $d_G^x(\Delta(\zeta))=0$ par a). On obtient ainsi l'inclusion $\Delta(e_B(\mathfrak{P}_\mathfrak{D}(B))\subset e_{\tilde{B}}(\mathfrak{P}_\mathfrak{D}(\tilde{B}))$ puis l'égalité en échangeant B et $\tilde{B}$ et donc un diagramme commutatif

$$\begin{array}{ccc} \mathfrak{P}_\mathfrak{D}(B) & \xrightarrow{\delta} & \mathfrak{P}_\mathfrak{D}(\tilde{B}) \\ \downarrow{\scriptstyle e_B} & & \downarrow{\scriptstyle e_{\tilde{B}}} \\ \mathfrak{R}(B) & \xrightarrow{\Delta} & \mathfrak{R}(\tilde{B}) \end{array}$$

où δ est isométrique. Ce diagramme a pour dual le diagramme commutatif

$$\begin{array}{ccc} \mathfrak{R}(B) & \xrightarrow{\Delta} & \mathfrak{R}(\tilde{B}) \\ \downarrow{\scriptstyle d_B} & & \downarrow{\scriptstyle d_{\tilde{B}}} \\ \mathfrak{R}_\mathfrak{F}(B) & \xrightarrow{{}^t\delta^{-1}} & \mathfrak{R}_\mathfrak{F}(\tilde{B}). \end{array}$$

Or d_G^1 n'est autre que l'extension de d_G aux espaces de fonctions centrales; la relation de commutativité annoncée en a) est donc vraie pour $x=1$ sans tensoriser par $\mathfrak{L}$.

Démontrons maintenant l'assertion a) par récurrence sur N et pour $x\neq 1$. Si $N=0$, $\mathfrak{R}(B)$ est engendré par un caractère irréductible et projectif ξ, donc $d_G^x(\mathfrak{R}(B))=\{0\}$ pour tout $x\in(G_l-\{1\})$ tandis que $d_G^1(\xi)=\xi$. La commutativité du diagramme est alors évidente pour tout $x\in G_l$.

La calcul de d_G^x se déduit du calcul des ${}^*R_{C_G(x)}^G$ grâce à la formule ([2], (4.3))

$$d_G^x=d_{C_G(x)}^x\circ{}^*R_{C_G(x)}^G.$$

Soit $\zeta\in\mathfrak{R}(B)$ et décomposons ${}^*R_{C_G(x)}^G$ comme en 2.5. Si (y) est une l-classe d'éléments de $C_G(x)$, soit $m'_{x,y}=m'$: $\mathscr{F}^2\to\mathbb{N}$ la fonction associée (voir 1.6) et $\theta_y\in\mathrm{Irr}(C_G(x)\cap C_G(y))$ défini par m'. Comme x est central, on a

$$d_{C_G(x)}^x(\zeta^{(y)})=\theta_y(x)\,d_{C_G(x)}^1(\zeta^{(y)}).$$

Puisque $e\neq 1$ (1.13), G ne contient aucun élément central d'ordre l, donc l'hypothèse d'induction s'applique aux isométries Δ_x pour tout $x\in(G_l-\{1\})$. On a

donc $\Delta_x \circ d^1_{C_G(x)}(\xi^{(y)}) = d^1_{C_{\tilde{G}}(\tilde{x})} \circ \Delta_x(\xi^{(y)})$. D'autre part $\tilde{x}$ et $\tilde{y}$ définissent $m'_{\tilde{x},\tilde{y}} = \tilde{m}'$, d'où $\tilde{\theta}_{\tilde{y}}$ tel que $\tilde{\theta}_{\tilde{y}}(\tilde{x}) = \theta_y(x)$. On obtient par 2.5

$$\begin{aligned}(1 \otimes \Delta_x)(d^x_G(\zeta)) &= \sum_{(y)} \theta_y(x)\, \Delta_x(d^1_{C_G(x)}(\xi^{(y)})) \\ &= \sum_{(y)} \tilde{\theta}_{\tilde{y}}(\tilde{x})(d^1_{C_G(x)}(\Delta_x(\xi^{(y)})) \\ &= \sum_{(y)} d^{\tilde{x}}_{C_{\tilde{G}}(\tilde{x})}(\Delta_x(\xi^{(y)})) \\ &= (d^{\tilde{x}}_{C_{\tilde{G}}(\tilde{x})} \circ {}^*R^{\tilde{G}}_{C_{\tilde{G}}(\tilde{x})})(\Delta(\zeta))\end{aligned}$$

ce qui démontre a). □

3.2 Fin de la démonstration

Nous sommes dans une situation analogue à celle décrite dans [5] à propos des groupes symétriques. Rappelons à quelle condition l'isométrie Δ est dite parfaite (selon [3], 1.1); Δ est caractérisée par un caractère généralisé du produit $G \times \tilde{G}$, appartenant à $B \otimes \tilde{B}$ à savoir

$$\Theta := \sum_{\zeta \in \mathrm{Irr}\, B} \zeta \otimes \Delta(\zeta)^*$$

(où l'on désigne par χ^* le contragrédient d'un caractère χ). Les conditions sont

(P.1) *Pour tout* $(g, h) \in G \times \tilde{G}$, $\Theta(g, h) \in (|C_G(g)|.\mathfrak{O}) \cap (|C_{\tilde{G}}(h)|.\mathfrak{O})$,

(P.2) *Si* $\Theta(g, h) \neq 0$, *alors g est d'ordre premier à l si, et seulement si, h est d'ordre premier à l.*

Via les applications d^x_G l'espace des fonctions centrales sur G décompose en somme directe orthogonale, et cette décomposition croise la décomposition selon les blocs. Posons

$$\begin{aligned}d^x_B &:= (1 \otimes d_{b_x}) \circ d^x_G \colon \mathfrak{L} \otimes \mathfrak{R}(B) \to \mathfrak{L} \otimes \mathfrak{R}_{\mathfrak{F}}(b_x), \\ d^{\tilde{x}}_{\tilde{B}} &:= (1 \otimes d_{b_{\tilde{x}}}) \circ d^{\tilde{x}}_{\tilde{G}} \colon \mathfrak{L} \otimes \mathfrak{R}(\tilde{B}) \to \mathfrak{L} \otimes \mathfrak{R}_{\mathfrak{F}}(b_{\tilde{x}})\end{aligned}$$

(la composition avec d_{b_x} n'est qu'un changement de base, c'est l'identité sur les espaces de fonctions centrales, car toute image par d^x_G est nulle sur les éléments non p-réguliers de $C_G(x)$). On obtient, avec les notations introduites précédemment, des décompositions parallèles (une telle décomposition a été signalée en toute généralité par M. Broué, «On characters of height zero», The Santa-Cruz Conference on finite groups, AMS, 1980)

$$\begin{aligned}D_B &:= \oplus_{(x)}\, d^x_B \colon \mathfrak{L} \otimes \mathfrak{R}(B) \to \oplus_{(x)}\, \mathfrak{L} \otimes \mathfrak{R}_{\mathfrak{F}}(b_x), \\ D_{\tilde{B}} &:= \oplus_{(\tilde{x})}\, d^{\tilde{x}}_{\tilde{B}} \colon \mathfrak{L} \otimes \mathfrak{R}(\tilde{B}) \to \oplus_{(\tilde{x})}\, \mathfrak{L} \otimes \mathfrak{R}_{\mathfrak{F}}(b_{\tilde{x}}).\end{aligned}$$

Dans ces égalités (x) et $(\tilde{x})$ parcourent l'ensemble des classes décrites en 2.1. Soient $e^x_B \colon \mathfrak{L} \otimes \mathfrak{P}_{\mathfrak{O}}(b_x) \to \mathfrak{L} \otimes \mathfrak{R}(B)$ les adjointes des applications d^x_B. Alors la somme orthogonale

$$E_B := \otimes_{(x)}\, e^x_B \colon \otimes_{(x)}\, \mathfrak{L} \otimes \mathfrak{P}_{\mathfrak{O}}(b_x) \to \mathfrak{L} \otimes \mathfrak{R}(B)$$

est adjointe de D_B. Selon la Prop. 3.1 et sa démonstration appliquées à $C_G(x)$ pour tout x on a des diagrammes commutatifs duaux

$$\begin{array}{ccc} \mathfrak{P}_{\mathfrak{O}}(b_x) & \xrightarrow{\delta_x} & \mathfrak{P}_{\mathfrak{O}}(b_{\tilde{x}}) \\ \downarrow{\scriptstyle e_{b_x}} & & \downarrow{\scriptstyle e_{b_{\tilde{x}}}} \\ \mathfrak{R}(b_x) & \xrightarrow{\Delta_x} & \mathfrak{R}(b_{\tilde{x}}) \end{array} \qquad \begin{array}{ccc} \mathfrak{R}(b_x) & \xrightarrow{\Delta_x} & \mathfrak{R}(b_{\tilde{x}}) \\ \downarrow{\scriptstyle d_{b_x}} & & \downarrow{\scriptstyle d_{b_{\tilde{x}}}} \\ \mathfrak{R}_{\mathfrak{F}}(b_x) & \xrightarrow{{}^t\delta_{\tilde{x}}^{-1}} & \mathfrak{R}_{\mathfrak{F}}(b_{\tilde{x}}). \end{array}$$

Posons $\Delta' = \oplus_{(x)}(1 \otimes \delta_x)$, soit ${}^t\Delta'^{-1} = \oplus_{(x)}(1 \otimes {}^t\delta_x^{-1})$. Il résulte encore de la Prop. 3.1 que les diagrammes suivants, où les sommes directes sont orthogonales, sont commutatifs et duaux l'un de l'autre:

$$\begin{array}{ccc} \oplus_{(x)}\, \mathfrak{L} \otimes \mathfrak{P}_{\mathfrak{O}}(b_x) & \xrightarrow{\Delta'} & \oplus_{(\tilde{x})}\, \mathfrak{L} \otimes \mathfrak{P}_{\mathfrak{O}}(b_{\tilde{x}}) \\ \downarrow{\scriptstyle E_B} & & \downarrow{\scriptstyle E_{\tilde{B}}} \\ \mathfrak{L} \otimes \mathfrak{R}(B) & \xrightarrow{1 \otimes \Delta} & \mathfrak{L} \otimes \mathfrak{R}(\tilde{B}) \\ \mathfrak{L} \otimes \mathfrak{R}(B) & \xrightarrow{1 \otimes \Delta} & \mathfrak{L} \otimes \mathfrak{R}(\tilde{B}) \\ \downarrow{\scriptstyle D_B} & & \downarrow{\scriptstyle D_{\tilde{B}}} \\ \oplus_{(x)}\, \mathfrak{L} \otimes \mathfrak{R}_{\mathfrak{F}}(b_x) & \xrightarrow{{}^t\Delta'^{-1}} & \oplus_{(\tilde{x})}\, \mathfrak{L} \otimes \mathfrak{R}_{\mathfrak{F}}(b_{\tilde{x}}). \end{array}$$

L'application Θ définie par Δ peut alors s'écrire dans $\oplus_{(x)}\, \mathfrak{P}_{\mathfrak{O}}(b_x) \otimes \mathfrak{R}_{\mathfrak{F}}(b_{\tilde{x}})$. Soit pour toute classe (x) comme en 2.1 $\Phi(x)$ une base de $\mathfrak{R}_{\mathfrak{F}}(b_{\tilde{x}})$ et soit $\{\hat{\phi} \mid \phi \in \Phi(x)\}$ une base de $\mathfrak{P}_{\mathfrak{O}}(b_x)$ dont l'image par δ_x soit duale de $\Phi(x)$. On a alors $\Theta = \sum_{(x)} \sum_{\phi \in \Phi(x)} \hat{\phi} \otimes \phi$. Or il résulte des définitions de e_B^x et d_B^x que pour $\phi \in \Phi(x)$ et pour tout $(g, h) \in G \times \tilde{G}$, on a

(i) $(\hat{\phi} \otimes \phi)(xy, \tilde{x}z) = \hat{\phi}(z)$ si $y \in C_G(x)_{l'}$ et $z \in C_{\tilde{G}}(\tilde{x})_{l'}$;

(ii) $(\hat{\phi} \otimes \phi)(g, h) = 0$ si g_l *n'est pas conjugué de* x *dans* G. *ou* h_l *n'est pas conjugué de* $\tilde{x}$ *dans* $\tilde{G}$.

L'assertion (ii) implique la condition (P.2). Comme $\hat{\phi}$ est projectif, $\hat{\phi}(y)$ est divisible par l'ordre de $C_{C_G(x)}(y) = C_G(xy)$ dans $\mathfrak{O}$. La condition de divisibilité (P.1) est donc démontrée pour $g = xy$ et elle l'est pour h par symétrie. □

Bibliographie

1. Broué, M., Michel, J.: Blocs et séries de Lusztig dans un groupe réductif fini. J. Reine Angew. Math. **395**, 56–67 (1989)

2. Broué, M.: Les l-blocs des groupes $GL(n, q)$ et $U(n, q^2)$ et leurs structures locales. Astérisque **133–134**, 159–188 (1986)
3. Broué, M.: Blocs, isométries parfaites, catégories dérivées. C.R. Acad. Sci., Paris **307**, 13–18 (1988)
4. Digne, F., Michel, J.: Théorie de Deligne-Lusztig et caractères des groupes linéaires et unitaires. J. Algebra **107**, 217–255 (1987)
5. Enguehard, M.: Isométries parfaites dans les groupes symétriques. Rapport de recherche du LMENS, DMI, Ecole Normale Supérieure (1988)
6. Farahat, H.: On the representations of the symmetric group. Proc. London Math. Soc. **4**, 303–316 (1954)
7. Fong, P., Srinivasan, B.: The blocks of finite general and unitary groups. Invent. Math. **69**, 109–153 (1982)
8. Green, J.A.: The characters of the finite general linear groups. Trans. A.M.S. **80**, 402–447 (1955)
9. Deligne, P., Lusztig, G.: Representations of reductive groups over finite fields. Ann. Math. **103**, 103–161 (1976)
10. Lusztig, G., Srinivasan, B.: The characters of the finite unitary groups. J. Algebra **49**, 167–171 (1977)
11. Michler, G.O., Olsson, J.B.: Character correspondences in finite general linear, unitary and symmetric groups. Math. Z. **184**, 203–233 (1983)
12. Srinivasan, B.: Representations of Finite Chevalley Groups. (Lecture Notes Math. vol. 764) Berlin Heidelberg New York: Springer 1979
13. Zelevinsky, A.V.: Representations of Finite Classical Groups. (Lecture Notes Math. vol. 869) Berlin Heidelberg New York: Springer 1980

Math. Z. 206, 25–34 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

The Auslander-Reiten quiver of a modular group algebra revisited *

Christine Bessenrodt
Fachbereich Mathematik, Universität Duisburg, W-4100 Duisburg, FRG

Received April 14, 1989, in final form October 9, 1989

1 Introduction

Let G be a finite group and F a field of characteristic $p>0$. In an important paper [14] P. Webb described the Auslander-Reiter quiver $A(FG)$ of the modular group algebra FG. He proved that the tree class of a connected component Δ of the stable Auslander-Reiten quiver $A(FG)_s$ is either a Dynkin diagram A_n, a Euclidean diagram or one of the five infinite trees A_∞, B_∞, C_∞, D_∞ and A_∞^∞. For the reduced graph of Δ we have as an additional possibility the Euclidean diagram $\tilde{A}_n$ (see [2]). Moreover, if Δ is a component of $A(FG)$ which contains no projective modules, then the tree class of Δ is one of the infinite trees.

For odd primes p, Okuyama could show that the tree class is never a Euclidean diagram [13]. The purpose of this paper is to prove that for $p=2$ no Euclidean diagrams apart from the known examples occur as tree classes, and we will consider this question more generally for the reduced graph of Δ. More precisely, the main result of this article is the following:

Theorem 1.1 *Let Δ be a connected component of the stable Auslander-Reiten quiver $A(FG)_s$, and let B be the block containing Δ.*

If the reduced graph or the tree class of Δ is a Euclidean diagram, then B has a Klein four group V_4 as a defect group and Δ is isomorphic to either $\mathbb{Z}\tilde{A}_{1,2}$ or $\mathbb{Z}\tilde{B}_3$ or $\mathbb{Z}\tilde{A}_5$.

Furthermore, if F is algebraically closed then either $\Delta \simeq \mathbb{Z}\tilde{A}_{1,2}$ and B is Morita equivalent to FV_4, or $\Delta \simeq \mathbb{Z}\tilde{A}_5$ and B is Morita equivalent to FA_4 or the principal block of FA_5.

In the course of the proof we will obtain some more information, but we do not want to put all these details into the statement of the theorem.

As a consequence of our theorem we have the following bounds which we know to be sharp in the case of the group A_4 and a field F not containing a primitive cube root of unity (compare [2, 2.31.4 and 2.31.5]).

* This work has been supported by the DFG project "Darstellungstheorie" at the University of Essen.

Corollary 1.2 *The number of indecomposable summands in the middle of an Auslander-Reiten sequence of an indecomposable FG-module is at most four, and if it is equal to four, then one of the summands is projective.*

In particular, we obtain:

Corollary 1.3 *The heart of any projective indecomposable FG-module decomposes into at most three indecomposable summands.*

Even better, we know from the results above that the case of three summands in the heart of a projective indecomposable FG-module is rather exceptional. Apart from the case $\tilde{B}_3$ which occurs for blocks with a Klein four group as a defect group and a "small" field, it can only happen for components of type D_∞ resp. B_∞. Note that the second case does not occur for algebraically closed fields.

For the basic facts on Auslander-Reiten sequences and the Auslander-Reiten quiver and the terminology and notations we refer the reader to [2] and [14].

2 Preliminaries

We start with a slightly improved version of the result by Okuyama mentioned in the introduction [13, Theor. 2]. If Δ is a component of the stable Auslander-Reiten quiver for FG, we call a p-subgroup $Q \leqq G$ *a minimal vertex on* Δ if it is minimal among the vertices of the indecomposable FG-modules in Δ.

Theorem 2.1 *Let Δ be a connected component of the stable Auslander-Reiten quiver for FG, and suppose that its reduced graph is a Euclidean diagram. Let $Q \leqq G$ be a minimal vertex on Δ. Then the following holds:*

(i) *Δ contains a module with source F_Q or $\Omega(F_Q)$.*
(ii) *$p=2$, and $Q \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.*

Proof. For the first assertion we follow the lines of the first part of the proof of Theorem 2 in [13]. Consider the function d on Δ given by

$$d(V) = \dim \operatorname{Ext}^1_{FQ}(F_Q, V_Q) = \dim \operatorname{Ext}^1_{FG}({F_Q}^G, V)$$

for $V \in \Delta$. If there is no M in Δ such that M is a direct summand of ${F_Q}^G$ or $\Omega(F_Q)^G$ then d is an additive function [13]. This induces an additive function on $\mathbb{Z}\Gamma$, where Γ is the reduced graph of Δ. By [14, Cor. 2.4] resp. [9] such a function takes bounded values, and hence for every $V \in \Delta$ the restriction V_Q is periodic. But Δ contains a module with vertex Q and so this module must be periodic. This gives a contradiction by [2, 2.31.11].

For the second assertion we slightly modify the proof in [13]. Consider the set

$$\mathcal{S} = \{U \mid U \text{ is an indecomposable summand of } V_Q, \text{ for some } V \in \Delta\}$$

As the reduced graph of Δ is a finite diagram, $\mathcal{S}$ has only finitely many Ω^2-orbits. Certainly, Q is not cyclic, so by [14] we know that already the component of the Auslander-Reiten quiver containing F_Q contains infinitely many Ω^2-orbits as long as Q is not isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. So if Q is not a Klein four group we can choose an indecomposable non-periodic FQ-module X such that neither

X nor $\Omega(X)$ are sources of any module in Δ. Now consider the function d on Δ defined by

$$d(V)=\dim \operatorname{Ext}^1_{FQ}(X, V_Q)=\dim \operatorname{Ext}^1_{FG}(X^G, V)$$

for any $V\in\Delta$. As above, this function takes bounded values on Δ. In particular, since F_Q or $\Omega(F_Q)$ is a direct summand of M_Q for some $M\in\Delta$, the sequence $\dim \operatorname{Ext}^1_{FQ}(X, \Omega^n(F_Q))$, $n\geqq 0$, is bounded. This implies that X is periodic [13, Lemma 2], contradicting the choice of X. □

For the convenience of the reader we want to recall some results which we will use in the sequel. An important tool will be the following theorem by Kawata [10]:

Theorem 2.2 *Let Δ be a connected component of $A(FG)_s$, and let $Q\leqq G$ be a minimal vertex on Δ. Let $N=N_G(Q)$ and let f denote the Green correspondence with respect to (G, Q, N). Let U be a module in Δ with vertex Q, and let Θ be the component of $A(FN)_s$ containing fU. Define Λ as the subquiver of Θ consisting of those FN-modules $L\in\Theta$ which are connected to fU by a walk along modules with vertices containing Q.*

Then there is a graph isomorphism $\psi\colon \Lambda \xrightarrow{\cong} \Delta$ which behaves like the Green correspondence. More precisely, for $L\in\Lambda$ the module L^G has a unique indecomposable summand $\psi L=M$ whose vertex contains Q. Moreover, $M\in\Delta$. Furthermore, L is a direct summand of $(L^G)_N$ with multiplicity 1.

We will also need the following properties of Auslander-Reiten sequences (see e.g. [2, 2.17.10] and [8, Prop. 3.9]).

Proposition 2.3 *An Auslander-Reiten sequence $0\to\Omega^2 A\to C\to A\to 0$ splits on restriction to a subgroup H of G if and only if H does not contain a vertex of A.*

Proposition 2.4 *Let H be a normal subgroup of G, and let W be a G-stable indecomposable FH-module. Let r be the multiplicity of an indecomposable FG-module V in W^G, and let $\delta=\delta_{FG}(V)/\delta_{FH}(W)$, where $\delta_R(M)$ denotes the F-dimension of $\operatorname{End}_R(M)/J\operatorname{End}_R(M)$.*

Then restricting the Auslander-Reiten sequence ending with V to H gives δr times the Auslander-Reiten sequence ending with W plus a split sequence.

3 Proof of the Theorem

It is the aim of this section to prove the theorem stated in the introduction.

So let Δ be a connected component of the stable Auslander-Reiten quiver $A(FG)_s$, and suppose that its reduced graph is a Euclidean diagram.

Let $Q\leqq G$ be a minimal vertex on Δ. By Kawata's Theorem we can pass to the normaliser $N=N_G(Q)$ and to the corresponding quiver $\Lambda=\Theta$ of $A(FN)_s$, since a Euclidean diagram cannot be embedded into any other Dynkin or Euclidean diagram. Let b denote the block of FN containing Λ.

Let us now consider the quiver Λ. If its reduced graph is a Euclidean tree then also the tree class is Euclidean and we know by Webb's Theorem that Λ as well as $\Omega\Lambda$ have to contain a projective mesh, i.e. an Auslander-Reiten sequence which has a projective summand in the middle. By [4] this also holds

in the case where the reduced graph is of type $\tilde{A}_n$. Thus A contains some simple module T and some module of the form ΩS, where S is simple.

Moreover, we already know by Theorem 2.1 that $p=2$ and $Q \cong \mathbb{Z}_2 \times \mathbb{Z}_2$, and that A contains a module with source F_Q or $\Omega(F_Q)$.

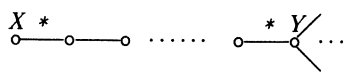
Let us note first that no module in A is periodic on Q, because otherwise all modules in A would have to be periodic on Q and hence A would contain a periodic module. In particular, any module in A with vertex Q has a source $\Omega^n F_Q$, for some $n \in \mathbb{Z}$. Since these modules are N-stable, we have for any module $X \in A$ with vertex Q

$$X_Q \cong (\Omega^n F_Q)^m, \quad \text{for some } n \in \mathbb{Z}, m \in \mathbb{N}.$$

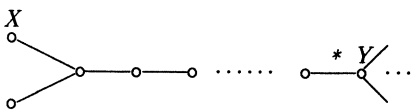
In particular, these restrictions are "homogeneous" and projective-free.

These properties are rather strong and come to good use together with the splitting property given in Proposition 2.3, as we will see in the following lemmas.

Lemma 3.1 *Suppose a part of the diagram for the reduced graph looks like*



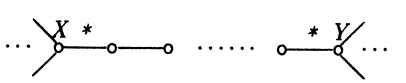
or



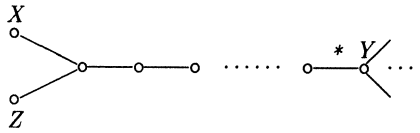
(*That is, we have indicated the position of the orbits of the modules X and Y, instead of * we may have any label, and to the right of Y the diagram is arbitrary.*) *If X has vertex $> Q$, then all modules on the walk from X to Y and including Y are inhomogeneous on Q, and hence have also vertex $> Q$.*

Proof. Start with the Auslander-Reiten sequence for X. Note that in both cases the projective-free part of the middle term is indecomposable or a sum of copies of one indecomposable module, say E. Since X has vertex $> Q$, this sequence splits on Q by Proposition 2.3. As X_Q is not periodic, the projective-free part of the sum $X_Q \oplus \Omega^2(X)_Q$ is not N-homogeneous, and hence the core of E_Q is not N-homogeneous. This implies that E also has vertex $> Q$, so we can next split the Auslander-Reiten sequence for E on Q. For the upper diagram it is then easy to see that the indecomposable projective-free module in the middle which is not in the orbit of X is again N-inhomogeneous on Q. For the other diagram one has to be a little bit more careful in keeping track of the summands on Q to get with similar arguments that the next module on the walk is again N-inhomogeneous on Q. Continuing like this along the walk to Y it is straightforward to obtain the assertion. □

Lemma 3.2 *Suppose a part of the diagram looks like*



or



where X, Y, Z all have vertex Q and the modules on the walk from X to Y have vertices > Q. Then all these modules are also projective-free on Q.

Proof. It is easy to see that the only critical part is the one containing projective meshes, since outside this region we have additivity on Q. So consider the convex hull of the projective meshes lying in the orbits between X and Y:

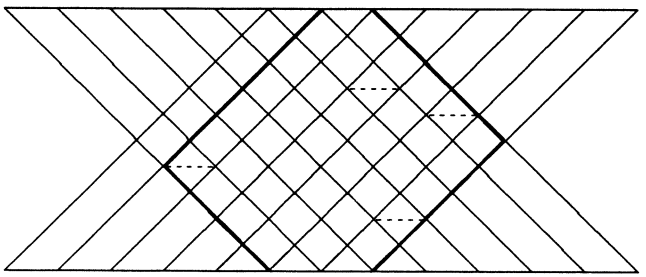
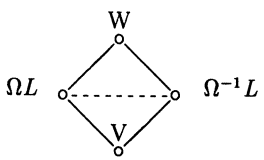


Fig. 1

Here the modules on the upper and lower horizontal lines are projective-free on Q. Since the labels are irrelevant for the following arguments we have omitted them in the diagram above for simplicity. Then we could only have projectives on Q inside the marked region; up to and including the borderline the modules have to be projective-free on Q. Now a projective mesh looks like

$$0 \to \Omega L \to ht\, P_L \oplus P_L \to \Omega^{-1} L \to 0$$

where L is simple and $ht\ P_L = \operatorname{rad} P_L / \operatorname{soc} P_L$ denotes the heart of P_L. We picture such a projective mesh as



where $ht\ P_L \cong V \oplus W$. Here (and in the following) the arrows are omitted; they are all directed from the left to the right. Also, we only indicate the presence of a projective by a dashed line, since it is clear what the corresponding Auslander-Reiten sequence looks like.

Since Q is normal in N, we have $L_Q \cong F_Q{}^n$ for some $n \in \mathbb{N}$. Hence $(\Omega L)_Q \cong (\Omega F_Q)^n \oplus FQ^m$ and $(\Omega^{-1} L)_Q \cong (\Omega^{-1} F_Q)^n \oplus FQ^l$ for some $m, l \in \mathbb{N}_0$.

Since $\dim \Omega L = \dim P_L - \dim L = \dim \Omega^{-1} L$ and $\dim \Omega F_Q = |Q| - 1 = \dim \Omega^{-1} F_Q$, this implies $m = l$.

Now consider a projective mesh next to the border of the convex region. Since one end is projective-free on Q, so has to be the other. But on the other hand, $P_{L|Q} | \Omega L_Q \oplus \Omega^{-1} L_Q$, giving a non-zero projective summand on Q – a contradiction.

Thus there is no projective mesh, and all modules on the walk from X to Y are projective-free on Q. □

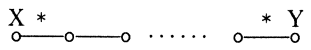
Lemma 3.3 *On any walk from the orbit of ΩS to the orbit of T there have to be at least two modules with vertex Q.*

Proof. Restricting to Q, the modules in the orbit of ΩS resp. T decompose into copies of $\Omega^{2n+1} F_Q$ resp. $\Omega^{2n} F_Q$ for suitable $n \in \mathbb{Z}$ (modulo projectives). Let us call modules which are sums of odd resp. even Ω's of F_Q (modulo projectives) to be of "odd" resp. "even" type on Q. Starting from the orbit of ΩS, all the modules on the walk up to and including the first one with vertex Q are of odd type. Similarly, all the modules on the walk from the last with vertex Q onwards to the orbit of T are of even type. □

With all these preparations let us now look at all the Euclidean diagrams in turn.

Case 1. Let us first assume that the simple FN-modules T and S in Λ and $\Omega\Lambda$ both have vertex $> Q$. By Lemma 3.2 this cannot happen if the diagram is of type $\tilde{A}_n$, so we only have to consider the Euclidean trees.

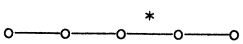
(i) Diagrams of the form



i.e.: $\tilde{A}_{1,1}$, $\tilde{A}_{1,2}$, $\tilde{B}_n$, $\tilde{C}_n$, $\widetilde{BC}_n$, $\tilde{G}_{21}$, $\tilde{G}_{22}$.

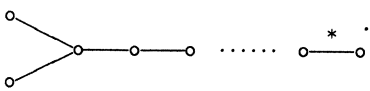
Because of Lemma 3.1, we must have $v_X X = Q = v_X Y$. But then the projective mesh ending with ΩS lies between two Q-orbits, contradicting Lemma 3.2.

(ii) The diagrams $\tilde{F}_{41}$ and $\tilde{F}_{42}$, which are both of the form

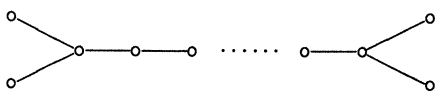


are also easily excluded by using the lemmas above.

(iii) Diagrams of the form

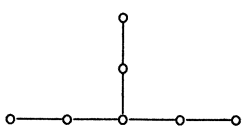


resp.



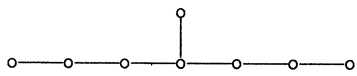
i.e. the diagrams $\tilde{D}_n$, $\widetilde{BD}_n$ and $\widetilde{CD}_n$, are again impossible by the lemmas above.

(iv) $\tilde{E}_6$



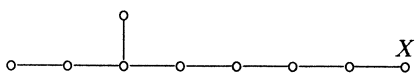
Let us first look at ΩS. No matter where it sits, it induces odd type on Q for the modules in the orbit corresponding to the central node. But similarly, T induces even type on Q for the center modules, which gives a contradiction.

(v) $\tilde{E}_7$



Again, this is easily killed by the lemmas above.

(vi) $\tilde{E}_8$



Here, a module at the position of X must have vertex Q because of Lemma 3.1, and then Lemmas 3.2 and 3.3 give a contradiction.

This finishes Case 1, showing that this situation can not occur.

Case 2. We may now assume that at least one of the simple modules T and S in A and ΩA has vertex Q.

First we look at the case where the reduced graph is of type $\tilde{A}_n$. We may assume that A contains a module U with source F_Q. Then by Proposition 2.4 the restriction of the Auslander-Reiten sequence of U to Q gives copies of the Auslander-Reiten sequence of F_Q, plus possibly a split sequence. But the occurrence of a split sequence leads to a contradiction by homogeneity arguments and Proposition 2.3. So we just have copies of the Auslander-Reiten sequence of F_Q:

$$0 \to \Omega^2 F_Q \to V \oplus W \to F_Q \to 0,$$

where V, W are indecomposable and $W=0$ if Q is not dihedral (see [14]). Hence the modules in the middle of the Auslander-Reiten sequence of U must also have vertex Q, and they have V resp. W as their sources. Let M be one of these modules, say with source V. If the restriction of its Auslander-Reiten sequence to Q has a split sequence as a summand, then the summand $K \ncong \Omega^2 U$ in the middle is not N-homogeneous on Q (note that V and $\Omega^2 V$ can not be conjugate because V is not periodic).

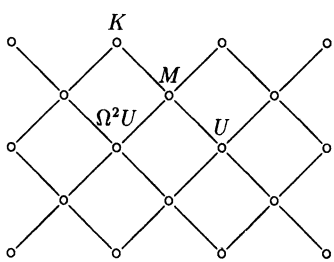


Fig. 2

Thus K has vertex $> Q$, and so its Auslander-Reiten sequence splits on Q. It is easy to see that then all the modules in the following rows are not N-homogeneous on Q which gives a contradiction, because we have to come back to the orbit of U.

So the restriction of the Auslander-Reiten sequence of M to Q gives just copies of the Auslander-Reiten sequence of V, and K must have the middle term of this latter sequence which is not isomorphic to $\Omega^2 F$ as a source. Continuing like this with similar arguments as above, we see that all modules in A have vertex Q and their sources are the modules in the Auslander-Reiten component of F_Q. As we know the Auslander-Reiten component of F_Q for the Klein four group Q, from the discussion above we learn that we must have the following picture, where the i at a node means that the corresponding module has source $\Omega^i F_Q$:

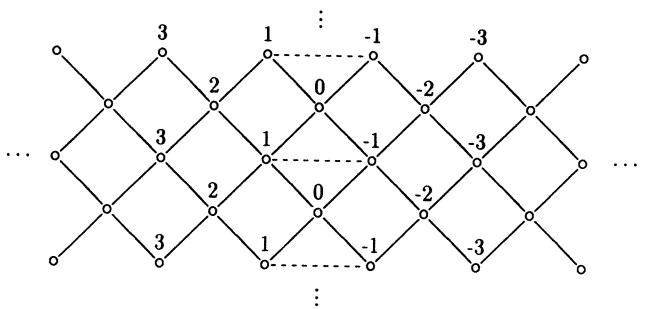


Fig. 3

In particular, this tells us that the quiver A must have projective meshes "above" the projective meshes on Q. We also deduce that all modules in A with source F_Q are simple (consider the component ΩA!).

It is easy to count the number of summands on Q: every module in A decomposes into $r = \dim T$ summands on Q. In particular, all simples in A have the same dimension.

Since the block b containing A has a simple module with vertex $Q \cong \mathbb{Z}_2 \times \mathbb{Z}_2$, we know by [11] that b has a defect group D with $C_D(Q) = Q$. On the other hand, as Q is normal in N, $N/C_G(Q)$ is isomorphic to a subgroup of S_3, so $|D/C_D(Q)| \leqq 2$. Hence $D = Q \cong \mathbb{Z}_2 \times \mathbb{Z}_2$ or D is dihedral of order 8.

Let us now assume that F is algebraically closed. Then by a result of Erdmann [6], we can only have the Euclidean types $\tilde{A}_{1,2}$ and $\tilde{A}_n$ occurring for the components of b. Furthermore, it is well-known that a block with dihedral defect groups has at most three simple modules, hence by the picture above

$$n+1 = 2|\{\text{simple modules in } A\}| \leqq 6,$$

and thus $n \leqq 5$.

If D is dihedral of order 8, then either b is Morita equivalent to FD or b is Morita equivalent to FS_4 [6]. The Auslander-Reiten quiver for FD is well-known [2]; it does not have a Euclidean component. For the group algebra

FS_4 it is easy to exclude the only possibilities $\tilde{A}_{1,2}$, $\tilde{A}_1$ and $\tilde{A}_3$ using the known structure of the projective indecomposable modules.

Thus $D=Q$ has to be a Klein four group.

Now let us go back from the quiver component Δ of b to the component Δ of B. First note that the block b is the Brauer correspondent of B and so B has also defect group Q (see e.g. [1, §14]).

By the results of Erdmann [5], [6] we thus have one of the following cases:

(i) B is Morita equivalent to FQ and $\Delta \simeq \mathbb{Z}\tilde{A}_{1,2}$.

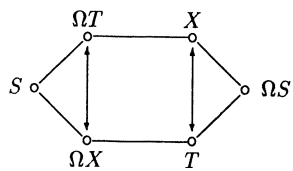
(ii) B is Morita equivalent to FA_4 and $\Delta \simeq \mathbb{Z}\tilde{A}_5$ contains all 3 simple B-modules.

(iii) B is Morita equivalent to the principal block of FA_5 and $\Delta \simeq \mathbb{Z}\tilde{A}_5$ is the component of the unique non-periodic simple B-module.

In all cases, Δ is the unique non-periodic component of B.

Using the explicit correspondence between Λ and Δ one may also reprove the information on the Green correspondents of the simple modules given by Erdmann [5].

For the proof of Case 2, we have assumed so far that F is algebraically closed. Since our component has only finitely many Ω^2-orbits, we can handle the general case by using that the behaviour of the reduced graph with respect to Galois extensions is well understood [2] and that Auslander-Reiten sequences are well behaved with respect to extensions from perfect fields [12]. The only new type that can occur is $\tilde{B}_3$, which arises from a Galois action on $\tilde{A}_5$ (see [2]). On the level of the normaliser (with S, $T$and X denoting the simple modules in b) this can be pictured as:



So the block b containing such a component Λ of type $\tilde{B}_3$ has again a Klein four defect group and two simple modules S and Y with projective covers

$$P_S\colon \begin{matrix} S \\ Y \\ S \end{matrix} \qquad P_Y\colon \begin{matrix} Y \\ Y \oplus S \oplus S \\ Y \end{matrix}$$

and $\dim Y = 2 \dim S$. Furthermore, we know that Y is not absolutely irreducible. This is exactly the situation in the case of the group algebra FA_4 of the alternating group A_4 over a field F not containing a primitive cube root of unity.

On the level of the block B we have again two possible cases as in the situation where the diagram is $\tilde{A}_5$. In the first case we get the same answer as above, i.e. the same structure of the two projective indecomposable modules. In the second case we have a block B with Klein four defect group and two simple modules, one of which is periodic.

This finishes Case 2 and hence the proof of the theorem.

References

1. Alperin, J.L.: Local representation theory. Cambridge: Cambridge University Press 1986
2. Benson, D.: Modular representation theory: New trends and methods. (Lecture Notes Math., vol. 1081) Berlin Heidelberg New York: Springer 1984
3. Brauer, R.: Notes on representations of finite groups. I, J. London Math. Soc., II. Ser. **13**, 162–166 (1976)
4. Butler, M.C.R., Ringel, C.M.: Auslander-Reiten sequences with few middle terms and applications to string algebras. Commun. Algebra **15**, 145–179 (1987)
5. Erdmann, K.: Blocks whose defect groups are Klein four groups: A correction. J. Algebra **76**, 505–518 (1982)
6. Erdmann, K.: Algebras and dihedral defect groups. Proc. London Math. Soc., III. Ser. **54**, 88–114 (1987)
7. Erdmann, K.: On the vertices of modules in the Auslander-Reiten quiver of p-groups. Math. Z. **203**, 321–334 (1990)
8. Green, J.A.: Functors on categories of finite group representations. J. Pure Appl. Algebra **37**, 265–298 (1985)
9. Happel, D., Preiser, U., Ringel, C.M.: Vinberg's characterization of Dynkin diagrams using subadditive functions with application to DTr-periodic modules. In: Dlab, V., Gabriel, P. (eds.) Representation Theory II, Proc. Ottawa Conf. 1979. (Lecture Notes Math. vol. 832) Berlin Heidelberg New York: Springer 1980
10. Kawata, S.: Module correspondence in Auslander-Reiten quivers for finite groups. Preprint 1988
11. Knörr, R.: On the vertices of simple modules. Ann. Math., II. Ser. **110**, 487–499 (1979)
12. Linnell, P.A.: The Auslander-Reiten quiver of a finite group. Arch. Math. **45**, 289–295 (1985)
13. Okuyama, T.: On the Auslander-Reiten quiver of a finite group. J. Algebra **110**, 425–430 (1987)
14. Webb, P.J.: The Auslander-Reiten quiver of a finite group. Math. Z. **179**, 97–121 (1982)

Math. Z. 206, 35–56 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On symplectic groups over polynomial rings

Fritz Grunewald [1], **Jens Mennicke** [2], **and Leonid Vaserstein** [3]

[1] Fakultät für Mathematik, Universität Bonn, Beringstrasse 4, W-5300 Bonn 1, Federal Republic of Germany

[2] Fakultät für Mathematik, Universität Bielefeld, W-4800 Bielefeld, Federal Republic of Germany

[3] Department of Mathematics, Penn State University, University Park, PA 16802, USA

Received April 4, 1989; in final form November 23, 1989

1 Introduction

By Suslin [17], $\mathrm{SL}_n R = \mathrm{E}_n R$ for all $n \geqq 3$ when $R = \mathbb{Z}[x_1, \ldots, x_m]$ is the polynomial ring in any number m of variables with integral coefficients. Here $\mathrm{E}_n R$ is the group generated by all elementary matices over R. The corresponding result for the symplectic groups has been unknown even for one variable ($m = 1$) in spite of some work done [1, 11, 12]. We prove here that $\mathrm{Sp}_{2n} R = \mathrm{Ep}_{2n} R$ when $R = \mathbb{Z}[x_1, \ldots, x_m]$. By $\mathrm{Ep}_{2n} R$ we denote the group generated by the symplectic elementary matrices [5, 11, 12, 18]. In fact, we obtain a more general result allowing a more general ring of coefficients A instead of the integers $\mathbb{Z}$.

Definition 1.1 We call a (commutative associative) ring A with 1 *locally principal* if for every maximal ideal μ of A the localization A_μ at μ is a principal ideal ring.

For example, every Dedekind ring is locally principal. Another example is $A = \mathbb{Z}/4\mathbb{Z}$. Note that a locally principal ring has Krull dimension $\leqq 1$. From the exercises in [8] it can be seen that a locally principal ring need not be Noetherian.

Our main result is

Theorem 1.2 *Let A be a locally principal ring. For an integer $m \geqq 0$ put $R = A[x_1, \ldots, x_m]$. Then $\mathrm{Sp}_{2n} R = \mathrm{Sp}_{2n} A \cdot \mathrm{Ep}_{2n} R$ for all integers $n \geqq 2$.*

Note that for Euclidean rings or local principal ideal rings A we have: $\mathrm{Sp}_{2n} A = \mathrm{Ep}_{2n} A$ (see Section 4 for more comments).

Using Theorem 1.2 and results of [18, 16], similar conclusions can be easily obtained for other classical simple algebraic groups of relative rank $\geqq 2$ (resp. for the simply connected Chevalley group of type G_2).

As a byproduct of our proof of Theorem 1.2, we also obtain the following result.

Theorem 1.3 *Let A be a locally principal ring. For an integer $m \geqq 0$ put $R = A[x_1, \ldots, x_m]$. Then $\mathrm{SL}_n R = \mathrm{SL}_n A \cdot \mathrm{E}_n R$ for all integers $n \geqq 3$.*

For a regular ring A of Krull dimension $\leqq 1$, Theorem 1.3 is contained in Suslin [17].

We use localization and patching (local global principle, cf. [17, 21]) which allows us to reduce Theorems 1.2 and 1.3 to similar statements for local rings A. By contrast with [17, 21], we need not assume that A is regular or even Noetherian, since we do not use [7]. In the case of a local ring A, following [14, 6, 18], we use elementary manipulations with symplectic symbols

$$\mathrm{Um}_2 R \to \mathrm{Sp}_4 R/\mathrm{Ep}_4 R.$$

Here $\mathrm{Um}_2 R = \{(a, b)\colon a, b \in R, aR + bR = R\}$ stands for the set of unimodular rows of length 2 over R. We wish to mention that in case the number of variables m in Theorems 1.2 or 1.3 is 1 our proofs are particularly elementary.

Concerning the coefficient ring A, our principal interest lay in the case $A = \mathbb{Z}$. Consider a factor ring B of $R = \mathbb{Z}[x]$ and the corresponding group homomorphism

$$\phi\colon \mathrm{SL}_2(\mathbb{Z}[x]) \to \mathrm{SL}_2 B.$$

Theorem 1.2 was used in [10] to describe the image of ϕ. As a result we were able to deduce certain group theoretic properties of $\mathrm{SL}_2(\mathbb{Z}[x])$. For example that $\mathrm{SL}_2(\mathbb{Z}[x])$ has free quotients of any finite rank.

Recall [16, 12] that the symplectic K_1-group $\mathrm{K}_1\mathrm{Sp}(R) = \mathrm{Sp}(R)/\mathrm{Ep}(R)$ is defined for any commutative ring R with 1 as a direct limit of $\mathrm{Sp}_{2n} R/\mathrm{Ep}_{2n} R$ as $n \to \infty$. By [18], $\mathrm{Ep}_{2n} R$ is normal in $\mathrm{Sp}_{2n} R$ for all $n \geqq 2$. As a consequence of Theorem 1.2 we obtain the following:

Corollary 1.4 *For any locally principal ring A, the inclusion $A \to R = A[x_1, \ldots, x_m]$ induces isomorphisms*

$$\mathrm{Sp}_{2n} A/\mathrm{Ep}_{2n} A = \mathrm{Sp}_{2n} R/\mathrm{Ep}_{2n} R$$

for all integers $n \geqq 2$. So in the limit we have $\mathrm{K}_1\mathrm{Sp}(A) = \mathrm{K}_1\mathrm{Sp}(R)$.

The last conclusion of this corollary has been unknown even in the case $A = \mathbb{Z}$, $m = 1$. It seems that if 2 in invertible in A, then the last conclusion in Corollary 1.4 can be obtained from results of Karoubi and Sansuc, whose methods are based on localization sequences in hermitian K-theory. We would like to mention that the corresponding problem for the symplectic K_0-functor (Witt group) is still open if $2A \neq A$.

Theorem 1.3 has the following consequence for $\mathrm{SK}_1(R) = \mathrm{SL}(R)/\mathrm{E}(R)$.

Corollary 1.5 *For any locally principal ring A, the inclusion $A \to R = A[x_1, \ldots, x_m]$ induces isomorphisms*

$$\mathrm{SL}_n A/\mathrm{E}_n A = \mathrm{SL}_n R/\mathrm{E}_n R$$

for all integers $n \geqq 3$. So in the limit we have $\mathrm{SK}_1(A) = \mathrm{SK}_1(R)$

If A is regular, the last conclusion of this corollary is contained in [4]. However, we do not use [7] and so we do not need the regularity of A. Moreover note that in our situation $\mathrm{K}_1(A)$ need not be isomorphic to $\mathrm{K}_1(R)$.

Our method also gives a new proof of the following results of Suslin [17] and Kopeiko [12]. For any field F and any integers $n \geqq 2$, $m \geqq 0$ we have, as a special case of Theorems 1.2 and 1.3 above:

(i) $\mathrm{SL}_{n+1}(F[x_1, \ldots, x_m]) = \mathrm{E}_{n+1}(F[x_1, \ldots, x_m])$,

(ii) $\mathrm{Sp}_{2n}(F[x_1, \ldots, x_m]) = \mathrm{Ep}_{2n}(F[x_1, \ldots, x_m])$.

Note that $\mathrm{SL}_2 R \neq \mathrm{E}_2 R$ for $R = A[x_1, \ldots, x_m]$ and $m \geqq 2$, for any commutative ring A with 1, $A \neq 0$. However for a locally principal ideal ring A we find a set of generators for the group $\mathrm{SL}_2 R = \mathrm{Sp}_2 R$.

Theorem 1.6 *Let A be a local principal ideal ring. Let $R = A[x]$. Then the group $\mathrm{SL}_2 R = \mathrm{Sp}_2 R$ is generated by matrices of the form*

$$(1_2 + XY)(1_2 + YX)^{-1}$$

where X is a 2 by 2 matrix over R, Y is diagonal of the form

$$Y = \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix}$$

with $y \in R$ and $1_2 + XY \in \mathrm{SL}_2 R$.

In Section 6 below we consider Laurent polynomial rings $R' = A[x, x^{-1}]$ and prove the following two theorems.

Theorem 1.7 *Let A be a local principal ideal ring. Set $R' = A[x, x^{-1}]$. Then for all integers $n \geqq 2$:*

(i) $\mathrm{Sp}_{2n} R' = \mathrm{Ep}_{2n} R'$

and

(ii) $\mathrm{SL}_{n+1} R' = \mathrm{E}_{n+1} R'$.

Therefore

$$\mathbf{1} = \mathrm{Sp}_{2n} A / \mathrm{Ep}_{2n} A = \mathrm{K}_1 \mathrm{Sp}(A) = \mathrm{Sp}_{2n} R' / \mathrm{Ep}_{2n} R' = \mathrm{K}_1 \mathrm{Sp}(R')$$

for all integers $n \geqq 2$ and

$$\mathbf{1} = \mathrm{SL}_n A / \mathrm{E}_n A = \mathrm{SK}_1(A) = \mathrm{SL}_n R' / \mathrm{E}_n R' = \mathrm{SK}_1(R')$$

for all integers $n \geqq 3$.

Theorem 1.8 *Let A be a principal ideal domain. Set $R' = A[x, x^{-1}]$. Then $\mathrm{Sp}_{2n} R' = \mathrm{Sp}_{2n} A \cdot \mathrm{Ep}_{2n} R'$ for all integers $n \geqq 2$. Therefore $\mathrm{Sp}_{2n} A / \mathrm{Ep}_{2n} A = \mathrm{K}_1 \mathrm{Sp}(A) = \mathrm{Sp}_{2n} R' / \mathrm{Ep}_{2n} R' = \mathrm{K}_1 \mathrm{Sp}(R')$ for all integers $n \geqq 2$.*

Note on notation. All our rings will be associative. We use letters A, F, R to denote commutative rings with 1. We write e_i for the row vector having a 1 in the i's column and 0 elsewhere. For a ring R and a natural number n we write $\mathrm{M}_n R$ for the ring of n by n matrices with entries from R. The symbol $\mathrm{e}_{i,j}$ stands for the n by n matrix which has a 1 in the i, j position and 0 elsewhere. See above for the definition of $\mathrm{Um}_2 R$. If g, h are elements of a group G we write $[\mathrm{g}, \mathrm{h}] = \mathrm{ghg}^{-1}\mathrm{h}^{-1}$ for their commutator.

2 Symplectic symbols

This section contains some technical results on symplectic (Mennicke) symbols over local rings. The following lemma is an easy result from commutative algebra. We use the convention that $a^0 = 1$ for any element a of a ring R.

Lemma 2.1 *Let R be a local principal ideal ring. Let $\mathfrak{m} = \pi R$ be the maximal ideal of R with some choice of a generator π. Then we have:*

(i) $\bigcap_{i=1}^{\infty} \mathfrak{m}^i = 0$,

(ii) *for any element $r \in R \setminus \{0\}$ there is a unique integer $j \geqq 0$ and a unit u of R such that $r = \pi^j u$.*

Proof. The first part of the lemma is a version of the Krull intersection theorem. For the second part, choose j maximal such that $r \in \mathfrak{m}^j$. Then $r = \pi^j u$, and u has to be a unit since it cannot be in $\mathfrak{m}$. □

Next we consider the following equivalence relation on unimodular rows over a ring R.

Notation 2.2. We denote by $\leftrightarrow_R$ the equivalence relation on $\mathrm{Um}_2 R$ generated by the following operations:

(S 1) $$(a, b) \leftrightarrow_R (a, b + ac),$$

(S 2) $$(a, b) \leftrightarrow_R (a + bc, b),$$

(S 3) $$(1 + ac, b) \leftrightarrow_R (1 + ac, bc^2)$$

for all $(a, b) \in \mathrm{Um}_2 R$ and $c \in R$.

In connection with (S 3), note that if $a, b, c \in R$, then the following two statements are equivalent:

(2.3) $$(1 + ac, b) \in \mathrm{Um}_2 R,$$

(2.4) $$(1 + ac, bc) \in \mathrm{Um}_2 R.$$

Indeed, (2.3) follows from (2.4). For the reverse implication, note that for any p, q in R, $(1 + ac)\, p + bq = 1$ implies that $(1 + ac)(p + qb) - qabc = 1$.

The above equivalence relation occurs if one considers the rows of matrices from $\mathrm{SL}_2 R = \mathrm{Sp}_2 R \subset \mathrm{Sp}_4 R$ up to multiplication by elementary symplectic matrices in $\mathrm{Ep}_4 R \cap \mathrm{Sp}_2 R$. See the next section for explanation.

If u is a unit from R, then we have

(S 4) $$(a, b) \leftrightarrow_R (ua, u^{-1} b)$$

by the following sequence of operations of type (S 1) or (S 2):

$$\begin{aligned}(a, b) &\leftrightarrow_R (a, b + (u-1)\, a) \\ &\leftrightarrow_R (a + b + (u-1)\, a, b + (u-1)\, a) = (b + ua, b - a + ua) \\ &\leftrightarrow_R (b + ua, b - a + ua + (u^{-1} - 1)(b + ua)) = (b + ua, u^{-1} b) \\ &\leftrightarrow_R (ua, u^{-1} b).\end{aligned}$$

We consider now the above equivalence relation for polynomial rings over local rings.

Lemma 2.5 *Let A be a local principal ideal ring. Then for any $(a, b)\in \mathrm{Um}_2(A[x])$ there is a pair $(a', b')\in \mathrm{Um}_2(A[x])$ with $(a, b)\leftrightarrow_{A[x]} (a', b')$ and so that a' is a monic polynomial.*

Proof. Let $\mathfrak{m}=\pi A$ be the maximal ideal of A and $K=A/\mathfrak{m}$ the residue field. We write

$$a=a_0+\ldots+a_n x^n \quad \text{and} \quad b=b_0+\ldots+b_l x^l$$

with $a_i, b_i\in A$. Since (a, b) is an unimodular row, the pair (a_0, b_0) is unimodular over A. So either a_0 or b_0 has to be a unit of A. We may use (S2) and (S4) to insure that $a_0=1$.

Furthermore, by repeated use of (S1) and (S3), we can assume that $l<n+1$ and that $b_0=0$.

We shall proceed by induction on the degree n of a. If $n=0$ then there is nothing to prove.

If $n=1$, then we have; $a=1+a_1 x$ and $b=b_1 x$. If $b_1=0$, then clearly (a, b) is equivalent to $(1, 0)$ and we are finished. Otherwise, by Lemma 2.1, we may write $a_1=\pi^f u_1$ and $b_1=\pi^g u_2$ with integers f, $g\geqq 0$ and units u_1, u_2 of A. If either f or $g=0$, we are finished using (S1)–(S3). Otherwise, we may assume using (S3) that either $g=1$ or $g=0$. In any case a_1 is a multiple of b_1 and we use (S2) to make $a=1$.

Let now $n>1$. In addition, we may assume that $a_n\in\mathfrak{m}$, since otherwise we are finished using (S4). By repeatedly using (S2) and (S3), we arrange that all the b_i are multiples of a_n. It follows that all the coefficients b_i are in $\mathfrak{m}$. Since the reduction of the row (a, b) modulo $\mathfrak{m}$ is still unimodular (over the field K), we conclude that $a_2, \ldots, a_n\in\mathfrak{m}$. If all the b_i now happen to be divisible by an even power of π we may use (S3) to divide through by it. The case $b=0$ creates a situation as already dealt with above. Else we may now assume that either

(2.6) there is a unit amongst the b_i,

or

(2.7) one of the b_i is divisible exactly by π, but there is no unit amongst the b_i.

In case (2.6) we choose k such that k is the largest number so that b_k is a unit. Using (S1) and (S3) with $c=x$, we replace b by

$$b'=b'_0+\ldots+b'_n x^n$$

with $b'_0, \ldots, b'_n\in\mathfrak{m}$ and either $b'_n=0$ and $b'_{n-1}=a_n b_k$ or $b'_n=a_n b_k$. In any case we may reduce the degree of a and keep the property $a_0=1$. Thus, our proof in case (2.6) is completed by induction.

In case (2.7) we choose the largest k such that b_k is divisible exactly by π. Using (S1) and (S3) with $c=x$, we replace b by

$$b'=b'_0+\ldots+b'_n x^n$$

with $b'_0, \ldots, b'_n \in \mathfrak{m}^2$ and either $b'_n = 0$ and $b'_{n-1} = a_n b_k$ or $b'_n = a_n b_k$. In any case we may reduce the degree of a and keep the property $a_0 = 1$. Thus, our proof in case (2.6) is completed by induction.

In case (2.7) we choose the largest k such that b_k is divisible exactly by π. Using (S1) and (S3) with $c = x$, we replace b by

$$b' = b'_0 + \ldots + b'_n x^n$$

with $b'_0, \ldots, b'_n \in \mathfrak{m}^2$ and either $b'_n = 0$ and $b'_{n-1} = a_n b_k$ or $b'_n = a_n b_k$.

If $a_n b_k = 0$, then there is an integer $j \geqq 1$ such that $\pi^j = 0$. Using repeatedly (S3) with $c = \pi$, we may then replace b by 0 and hence complete our proof. If $a_n b_k \neq 0$, we divide (using (S3)) the coefficients of b' by π^2. After this we have either

(2.8) $b'_{n=0}$ and a_n is a multiple of $\pi b'_{n-1}$ in A,

or

(2.9) a_n is a multiple of $\pi b'_n$ in A.

In any case we may reduce the degree of a and keep the property that $a_0 = 1$. This finishes the induction step. □

In the following lemma we do not require that A is a principal ideal ring.

Lemma 2.10 *Let A be a local ring. Then $(a, b) \leftrightarrow_{A[x]} (1, 0)$ for any $(a, b) \in \mathrm{Um}_2(A[x])$ with a monic polynomial a.*

Proof. We write $\mathfrak{m}$ for the maximal ideal of A and $K = A/\mathfrak{m}$ for the residue field. We proceed by induction on the degree n of the monic polynomial a.

If $n = 0$, then $a = 1$ and we are finished using (S1). If $n = 1$, then using (S1) we may assume that $b \in A$. Reduction modulo $\mathfrak{m}$ shows that b has to be a unit of A. So we can use (S2) and (S4) to complete our proof.

Now suppose $n > 1$.

$$a = a_0 + \ldots + a_{n-1} x^{n-1} + a_n x^n \quad \text{and} \quad b = b_0 + \ldots + b_l x^l$$

with $a_i, b_i \in A$ and $a_n = 1$. Performing, if necessary, an addition operation (S1), we may assume that $l < n$ for the degree l of b. Putting $x = 0$, we see that either a_0 or b_0 has to be a unit in A. Adding if necessary b to a, we arrange that a_0 is a unit. Using (S4), we make $a_0 = 1$. Now a_n is a unit of A.

Let k be the largest number so that b_k is a unit in A. Reduction modulo $\mathfrak{m}$ shows that there is such a number k. By (S2) and by the now admissible operations $(a, bx^2) \leftrightarrow_{A[x]} (a, b)$, we can arrange that either $k = n-1$, or $k = n-2$. In the first case we use our operations (S2) and (S4) to make a monic of degree $n-1$, and we are done by induction.

So let $k = n-2$. Thus,

$$b = b_0 + \ldots + b_{n-2} x^{n-2} + b_{n-1} x^{n-1}$$

with $b_{n-1} \in \mathfrak{m}$ and a unit b_{n-2} in A. We use operation (S2) to insure that

$$a_{n-1} = 1 \quad \text{if } b_{n-3} - a_n^{-1} b_{n-1} a_{n-2} \in \mathfrak{m},$$
$$a_{n-1} = 0 \quad \text{if } b_{n-3} - a_n^{-1} b_{n-1} a_{n-2} \text{ is not in } \mathfrak{m}.$$

Note that after these operations we still have $a_0=1$ and a_n is still a unit. By the operation $(a, b)\leftrightarrow_{A[x]}(a, bx^2)$ followed by $(a, b)\leftrightarrow_{A[x]}(a, b-aa_n^{-1}b_{n-1})$, we replace b by

$$b'=b'_0+\ldots+b'_n x^n.$$

We have

$$b'_n=b_{n-2}-a_n^{-1}b_{n-1}a_{n-1}$$

and

$$b'_{n-1}=b_{n-3}-a_n^{-1}b_{n-1}a_{n-2}.$$

Since b_{n-2} was a unit, now so is b'_n. We proceed by the operation

$$(a, b')\leftrightarrow_{A[x]}(a-b'b_n'^{-1}a_n, b')=:(a', b').$$

We write $a'=a'_0+\ldots+a'_{n-1}x^{n-1}+a'_n x^n$ with $a'_i\in A$. Note that $a'_n=0$ and

$$a'_{n-1}=a_{n-1}-b'_{n-1}b_n'^{-1}=a_{n-1}-b_n'^{-1}(b_{n-3}-a_n^{-1}b_{n-1}a_{n-2})a_n.$$

By our choice of a_{n-1}, we are finished now by induction. □

Putting together Lemmas 2.5 and 2.10, we obtain the following

Corollary 2.11 *Let A be a local principal ideal ring. Set $R=A[x]$. Then $(a, b)\leftrightarrow_R(1, 0)$ for every $(a, b)\in \mathrm{Um}_2 R$.*

To prove Theorem 1.3, it suffices to work with a coarser than $\leftrightarrow_R$ equivalence relation $\sim_R$ on $\mathrm{Um}_2 R$. Namely, for any R, let $\sim_R$ be the equivalence relation on $\mathrm{Um}_2 R$ generated by the operations (S1), (S2) and the following operation (instead of (S3)):

(S5) $(1+ac, b)\sim_R(1+ac, bc)$

for all $(a, b)\in\mathrm{Um}_2 R$ and $c\in R$.

Since (S3) can be obtained by applying (S5) twice, the equivalence $\leftrightarrow_R$ implies the equivalence $\sim_R$. So Corollary 2.11 implies the following

Corollary 2.12 *Let A be a local principal ideal ring. Set $R=A[x]$. Then $(a, b)\sim_R(1, 0)$ for every $(a, b)\in\mathrm{Um}_2 R$.*

Note that a direct proof of this statement, using the above line of argument, would be much easier and shorter than the proof of Lemmas 2.5 and 2.10 above.

3 Action of elementary matrices on unimodular rows

In this section we shall analyze the action of the elementary groups $\mathrm{E}_n(A[x_1, \ldots, x_m])$ on the set of unimodular rows $\mathrm{Um}_n(A[x_1, \ldots, x_m])$. We recall the following notation.

Notation 3.1. Let R be a ring and let $n\geqq 1$ be an integer we write

$$\mathrm{Um}_n(R)=\{(a_1, \ldots a_n)\in R^n\text{: the } a_i \text{ generate R as an ideal}\}$$

for the set of unimodular rows over R. The group $\mathrm{SL}_n R$ acts by matrix multiplication on $\mathrm{Um}_n(\mathrm{R})$. As usual if $a \in R$ and if $n \geqq 1$ is an integer we define the following elementary matrix

$$\mathrm{E}_{i,j}(a) = 1_n + a\, e_{i,j}.$$

Let r be an ideal of R then the elementary group E_n r is defined as the subgroup of $\mathrm{SL}_n R$ generated by the $E_{i,j}(a)$, $a \in$ r.

As usual if R is a ring and $S \subset R$ is a multiplicative subset we write $S^{-1}R$ for the ring of fractions. We put $F_S\colon R \to S^{-1}R$ for the canonical homomorphism. F_S induces maps between the various groups $\mathrm{E}_n(R)$ and $\mathrm{E}_n(S^{-1}R)$, $\mathrm{SL}_n(R)$ and $\mathrm{SL}_n(S^{-1}R)$, ... We say that $F_S(a)$ is obtained by localization from $a \in R$.

Theorem 3.2 *Let A be a locally principal ring, and let $m \geqq 0$ and $n \geqq 3$ be integers. Then the group $\mathrm{E}_n(A[x_1, \ldots, x_m])$ acts transitively on the set $\mathrm{Um}_n(A[x_1, \ldots, x_m])$.*

We deduce Theorem 3.2 by the patching technique from the following result of Suslin [17].

Proposition 3.3 *Let A be a noetherian ring of dimension $\leqq 1$, and let $m \geqq 0$ and $n \geqq 3$ be integers. Then the group $\mathrm{E}_n(A[x_1, \ldots, x_m])$ acts transitively on the set $\mathrm{Um}_n(A[x_1, \ldots, x_m])$.*

We need the following fact which also appears in a somewhat different form as Lemma 11 in [20]. We include a proof for convenience.

Lemma 3.4 *Let $n \geqq 3$ be an integer. Let R be a ring and $S \subset R$ a multiplicative subset and g an element of $\mathrm{E}_n(S^{-1}R)$. Then there is an $s \in S$ so that*

$$g F_S(\mathrm{E}_n(s s_0 R))\, g^{-1} \subset F_S(\mathrm{E}_n(s_0 R))$$

for any $s_0 \in S$.

Proof. By induction on the length of an expression for g as a product of elementary matrices we may assume that $g = \mathrm{E}_{i,j}(a)$ for some $a \in S^{-1}R$. We choose $t \in R$ so that $ta \in F_S(R)$. We prove that

$$g \mathrm{E}_{k,l}(b)\, g^{-1} \in F_S(\mathrm{E}_n(s_0 R))$$

for every $b \in s^2 s_0 R$ and any $1 \leqq k, l \leqq n$. This statement is clear from the Steinberg relations (chapter 5 of [6]) if (i, j) is not equal to (l, k). In case $n \geqq 3$ we may choose an m unequal to k, l and write

$$\mathrm{E}_{k,l}(b) = \mathrm{E}_{k,m}(s)\, \mathrm{E}_{m,l}(s s_0 b_0)\, \mathrm{E}_{k,m}(s)^{-1}\, \mathrm{E}_{m,l}(s s_0 b_0)^{-1}$$

where $b = s^2 s_0 b_0$. The result follows as in the previous case. □

The following line of argument was originally devised by the third author to give an elementary proof for the result of Quillen and Suslin on the freeness of projective modules over polynomial rings. See the book of Lam [13] for comments.

Lemma 3.5 *Let A be a ring and $S \subset A$ a multiplicative subset. Let $f = (f_1, \ldots, f_n) \in \mathrm{Um}_n(A[x_1, \ldots, x_m])$ the following two statements are equivalent:*

(i) *there is a $g \in \mathrm{E}_n(S^{-1}A[x_1, \ldots, x_m])$ so that $fg = f(0) = (f_1(0), \ldots, f_n(0))$ over $S^{-1}A[x_1, \ldots, x_m]$,*

(ii) *there is an* $h \in E_n(A[x_1, \ldots, x_m, t_1, \ldots, t_m])$ *and an* $s \in S$ *so that* $f(x_1 + st_1, \ldots, x_m + st_m)\, h = f(x_1, \ldots, x_m)$.

Proof. (ii) $\rightarrow$ (i): Apply the ring homomorphism $A[x_1, \ldots, x_m, t_1, \ldots, t_m] \rightarrow S^{-1}A[x_1, \ldots, x_m]$ which maps x_i to 0 and t_i to $s^{-1}x_i$.
(i) $\rightarrow$ (ii): We choose $g = g(x_1, \ldots, x_m) \in E_n(S^{-1}A[x_1, \ldots, x_m])$ with $fg = f(0)$. We put

$$\alpha_g = \alpha_g(x_1, \ldots, x_m, t_1, \ldots, t_m) = g(x_1 + t_1, \ldots, x_m + t_m)\, g(x_1, \ldots, x_m)^{-1},$$

and find $\alpha_g \in SL_n(S^{-1}A(x_1, \ldots, x_m, t_1, \ldots, t_m])$. The element α_g has by definition the property:

$$f(x_1 + t_1, \ldots, x_m + t_m)\, \alpha_g = f(x_1, \ldots, x_m).$$

The matrix $\alpha_g(x_1, \ldots, x_m, t_1, \ldots, t_m)$ also satisfies:

$$\alpha_g(x_1, \ldots, x_m, 0, \ldots, 0) = 1_n.$$

We shall now show that there is an $s \in S$ so that

$$\alpha_g(x_1, \ldots, x_m, st_1, \ldots, st_m) \in E_n(S^{-1}A[x_1, \ldots, x_m, t_1, \ldots, t_m])$$

and is the localisation of some

$$\beta_g(x_1, \ldots, x_m, t_1, \ldots, t_m) \in E_n(A[x_1, \ldots, x_m, t_1, \ldots, t_m]),$$

having the property $\beta_g(x_1, \ldots, x_m, 0, \ldots, 0) = 0$.

To do this we write

$$g(x_1, \ldots, x_m) = \varepsilon_1(x_1, \ldots, x_m) \ldots \varepsilon_r(x_1, \ldots, x_m)$$

as a product of r elementary matrices $\varepsilon_i(x_1, \ldots, x_m)$. Now we proceed by induction on r. For $r = 1$ the statement is clear. Let now be $r \geqq 2$ and put

$$b(x_1, \ldots, x_m) = \varepsilon_1(x_1, \ldots, x_m) \ldots \varepsilon_{r-1}(x_1, \ldots, x_m).$$

We find

$$\alpha_g = b(x_1 + t_1, \ldots, x_m + t_m)\, \varepsilon_r(x_1 + t_1, \ldots, x_m + t_m)\, \varepsilon_r(x_1, \ldots, x_m)^{-1}\, b(x_1, \ldots, x_m)^{-1}.$$

We choose $s \in S$ so that

$$\varepsilon_r(x_1 + st_1, \ldots, x_m + st_m)\, \varepsilon_r(x_1, \ldots, x_m)^{-1}$$

comes from $A[x_1, \ldots, x_m, t_1, \ldots, t_m]$ and is also 'big' enough to contain the s for b from Lemma 3.4 and the s obtained from the application of the induction hypothesis to b. We have now finished the induction and found the above $\beta_g(x_1, \ldots, x_m, t_1, \ldots, t_m)$. Put

$$F(x_1, \ldots, x_m, t_1, \ldots, t_m) = f(x_1 + st1, \ldots, x_m + st_m)\, \beta_g - f(x_1, \ldots, x_m).$$

F localizes to 0 and has the property $F(x_1, \ldots, x_m, 0, \ldots, 0) = 0$. We conclude that there is an $s_1 \in S$ so that $h(x_1, \ldots, x_m, t_1, \ldots, t_m) = \beta_g(x_1, \ldots, x_m, s_1 t_1, \ldots, s_1 t_m)$ satisfies (ii). □

Lemma 3.6 *Let A be a ring and let $n \geqq 3$ be an integer. Assume that $f=(f_1, \ldots, f_n)$ is in $\mathrm{Um}_n(A[x_1, \ldots, x_m])$. Then the following two statements are equivalent:*

(i) *there is a $g \in \mathrm{E}_n(A[x_1, \ldots, x_m])$ so that $fg=f(0)$,*

(ii) *for every maximal ideal μ of A there is a $g_\mu \in \mathrm{E}_n(A_\mu[x_1, \ldots, x_m])$ such that $fg_\mu=f(0)$.*

Proof. (i) $\rightarrow$ (ii) being obvious we show (ii) $\rightarrow$ (i). The set

$$\mathfrak{B}=\{b \in A: \text{there is } g \in \mathrm{E}_n(A[x_1, \ldots, x_m, t_1, \ldots, t_m]) \text{ so that } f(x_1+b_t, \ldots, x_m+bt_m))\, g=f(x_1, \ldots, x_m)\}$$

is clearly an ideal of A. From Lemma 3.5 it follows that

$$\mathfrak{C}=\{s \in A: \quad \text{there is an } h \in \mathrm{E}_n(S^{-1}A[t]) \text{ so that } fh=f(0) \text{ where } S=\{s^i: i \in \mathbb{Z}\}\}$$

is the radical ideal of $\mathfrak{B}$. Our assumption implies by lemma 3.5 that for every maximal ideal μ of A there is an element $s \in A \setminus \mu$ with $s \in \mathfrak{B}$. This implies $\mathfrak{B}=\mathfrak{C}=A$ and hence the result follows. □

Proof of Theorem 3.2. By Proposition 3.3 the group $E_n(A_\mu[x_1, \ldots, x_m])$ acts transitively on $\mathrm{Um}_n(A_\mu[x_1, \ldots, x_m])$ for every maximal ideal μ. Lemma 3.6 then finishes the argument. □

Whenever $n<m$ we have a monomorphism

$$\psi\colon \mathrm{SL}_n R \rightarrow \mathrm{SL}_m R$$

by putting the matrices from $\mathrm{SL}_n R$ in the upper left hand corner:

$$\psi(\mathrm{SL}_n R)=\begin{pmatrix} SL_n R & 0 \\ 0 & 1_{m-n} \end{pmatrix}.$$

The following obviously follows from Theorem 3.2.

Proposition 3.7 *Let $R=A[x_1, \ldots, x_m]$ where A is a locally principal ring and $m \geqq 0$ is an integer. Then $\mathrm{SL}_n R=\psi(\mathrm{SL}_2 R)$. $\mathrm{E}_n R$ for all integers $n \geqq 3$.*

In case A is a local principal ideal ring and if the number of variables is 1 Proposition 3.7 also follows since the ring $A[x]$ satisfies the second stable range condition ($sr(A[x]) \leqq 2$) in the sense of Bass [6]. This means that given a unimodular triple (a_1, a_2, a_3), there are $r_1, r_2 \in R$ such that the pair

$$(a_1-r_1 a_3, a_2-r_2 a_3)$$

is also unimodular. This follows from the fact that the space of maximal ideals of the ring $R=A[x]$ is the union of two noetherian subsets of Krull dimension 1, hence by [6, p. 173] $\mathrm{sr}(R) \leqq 2$.

We shall now explain how to use the equivalence relation $\sim_R$ and the results from section 2. For $(a, b) \in \mathrm{Um}_2 R$ and an integer $n \geqq 3$ we define

$$\eta(a, b)=\psi(g)\, \mathrm{E}_n R \in \mathrm{SL}_n R/\mathrm{E}_n R$$

where g is any matrix from $\mathrm{SL}_2 R$ having (a, b) as the first row. It is clear that this construction gives a well defined map $\eta\colon \mathrm{Um}_2 R \to \mathrm{SL}_n R/\mathrm{E}_n R$. We quote now a result from [6], chapter 5:

Lemma 3.8 *Let R be a ring and (a, b), $(a', b') \in \mathrm{Um}_2 R$ with $(a, b) \sim_R (a', b')$. Then $\eta(a, b) = \eta(a', b')$.*

From Corollary 2.12, Lemmas 3.7 and 3.8 we infer the following:

Corollary 3.9 *Let A be a local principal ideal ring and $n \geqq 3$ an integer, then:*

$$\mathrm{SL}_n(A[x]) = \mathrm{E}_n(\mathrm{A}[\mathrm{x}]).$$

4 Symplectic groups

First of all, we introduce some notation concerning the symplectic groups. For a natural number n we define

$$J_n = \sum_{i=1}^{n} e_{2i-1,\,2i} - \sum_{i=1}^{n} e_{2i,\,2i-1}$$

to be the standard $2n$ by $2n$ alternating matrix:

$$J_n = \begin{pmatrix} 01\,0 & \dots & 0 \\ -10\,0 & \dots & 0 \\ & \dots & \\ & \dots & \\ & \dots & \\ 0 \dots & 0 & 01 \\ 0 \dots & 0 & -10 \end{pmatrix}$$

The symplectic group over the ring R is defined as

$$\mathrm{Sp}_{2n} R = \{g \in \mathrm{M}_{2n} R : g^T J_n g = J_n\}.$$

Here g^T is the transpose of the matrix g. To define the elementary subgroup we need the following permutation of $\mathbb{Z}$: $(2i)' = 2i-1, (2i-1)' = 2i$. For an element a of a ring R we define the following (elementary symplectic) $2n$ by $2n$ matrix:

$$\mathrm{SE}_{i,j}(a) = \begin{cases} 1_{2n} + a e_{i,j} & \text{if } i = j', \\ 1_{2n} + a e_{i,j} - (-1)^{i+j} a e_{j',i'} & \text{if } i \neq j'. \end{cases}$$

It is easy to see that $\mathrm{SE}_{i,j}(a) \in \mathrm{Sp}_{2n} R$ for any $a \in R$. The elementary symplectic group $\mathrm{Ep}_{2n} R$ is defined as the group generated by all elementary symplectic matrices $\mathrm{SE}_{i,j}(a)$, for all suitable integers i, j. Note that $\mathrm{Ep}_4 R$ is normal in $\mathrm{Sp}_4 R$ [9]. In the proofs of one of the following Lemmas we need the Steinberg relations between the symplectic elementary matrices. These relations are contained in the Lemma 4.1. They can be proved by elementary computations.

Lemma 4.1 *Let R be a ring. The following relations hold for all integers i, j, k, l and all elements a, b of R.*

(1) $[\mathrm{SE}_{i,i'}(a), \mathrm{SE}_{j,j'}(b)] = 1$ *if* $j \neq i'$,

(2)
$$[\mathrm{SE}_{i,i'}(a), \mathrm{SE}_{k,l}(b)] = 1 \quad \text{if } k \neq l,\ k' \neq l,\ i' \neq k,\ i \neq l,$$
$$[\mathrm{SE}_{i,i'}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{i,l}(ab)\,\mathrm{SE}_{l',l}(-(-1)^{k+l}ab^2) \quad \text{if } k \neq l,\ k' \neq l,\ i' = k,\ i \neq l,\ l \neq i',$$
$$[\mathrm{SE}_{i,i'}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{i,i'}(2ab - (-1)^{k+l}ab^2) \quad \text{if } k \neq l,\ k' \neq l,\ i' \neq k,\ i \neq l,\ l = i',$$
$$[\mathrm{SE}_{i,i'}(a), \mathrm{SE}_{k,l}(b)] = SE_{k,i'}(-ab)\,\mathrm{SE}_{k,k'}(-(-1)^{k+l}ab^2) \quad \text{if } k \neq l,\ k' \neq l,\ i' \neq k,\ i = l,\ i \neq k,$$
$$[\mathrm{SE}_{i,i'}(a), \mathrm{SE}_{k,l}(b)] = SE_{k,k'}(-2ab - (-1)^{k+l}ab^2) \quad \text{if } k \neq l,\ k' \neq l,\ k' \neq k,\ i = l,\ i = k,$$

(3) For $i \neq j$, $i \neq j'$, $k \neq l$, $k \neq l'$, the following relations hold.

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = 1$ *if* $j \neq k, j \neq l', i' \neq k, i' \neq l'$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{i,l}(ab)$ *if* $j = k, j \neq l', i' \neq k, i' \neq l', i \neq l'$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{i,i'}(2ab)$ *if* $j = k, j \neq l', i' \neq k, i' \neq l', i = l'$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{k,i'}((-1)^{i+j}ab)$ *if* $j \neq k, j = l', i' \neq k, i' \neq l', i \neq k$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{i,i'}(2(-1)^{i+j}ab)$ *if* $j \neq k, j = l', i' \neq k, i' \neq l', i = k$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{l',j}((-1)^{k+l}ab)$ *if* $j \neq k, j \neq l', i' = k, i' \neq l', l \neq j$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{i,i'}(2(-1)^{k+l}ab)$ *if* $j \neq k, j \neq l', i' = k, i' \neq l', l = j$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{k,j}(-ab)$ *if* $j \neq k, j \neq l', i' \neq k, i' = l', k \neq j'$,

$[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)] = \mathrm{SE}_{k,k'}(-2ab)$ *if* $j \neq k, j \neq l', i' \neq k, i' = l', k = j'$.

Note that the following commutators cannot be computed in a simple way as products of elementary matrices:

(1) $[\mathrm{SE}_{i,i'}(a), \mathrm{SE}_{j,j'}(b)]$ if $j = i'$,
(2) $[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)]$ if $i \neq j, i \neq j', k \neq l, k \neq l, j = k, j \neq l', i' \neq k, i = l$,
(3) $[\mathrm{SE}_{i,j}(a), \mathrm{SE}_{k,l}(b)]$ if $i \neq j, i \neq j', k \neq l, k \neq l, j \neq k, j = l', i' = k, i \neq l$.

The following result makes it possible to to relate orbits of $\mathrm{E}_n(R)$ on $\mathrm{Um}_n(R)$ to orbits under the symplectic elementary group.

Lemma 4.2 *Let R be any ring and let $n \geqq 1$ be an integer. Then*

$$\mathrm{e}_{2n}\,\mathrm{E}_{2n}(R) = \mathrm{e}_{2n}\,\mathrm{Ep}_{2n}(R).$$

Proof. A proof can be found in [21]. □

From Theorem 3.2 we infer:

Proposition 4.3 *Let A be a locally principal ring, and let $m \geqq 0$ and $n \geqq 2$ be integers. Then the group $\mathrm{Ep}_{2n}(A[x_1, \ldots, x_m])$ acts transitively on the set $\mathrm{Um}_{2n}(A[x_1, \ldots, x_m])$.*

From our definition it is clear that whenever $m \geqq n$ we have a monomorphism

$$\psi\colon \mathrm{Sp}_{2n}R \to \mathrm{Sp}_{2m}R$$

by putting the matrices from $\mathrm{Sp}_{2n}R$ in the upper left hand corner. $\psi(\mathrm{Sp}_{2n}R) =$

$$\begin{pmatrix} \mathrm{Sp}_{2n}R & 0 \\ 0 & 1_{2m-2n} \end{pmatrix}.$$

Proposition 4.3 implies:

Lemma 4.4 *Let $R = A[x_1, \ldots, x_m]$ where A is a locally principal ring and $m \geqq 0$ is an integer. Then $\mathrm{Sp}_{2n}R = \psi(\mathrm{SL}_2 R)$. $\mathrm{Ep}_{2n}R$ for all integers $n \geqq 2$.*

Proof. Assume that $n \geqq 2$. We show that for every element $A \in \mathrm{Sp}_{2n}(R)$ there is an $X \in \mathrm{Ep}_{2n}(R)$ with $XA \in \psi(\mathrm{Sp}_{2(n-1)}(R))$. By induction it follows that

$$\mathrm{Sp}_{2n}(R) = \mathrm{Ep}_{2n}(R) \cdot \psi(\mathrm{SL}_2(R)).$$

This statement is equivalent to the lemma.

By Proposition 4.3 we find the element $X_1 \in \mathrm{Ep}_{2n}(R)$ with

$$X_1 A = \begin{pmatrix} B & b & 0 \\ c & d & 0 \\ e & f & 1 \end{pmatrix}$$

where B is a $(2n-1) \times (2n-1)$ matrix with entries in R and b, c, d, e, f are column or row vectors of appropriate sizes. Since $X_1 A \in \mathrm{Sp}_{2n}(R)$ we find that $c=0$ and $d=1$. We also infer that $B \in \mathrm{Sp}_{2n-1}(R)$. It follows from our definitions that

$$X_1 A \cdot \begin{pmatrix} B^{-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{Ep}_{2n}(R).$$

This finishes the inductive argument. □

We shall now explain the relation of the symbols studied in section 2 and the symplectic group. For $(a, b) \in \mathrm{Um}_2 R$ we define

$$\eta(a, b) = \psi(g)\, \mathrm{Ep}_4 R \in \mathrm{Sp}_4 R / \mathrm{Ep}_4 R$$

where g is any matrix from $\mathrm{SL}_2 R$ having (a, b) as the first row. It is clear that this construction gives a well defined map η: $\mathrm{Um}_2 R \to \mathrm{Sp}_4 R / \mathrm{Ep}_4 R$. We quote now a result from [18, Lemma 1.5]:

Lemma 4.5 *Let R be a ring and (a, b), $(a', b') \in \mathrm{Um}_2 R$ with $(a, b) \leftrightarrow_R (a', b')$. Then $\eta(a, b) = \eta(a', b')$.*

Putting Lemma 4.5, Lemma 4.4 and Corollary 2.11 together we get:

Corollary 4.6 *Let A be a local principal ideal ring and $n \geqq 2$ an integer, then:*

$$\mathrm{Sp}_{2n}(A[x]) = \mathrm{Ep}_{2n}(A[x]).$$

Finally we need the following analogue of Lemma 3.4.

Lemma 4.7 *Let $n \geqq 2$ be an integer. Let R be a ring and $S \subset R$ a multiplicative subset and g an element of $\mathrm{Ep}_{2n}(S^{-1}R)$. Then there is an $s \in S$ so that*

$$g F_S(\mathrm{Ep}_{2n}(s s_0 R))\, g^{-1} \subset F_S(\mathrm{Ep}_{2n}(s_0 R))$$

for any $s_0 \in S$.

Proof. The proof follows the same line of argument as in Lemma 3.4. Instead of the Steinberg relations for SL_n we use the relations appearing in Lemma 4.1. □

5 Proofs of the main results

In this section we shall prove the main results mentioned in the introduction. In particular we give proofs of Theorems 1.2 and 1.3. The following lemma from commutative algebra is important in our approach. It appears in the special case $A = \mathbb{Z}$ in [9]. We thank S. Novikov for giving this reference.

Proposition 5.1 *Let A be any principal ideal ring. Let $S_1 \subset R = A[x]$ be the multiplicative set consisting of all monic polynomials. Then $S_1^{-1} R$ is also a principal ideal ring.*

Proof. We consider the polynomial ring $A[x^{-1}]$ and define S_0 to be the multiplicative subset of $A[x^{-1}]$ consisting of all polynomials in x^{-1} with constant coefficient equal to 1. We put $B = S_0^{-1} A[x^{-1}]$. Note that x^{-1} is contained in the Jacobson radical of B since for any $a \in B$ the element $1 - ax^{-1}$ is a unit of B. Clearly B is a subring of $S_1^{-1} A[x]$ and

$$S_1^{-1} A[x] = B[x].$$

Putting x^{-1} equal to 0 gives a ring homomorphism

$$\phi\colon B \to A.$$

Let now $\mathfrak{a}$ be an ideal of $S_1^{-1} A[x]$. Put $\mathfrak{a}_0 = \mathfrak{a} \cap B$. Choose an element $a \in \mathfrak{a}_0$ so that $\phi(a)$ generates $\phi(\mathfrak{a}_0)$. The ideal $\mathfrak{a}_0$ is generated by a and elements from $x^{-1} B$. We have

$$x^{-1} B \cap \mathfrak{a}_0 = x^{-1} \mathfrak{a}_0$$

since $\mathfrak{a}_0$ comes from an ideal in $S_1^{-1} A[x]$. The lemma of Nakayama implies that $\mathfrak{a}_0$ is already generated by a. Since the ring $S_1^{-1} A[x]$ is a localization of B it follows that $\mathfrak{a}$ is generated by a as an ideal of $S_1^{-1} A[x]$. □

The following results show that we may in the proofs of Theorems 1.2 and 1.3 invert all monic polynomials.

Proposition 5.2 *Let A be a ring, f a monic polynomial form $A[x]$ and S the multiplicative subset generated by f. For every integer $n \geqq 3$ and every element $g \in \mathrm{SL}_n(A[x])$ the following statements are equivalent:*

(i) $g \in \mathrm{E}_n(A[x])$,

(ii) $g \in \mathrm{E}_n(S^{-1} A[x])$.

This is corollary 5.7 of [17].

Proposition 5.3 *Let A be a ring, f a monic polynomial from $A[x]$ and S the multiplicative subset generated by f. For every integer $n \geqq 2$ and every element $g \in \mathrm{Sp}_{2n}(A[x])$ the following statements are equivalent:*

(i) $g \in \mathrm{Ep}_{2n}(A[x])$,

(ii) $g \in \mathrm{Ep}_{2n}(S^{-1} A[x])$.

For the proof see [12] Theorem 3.10.

Notation 5.4. For an ideal $\mathfrak{a}$ of the ring R, the elementary group $\mathrm{E}_n \mathfrak{a}$ is defined as the subgroup of $\mathrm{SL}_n R$ generated by all $\mathrm{E}_{i,j}(a)$, $a \in \mathfrak{a}$. We write

$$\mathrm{SL}_n(R, \mathfrak{a}) = \{g \in \mathrm{SL}_n R : g - 1_n \in M_n \mathfrak{a}\},$$
$$\mathrm{Sp}_{2n}(R, \mathfrak{a}) = \{g \in \mathrm{Sp}_{2n} R : g - 1_{2n} \in M_{2n} \mathfrak{a}\}$$

for the principal congruence subgroup for the ideal $\mathfrak{a}$ in $\mathrm{SL}_n R$ or $\mathrm{Sp}_{2n} R$. If A is a ring and $m \geqq 0$ an integer we put

$$\mathfrak{a}(m) \subset A[x_1, \ldots, x_m]$$

For the ideal generated by the variables $x_1, \ldots, x_m$.

The following two results make it possibel to localize the coefficient rings. They are appropriate generalizations of the patching results from [17, 12] to the case of many variables.

Proposition 5.5 *Let A be a ring and $n \geqq 3$ an integer. Then for any $g = g(x_1, \ldots, x_m) \in \mathrm{SL}_n(A[x_1, \ldots, x_m], \mathfrak{a}(m))$ the following statements are equivalent:*

(i) $g \in \mathrm{E}_n(A[x_1, \ldots, x_m])$,

(ii) the image of g in $\mathrm{SL}_n(A_\mu[x_1, \ldots, x_m])$ lies in $\mathrm{E}_n(A_\mu[x_1, \ldots, x_m])$ for every maximal ideal μ of A.

Proof. Follows as Lemma 3.6 does from Lemma 3.4. □

Proposition 5.6 *Let A be a ring and $n \geqq 2$ an integer. Then for any $g = g(x_1, \ldots, x_m) \in \mathrm{Sp}_{2n}(A[x_1, \ldots, x_m], \mathfrak{a}(m))$ the following statements are equivalent:*

(i) $g \in \mathrm{Ep}_{2n}(A[x_1, \ldots, x_m])$,

(ii) the image of g in $\mathrm{Sp}_{2n}(A_\mu[x_1, \ldots, x_m]$ lies in $\mathrm{Ep}_{2n}(A_\mu[x_1, \ldots, x_m])$ for every maximal ideal μ of A.

Proof. To prove that (ii) implies (i) we use the same line of argument as in lemma 3.6. First of all we prove anagolously with lemma 3.5 that for a multiplicative subset S of ring A and any $g = g(x_1, \ldots, x_m) \in \mathrm{Sp}_{2n}(A[x_1, \ldots, x_m], \mathfrak{a}(m))$ the following statements are equivalent:

(a) $g \in \mathrm{Ep}_{2n}(S^{-1}A[x_1, \ldots, x_m])$,

(aa) there is an $h \in \mathrm{Ep}_{2n}(A[x_1, \ldots, x_m, t_1, \ldots, t_m])$ and an $s \in S$ so that $g(x_1 + st_1, \ldots, x_m + st_m)\, h = g(x_1, \ldots, x_m)$.

For the proof we use lemma 4.7. As in the proof of Lemma 3.6 we may finish the argument. □

Proof of Theorem 1.2. We have

$$\mathrm{Sp}_{2n}(A[x_1, \ldots, x_m]) = \mathrm{Sp}_{2n}A \cdot \mathrm{Sp}_{2n}(A[x_1, \ldots, x_m], \mathfrak{a}(m)).$$

To prove our result it suffices to show

$$\mathrm{Sp}_{2n}(A[x_1, \ldots, x_m], \quad \mathfrak{a}(m)) \subset \mathrm{Ep}_{2n}(A[x_1, \ldots, x_m]).$$

By Proposition 5.6 we are finished if we prove this in case A is a local principal ideal ring. From now on we assume that A is a local principal ideal ring. We shall now prove by induction on the number $k \geqq 0$ of variables that for $n \geqq 2$

$$\mathrm{Sp}_{2n}(A[x_1, \ldots, x_k]) = \mathrm{Ep}_{2n}(A[x_1, \ldots, x_k]).$$

For $k = 0, 1$ the statement follows from Lemma 4.4 and corollary 4.6. We proceed with the induction step.

For a ring B and an integer $k \geqq 0$ we define the rings $\mathfrak{W}^k(B)$ inductively in the following way

$$\mathfrak{W}^0(B) = B, \quad \mathfrak{W}^1(B) = B[x_1],$$
$$\mathfrak{W}^{k+1}(B) = S_k^{-1}\, \mathfrak{W}^k(B)[x_{k+1}].$$

Here $S_k \subset \mathfrak{W}^k(B)$ is the multiplicative set of polynomials which are monic in x_k. Note that for $k \geqq 1$ $\mathfrak{W}^k(B)$ is a polynomial ring over a ring of fractions of $\mathfrak{W}^{k-1}(B)$. We furthermore have $A[x_1, \ldots, x_{k+1}] \subset \mathfrak{W}^{k+1}(A)$.

Since A is a local principal ideal ring it follows from proposition 5.1 that for $k \geqq 1$ the rings $\mathfrak{W}^k(A)$ are polynomial rings over principal ideal rings.

By Proposition 5.6 and Corollary 4.6 we have:

$$\mathrm{Sp}_{2n}(\mathfrak{W}^{k+1}(A)) = \mathrm{Sp}_{2n}(S_k^{-1}\,\mathfrak{W}^k(A))\,\mathrm{Ep}_{2n}(\mathfrak{W}^{k+1}(A)).$$

This implies that an element $g \in \mathrm{Sp}_{2n}(A[x_1, \ldots, x_{k+1}])$ can be written as $g = g_0\, g_1$ with $g_0 \in \mathrm{Sp}_{2n}(S_k^{-1}\,\mathfrak{W}^k(A))$ and $g_1 \in \mathrm{Ep}_{2n}(\mathfrak{W}^{k+1}(A))$. Let $g_2 \in \mathrm{Sp}_{2n}(S_k^{-1}\,\mathfrak{W}^k(A))$ be the element obtained by putting the variable x_{k+1} in g_1 equal to 0. After multiplying g_1 with an element from $\mathrm{Ep}_{2n}(S_k^{-1}\,\mathfrak{W}^k(A))$ we may assume that $g_2 = 1$. After this we must have $g_0 \in \mathrm{Sp}_{2n}(A[x_1, \ldots, x_k])$. By induction it follows that

$$\mathrm{Sp}_{2n}(A[x_1, \ldots, x_{k+1}]) \subset \mathrm{Ep}_{2n}(\mathfrak{W}^{k+1}(A))$$

for all $k \geqq 0$.

In Proposition 5.3 the multiplicative set S can be taken to be all monic polynomials in $A[x]$, since each matrix in $\mathrm{Ep}_{2n}(S^{-1}A[x])$ is a product of finitely many elementary symplectic matrices, each entry of which has a denominator involving only finitely many elements of S. Now use proposition 5.3 k times with the order of variables reversed to conclude the result. □

Proof of Theorem 1.3. The proof of theorem follows the same line of argument as in the proof of Theorem 1.2. □

Proof of Theorem 1.6. We define H to be the subgroup of $\mathrm{SL}_2 R$ generated by the matrices of the form

$$(1_2 + XY)(1_2 + YX)^{-1}$$

where X is a 2 by 2 matrix with entries in R, Y is a diagonal matrix with

$$Y = \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix}$$

and $1_2 + XY \in \mathrm{SL}_2 R$. First of all we establish

$$\mathrm{E}_2(R) \subset H. \tag{5.7}$$

In fact if g is an elementary matrix we find $g = (1_2 + XY)(1_2 + YX)^{-1}$ where $X = 1_2 - g$ and $y = 2$ or $y = -2$. Notice that it follows that

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in H,$$

and hence every matrix from $\mathrm{SL}_2 R$ with a 1 in some entry is in H.

Secondly we have

$$[\mathrm{E}_2(R), \mathrm{SL}_2 R] \subset H. \tag{5.8}$$

If $g = E_{1,2}(t)$ and

$$h = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then the commutator of g and h is equal to

$$\begin{pmatrix} 1+tac+t^2c^2 & t-ta^2-t^2ac \\ tc^2 & 1-tac \end{pmatrix}$$

This matrix is easily found to be equal to the product of two matrices of the type $(1_2+XY)(1_2+YX)^{-1}$ and a matrix which has a 1 in the lower lefthand corner. The argument for $g = E_{2,1}(t)$ goes similarly.

Next define a map

$$\eta\colon \mathrm{Um}_2 R \to \mathrm{SL}_2 R/H$$

by putting $\eta(a, b)$ equal to the residue class of some matrix in $\mathrm{SL}_2 R$ having (a, b) as first row. It follows from (5.7) and (5.8) that η is well defined and that rows which are equivalent under the moves (S1) and (S2) of section 2 have the same image under η.

It is also a simple matter to find a matrix $f=(1_2+XY)(1_2+YX)^{-1}$ with $\eta(1+ac, b)\, f = \eta(1+ac, bc)$. Hence

(5.9) if $(a, b), (a', b') \in \mathrm{Um}_2 R$ satisfy $(a, b) \sim_R (a', b')$ then $\eta(a, b) = \eta(a', b')$.

Of course (5.9) is true for any ring R. Using Corollary 2.12 we conclude Theorem 1.6. □

6 Laurent polynomial rings

The following lemmas will be needed in the proofs of our results for Laurent polynomial rings.

Lemma 6.1 *Let A be a local principal ideal ring. For any $(a, b) \in \mathrm{Um}_2(A[x, x^{-1}])$ there is either a pair $(a', b') \in \mathrm{Um}_2(A[x])$ or $(a', b') \in \mathrm{Um}_2(A[x^{-1}])$ with $(a, b) \leftrightarrow_{A[x,x^{-1}]} (a', b')$ and so that the constant coefficient of a' is a unit of A.*

Proof. We write $\mathfrak{m}$ for the maximal ideal of A and $K = A/\mathfrak{m}$ for the residueclass field. We furthermore choose a generator π of $\mathfrak{m}$.

Since x is a unit in $A[x, x^{-1}]$, we may write

$$(a, b) = (1 - x^{-1}x(1-a), b)$$

and find by (S3) that

(6.2) $$(a, b) \leftrightarrow_{A[x,x^{-1}]} (a, bx^{-2})$$

and

$$(a, b) \leftrightarrow_{A[x,x^{-1}]} (a, bx^2).$$

In addition, $(a, b) \leftrightarrow_R (-b, a)$ for any ring R and any $(a, b) \in \mathrm{Um}_2 R$ (by applying operations (S1) and (S2)).

Using this as well as (S4) and (6.2), we may replace our initial unimodular row (a, b) by (a, b) with

$$a = a_0 + \ldots + a_n x^n \quad \text{and} \quad b = b_0 + \ldots + b_l x^l \tag{6.3}$$

with $a_i, b_i \in A$.

We may furthermore assume that

(6.4) $Aa_0 \neq A$ and the ideal Aa_0 is the largest possible in all expressions of the form (6.3) equivalent to (a, b).

We now proceed by induction on the degree n of an element a satisfying (6.4). We start off by arranging some further properties of our pair (a, b).

It follows from (6.4) that a_0 divides b_0, i.e., $b_0 = a_0 s$ for some $s \in A$. Using (S1), we replace b by $b - sa$. Now $b_0 = 0$. The element a_0 also divides b_1, because otherwise we may use (S2) and replace a by $a - bx^{-1}$ which would have a constant coefficient strictly dividing a_0. Adding a multiple of a to b, we arrange that both b_0 and b_1 are 0. We now use (6.2) to divide b by x^2. Repeating this argument, we see that we can assume the following:

(6.5) the degree of b is not more than the degree of a, i.e., $l < n + 1$.

Starting again, we have as above $b_0 = a_0 s$ and $b_1 - sa_1 = a_0 t$ for some s, $t \in A$. Moreover

$$(a, b) \leftrightarrow_{A[x, x^{-1}]} (a, (b - sa - txa) x^{-2}) \tag{6.6}$$

Repeating the move (6.6) n times, we see that we may assume that

(6.7) $b_0 = 0$ and each b_i is contained in the ideal generated by the $a_0, \ldots, a_n$.

It follows by reduction modulo $\mathfrak{m}$ that at least one a_i is a unit.

We start the induction now. If $n = 0$, then we have already finished.

If $n = 1$, then

$$a = a_0 + a_1 x \quad \text{and} \quad b = b_1 x.$$

If a_0 is a unit, we are finished. Otherwise, the above a_1 is a unit. We use (S4) together with a move (6.2) and obtain that

$$(a, b) \leftrightarrow_{A[x, x^{-1}]} (a_1 + a_0 x^{-1}, b_1)$$

which proves our claim.

Let now $n > 1$. We have

$$a = a_0 + \ldots + a_n x^n \quad \text{and} \quad b = b_0 + \ldots + b_n x^n.$$

Since $b_0 = 0$,

$$(a, b) \leftrightarrow_{A[x, x^{-1}]} (a, (b - txa) x^{-2} - sa) = (a, b') \tag{6.8}$$

for suitable $s, t \in A$ satisfying $b_1 = a_0 t$ and $b_2 - ta_1 = a_0 s$ with

$$b' = b'_1 x + \ldots + (-sa_{n-1} - ta_n) x^{n-1} - sa_n x^n \in A[x].$$

If s or t is a unit, then using (S2) we may replace a by a polynomial of lower degree but with the same constant coefficient, and we are finished by induction. Repeating this argument n times, we may insure that

(6.9) $$b_0 = 0 \text{ and each } b_i \text{ lies in } \mathfrak{m}.$$

Reducing a, b modulo $\mathfrak{m}$, we see that the reduction of a is a unit in $K[x, x^{-1}]$. It follows that exactly one a_i is a unit. If a_n is this unit, we may finish as above by switching to polynomials in x^{-1}. So we may assume

(6.10) exactly one a_i, say a_g is a unit in A, where $0 < g < n$.

The following sequence of operations divides all b_i by π^2 if this is possible in A:

(6.11) (S4) with $u = a_g^{-1} x^{-g}$; (S3) with $c = \pi$; (S4) with $u = a_g x^g$.

Disposing first of the case $b = 0$ (then a is a unit in A and we use (S4)), we may assume that

(6.12) $b_0 = 0$ and at least one of the coefficients $b_1, \ldots, b_n$ is a unit in A or is exactly divisible by π.

Let f be the smallest index for b_i with this property. We may assume $f > 1$ and that b_n is divisible by π^2. The last assumption can be made since otherwise a_n is a multiple of b_n and we may lower the degree of a without changing its constant coefficient. If $f = 1$, then $b_1 = t a_0$ with a unit t of A. As before we are finished since we may decrease the degree of a without changing its constant coefficient.

This argument has also finished the case $n = 2$, because in this case there is no pair $a = a_0 + a_1 x + a_2 x^2$, $b = b_1 x + b_2 x^2$ satisfying (6.12).

Let now $n > 2$. We consider first the case when b_f is a unit. If $f = 2$, we write $b_1 = t a_0$ with an element t of A. If t is a unit we finish as before. If not, the constant coefficient of $(b - ta)\, x^{-2}$ which is $b_2 - t a_1$ is a unit and we have the result. If $f > 2$, we carry out move (6.7) and obtain

$$(a, b) \leftrightarrow_{A[x, x^{-1}]} (a, b') = (a, b'_1 x + \ldots + b'_n x^n)$$

with $b'_{f-2} = b_f - t a_{f-1} - s a_{f-2}$. If s or t is a unit we finish by reducing the degree of a. If both are in $\mathfrak{m}$ then b'_{f-2} is a unit and we have proved the result by induction.

We now consider the set L of all $d \in A[x]$ such that

$$(a, d) \leftrightarrow_{A[x, x^{-1}]} (a, b) \qquad d = d_1 x + \ldots + d_n x^n$$

with $d_i \in A$, no d_i is a unit, at least one d_i is exactly divisible by π, d_n is divisible by π^2.

For $d \in L$ we define $f(d)$ as the least i such that d_i is exactly divisible by π. We have already shown that we may assume that L is not empty and that $f(d) > 1$ for all d in L. For $b \in L$ we write out the coefficients of b' after move (6.8):

(6.13) $$b'_{r-2} = b_r - t a_{r-1} - s a_{r-2} \quad \text{for } r = 3 \text{ to } n,$$
$$b'_{n-1} = -t a_n - s a_{n-1}, \quad \text{and} \quad b'_n = -s a_n.$$

We first consider the case that $b \in L$ but b' is no longer in L. We may assume that $t, s \in \mathfrak{m}$. So all $b_i' \in \mathfrak{m}^2$. This can only happen if s or t is exactly divisible by π. If s has this property then we may by (6.10) divide by π^2 and after that lower the degree of a. Note that a_n is a multiple of sa_n/π^2. If $s \in \mathfrak{m}^2$, then already $(b - txa)\, x^{-2}$ has all its coefficients in $\mathfrak{m}^2$. We now use (6.11) to lower the degree of a.

In all further moves (6.7) we may assume that $b' \in L$. We now consider the case when $g = n-1$. Recall that g was the unique natural number so that a_g is a unit. Now $f(b) < g+1$ for any $b \in L$. If $b' \in L$, then $f(b') < f(b)$ by a formulas (6.13). So we may assume that $f(b) = 2$. Then s has to be a unit unless $g = 1$ and hence $n = 2$, a case which we have already considered.

From now on we assume that $a_{n-1} \in \mathfrak{m}$, hence $b_n', b_{n-1}' \in \mathfrak{m}^2$. Applying move (2.3) a few times, we may now arrange that

$$b_i \in \mathfrak{m}^2 \quad \text{for } g < i < n. \tag{6.17}$$

If there is an element $b \in L$ satisfying (6.13) and so that $f(b) = 2$, then we set $b_1 = ta_0$ with an element t of A which may be assumed to be in $\mathfrak{m}$. If the constant coefficient of $(b - ta)\, x^{-2}$, which is $b_2 - ta_1$, is exactly divisible by π, then $b_2 - ta_1 = sa_0$ with a unit s of A and the degree of a may be lowered. The element $b_2 - ta_1$ cannot be divisible by π^2 since then $g = 1$ contradicts (6.14).

But if $b \in L$ satisfies (6.14) and if after the move (6.8) we have $b' \in L$, then $f(b) > f(b')$. So after sufficiently many moves (6.7) we end up with an element $b \in L$ satifying (6.14) with $f(b) = 1$ or 2. These cases have been considered before. □

Lemma 6.15 *Let A be a local ring. Set $R = A[x]$ and $R' = A[x, x^{-1}]$. Let $a = a_0 + \ldots + a_n x^n$, $b = b_0 + \ldots + b_l x^l \in R$. Assume that*

(i) $(a_0, b_0) \in \mathrm{Um}_2 A$, *(ii)* $(a, b) \in \mathrm{Um}_2 R'$.

Then $(a, b) \in \mathrm{Um}_2 R$.

Proof. From hypothesis (i) it follows that we may without loss of generality assume that $a_0 = 1$ and $b_0 = 0$. Hypothesis (ii) shows that there are $p, q \in R$ such that

$$pa + qb = x^k \quad \text{for some } k \geqq 0. \tag{6.16}$$

We proceed by induction on k. If $k = 0$, then we are finished. Let now $k > 0$.

Comparing constant coefficients in equation (6.16) we see that we can find $a', b', p' \in R = A[x]$ such that

$$a = 1 + a'x, \quad b = b'x, \quad p = p'x.$$

The pair (a, b') has a unimodular pair of constant coefficients and satisfies the equality $p'a + qb' = x^{k-1}$. By induction $(a, b') \in \mathrm{Um}_2 R$. By (2.3), (2.4) we infer $(a, b) \in \mathrm{Um}_2 R$. □

The above lemmas give the possibility to reduce problems about symplectic symbols from the Laurent polynomial ring to the ordinary polynomial ring. Of course the same applies to our equivalence relation $\sim_R$. The following proposition follows formally from the above. A shorter proof can be obtained by applying the above proof directly to this equivalence relation.

Proposition 6.17 *Let A be a local principal ideal ring and put $R'=A[x, x^{-1}]$. Then*

$$(a, b) \sim_{R'} (1, 0) \quad \text{and} \quad (a, b) \leftrightarrow_{R'} (1, 0)$$

for every $(a, b) \in \mathrm{Um}_2 R'$.

Proof of Theorem 1.7. The space of maximal ideals of the ring $R'=A[x, x^{-1}]$ is the union of two noetherian subsets of Krull dimension 1. This implies [4, p. 173] that $\mathrm{sr}(R') \leqq 2$. Now as in Section 3 we use the appropriate stability results from [19] to see that

$$\psi(\mathrm{SL}_2 R') \cdot \mathrm{E}_{n+1} R' = \mathrm{SL}_{n-1} R'$$

and

$$\psi(\mathrm{SL}_2 R') \cdot \mathrm{Ep}_{2n} R' = \mathrm{Sp}_{2n} R'$$

for $n \geqq 2$.

Proposition 6.17 finishes the proof. □

To prove Theorem 1.8, we need the following:

Proposition 6.18 *Let A be a principal ideal domain. Then for all integers $n \geqq 1$ the group $\mathrm{Sp}_{2n}(A[x, x^{-1}])$ is generated by $\mathrm{Sp}_{2n}(A[x])$ and $\mathrm{Ep}_{2n}(A[x, x^{-1}])$.*

Proof. Consider the group $\mathrm{Sp}_2(A[x, x^{-1}]) = \mathrm{SL}_2(A(x, x^{-1}])$ as acting on the tree for $\mathrm{SL}_2(K[[x]])$ as described in [15]. Our result follows for $n=1$ from Yharas theorem [15], [3]. The ring $R'=A[x, x^{-1}]$ satisfies $\mathrm{sr}(R') \leqq 3$. As in the proof of Theorem 1.7 we may finish by using the stability result from [14]. □

Proof of Theorem 1.8. The previous Proposition reduces us to the case of a polynomial ring. We then use Theorem 1.2. □

Acknowledgements. The authors thank the Max Planck Institute for Mathematics at Bonn for supporting their work. We also thank J.P. Serre for his useful comments.

References

1. Abe, E.: Whitehead groups of Chevalley groups over polynomial rings. Commun. Algebra **11**, 1271–1307 (1983)
2. Atiyah, M., Macdonald, I.G.: Introduction to commutative algebra. Reading, M.A.: Addison-Wesley 1969
3. Bachmuth, S., Mochizuki, H.Z.: IA-Automorphisms of the free metabelian group of rank 3. J. Algebra **55**, 106–115 (1978)
4. Bass, H.: K-theory and stable algebra. Publ. Math., Inst. Hautes Étud. Sci. **22**, 485–544 (1964)
5. Bass, H., Milnor, J., Serre, J.-P.: Solution of the congruence subgroup problem for $\mathrm{SL}_n A (n \geqq 3)$ and $\mathrm{Sp}_{2n} A (n \geqq 2)$. Publ. Math. Inst. Hautes Étud. Sci. **33**, 59–137 (1967)
Erratum: On a functorial property of power residue symbols, Publ. Math. Inst. Hautes Étud. Sci. **44**, 241–244 (1974)
6. Bass, H.: Algebraic K-theory. Benjamin, New York, 1968
7. Bass, H., Heller, A., Swan, R.G.: The Whitehead group of a polynomial extension, Publ. Math. Hautes Étud. Sci. **22**, 61–79 (1964)
8. Bourbaki, N.: Algebre commutative. Chap. 2, Localisation. Paris: Herman 1961
9. Farber, M.Sh.: Exactness of the Novikov inequalities. Funk. An. **19**, 40–48 (1985)

10. Grunewald, F., Mennicke, J., Vaserstein, L.: On the groups $SL_2(\mathbb{Z}[x])$ and $SL_2(K[x, y])$. Preprint
11. Karoubi, M.: Périodicité de la K-théorie hermitienne. (Lecture Notes Math. vol. 343, 301–411) Berlin Heidelberg New York: Springer 1973
12. Kopeiko, V.I.: Stabilization of symplectic groups over polynomial rings. Math. Sbornik **106**, 94–107 (1978)
13. Lam, T.Y.: Serre's problem. (Lecture Notes Math. vol. 635) Berlin Heidelberg New York: Springer 1978
14. Mennicke, J.L.: Zur Theorie der Siegelschen Modulgruppe. Math. Ann. **159**, 115–129 (1965)
15. Serre, J.-P.: Trees. Berlin Heidelberg New York: Springer 1980
16. Stein, M.: Stability theorems for K_1, K_2, and related functors modeled on Chevalley groups. Japan J. Math., New Ser. **4**, 77–108 (1978)
17. Suslin, A.: On the structure of the special linear group over polynomial rings. Izv. Akad. Nauk SSSR, Ser. Mat. **41**, 235–252 (1977) (Translated as: Math. USSR, Izv. **11**, 221–238 (1977)
18. Vaserstein, L.N.: Stabilization for unitary and orthogonal groups over a ring with involution. Mat. Sbornik **81**, 328–351 (1970) (Translated as: Math. USSR, Sb. **10**, 307–326 (1970)
19. Vaserstein, L.N.: Stabilization for classical groups over rings. Mat. Sbornik **93**, 268–295 (1974) (Translated as: Math. USSR, Sb. **22**, 271–303 (1974)
20. Vaserstein, L.N.: On the normal subgroups of GL_n over a ring. (Lecture Notes Math. vol. 854 pp. 454–465 Berlin Heidelberg New York: Springer 1981
21. Vaserstein, L.N., Suslin, A.A.: Serre's problem on projective modules over polynomial rings and algebraic K-theory. Izv. Akad. Nauk SSSR Ser. Mat. **40**, 993–1054 (1976) (Translated as Math. USSR, Izv. **5**, 931–1001 (1971)

Math. Z. 206, 57–66 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Symplectic K_2 of Laurent polynomials, associated Kac-Moody groups and Witt rings

Jun Morita *[1] **and Ulf Rehmann** [2]

[1] Institute of Mathematics, University of Tsukuba, Tsukuba, Japan
[2] Fakultät für Mathematik, Universität Bielefeld, W-4800 Bielefeld, Federal Republic of Germany

Received May 17, 1989; in final form November 6, 1989

0 Introduction

In this note, we will show the following three theorems. Here F denotes an arbitrary field, F^* the multiplicative group of F, and $F[\xi, \xi^{-1}]$ the ring of Laurent polynomials in ξ with coefficients in F.

Let $W(F)$ be the Witt ring over F with the maximal ideal $I(F)$ consisting of classes with even rank (cf. [11]). Then there are two exact sequences as follows (cf. Suslin [20]):

$$1 \to I^3(F) \to K_2 \operatorname{Sp}(F) \to K_2(F) \to 1,$$
$$1 \to N(F) \to K_2(F) \to K_2 \operatorname{Sp}(F) \to I^2(F) \to 1,$$

where $N(F) = \langle \{u, -1\} | u \in F^* \rangle \subseteq K_2(F)$ and $N(F)$ is the 2-torsion part of $K_2(F)$ (cf. [20]). We will study an $F[\xi, \xi^{-1}]$-version. As is well-known, the tame symbol induces $K_2(F[\xi, \xi^{-1}]) \simeq K_2(F) \oplus F^*$. Then, we obtain the following.

Theorem A *There are the following exact sequences*:

$$1 \to I^3(F) \oplus I^2(F) \to K_2 \operatorname{Sp}(F[\xi, \xi^{-1}]) \to K_2(F[\xi, \xi^{-1}]) \to 1,$$
$$1 \to N(F) \oplus \{\pm 1 \in F\} \to K_2(F[\xi, \xi^{-1}]) \to K_2 \operatorname{Sp}(F[\xi, \xi^{-1}]) \to I^2(F) \oplus I(F) \to 1.$$

Remark. The following description might show how the two exact sequences are interrelated:

* The first author was partially supported by the Alexander von Humboldt-foundation in 1988–89. The authors want to thank Professor A. Bak for his valuable advice on $K_2 \operatorname{Sp}(F)$, Witt rings and their relation.

Let $Br_2(F)$ be the 2-torsion part of the Brauer group $Br(F)$. Then, in char $F \neq 2$, we obtain the next diagram of exact sequences.

$$\begin{array}{ccccccccc}
 & & & & 1 & & 1 & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & 1 & \longrightarrow & \dfrac{K_2 F[\xi, \xi^{-1}]}{N(F) \oplus \{\pm 1\}} & \xrightarrow{2} & 2K_2 F[\xi, \xi^{-1}] & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & I^3(F) \oplus I^2(F) & \longrightarrow & K_2 \operatorname{Sp} F[\xi, \xi^{-1}] & \xrightarrow{\text{nat}} & K_2 F[\xi, \xi^{-1}] & \longrightarrow & 1 \\
 & & \| & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & I^3(F) \oplus I^2(F) & \xrightarrow{\text{nat}} & I^2(F) \oplus I(F) & \longrightarrow & Br_2 F \oplus F^*/F^{*2} & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 1 & & 1 & & 1 & &
\end{array}$$

Note that $2K_2 F[\xi, \xi^{-1}] \simeq 2K_2 F \oplus F^{*2}$, and $N(F) \oplus \{\pm 1\}$ is the 2-torsion part of $K_2 F[\xi, \xi^{-1}]$. Here τ is the map induced by $\{u, v\} \mapsto \{u^2, v\} = \{u, v^2\}$, which is equal to the so-called hyperbolic map (see Section 6). One should also notice that every object in this diagram splits into two components, and the situation over a field F is well-known (cf. [10, 20]).

We also obtain the group structure and the stability of

$$K_2 \operatorname{Sp}_{2n}(F[\xi, \xi^{-1}]) = K_2(C_n, F[\xi, \xi^{-1}])$$

as follows. Let $P(F)$ be the set $F^* \times I^2(F)$ with the product operation

$$(u, y) \cdot (v, z) = (uv, y + z - \langle\langle u, v \rangle\rangle)$$

for all (u, y), $(v, z) \in P(F)$, which makes $P(F)$ into a group (cf. [20]), in other words, $P(F)$ is a central extension of F^* with the so-called Pfister cocycle $\langle\langle \cdot, \cdot \rangle\rangle$:

$$1 \to I^2(F) \to P(F) \to F^* \to 1.$$

Then, we obtain the next result.

Theorem B $K_2 \operatorname{Sp}_{2n}(F[\xi, \xi^{-1}]) \simeq K_2 \operatorname{Sp}(F) \oplus P(F)$ *for all* $n \geqq 1$.

To prove these two theorems, we will study the associated Kac-Moody groups. Let $K_2(\tilde{C}_n, F)$ be the K_2-group associated with a (universal) Kac-Moody group of type $\tilde{C}_n$ (cf. [21]). Then we obtain the following.

Theorem C $K_2(\tilde{C}_n, F) \simeq K_2 \operatorname{Sp}(F) \oplus I^2(F)$ *for all* $n \geqq 1$.

As a direct consequence of our discussion, we can obtain some characterization of $W(F)$ and $I^r(F)$ for all $r \geqq 1$ (cf. Section 6).

Contents

1. Preliminary: $\mathrm{St}(X, R)$
2. Relations between $K_2(\tilde{X}_n, F)$ and $K_2(X_n, F[\xi, \xi^{-1}])$
3. Structure theorem for $K_2(\tilde{C}_n, F)$
4. Structure theorem for $K_2(C_n, F[\xi, \xi^{-1}])$
5. Relations between $K_2\,\mathrm{Sp}(F[\xi, \xi^{-1}])$ and $K_2(F[\xi, \xi^{-1}])$
6. Remarks

For elements x, y of a group, $[x, y]=xyx^{-1}y^{-1}$ denotes the commutator of x and y. For groups G_1, G_2 such that G_1 acts on G_2, we let $G_1 \ltimes G_2$ denote the semidirect product of G_1 and G_2. We always use 1 as a trivial group. In a group, the symbol $\langle \ldots \rangle$ means the subgroup generated by ....

1 Preliminary: $\mathrm{St}(X, R)$

Let $\mathbf{g}$ be a Kac-Moody algebra over $\mathbb{C}$ (=the field of complex numbers), Δ the root system of $\mathbf{g}$, and X the Dynkin diagram of $\mathbf{g}$ (cf. [7, 8, 12]). Using a Chevalley basis $\{e_\alpha | \alpha \in \Delta^{\mathrm{re}}\}$ for Δ^{re} (cf. [13, 22]), we can define the Steinberg group $\mathrm{St}(X, R)$ over a commutative ring R (with 1), where Δ^{re} is the set of real roots. That is, $\mathrm{St}(X, R)$ is the group generated by $x_\alpha(t)$ for all $\alpha \in \Delta^{\mathrm{re}}$ and $t \in R$ with the defining relations:

$$x_\alpha(s)\, x_\alpha(t) = x_\alpha(s+t);$$

$$[x_\alpha(s), x_\beta(t)] = \prod x_{i\alpha+j\beta}(N_{\alpha\beta ij}\, s^i t^j)$$

if $(\mathbb{Z}_{>0}\,\alpha + \mathbb{Z}_{>0}\,\beta) \cap \Delta \subseteq \Delta^{\mathrm{re}}$;

$$w_\alpha(u)\, x_\beta(t)\, w_\alpha(-u) = x_{\beta'}(C_{\alpha\beta}\, u^{-\beta(h_\alpha)} t)$$

for $\alpha, \beta \in \Delta^{\mathrm{re}}$, $s, t \in R$, $u \in R^*$ (=the group of units), where the product at the right hand side in the second relation is taken over all roots of the form $i\alpha + j\beta \in \Delta^{\mathrm{re}}$ with $i, j \geqq 1$, the $N_{\alpha\beta ij}$ are certain integers depending only on the structure of $\mathbf{g}$ (cf. [13]), $w_\alpha(u) = x_\alpha(u) x_\alpha(-u^{-1}) x_\alpha(u)$, $\beta' = \beta - \beta(h_\alpha)\alpha$, h_α is the coroot of α, and $C_{\alpha\beta} = \pm 1$ is given by

$$(\exp \mathrm{ad}\, e_\alpha)(\exp -\mathrm{ad}\, e_{-\alpha})(\exp \mathrm{ad}\, e_\alpha)\, e_\beta = C_{\alpha\beta}\, e_{\beta'}.$$

Let T be the subgroup of $\mathrm{St}(X, R)$ generated by $h_\alpha(u)$ for all $\alpha \in \Delta^{\mathrm{re}}$ and $u \in R^*$, and $C(X, R)$ the subgroup of T generated by $\{u, v\}_\alpha$ for all $u, v \in R^*$ and $\alpha \in \Delta^{\mathrm{re}}$, where $h_\alpha(u) = w_\alpha(u)\, w_\alpha(-1)$ and

$$\{u, v\}_\alpha = h_\alpha(uv)\, h_\alpha(u)^{-1} h_\alpha(v)^{-1}.$$

Let U be the subgroup of $\mathrm{St}(X, R)$ generated by $x_\alpha(t)$ for all $\alpha \in \Delta^{\mathrm{re}}_+$ and $t \in R$, where Δ^{re}_+ is the set of positive real roots.

If $X = X_n$ is of finite type (one of A_n, B_n, C_n, ..., G_2, assumed to be simple), then

$$K_2(X_n, F) = C(X_n, F) \quad \text{and} \quad K_2(X_n, F[\xi, \xi^{-1}]) = C(X_n, F[\xi, \xi^{-1}])$$

(cf. [1, 6, 9, 17, 18]). Put $K_2(\cdot)=\lim K_2(A_n,\cdot)$ and $K_2\,\mathrm{Sp}(\cdot)=\lim K_2(C_n,\cdot)$. Matsumoto [9] gives a presentation of $K_2(X_n, F)$ and shows

$$K_2(X_n,F)\simeq\begin{cases}K_2\,\mathrm{Sp}(F) & \text{if } X_n=C_n(n\geqq 1);\\ K_2(F) & \text{otherwise.}\end{cases}$$

In case of $F[\xi,\xi^{-1}]$, we also have the tame splitting (for this and the following, see [16], Theor. 11 and [6], Kor. 4):

$$K_2(X_n,F[\xi,\xi^{-1}])\simeq K_2(F)\oplus F^* \quad \text{if } X_n\neq C_n(n\geqq 1).$$

We will study the group structure of $K_2(C_n, F[\xi,\xi^{-1}])$ in Section 4. Sometimes we also write $K_2\,\mathrm{Sp}_{2n}(\cdot)=K_2(C_n,\cdot)$. Now let $b\in\Delta$ be the negative highest root. Let $L(C_n, F)$ be the subgroup of $K_2(C_n, F[\xi,\xi^{-1}])$ generated by $\{\xi,u\}_b$ for all $u\in F^*$, and $M(C_n, F)$ the subgroup of $L(C_n, F)$ generated by

$$m_b(u,v)=\{\xi,uv\}_b\cdot\{\xi,u\}_b^{-1}\cdot\{\xi,v\}_b^{-1},$$

called a multiplicator, for all $u, v\in F^*$.

Then $K_2(C_n, F[\xi,\xi^{-1}])\simeq K_2(C_n, F)\oplus L(C_n, F)$, and the sequence:

$$1\to M(C_n,F)\to L(C_n,F)\to F^*\to 1$$

is exact.

Recently, for general X (of infinite type) and for an arbitrary field F, Tits [21] has established that $K_2(X, F)$ can be uniquely defined and $K_2(X, F)=C(X, F)$. In particular, if $X=\tilde{X}_n$ is the extended Dynkin diagram of X_n (cf. [3]), then we obtain

$$K_2(\tilde{X}_n,F)\simeq K_2(F) \quad \text{if } \tilde{X}_n\neq\tilde{C}_n(n\geqq 1).$$

We will study the group structure of $K_2(\tilde{C}_n, F)$ in Section 3.

The relation between $K_2(X_n, F[\xi,\xi^{-1}])$ and $K_2(\tilde{X}_n, F)$ will be discussed in Section 2. And we will study the relation between $K_2\,\mathrm{Sp}(F[\xi,\xi^{-1}])$ and $K_2(F[\xi,\xi^{-1}])$ in Section 5.

2 Relations between $K_2(\tilde{X}_n, F)$ and $K_2(X_n, F[\xi,\xi^{-1}])$

Let $\mathfrak{g}$ be a finite dimensional simple Lie algebra over $\mathbb{C}$ (of type X_n), and $\tilde{\mathfrak{g}}$ the associated affine Kac-Moody algebra (of type $\tilde{X}_n$). Then we obtain the following exact sequence of Lie algebras over $\mathbb{C}$:

$$0\to\mathbb{C}\to\tilde{\mathfrak{g}}\to\mathbb{C}[\xi,\xi^{-1}]\otimes\mathfrak{g}\to 0.$$

Let Δ (resp. $\tilde{\Delta}$) be the root system of $\mathfrak{g}$ (resp. $\tilde{\mathfrak{g}}$). Then $\tilde{\Delta}$ (resp. $\tilde{\Delta}^{re}$) can be identified with $\Delta\times\mathbb{Z}$ (resp. $\Delta^{re}\times\mathbb{Z}$). Here 0 is also regarded as a root. We fix a Chevalley basis $\mathscr{C}=\{e_a\mid a\in\Delta^{re}\}$ for Δ^{re} and then $\{t^m\otimes e_a\mid(a,m)\in\Delta^{re}\times\mathbb{Z}\}$ gives a Chevalley basis $\tilde{\mathscr{C}}$ for $\tilde{\Delta}^{re}$ (cf. [4, 8]). We use the notation $e_\alpha=t^m\otimes e_a$ for $\alpha=(a,m)\in\tilde{\Delta}^{re}=\Delta^{re}\times\mathbb{Z}$. Using $\mathscr{C}$ and $\tilde{\mathscr{C}}$, we can construct a universal Chevalley group, $G(X_n, F[\xi,\xi^{-1}])$, over $F[\xi,\xi^{-1}]$ (of type X_n), and a universal Kac-Moody group, $G(\tilde{X}_n, F)$, over F (of type $\tilde{X}_n$). Also we can define the associated

Steinberg groups $\mathrm{St}(X_n, F[\xi, \xi^{-1}])$ and $\mathrm{St}(\tilde{X}_n, F)$ as in Section 1. Then we obtain the following commutative diagram of exact sequences (cf. [15, 22]).

$$\begin{array}{ccccccccc}
 & & 1 & & 1 & & F^* & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & K_2(\tilde{X}_n, F) & \longrightarrow & \mathrm{St}(\tilde{X}_n, F) & \longrightarrow & G(\tilde{X}_n, F) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow{\scriptstyle \phi} & & \downarrow & & \\
1 & \longrightarrow & K_2(X_n, F[\xi, \xi^{-1}]) & \longrightarrow & \mathrm{St}(X_n, F[\xi, \xi^{-1}]) & \longrightarrow & G(X_n, F[\xi, \xi^{-1}]) & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & F^* & & 1 & & 1 & & \\
 & & \downarrow & & & & & & \\
 & & 1 & & & & & &
\end{array}$$

Here ϕ is the isomorphism given as follows:

$$\phi(x_\alpha(t)) = x_a(t\xi^m)$$

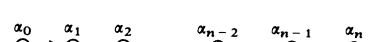
for all $\alpha = (a, m) \in \tilde{\Delta}^{\mathrm{re}} = \Delta^{\mathrm{re}} \times \mathbb{Z}$ and $t \in F$. Then

$$\phi(K_2(\tilde{X}_n, F)) = \begin{cases} K_2(C_n, F) \oplus M(C_n, F) & \text{if } \tilde{X}_n = \tilde{C}_n (n \geqq 1); \\ K_2(X_n, F) & \text{otherwise;} \end{cases}$$

(cf. Section 4).

3 Structure theorem for $K_2(\tilde{C}_n, F)$

Let $\tilde{C}_n$ be the following Dynkin diagram:



(usually called $C_n^{(1)}$ cf. [8; p. 44]) with simple roots $\tilde{\Pi} = \{a_0, a_1, \ldots, \alpha_n\}$ as vertices. Let $G(\tilde{C}_n, F)$ be a universal Kac-Moody group of type $\tilde{C}_n$, and $H = \prod_{i=0}^{n} H_i$ (direct product) the standard torus of $G(\tilde{C}_n, F)$, where $H_i \simeq F^*$ is the one dimensional torus corresponding to α_i.

Put

$$K = \frac{K_2(A_1, F) \times K_2(C_n, F)}{J},$$

where A_1 (resp. C_n) is the Dynkin subdiagram of $\tilde{C}_n$ generated by $\{\alpha_0\}$ (resp. $\{\alpha_1, \ldots, \alpha_n\}$), and J is the subgroup of $K_2(A_1, F) \times K_2(C_n, F)$ generated by $\{u^2, v\}_{\alpha_0} \cdot \{v^2, u\}_\beta$ with $\beta = 2\alpha_1 + 2\alpha_2 + \ldots + 2\alpha_{n-1} + \alpha_n$ for all $u, v \in F^*$.

Let $\tilde{T}$ be the set $H \times K$ with the following product operation:

$$(d \cdot z) \cdot (d', z') = (d d', z z' z'')$$

for all $(d, z), (d', z') \in H \times K$, with

$$d = (u_0, \ldots, u_n), d' = (u_0', \ldots, u_n') \in H, \qquad z, z' \in K,$$

$$z'' = \prod_{i=0}^{n} \{u_i', u_i\}_{\alpha_i}^{-1} \cdot \prod_{0 \leqq i < j \leqq n} \{u_j^{\alpha_i(h_j)}, u_i'\}_{\alpha_i} \bmod J,$$

where h_j is the coroot of α_j. Then $\tilde{T}$ is a group, and

$$1 \to K \to \tilde{T} \to H \to 1$$

is a central extension. Put $\tilde{B} = \tilde{T} \leftthreetimes U$.

Now we will construct a group G.

(1) Suppose $n > 1$.

For $0 \leqq i < j \leqq n$, put

$$\tilde{\Delta}^{\mathrm{re}}(i,j) = \tilde{\Delta}^{\mathrm{re}} \cap (\mathbb{Z}\alpha_i \oplus \mathbb{Z}\alpha_j),$$

$$\tilde{\Delta}_+^{\mathrm{re}}(i,j) = \tilde{\Delta}^{\mathrm{re}}(i,j) \cap \tilde{\Delta}_+,$$

$$U_{ij} = \langle x_\alpha(t) \mid \alpha \in \tilde{\Delta}_+^{\mathrm{re}}(i,j), t \in F \rangle \subseteq \mathrm{St}(\tilde{C}_n, F),$$

$$V_{ij} = \langle y x_\alpha(t) y^{-1} \mid \alpha \in \tilde{\Delta}_+^{\mathrm{re}} \setminus \tilde{\Delta}_+^{\mathrm{re}}(i,j), t \in F, y \in U_{ij} \rangle \subseteq \mathrm{St}(\tilde{C}_n, F),$$

$$\tilde{T}_{ij} = \langle H_i, H_j \rangle \subseteq \tilde{T},$$

$$\tilde{S}_{ij} = \langle x_{\pm\alpha_i}(t), x_{\pm\alpha_j}(t) \mid t \in F \rangle \subseteq \mathrm{St}(\tilde{C}_n, F) \quad \text{for } (i,j) \neq (0,n),$$

$$\tilde{S}_{0n} = \frac{\mathrm{St}(A_1, F) \times \mathrm{St}(A_1, F)}{J'},$$

where $J' = \langle (\{u^2, v\}_a, \{v^2, u\}_a) \mid u, v \in F^* \rangle$ and a is a fixed root of $\mathrm{St}(A_1, F)$. Then $U = U_{ij} \leftthreetimes V_{ij}$, and we claim that there is a canonical injection v_{ij} of $\tilde{T}_{ij}$ into $\tilde{S}_{ij}$, which sends $d \in H$ having coefficient $u \in F^*$ at the k-th component and 1's everywhere else to $h_k(u) \in \mathrm{St}(\tilde{C}_n, F)$:

By our construction and by the presentation of symplectic K_2 over fields due to Matsumoto ([9], Theor. 5.10, Prop. 5.5) it follows that this map factors through $K_2(A_1, F) \times K_2(C_n, F)$, and from [9], Lemme 5.4(c) we conclude that it vanishes on J, which proves our claim.

Let, for $0 \leqq i < j \leqq n$,

$$J_{ij} = \langle y \cdot v_{ij}(y)^{-1} \mid y \in \tilde{T}_{ij} \rangle \subseteq \tilde{T} \leftthreetimes \tilde{S}_{ij},$$

$$\tilde{P}_{ij} = \left(\frac{\tilde{T} \leftthreetimes \tilde{S}_{ij}}{J_{ij}} \right) \leftthreetimes V_{ij}.$$

Then, by Steinberg's relations in Chevalley groups (cf. [9], Lemme 5.2), there is a canonical homomorphism μ_{ij} of $\tilde{P}_{ij}$ into $\mathrm{St}(\tilde{C}_n, F)$, and $\tilde{B}$ can be considered as a subgroup of $\tilde{P}_{ij}$. If $\{i,j\} \cap \{k,l\} = \{m\}$ with $0 \leqq i < j \leqq n$ and $0 \leqq k < l \leqq n$, then we can identify the subgroup

$$\langle \tilde{B}, \mu_{ij}^{-1}(x_{-\alpha_m}(t)) \mid t \in F \rangle \subseteq \tilde{P}_{ij}$$

with the subgroup

$$\langle \tilde{B}, \mu_{kl}^{-1}(x_{-\alpha_m}(t)) | t \in F \rangle \subseteq \tilde{P}_{kl}.$$

We call both of them $\tilde{P}_m$. Taking, for $0 \leqq i < j \leqq n$ and $0 \leqq k < l \leqq n$ with $(i, j) \neq (k, l)$,

$$\tilde{P}_{ij} \cap \tilde{P}_{kl} = \tilde{B} \quad \text{if } \{i,j\} \cap \{k,l\} = \emptyset;$$
$$\tilde{P}_{ij} \cap \tilde{P}_{kl} = \tilde{P}_m \quad \text{if } \{i,j\} \cap \{k,l\} = \{m\},$$

we let

$$\tilde{G} = *\langle \tilde{P}_{ij} | 0 \leqq i < j \leqq n \rangle$$

be the amalgamated free product of the $\tilde{P}_{ij}$ along with their intersections.

(2) Suppose $n = 1$.

In this case, the Dynkin diagram is $\tilde{A}_1 = \tilde{C}_1$:

(usually called $A_1^{(1)}$ cf. [8; p. 44]). For $i = 0, 1$, put

$$U_i = \langle x_{\alpha_i}(t) | t \in F \rangle \subseteq \mathrm{St}(\tilde{A}_1, F),$$
$$V_i = \langle y x_\alpha(t) y^{-1} | \alpha \in \tilde{\Delta}_+^{\mathrm{re}} \setminus \{\alpha_i\}, t \in F, y \in U_i \rangle \subseteq \mathrm{St}(\tilde{A}_1, F),$$
$$\tilde{T}_i = \langle H_i \rangle \subseteq \tilde{T},$$
$$\tilde{S}_i = \langle x_{\pm\alpha_i}(t) | t \in F \rangle \subseteq \mathrm{St}(\tilde{A}_1, F).$$

Then $U = U_i \ltimes V_i$, and there is a canonical injection v_i of $\tilde{T}_i$ into $\tilde{S}_i$. Let, for $i = 0, 1$,

$$J_i = \langle y \cdot v_i(y)^{-1} | y \in \tilde{T}_i \rangle \subseteq \tilde{T} \ltimes \tilde{S}_i,$$
$$\tilde{P}_i = \left(\frac{\tilde{T} \ltimes \tilde{S}_i}{J_i} \right) \ltimes V_i.$$

Then $\tilde{B}$ can be considered as a subgroup of $\tilde{P}_i$. Let

$$\tilde{G} = \tilde{P}_0 *_{\tilde{B}} \tilde{P}_1$$

be the amalgamated free product of $\tilde{P}_0$ and $\tilde{P}_1$ along with $\tilde{B}$.

In any case of $n \geqq 1$, there is a natural homomorphism, say ρ, of $\tilde{G}$ onto $\mathrm{St}(\tilde{C}_n, F)$. In the standard way as in Steinberg [18], the relations of $\mathrm{St}(\tilde{C}_n, F)$ can be lifted to $\tilde{G}$ using ρ^{-1}, since $\mathrm{St}(\tilde{C}_n, F)$ has an analogous decomposition (cf. [19]) and *ker* ρ is central. (In the much more general context of Kac-Moody groups, a complete proof of the lifting of the relations is given in [15], section 6, IV, however, in our case, the 'classical' construction of Steinberg [18], proof of Theor. 10, gives the result.) Hence ρ is an isomorphism, and

$$K_2(\tilde{C}_n, F) \simeq K$$

(cf. [9, 14]). On the other hand, we see easily $K \simeq K_2(A_1, F) \oplus Q(F)$, where

$$Q(F) = \frac{K_2(A_1, F)}{\langle \{u^2, v\}_a | u, v \in F^* \rangle}$$

with a fixed root a of $\mathrm{St}(A_1, F)$. By Suslin [20], $Q(F) \simeq I^2(F)$. Therefore we have the following.

Theorem C $K_2(\tilde{C}_n, F) \simeq K_2\,\mathrm{Sp}(F) \oplus I^2(F)$ *for all* $n \geqq 1$.

4 Structure theorem for $K_2(C_n, F[\xi, \xi^{-1}])$

Let $\Pi = \{a_1, \ldots, a_n\}$ be the set of fixed simple roots of Δ (of type C_n), and b the negative highest root in Δ. Put $\alpha_0 = (b, 1)$, $\alpha_1 = (a_1, 0)$, $\ldots$, $\alpha_n = (a_n, 0) \in \Delta \times \mathbb{Z}$. Then $\tilde{\Pi} = \{\alpha_0, \alpha_1, \ldots, \alpha_n\}$ becomes the set of simple roots of $\tilde{\Delta} = \Delta \times \mathbb{Z}$ (of type $\tilde{C}_n$). By the definition of the isomorphism ϕ of $\mathrm{St}(\tilde{C}_n, F)$ onto $\mathrm{St}(C_n, F[\xi, \xi^{-1}])$, we obtain

$$\phi(h_{\alpha_0}(u)) = \{\xi, u\}_b \cdot h_b(u),$$
$$\phi(h_{\alpha_i}(u)) = h_{a_i}(u)$$

for all $u \in F^*$ and $1 \leqq i \leqq n$. Hence, the isomorphism ϕ maps $K_2(\tilde{C}_n, F)$ onto $K_2(C_n, F) \oplus M(C_n, F)$. By Theorem C, we see the following.

Proposition 1 $M(C_n, F) \simeq I^2(F)$ *for all* $n \geqq 1$.

Let $P(F)$ be the set $F^* \times I^2(F)$ with the product operation:

$$(u, y) \cdot (v, z) = (uv, y + z - \langle\langle u, v \rangle\rangle)$$

for all (u, y), $(v, z) \in P(F)$, which makes $P(F)$ into a group (cf. [20]), where $\langle\langle \cdot, \cdot \rangle\rangle$ is the so-called Pfister cocycle. Then, via the isomorphism in Proposition 1, the following two exact sequences are equivalent:

$$1 \to M(C_n, F) \to L(C_n, F) \to F^* \to 1,$$
$$1 \to I^2(F) \to P(F) \to F^* \to 1.$$

Hence we obtain the following.

Proposition 2 $L(C_n, F) \simeq P(F)$ *for all* $n \geqq 1$.

Since $K_2(C_n, F[\xi, \xi^{-1}]) \simeq K_2(C_n, F) \oplus L(C_n, F)$ as in Section 2, the next theorem comes from Proposition 2.

Theorem B $K_2\,\mathrm{Sp}_{2n}(F[\xi, \xi^{-1}]) \simeq K_2\,\mathrm{Sp}(F) \oplus P(F)$ *for all* $n \geqq 1$.

Hence we obtain the stability of $K_2\,\mathrm{Sp}_{2n}(F[\xi, \xi^{-1}])$ from $n = 1$. In general, the stability of this kind of groups is known if n is large enough (for example, see Habdank [5]).

5 Relations between $K_2\,\mathrm{Sp}(F[\xi, \xi^{-1}])$ and $K_2(F[\xi, \xi^{-1}])$

As is well-known, the tame symbol induces $K_2(F[\xi, \xi^{-1}]) \simeq K_2(F) \oplus F^*$. By Theorem B, we know already $K_2\,\mathrm{Sp}(F[\xi, \xi^{-1}]) \simeq K_2\,\mathrm{Sp}(F) \oplus P(F)$. Then the

canonical homomorphism ψ of $K_2\,\mathrm{Sp}(F[\xi, \xi^{-1}])$ onto $K_2(F[\xi, \xi^{-1}])$ naturally induces the corresponding homomorphism on their components:

$$\psi_1\colon\ K_2\,\mathrm{Sp}(F) \to K_2(F),$$
$$\psi_2\colon\ P(F) \to F^*.$$

Then the kernel of ψ_1 is isomorphic to $I^3(F)$ (cf. [20], and it follows from Propositions 1 and 2 that the kernel of ψ_2 is isomorphic to $I^2(F)$. On the other hand, it is now easy to compute the cokernel of the hyperbolic map of $K_2(F[\xi, \xi^{-1}])$ to $K_2\,\mathrm{Sp}(F[\xi, \xi^{-1}])$ (cf. Section 6). Hence we obtain the following.

Theorem A *There are the following exact sequences*:

$$1 \to I^3(F) \oplus I^2(F) \to K_2\,\mathrm{Sp}(F[\xi, \xi^{-1}]) \to K_2(F[\xi, \xi^{-1}]) \to 1,$$
$$1 \to N(F) \oplus \{\pm 1 \in F\} \to K_2(F[\xi, \xi^{-1}]) \to K_2\,\mathrm{Sp}(F[\xi, \xi^{-1}]) \to I^2(F) \oplus I(F) \to 1.$$

6 Remarks

(1) Put $M(F) = M(C_n, F)$ (cf. Section 1). Set

$$m^{(2)}(u_1, u_2) = m_b(u_1, u_2),$$
$$m^{(r+1)}(u_1, \ldots, u_{r+1}) = m^{(r)}(u_1, \ldots, u_{r-1}, u_r \cdot u_{r+1}) \cdot m^{(r)}(u_1, \ldots, u_{r-1}, u_r)^{-1} \cdot m^{(r)}(u_1, \ldots, u_{r-1}, u_{r+1})^{-1}$$

for $r \geqq 2$. Let $M^{(r)}(F)$ be the subgroup of $M(F)$ generated by $m^{(r)}(u_1, \ldots, u_r)$ for all $u_1, \ldots, u_r \in F^*$. Then, by Proposition 1, we obtain

$$M^{(r)}(F) \simeq I^r(F),$$

as abelian groups, for all $r \geqq 2$.

(2) Let $\mathrm{Cok}\,\mathscr{H}(\cdot)$ be the cokernel of the hyperbolic map $\mathscr{H}(\cdot)$ of $K_2(\cdot)$ to $K_2\,\mathrm{Sp}(\cdot)$ (cf. [2]). Then, our result implies

$$\mathrm{Cok}\,\mathscr{H}(F[\xi, \xi^{-1}]) \simeq \mathrm{Cok}\,\mathscr{H}(F) \oplus \frac{L(F)}{L'(F)} \simeq I^2(F) \oplus I(F).$$

Here $L'(F)$ is the subgroup of $L(F) = L(C_n, F)$ generated by $\{\xi, u^2\}_b$ for all $u \in F^*$, and $I(F) \simeq L(F)/L'(F)$.

(3) As it has been pointed out by the referee, there is the following reformulation of some aspects of theorems A and B:

Let $\Sigma_n(F)$, $n = 1, 2$ denote the kernel of the map $K_n(K) \oplus I^n(F) \to K_n(F)/2$ given by projection and sending an n-fold Pfister form to the corresponding symbol. Then, Suslin's result says that $K_2\,\mathrm{Sp}(F) = \Sigma_2(F)$, and Theorem B gives $K_2\,\mathrm{Sp}_{2n}F[\xi, \xi^{-1}] = \Sigma_2(F) \oplus \Sigma_1(F)$.

References

1. Abe, E., Morita, J.: Some Tits systems with affine Weyl groups in Chevalley groups over Dedekind domains. J. Algebra **115**, 450–465 (1988)
2. Bak, A.: K-theory of forms. Ann. Math. Stud. **98**, (1981)

3. Bourbaki, N.: Groupes et algèbres de Lie. Chaps. 4–6. Paris: Hermann 1968
4. Garland, H.: The arithmetic theory of loop algebras. J. Algebra **53**, 480–551 (1978)
5. Habdank, G.: Homology stability for Λ-quadratic groups. (to appear)
6. Hurrelbrink, J.: Endlich präsentierte arithmetische Gruppen und K_2 über Laurent-Polynomringen. Math. Ann. **225**, 123–129 (1977)
7. Kac, V.G.: Simple irreducible graded Lie algebras of finite growth. Math. USSR, Izv. **2**, 211–230 (1968)
8. Kac, V.G.: Infinite dimensional Lie algebras. Prog. Math. **44**, (1983)
9. Matsumoto, H.: Sur les sous-groupes arithmétiques des groupes semi-simple déployés. Ann. Scient. Ec. Norm. Sup. (4) **2**, 1–62 (1969)
10. Merkurjev, A.S., Suslin, A.A.: K-cohomology of Severi-Bauer varieties and the norm residue homomorphism. Math. USSR, Izv. **21**, 307–340 (1983)
11. Milnor, J., Husemoller, D.: Symmetric bilinear forms. Berlin Heidelberg New York: Springer 1973
12. Moody, R.V.: A new class of Lie algebras. J. Algebra **10**, 211–230 (1968)
13. Morita, J.: Commutator relations in Kac-Moody groups. Proc. Jpn. Acad., Ser. A **63**, 21–22 (1987)
14. Morita, J., Rehmann, U.: A Matsumoto-type theorem for Kac-Moody groups. Tōhoku Math. J. (in press)
15. Peterson, D.H., Kac, V.G.: Infinite flag varieties and conjugacy theorems. Proc. Natl. Acad. Sci. USA **80**, 1778–1782 (1983)
16. Quillen, D.: Higher K-theory for categories with exact sequences. Lond. Math. Soc. Lect. Note Ser. **11**, 95–103 (1974)
17. Rehmann, U.: Präsentation von Chevalley Gruppen über $K[t]$. Preprint 1975
18. Steinberg, R.: Lectures on Chevalley groups. Yale Univ. Lect. Notes, New Haven CT, 1968
19. Serre, J.-P.: Trees. Berlin Heidelberg New York: Springer 1980
20. Suslin, A.A.: Torsion in K_2 of fields. K-Theory **1**, 5–29 (1987)
21. Tits, J.: Uniqueness and presentation of Kac-Moody groups over fields. J. Algebra **105**, 542–573 (1987)
22. Tits, J.: Groupes associés aux algèbres de Kac-Moody. Séminaire Bourbaki 41ème année, 1988/89, no. 700

Math. Z. 206, 67–69 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Subnormal subgroups of 3-dimensional Poincaré duality groups

R. Bieri[1] and J.A. Hillman[2]

[1] Fachbereich Mathematik, J.W. Goethe-Universität, W-6000 Frankfurt a/M., Federal Republic of Germany

[2] Department of Pure Mathematics, The University of Sydney, Sydney, NSW 2006, Australia

Received August 8, 1988; in final form April 15, 1990

In this note we shall show that if a 3-dimensional Poincaré duality group G has a nontrivial almost finitely presentable subnormal subgroup of infinite index then either G is virtually the group of a surface bundle over the circle or the subgroup is infinite cyclic and normal in G. (In the latter case it remains unknown whether G must be the group of a Seifert fibred 3-manifold, although this is true if G has a subgroup of finite index with infinite abelianization [5]). A related result about 3-manifold groups with nontrivial subnormal subgroups has been given by Elkalla [4].

Theorem. *Let G be a 3-dimensional Poincaré duality group which has a nontrivial almost finitely presentable subgroup H which is subnormal and of infinite index in G. Then either H is infinite cyclic and is normal in G or G is virtually poly-Z, or H has finite index in a normal subgroup which is a surface group and G is virtually the group of a surface bundle over the circle.*

Proof. Since H is subnormal in G there is a finite sequence $\{J_i | 0 \leqq i \leqq n\}$ of subgroups of G with $J_0 = H$, $J_n = G$ and J_i normal in J_{i+1} for each $i < n$. Since H has infinite index in G it has cohomological dimension at most 2 [12], and so is either infinite cyclic, a nonabelian free group or has cohomological dimension 2.

Suppose first that H is infinite cyclic. Each locally-nilpotent normal subgroup of (any) group J is contained in the Hirsch-Plotkin radical $\sqrt{J}$, the unique maximal locally-nilpotent normal subgroup [9: 12.1.3]. By the uniqueness, if J_i is normal in J_{i+1} then $\sqrt{J_i}$ is also normal in J_{i+1}, and so is contained in $\sqrt{J_{i+1}}$. A finite induction now shows that H is contained in $\sqrt{G}$. If $\sqrt{G}$ is abelian and of rank 1 then $\mathrm{Aut}(\sqrt{G})$ is isomorphic to a subgroup of $Q^\times$, and so $C_G(\sqrt{G})$ (the centralizer of $\sqrt{G}$ in G) contains the commutator subgroup G'. If moreover G/G' is infinite then $\sqrt{G}$ must be finitely generated, and so infinite cyclic [5]. Thus, whether or no G/G' is finite, $G/C_G(\sqrt{G})$ must be finite and hence of order at most 2, and so H is normal in G. If $\sqrt{G} = Z^2$ then G

is virtually poly-Z by [5] (or [6]). Otherwise c.d. $\sqrt{G}=3$ so $[G:\sqrt{G}]$ is finite [12] and $\sqrt{G}$ is finitely generated. Therefore $\sqrt{G}$ is finitely generated torsion free nilpotent and hence poly-Z by 5.2.20 of [9], and so G is again virtually poly-Z.

Next suppose that c.d. $H=2$. Let $k=\min\{i|[J_i:H]=\infty\}$. Then H has finite index in J_{i-1}, which therefore is also almost finitely presentable. Suppose that c.d. $J_k=2$. If K is a finitely generated subgroup of J_k which contains J_{k-1} then $[K:J_{k-1}]$ is finite, by Corollary 8.6 of [1], and so J_k is the union of a strictly increasing sequence of finite extensions of J_{k-1}. But by Lemma 1.4 of [11] the number of indecomposable factors in such intermediate groups must be strictly decreasing unless one (and hence all) is indecomposable. Thus J_{k-1} is indecomposable, and so has one end (since it is torsion free but not infinite cyclic). By Proposition 8.2.3 of [2] there is a free $Z[J_k]$-module W such that $H^2(J_k;W)\neq 0$. Since J_{k-1} is almost finitely presentable and has one end $H^q(J_{k-1};W)=0$ for $q=0$ or 1 and $H^2(J_{k-1};W)$ is an induced $Z[J_k/J_{k-1}]$-module. As $[J_k:J_{k-1}]$ is infinite, $H^0(J_k/J_{k-1};H^2(J_{k-1};W))=0$, by Lemma 8.1 of [1]. The Lyndon-Hochschild-Serre spectral sequence for J_k as an extension of J_k/J_{k-1} by the normal subgroup J_{k-1} now gives $H^2(J_k;W)=0$, which is a contradiction. Therefore c.d. $J_k=3$, so $[G:J_k]$ is finite, and J_k is a 3-dimensional Poincare duality group with an almost finitely presentable normal subgroup of infinite index which is not cyclic, namely J_{k-1}. It now follows from [6] that J_{k-1} must be a surface group and G is virtually the group of a surface bundle over the circle.

Suppose finally that G has a finitely generated nonabelian free subnormal subgroup. We may assume that $\{J_i|0\leqq i\leqq n\}$ is a chain of minimal length n among subnormal chains in G with $H=J_0$ a finitely generated nonabelian free group. In particular $[J_1:H]$ is infinite, for otherwise J_1 would also be a finitely generated nonabelian free group. We may also assume that H is maximal in the partially ordered set of finitely generated free normal subgroups of J_1. (Note that ascending chains of such subgroups are always finite, for if F is a nontrivial finitely generated normal subgroup of a free group G then $[G:F]$ is finite and $[G:F]\,(1-\text{rank } G)=1-\text{rank } F$).

Since J_1 has a finitely generated nonabelian free normal subgroup of infinite index it is not free, and nor is it a 3-dimensional Poincaré duality group [6]. Therefore c.d. $J_1=2$. The kernel of the homomorphism from J_1/H to $\text{Out}(H)$ determined by the conjugation action of J_1 on H is $HC_{J_1}(H)/H$, which is isomorphic to $C_{J_1}(H)$ since H has trivial centre. As $\text{Out}(H)$ is virtually of finite cohomological dimension [3] and c.d. $C_{J_1}(H)$ is finite, v.c.d. J_1/H is finite. Therefore c.d. $J_1=$c.d. $H+$v.c.d. J_1/H, by Theorem 5.6 of [1], so v.c.d. $J_1/H=1$ and J_1/H is virtually free.

If g normalizes J_1 then $HH^g/H=H^g/H\cap H^g$ is a finitely generated normal subgroup of J_1/H and so either has finite index or is finite. In the former case J_1/H would be finitely presentable (since it is then an extension of a finitely generated virtually free group by a finitely generated free normal subgroup) and as it is subnormal in G it must be a surface group, by our earlier work. But surface groups do not have finitely generated nonabelian free normal subgroups. Therefore HH^g/H is finite and so $HH^g=H$, by the maximality of H. Since this holds for any g in J_2, H is normal in J_2 and so is the initial term of a subnormal chain of length $n-1$ terminating with G, contradicting our

assumption that n was minimal. Therefore G has no finitely generated nonabelian free subnormal subgroups. □

Elkalla has shown that if the group G of a closed 3-manifold has a subnormal subgroup H which is contained in a finitely generated indecomposable subgroup U of infinite index and if G is "U-residually finite" then either H is infinite cyclic and is normal in G or G is virtually the group of a surface bundle, and U is commensurate with a normal surface subgroup [4]. (Note that finitely generated subgroups of 3-manifold groups are finitely presentable [10]).

For other recent work on 3-dimensional Poincaré duality groups see [5, 6, 7, 8, 13].

References

1. Bieri, R.: Homological Dimension of Discrete Groups. Queen Mary College Mathematics Notes, London 1976
2. Brown, K.S.: Cohomology of Groups. Graduate Texts in Mathematics vol. 87. Berlin Heidelberg New York: Springer 1982
3. Culler, M., Vogtmann, K.: Moduli of graphs and automorphisms of free groups. Invent. Math. **84**, 91–120 (1986)
4. Elkalla, H.S.: Subnormal subgroups in 3-manifold groups. J. Lond. Math. Soc. **30**, 342–360 (1984)
5. Hillman, J.A.: Seifert fibre spaces and Poincaré duality groups. Math. Z. **190**, 365–369 (1985)
6. Hillman, J.A.: Three-dimensional Poincaré duality groups which are extensions. Math. Z. **195**, 89–92 (1987)
7. Kropholler, P.H.: An analogue of the torus decomposition theorem for certain Poincaré duality groups. Proc. Lond. Math. Soc. **60**, 503–529 (1990)
8. Kropholler, P.H., Roller, M.A.: Splittings of Poincaré duality groups. Math. Z. **197**, 421–438 (1988)
9. Robinson, D.J.S.: A Course in the Theory of Groups. Graduate Texts in Mathematics vol. 80. Berlin Heidelberg New York: Springer 1982
10. Scott, G.P.: Compact submanifolds of 3-manifolds. J. Lond. Math. Soc. **7**, 437–440 (1973)
11. Scott, G.P.: Normal subgroups in 3-manifold groups. J. London Math. Soc. **13**, 5–12 (1976)
12. Strebel, R.: A remark on subgroups of infinite index in Poincaré duality groups. Comment. Math. Helvet. **52**, 317–324 (1977)
13. Thomas, C.B.: Splitting theorems for certain PD^3-groups. Math. Z. **186**, 201–209 (1984)

Math. Z. 206, 71–80 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Extension a la dimension n d'un théorème de Ortel et Schneider

Y. Dupain

CeReMaB-URA226, Université Bordeaux I, 351, cours de la Libération, F-33405 Talence Cedex, France

Received February 19, 1990

I. Introduction

Soit φ une fonction continue dans le disque unité du plan complexe $\mathbb{C}$. Bagemihl et Seidel [1] ont démontré qu'il existe une fonction holomorphe F dans ce disque, telle que, pour presque tout point ξ du cercle unité: $F(r\xi)-\varphi(r\xi)\to 0$ quand $r\to 1$. Pour toute fonction φ mesurable sur le cercle unité, il existe donc une fonction holomorphe F dans le disque unité, admettant φ pour limite radiale presque partout. Sauf cas particulier, F n'est pas bornée, mais Kahane et Katznelson [5] ont démontré que, quel que soit l'ordre de croissance (non borné) ω, on pouvait choisir F de croissance inférieure à ω.

Enfin, Ortel et Schneider [6], démontrent l'existence de fonctions holomorphes d'ordre de croissance imposé et admettant pour limites radiales inférieures et supérieures, des fonctions mesurables quelconques compatibles (presque partout).

L'existence de fonctions holomorphes dans la boule unité de $\mathbb{C}^n$, admettant pour limite radiale presque partout une fonction donnée n'a été étudiée que récemment pour $n>1$. En étendant le théorème de Mergelyan à certains compacts de $\mathbb{C}^n$, Hakim et Sibony [3] démontrent l'extension naturelle du théorème de Bagemihl-Seidel. Ce problème a aussi été résolu par Dupain [2] en utilisant la méthode «d'approximation sur les grands cercles».

Le seul résultat connu, contenant des conditions de croissance, est celui de Iordan [4], qui en utilisant la construction de Hakim et Sibony démontre:

Théorème [Iordan]. *Soit φ une fonction mesurable sur la sphère unité S de $\mathbb{C}^n$. Soit ω un ordre de croissance non borné. Alors il existe une fonction holomorphe F dans la boule unité de $\mathbb{C}^n$, de croissance inférieure à ω, et une suite croissante r_n tendant vers 1, telles que, pour presque tout point ξ de S, $F(r_n\xi)\to\varphi(\xi)$.*

En utilisant les résultats et méthodes développés dans [2], nous allons démontrer le théorème suivant:

Théorème. *Soit φ une fonction continue dans la boule unité B de $\mathbb{C}^n$. Soit ω un ordre de croissance non borné. Alors il existe une fonction holomorphe F dans B de croissance inférieure à ω, et une bijection croissante σ de $[0, 1[$ sur $[0, 1[$ telles que: Pour tout point Z de S, $f(\sigma(r)e^{i\theta}Z)-\varphi(re^{i\theta}Z)\to 0$ pour presque tout θ.*

Remarques. (1) Le comportement radial est estimé presque partout pour la fonction F, mais aussi pour toutes ses slice-fonctions.
(2) En utilisant le théorème de Lusin, on peut obtenir comme limites radiales inférieures et supérieures n'importe quelles fonctions mesurables, mais le résultat est plus précis car on peut imposer le comportement asymptotique.

II. Notations

n désigne un entier supérieur ou égal à 2, $\mathbb{C}^n$ l'espace vectoriel des n-uples

$$z=(z_1, z_2, \ldots z_n)$$

de nombres complexes muni du produit scalaire $\langle z, w\rangle=\sum z_i \bar{\omega}_i$ et de la norme

$$|z|=\langle z, z\rangle^{1/2}.$$

La boule unité de $\mathbb{C}^n$ est $B=\{z\in\mathbb{C}^n; |z|<1\}$ et la sphère unité $S=\{\xi\in\mathbb{C}^n; |\xi|=1\}$, D est le disque unité de $\mathbb{C}$, $D=\{z\in\mathbb{C}, |z|<1\}$ et sa frontière $T=\{\xi\in\mathbb{C}; |\xi|=1\}$.

Pour $P\in S$, nous désignons par $\mathscr{C}_P$, le «grand cercle» passant par P,

$$\mathscr{C}_P=\{e^{i\theta}P; \theta\in[0, 2\pi[\}.$$

La mesure de Lebesgue normalisée de S est notée σ, et celle de $[0, 2\pi[$, de T et de $\mathscr{C}_P$ est notée μ. Seules ces mesures seront considérées.

d est l'écart défini sur S par: $d(\eta, \xi)=(1-|\langle\eta, \xi\rangle|^2)^{1/2}$.

L'algèbre des fonctions continues sur $B\cup S$ (respectivement $D\cup T$) et holomorphes dans B (respectivement D) est notée $A(B)$ (respectivement $A(D)$).

Un ordre de croissance est une fonction croissante ω: $[0, 1[\to]0, +\infty[$. Dans toute la suite, nous ne considérerons que des fonctions ω non bornées.

III. Approximations de fonctions continues sur S par des fonctions holomorphes dans B à croissance contrôlée

Let but essentiel de ce paragraphe est de démontrer la proposition suivante.

Proposition I. *Il existe une constante $c<1$ telle que:*
Soient φ une fonction continue sur S et γ et ν des réels de $]0, 1[$.
Alors il existe un réel M tel que, pour tout k assez grand et tout point Z de S il existe des sous-ensembles W_Z^k contenus dans $\mathscr{C}_Z$ de mesure supérieure à $1-\gamma$, et une fonction h_k de $A(B)$ vérifiant

(1) $|h_k(z)|\leqq M|z|^k$

(2) si $z\in W_Z^k$; $|h_k(rz)|<|\varphi(z)|+\nu$ *pour* $r\in[0, 1[$
$|h_k(rz)|<\nu$ *pour* $r<1-1/k$
$|h_k(rz)-\varphi(z)|<c|\varphi(z)|+\nu$ *pour* $r>1-1/4k$.

Il est facile, à partir du Lemme 3 de [2] de démontrer le lemme

Lemme 0. *Il existe une constante $c_1<1$ telle que:*
Soient φ une fonction continue sur S et γ et η des réels de $]0, 1[$.

Alors pour tout δ assez petit, il existe une fonction f de $A(B)$ vérifiant
(1) $|z|<1-4\delta \Rightarrow |f(z)|<\eta$
(2) *Pour tout point Z de S, il existe un sous-ensemble W_Z de $\mathscr{C}_Z$, de mesure supérieure à $1-\gamma$ tel que*

si $z\in W_z$; $\quad |f(rz)|<|\varphi(z)|+\eta \quad$ *pour* $r\in[0,1[$
$\quad |f(rz)-\varphi(z)|<c_1|\varphi(z)|+\nu \quad$ *pour* $r>1-\delta$.

Il suffit en effet de suivre les démonstrations des Lemmes 4 et 5, sans faire intervenir la restriction des classes S_α à un ensemble A dénombrable. La dénombrabilité de l'ensemble est simplement imposée pour conserver des propriétés de mesure afin d'obtenir à la fin des ensembles W_Z indépendants de la fonction φ choisie au départ, ce dont nous n'avons aucun besoin ici.

Pour démontrer la Proposition I, on utilise une fonction F vérifiant le Lemme 0. Le contrôle de la croissance est obtenu en multipliant F par une fonction «approchant la fonction 1» et de croissance contrôlée. Nous allons maintenant construire une telle fonction.

III. 1 Approximations de la fonction 1

Notre but est de démontrer le lemme suivant:

Lemme 1. *Il existe un réel $\Delta>0$ tel que:*
Pour tous réels de $]0,1[$ *ε et γ, il existe un réel M tel que:*
Pour tout k assez grand, il existe une fonction w: $B\to[0,1/2]$ et une fonction g de $A(B)$ vérifiant:
(1) $|z|>1-1/4k \Rightarrow w(z)>\Delta$
(2) $|g(z)|\leqq M|z|^k$
(3) *pour tout point Z de S, il existe un sous ensemble W_Z de $\mathscr{C}_Z$ de mesure supérieure à $1-\gamma$ tel que: $z\in W_Z \Rightarrow |g(rz)-w(rz)|<\varepsilon$ pour $r\in[0,1[$.*

III 1.1. La fonction de base en dimension 1

Soient a et b deux réels tels que $e^{-1}<a<b<e^{-1/2}$. Soit K appartenant à $\mathbb{N}^*$ tel que $k\geqq K$ entraine $(1-1/k)^k<a$ et $(1-1/2k)^k>b$.

Soient ε_1 et ε_2 deux réels de $]0,1[$. D'après le théorème de Mergelyan, il existe un compact E contenu dans T, de mesure supérieure à $1-\varepsilon_2$ et une fonction h de $A(D)$ vérifiant:
(1) $|z|<a \Rightarrow |h(z)|<\varepsilon_1$
(2) $z=e^{i\theta}\in E \Rightarrow |h(rz)-c(r)\bar{z}|<\varepsilon_1$
où $c(r)$ est la fonction définie sur $[0,1]$ par $c(r)=0$ pour $r\leqq a$; $c(r)=1$ pour $r\geqq b$ et $c(r)$ est linéaire sur $[a,b]$.

Fixons une fonction h vérifiant les propriétés précédentes et posons $M_1=\operatorname{Max}_{\bar{D}}(h)$.

Pour $k>K$ posons $f_k(z)=z^k h(z^k)$.
Alors $|z|<1-1/k \Rightarrow |z|^k<a \Rightarrow |f_k(z)|<\varepsilon_1|z|^k$.
Définisons le sous-ensemble E_k de T par $z\in E_k \Leftrightarrow z^k\in E$.

Alors $\mu(E_k)=\mu(E)>1-\varepsilon_2$.

Soit $z=e^{i\theta}$ un élément de E_k; alors $z^k\in E$ et $r^k\geqq b$ entraine $|h(r^kz^k)-\bar{z}^k|<\varepsilon_1$, donc, en particulier, si $r>1-1/2k$, $|f_k(rz)-r^k|<\varepsilon_1$.

III 1.2. Une fonction approchant 1 en dimension n

Construction de la fonction

Soit k un entier supérieur à K. Posons $\delta=(4k)^{-1/2}$. Soient ε et γ des réels de $]0,1[$.

Soit $\mathscr{C}_1, \mathscr{C}_2, \ldots \mathscr{C}_N$ un système de grands cercles, maximal pour la propriété $i\neq j\Rightarrow d(\mathscr{C}_i,\mathscr{C}_j)\geqq\delta$. Soient $P_1, P_2, \ldots, P_N$ des points respectivement de $\mathscr{C}_1, \mathscr{C}_2, \ldots \mathscr{C}_N$.

Posons $f(z)=\sum_{l=1}^{N} f_k(\langle z, P_l\rangle)$, où f_k est la fonction définie en III 1.1 pour les paramètres $\varepsilon_2=2^{-12n}\gamma$ et $\varepsilon_1=\varepsilon/(C+2^{12n})$ où C est la constante:

$$C=4^{2n-2}+2^{6n}\sum_{s=0}^{\infty}2^{6sn}/e^{4s/8}.$$

Majoration de $\sum_{l=1}^{N}|\langle z, P_l\rangle|^k$ *pour* $z\in S$

Soit z un point fixé de S. Pour $s=1, 2, \ldots, s_0=[-\mathrm{Log}_2\,\delta]$ posons

$$A_s=\{l;|\langle z, P_l\rangle|^2\geqq 1-2^{2s}\delta^2\}. \qquad (A_s=A_s(z))$$

Alors

$$l\in A_s\Leftrightarrow|\langle z, P_l\rangle|^2\geqq 1-2^{2s}\delta^2\Leftrightarrow 1-|\langle z, P_l\rangle|^2\leqq c^{2s}\delta^2\Leftrightarrow d(z, P_l)\leqq 2^s\delta$$

Or $\sigma(\{Z; d(z, Z)\leqq t\})=(t)^{2n-2}$ pour $t\leqq 1$ (voir, par exemple Rudin [7]). d vérifiant l'inégalité triangulaire, les ensembles $\{Z; d(Z, P_l)<\delta/2\}_{l\in A_s}$ sont disjoints, de mesure $(\delta/2)^{2n-2}$ et contenus dans l'ensemble $\{Z: d(Z, z)\leqq(2^{s+1}+1)\,\delta/2\}$.

Le cardinal de A_s est donc inférieur ou égal à $(2^{s+1}+1)^{2n-2}$.

Donc $\mathrm{card}(A_0)\leqq 4^{2n-2}$ et pour $s>0$, $\mathrm{card}(A_s)\leqq E^s$ en posant $E=2^{6n}$.

D'où

$$\sum_{l=1}^{N}|\langle z, P_l\rangle|^k\leqq 4^{2n-2}+\sum_{s=1}^{s_0}(1-2^{2(s-1)}\delta^2)^{k/2}\,\mathrm{card}(A_s)$$

$$\leqq 4^{2n-2}+\sum_{s=1}^{s_0}(1-2^{2(s-1)}/4k)^{k/2}E^s$$

$$\leqq 4^{2n-2}+E\sum_{s=0}^{\infty}(1-4^s/4k)^{k/2}E^s$$

or $(1-4^s/4k)^{k/2}=e^{(k/2)\,\mathrm{Log}(1-4^s/4k)}<e^{(k/2)(-4^s/4k)}=e^{-4^s/8}$ d'où $\sum_{l=1}^{N}|\langle z, P_l\rangle|^k<C$.

Définition des ensembles W_Z

Soit Z un point de S. Posons $D_Z=\{l; d(Z,P_l)\leqq 4\delta\}$. Nous avons vu que $d(Z,P_l)\leqq 4\delta$ est équivalent à $l\in A_2(Z)$ et donc, card $D_Z\leqq E^2$. Posons $\langle Z,P_l\rangle = r_l e^{i\theta_l}$ (avec $r_l\geqq 0$) et $W_Z=\bigcap_{l\in D_Z}\{e^{-i\theta_l}E_k Z\}$ où E_k est le sous ensemble du cercle unité défini dans le paragraphe précédent.

W_Z est de mesure supérieure à $1-E^2\varepsilon_2=1-\gamma$.

III 1.3. Evaluations de f. Démonstration du Lemme 1

Soit z un point de B. Alors

$$|f(z)|=\left|\sum_{l=1}^{N}\langle z,P_l\rangle^k h(\langle z,P_l\rangle^k)\right|\leqq \operatorname*{Max}_{\bar{D}}(h)\cdot|z|^k \operatorname*{Max}_{S}\sum_{l=1}^{N}|\langle\zeta,P_l\rangle|^k$$
$$\leqq M_1 C|z|^k$$

(Remarquons que M_1 dépend de ε_1 et ε_2 et donc en fait de ε et γ mais bien sur pas de k).

Soient Z un point de S et z un point de W_Z, nous allons évaluer $f(rz)$.

$$f(rz)=\sum_{l\in D_Z} f_k(\langle rz,P_l\rangle)+\sum_{l\notin D_Z} f_k(\langle rz,P_l\rangle)$$

Si $l\notin D_Z$, alors $d(Z,P_l)>4\delta$ et $|\langle Z,P_l\rangle|^2<1-2^4\delta^2$, ce qui entraine

$$|\langle z,P_l\rangle|<1-2^2\delta^2=1-1/k,$$

et d'après les propriétés de h (voir III 1.1)

$$|f_k(\langle rz,P_l\rangle)|\leqq\varepsilon_1|\langle rz,P_l\rangle|^k.$$

Alors $|\sum_{l\notin D_Z} f_k(\langle rz,P_l\rangle)|\leqq Cr^k\varepsilon_1<C\varepsilon_1$.

Remarquons que, dans cette majoration, nous n'avons pas utilisé l'appartenance de z à W_Z.

Evaluons maintenant $f_k(\langle rz,P_l\rangle)$ pour l appartenant à D_Z.

z appartient à W_Z, il existe donc un élément z_l de E_k tel que $z=e^{-i\theta_l}z_l Z$.

Alors $\langle rz,P_l\rangle = re^{-i\theta_l}z_l\langle Z,P_l\rangle = rr_l z_l$ et, $f_k(\langle rz,P_l\rangle)=(rr_l)^k z_l^k h((rr_l z_l)^k)$. Mais $|h((rr_l z_l)^k)-c((rr_l)^k)\bar{z}_l^k|<\varepsilon_1$ et $|f_k(\langle rz,P_l\rangle)-(rr_l)^k c((rr_l)^k)<\varepsilon_1$.

En combinant les 2 évaluations nous obtenons:

$$|f(rz)-\sum_{l\in D_Z}(rr_l)^k c((rr_l))|<(C+E^2)\varepsilon_1=\varepsilon$$

Définissons w_1 sur $[0,1[$. $\mathscr{C}_Z$ par $w_1(rz)=\sum_{l\in D_Z}(rr_l)^k c((rr_l)^k)$.

Nous venons de montrer que si $z\in W_Z$, alors $|f(rz)-w_1(rz)|<\varepsilon$.

Or pour tout r, $c(r)\in[0,1]$; w_1 envoie donc la boule dans $[0,E^2]$. Mais le système de cercles étant maximal, il existe un indice l_0 tel que $d(Z,P_{l_0})<\delta$,

qui est équivalent à $1-r_{l_0}^2<\delta^2=1/4k$; alors $r_{l_0}>1-1/4k$. Supposons de plus que $r>1-1/4k$, alors $rr_{l_0}>1-1/2k$, ce qui entraine $c((rr_{l_0})^k)=1\,\mathrm{car}(rr_{l_0})^k>b$, et finalement $w(rz)\in[b, E^2]$.

Considérons maintenant la fonction g de $A(B)$ définie par $g=f/2E^2$. Il est clair que la fonction g vérifie le Lemme 1. (La constante Δ étant égale à $b/2E^2$).

III 2. Approximation d'une fonction continue sur S-Démonstration de la Proposition I

Soient φ une fonction continue sur S et γ et ν des réels de $]0, 1[$.

D'après le Lemme 0, il existe pour tout k assez grand, une fonction f de $A(B)$ vérifiant

(1) $|z|<1-1/k\Rightarrow|f(z)|<\nu/2$

(2) Pour tout point Z de S, il existe un sous-ensemble W_Z' de $\mathscr{C}_Z$ de mesure supérieure à $1-\gamma/2$ tel que

si $z\in W_Z'$; $\quad|f(rz)|<|\varphi(z)|+\nu/2$ pour $r\in[0, 1[$

$\quad|f(rz)-\varphi(z)|<c_1|\varphi(z)|+\nu/2$ pour $r>1-1/4k$

Posons $M'=\underset{\bar{B}}{\mathrm{Max}}|f|+2$.

D'après le Lemme 1, il existe un réel M, tel que pour tout k assez grand, il existe une fonction g de $A(B)$ vérifiant

(3) $|g(z)|\leqq M|z|^k$

(4) Pour tout point Z de S, il existe un sous-ensemble W_Z'' de $\mathscr{C}_Z$ de mesure supérieure à $1-\gamma/2$ tel que: $z\in W_Z''\Rightarrow|g(rz)-w(rz)|<\nu/2M'$.

Remarquons que M' dépend de γ et ν, que M dépend de M' et donc de γ et ν. M et M' sont par contre indépendants de k. Par contre g dépend bien sûr de k de même que les ensembles W_Z''.

Posons $h=fg$. Alors $|h(z)|\leqq MM'|z|^k$ et la première propriété de la Proposition I est vérifiée.

Soit Z un point de S; Posons: $W_Z=W_Z'\cap W_Z''$. Alors W_Z est de mesure supérieure à $1-\gamma$.

Soit $z\in W_Z$.

Comme $z\in W_Z''$, $\quad|g(rz)|\leqq|w(rz)|+\nu/M'\leqq 1/2+1/2=1$,

et $|h(rz)|\leqq|f(rz)|<|\varphi(z)|+\nu$ (d'après (2), car $z\in W_Z'$) et, si $r<1-1/k$, $|h(rz)|<\nu$ d'après (1).

Supposons maintenant $r>1-1/4k$, alors $g(rz)=w(rz)+R(rz)$ où $w(rz)\in[\Delta, 1/2]$ et $|R(rz)|<\nu/2M'$

$$|h(rz)-\varphi(z)|\leqq w(rz)\,|f(rz)-\varphi(z)|+(1-w(rz))\,|\varphi(z)|+|f(rz)|\,|R(rz)|$$

Mais $z\in W_Z'$ et $r>1-1/4k$ entraînent $|f(rz)-\varphi(z)|<c_1|\varphi(z)|+\nu/2$.

Alors

$$\begin{aligned}|h(rz)-\varphi(z)|&\leqq|\varphi(z)|\,(1-w(rz)\,(1-c_1))+w(rz)\cdot\nu/2+|f(rz)|\cdot\nu/2M'\\&\leqq c|\varphi(z)|+\nu,\qquad\text{avec } c=1-\Delta(1-c_1)<1\end{aligned}$$

ce qui achève la démonstration de la Proposition I.

IV. Démonstration du théorème

IV 1. Le lemme fondamental

Des applications successives de la Proposition I conduisent au résultat suivant:

Lemme 2. *Soit φ une fonction continue sur S. Soient α, η et ε des réels de $]0,1[$ et ω un ordre de croissance. Alors il existe une fonction h de $A(B)$ et deux réels β et β', $\alpha<\beta<\beta'<1$ tels que:*

(1) $|h(z)|\leqq w(|z|)$

(2) *Pour tout point Z de S, il existe un sous-ensemble W_Z de $\mathscr{C}_Z$, de mesure supérieure à $1-\eta$ tel que*

$$\begin{array}{lll} \text{si } z\in W_Z; & |h(rz)|<\varepsilon & \text{pour } r\in[0,\beta[\\ & |h(rz)-\varphi(z)|<|\varphi(z)|+\varepsilon & \text{pour } r\in[\beta,\beta'[\\ & |h(rz)-\varphi(z)|<\varepsilon & \text{pour } r\in[\beta',1[. \end{array}$$

Construction de h

Soit N un entier vérifiant $c^{N-1}\max_S|\varphi|<\varepsilon/3=\varepsilon'$; Nous construisons h sous la forme: $h=\sum_{l=1}^{N} f_l$, où (f_l) est une suite de fonctions vérifiant la Proposition I, définie par récurrence.

f_1 vérifie la Proposition 1, pour la fonction $\varphi_1=\varphi$, les paramètres $\gamma_1=\eta/2$ et $\nu_1=\varepsilon'/2$, k_1 étant choisi assez grand pour que f_1 soit de croissance inférieure à $\omega_1=\omega/2$ et que α soit inférieur à $1-1/k_1$. Posons $\delta_1=1/k_1$ et $\delta'_1=1/4k_1$ et $W_{Z,1}$ les sous ensembles de $\mathscr{C}_Z$ correspondants.

Supposons construites les fonctions de $A(B)$, f_s pour $s=1,2,\ldots l$; (Les ensembles et réels correspondants étant notés $W_{Z,s}$ et δ_s, δ'_s.) Ces fonctions f_s étant uniformément continues sur $\bar{B}$, il existe un réel R_l tel que:

$$r\geqq R_l\Rightarrow\forall s=1,2,\ldots l,\quad \forall z\in S,\quad |f_s(rz)-f_s(z)|<\varepsilon'/2^l$$

f_{l+1} vérifie la Proposition I pour la fonction

$$\varphi_{l+1}=\varphi_l-f_l=\varphi-\sum_{s=1}^{l} f_s,$$

les paramètres $\gamma_{l+1}=\eta/2^{l+1}$ et $\nu_{l+1}=\varepsilon'/2^{l+1}$; k_{l+1} étant choisi assez grand pour que $\delta_{l+1}=1/k_{l+1}<\delta'_l$, $1-\delta_{l+1}>R_l$ et que f_{l+1} soit de croissance inférieure à $\omega_{l+1}=\omega/2^{l+1}$.

Démonstration du lemme.
Par construction, $|h(z)|\leqq\sum_{l=1}^{N}|f_l(z)|\leqq\sum_{l=1}^{N}\omega_l(|(z)|)<\omega(|z|)$.

Soit $\mathbb{Z}$ un point de S.

Posons $W_Z=\bigcap_{l=1}^{N} W_{Z,l}$. Chaque $W_{Z,l}$ étant de mesure supérieure à $1-\eta/2^l$, W_Z est de mesure supérieure à $1-\eta$.

Posons $\beta=1-\delta_1$ et $\beta'=1-\delta_N$.

Soit z un point de W_Z; évaluons $h(rz)$ pour les différentes valeurs de r.
Si $r<\beta$, pour tout $l=1, 2, \ldots N$; $r<1-\delta_l$ et $|f_l(rz)|<v_l=\varepsilon'/2^l$, et,

$$|h(rz)|<\varepsilon'<\varepsilon.$$

Si $\beta\leqq r<\beta'$; il existe un indice unique l_0 tel que $1-\delta_{l_0}\leqq r<1-\delta_{l_0+1}$.
Alors

$$|h(rz)-\varphi(z)|\leqq\sum_{l=1}^{l_0-1}|f_l(rz)-f_l(z)|+|f_{l_0}(rz)-\varphi_{l_0}(z)|+\sum_{l=l_0+1}^{N}|f_l(rz)|$$

or

$$r\geqq 1-\delta_{l_0}\Rightarrow\forall l=1, \ldots l_{0-1};\quad |f_l(rz)-f_l(z)|<\varepsilon'/2^l \text{ car } 1-\delta_{l_0}>R_{l_0-1}$$

et

$$r<1-\delta_{l_0+1}\Rightarrow\forall l\geqq l_{0+1};\quad |f_l(rz)|<\varepsilon'/2^l$$

donc

$$|h(rz)-\varphi(z)|\leqq|f_{l_0}(rz)-\varphi_{l_0}(z)|+\varepsilon'\sum_{l\neq l_0}1/2^l$$

mais puisque

$$r\geqq 1-\delta_{l_0},\quad |f_{l_0}(rz)-\varphi_{l_0}(z)|\leqq|\varphi_{l_0}(z)|+\varepsilon'/2^{l_0}$$

et

$$|h(rz)-\varphi(z)|<|\varphi_{l_0}(z)|+\varepsilon'$$

or

$$\begin{aligned}|\varphi_{l_0}(z)|&=|\varphi_{l_0-1}(z)-f_{l_0-1}(z)|<c|\varphi_{l_0-1}(z)|+\varepsilon'/2^{l_0-1}\\&<c[c|\varphi_{l_0-2}(z)|+\varepsilon'/2^{l_0-2}]+\varepsilon'/2^{l_0-1}=c^2|\varphi_{l_0-2}(z)|+c\varepsilon'/2^{l_0-2}+\varepsilon'/2^{l_0-1}\\&\ldots\\&<\varepsilon'(2^{1-l_0}+c2^{2-l_0}+\ldots+c^{l_0-2}2^{-1})+c^{l_0-1}|\varphi(z)|<\varepsilon'+c^{l_0-1}|\varphi(z)|\end{aligned}$$

et finalement $|h(rz)-\varphi(z)|<|\varphi(z)|+2\varepsilon'<|\varphi(z)|+\varepsilon$

Enfin si $r>\beta'$; $|h(rz)-\varphi(z)|\leqq\sum_{l=1}^{N-1}|f_l(rz)-f_l(z)|+|f_N(rz)-\varphi_N(z)|$ et comme precedemment

$$|h(rz)-\varphi(z)|\leqq 2\varepsilon'+c^{N-1}|\varphi(z)|<3\varepsilon'=\varepsilon$$

ce qui achève la démonstration du lemme.

IV 2. Construction de la fonction f

Soient φ une fonction continue dans B et ω un ordre de croissance. Nous allons construire f sous la forme $f=\sum_{k=1}^{\infty}f_k$. Les fonctions f_k sont obtenues de façon recurrente, par application du Lemme 2 sur des blocs d'indices consecutifs $I_l=]s_l, s_{l+1}]$ où elles rendent compte du comportement de φ sur l'ensemble $\{|z|\in[1-1/2^{l-1}, 1-1/2^l[=J_l\}$.

Pour r appartenant à $]0, 1[$, definissons sur S les fonctions φ_r par $\varphi_r(z)=\varphi(rz)$.

Le Lemme 2 est appliqué à des fonctions proches de $\varphi_r-\varphi_{r'}$ pour r et r' dans J_l. Le controle donné par le Lemme 2 est de l'ordre de $|\varphi_r-\varphi_{r'}|$ qui peut

ne pas être petit, c'est ce qui nous amène à découper cet intervalle en N_l morceaux $(S_{l+1}=S_l+N_l)$ tels que sur chacun, $|\varphi_r-\varphi_{r'}|$ soit petit.

Posons $I_1=1$; f_1 est une fonction vérifiant le Lemme 2 pour la fonction $\varphi_{1/2}$, les paramètres $\alpha=\eta=\varepsilon=1/2$ et l'ordre de croissance $\omega/2$. Les ensembles W_Z et les réels correspondants sont notés W_Z^1, β_1 et β'_1.

Supposons construites les fonctions de $A(B)$ f_s sur $I_0, I_1, \ldots, I_{l-1}$. (Les ensembles W_Z et les réels β et β' correspondants étant notés W_Z^s, β_s et β'_s). L'uniforme continuité de φ sur $(1-1/2^l)\,\bar{B}$ entraine qu'il existe un entier N_l tel que, pour tout point z de S et r et r' dans $[0, 1-1/2^l]$;

$$|r-r'|<1/2^l N_l \Rightarrow |\varphi(rz)-\varphi(r'z)|<1/2^l.$$

Posons alors $S_{l+1}=S_l+N_l$.

Soit k un indice de $[S_l, S_{l+1}[$. Supposons construites les fonctions f_s de $A(B)$ pour $s\leqq k$. Ces fonctions étant absolument continues sur $\bar{B}$, il existe un réel $R_k<1$ tel que:

$$\forall s=1,2,\ldots,k; \quad r>R_k \Rightarrow \forall z\in S, \quad |f_s(rz)-f_s(z)|<1/2^k.$$

Choisissons la fonction f_{k+1} satisfaisant le Lemme 2 pour la fonction $\varphi_{r_k}-\sum_{s=1}^{k} f_s$ (où $r_k=1-1/2^{l-1}+(k+1-S_l)/2^l N_l$), l'ordre de croissance $\omega/2^{k+1}$, les paramètres $\eta=\varepsilon=1/2^{k+1}$ et $\alpha=\mathrm{Max}(\beta'_k, R_k)$. (Alors β_{k+1} est supérieur à β'_k et à R_k.)

IV 3. Estimations de f. Fin de la démonstration du théorème

Les ordres de croissance imposés aux fonctions f_k entrainent la convergence uniforme de la série sur tout compact de B. f est holomorphe et de croissance inférieure à ω.

Soit Z un point de S. Posons $W_Z=\underline{\lim}\, W_Z^s$: W_Z est de mesure 1.

Pour tout entier $k\geqq 1$, il existe un indice l et un seul tel que $k\in[S_l, S_{l+1}[$.

Posons $r'_k=1-1/2^{l-1}+(k-S_l/2^l N_l)$; $r'_{k+1}=r'_k+1/2^l N_l$ et $r'_{k-1}=r'_k-1/2^l N_l$ si $k>S_l$ et $r'_{k-1}=r'_k-1/2^{l-1}N_{l-1}$ si $k=S_l$.

Nous allons montrer que si z appartient à W_Z, r à $[\beta_k, \beta_{k+1}[$ et r' à $[r'_k, r'_{k+1}[$ alors $f(rz)$ est proche de $\varphi(r'z)$ pour k assez grand.

z appartient à W_Z entraine: $\exists k_0/\forall k\geqq k_0$, $z\in W_Z^k$.

Choisissons $k>k_0$ et estimons $|f(rz)-\varphi(r'z)|$ pour r et r' dans les intervalles définis ci-avant.

$$|f(rz)-\varphi(r'z)|\leqq \sum_{s=1}^{k-1}|f_s(rz)-f_s(z)|+|\varphi(r'_k z)-\sum_{s=1}^{k-1} f_s(z)-f_k(rz)|$$
$$+|\varphi(r'_k z)-\varphi(r'z)|+\sum_{s=k+1}^{\infty}|f_s(rz)|$$

Or, $r\geqq\beta_k \Rightarrow r>R_{k-1}$ et $\forall s=1,2,\ldots,k-1$, $|f_s(rz)-f_s(z)|\leqq 1/2^{k-1}$. De plus, comme $r<\beta_{k+1}$, la croissance de la suite β_s entraine: $\forall s\geqq k+1$, $r<\beta_s$; Mais pour tout $s\geqq k+1$, $z\in W_Z^s$ et $|f_s(rz)|<1/2^s$.

Enfin, par définition de N_l; $|\varphi(r'_k z)-\varphi(r'z)|<1/2^l$.

Alors

$$|f(rz)-\varphi(r'z)| \leqq |\varphi(r'_k z)-\sum_{s=1}^{k-1} f_s(z)-f_k(rz)|+(k-1)/2^{k-1}+1/2^k+1/2^l$$

Mais $z \in W_Z^k$, alors

$$|\varphi(r'_k z)-\sum_{s=1}^{k-1} f_s(z)-f_k(rz)| < |\varphi(r'_k z)-\sum_{s=1}^{k-1} f_s(z)|+1/2^k$$

Or comme z appartient aussi à W_Z^{k-1}:

$$|\varphi(r'_{k-1} z)-\sum_{s=1}^{k-1} f_s(z)| < 1/2^{k-1}$$

et donc

$$|\varphi(r'_k z)-\sum_{s=1}^{k-1} f_s(z)| < 1/2^{k-1}+|\varphi(r'_{k-1} z)-\varphi(r'_k z)| \leqq 1/2^{k-1}+1/2^{l-1}$$

($1/2^{l-1}$ peut être remplacé par $1/2^l$ si $k > S_l$).

Quand k tend vers l'infini, l tend aussi vers l'infini et $f(rz)-\varphi(r'z)$ tend vers 0, ce qui achève la démonstration du théorème en prenant pour σ, la fonction affine par morceaux envoyant $[r'_k, r'_{k+1}[$ sur $[\beta_k, \beta_{k+1}[$.

References

1. Bagemihl, F., Seidel, W.: Some boundary properties of analytic functions. Math. Z. **61**, 186–199 (1954)
2. Dupain, Y.: Limites radiales de fonctions holomorphes dans la boule unité de $\mathbb{C}^n$. J. d'Anal. Math. **52**, 107–116 (1989)
3. Hakim, M., Sibony, N.: Boundary properties of analytic functions in the ball of $\mathbb{C}^n$. Math. Ann. **276**, 549–555 (1987)
4. Iordan, A.: On the radial cluster sets for holomorphic functions in the unit ball of $\mathbb{C}^n$. Prépublication Orsay 1988
5. Kahane, J.P., Katznelson, Y.: Sur le comportement radial des fonctions analytiques. C.R. Acad. Sci., Paris, Ser. A, **372**, 718–719 (1971)
6. Ortel, M., Schneider, W.: Radial limits of functions of slow growth in the unit disk. Math. Scand. **56**, 287–310 (1985)
7. Rudin, W.: New constructions of functions holomorphic in the unit ball of $\mathbb{C}^n$. NSF-CBMS Regional Conference N 63, Michigan State University 1986

Math Z. 206, 81–88 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Sobolev estimates for the $\bar{\partial}$-Neumann operator on domains in C^n admitting a defining function that is plurisubharmonic on the boundary*

Harold P. Boas and Emil J. Straube

Department of Mathematics, Texas A&M University, College Station, TX 77843, USA

Received June 11, 1990

We present here a technique for proving Sobolev estimates for the Bergman projection and the $\bar{\partial}$-Neumann operator that applies to every smooth bounded (weakly) convex domain in $\mathbf{C}^n$, and more generally to every smooth bounded (weakly) pseudoconvex domain in $\mathbf{C}^n$ admitting a defining function that is plurisubharmonic on the boundary. Rather than state a general technical theorem, we give the results directly for these important classes of domains. We discuss some extensions in the remarks at the end of the paper, where we also discuss related work.

For background about the $\bar{\partial}$-Neumann problem and the Bergman projection, see [22, 26, 29, 31]. We recall that the $\bar{\partial}$-Neumann operator N_q is the inverse of the complex Laplacian $\bar{\partial}^*\bar{\partial}+\bar{\partial}\bar{\partial}^*$ on $(0, q)$ forms when $1 \leqq q \leqq n$. When $0 \leqq q \leqq n$, the Bergman projection P_q is the orthogonal projection from the space of square-integrable $(0, q)$ forms onto the subspace of $\bar{\partial}$-closed $(0, q)$ forms; P_0 is the usual Bergman projection on functions.

We recall that a smooth bounded domain Ω in $\mathbf{C}^n$ in pseudoconvex if the complex Hessian of any smooth defining function for Ω is positive semi-definite on the complex tangent space at each boundary point. We say that Ω has a defining function that is plurisubharmonic on the boundary if some smooth defining function has its complex Hessian positive semi-definite on all vectors at each boundary point. (We require plurisubharmonicity only *on* the boundary, and *not* in a neighborhood of the boundary.)

Theorem. *Let Ω be a smooth bounded domain in $\mathbf{C}^n$ admitting a defining function that is plurisubharmonic on the boundary. Then*
(1) *when $1 \leqq q \leqq n$, the $\bar{\partial}$-Neumann operator N_q is continuous on the space of $(0, q)$ forms with coefficients in $W^s(\Omega)$ for every nonnegative real number s; and*
(2) *when $0 \leqq q \leqq n$, the Bergman projection P_q is continuous on the space of $(0, q)$ forms with coefficients in $W^s(\Omega)$ for every nonnegative real number s.*

Here $W^s(\Omega)$ denotes the usual Sobolev space of order s. Of course, the Sobolev lemma implies as usual that the operators are continuous on the spaces of forms with coefficients in $C^\infty(\bar{\Omega})$. We have shown earlier [10] that the two parts of the theorem are equivalent; we shall prove (2) below.

* This paper is a revision (May 1990) of our preprint formerly titled "Sobolev estimates for the $\bar{\partial}$-Neumann operator on convex domains in $\mathbf{C}^n$."

Remark. The theorem applies, in particular, to convex domains, for every smooth bounded convex domain admits a plurisubharmonic defining function. Indeed, the so-called Minkowski functional or gauge function $\mu_\Omega(z) := \inf\{\lambda > 0 : z \in \lambda\Omega\}$ is convex [32, p. 35] and *a fortiori* plurisubharmonic, and it is easy to see that it is smooth when Ω is.

In view of work of Bell and Catlin [6], Diederich and Fornæss [21], and Bell [4, 5], our theorem immediately gives boundary smoothness results for holomorphic mappings when one of the domains involved admits a defining function that is plurisubharmonic on the boundary. For example, from [6] or [21] we get the following.

Corollary. *Every proper holomorphic mapping from a smooth bounded domain in* $\mathbf{C}^n$ *admitting a defining function that is plurisubharmonic on the boundary onto another smooth bounded pseudoconvex domain in* $\mathbf{C}^n$ *extends smoothly to the boundary of the first domain.*

In particular, biholomorphic mappings between smooth bounded domains in $\mathbf{C}^n$ extend smoothly to the boundaries if one of the domains is convex. For further discussion of proper mappings, see [3] and [23].

Of course, the theorem also has immediate consequences for the theory of the $\bar\partial$-equation on domains admitting a defining function that is plurisubharmonic on the boundary. In particular, the canonical Kohn solution orthogonal to holomorphic functions of the equation $\bar\partial u = \alpha$ admits the estimate $\|u\|_{W^s(\Omega)} \leqq C_s \|\alpha\|_{W^s(\Omega)}$ for every nonnegative real number s, with C_s independent of α.

We thank Alan Noell, whose paper [30] was a stimulating influence. His observation that the unit normal to the boundary of a convex domain is constant along Levi-null curves was the key to our construction of certain special vector fields (analogous to the ones in the Lemma below) in the case when Ω is convex. Our present construction is a variant.

Proof of the theorem. As mentioned above, the two parts of the theorem are equivalent [10], and we shall prove (2). By standard interpolation theory, it is enough to prove continuity in $W^k(\Omega)$ when k is a positive integer.

The proof is by a downward induction on the degree q. The fundamental formula relating the projection to the $\bar\partial$-Neumann operator is $P_{q-1} = I - \bar\partial^* N_q \bar\partial$. The basis step of the induction (the case $q = n$) holds because in top degree the $\bar\partial$-Neumann boundary conditions reduce to ordinary Dirichlet boundary conditions [22, p. 63]. Consequently N_n gains two derivatives, and so when $k \geqq 0$ the projection $P_{n-1} = I - \bar\partial^* N_n \bar\partial$ satisfies the estimate $\|P_{n-1} f\|_k \leqq C \|f\|_k$ for every $(0, n-1)$-form f.

Now fix a positive integer k. We suppose that $1 \leqq q \leqq n-1$, and we make the induction hypothesis that P_{r-1} is continuous on $W^k(\Omega)$ when $r > q$. We wish to prove the estimate

$$\|P_{q-1} f\|_k \leqq C_k \|f\|_k \tag{1}$$

with C_k independent of the $(0, q-1)$ form f. (For simplicity of notation, we will henceforth drop the index k on the constant C.) We will first prove this estimate under the assumption that both f and $P_{q-1} f$ have coefficients in $C^\infty(\bar\Omega)$. At the end of the proof we will show how to remove this *a priori* assumption.

Let ρ be a smooth defining function for Ω that is plurisubharmonic on the boundary, and let L denote the complex normal vector field defined in a neighborhood of the boundary $b\Omega$ by

$$L := \frac{4}{|V\rho|^2} \sum_{j=1}^{n} \frac{\partial \rho}{\partial \bar{z}_j} \frac{\partial}{\partial z_j}.$$

Then $L\rho = 1$ in a neighborhood of the boundary, and $L - \bar{L}$ is the special tangential direction transverse to the complex tangent space of the boundary. It will be enough to estimate derivatives of $P_{q-1} f$ in this special direction for the following reason.

First observe that since the complex $\bar{\partial} \oplus \vartheta$ is elliptic (here ϑ is the formal adjoint of $\bar{\partial}$), and the boundary is noncharacteristic, derivatives of order k of components of $P_{q-1} f$ can be written in terms of derivatives of order $k-1$ of the components of $\bar{\partial} P_{q-1} f$ (which is zero), of $\vartheta P_{q-1} f$ (which is ϑf), and of $P_{q-1} f$ itself; and of derivatives of order k, tangential at the boundary, of the components of $P_{q-1} f$. Now derivatives of type (0, 1) are controlled by the estimate

$$\sum_{j,J} \left\| \frac{\partial u_J}{\partial \bar{z}_j} \right\|_{k-1}^2 \leqq C(\|\bar{\partial} u\|_{k-1}^2 + \|\bar{\partial}^* u\|_{k-1}^2 + \|u\|_{k-1}^2) \tag{2}$$

for u in the common domain of $\bar{\partial}$ and $\bar{\partial}^*$, and if Y is a vector field of type (1, 0) such that $Y\rho = 0$ on the boundary, then

$$\|Yu\|_{k-1}^2 \leqq C(\|\bar{\partial} u\|_{k-1}^2 + \|\bar{\partial}^* u\|_{k-1}^2 + \|u\|_{k-1} \|u\|_k). \tag{3}$$

Consequently (since $(I - P_{q-1}) f$ is in the common domain of $\bar{\partial}$ and $\bar{\partial}^*$),

$$\|P_{q-1} f\|_k^2 \leqq C(\|(L - \bar{L})^k P_{q-1} f\|_0^2 + \|f\|_k^2 + \|P_{q-1} f\|_{k-1}^2). \tag{4}$$

For the reader's convenience, we sketch the derivation of estimates (2) and (3) (see also [7, pp. 472–473]). The case $k - 1 = 0$ of (2) is standard (see [22] and [24, Prop. 2.1.2]), and the corresponding case of (3) follows from it by integrating twice by parts. Both estimates lift to higher derivatives by using again that the boundary is noncharacteristic for $\bar{\partial} \oplus \vartheta$ and by working in a special boundary chart [see 22, p. 33].

If X is any smooth vector field of type (1, 0) such that $X\rho = 1$ on $b\Omega$, then $L - X$ is complex tangential, so estimates (2)–(4) imply

$$\|P_{q-1} f\|_k^2 \leqq C \|(X - \bar{X})^k P_{q-1} f\|_0^2 + C_X(\|f\|_k^2 + \|P_{q-1} f\|_{k-1}^2),$$

with the first constant independent of X. We claim that a field X can be chosen that has good commutation properties with $\bar{\partial}$.

Lemma. *Let Ω be a smooth bounded domain in $\mathbf{C}^n$ admitting a defining function ρ that is plurisubharmonic on the boundary. For every positive ε there exists a vector field X of type (1, 0) with coefficients in $C^\infty(\bar{\Omega})$ such that $X\rho = 1$ on $b\Omega$ and such that when $1 \leqq j \leqq n$, the Hermitian inner product of the commutator $[X, \partial/\partial \bar{z}_j]$ with the complex normal L has modulus less than ε on the boundary.*

This lemma is the only point in our argument where the plurisubharmonicity of the defining function – rather than the pseudoconvexity of the domain – is used. Our proof is consequently valid in any pseudoconvex domain for which one can construct an appropriate vector field. We postpone the proof of the lemma and proceed with the estimates for P_{q-1}.

The value of ε will be specified later; for now we view ε as fixed. Let φ be a smooth cutoff function that is equal to one in a neighborhood of the boundary and that has support so thin that the commutator property of the lemma holds (by continuity) on the support of φ. Because of interior elliptic regularity, it will suffice to estimate $\|\varphi(X-\bar{X})^k P_{q-1} f\|_0^2$.

Writing this as an inner product, replacing P_{q-1} in one factor by $I-\bar{\partial}^* N_q \bar{\partial}$, and integrating by parts k times (using that $X-\bar{X}$ is tangential at the boundary, so that no boundary terms arise) we get

$$-\langle \bar{\partial}^* N_q \bar{\partial} f, \varphi^2 (X-\bar{X})^{2k} P_{q-1} f\rangle + O(\|P_{q-1} f\|_k (\|f\|_k + \|P_{q-1} f\|_{k-1})).$$

In the main term, we move the $\bar{\partial}^*$ to the other side of the inner product as $\bar{\partial}$ and expand the commutator $[\bar{\partial}, (X-\bar{X})^{2k}]$ as in [19, Lemma 2, p. 418] to obtain

$$-2k\langle N_q \bar{\partial} f, \varphi^2 (X-\bar{X})^{2k-1} [\bar{\partial}, X-\bar{X}] P_{q-1} f\rangle$$
$$+ O(\|N_q \bar{\partial} f\|_{k-1} \|P_{q-1} f\|_k + \|f\|_k^2).$$

(Note that we can always integrate up to $k-1$ tangential derivatives back to the left-hand side before applying the Cauchy-Schwarz inequality.)

In view of the lemma, the commutator $[\bar{\partial}, X-\bar{X}]$ can be written on the support of φ as the sum of derivatives of type (1, 0) that are complex tangential at the boundary, arbitrary derivatives of type (0, 1), and a function of modulus less than ε times L. Also, replacing $(X-\bar{X})^{2k-1}$ by $(L-\bar{L})^{2k-1}$ makes an error involving at least one complex tangential derivative. Therefore, using estimates (2) and (3) and the inequalities $\|u\|_{k-1} \leqq \varepsilon \|u\|_k + C_\varepsilon \|u\|_0$ and $|ab| \leqq \delta |a|^2 + (4\delta)^{-1} |b|^2$, we obtain the estimate

$$\|P_{q-1} f\|_k^2 \leqq \varepsilon C (\|N_q \bar{\partial} f\|_k^2 + \|P_{q-1} f\|_k^2) + C_\varepsilon \|f\|_k^2$$

with the first constant independent of ε (as a result of replacing $X-\bar{X}$ by $L-\bar{L}$ in the main term).

In a previous paper [10], we showed how to relate $N_q \bar{\partial}$ to P_{q-1}, P_q, and Kohn's weighted $\bar{\partial}$-Neumann operator $N_{t,q}$. (We recall that for any smooth bounded pseudoconvex domain, Kohn showed [25] global regularity in W^k of the $\bar{\partial}$-Neumann problem set up with respect to the weight $w_t(z) := \exp(-t|z|^2)$ when t is sufficiently large, depending on k.) Indeed,

$$\bar{\partial}^* N_q = (I - P_{q-1})(\bar{\partial}_t^* N_{t,q}) P_q,$$

from which it follows that

$$N_q \bar{\partial} = P_q w_t N_{t,q} \bar{\partial} [w_{-t} (I - P_{q-1})].$$

By the induction hypothesis, P_q is continuous on forms with coefficients in $W^k(\Omega)$, and using Kohn's theory we may assume that t is so large that the weighted

operator $N_{t,q}\bar{\partial}$ is too. (Only the continuity of $N_{t,q}$ is explicit in [25], but the continuity of $N_{t,q}\bar{\partial}$ follows from an argument analogous to the one in [26, p. 140], with the compactness estimate replaced by [24, Proposition 2.1.2].) Consequently, $\|N_q\bar{\partial}f\|_k \leqq C(\|P_{q-1}f\|_k + \|f\|_k)$. Inserting this estimate in the previous one and taking ε sufficiently small gives the required *a priori* estimate (1).

We now remove the assumption that $P_{q-1}f$ is *a priori* smooth. In view of [25, proof of Theorem 3.7], we may choose a number M so large that for all sufficiently small positive δ, the approximating subdomains Ω_δ that have defining functions $\rho_\delta(z) = \rho(z) + \delta \exp(M|z|^2)$ are smooth and strictly pseudoconvex. The above argument can be carried out uniformly on these domains as follows. (We emphasize that we are making no assumption about the plurisubharmonicity of ρ_δ.)

We modify the induction hypothesis by requiring that the estimates hold for the projection of the domain Ω_δ uniformly in δ for sufficiently small positive δ. Uniform estimates for the $\bar{\partial}$-Neumann operator in top degree (the basis step) and for the weighted $\bar{\partial}$-Neumann operator do hold uniformly in δ. The analogue of estimate (4) holds uniformly on Ω_δ for all small δ with the field L replaced by $L_\delta := 4|V\rho_\delta|^{-2}\sum(\partial\rho_\delta/\partial\bar{z}_j)(\partial/\partial z_j)$. At the end of the proof of the lemma, we will indicate how to construct fields X_δ on the approximating domains Ω_δ so that the above estimates are uniform in δ. We conclude by induction that *a priori* estimates hold uniformly in δ for the Bergman projections of the domains Ω_δ.

Of course, the projection $P_{q-1,\delta}$ for the strictly pseudoconvex domain Ω_δ preserves $C^\infty(\bar{\Omega}_\delta)$. Consequently, if f is smooth up to the boundary of Ω, then the *a priori* estimate $\|P_{q-1,\delta}f\|_{W^k(\Omega_\delta)} \leqq C\|f\|_{W^k(\Omega)}$ (uniform in δ) is a real estimate. Passing to the limit as $\delta \to 0$, we deduce that $P_{q-1}f \in W^k(\Omega)$ and $\|P_{q-1}f\|_k \leqq C\|f\|_k$. Since $C^\infty(\bar{\Omega})$ is dense in $W^k(\Omega)$, we conclude that estimate (1) holds as a real estimate on Ω.

Proof of the Lemma. The proof is based on the observation that if ρ is a defining function that is plurisubharmonic on the boundary, then for each j, derivatives of $\partial\rho/\partial z_j$ of type (0, 1) in directions that lie in the null space of the Levi form must vanish. Indeed, if coordinates are chosen so that $\partial/\partial z_1$ is in the null space of the Levi form at a boundary point p, then $(\partial^2\rho/\partial z_1\,\partial\bar{z}_1)(p) = 0$. Since the matrix $\partial^2\rho/\partial z_j\,\partial\bar{z}_k(p)$ is positive semidefinite, its whole first column must be zero, that is, $(\partial/\partial\bar{z}_1)(\partial\rho/\partial z_j)(p) = 0$ for each j.

Fix a point p in the boundary. Since the gradient of ρ does not vanish at p, there is some j for which $\partial\rho/\partial z_j(p) \neq 0$. Choose such a j and set $X_p = (\partial\rho/\partial z_j)^{-1}(\partial/\partial z_j)$ in a neighborhood of p. Let $L_1, \ldots, L_{n-1}$ be a basis near p for the complex tangential vector fields of type (1, 0). We may assume that the Levi form is diagonal at p in this basis, in other words, that $\langle[\bar{L}_k, L_j], L\rangle(p)$ is a diagonal $(n-1)\times(n-1)$ matrix, where the angle brackets denote the standard pointwise Hermitian inner product. If the k^{th} eigenvalue is zero, then by the above observation it follows that $\bar{L}_k$ annihilates the coefficient of X_p at p. Hence the linear system of equations $\sum_{j=1}^{n-1}\langle[\bar{L}_k, L_j], L\rangle(p)\,a_j = \langle[\bar{L}_k, X_p], L\rangle(p)$ is solvable for scalars $a_1, \ldots, a_{n-1}$. This means that the commutator of every complex tangential vector field of type (0, 1) with $X_p - \sum_{j=1}^{n-1} a_j L_j$ has vanishing component in the L direction at p. Since $X_p - L$ is complex tangential

(because $X_p\rho = 1 = L\rho$), there are scalars $b_1, \ldots, b_{n-1}$ such that $X_p - \sum_{j=1}^{n-1} a_j L_j = L - \sum_{j=1}^{n-1} b_j L_j$.

Now fix a positive ε. By continuity, there is a neighborhood of p on which the commutator of every unit complex tangential vector field of type (0, 1) with $L - \sum_{j=1}^{n-1} b_j L_j$ has component in the L direction of magnitude less than ε. Patching the $\sum b_j L_j$ with a partition of unity gives a complex tangential vector field Y of type (1, 0) on the whole boundary such that the commutator of every unit complex tangential vector field of type (0, 1) with $L - Y$ has component in the L direction of magnitude less than ε at every point of the boundary. (Terms in the commutator coming from derivatives of the partition of unity do not contribute to the L component since they are complex tangential.)

Any function u on $b\Omega$ can be extended to $\bar{\Omega}$ in such a way that $\bar{L}u$ vanishes on the boundary: one only has to solve the boundary value problem that twice the real unit normal derivative of u equal the tangential derivative $(\bar{L} - L)u$ on the boundary. Consequently there is a vector field X of type (1, 0) on $\bar{\Omega}$ that matches $L - Y$ on the boundary and in addition satisfies $\langle [\bar{L}, X], L \rangle|_{b\Omega} = 0$. This is the required vector field.

In the approximation argument at the end of the proof of the theorem, we needed corresponding fields X_δ on the approximating domains Ω_δ. To construct these fields uniformly in δ, note that the commutator with $L_\delta - Y$ of every unit complex tangential vector field to the boundary of Ω_δ has small component in the L_δ direction on $b\Omega_\delta$ for δ small. We may also assume that Y annihilates ρ_δ on $b\Omega_\delta$ (by letting $L_1, \ldots, L_{n-1}$ be a basis near p for the complex tangential fields to $b\Omega_\delta$ for all small δ). To obtain X_δ, restrict $L_\delta - Y$ to the boundary of Ω_δ and solve the corresponding boundary value problem in Ω_δ. The estimates in the proof of the theorem can then be done uniformly in δ.

Remarks. (1) The only general result previously known for domains admitting plurisubharmonic defining functions was continuity in $W^{1/2}(\Omega)$ of the Bergman projection and $\bar{\partial}$-Neumann operator [12], [13]. At the same time as we circulated the first version of this paper, which covered only the case of convex domains in $\mathbf{C}^n$, So-Chin Chen [17] independently obtained a proof of the regularity of the $\bar{\partial}$-Neumann problem in convex domains in dimension $n = 2$.

(2) The geometric property that makes the construction of the special vector field X possible is that the normal to $b\Omega$ (or a modification of the normal by a complex tangential field) does not wind wildly along directions which lie in the null space of the Levi form. This property is related to the notion of "nowhere wormlike" that we have introduced in [11] in the context of Hartogs domains in $\mathbf{C}^2$. Once the field X is constructed, the proof uses techniques employed in our earlier papers [7], [9]–[11].

(3) Commutator conditions have been used by other authors in studying the $\bar{\partial}$-Neumann problem. We mention the work of Derridj and Tartakoff [19], Derridj [18], Komatsu [28], and Chen [14]–[16]. Vector fields having nice commutation properties with the Bergman projection were used in [34].

(4) As mentioned above, our arguments also apply when the pseudoconvex domain Ω does not admit a defining function that is plurisubharmonic on the

boundary, if the required vector fields exist for some other reason. Also, the condition $X\rho = 1$ can be relaxed to $X\rho$ real, bounded and bounded away from zero independently of ε. In fact, for the estimation of $\|(X - \bar{X})^k P_{q-1} f\|_0$ it is allowed for $X\rho$ to vanish: the transversality condition is only needed when we estimate $\|P_{q-1} f\|_k$. This means that our arguments apply in the case where we have different transverse vector fields on different patches of $b\Omega$, as long as they are all globally defined and satisfy the commutator condition on the whole boundary. In particular, this covers the case where Ω has "transverse symmetries" (see [1, 34]). Furthermore, it is clear that the field X need exist only near the boundary points of infinite type. Finally, we note that $X\rho$ does not have to be exactly real on the boundary; a smallness condition on the argument of $X\rho$ (depending on the level k in the Sobolev scale) will do (see [2, 9, 11]). Thus one can use approximation theorems to approximate the normal field by fields satisfying the commutator properties (see [9] and [11] in the context of fields with holomorphic coefficients).

(5) Although the technique used in this paper is of considerable generality, it breaks down on the Diederich-Fornæss worm domains [20], for which global regularity of the $\bar{\partial}$-Neumann problem remains an open question.

(6) For completeness, we mention that the theorem holds as well for (p, q) forms, as the index p plays no role in solving the $\bar{\partial}$-equation.

(7) We conclude by observing that our method can also be used to study operators on the boundary. The proof of the theorem can be modified to show that the Szegö projection on the boundary of a smooth bounded pseudoconvex domain in $\mathbf{C}^n$ admitting a defining function that is plurisubharmonic on the boundary is regular in Sobolev norms. Consequently, the canonical solution to the $\bar{\partial}_b$-equation admits Sobolev estimates as well, in view of the solvability results in [33], [27], and [8]. We hope to return to these questions elsewhere.

Acknowledgement. The second author gratefully acknowledges the support and hospitality of the Max-Planck-Institut für Mathematik during the completion of this work.

References

1. Barrett, D.E.: Regularity of the Bergman projection on domains with transverse symmetries. Math. Ann. **258**, 441–446 (1982)
2. Barrett, D.E.: Regularity of the Bergman projection and local geometry of domains. Duke Math. J. **53**, 333–343 (1986)
3. Bedford, E.: Proper holomorphic mappings. Bull. Am. Math. Soc. **10**, 157–175 (1984)
4. Bell, S.R.: Biholomorphic mappings and the $\bar{\partial}$-Neumann problem. Ann. Math. **114**, 103–113 (1981)
5. Bell, S.R.: Boundary behavior of proper holomorphic mappings between non-pseudoconvex domains. Am. J. Math. **106**, 639–643 (1984)
6. Bell, S., Catlin, D.: Boundary regularity of proper holomorphic mappings. Duke Math. J. **49**, 385–396 (1982)
7. Boas, H.P., Chen, S.C., Straube, E.J.: Exact regularity of the Bergman and Szegö projections on domains with partially transverse symmetries. Manuscr. Math. **62**, 467–475 (1988)
8. Boas, H.P., Shaw, M.C.: Sobolev estimates for the Lewy operator on weakly pseudo-convex boundaries. Math. Ann. **274**, 221–231 (1986)
9. Boas, H.P., Straube, E.J.: Complete Hartogs domains in $\mathbf{C}^2$ have regular Bergman and Szegö projections. Math. Z. **201**, 441–454 (1989)
10. Boas, H.P., Straube, E.J.: Equivalence of regularity for the Bergman projection and the $\bar{\partial}$-Neumann operator, Manuscr. Math. **67**, 25–33 (1990)

11. Boas, H.P., Straube, E.J.: The Bergman projection on Hartogs domains in $\mathbf{C}^2$. Trans. Am. Math. Soc., (to appear)
12. Bonami, A., Charpentier, P.: Une estimation Sobolev 1/2 pour le projecteur de Bergman. C. R. Acad. Sci. Paris **307**, 173–176 (1988)
13. Bonami, A., Charpentier, P.: Boundary values for the canonical solution to $\bar{\partial}$-equation and $W^{1/2}$ estimates. (preprint)
14. Chen, S.-C.: Global analytic hypoellipticity of the $\bar{\partial}$-Neumann problem on circular domains. Invent. Math. **92**, 173–185 (1988)
15. Chen, S.-C.: Global regularity of the $\bar{\partial}$-Neumann problem on circular domains. Math. Ann. **285**, 1–12 (1989)
16. Chen, S.-C.: Global regularity of the $\bar{\partial}$-Neumann problem: a sufficient condition. (preprint)
17. Chen, S.-C.: Global regularity of the $\bar{\partial}$-Neumann problem in dimension two. (preprint)
18. Derridj, M.: Regularité pour $\bar{\partial}$ dans quelques domaines faiblement pseudoconvexes. J. Differ. Geom. **13**, 559–576 (1978)
19. Derridj, M., Tartakoff, D.S.: On the global real analyticity of solutions to the $\bar{\partial}$-Neumann problem. Commun. Partial Differ. Equations **1**, 401–435 (1976)
20. Diederich, K., Fornæss, J.E.: Pseudoconvex domains: an example with nontrivial Nebenhülle. Math. Ann. **225**, 275–292 (1978)
21. Diederich, K., Fornæss, J.E.: Boundary regularity of proper holomorphic mappings. Invent. Math. **67**, 363–384 (1982)
22. Folland, G.B., Kohn, J.J.: The Neumann Problem for the Cauchy-Riemann Complex. Ann. Math. Stud. Vol. 75, Princeton: Princeton Univ. Press 1972
23. Forstnerič, F.: Proper holomorphic mappings: a survey. In: J.E. Fornaess (ed.) Several Complex Variables. Proc. Mittag-Leffler Inst. 1987–1988. Princeton: Princeton University Press 1990
24. Hörmander, L.: L^2 estimates and existence theorems for the $\bar{\partial}$ operator. Acta Math. **113**, 89–152 (1965)
25. Kohn, J.J.: Global regularity for $\bar{\partial}$ on weakly pseudo-convex manifolds. Trans. Am. Math. Soc. **181**, 273–292 (1973)
26. Kohn, J.J.: A survey of the $\bar{\partial}$-Neumann problem. In: Proc. Symp. Pure Math, vol. 41, Am. Math. Soc., Providence, 1984, pp. 137–145
27. Kohn, J.J.: The range of the tangential Cauchy-Riemann operator. Duke Math. J. **53**, 525–545 (1986)
28. Komatsu, G.: Global analytic-hypoellipticity of the $\bar{\partial}$-Neumann problem. Tôhoku Math. J. **28**, 145–156 (1976)
29. Krantz, S.G.: Function Theory of Several Complex Variables. New York: Wiley 1982
30. Noell, A.: Local versus global convexity of pseudoconvex domains. (preprint)
31. Range, R.M.: Holomorphic Functions and Integral Representations in Several Complex Variables. Berlin Heidelberg New York: Springer 1986
32. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, N.J. 1970
33. Shaw, M.-C.: L^2 estimates and existence theorems for the tangential Cauchy-Riemann complex. Invent. Math. **82**, 133–150 (1985)
34. Straube, E.J.: Exact regularity of Bergman, Szegö, and Sobolev space projections in non pseudoconvex domains. Math. Z. **192**, 117–128 (1986)

Notes added in proof

(1) David Barrett has recently shown ("Behavior of the Bergman projection on the Diederich-Fornaess worm," preprint) that the conclusion of the theorem does not hold on the worm domains; regularity in the space $C^\infty(\bar{\Omega})$ remains open.

(2) We have proved Sobolev estimates for the inverse of $\Box_b$, for the canonical solution of the $\bar{\partial}$-equation, and for the Szegö projection on the class of domains considered in the theorem ("Sobolev estimates for the complex Green operator on a class of weakly pseudoconvex boundaries", preprint).

Math. Z. 206:89–102 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

On the extension of C.R. functions

E. Amar

Université Bordeaux, U.F.R. Mathématiques et Informatique, 351, Cours de la Libération, F-33405 Talence, France

Received June 12, 1989; in final form March 28, 1990

In this work we study the following question:
Let M be a $\mathscr{C}^\infty$ manifold in $\mathbb{C}^n$ and f a Cauchy-Riemann (C.R. for short) function on M, is it possible to write f as a sum on M of holomorphic functions in wedges of edge M?

Of course, stated in this way, this question asks for an other one: what is a C.R. function, or a C.R. form, on a general manifold M? Hence we must start with the following natural definition:

Definition. Let M be a closed set in $\mathbb{C}^n$ and f a (p, q) form on M; we say that f is C.R. on M if there is an extension F of f in a neighborhood of M in $\mathbb{C}^n$, such that: $\bar{\partial}F$ is flat on M.

In the case M is an analytic set, then a function, i.e. a $(0, 0)$ form, C.R. in this sense is just a holomorphic function [3]; in the case where M is a C.R. manifold i.e. the dimension of the complex tangent space to M is constant, then this definition is equivalent to the classical one as it was proved by Andreotti and Hill [4].

Hence this definition makes a simple link between somewhat different notions of holomorphy. The aim of this paper is also to give simple proofs of the extension property which work in this general frame.

The literature on the extension property for C.R. functions on C.R. manifold is vaste since H. Lewy [18], J.J. Kohn and H. Rossi [17], A. Andreotti and C.D. Hill [4] [5], G. Henkin [14] and more recent papers by J.P. Rosay [20]; M.S. Baouendi, C.H. Chang and F. Trèves [6]; M.S. Baouendi, L.P. Rothschild and F. Trèves [7]; M.S. Baouendi and L.P. Rothschild [8]; J.M. Trépreau [21]. (There is an extensive bibliography in the paper by G. Henkin [14] for the first period and in the nice survey by F. Forstneric [13] for now.)

The ideas here are coming from two sources: the work I have done on extension and division problems for holomorphic functions [1] [2] [3] on one side and the talks I had with Salah Baouendi, Linda Rothschild and François Trèves on classical C.R. functions on the other side and I am happy to thank them for these interesting talks.

As in [2] and [3], the aim is to solve some $\bar{\partial}$-equations in open sets of $\mathbb{C}^n$ with $\mathscr{C}^\infty$ bounds, in order to solve in fact $\bar{\partial}$-equations for currents. In [5],

the authors studied also the case of manifolds of codimension higher than one but their point of view is very different, the results are different and they worked only in the case of C.R. manifolds. In § 1 we prove:

Theorem 1.1: *Let M be a closed set in $\mathbb{C}^n$ such that $0 \in M$ and M cuts the space near 0 in two parts M^- and M^+; if f is a C.R. (p, q) form on M near 0, then there are two (p, q) forms, f^-, $\mathscr{C}^\infty$ on $\bar{M}^-$ near 0 and f^+, $\mathscr{C}^\infty$ on $\bar{M}^+$ near 0, such that: $\bar{\partial} f^- = 0$ on M^-, $\bar{\partial} f^+ = 0$ on M^+ and $f = f^- + f^+$ on M near 0.*

The proof is very simple and almost contained in [4], with a change of language: the current $f \wedge \bar{\partial}\chi$ is $\bar{\partial}$-closed in a ball near 0, and there is a distribution u such that, near 0, $\bar{\partial} u = f \wedge \bar{\partial}\chi$ where χ is the indicatrix of M^-; then the restriction u^- of u to M^- is $\mathscr{C}^\infty$ in $\bar{M}^-$ and $\bar{\partial}$-closed there; the restriction u^+ of u to M^+ is $\mathscr{C}^\infty$ in $\bar{M}^+$ and $\bar{\partial}$-closed there and we have $f = u^- - u^+$. In the case $q = 0$ then *any* solution u verifies $f = u^- - u^+$.

This proof relies only on the solution of the $\bar{\partial}$-equation in an open ball of $\mathbb{C}^n$. This case is a good illustration for the idea behind the more complicated codimension two case.

Much more precise results in the case of C.R. manifolds with extension in only *one* wedge were done by Trépreau [21] and Tumanov [22], using the method of discs.

On the other hand Trépreau proved [unpublished] that, even in the case of C.R. manifolds, there is not always extension, no matter the number of wedges you choose. So we must put conditions in order to get the extension property.

In § 2, we prove a general local theorem of extension for the codimension two case.

In § 3, we give some applications; let S_1 and S_2 be two real $\mathscr{C}^\infty$ hypersurfaces in $\mathbb{C}^n$, then we get first a local result

Theorem 3.2. *Let $M = S_1 \cap S_2$, where S_1 is strictly pseudo-convex near 0, and S_1 and S_2 are transversal; then if $n \geqq 3$, any C.R. function on M near 0 is a sum of boundary values of holomorphic functions in wedges of edge M.*

In the case $n = 2$, we have to deal with the H. Lewy's obstruction, hence we get the same result provided that the C.R. function f verifies an extra regularity condition (Theor. 3.3).

In special cases, when $M \subset \mathscr{C}^2$ has only one complex tangency of elliptic type, Kenig and Webster [15], and Moser and Webster [19] obtained a better result: they characterized the local enveloppe of holomorphy of M.

In § 4 and § 5, we study the compact case and we get

Theorem 4.1. *Let $M = \partial\Omega_1 \cap \partial\Omega_2$, with Ω_i bounded pseudo convex domains with smooth $\mathscr{C}^\infty$ boundaries and defined by global p.s.h. functions; suppose that $\partial\Omega_1$ and $\partial\Omega_2$ are transversal, then if $n \geqq 3$, any C.R. function f on M is a sum of holomorphic functions defined on wedges of edge M.*

Theorem 5.2. *If $n = 2$, then there is a measure $d\pi$ on M such that if: $\int f \varphi \, d\pi = 0$ for all φ holomorphic near $\bar{\Omega}_i$, $i = 1, 2$ then f is a sum of holomorphic functions defined on wedges of edge M.*

The condition in the case $n = 2$ is naturally coming from the necessary condition to solve the $\bar{\partial}_b$-equation on $\partial\Omega_i$.

Again in some special cases, when $M \subset \mathbb{C}^2$, Bedford and Gaveau [9] characterized the enveloppe of holomorphy of M.

§ 1 Codimension one

Suppose that the closed set M cuts the space in two parts M^+ and M^- and that $0 \in M$ then we have the following local result

Theorem 1.1. *If f is a C.R. (p, q) form on M near 0, then there are two (p, q) forms, f^+, $\mathscr{C}^\infty$ on $\bar{M}^+$ and f^-, $\mathscr{C}^\infty$ on $\bar{M}^-$, such that $\bar{\partial} f^+ = 0$ on M^+ and $\bar{\partial} f^- = 0$ on M^- and with $f = f^+ + f^-$ on M.*

Proof. The form

$$\omega := \chi \cdot \bar{\partial} F \quad \text{is a } \mathscr{C}^\infty(p, q+1) \quad \text{form near 0 in } \mathscr{C}^n, \tag{1.1}$$

where χ is the characteristic function of the set M^-, because $\bar{\partial} F$ is flat on M. Moreover ω is $\bar{\partial}$-closed and, choosing a ball centered at 0 and contained in the set where ω is smooth, we can solve

$$\bar{\partial} v = \omega \quad \text{with } v \text{ in } \mathscr{C}^\infty \text{ in this ball.} \tag{1.2}$$

Let now

$$f^+ := v \quad \text{and} \quad f^- := F - v \tag{1.3}$$

then we have

$$\bar{\partial} f^+ = \bar{\partial} v = \omega = 0 \quad \text{in } \bar{M}^{-c}, \quad \text{hence in } M^+ \tag{1.4}$$

$$\bar{\partial} f^- = \bar{\partial} F - \bar{\partial} v = 0 \quad \text{in } M^- \tag{1.5}$$

and

$$F = f^- + f^+, \quad \text{hence,} \quad f = f^+ + f^- \quad \text{on } M. \tag{1.6}$$

Now the current $\omega' = F \wedge \bar{\partial} \chi$ verifies $\bar{\partial} \omega' = \bar{\partial} F \wedge \bar{\partial} \chi = 0$ because $\bar{\partial} F$ is flat on M and the support of $\bar{\partial} \chi$ is contained in M; moreover we have that

$$u = F\chi - v \quad \text{verifies } \bar{\partial} u = \omega' \quad \text{by (1.2)} \tag{1.7}$$

and

$$u^+ = -v = -f^+, \quad u^- = F - v = f^- \quad \text{by (1.3),} \tag{1.8}$$

where u^- is the restriction of u to M^- and u^+ is the restriction of u to $\bar{M}^{-c}$; hence we get

$$f = u^- - u^+ \quad \text{on } M. \tag{1.9}$$

Moreover, if $q = 0$ the difference between two solutions of

$$\bar{\partial} u = \omega' \tag{1.10}$$

is a $\bar{\partial}$-closed distribution, hence a holomorphic function near 0, hence $\mathscr{C}^\infty$ through M. So we get in that case that *any* solution u of (1.10) verifies (1.9).

§ 2-Codimension two

Let Ω be an open set in $\mathbb{C}^n$; we say that Ω has the E_q *property at* 0 if $0 \in \partial\Omega$ and it is possible to solve locally the $\bar{\partial}$-equation with $\mathscr{C}^\infty$ bounds up to $\partial\Omega$ for (p, q) forms; precisely:

let $\mathscr{U}$ be a neighbourhood of 0 in $\mathscr{C}^n$ and ω a (p, q) form in $\mathscr{C}^\infty(\mathscr{U} \cap \bar{\Omega})$, $\bar{\partial}$-closed, then there is a neighbourhood $\mathscr{V}$ of 0 in $\mathbb{C}^n$ and a $(p, q-1)$ form η in $\mathscr{C}^\infty(\mathscr{V} \cap \bar{\Omega})$ such that $\bar{\partial}\eta = \omega$ in $\mathscr{V} \cap \Omega$.

Examples:

Ω a $\mathscr{C}^\infty$ smoothly bounded pseudo-convex open set in $\mathbb{C}^n$, the existence of admissible neighbourhoods [1] and the results of J.J. Kohn [16] prove the property E_q for $1 \leqq q \leqq n$;

Ω with the property $Z(q)$, the Levi form of $\partial\Omega$ at 0 has at least $q+1$ negative eigenvalues or at least $n-q$ positive ones [4];

Ω a convex open set, proved in [12].

Now we are in position to prove the following local result

Theorem 2.1. *Let M be a closed set in $\mathbb{C}^n$ such that $0 \in M$ and there are open sets Ω_i, Ω_i', $i = 1, 2$ with:*

Ω_i and Ω_i' verify the property E_{q+1}, $i = 1, 2$ at 0;

$M = \partial\Omega_1 \cap \partial\Omega_2 = \partial\Omega_1' \cap \partial\Omega_2'$ near 0;

$M = \bar{\Omega}_1 \cap \bar{\Omega}_2'$ near 0;

$\partial\Omega_1$ and $\partial\Omega_2$ are transversal, $\partial\Omega_i$ and $\partial\Omega_i'$ are regularly situated. Then any C.R. form $\mathscr{C}^\infty$ on M near 0 is a sum of $\bar{\partial}$-closed forms defined in wedges of edge M.

In fact the wedges are precisely the sets: $\Omega_1 \cap \Omega_2$, $\Omega_1 \cap \Omega_2'$, $\Omega_1' \cap \Omega_2$ and $\Omega_1' \cap \Omega_2'$.

Let γ_1 a function such that

$$\gamma_1 = 1 \quad \text{in } \Omega_1; \quad \gamma_1 = 0 \quad \text{in } \Omega_1' \tag{2.1}$$

moreover γ_1 can be chosen such that $\gamma_1 \in \mathscr{C}^\infty(M^c)$ because $\partial\Omega_1$ and $\partial\Omega_1'$ are regularly situated.

Similarly, let γ_2 a function such that

$$\gamma_2 = 1 \quad \text{in } \Omega_2; \quad \gamma_2 = 0 \quad \text{in } \Omega_2' \quad \text{and} \quad \gamma_2 \in \mathscr{C}^\infty(M^c). \tag{2.2}$$

Now let f a C.R. (p, q) form on M and F a $\mathscr{C}^\infty$-extension of f in $\mathscr{C}^n$ which is $\bar{\partial}$-flat on M.

Let

$$\omega := F \wedge \bar{\partial}\gamma_1 \wedge \bar{\partial}\gamma_2, \tag{2.3}$$

this $(p, q+2)$ current has M as support and we have

$$\bar{\partial}\omega = \bar{\partial}F \wedge \bar{\partial}\gamma_1 \wedge \bar{\partial}\gamma_2 = 0 \tag{2.4}$$

because $\bar{\partial}F$ is flat on M and $\bar{\Omega}_i$ and $\bar{\Omega}_i'$ are regularly situated.

Now $\bar{\partial}F\cdot\gamma_1\wedge\bar{\partial}\gamma_2$ is $\bar{\partial}$-closed for the same reason as above and this form is $\mathscr{C}^\infty$ in a full neighbourhood of 0. Hence we can find in a ball $\mathbb{B}$ centered at 0 a $\mathscr{C}^\infty(p, q+1)$ form X such that

$$\bar{\partial}X=-\bar{\partial}F\cdot\gamma_1\wedge\bar{\partial}\gamma_2. \tag{2.5}$$

Let us put

$$\mu_0:=F\gamma_1\wedge\bar{\partial}\gamma_2+X \tag{2.6}$$

then we have $\bar{\partial}\mu_0=F\wedge\bar{\partial}\gamma_1\wedge\bar{\partial}\gamma_2$.

We can solve the equation $\bar{\partial}v=X$ where $\bar{\partial}X=0$, hence in Ω_2 and in Ω_2', so we get

$$v_2\in\mathscr{C}^\infty(\bar{\Omega}_2),\quad\text{with }\bar{\partial}v_2=X\quad\text{in }\Omega_2 \tag{2.7}$$

and

$$v_2'\in\mathscr{C}^\infty(\bar{\Omega}_2'),\quad\text{with }\bar{\partial}v_2'=X\quad\text{in }\Omega_2'. \tag{2.8}$$

Now let τ_2 a smooth form such that $\tau_2=v_2$ in Ω_2, $\tau_2=v_2'$ in Ω_2'; we can choose τ_2 to be in $\mathscr{C}^\infty(M^c)\cap L^\infty(\mathbb{B})$. Then clearly

$$\mu=\mu_0-\bar{\partial}\tau_2=F\wedge\gamma_1\cdot\bar{\partial}\gamma_2+X-\bar{\partial}\tau_2 \tag{2.9}$$

still verifies $\bar{\partial}\mu=\omega$, and the support of μ is contained in $(\Omega_1^c\cap\Omega_1'^c)\cup(\Omega_2^c\cap\Omega_2'^c)$.

Let us see what happen in Ω_1; we have that $\gamma_1=1$ there and

$$\mu=F\wedge\bar{\partial}\gamma_2+X-\bar{\partial}\tau_2; \tag{2.10}$$

this can be rewritten in Ω_1

$$\mu=\bar{\partial}(F\cdot\gamma_2)+X-\gamma_2\cdot\bar{\partial}F-\bar{\partial}\tau_2 \tag{2.11}$$

but the form $\gamma_2\cdot\bar{\partial}F$ is smooth on $\bar{\Omega}_1$ because $\bar{\partial}F$ is flat on M, so we get that $X-\gamma_2\cdot\bar{\partial}F$ is a $\mathscr{C}^\infty(\bar{\Omega}_1)$ $\bar{\partial}$-closed smooth $(p, q+1)$ form and thus there is a (p, q) form α_1 such that

$$\bar{\partial}\alpha_1=X-\gamma_2\cdot\bar{\partial}F\quad\text{and}\quad\alpha_1\in\mathscr{C}^\infty(\bar{\Omega}_1), \tag{2.12}$$

hence the form $h_1:=F\cdot\gamma_2-\tau_2+\alpha_1$ verifies

$$\bar{\partial}h_1=\mu\quad\text{on }\Omega_1. \tag{2.13}$$

Similarly in Ω_1' we have $\gamma_1=0$ hence $\mu=X-\bar{\partial}\tau_2$; still X is in $\mathscr{C}^\infty(\bar{\Omega}_1')$ hence there is a (p, q) form α_1' in $\mathscr{C}^\infty(\bar{\Omega}_1')$ such that

$$\bar{\partial}\alpha_1'=X,\quad\text{in }\Omega_1', \tag{2.14}$$

hence the form $h_2:=-\tau_2+\alpha_1'$ verifies

$$\bar{\partial}h_2=\mu\quad\text{on }\Omega_1'. \tag{2.15}$$

But recalling that μ is supported on $(\Omega_1^c \cap \Omega_1'^c) \cup (\Omega_2^c \cap \Omega_2'^c)$ we get that

(2.16) $h_1 = F \cdot \gamma_2 - \tau_2 + \alpha_1$ is $\bar{\partial}$-closed in $\Omega_1 \cap \Omega_2$ and in $\Omega_1 \cap \Omega_2'$.

So we get that

(2.17) $h_1^+ := F - \upsilon_2 + \alpha_1$ is $\bar{\partial}$-closed in Ω_1 in $\Omega_1 \cap \Omega_2$,

(2.18) $h_1^- := -\upsilon_2' + \alpha_1$ is $\bar{\partial}$-closed in $\Omega_1 \cap \Omega_2'$.

Similarly

(2.19) $h_2^+ := -\upsilon_2 + \alpha_1'$ is $\bar{\partial}$-closed in $\Omega_2 \cap \Omega_1'$,

(2.20) $h_2^- := -\upsilon_2' + \alpha_2'$ is $\bar{\partial}$-closed in $\Omega_2' \cap \Omega_1'$.

Because all the forms involved are $\mathscr{C}^\infty$ up to the boundary of their domain, we get that

(2.21) $F = f = -h_2^+ + h_2^- + h_1^+ - h_1^-$, on M.

Hence the theorem is proved.

§ 3 Applications

In all the applications we shall have to make the hypothesis that our domains will be defined by a p.s.h. function in a neighbourhood of their closure.

In order to use the § 2, we need the following lemma

Lemma 3.1. *Let $M = S_1 \cap S_2$, where S_2 is strictly pseudo-convex and S_1 and S_2 are transversal, then there are two pairs of domains Ω_i and Ω_i', $i = 1, 2$, such that:*
a) *Ω_i and Ω_i' are smoothly bounded and either pseudo-convex or strictly pseudo-concave,*
b) $\bar{\Omega}_i \cap \bar{\Omega}_i' = M$, $i = 1, 2$,
c) *$\bar{\Omega}_i$ and $\bar{\Omega}_i'$ are regularly situated, $i = 1, 2$.*

Proof of the lemma. a) S_1 and S_2 are strictly pseudo-convex; then we take $\Omega_1 = \{\rho_1 > 0\}$ then Ω_1 is strictly pseudo-concave. Now let $\rho_1' = \rho_1 + \varepsilon \rho_2^2$, where ρ_i is the defining function of S_i; a simple computation then shows that if ε is small enough then $\Omega_1' := \{\rho_1' < 0\}$ is still s.p.c. We do the same to get Ω_2 and Ω_2', just reversing the role of ρ_1 and ρ_2.

The other properties of the lemma are easy to see.

b) S_1 s.p.c. then let $r_2 := t\rho_1 + (1-t)\rho_2$, again if $(1-t)$ is small enough, S_2' $:= \{r_2 = 0\}$ is s.p.c. and because $M = S_1 \cap S_2'$ we now are in the case a).

a) Local case and $n \geqq 3$

Let S_i be smooth real hypersurfaces, then we have the following

Theorem 3.2. *Let $M = S_1 \cap S_2$, where S_1 is strictly pseudo-convex and S_1 and S_2 are transversal. Then any C.R. function on M is a sum of boundary values of holomorphic functions defined in wedges of edge M.*

Proof. We apply the lemma above to get the sets Ω_i and Ω_i'; if Ω_i (or Ω_i') is pseudo-convex then they verify the E_1 property; if Ω_i is strictly pseudo-concave, because $n \geqq 3$, it still verifies the E_1 property. Hence in any case we are in position to apply the results of § 2.

b) Local case and $n=2$

In this case we have to distinguish between the two possibilities:

S_1 and S_2 are complex transversal,

then $M=S_1 \cap S_2$ is totally real, hence the result is well known.

S_1 and S_2 are not complex transversal,

then suppose that S_i verify the same hypothesis as in the theorem above; using the lemma again we introduce the sets Ω_i and Ω_i'. So suppose that Ω_2 is strictly pseudo-concave, i.e. S_2 is strictly pseudo-convex; to give a theorem in that case we must return to the proof.

Let f be a C.R. function on $M=S_2 \cap S_2$. Using the lemma 3.1 we can introduce the sets Ω_i' and γ_1 and γ_2 as before and let F be the extension of f in a ball $\mathbb{B}$ centered at 0, $\bar{\partial}F$ flat on M. Again we can solve the problem

$$\bar{\partial}X = \bar{\partial}F \cdot \gamma_1 \wedge \bar{\partial}\gamma_2$$

with $X \in \mathscr{C}^{\infty}_{(0,1)}(\mathbb{B})$ and the question now is to solve

$$\bar{\partial}v = X \quad \text{in } \Omega_2. \tag{3.7}$$

This time we are in a strictly pseudo-concave set in $\mathbb{C}^2$ and, in order to solve this equation we must overcome the Lewy's obstruction

it is necessary and sufficient that X verifies a condition given by G. Henkin [14, § 5]: there is a kernel K' acting on X such that the condition is

$$K'(X) \quad \text{is real analytic near } 0; \tag{3.8}$$

but X is a solution of a $\bar{\partial}$-equation in a ball, hence using $\bar{\partial}$-solving kernels, K'', [11], we have

$$X = K''(-\gamma_1 \bar{\partial}F \wedge \bar{\partial}\gamma_2) \tag{3.9}$$

so, composing K' and K'', we get the existence of a kernel K acting on $\bar{\partial}F$ such that the necessary and sufficient condition for solving (3.7) locally can be written

$$K(\bar{\partial}F) \quad \text{is real-analytic near } 0; \tag{3.10}$$

Once this is done, the rest of the proof is exactly the same.
Hence in that case we have

Theorem 3.3. *Let $M=S_1 \cap S_2$, $0 \in M$, where S_1 is strictly pseudo-convex and S_1 and S_2 are transversal, in $\mathbb{C}^2$. Let f be a C.R. function on M near 0 and F an extension of f in $\mathbb{C}^2$ near 0 such that $\bar{\partial}F$ is flat on M. Suppose moreover that* (3.10) *holds then f is a sum of boundary values of holomorphic functions defined in wedges of edge M.*

Remark 3.4. If the C.R. function f on M is holomorphic in a thin tubular neighbourhood of M, then the condition (3.10) is fullfilled, meaning that independantly of the thinness of this neighbourhood, f is the sum on M of boundary values of holomorphic functions defined on fixed wedges.

Remark 3.5. The case of a surface $M \subset \mathbb{C}^2$ of real dimension two with an isolated complex tangency at the origin studied by Kenig and Webster [15] and by Moser and Webster [19] is in the field of application of theorem 3.1, but in the case when the origin is an elliptic point, they have a better result: they characterize the local hull of holomorphy of M near 0 as the union of a one parameter family of discs.

§ 4 Compact case and $n \geqq 3$

We shall need the following

Lemma 4.1. *Let $M = S_1 \cap S_2$, with S_i defined by global p.s.h. functions, and S_1 and S_2 transversal. Then there are 2 pairs of pseudo-convex domains Ω_i and Ω_i' such that $M = \partial\Omega_i \cap \partial\Omega_i'$, $i = 1, 2$, $\partial\Omega_i$ and $\partial\Omega_i'$ are regularly situated, and $\partial\Omega_1$ and $\partial\Omega_2$ are still transversal.*

Let us suppose, without loss of generality, that $\rho_i \geqq -1$, where ρ_i is the defining function for S_i; let $\varepsilon > 0$ be such that ρ_i is p.s.h. in $\{\rho_i < \varepsilon\}$ and let $t \in]0, 1[$ fixed small and finally let $A = 2(1-t)/t$.

Now let us take a function χ, convex and increasing on $\mathbb{R}$ such that

$$\chi \equiv 0 \quad \text{on } [-\infty, \varepsilon/3] \quad \text{and} \quad \chi(2\varepsilon/3) = 1/2.$$

Now let $\rho_i' := \rho_i + A\chi(\rho_i)$, $i = 1/2$ and $\rho_t := (1-t)\rho_1' + t\rho_2'$ then we have

Claim. ρ_t is a p.s.h. function near the closure of $\Omega_t := \{\rho_t < 0\}$.

The point is to see that ρ_i' are p.s.h. functions near Ω_t; we have $\rho_t < 0 \Rightarrow (1-t)\rho_1' + t\rho_2' < 0$, hence

$$\rho_1' < t/(1-t) \Rightarrow \rho_1 + A\chi(\rho_1) < t/(1-t) \Rightarrow \chi(\rho_1) < t^2/2(1-t)^2 \Rightarrow \rho_1 < 2\varepsilon/3;$$

ρ_1 is a p.s.h. function, hence ρ_1' too because χ is convex increasing.

For ρ_2 we have $\rho_2' < (1-t)/t \Rightarrow \chi(\rho_2) < 1/2 \Rightarrow \rho_2 < 2\varepsilon/3$, hence the same conclusion. Let

$$(*) \qquad \Omega_t := \{\rho_t < 0\}$$

Then by the claim Ω_t is pseudo-convex.

Now let (with a new ε!)

$$\Omega_t' := \{\rho_t' := \rho_t + \varepsilon\rho_2^2 < 0\}$$

We can calculate the Levi form for ρ_t'

$$\partial\bar{\partial}\rho_t' = (1-t)\,\partial\bar{\partial}\rho_t + (t + 2\varepsilon\rho_2')\,\partial\bar{\partial}\rho_2' + 2\varepsilon\partial\rho_2' \wedge \bar{\partial}\rho_2'.$$

Now we can choose ε so small that $(t + 2\varepsilon\rho_2') \geqq 0$, hence ρ_t' is a p.s.h. function and hence Ω_t' is still p.c. with smooth boundary.

We have that

$$\rho_t = \rho_t' = 0 \Rightarrow \rho_2' = 0 \quad \text{and hence} \quad \rho_1' = 0,$$
$$\text{but} \quad \rho_i' = 0 \Rightarrow \rho_i = 0 \quad \text{hence} \quad \Omega_t \cap \Omega_t' = M.$$

Moreover, the order of contact is clearly two between Ω_t and Ω_t', hence they are regularly situated.

This pair Ω_t and Ω_t' may play the role of the pair Ω_1 and Ω_1' of the lemma, and in order to get the pair Ω_2 and Ω_2', we have just to play the same game, exchanging the role of 1 and 2. Hence the lemma.

Theorem 4.1. *Let $M = \partial\Omega_1 \cap \partial\Omega_2$, with Ω_i bounded pseudo-convex domains with smooth $\mathscr{C}^\infty$ boundaries and defined by global p.s.h. functions; suppose that $\partial\Omega_1$ and $\partial\Omega_2$ are transversal then any C.R. function on M is a sum of holomorphic functions defined in wedges of edge M.*

Proof. The proof is very analogous to the one of § 2 but we have to use an extra result by J.P. Rosay [20] and there is a change of notations so we prefer to do it completely.

Introduce the pairs Ω_i, Ω_i' given by lemma 4.1: then Ω_i and Ω_i' are pseudo-convex; moreover we have

$$\bar{\Omega}_i' \setminus M \subset \Omega_i, \qquad i = 1, 2. \tag{4.1}$$

Let γ_1 a function such that

$$\gamma_1 = 0 \quad \text{in} \quad \Omega_1^c; \qquad \gamma_1 = 1 \quad \text{in} \quad \Omega_1'; \qquad \gamma_1 \in \mathscr{C}^\infty(\mathbb{C}^n \setminus M). \tag{4.2}$$

Let γ_2 the analogous function for Ω_2

$$\gamma_2 = 0 \quad \text{in} \quad \Omega_2^c; \qquad \gamma_2 = 1 \quad \text{in} \quad \Omega_2'; \qquad \gamma_2 \in \mathscr{C}^\infty(\mathbb{C}^n \setminus M).$$

Let F the extension of f such that $\bar{\partial}F$ is flat on M and let

$$\omega := F\,\bar{\partial}\gamma_1 \wedge \bar{\partial}\gamma_2 \tag{4.3}$$

it is a $\bar{\partial}$-closed current as before.

Now we have that $\bar{\partial}F\gamma_1 \wedge \bar{\partial}\gamma_2$ is a $\mathscr{C}^\infty(\mathbb{C}^n)$ $\bar{\partial}$-closed form; moreover we can manage to have F with compact support, so it is possible to solve

$$\bar{\partial}G = \bar{\partial}F\gamma_1 \wedge \bar{\partial}\gamma_2 \quad \text{with} \quad G \quad \text{in} \quad \mathscr{C}^\infty(\mathbb{C}^n). \tag{4.4}$$

So we get that $\mu_0 := F\gamma_1 \cdot \bar{\partial}\gamma_2 - G$ verifies $\bar{\partial}\mu_0 = \omega$.

Now we want to diminish the support of G. Clearly we have that

$$G \in \mathscr{C}^\infty_{(0,1)}(\bar{\Omega}_2'), \qquad \bar{\partial}G \equiv 0 \quad \text{in} \quad \Omega_2'. \tag{4.5}$$

Hence, using J. Kohn's $\mathscr{C}^\infty$ estimates for $\bar{\partial}$, we can find

$$v_2' \in \mathscr{C}^\infty(\bar{\Omega}_2'), \qquad \text{such that} \quad \bar{\partial}v_2' = G. \tag{4.6}$$

G is still in $\mathscr{C}^\infty(\Omega_2^c)$ and $\bar{\partial}$-closed, so on the boundary $\partial\Omega_2$, G induces a $\mathscr{C}^\infty$, $\bar{\partial}_b$-closed form; hence using J.P. Rosay's result [20] with $n \geqq 3$, we can extend

$$G \quad \text{in} \quad \bar{\Omega}_2 \quad \text{as a} \quad \mathscr{C}^\infty\,\bar{\partial}\text{-closed}\,(0,1) \quad \text{form} \quad G_2. \tag{*}$$

Now let us take a big ball $\mathbb{B}$ which contains $\bar{\Omega}_2$ and $\bar{\Omega}_1$, and call again G_2 the (0, 1) form G extended by G_2; then G_2 is in $\mathscr{C}^\infty(\mathbb{B})$ and $\bar{\partial}$-closed there.

We can solve the $\bar{\partial}$-equation $\bar{\partial}v_2 = G_2$ with v_2 in $\mathscr{C}^\infty(\mathbb{B})$. In particular we have

$$\bar{\partial}v_2 = G \quad \text{in } \Omega_2^c \cap \mathbb{B}. \tag{4.7}$$

Let us still denote by v_2' any $\mathscr{C}^\infty$ continuation of v_2' to $\mathbb{B}$ and define

$$v := \gamma_2 v_2' + (1-\gamma_2) v_2. \tag{4.8}$$

We are now in a position to consider the following current

$$\mu := F\gamma_1 \cdot \bar{\partial}\gamma_2 - G + \bar{\partial}v. \tag{4.9}$$

Clearly this current still verifies $\bar{\partial}\mu = \omega$ but now μ is supported by $(\bar{\Omega}_1 \cap \Omega_1'^c) \cup (\bar{\Omega}_2 \cap \Omega_2'^c)$.

Let us see what happens in Ω_1'; there $\gamma_1 = 1$ and $\mu = F \cdot \bar{\partial}v$; this can be rewritten in Ω_1'

$$\mu = \bar{\partial}(F \cdot \gamma_2) - \gamma_2 \bar{\partial}F - G + \bar{\partial}v, \tag{4.10}$$

but the form $\gamma_2 \cdot \bar{\partial}F$ is smooth in $\bar{\Omega}_1'$ because $\bar{\partial}F$ is flat on M, so we get that $G + \gamma_2 \cdot \bar{\partial}F$ is a $\mathscr{C}^\infty(\bar{\Omega}_1')$ $\bar{\partial}$-closed smooth (0, 1) form and thus there is a function α_1' such that

$$\bar{\partial}\alpha_1' = G + \gamma_2 \cdot \bar{\partial}F, \quad \text{and} \quad \alpha_1' \in \mathscr{C}^\infty(\bar{\Omega}_1'); \tag{4.11}$$

hence the function $h_1' := F \cdot \gamma_2 + v - \alpha_1'$ verifies

$$\bar{\partial}h_1' = \mu \quad \text{on } \Omega_1'. \tag{4.12}$$

Similarly in Ω_1^c we have $\gamma = 0$ hence $\mu = -G + \bar{\partial}v$; still G is in $\mathscr{C}^\infty(\bar{\Omega}_1^c)$ hence using again Rosay's result, there is a function α_1 in $\mathscr{C}^\infty(\bar{\Omega}_1^c)$ such that

$$\bar{\partial}\alpha_1 = G, \quad \text{in } \Omega_1^c; \tag{**}$$

hence the function $h_1 := v - \alpha_1$ verifies

$$\bar{\partial}h_1 = \mu \quad \text{in } \Omega_1^c. \tag{4.13}$$

But recalling that μ is supported on $(\Omega_1'^c \cap \bar{\Omega}_1) \cup (\bar{\Omega}_2 \cap \Omega_2'^c)$ we get that

$$h_1 = F \cdot \gamma_2 + v - \alpha_1 \quad \text{is holomorphic in } \Omega_2' \cap \bar{\Omega}_1^c \quad \text{and in } \bar{\Omega}_1^c \cap \bar{\Omega}_2^c. \tag{4.14}$$

So

$$h_1^+ := F + v_2' - \alpha_1 \quad \text{is holomorphic in } \Omega_2' \cap \Omega_1^c; \tag{4.15}$$

$$h_1^- := v_2 - \alpha_1 \quad \text{is holomorphic in } \bar{\Omega}_1^c \cap \bar{\Omega}_2^c. \tag{4.16}$$

Similarly

$$h_2^+ := v_2 - \alpha_1' \quad \text{is holomorphic in } \Omega_2^c \cap \Omega_1' \tag{4.17}$$

$$h_2^- := v_2' - \alpha_1' \quad \text{is holomorphic in } \Omega_2' \cap \Omega_1'. \tag{4.18}$$

Because all the functions involved are $\mathscr{C}^\infty$ up to the boundaries of their domain we get

$$F=f=h_2^+-h_2^-+h_1^+-h_1^-, \quad \text{on } M, \tag{4.19}$$

and the theorem.

§ 5 Compact case and $n=2$

The only problem is how to obtain (*) and (**) in the previous proof, because we are now in the critical case with $n=2$. In order to study what happens here we have to introduce some notations

$$\Omega_i \quad \text{is defined by } \rho_i<0, \quad i=1,2, \tag{5.1}$$

$$\Omega_i' \quad \text{is defined by } \rho_i+\varepsilon_i^2<0, \quad i=1,2, \tag{5.2}$$

using Lemma 4.1; ε_i being a smooth function. Ω_i are bounded pseudo-convex domains with smooth $\mathscr{C}^\infty$ boundaries and $\partial\Omega_1$ and $\partial\Omega_2$ are transversal.

Let f be a C.R. function on M and F be its $\mathscr{C}^\infty$ $\bar\partial$-flat extension. We have to be precise in defining the functions γ_i

$$\gamma_i := \chi(-\rho_i/\varepsilon_i^2), \quad i=1,2 \tag{5.3}$$

where χ is a function such that

$$\chi(t)=0 \quad \text{if } t\leqq 0; \quad \chi(t)=1 \quad \text{if } t\geqq 1; \quad \chi\in\mathscr{C}^\infty(\mathbb{R}). \tag{5.4}$$

Let us start in the same way as above up to the point where we have the (0, 1) form G such that

$$\bar\partial G=\bar\partial F\gamma_1\wedge\bar\partial\gamma_2 \quad \text{with } G \text{ in } \mathscr{C}^\infty(\mathbb{C}^n). \tag{5.5}$$

Of course, because Ω_2' is pseudo-convex, we can solve

$$\bar\partial v_2'=G \quad \text{in } \Omega_2' \quad \text{with } v_2'\in\mathscr{C}^\infty(\bar\Omega_2'). \tag{5.6}$$

We want to solve the same equation in Ω_2^c and now we have the global condition [20]

$$\int_{\partial\Omega_2}\varphi\cdot G\wedge\beta=0, \quad \text{for all } \varphi \text{ holomorphic near } \bar\Omega_2, \tag{*'}$$

where $\beta := dz_1\wedge dz_2$.

The idea is to show that this condition is in fact an integral condition on f over M.

Proposition 5.1. There is a measure $d\pi$ on M such that the condition (*') is equivalent to the condition (A):

$$\int_M f\cdot\varphi\, d\pi=0, \quad \text{for all } \varphi \text{ holomorphic near } \bar\Omega_2. \tag{A}$$

Proof. By Stokes' theorem we have

$$\int_{\partial\Omega_2} \varphi \cdot G \wedge \beta = \int_{\Omega_2} \varphi \cdot \bar{\partial} G \wedge \beta. \tag{5.7}$$

We can replace $\bar{\partial}G$ by its value: $\gamma_1 \cdot \bar{\partial}F \wedge \bar{\partial}\gamma_2$ hence we have to estimate

$$I := \int_{\Omega_2} \varphi \gamma_1 \bar{\partial} F \wedge \bar{\partial}\gamma_2 \wedge \beta. \tag{5.8}$$

Let us introduce smooth approximations for the γ_i's

$$\gamma_1^\varepsilon := \chi\left[\frac{-\rho_1}{\varepsilon_1^2+\varepsilon^2}\right]; \quad \gamma_2^\delta := \chi\left[\frac{-\rho_2}{\varepsilon_2^2+\delta^2}\right]. \tag{5.9}$$

We can choose F with compact support in $\mathbb{C}^2$, hence by Stokes' formula applied in a domain Ω containing the support of $\bar{\partial}\gamma_2^\delta$ we get

$$\int_{\mathbb{C}^2} \varphi \gamma_1^\varepsilon \bar{\partial} F \wedge \bar{\partial}\gamma_2^\delta \wedge \beta + \int_{\mathbb{C}^2} \varphi F \bar{\partial}\gamma_1^\varepsilon \wedge \bar{\partial}\gamma_2^\delta \wedge \beta = \int_{\partial\Omega} \varphi F \gamma_1^\varepsilon \bar{\partial}\gamma_2^\delta \wedge \beta = 0, \tag{5.10}$$

provided that φ is holomorphic near $\bar{\Omega}$.

Because of the transversality of $\partial\Omega_1$ and $\partial\Omega_2$, when $\varepsilon \to 0$, $\delta \to 0$, we have that

$$\bar{\partial}\gamma_1^\varepsilon \wedge \bar{\partial}\gamma_2^\delta \wedge \beta \to \bar{\partial}\gamma_1 \wedge \bar{\partial}\gamma_2 \wedge \beta, \tag{5.11}$$

which is a measure supported by M. Moreover, if φ is holomorphic near $\bar{\Omega}_2$ for ε and δ small enough, φ will be holomorphic near the support of $\bar{\partial}\gamma_2$ hence we have

$$I = \int_{\Omega_2} \varphi \gamma_1 \bar{\partial} F \wedge \bar{\partial}\gamma_2 \wedge \beta = -\int_M \varphi f \bar{\partial}\gamma_1 \wedge \bar{\partial}\gamma_2 \wedge \beta, \tag{5.12}$$

$$\forall \varphi \quad \text{holomorphic near } \bar{\Omega}_2$$

Hence the Proposition 5.1 with $d\pi := \bar{\partial}\gamma_1 \wedge \bar{\partial}\gamma_2 \wedge \beta$.

Of course, doing exactly the same thing for (**'), if

$$\int_M \varphi f \, d\pi = 0, \tag{5.18}$$

for all φ holomorphic near $\bar{\Omega}_1$, then we can solve

$$\bar{\partial}\alpha_1 = G, \quad \text{in } \Omega_1^c; \tag{**'}$$

the end of the proof is the same as in § 4, hence we proved the following

Theorem 5.2. *Let $M=\partial\Omega_1\cap\partial\Omega_2$, with Ω_i bounded pseudo-convex domains with $\mathscr{C}^\infty$ smooth boundaries; suppose that $\partial\Omega_1$ and $\partial\Omega_2$ are transversal then any C.R. function f on M such that:*

$$\int_M \varphi f\, d\pi = 0 \quad \forall\, \varphi \quad \text{holomorphic near } \bar{\Omega}_i, \quad i=1,2$$

is a sum of holomorphic functions defined in wedges of edge M.

Remark 5.3. In special cases, Bedford and Gaveau [9] obtained a better result: they characterized the enveloppe of holomorphy of M.

Acknowledgements. This work was partially done when I was visiting professor at U.C.S.D. and it is my pleasure to thank this organisation for its support.

References

1. Amar, E.: Cohomologie Complexe at Applications. J. London Math. Soc. 29, 127–140 (1984)
2. Amar, E.: Extension de fonctions holomorphes et courants. Bull. Soc. Math., 2ième série **107**, 25–48 (1983)
3. Amar, E.: Extension of holomorphic functions and residue current. Université de Bordeaux L. Preprint (1988)
4. Andreotti, A., Hill, C.D.: Levi convexity and the Hans Lewy problem. III. Ann. Sc. Norm. Super. Pisa **26**, 325–363; ibid. 747–806 (1972)
5. Andreotti, A., Hill, C.D., Lojasiewicz, S., Mackichan, B.: Complexes of Differential Operators. Invent Math. **35**, 43–86 (1976)
6. Baouendi, S.M., Chang, C.H., Treves, F.: Microlocal hypo-ellipticity and extension of C.R. functions. J. Differ. Geom. **18**, 331–391 (1983)
7. Baouendi, S.M., Rothschild, L.P., Treves, F.: C.R. structures with group action and extendability of C.R. functions. Invent Math. **82**, 359–396 (1985)
8. Bouendi, S.M., Rothschild, L.P.: Transversal Lie group actions on abstract C.R. manifolds. Preprint
9. Bedford, E., Gaveau, B.: Envelopes of holomorphy of certain 2-spheres in $\mathbb{C}_2$. Am. J. Math. **105**, 975–1009 (1983)
10. Chaumat, J., Chollet, A.M.: Ensembles pics pour $A^\infty(\mathbf{D})$. Ann. Inst. Fourier **29**, 171–200 (1979)
11. Charpentier, P.: Formules explicites pour les solutions minimales de l'équation $\bar{\partial}u=f$... Ann. Inst. Fourier **30**, 121–154 (1980)
12. Dufresnoy, A.: Sur l'opérateur d" et et les fonctions différentiables au sens de Whitney. Ann. Inst. Fourier **29**, 229–238 (1979)
13. Forstneric, F.: Proper holomorphic mappings: a survey. Preprint
14. Henkin, G.M.: The Lewy equation and analysis on pseudoconvex manifolds. Russ. Math. Surv. **32**, 57–118 (1977)
15. Kenig, C., Webster, S.: The local hull of holomorphy of a surface in the space of two variables. Invent. Math. **67**, 1–21 (1982)
16. Kohn, J.J.: Global regularity of $\bar{\partial}$ on weakly pseudo-convex manifolds. T.A.M.S. **181**, 273–292 (1973)
17. Kohn, J.J., Rossi, H.: On the extension of holomorphic functions from the boundary of a complex manifold. Ann. Math. **81**, 451–472 (1965)
18. Lewy, H.: An example of a smooth linear partial differential equation without solution. Ann. Math. **64**, 155–158 (1957)
19. Moser, J.K., Webster, S.: Normal forms for real surfaces in $\mathbb{C}^2$ near complex tangents and hyperbolic surface transformations. Acta. Math. **150**, 255–296 (1983)

20. Rosay, J.P.: Equation de Lewy-résolubilité globale de l'équation $\bar{\partial}_b u = f$ sur la frontière de domaines faiblement pseudo-convexes de $\mathbb{C}^2$ (ou de $\mathbb{C}^n$). Duke Math. J. **49**, 121–128 (1982)
21. Trepreau, J.M.: Sur le prolongement des fonctions C.R. définies sur une hypersurface réelle de classe $\mathscr{C}^2$ dans $\mathbb{C}^n$. Invent. Math. **83**, 583–592 (1986)
22. Tumanov, A.E.: Extending C.R. functions to a wedge from manifolds of finite type. Mat. Sb. **136**, 128–139 (1988)

Math. Z. 206, 103–117 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

Cohomology groups of reduced enveloping algebras

Rolf Farnsteiner
Department of Mathematics, University of Wisconsin, Milwaukee, WI 53201, USA

Received March 12, 1990

0 Introduction

In this paper we shall be studying the cohomology theory of finite dimensional restricted Lie algebras $(L, [p])$ defined over a field F of positive characteristic p. The classification theory of their irreducible representations naturally leads to the consideration of a family $u(L, S)$ of associative algebras that are parametrized by linear forms $S \in L^*$. Since these so-called reduced enveloping algebras are Frobenius [cf. 20, 1.2 and 30, V.4.3] their cohomology theory shares many features with the cohomology theory of finite groups.

In Sect. 1 we provide subsidiary results concerning extension functors and projective modules of Frobenius algebras. Proposition 1.1 is applied in the succeeding section to illustrate that, in contrast to classical semisimple Lie algebras [cf. 20, 3.6], $u(L, S)$ usually fails to be semisimple if L is a restricted Lie algebra of Cartan type. The connection between the Hochschild cohomology of $u(L, S)$ and the cohomology theory of $u(L) := u(L, 0)$ is established in Theorem 2.4.

The third section investigates various aspects of the interplay between structural features of L and its restricted cohomology. We augment Hochschild's early work [24] by providing new results, most in form of non-vanishing theorems, for nilpotent and non-toral completely solvable algebras. In particular, it is shown in (3.4) that the cohomology ring $H^*(u(L), F)$ of a unipotent Lie algebra has homogeneous elements of all possible degrees. The universal p-envelope is employed to illustrate that our results do not obtain for infinite dimensional Lie algebras.

The passage from the restricted to the ordinary cohomology groups of L can be described by means of a Grothendieck spectral sequence (cf. [20]). In Theorem 4.1 we elaborate on this result and establish a 5-term exact sequence that is closely related to the one obtained by Hochschild ([23, p. 575]).

1 Cohomology of Frobenius algebras

Throughout this section A shall denote an associative algebra over a field F of arbitrary characteristic. A pair consisting of such an algebra and an F-algebra

homomorphism $\varepsilon\colon A \to F$ is usually referred to as a *supplemented algebra.* For an A-module M one then defines the cohomology groups of A with coefficients in M via

$$H^n(A, M) := \mathrm{Ext}_A^n(F, M) \qquad \forall n \geqq 0.$$

Proposition 1.1 *Let (B, ε) be a supplemented subalgebra of A with supplementation ideal B^+. Suppose that A is B-projective and let $M \neq (0)$ be a projective A-module. Then the following statements hold:*

(1) *If B^+ is nilpotent, then $\dim_F M \geqq \dim_F B$.*

(2) *If B^+ is the Jacobson radical of B and M is a finitely generated B-module, then $\dim_F M \geqq \dim_F B$.*

Proof. Consider the exact sequence

$$(0) \to B^+ \to B \xrightarrow{\varepsilon} F \to (0)$$

of B-modules. Since A is a projective B-module, Proposition 6.2 of [4, p. 30] ensures the B-projectivity of M. Consequently, the sequence

$$(0) \to \mathrm{Hom}_B(M, B^+) \to \mathrm{Hom}_B(M, B) \to \mathrm{Hom}_B(M, F) \to (0)$$

is exact. The F-vector space $\mathrm{Hom}_B(M, F)$ is isomorphic to the dual of $M/(B^+ \cdot M)$. If B^+ is nilpotent, then this space obviously does not vanish. Alternatively, we arrive at the same conclusion by invoking Nakayama's Lemma [cf. 27, p. 137]. Consequently, the mapping $\mathrm{Hom}_B(M, B^+) \to \mathrm{Hom}_B(M, B)$ is not surjective in either case and there exists $f \in \mathrm{Hom}_B(M, B)$ such that $f(M)$ is not contained in B^+. Let m_0 be an element of M satisfying $\varepsilon(f(m_0)) = 1$. Suppose that $b \in B$ annihilates m_0. Then $0 = bf(m_0)$. On the other hand, we may write $f(m_0) = 1 - r$, where $r \in B^+$. It follows that $f(m_0)$ is invertible, whence $b = 0$. As a result, the linear mapping $B \to M$; $b \mapsto b \cdot m_0$ is injective and $\dim_F B \leqq \dim_F M$. □

Definitions. A finite dimensional F-algebra A is said to be a *Frobenius algebra* if there exists a nondegenerate F-bilinear form $(,)\colon A \times A \to F$ such that

$$(ab, c) = (a, bc) \qquad \forall a, b, c \in A.$$

The *left socle* $soc(A)$ is the sum of all irreducible submodules of A. It is known to coincide with $r(\mathrm{Jac}(A))$, the right annihilator of the Jacobson radical $\mathrm{Jac}(A)$ of A [cf. 6, 58.3].

Recall that for Frobenius algebras the notions of projectivity and injectivity of A-modules coincide [6, 62.11].

Theorem 1.2 *Let A be a Frobenius algebra, M an A-module. Then M is either injective or $\mathrm{Ext}_A^n(\mathrm{soc}(A), M) \neq (0)$ $\forall n \geqq 1$.*

Proof. Suppose that M is not injective. Since A is a Frobenius algebra, Ikeda's Theorem [6, 62.11] implies that the injective dimension $\mathrm{id}(M)$ of M is infinite. Let $n \geqq 1$. As $\mathrm{id}(M) \geqq n$, there exists an A-module Q such that $\mathrm{Ext}_A^n(Q, M) \neq (0)$. For any A-module X put

$$k(X) := \min\{k \in \mathbb{N};\ \mathrm{Jac}(A)^k \cdot X = (0)\},$$

and let $r := \min\{k(X);\ \mathrm{Ext}^n_A(X, M) \neq (0)\}$. Choose an A-module N such that $r = k(N)$. We consider the exact sequence

$$(0) \to \mathrm{Jac}(A)\cdot N \to N \to N/(\mathrm{Jac}(A)\cdot N) \to (0),$$

as well as the resulting exact sequence

$$\to \mathrm{Ext}^n_A(N/(\mathrm{Jac}(A)\cdot N), M) \to \mathrm{Ext}^n_A(N, M) \to \mathrm{Ext}^n_A(\mathrm{Jac}(A)\cdot N, M).$$

As $\mathrm{Ext}^n_A(N, M) \neq (0)$ the extreme terms of the above sequence cannot both vanish. By definition of r the group $\mathrm{Ext}^n_A(\mathrm{Jac}(A)\cdot N, M)$ is trivial. This in turn implies $\mathrm{Ext}^n_A(N/(\mathrm{Jac}(A)\cdot N), M) \neq (0)$, whence $r = 1$. As a result, there exists an A-module N such that $\mathrm{Ext}^n_A(N, M) \neq (0)$ and $\mathrm{Jac}(A)\cdot N = (0)$. Since $A/\mathrm{Jac}(A)$ is finite dimensional semisimple, N is a direct sum of irreducible A-modules: $N = \bigoplus_{i \in I} N_i$. According to (9.4) of [4, p. 98] we have $(0) \neq \mathrm{Ext}^n_A(N, M) \cong \prod_{i \in I} \mathrm{Ext}^n_A(N_i, M)$, proving the existence of an irreducible (and thereby necessarily finite dimensional) A-module V_n such that $\mathrm{Ext}^n_A(V_n, M) \neq (0)$. Since A is Frobenius, [6, 58.13] applies, and we may assume that V_n is a submodule of $\mathrm{soc}(A)$. The complete reducibility of $\mathrm{soc}(A)$ entails the existence of $U \subset \mathrm{soc}(A)$ such that $\mathrm{soc}(A) = V_n \oplus U$. Consequently, $\mathrm{Ext}^n_A(\mathrm{soc}(A), M) \cong \mathrm{Ext}^n_A(V_n, M) \oplus \mathrm{Ext}^n_A(U, M) \neq (0)$, as desired. □

Let (A, ε) be a supplemented F-algebra. An element $x \in A$ is called a *left integral* if $a \cdot x = \varepsilon(a)\, x\ \forall a \in A$. The subspace of left integrals will be denoted by $\int$. If A is a Frobenius algebra, then an application of [6, 58.13] shows that $\int \neq (0)$.

We say that A is a *local* algebra if it contains exactly one maximal left ideal. In that case A possesses, up to isomorphisms, only one irreducible module V. If A is also supplemented, then F is the only irreducible A-module. For future reference we note the following

Corollary 1.3 *Suppose that A is a supplemented, local Frobenius algebra. Let M be an A-module. Then M is either injective or $H^n(A, M) \neq (0)\ \forall n \geqq 1$.*

Proof. In this case we have $\mathrm{soc}(A) = \int$. □

The Frobenius algebras we shall be concerned with, possess an involution $\omega\colon A \to A$, that is, an anti-automorphism of order 2. In that case, the dual space M^* of an A-module M obtains the structure of a left A-module via $(a \cdot f)(m) := f(\omega(a)\, m)$. The natural equivalence $\mathrm{Hom}_A(N, M^*) \cong \mathrm{Hom}_A(M, N^*)$ entails the injectivity of the dual of any projective A-module.

Lemma 1.4 *Let A be a finite dimensional F-algebra with involution ω, N a finite dimensional A-module, M an arbitrary A-module.*
(1) *There exists a finite dimensional A-module Q such that $\mathrm{Ext}^n_A(M, Q) \cong \mathrm{Ext}^{n+1}_A(M, N)\ \forall n > 0$.*

(2) *If A is a Frobenius algebra, then there exists a finite dimensional A-module R such that $\mathrm{Ext}^n_A(M, N) \cong \mathrm{Ext}^{n+1}_A(M, R)\ \forall n > 0$.*

Proof. (1) By dualizing a finite dimensional projective presentation

$$(0) \to T \to P \to N^* \to (0)$$

we obtain, in accordance with the above observation, a finite dimensional injective presentation

$$(0) \to N \to I \to Q \to (0)$$

of N. As $\mathrm{Ext}_A^n(M, I) = (0)\ \forall n > 0$, the long exact cohomology sequence collapses to isomorphisms $\mathrm{Ext}_A^n(M, Q) \cong \mathrm{Ext}_A^{n+1}(M, N)\ \forall n > 0$.

(2): Let $(0) \to R \to P \to N \to (0)$ be a finite dimensional projective presentation of N. Theorem 62.11 of [6] ensures the injectivity of P and the long exact cohomology sequence furnishes the asserted isomorphisms. □

Proposition 1.5 *Let A be a supplemented Frobenius algebra. Then the following statements are equivalent:*

(1) $\varepsilon(\int) \neq (0)$ (2) $H^n(A, M) = (0)\ \forall n \geqq 1$ (3) *F is projective.*

Proof. (1)⇒(2) Let $\varphi\colon A \to A^*$ be the defining Frobenius isomorphism, $\lambda := \varphi(1) \in A^*$ the corresponding Frobenius homomorphism. Then the pairing

$$(x, y) := \lambda(yx)$$

is nondegenerate. Suppose that x_1 is a left integral such that $\varepsilon(x_1) = 1$. Let $\{x_1, \ldots, x_n\}$ be a basis for A, $\{y_1, \ldots, y_n\}$ the dual basis with respect to $(,)$. Consider the "Casimir element" $c_0 := \sum_{i=1}^{n} x_i\, y_i$. We obtain

$$\varepsilon(y_i)\, \lambda(x_1) = (x_1, y_i) = \delta_{i,1} \qquad 1 \leqq i \leqq n,$$

proving that $y_i \in \ker \varepsilon\ \forall i \neq 1$ and $\varepsilon(y_1) \neq 0$. Consequently, $\varepsilon(c_0) = \sum_{i=1}^{n} \varepsilon(x_i)\ \varepsilon(y_i) = \varepsilon(y_1) \neq 0$. According to [16, 3.6] this entails the vanishing of the relevant cohomology groups. The implication (2)⇒(3) follows directly from the definition. To verify the remaining statement, we consider the short exact sequence $(0) \to \ker \varepsilon \to A \to F \to (0)$ of A-modules. As F is projective this sequence splits, thereby ensuring the existence of a direct complement $V \cong F$ of $\ker \varepsilon$. Since the elements of V are left integrals, the assertion follows. □

For a finite dimensional Hopf algebra H the above conditions are known to be equivalent with the semisimplicity of H [cf. 31].

2 Cohomology groups of reduced enveloping algebras

This section is concerned with the cohomology theory of the reduced enveloping algebras of a restricted Lie algebras. Unless noted otherwise $(L, [p])$ is assumed to be a finite dimensional restricted Lie algebra over a base field F of positive characteristic p. We fix a p-subalgebra $K \subset L$ and a linear form $S \in L^*$ and denote the reduced universal enveloping algebras of K and L by $u(K, S|_K)$ and $u(L, S)$, respectively [cf. 30, § 5.3 for definitions]. One customarily writes $u(L)$ in case of $S = 0$. According to [30, 4.3, p. 218] $u(L, S)$ is a Frobenius algebra.

The adjoint representation induces an operation $K \to \mathrm{gl}(L/K)$ which we shall denote by $\mathrm{ad}\, L/K$. Since the mapping $K \to F$; $x \mapsto \mathrm{tr}(\mathrm{ad}\, L/K(x))$ sends $x^{[p]}$ to $(\mathrm{tr}(\mathrm{ad}\, L/K(x)))^p$, the correspondence $K \to u(K, S|_K)$; $x \mapsto x - \mathrm{tr}(\mathrm{ad}\, L/K(x))\, 1$

extends to an F-algebra automorphism σ of $u(K, S|_K)$. If V is a K-module V with character $S|_K$, then the L-module V_σ with underlying vector space V and action

$$x \cdot v := \sigma(x)\, v$$

is easily seen to have the same character. The automorphism σ endows $u(L, S)$: $u(K, S|_K)$ with the structure of a free Frobenius extension [cf. 17]. In particular, induced and coinduced modules are isomorphic and we obtain the ensuing reciprocity relations.

Lemma 2.1 *Let M be an L-module with character $S \in L^*$, V a K-module with character $S|_K$. Then there exist isomorphisms*

(1) $\mathrm{Ext}^n_{u(L,S)}(M, u(L,S) \otimes_{u(K,S|_K)} V) \cong \mathrm{Ext}^n_{u(K,S|_K)}(M, V_\sigma)\ \forall n \geqq 0.$

(2) $\mathrm{Ext}^n_{u(L,S)}(u(L,S) \otimes_{u(K,S|_K)} V, M) \cong \mathrm{Ext}^n_{u(K,S|_K)}(V, M)\ \forall n \geqq 0.$

Proof. (1): Since $u(L, S)$ is a free left $u(K, S|_K)$-module, it follows that $\mathrm{Ext}^n_{u(K,S|_K)}(u(L,S), V_\sigma) = (0)\ \forall n > 0$. Consequently, (4.1.4) of [4, p. 118] applies and we obtain isomorphisms

$$\mathrm{Ext}^n_{u(L,S)}(M, \mathrm{Hom}_{u(K,S|_K)}(u(L,S), V_\sigma)) \cong \mathrm{Ext}^n_{u(K,S|_K)}(M, V_\sigma) \qquad \forall n \geqq 0.$$

By invoking (1.5) and (1.4) of [17] consecutively, we see that the $u(L, S)$-modules $u(L,S) \otimes_{u(K,S|_K)} V$ and $\mathrm{Hom}_{u(K,S|_K)}(u(L,S), V_\sigma)$ are isomorphic.

(2): As $u(L, S)$ is free as a right $u(K, S|_K)$-module and $\otimes_{u(K,S|_K)}$ is a left balanced functor, the torsion functors $\mathrm{Tor}^{u(K,S|_K)}_n(u(L,S), M)$ vanish for $n > 0$. The isomorphisms can now be obtained from (4.1.3) of [4, p. 118]. □

In [20] Friedlander and Parshall extensively studied the reduced enveloping algebras of classical simple Lie algebras. Their algebraic groups act on L^* and every character S lies in the orbit of a character that annihilates the unipotent radical of the Borel subalgebra. This circumstance enables them to exhibite a necessary and sufficient condition for $u(L, S)$ to be semisimple. We shall study here the same problem in the context of restricted Lie algebras of Cartan type.

Suppose that $L = \bigoplus_{i=-r}^{s} L_i$ is restricted graded, i.e. $L_i^{[p]} \subset L_{ip}$ for every i. Let $L^+ := \sum_{i>0} L_i$; $L^- := \sum_{i<0} L_i$.

Recall that a restricted Lie algebra is referred to as *p-unipotent* if for every $x \in L$ there is $n(x) \in \mathbb{N}$ such that $x^{[p]^{n(x)}} = 0$.

Proposition 2.2 *Let L be a restricted graded Lie algebra of dimension l, $K \subset L_0$ a p-subalgebra. Suppose that $S \in L^*$ is a linear form such that $K + L^+ \subset \ker S$. If $u(L, S)$ is semisimple, then the following statements hold:*

(1) $l \leqq 2(\dim_F L^- + \dim_F L_0 - \dim_F K)$

(2) *Let V be an L_0-module with character $S|_{L_0}$. Then* $\dim_F V \geqq p^{\dim_F L^+ - \dim_F L^-}$.

Proof. As $u(L, S)$ is semisimple every L-module M with character S is projective. The assumption concerning S ensures that the subalgebra of $u(L, S)$ generated by $K + L^+$ is isomorphic to $u(K + L^+)$ and thereby supplemented. Owing to the Engel-Jacobson Theorem [30, I.3.6] $K + L^+$ is p-unipotent. Consequently, (1.1) applies and we obtain $\dim_F M \geqq p^{\dim_F L^+ + \dim_F K}$ or $M = (0)$.

Since $u(L, S) \otimes_{u(K+L^+)} F$ is a module with character S, the first assertion readily follows. If V is given as in (2), then V obtains the structure of a $\mathscr{B} := L_0 + L^+$-module with character S by letting L^+ operate trivially. The desired estimate now follows by specializing $M = u(L, S) \otimes_{u(\mathscr{B}, S|_{\mathscr{B}})} V$. □

Remark. If $(L, [p])$ is classical semisimple with triangular decomposition $L = \mathscr{N}^- \oplus H \oplus \mathscr{N}^+$ and $S \in L^*$ is a nilpotent character, i.e. $S(H \oplus \mathscr{N}^+) = (0)$, then (1) shows that $u(L, S)$ is not semisimple. This fact can be used to simplify the proof of [20, 3.6].

By inspection, we now obtain the ensuing result:

Corollary 2.3 *Let L be a restricted Lie algebra of Cartan type, $S \in L^*$ a linear form that vanishes on L^+. If $\operatorname{char}(F) > 3$, then $u(L, S)$ is not semisimple.* □

In the sequel we shall study the Hochschild cohomology of the algebra $u(L, S)$ [cf. 22]. Our results are contingent upon an appropriate application of the "Mapping Theorem" [cf. 4, p. 150]. We consider $u(L, S)^e := u(L, S) \otimes_F u(L, S)^{op}$, the enveloping algebra of the reduced enveloping algebra $u(L, S)$. Let M be a $u(L, S)$-bimodule. Then M obtains the structure of a restricted L-module via

$$x \cdot m := xm - mx \qquad \forall x \in L, \qquad \forall m \in M.$$

Throughout this section, $u(L, S)$-bimodules shall be considered restricted L-modules in this fashion.

Following Hochschild [23] we define the restricted cohomology groups $H^n_*(L, M)$ of L with coefficients in the restricted L-module M to be those of the Hopf algebra $u(L)$.

Theorem 2.4 *Let M be a $u(L, S)$-bimodule. Then there exists a natural isomorphism*

$$H^n(u(L, S), M) \cong H^n_*(L, M).$$

Proof. Let $\mathscr{S} \in (L \oplus L)^*$ be given by $\mathscr{S}(x, y) = S(x) - S(y)$ $\forall x, y \in L$. The mapping $\chi\colon L \oplus L \to u(L, S)^e$; $\chi(x, y) := x \otimes 1 - 1 \otimes y$ is easily seen to be a homomorphism of Lie algebras satisfying

$$\chi((x, y))^p - \chi((x, y)^{[p]}) = \mathscr{S}(x, y)^p 1 \otimes 1 \qquad \forall x, y \in L,$$

thereby admitting an associative extension $u(L \oplus L, \mathscr{S}) \to u(L, S)^e$ which we shall also denote by χ. The natural inclusions $x \mapsto (x, 0)$ and $x \mapsto (0, -x)$ induce algebra homomorphisms $\psi_1\colon u(L, S) \to u(L \oplus L, \mathscr{S})$ and $\psi_2\colon u(L, S) \to u(L \oplus L, \mathscr{S})^{op}$, respectively. Taking opposite algebras, we see that ψ_2 may also be considered an algebra homomorphism $u(L, S)^{op} \to u(L \oplus L, \mathscr{S})$. An elementary computation ensures the validity of

$$\psi_1(u)\,\psi_2(v) = \psi_2(v)\,\psi_1(u) \qquad \forall u, v \in u(L, S).$$

Consequently, there exists an F-algebra homomorphism $\psi\colon u(L, S)^e \to u(L \oplus L, \mathscr{S})$ such that $\psi(u \otimes v) = \psi_1(u)\,\psi_2(v)$ $\forall u, v \in u(L, S)$. This mapping is, as standard arguments show, the inverse of χ.

Note that the map $L \to u(L \oplus L, \mathscr{S})$; $x \mapsto (x, x)$ is a homomorphism of restricted Lie algebras. The composite of its extension $\gamma\colon u(L) \to u(L \oplus L, \mathscr{S})$ with χ defines an F-algebra homomorphism $\Delta\colon u(L) \to u(L, S)^e$ that satisfies $\Delta(x)$

$= x \otimes 1 - 1 \otimes x \ \forall x \in L$. The mappings Δ and γ endow $u(L, S)^e$ and $u(L \oplus L, \mathscr{S})$ with the structure of a right $u(L)$-module, respectively. Evidently, χ induces an isomorphism between these $u(L)$-modules. According to Jacobson's refinement of the Poincaré-Birkhoff-Witt Theorem $u(L \oplus L, \mathscr{S})$ is free over $u(L)$. Hence $u(L, S)^e$ enjoys the same property and the torsion functors $\mathrm{Tor}_n^{u(L)}(u(L, S)^e, F)$ vanish for $n > 0$.

Let τ: $u(L, S)^e \to u(L, S)$ be the canonical augmentation of $u(L, S)^e$, that is, the linear map given by $\tau(a \otimes b) = ab$. Furthermore, let ε and ι denote the canonical supplementation of $u(L)$ and the embedding $F \hookrightarrow u(L, S)$, respectively. Since Δ maps $\ker \varepsilon$ into $\ker \tau$, we have $\tau \circ \Delta = \iota \circ \varepsilon$. This qualifies Δ as a mapping of augmented rings [cf. 4, p. 149]. Consider the linear mapping g: $u(L, S)^e \otimes_{u(L)} F \to u(L, S)$ that assigns $\alpha\tau(u)$ to $u \otimes \alpha$. The kernel of τ is obviously generated by $\{u \otimes 1 - 1 \otimes u;\ u \in u(L, S)\}$. The relation

$$uv \otimes 1 - 1 \otimes uv = (u \otimes 1)(v \otimes 1 - 1 \otimes v) + (1 \otimes v)(u \otimes 1 - 1 \otimes u)$$

then shows that $\ker \tau = u(L, S)^e \Delta(L)$. This readily implies the bijectivity of g. Consequently, the requirements of the Mapping Theorem [cf. 4, 3.1, p. 150] are satisfied and we conclude that the natural homomorphism

$$F_\Delta \colon \mathrm{Ext}_{u(L,S)^e}(u(L, S), M) \to \mathrm{Ext}_{u(L)}(F, M)$$

induced by Δ is an isomorphism of degree 0 for every $u(L, S)$-bimodule M. The asserted result now follows from the usual identifications $\mathrm{Ext}^n_{u(L,S)^e}(u(L, S), M) \cong H^n(u(L, S), M)$ and $\mathrm{Ext}^n_{u(L)}(F, M) \cong H^n_*(L, M)$, respectively. □

The subsequent result constitutes a slight generalization of the first part of Proposition 5.1 of [20].

Corollary 2.5 *Let M and N be two L-modules with character S. Then there exist natural isomorphisms* $\mathrm{Ext}^n_{u(L,S)}(M, N) \cong H^n_*(L, \mathrm{Hom}_F(M, N))$.

Proof. Owing to [4, 4.4 p. 170] the groups $\mathrm{Ext}^n_{u(L,S)}(M, N)$ are isomorphic to $H^n(u(L, S), \mathrm{Hom}_F(M, N))$, where $\mathrm{Hom}_F(M, N)$ is given the natural structure of a bimodule via

$$(u \cdot f)(m) = uf(m); \qquad (f \cdot u)(m) = f(um).$$

According to (2.4) the latter groups may be identified with the restricted cohomology groups $H^n_*(L, \mathrm{Hom}_F(M, N))$. Since the L-module structure induced on $\mathrm{Hom}_F(M, N)$ by Δ coincides with the natural restricted representation $L \to \mathrm{gl}(\mathrm{Hom}_F(M, N))$, our assertion follows. □

3 Restricted cohomology

We adopt the conventions of the preceding section. Suppose F to be algebraically closed and assume that $H^1_*(L, M) = (0)$ for every finite dimensional L-module M. Given $x \in L$ such that $x^{[p]} = 0$, we consider the p-subalgebra $K := Fx$ and note that, owing to (2.1), $H^1_*(K, F) \cong H^1_*(L, u(L) \otimes_{u(K)} F) = (0)$. According to Theorem 2.1 of [23], the former space may be identified with $\{f \in K^*;\ f(x^{[p]}) = 0\ \forall x \in K\}$. Consequently, K is the trivial subspace and the p-mapping is nonsingu-

lar. It now follows from [7] that L is a torus in the sense of the definition given below.

In the sequel we shall be concerned with various generalizations of the above argument, which due to its appeal to Chwe's result, is of no avail for general fields.

For a subalgebra H of L let H_p denote the smallest p-subalgebra of L containing H.

Definition A restricted Lie algebra $(L, [p])$ is called a *torus* if

(1) L is abelian

(2) $x \in (Fx^{[p]})_p \ \forall x \in L$.

The subsequent result elaborates on a theorem due to G. Hochschild [cf. 24].

Theorem 3.1 *The following statements are equivalent:*

(1) *L is a torus*

(2) *For every finite dimensional restricted L-module M we have $H^n_*(L, M) = (0)$ $\forall n > 0$*

(3) *There is $n_0 > 0$ such that $H^{n_0}_*(L, M) = (0)$ for every finite dimensional restricted L-module M.*

Proof. (1)$\Rightarrow$(2): If L is a torus, then L is, as a vector space, generated by the set $\{x^{[p]};\ x \in L\}$. Hence the arguments of [24] ensure the semisimplicity of $u(L)$. Consequently, every $u(L)$-module M is injective, and $H^n_*(L, M) = \mathrm{Ext}^n_{u(L)}(F, M) = (0)\ \forall n > 0$.

(2)$\Rightarrow$(3): trivial

(3)$\Rightarrow$(1): According to Berkson [2], $u(L)$ is a Frobenius algebra. Hence (1.4) applies and we obtain

(a) $H^1_*(L, M) = (0)$ for every finite dimensional restricted L-module M

(b) $H^2_*(L, F) = (0)$.

Condition (b) implies by virtue of [10, 3.1] [cf. also 4.3 below], that L is abelian. Let x be an element of L and consider $K := (Fx)_p$. According to (2.1) we have $H^1_*(K, F) \cong H^1_*(L, u(L) \otimes_{u(K)} F_\sigma) = (0)$. If x does not belong to $(Fx^{[p]})_p$, then $K = Fx \oplus (Fx^{[p]})_p$ and there is $f \in K^*$ such that $f(x) = 1$, $f((Fx^{[p]})_p) = (0)$. Owing to (2.1) of [23] f defines a nonzero element of $H^1_*(K, F)$, a contradiction. Consequently, L is a torus. □

We briefly digress in order to illustrate that the preceding result is not transferable to infinite dimensional algebras. Following [30] we let $\hat{L}$ and $\mathscr{U}(L)$ denote the universal p-envelope of L and the universal enveloping algebra of L, respectively.

Proposition 3.2 *Let L be an arbitrary Lie algebra. Then*

(1) *$\mathscr{U}(L)$ and $u(\hat{L})$ are isomorphic as supplemented algebras.*

(2) *For every L-module M, we have $H^n(L, M) \cong H^n_*(\hat{L}, M)\ \forall n \geqq 0$.*

Proof. (1): The canonical inclusion $L \hookrightarrow \hat{L}$ may be extended to a homomorphism $f\colon L \to u(\hat{L})$ of Lie algebras. Note that $\delta(f(x)) = 0 = \varepsilon(x)\ \forall x \in L$, where δ and ε denote the supplementations of $u(\hat{L})$ and $\mathscr{U}(L)$, respectively. The mapping f extends uniquely to an F-algebra homomorphism $\varphi\colon \mathscr{U}(L) \to u(\hat{L})$. By the above, the subalgebra $A := \{u \in \mathscr{U}(L);\ \delta(\varphi(u)) = \varepsilon(u)\}$ contains L and thereby coincides with $\mathscr{U}(L)$. Consequently, φ is a homomorphism of supplemented algebras. Since the canonical inclusion $\hat{L} \hookrightarrow \mathscr{U}(L)$ is a homomorphism of restricted Lie

algebras it extends to an F-algebra homomorphism $\psi\colon u(\hat{L})\to\mathscr{U}(L)$. For $x\in L$ we have $\psi(\varphi(x))=\psi(f(x))=x$, proving that $\psi\circ\varphi=\mathrm{id}_{\mathscr{U}(L)}$. If x lies in L, then $(\varphi\circ\psi)(x)=\varphi(x)=f(x)=x$ and the p-subalgebra $\{x\in\hat{L};\ (\varphi\circ\psi)(x)=x\}$ of $\hat{L}$ contains L, hence coincides with $\hat{L}$. Consequently, $A:=\{u\in u(\hat{L});\ (\varphi\circ\psi)(u)=u\}$ contains the generating set $\hat{L}$. We therefore also have $\varphi\circ\psi=\mathrm{id}_{u(\hat{L})}$, qualifying φ as an isomorphism.

(2): Let M be an L-module. Then M is a $\mathscr{U}(L)$-module and hence a restricted $\hat{L}$-module. The $\mathscr{U}(L)$-module structure on M induced by φ coincides with the originally given one and we thus obtain $H^n(L,M)\cong H^n_*(\hat{L},M)\ \forall n\geqq 0$. □

Example. Let L be a Lie algebra of dimension k. The above result implies that $H^n_*(\hat{L},M)=(0)\ \forall n\geqq k+1$. In particular, let L be the two-dimensional non-abelian Lie algebra. Then we obtain

(1) $H^3_*(\hat{L},M)=(0)$ for every restricted $\hat{L}$-module M.

(2) $H^2_*(\hat{L},F)=(0)$.

Consequently, parts of (3.1) of [10] lose their validity for infinite dimensional restricted Lie algebras.

Given a finite dimensional restricted Lie algebra $(L,[p])$, the integer $rk(L):=\max\{\dim_F T;\ T\subset L \text{ torus}\}$ is called the *rank* of L. Owing to [30, 3.4 p. 80] L is p-unipotent if and only if $rk(L)=0$.

Proposition 3.3 *Let M be a restricted L-module. Then the following statements hold*:

(1) *M is either injective or $H^n_*(L,\mathrm{soc}(u(L))\otimes_F M)\neq(0)\ \forall n\geqq 1$.*

(2) *If L is p-unipotent, then M is either injective or $H^n_*(L,M)\neq(0)\ \forall n\geqq 1$.*

Proof. (1): Let $\{V_1,\ldots,V_t\}$ be a complete set of representatives of irreducible $u(L)$-modules. Since the $(V_i)^*$ also form such a set, it follows that the modules $\bigoplus_{i=1}^{t} V_i$ and $\bigoplus_{i=1}^{t}(V_i)^*$ are isomorphic. Owing to [6, 58.13] these modules are direct summands of $\mathrm{soc}(u(L))$ and $\mathrm{soc}(u(L))^*$, respectively. Suppose that M is not injective. According to (1.2) we obtain $\mathrm{Ext}^n_{u(L)}(\mathrm{soc}(u(L)),M)\neq(0)$ for $n\geqq 1$. The above observations then show that $\mathrm{Ext}^n_{u(L)}(\mathrm{soc}(u(L))^*,M)\neq(0)$ for $n\geqq 1$. Corollary 2.5 now provides isomorphisms

$$H^n_*(L,\mathrm{soc}(u(L))\otimes_F M)\cong H^n_*(L,\mathrm{Hom}_F(\mathrm{soc}(u(L))^*,M))\cong\mathrm{Ext}^n_{u(L)}(\mathrm{soc}(u(L))^*,M)$$

for every n.

(2): Since L is p-unipotent, the Engel-Jacobson theorem [cf. 30, 3.6 p. 18] ensures the nilpotency of the supplementation ideal $u(L)^+$. Consequently, $u(L)$ is a local algebra and an application of (1.3) yields the desired result. □

Given a restricted Lie algebra $(L,[p])$ with nilradical $N(L)$, we consider $\Omega:=\{u\in\mathrm{soc}(u(L));\ N(L)\cdot u=0\}$. Note that Ω is a $u(L)$-submodule of $\mathrm{soc}(u(L))$. Let $\mathscr{G}$ be the set of all F-algebra homomorphisms from $u(L)$ to F.

Theorem 3.4 *Let $(L,[p])$ be finite a dimensional non-toral Lie algebra. Then the following statements hold*:

(1) *$H^n_*(L,\Omega)\neq(0)\ \forall n\geqq 0$.*

(2) *If L is completely solvable and F is algebraically closed, then $H^n_*(L,\bigoplus_{\lambda\in\mathscr{G}}F_\lambda)\neq(0)\ \forall n\geqq 0$.*

(3) *If L is nilpotent, then $H^n_*(L, F) \neq (0)\ \forall n \geqq 0$.*

Proof. (1) Since $H^0_*(L, F) \cong \mathrm{Hom}_{u(L)}(F, F) \neq (0)$ it readily follows from (3) of (3.1) that there exists a sequence $(Q_n)_{n \in \mathbb{N}_0}$ of finite dimensional restricted L-modules such that $H^n_*(L, Q_n) \neq (0)\ \forall n \in \mathbb{N}_0$. Each of the Q_n possess an irreducible subquotient M_n satisfying $H^n_*(L, M_n) \neq (0)$. Corollary 1.4 of [15] now ensures that the Fitting null space of M_n relative to $N(L)$ does not vanish. As M_n is irreducible, it coincides with its Fitting null space and $N(L)$ operates nilpotently on M_n. An application of Engel's theorem then yields $N(L) \cdot M_n = (0)\ \forall n \geqq 0$. Hence there is a direct summand I_n of Ω such that $H^n_*(L, I_n) \neq (0)$. Consequently, $H^n_*(L, \Omega) \neq (0)\ \forall n \geqq 0$.

(2) If L is completely solvable, then the first derived algebra $L^{(1)}$ is by definition nilpotent. As F is algebraically closed, every irreducible submodule of Ω is isomorphic to F_λ for a suitably chosen $\lambda \in \mathscr{G}$. Consequently, $\Omega = \bigoplus_{\lambda \in \mathscr{G}} (F_\lambda)^{n_\lambda}$.

(3) If L is nilpotent, then $\Omega = \int$, a direct sum of copies of F. □

Remark. Statement (3) loses its validity for completely solvable Lie algebras. Consider for example $\mathscr{B} := Fh \oplus Fx$, $[h, x] = x$, $h^{[p]} = h$, $x^{[p]} = 0$, the Borel subalgebra of $\mathrm{sl}(2, F)$. Using (3.1) of [23] one computes $H^1_*(\mathscr{B}, F) = (0)$. By the same token we have $H^1_*(Fx, F) = F$. The latter observation falsifies (1.4) of [26, p. 321], where the canonical action of Fh on $H^1_*(Fx, F)$ is erroneously claimed to be trivial.

The $u(L)$-module $u(L)$ is projective and therefore injective. Thus, the cohomology groups $H^n_*(L, u(L))$ vanish for $n > 0$ and

$$q(L) := \min\{\dim_F M;\ M \neq (0),\ H^n_*(L, M) = (0)\ \forall n > 0\}$$

is well defined and bounded by the dimension of $u(L)$.

The injectivity of $u(L)$ also shows that Dixmier's result [9, Theor. 2] possesses no analogue in the theory of restricted Lie algebras. If $(L, [p])$ is p-unipotent, then L acts nilpotently on $u(L)$ via the left multiplication, while $H^n_*(L, u(L)) = (0)$ for $n > 0$. We put

$$N(L) := \{x \in L;\ \exists n \in \mathbb{N}: x^{[p]^n} = 0\}$$
$$S(L) := \{x \in L;\ x \in (Fx^{[p]})_p\}.$$

The elements of $S(L)$ are usually referred to as *semisimple elements*. Let $\mathscr{T}$ be the set of maximal tori of L. For $T \in \mathscr{T}$ we define

$$\mathscr{H}(L, T) := \{P;\ P \subset L \text{ nilpotent } p\text{-subalgebra},\ T \subset P,\ T \notin [P, P]_p\}$$

as well as $r(L) := \max\{\dim_F P;\ P \in \bigcup_{T \in \mathscr{T}} \mathscr{H}(L, T)\}$ if $\bigcup_{T \in \mathscr{T}} \mathscr{H}(L, T)$ is not empty and $r(L) := 0$ otherwise. Let T be an element of $\mathscr{T}$. For $P \in \mathscr{T}$ one has $P \subset C_L(T)$, the centralizer of T in L. The latter algebra is known to be a Cartan subalgebra of L. Consequently, $r(L)$ is bounded by the maximal dimension $c(L)$ of Cartan subalgebras of L. Following Block-Wilson [3] we call a maximal torus T *standard* if $[C_L(T), C_L(T)]_p$ is p-unipotent. If $T \neq (0)$, then $C_L(T)$ belongs to $\mathscr{H}(L, T)$. Wilson proved in [32] that for simple L and $p > 7$ every maximal torus of L is standard. The same applies to completely solvable, non-nilpotent Lie algebras satisfying $C(L)^{[p]} = (0)$. In either case we obtain $r(L) = c(L)$.

Theorem 3.5 *Let $(L, [p])$ be a restricted Lie algebra of dimension l over an algebraically closed field F of characteristic $p>2$. Then the following statements hold:*

(1) $q(L) \leqq p^{l-r(L)}$

(2) *If L is nilpotent, then $q(L)=p^{l-r(L)}$.*

Proof. (1): If $\bigcup_{T\in\mathscr{T}} \mathscr{H}(L, T)$ is empty, then (1) is trivial. Otherwise let $T\in\mathscr{T}$ be a maximal torus, $P\in\mathscr{H}(L, T)$ a nilpotent p-subalgebra of dimension $r(L)$ and consider the canonical projection $\pi\colon P\to P/[P, P]_p$. As P is nilpotent, $T\subset P$ is the unique maximal torus of P. According to [30, II.4.5] $\pi(T)$ is a maximal torus of the abelian Lie algebra $P/[P, P]_p$. An application of [30, II.3.6] now ensures the existence of a non-trivial homomorphism $\lambda\colon u(P/[P, P]_p)\to F$. Let F_λ denote the corresponding one-dimensional P-module. Owing to (2.1) we have isomorphisms

$$H^n_*(L, u(L)\otimes_{u(P)} F_{\lambda-\sigma})\cong H^n_*(P, F_\lambda) \quad \text{for } n\geqq 0.$$

By construction F_λ has a trivial Fitting-Null-Space. Therefore the latter groups vanish on account of [15, 1.4]. Consequently, $q(L)\leqq \dim_F u(L)\otimes_{u(P)} F_{\lambda-\sigma}=p^{l-r(L)}$.

(2): Suppose L to be nilpotent. Let M be a finite dimensional $u(L)$-module such that $\dim_F M=q(L)$ and $H^n_*(L, M)=(0)$ $\forall n>0$. Since L is nilpotent, $T:=S(L)\subset C(L)$ is the unique maximal torus of L. Moreover [30, II.4.5] entails that the factor algebra L/T is p-unipotent. Since $H^n_*(L, M^T)\cong H^n_*(L, M)=(0)$ for all $n>0$ we obtain $M^T=(0)$ or $M^T=M$. We first discuss the latter case. As T is a torus, the Grothendieck spectral sequence

$$H^p_*(L/T, H^q_*(T, M))\Rightarrow H^n_*(L, M),$$

collapses to isomorphisms $H^n_*(L/T, M)\cong H^n_*(L, M)$. By [3.3 (2)] and [6, 62.11], M is a projective $u(L/T)$-module. According to the Engel-Jacobson Theorem the supplementation ideal $u(L/T)^+$ is nilpotent. Thus Proposition 1.1 implies the estimate $\dim_F M\geqq p^{l-rk(L)}\geqq p^{l-r(L)}$. Now assume that $M^T=(0)$ and let $V\subset M$ be an irreducible submodule. Then $V^T=(0)$ and [15, 1.4] gives rise to $H^n_*(L, V)=(0)$ $\forall n>0$. Consequently, $M=V$ is irreducible. According to [29, Satz 3] there exists a p-subalgebra $P\subset L$ and an F-algebra homomorphism $\lambda\colon u(P)\to F$ such that $M\cong u(L)\otimes_{u(P)} F_\lambda$. Lemma 2 of [29] readily implies $C(L)\subset P$. The assumption $T\subset[P, P]_p$ entails the inclusion $T\subset\ker\lambda$. Hence $t\cdot(1\otimes 1)=t\otimes 1=1\otimes t\cdot 1=\lambda(t)\cdot(1\otimes 1)=0$ for every $t\in T$, contradicting the triviality of M^T. It now follows that $r:=\dim_F P\leqq r(L)$, whence $\dim_F M=p^{l-r}\geqq p^{l-r(L)}$. We therefore have $q(L)\geqq p^{l-r(L)}$ in either case, which in view of (1) proves the assertion. □

Given an arbitrary L-module M, we denote the n-th cohomology group of L with coefficients in M by $H^n(L, M)$ [cf. 5]. These groups are known to coincide with $\mathrm{Ext}^n_{u(L)}(F, M)$.

Remarks. (1) It follows from (1) of (3.5) that L is p-unipotent, whenever $q(L)=p^l$.

(2) Suppose that $p>2$ and let $(L, [p])$ be a restricted Lie algebra such that

(a) $\mathrm{rk}(L)\neq 0$ (b) every maximal torus $T\in\mathscr{T}$ is standard.

Let H be a Cartan subalgebra of dimension $c(L)$. Then there exists $T\in\mathscr{T}$ such that $H\in\mathscr{H}(L, T)$. Since T is standard, the roots of H are linear functionals. Let α be a root. As $[H, H]$ acts nilpotently on the root space L_α, there exists an H-eigenvector $x\in L_\alpha$. Note that $x^{[p]}\in T$, $\alpha(x^{[p]})=0$. Put $\lambda:=2\alpha$ and consider

$\mathscr{B} := H \oplus Fx$. We extend λ to $\mathscr{B}$ by setting $\lambda(x)=0$. Then λ determines an F-algebra homomorphism $u(\mathscr{B}) \to F$, which we shall also denote by λ. Since there are H-module isomorphisms $H^0(Fx, F_\lambda) \cong F_\lambda$ and $H^1(Fx, F_\lambda) \cong F_\alpha$, [15, 1.4] shows that the Hochschild-Serre spectral sequence $H^p(H, H^q(Fx, F_\lambda)) \Rightarrow H^n(L, M)$ collapses, whence we conclude the vanishing of the latter groups. Remark 1.2 of [18] then implies $H^n_*(\mathscr{B}, F_\lambda)=(0)\ \forall n \geqq 0$. It now follows from (2.1) that $q(L) \leqq \dim_F u(L) \otimes_{u(\mathscr{B})} F_{\lambda-\sigma} = p^{l-c(L)-1}$.

4 Connections with ordinary cohomology

The succeeding results elaborate on those obtained by Friedlander and Parshall in [20]. The exact sequence displayed in (4.3) essentially coincides with the first 5 terms of Hochschild's exact sequence [cf. 23, p. 575]. In the following let $(L, [p])$ be a restricted Lie algebra of arbitrary dimension over a (not necessarily algebraically closed) field F of positive characteristic $p>0$.

Theorem 4.1 *Let $K \subset L$ be a p-ideal, $\{e_i;\ i \in I\}$ a cobasis of K in L. Suppose that M is an L-module such that $e_i^p - e_i^{[p]}$ annihilates M for every i. Then there exists a third quadrant spectral sequence*

$$\bigoplus_{i+j=q} \operatorname{Hom}_F(\Lambda^i(L/K), H^p_*(L/K, H^j(K, M))) \Rightarrow H^n(L, M).$$

Proof. We consider the subalgebra $\mathscr{O}(L, K)$ of $\mathscr{U}(L)$ that is generated by $K \cup \{e_i^p - e_i^{[p]};\ i \in I\}$, and note that $\mathscr{O}(L, K)$ is a sub Hopf algebra of $\mathscr{U}(L)$. Let $\mathscr{R}$ and $\mathscr{J}$ be the subalgebra and the ideal generated by $\{e_i^p - e_i^{[p]};\ i \in I\}$, respectively. For any F-algebra A, we let Mod_A denote the category of left A-modules. Define functors $\mathscr{G}\colon \operatorname{Mod}_{\mathscr{U}(L)} \to \operatorname{Mod}_{u(K/L)}$ and $\mathscr{F}\colon \operatorname{Mod}_{u(L/K)} \to \operatorname{Mod}_F$ by means of $\mathscr{G}(\mathscr{M}) := \operatorname{Hom}_{\mathscr{O}(L,K)}(F, \mathscr{M})$ and $\mathscr{F}(\mathscr{N}) := \operatorname{Hom}_{u(L/K)}(F, \mathscr{N})$, respectively. Since $\mathscr{F} \circ \mathscr{G} = \operatorname{Hom}_{\mathscr{U}(L)}(F, \cdot)$, there results a Grothendieck spectral sequence

$$H^p_*(L/K, H^q(\mathscr{O}(L, K), \mathscr{M})) \Rightarrow H^n(L, \mathscr{M})$$

for every L-module $\mathscr{M}$.

According to Jacobson's refinement of the Poincaré-Birkhoff-Witt theorem the multiplication von $\mathscr{U}(L)$ induces an F-algebra isomorphism

$$\mathscr{O}(L, K) \cong \mathscr{R} \otimes_F \mathscr{U}(K),$$

with $\mathscr{R}$ being isomorphic to the universal enveloping algebra of the abelian Lie algebra with underlying vector space L/K. Our assumption concerning M ensures that M is $\mathscr{O}(L, K)$-isomorphic to $\operatorname{Hom}_F(F, M)$, where the latter space is given an $\mathscr{O}(L, K)$-structure via

$$((r \otimes u) \cdot f)(\alpha) := uf(r\alpha).$$

Given L-modules A, A', C and C', we note that the natural transformation

$$\varphi\colon \operatorname{Hom}_{\mathscr{O}(L,K)}(A \otimes_F A; \operatorname{Hom}_F(C, C')) \to \operatorname{Hom}_F(C \otimes_{\mathscr{R}} A, \operatorname{Hom}_{\mathscr{U}(K)}(A', C'));$$

defined via

$$\varphi(f)(c\otimes a)(a') := f(a\otimes a')(c)$$

is an isomorphism of L-modules. Let $X := (X_i)_{i\geqq 0}$ be the standard $\mathscr{U}(L/K)$-projective resolution of F. Jacobson's refinement of the Poincaré-Birkhoff-Witt theorem shows that the canonical projection $\mathscr{U}(L)\to\mathscr{U}(L/K)$ induces an imbedding $\mathscr{R}\hookrightarrow\mathscr{U}(L/K)$ such that $\mathscr{U}(L/K)$ is a free $\mathscr{R}$-module. It follows that the pull-back from L/K to L endows each X_i with the structure of a projective $\mathscr{R}$-module. An analogous argument entails that a $\mathscr{U}(L)/\mathscr{J}$-projective resolution $Y := (Y_j)_{j\geqq 0}$ of F is $\mathscr{U}(K)$-projective. Since the restriction of the L-action on $X_i\otimes_F Y_j$ to $\mathscr{O}(L, K)$ coincides with the one defined by

$$(r\otimes u)\cdot(x\otimes y) := rx\otimes uy,$$

an application of [4, 1.1 p. 203] ensures that the complex $X\otimes_F Y$ is an $\mathscr{O}(L, K)$-projective resolution of F. The mapping φ induces an isomorphism

$$\mathrm{Hom}_{\mathscr{O}(L,K)}(X\otimes_F Y, \mathrm{Hom}_F(F, M))\cong\mathrm{Hom}_F(F\otimes_{\mathscr{R}} X, \mathrm{Hom}_{\mathscr{U}(K)}(Y, M))$$

of complexes. Note that this isomorphism respects the canonical $u(L/K)$-structures. Künneth's Theorem now provides isomorphisms

$$H^q(\mathscr{O}(L, K), M)\cong\bigoplus_{i+j=q}\mathrm{Hom}_F(\mathrm{Tor}_i^{\mathscr{R}}(F, F), H^j(K, M)).$$

In order to determine the action of L/K on $\mathrm{Tor}_i^{\mathscr{R}}(F, F)$ we consider the functor $\mathscr{P}$: $\mathrm{Mod}_{\mathscr{R}}\to\mathrm{Mod}_F$; $\mathscr{P}(M)=F\otimes_{\mathscr{R}} M$. Suppose that V is a left L/K-module. As $\mathscr{R}$ lies in the center of $\mathscr{U}(L)$, every $x\in L/K$ defines a transformation τ_x: $\mathscr{P}(V)\to\mathscr{P}(V)$; $\tau_x(\alpha\otimes v)=\alpha\otimes x\cdot v$, which is natural with respect to $\mathscr{U}(L/K)$-linear maps. Let $L_i\mathscr{P}$ denote the i-th left derived functor of $\mathscr{P}$. Since $\mathscr{U}(L/K)$-projective resolutions are $\mathscr{R}$-projective, the transformation τ_x induces actions on $L_i\mathscr{P}(V)$ for all $i\geqq 0$. Note that τ_x coincides on $\mathscr{P}(V)$ with $\mathscr{P}(\mu_x)$, where μ_x: $V\to V$ is the $\mathscr{R}$-linear map given by $\mu_x(v)=x\cdot v$. Consequently, the L/K-action induced by τ_x on $L_i\mathscr{P}(V)$ coincides with $L_i\mathscr{P}(\mu_x)$. The specialization $V=F$ now shows that L/K operates trivially on $\mathrm{Tor}_i^{\mathscr{R}}(F, F)=L_i\mathscr{P}(F)$. Since $\mathscr{R}$ is the enveloping algebra of the abelian Lie algebra with underlying vector space L/K, we obtain linear isomorphisms $\mathrm{Tor}_i^{\mathscr{R}}(F, F)\cong\Lambda^i(L/K)$.

Now let $Z := (Z_i)_{i\geqq 0}$ be a $u(L/K)$-projective resolution of F. Owing to [4, p. 32] there are isomorphisms of complexes

$$\mathrm{Hom}_{u(L/K)}(Z, \mathrm{Hom}_F(\Lambda^i(L/K), Q))\cong\mathrm{Hom}_F(\Lambda^i(L/K), \mathrm{Hom}_{u(L/K)}(Z, Q)),$$

where $Q := H^j(K, M)$. By passing to the corresponding cohomology groups, while observing [4, 7.2 p. 68], we obtain isomorphisms

$$\begin{aligned}&\mathrm{Ext}_{u(L/K)}(F, \mathrm{Hom}_F(\Lambda^i(L/K), H^j(K, M)))\\ &\quad\cong H(\mathrm{Hom}_F(\Lambda^i(L/K), \mathrm{Hom}_{u(L/K)}(Z, H^j(K, M))))\\ &\quad\cong\mathrm{Hom}_F(\Lambda^i(L/K), \mathrm{Ext}_{u(L/K)}(F, H^j(K, M))).\end{aligned}$$

Evidently, this results in isomorphisms

$$H^p_*(L/K, \mathrm{Hom}_F(\Lambda^i(L/K), H^j(K, M))) \cong \mathrm{Hom}_F(\Lambda^i(L/K), H^p_*(L/K, H^j(K, M))) \quad \forall p \geqq 0.$$

As a result the initial terms of the above spectral sequence are of the desired form. □

Corollary 4.2 *Let $K \subset L$ be a p-ideal, $\{e_i;\ i \in I\}$ a cobasis of K in L. Suppose that M is an L-module such that $e_i^p - e_i^{[p]}$ annihilates M for every i. If L/K is a finite dimensional torus, then there are isomorphisms*

$$\bigoplus_{i+j=n} \Lambda^i(L/K) \otimes_F H^j(K, M)^L \cong H^n(L, M).$$

Proof. Since L/K is a torus $u(L/K)$ is semisimple and $H^p_*(L/K, H^j(K, M)) = (0)$ for $p > 0$. Consequently, the spectral sequence of (4.1) collapses to the isomorphisms of (4.2). □

Corollary 4.3 *Let $K \subset L$ be a p-ideal, $\{e_i;\ i \in I\}$ a cobasis of K in L. Suppose that M is an L-module such that $e_i^p - e_i^{[p]}$ annihilates M for every i. Then there exists an exact sequence* $(0) \to H^1_*(L/K, M^K) \to H^1(L, M) \to H^1(K, M)^L \oplus \mathrm{Hom}_F(L/K, M^L) \to H^2_*(L/K, M^K) \to H^2(L, M)$.

Proof. With the notation of (4.1) we have $\mathcal{G}(M) = M^K$. The Cohomology Five Term Sequence [cf. 27, p. 305] therefore gives rise to an exact sequence $(0) \to H^1_*(L/K, M) \to H^1(L, M) \to H^1(\mathcal{O}(L, K), M)^{L/K} \to H^2_*(L/K, M^K) \to H^2(L, M)$. By the proof of (4.1) the L/K-module $H^1(\mathcal{O}(L, K), M)$ is isomorphic to $H^1(K, M) \oplus \mathrm{Hom}_F(L/K, M^K)$. Since L/K acts trivially on the first entry of the second summand, the assertion follows. □

Remark. By specializing $K = (0)$ one can retrieve the following well-known cohomological characterization of finite dimensional tori. Suppose that $H^2_*(L, F) = (0)$. Then (4.3) results in an exact sequence

$$(0) \to H^1_*(L, F) \to L/[L, L] \to L \to (0).$$

Consequently, L is abelian and $H^1_*(L, F)$ vanishes. Owing to [23, 2.1] we have $H^1_*(L, F) \cong L/\langle L^{[p]} \rangle$. According to [30, II.3.2] the identity $L = \langle L^{[p]} \rangle$ entails that L is a torus.

The following result improves (5.3) of [20] by relaxing a constraint on the relevant modules.

Corollary 4.4 *Let M and N be L-modules of character S. Then there exists a spectral sequence* $\mathrm{Hom}_F(\Lambda^q(L), \mathrm{Ext}^p_{u(L,S)}(M, N)) \Rightarrow \mathrm{Ext}^n_{\mathfrak{A}(L)}(M, N)$.

Proof. According to (2.5) there exist isomorphisms

$$\mathrm{Hom}_F(\Lambda^q(L), \mathrm{Ext}^p_{u(L,S)}(M, N)) \cong \mathrm{Hom}_F(\Lambda^q(L), H^p_*(L, \mathrm{Hom}_F(M, N))).$$

The usual equivalence $\mathrm{Ext}^n_{\mathfrak{A}(L)}(M, N) \cong H^n(L, \mathrm{Hom}_F(M, N))$ in conjunction with (4.1) yields the desired result. □

Remark. According to (4.4) the inferences of Corollary 5.4 of [20] continue to be valid for modules of arbitrary dimensions.

References

1. Abe, E.: Hopf algebras. Cambridge: Cambridge University Press 1980
2. Berkson, A.J.: The u-Algebra of a Restricted Lie Algebra is Frobenius. Proc. Am. Math. Soc. **15**, 14–15 (1964)

3. Block, R., Wilson, R.: Classification of the Restricted Simple Lie Algebras. J. Algebra **114**, 115–259 (1988)
4. Cartan, H., Eilenberg, S.: Homological Algebra. Princeton, N.J.: Princeton University Press 1956
5. Chevalley, C., Eilenberg, S.: Cohomology Theory of Lie Groups and Lie Algebras. Trans. Am. Math. Soc. **63**, 85–124 (1948)
6. Curtis, C., Reiner, I.: Representation Theory of Finite Groups and Associative Algebras. New York: Interscience 1962
7. Chwe, B.S.: On the Commutativity of Restricted Lie Algebras. Proc. Am. Math. Soc. **16**, 547 (1965)
8. Chwe, B.S.: Relative Homological Algebra and Homological Dimension of Lie Algebras. Trans. Am. Math. Soc. **117**, 477–493 (1965)
9. Dixmier, J.: Cohomologie des algèbres de Lie nilpotentes. Acta Math. (Szeged) **16**, 246–250 (1955)
10. Farnsteiner, R.: Conditions for the Commutativity of Restricted Lie Algebras. Commun. Algebra **13**, 1475–1489 (1985)
11. Farnsteiner, R.: On the Cohomology of Associative Algebras and Lie Algebras. Proc. Am. Math. Soc. **99**, 415–420 (1987)
12. Farnsteiner, R.: On the Vanishing of Homology and Cohomology Groups of Associative Algebras. Trans. Am. Math. Soc. **306**, 651–665 (1988)
13. Farnsteiner, R.: Cohomology Groups of Infinite Dimensional Algebras. Math. Z. **199**, 407–423 (1988)
14. Farnsteiner, R.: Beiträge zur Kohomologietheorie assoziativer Algebren. Habilitationsschrift Hamburg 1989
15. Farnsteiner, R.: Lie Theoretic Methods in Cohomology Theory. (Lect. Notes Math., vol. 1373 pp. 93–110) Berlin Heidelberg New York: Springer 1989
16. Farnsteiner, R.: On the Cohomology of Ring Extensions. Adv. Math (to appear)
17. Farnsteiner, R., Strade, H.: Shapiro's lemma and its consequences in the cohomology theory of modular Lie Algebras. Math. Z. **206**, 153–168 (1991)
18. Friedlander, E.M., Parshall, B.J.: Cohomology of Infinitesimal and Discrete Groups. Math. Ann. **273**, 353–374 (1986)
19. Friedlander, E.M., Parshall, B.J.: Cohomology of Lie Algebras and Algebraic Groups. Am. J. Math. **108**, 235–253 (1986)
20. Friedlander, E.M., Parshall, B.J.: Modular Representation Theory of Lie Algebras. Am. J. Math. **110**, 1055–1094 (1988)
21. Hilton, P.J., Stammbach, U.: A Course in Homological Algebra. (Graduate Texts, vol. 4) Berlin Heidelberg New York: Springer 1970
22. Hochschild, G.P.: On the Cohomology Groups of an Associative Algebra. Ann. Math. **46**, 58–67 (1945)
23. Hochschild, G.P.: Cohomology of Restricted Lie Algebras. Am. J. Math. **76**, 555–580 (1954)
24. Hochschild, G.P.: Representations of Restricted Lie Algebras of Characteristic *p*. Proc. Am. Math. Soc. **16**, 603–605 (1954)
25. Humphreys, J.E.: Restricted Lie Algebras (and beyond). Contemp. Math. **13**, 91–98 (1982)
26. Pareigis, B.: Kohomologie von Lie-*p*-Algebren. Math. Z. **104**, 281–336 (1968)
27. Rotman, J.: An Introduction to Homological Algebra. Orlando, Academic Press, vol. 85, 1979
28. Seligman, G.B.: Modular Lie Algebras. (Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 40) Berlin Heidelberg New York: Springer 1967
29. Strade, H.: Darstellungen auflösbarer Lie-*p*-Algebren. Math. Ann. **232**, 15–32 (1978)
30. Strade, H., Farnsteiner, R.: Modular Lie Algebras and their Representations. New York: Marcel Dekker Textbooks and Monographs 116 (1988)
31. Sweedler, M.: Integrals for Hopf Algebras. Ann. Math. **89**, 323–335 (1969)
32. Wilson, R.L.: Cartan Subalgebras of Simple Lie Algebras. Trans. Am. Math. Soc. **234**, 435–447 (1977)
33. Zassenhaus, H.: The Representations of Lie Algebras of Prime Characteristics. Proc. Glasgow Math. Ass. **21**, 1–36 (1954)

Math. Z. 206, 119–130 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

An inverse problem of the wave equation

A.G. Ramm [1] **and J. Sjöstrand** [2]

[1] Mathematics Department, Kansas State University, Manhattan, KS, 66506, USA
[2] Mathematics Department, University of Paris-Sud, U.R.A. 760-C.N.R.S., F-91405 Orsay Cedex, France

Received October 10, 1989; in final form March 3, 1990

0 Introduction

The inverse problems of finding unknown coefficients of a PDE from boundary data are of interest in many applications. One can mention geophysics, underwater acoustics, remote sensing, nondestructive evaluation and many other areas.

Recently such type of problems received much attention. In particular, in [1]–[2], [4]–[12] a number of problems of the above type has been considered. In this paper we are interested in the problem of finding the unknown coefficient $q(x, t)$ in the equation $(1.1)_q$ in Sect. 1, given the boundary data described in Theorem 1.1, in Sect. 1.

A similar problem in which $q=q(x)$ does not depend on t has been considered earlier in [6] by the method used in [13]. This method can be simplified considerably. This simplification is given in [5] and [7]. The referee of this paper has also brought to our attention that similar ideas are developed in the time-independent case in [15]. A general approach to the inverse problems based on completeness of the set of products of solution to PDE is developed in [8]–[11]. After this work has been completed we learned about the work of Stefanov [12] which essentially contains our Proposition 2.2. Stefanov [12] studies the inverse scattering problem for time-dependent potentials.

This paper is organized as follows: in Sect. 1 the statement of the inverse problem is formulated, the basic results are formulated and proved. In Sect. 2 some auxiliary material on the X-ray transform is given in a self-contained way for convenience of the reader. This material is used in section one.

1 Statement and proof of the main result

Let $D\subset\mathbb{R}^n$, $n\geqq 2$ be open connected bounded with smooth boundary. If $q\in C^\infty(\mathbb{R}\times\bar{D})$ and $f\in C_0^\infty(\mathbb{R}\times\partial D)$, the problem

$$(1.1)_q \qquad (\Box+q)u=0 \quad \text{in } \mathbb{R}\times D, \quad u|_{\mathbb{R}\times\partial D}=f,$$

has a unique solution $u \in C^\infty(\mathbb{R} \times \bar{D})$ with the property that u vanishes for large negative t. Here $\Box = \partial_t^2 - \sum_1^n \partial_{x_j}^2$ is the d'Alembertian. We call u the forward solution of $(1.1)_q$. Similarly, there is a unique backward solution $v \in C^\infty(\mathbb{R} \times \bar{D})$ of $(1.1)_q$, that is: a solution which vanishes for large positive t. Using the property of finite speed of propagation for supports, we know that u vanishes on $\{(t, x) \in \mathbb{R} \times D;\ \mathrm{dist}_D(x, \Pi_x(\mathrm{supp} f)) > t - t_-\}$, where dist_D denotes the distance for the standard metric in D, and $\Pi_x(\mathrm{supp} f) = \{x \in \partial D;\ \exists t \in \mathbb{R}$ with $(t, x) \in \mathrm{supp} f\}$. Moreover $t_- = \inf \Pi_t(\mathrm{supp} f)$, where $\Pi_t(\mathrm{supp} f) = \{t \in \mathbb{R};\ \exists x \in \partial D$ with $(t, x) \in \mathrm{supp} f\}$.

To q, we associate the set of data

$$\left\{ \left(u|_{\mathbb{R} \times \partial D}, \left. \frac{\partial u}{\partial N} \right|_{\mathbb{R} \times \partial D} \right);\quad u \in C^\infty(\mathbb{R} \times \bar{D}) \right.$$
$$\left. \text{is a forward solution of } (1.1)_q \text{ for some } f \in C_0^\infty(\mathbb{R} \times \partial D) \right\}.$$

Here N is the interior unit normal of $\partial\Omega$.

Theorem 1.1. *Let $q_1, q_2 \in C^\infty(\mathbb{R} \times \bar{D})$ have the property that $\tilde{q}_1 - \tilde{q}_2 \in \mathscr{S}'(\mathbb{R}^{n+1})$, where $\tilde{q}_j = q_j$ in $\mathbb{R} \times \bar{D}$ and $\tilde{q}_j = 0$ in $\mathbb{R}^{n+1} \setminus (\mathbb{R} \times \bar{D})$. If q_1 and q_2 have the same sets of data, then $q_1 = q_2$.*

We shall give two variants of a proof based on WKB solutions. The first variant will be more direct, the second variant is based on a method given in [8] and may turn out to work also under considerably relaxed regularity assumptions on q_j. In [16] $q \in C(\mathbb{R}; L^\infty(D))$.

The first steps are the same for the two proofs: We first need to know that there are "many" lines which intersect ∂D only transversally:

Lemma 1.2. *Let $\omega \in S^{n-1}$ be fixed. For $x \in (\omega)^\perp \subset \mathbb{R}^n$, we let $l_{x,\omega}$ be the line: $\mathbb{R} \ni t \mapsto x + t\omega$. Then the set of $x \in (\omega)^\perp$ such that $l_{x,\omega}$ intersects ∂D tangentially at some point, is of measure 0.*

Proof. After an orthogonal change of coordinates, we may assume that $\omega = (0, \ldots, 0, 1)$. We write $x = (x', x_n) \in \mathbb{R}^n$, $x' \in \mathbb{R}^{n-1}$. Consider ∂D as an abstract smooth manifold, and let $G: \partial D \to \mathbb{R}^n$ be the trivial injection. Let $g(y')$, $y' \in \partial\Omega$, be the x'-component of $G(y')$. The set of $x \in (\omega)^\perp$, such that $l_{x,\omega}$ intersects ∂D non-transversally at some point, can then be identified with $\{g(y') \in \mathbb{R}^{n-1};\ dg(y')$ is non-invertible$\}$. By Sard's theorem, this set is of measure 0. □

Let $(s_0, y_0), (t_0, x_0) \in \mathbb{R} \times \partial D$ have the properties:

a) $t_0 - s_0 = |x_0 - y_0| \neq 0$,

b) the line through x_0 and y_0 intersects ∂D transversally at every point of intersection (including x_0, y_0),

c) the segment $]y_0, x_0[= \{s y_0 + (1-s) x_0;\ 0 < s < 1\}$ is contained in D.

Put $\omega_0 = \dfrac{x_0 - y_0}{|x_0 - y_0|}$. Let $\varphi(t, x)$ be smooth, real valued, defined in a small neighborhood of $[(s_0, y_0), (t_0, x_0)] = \{s y_0 + (1-s) x_0;\ 0 \leqq s \leqq 1\}$, satisfying the eiconal equation:

$$\partial_t \varphi = -|\partial_x \varphi|, \tag{1.2}$$

and having the property that

$$\partial_x \varphi = \omega_0 \quad \text{on } [(s_0, y_0), (t_0, x_0)]. \tag{1.3}$$

We may take $\varphi(t, x) = -t + x \cdot \omega_0$, or more generally, $= -t + \tilde{\varphi}(x)$ where $|\nabla \tilde{\varphi}| = 1$ and $\partial_x \tilde{\varphi} = \omega_0$ on $[y_0, x_0]$.

Let $f \in C_0^\infty(\mathbb{R} \times \partial D)$ have its support in a small neighborhood of (s_0, y_0). Then in a neighborhood of $[(s_0, y_0), (t_0, x_0)]$, we can construct an asymptotic solution $u = u(t, x, \lambda)$, $\lambda \geqq 1$, of $(\Box + q)u = \mathcal{O}(\lambda^{-\infty})$ with boundary condition: $u|_{(\mathbb{R} \times \partial D) \cap \text{neigborhood of } (s_0, y_0)} = f e^{i\lambda\varphi}$, of the form

$$u(x, \lambda) = a(t, x, \lambda)\, e^{i\lambda\varphi(t,x)}, \tag{1.4}$$

where a is the symbol of order 0, ($\partial_t^k \partial_x^\alpha a = \mathcal{O}(1)$ for all k, α) with an asymptotic expansion

$$a(t, x, \lambda) \sim \sum_0^\infty a_j(t, x)\, \lambda^{-j}, \qquad a_j \in C^\infty. \tag{1.5}$$

The a_j are determined by the asymptotic equation

$$La + \frac{i}{2\lambda}(\Box + q)a = \mathcal{O}(\lambda^{-\infty}), \tag{1.6}$$

where $L = \partial_x \varphi \cdot \partial_x - \partial_t \varphi \cdot \partial_t - \frac{1}{2} \Box \varphi$, which leads to the sequence of transport equations:

$$\begin{aligned} La_0 &= 0 \\ La_1 &= -\frac{i}{2}(\Box + q)\, a_0 \\ La_2 &= -\frac{i}{2}(\Box + q)\, a_1 \\ &\vdots \end{aligned}$$

The transport equations can be solved uniquely in a fixed (suitable) neighborhood of $[(s_0, y_0), (t_0, x_0)]$, after imposing the initial conditions: $a_0|_{(\mathbb{R} \times \partial D) \cap \text{neighborhood of supp } f} = f$, $a_j|_{\ldots} = 0$, $j \geqq 1$.

We then define a to be an asymptotic sum as in (1.5), and u, defined by (1.4), will then have the following properties:

(1.7) supp $u \cap (\mathbb{R} \times D)$ is contained in a small neightborhood of $](s_0, y_0), (t_0, x_0)[$ (that can be made arbitrarily small, by shrinking the support of f sufficiently).

(1.8) $u|_{\mathbb{R} \times \partial D} = f e^{i\lambda\varphi}$ except in a small neighborhood of (t_0, x_0).

(1.9) $\partial_t^k \partial_x^\alpha (\Box + q)u = \mathcal{O}(\lambda^{-N})$, uniformly for $(x, t) \in \mathbb{R} \times D$, $\lambda \geqq 1$, for all fixed k, α, N.

If we want to improve (1.8), then near (t_0, x_0), we have to add a "reflected term". This is done in the following way. In a neighborhood of (t_0, x_0), consider the system:

$$\partial_t \psi = -|\partial_x \psi| \tag{1.10}$$

$$\psi|_{\mathbb{R}\times\partial D} = \varphi|_{\mathbb{R}\times\partial D} \text{ (near } (t_0, x_0)). \tag{1.11}$$

There are exactly two (local) solutions of (1.10), (1.11) and we choose the one which is different from φ. Then, we have

$$\frac{\partial \psi}{\partial N} = -\frac{\partial \varphi}{\partial N} \quad \text{on } \mathbb{R}\times\partial D. \tag{1.12}$$

We then construct $b(t, x, \lambda)\, e^{i\lambda\psi(t,x)}$ as above, with $b \sim \sum_0^\infty b_j \lambda^{-j}$ and $b|_{\mathbb{R}\times\partial D} = -a|_{\mathbb{R}\times\partial D}$ near (t_0, x_0). The support of b is then contained in a small neighborhood of the line segment starting at (t_0, x_0) and obtained from $[(s_0, y_0), (t_0, x_0)]$ by optical reflection.

Taking

$$u = a e^{i\lambda\varphi} + b e^{i\lambda\psi} \tag{1.13}$$

we obtain, with a constant $C>0$, independent of f:

$$u|_{]-\infty, t_0+1/C[\times D} = f e^{i\lambda\varphi} \tag{1.14}$$

$$\partial_t^k \partial_x^\alpha (\Box + q)u = \mathcal{O}(\lambda^{-N}) \text{ uniformly in } \left]-\infty, t_0+\frac{1}{C}\right]\times \bar{D}\times[1, +\infty[\tag{1.15}$$

for all k, α, N.

By considering Taylor expansions at $\mathbb{R}\times\partial D$, we may also arrange so that

$$(\Box + q)u \text{ vanishes to infinite order at } \left]-\infty, t_0+\frac{1}{C}\right]\times\partial D. \tag{1.16}$$

Let $\chi(t)\in C^\infty(\mathbb{R})$ be equal to 1 for $t \leqq t_0 + \frac{1}{2C}$ and equal to 0 for $t \geqq t_0 + \frac{2}{3C}$. Let $w\in C^\infty(\mathbb{R}\times\bar{D})$ be the forward solution of the system

$$(\Box + q)w = -(\Box + q)(\chi(t)u) \tag{1.17}$$

$$w|_{\mathbb{R}\times\partial D} = 0. \tag{1.18}$$

Here "forward" again means: "vanishing for large negative t".

Replacing u by $\chi(t)u + w$, we get the exact forward solution of

$$(\Box + q)u = 0 \tag{1.19}$$

$$u|_{\mathbb{R}\times\partial D} = f e^{i\lambda\varphi}, \tag{1.20}$$

and our construction together with standard a priori estimates (see for instance [14]) shows that

$$u = a e^{i\lambda\varphi} + b e^{i\lambda\psi} + r, \quad \text{for } t \leqq t_0 + \frac{1}{2C}, \tag{1.21}$$

where $\partial_t^k \partial_x^\alpha r = \mathcal{O}(\lambda^{-N})$, $\forall k, \alpha, N$.

Notice that the eiconal equations for φ and ψ do not depend on q. Similarly a_0 and b_0 are independent of q. Let us consider the second transport equation for a_1 along an integral curve of $\partial_x\varphi \cdot \partial_x = -\partial_t\varphi \cdot \partial_t$ close to $[(s_0, y_0), (t_0, x_0)]$. Such an integral curve is necessarily a straight line, and even a "light ray" of the form $t \mapsto (Ct, x + Ct\omega)$, $|\omega| = 1$, $C > 0$ (close to 1). Here t is the natural parametrization. The restriction to this curve of the second transport equation, then takes the form:

$$\left(\frac{d}{dt} + g(t)\right) a_1 = -\frac{i}{2}(h(t) + q(t))\, a_0, \tag{1.22}$$

where g, h are independent of q_1. Using the equation $\left(\frac{d}{dt} + g\right) a_0 = 0$, we obtain

$$a_1(t) = \left(-\frac{i}{2}\int_{t_1}^{t} h(s)\, ds - \frac{i}{2}\int_{t_1}^{t} q(s)\, ds\right) a_0(t) \tag{1.23}$$

where t_1 is the value of the parameter t, when the curve crosses $\mathbb{R} \times \partial D$ near (s_0, y_0).

We next look at $\left.\frac{\partial u}{\partial N}\right|_{\mathbb{R}\times\partial D}$ near (t_0, x_0). Using (1.21), (1.11), (1.12) and the fact that $a = -b$ on $\mathbb{R} \times \partial D$ near (t_0, x_0), we get:

$$\begin{aligned} &e^{-i\lambda\varphi} \left.\frac{\partial u}{\partial N}\right|_{\mathbb{R}\times\partial D \cap \text{neighborhood of } (t_0, x_0)} \\ &= \left(2i\frac{\partial\varphi}{\partial N} a_0\right)\lambda + \left(2i\frac{\partial\varphi}{\partial N} a_1 + \frac{\partial a_0}{\partial N} + \frac{\partial b_0}{\partial N}\right) + \mathcal{O}(\lambda^{-1}). \end{aligned} \tag{1.24}$$

The first term of the right-hand side is independent of q, and the only q-dependence in the second term is via a_1, and this dependence is given by (1.23) if we replace t by t_2: the time when the integral curve hits $\mathbb{R}/\partial D$ near (t_0, x_0).

Now replace q by q_j, $j = 1, 2$, when q_j satisfy the assumptions in the theorem. Let u_1, u_2 be the corresponding solution "u" with the same f. Assuming $f(s_0, y_0) \neq 0$, we get from (1.23), (1.24) together with the fact that $\frac{\partial u_1}{\partial N} = \frac{\partial u_2}{\partial N}$ on $\mathbb{R} \times \partial D$, that:

$$\int_0^{t_0 - s_0} (q_1 - q_2)(s_0 + s, y_0 + s\omega_0)\, ds = 0. \tag{1.25}$$

Combining this with Lemma 1.1 we see that for every $\omega \in S^{n-1}$ we have for almost every $x \in \mathbb{R}^n$:

$$\mathcal{R}(\tilde{q}_1 - \tilde{q}_2)(1, \omega, x) = 0. \tag{1.26}$$

(See the section on ray transforms).

By Fubini's theorem $\mathcal{R}(\tilde{q}_1, \tilde{q}_2)(1, \omega, x)$ vanishes for almost all $(\omega, x) \in S^{n-1} \times \mathbb{R}^n$, and hence vanishes in the sense of distributions. Since $\tilde{q}_1 - \tilde{q}_2$ has its support in a cylinder: $|x| \leqq R$, and is of class $\mathcal{S}'$, we can apply Proposition 2.2 in the next section and conclude that $\tilde{q}_1 - \tilde{q}_2 = 0$ in $\mathcal{S}'$. Hence $q_1 - q_2 = 0$, and this completes the first proof of the theorem. □

Second proof. We first apply the argument from [8]. Let u_1, u_2 be forward solutions of $(1.1)_{q_1}$, $(1.1)_{q_2}$, but with the same f. Taking the difference of $(1.1)_{q_1}$, $(1.1)_{q_2}$, we get with $u = u_1 - u_2$, $q = q_1 - q_2$:

$$(\Box + q_1) u = -q u_2 \tag{1.27}$$

$$u|_{\mathbb{R} \times \partial D} = \left.\frac{\partial u}{\partial N}\right|_{\mathbb{R} \times \partial D} = 0. \tag{1.28}$$

Let $w_1 \in C^\infty(\mathbb{R} \times \bar{D})$ be a backward solution of $(\Box + q_1) w_1 = 0$ with $w_1|_{\mathbb{R} \times \partial D} = g \in C_0^\infty(\mathbb{R} \times D)$, for some g.

Multiplying (1.27) by w_1, integrating by parts, and using (1.28), we get

$$\begin{aligned} -\iint_{\mathbb{R} \times D} q u_2 w_1 \, dt \, dx &= \iint_{\mathbb{R} \times D} [(\Box + q_1) u w_1 - u(\Box + q_1) w_1] \, dt \, dx \\ &= \iint_{\mathbb{R} \times \partial D} \left(\frac{\partial u}{\partial N} w_1 - u \frac{\partial w_1}{\partial N} \right) dt \, dS(x) = 0. \end{aligned} \tag{1.29}$$

We now take y_0, x_0, ω_0, $\varphi = -t + x \cdot \omega_0$ as in the first proof, and we replace f, g by $f e^{i\lambda\varphi}$, $g e^{i\lambda\varphi}$, where the new f, $g \in C_0^\infty(\mathbb{R} \times \partial D)$ are supported near (s_0, y_0), (t_0, x_0) respectively. Then u_2 has all the properties of "u" in the first proof (with q replaced by q_2), and similarly

$$w_1 = \tilde{a} e^{-i\lambda\varphi} + \tilde{b} e^{-i\lambda\tilde{\psi}} + \tilde{r}, \quad \text{in } t \geqq s_0 - \frac{1}{2C}, \quad \partial_t^k \partial_x^\alpha \tilde{r} = \mathcal{O}(\lambda^{-\infty}). \tag{1.30}$$

$u_2 w_1$ has its support in $s_0 - \dfrac{1}{2C} \leqq t \leqq t_0 + \dfrac{1}{2C}$, and we shall look at:

$$\begin{aligned} 0 &= \lim_{\lambda \to \infty} \iint_{\mathbb{R} \times D} q u_2 w_1 \, dt \, dx \\ &= \iint_{\mathbb{R} \times D} q a_0 \tilde{a}_0 \, dt \, dx + \lim_{\lambda \to \infty} \iint_{\mathbb{R} \times D} q b_0 \tilde{a}_0 e^{i\lambda(\tilde{\psi} - \varphi)} dt \, dx \\ &\quad + \lim_{\lambda \to \infty} \iint_{\mathbb{R} \times D} q a_0 \tilde{b}_0 e^{i\lambda(\varphi - \tilde{\psi})} dt \, dx. \end{aligned} \tag{1.31}$$

Here we used that b and $\tilde{b}$ have disjoint supports (assuming that the supports of f and g are sufficiently small). Now $\nabla(\tilde{\psi} - \varphi)$ and $\nabla(\varphi - \tilde{\psi})$ are non-vanishing,

so by an integration by parts, we see that the second and the third integral to the right in (1.31) are $\mathcal{O}\left(\frac{1}{\lambda}\right)$. Hence (1.31) becomes

$$0=\iint_{\mathbf{R}\times D} q a_0 \tilde{a}_0 \, dt\, dx. \tag{1.32}$$

Now we exploit the fact that φ is linear. Then $\Box\varphi=0$ and the first transport equation for a_0 says that $a_0(t+s, x+s\omega_0)=a_0(t, x)=f(t, x)$, if (t, x) is the boundary point close to (s_0, y_0), and $0\leqq s\leqq t_0-s_0+\text{const.}$ $\tilde{a}_0$ is also constant on every ray pointing in the direction $(1, \omega_0)$. Letting $f\geqq 0$, $g\geqq 0$ be suitable functions with supports shrinking to (s_0, y_0) and (t_0, x_0) respectively, and taking the corresponding limit in (1.32), we obtain (1.25) and the proof can be finished as before. □

One can show that even with (certain) more general choices of the phases in the second proof, one still gets (1.25) and nothing else.

Some refinements of Theorem 1.1 are possible. Let $d>0$ be the diameter of D. Examining any one of the two proofs, we see that in Theorem 1.1, we can replace the set of data by the restricted set of data:

$$\mathcal{D}(d+\varepsilon, q)=\left\{\left(u, \frac{\partial u}{\partial N}\right)\Big|_{[t_-(f), t_-(f)+d+\varepsilon]\times\partial D};\right. \tag{1.33}$$
$$\left. u \text{ is a forward solution of } (1.1)_q \text{ with } f\in C_0^\infty(\mathbf{R}\times\partial D)\right\}.$$

Here $t_-(f)=\inf \Pi_t(\operatorname{supp} f)$. More precisely we have the following result which will be further improved below:

Preliminary theorem. *Let q_1, $q_2\in C^\infty(\mathbf{R}\times\bar{D})$ have the property that $\tilde{q}_1$, $\tilde{q}_2\in\mathcal{S}'(\mathbf{R}^{n+1})$, with $\tilde{q}_j$ defined as in Theorem* 1.1. *If for some $\varepsilon>0$, we have $\mathcal{D}(d+\varepsilon, q_1)=\mathcal{D}(d+\varepsilon, q_2)$ (with $d=\operatorname{diam}(D)$), then $q_1=q_2$.*

Using the fact that only a small open set of directions ω is needed in Proposition 2.2, this result can be further improved: For $\omega\in S^{n-1}$, $x\in(\omega)^\perp$, let $s(\omega, x)$ be the supremum of the lengths of all segments contained in $\bar{D}\cap\{x+r\omega; r\in\mathbf{R}\}$. Put

$$\bar{s}=\inf_{\omega\in S^{n-1}}\sup_{x\in(\omega)^\perp} s(\omega, x),$$

and notice that for every $\varepsilon>0$, there is a non-empty open set $V_\varepsilon\subset S^{n-1}$ such that

$$\sup_{x\in(\omega)^\perp} s(\omega, x)<\bar{s}+\varepsilon \quad \text{for } \omega\in V_\varepsilon.$$

(This is due to the fact that $s(\omega, x)$ is upper semicontinuous).

Again by examining one of the proofs of Theorem 1.1, we get:

Theorem 1.3. *Let q_1, $q_2\in C^\infty(\mathbf{R}\times\bar{D})$ have the property that $\tilde{q}_1-\tilde{q}_2\in\mathcal{S}'(\mathbf{R}^{n+1})$, with $\tilde{q}_j$ defined as in Theorem* 1.1. *If for some $\varepsilon>0$, we have $\mathcal{D}(\bar{s}+\varepsilon, q_1)=\mathcal{D}(\bar{s}+\varepsilon, q_2)$, then $q_1=q_2$.*

In the case when $q_1(t, x)=q_2(t, x)$ for t outside a compact interval, we conclude that the boundary data over a fixed bounded time interval will determine the potential:

Corollary 1.4. *Assume that* $\bar{s}<a<b<T-\bar{s}$, $T<\infty$. *Let* q_1, $q_2\in C^\infty(\mathbb{R}\times\bar{D})$ *with* $(q_1-q_2)(t, x)=0$ *for* $t\in\mathbb{R}\setminus[a, b]$.

Suppose that for every $f\in C_0^\infty(]0, T[\times\partial D)$, *we have*

$$(u_1-u_2)|_{]0,T[\times\partial D}=\left(\frac{\partial u_1}{\partial N}-\frac{\partial u_2}{\partial N}\right)\Big|_{]0,T[\times\partial D}=0, \tag{1.33}$$

where u_j *is the forward solution of* $(1.1)_{q_j}$. *Then* $q_1=q_2$.

Proof. With $\varepsilon>0$ sufficiently small we verify the assumptions of Theorem 1.3. Let $f\in C_0^\infty(\mathbb{R}\times\partial D)$ and let u_j be the forward solution of $(1.1)_{q_j}$, $j=1, 2$. If $t_-(f)\leqq 0$ or if $t_-(f)\geqq b$, then using the fact that $q_1=q_2$ in $(\mathbb{R}\setminus[a, b])\times D$, we see that $u_1=u_2$ for $t_-(f)\leqq t\leqq t_-(f)+\bar{s}+\varepsilon$, and hence the boundary data coincide on the same time interval. If $0<t_-(f)\leqq b$, then u_j will not change in the region $t_-(f)\leqq t\leqq t_-(f)+\bar{s}+\varepsilon$, if we replace $f(t, x)$ by $\chi(t)f(t, x)$ if $\chi\in C_0^\infty(]0, T[)$ is equal to 1 on $[\delta, T-\delta]$ for $\delta>0$ small enough. The assumption (1.33) then implies that

$$(u_1-u_2)|_{[t_-(f),t_-(f)+\bar{s}+\varepsilon]\times\partial D}=\left(\frac{\partial u_1}{\partial N}-\frac{\partial u_2}{\partial N}\right)\Big|_{[t_-(f),t_-(f)+\bar{s}+\varepsilon]\times\partial D}=0. \quad \square$$

2 Ray transforms

In the beginning of this section we collect some simple (well known) facts. After this section had been written we learned that Stefanov essentially proved Proposition 2.2 (see [12], the argument leading from (6.7) to "$q=0$"), and that result should be attributed to him. We keep the proof (which differs in details only) for the sake of completeness.

For $x\in\mathbb{R}^n$, $\omega\in\mathbb{R}^n\setminus\{0\}=\dot{\mathbb{R}}^n$, $u\in\mathcal{S}(\mathbb{R}^n)$, we put:

$$\mathcal{R}u(\omega, x)=\int_{-\infty}^{+\infty} u(x+t\omega)\,dt. \tag{2.1}$$

Then:

$$\mathcal{R}u(\omega, x)=\mathcal{R}u(\omega, y), \quad \text{if } x-y \text{ is a multiple of } \omega, \tag{2.2}$$

$$\mathcal{R}u(-\omega, x)=\mathcal{R}u(\omega, x), \tag{2.3}$$

$$\mathcal{R}u(s\omega, x)=\frac{1}{s}\mathcal{R}u(\omega, x), \quad \text{if } s>0 \tag{2.4}$$

$$(\partial_{x_j}\partial_{\omega_k}-\partial_{x_k}\partial_{\omega_j})\,\mathcal{R}u(\omega, x)=0, \quad \forall j, k. \tag{2.5}$$

We note that (2.5) involves only $n-1$ independent equations since on the symbol level:

$$\xi_j \omega_k^* - \xi_k \omega_j^* = \frac{\xi_j}{\xi_l}(\xi_l \omega_k^* - \xi_k \omega_l^*) + \frac{\xi_k}{\xi_l}(\xi_j \omega_l^* - \xi_l \omega_j^*), \tag{2.6}$$

so, for instance, in the region $\xi_1 \neq 0$, we only need the symbols $\xi_1 \omega_j^* - \xi_j \omega_1^*$ for $2 \leqq j \leqq n$.

The distribution kernel of $\mathscr{R}$ can be written:

$$\int_{-\infty}^{+\infty} \delta(x+t\omega-y)\,dt = \int_{-\infty}^{+\infty} \int e^{i(x+t\omega-y)\cdot\theta} \frac{d\theta}{(2\pi)^n}\,dt. \tag{2.7}$$

For x, ω, y in a compact subset of $\mathbb{R}^n \times \dot{\mathbb{R}}^n \times \mathbb{R}^n$, the t-integration can be restricted to a compact, and calculating the critical set and the corresponding canonical relation for the phase, with (t, θ) as fiber variables, we obtain

$$WF'(\mathscr{R}) \subset \{(\omega, x, t\theta, \theta; y, \theta); x+t\omega = y, \theta \neq 0, \omega \cdot \theta = 0\}. \tag{2.8}$$

Here we refer to [3] for the definitions and basic properties of wavefront sets (WF and WF').

In particular the inverse image of the zero section over $\dot{\mathbb{R}}^n_\omega \times \mathbb{R}^n_x$ is empty, so $\mathscr{R}u$ is well defined as a distribution, when u is a distribution with compact support, (or satisfying some other natural support condition).

Let us examine the relation between $\mathscr{R}$ and the ordinary Fourier transform: $u \mapsto \mathscr{F}u(\xi) = \hat{u}(\xi) = \int e^{-ix\cdot\xi} u(x)\,dx$. Without loss of information (in view of (2.4)), we may restrict ω to S^{n-1}. Then $\mathscr{R}u(\omega, x) = (u \mid \delta_{\omega,x})$ (standard sesquilinear L^2 inner product), where $\delta_{\omega,x}$ is the δ-measure, carried by the line: $t \mapsto x + t\omega$. The Fourier-transform of this measure is given by

$$\delta_{\omega,x}(\xi) = e^{-ix\cdot\xi}\delta(\omega \cdot \xi), \tag{2.9}$$

with δ = standard Dirac measure at $0 \in \mathbb{R}$. If $\varphi \in \mathscr{S}(\mathbb{R}^n)$:

$$\langle \hat{\delta}_{\omega,x}, \varphi \rangle = \int_{\xi\cdot\omega=0} \varphi(\xi)\, e^{-ix\cdot\xi} S(d\xi), \tag{2.10}$$

$S = (n-1)-$dimensional standard volume on $(\omega)^\perp$.

By Plancherel's formula, we get for $u \in \mathscr{S}(\mathbb{R}^n)$:

$$\mathscr{R}u(\omega, x) = \frac{1}{(2\pi)^n} \int_{\xi\cdot\omega=0} \hat{u}(\xi)\, e^{ix\cdot\xi} S(d\xi). \tag{2.11}$$

If we assume that $\mathscr{R}u(\omega, x) = 0$ for a fixed ω and every x, then after multiplying the preceeding identity by $f(x) \in \mathscr{S}(\mathbb{R}^n)$ and integrating w.r.t. x, we get:

$$0 = \frac{1}{(2\pi)^n} \int_{\xi\cdot\omega=0} \hat{u}(\xi)\hat{f}(-\xi)\, S(d\xi). \tag{2.12}$$

Varying f, we see that $\hat{u}(\xi)=0$ for $\xi\cdot\omega=0$:

Lemma 2.1. *If $u\in\mathscr{S}(\mathbb{R}^n)$, $\omega\in S^{n-1}$, and $\mathscr{R}u(\omega, x)=0$ for every $x\in\mathbb{R}^n$, then $\hat{u}$ vanishes on the hyperplane $(\omega)^\perp=\{\xi\in\mathbb{R}^{n*};\ \omega\cdot\xi=0\}$.*

(It is easy to see this directly from the definitions on $\mathscr{R}$ and $\mathscr{F}$).

We now turn to the situation given by the wave equation. We replace n by $n+1$ and $\mathbb{R}^n_x$ by $\mathbb{R}^{n+1}_{t,x}=\mathbb{R}_t\times\mathbb{R}^n_x$. We will assume that $n\geqq 2$ (so that $n+1\geqq 3$), and we shall restrict our ray transform to light-rays of the form:

$$t\mapsto(0, x)+t(1, \omega),\qquad \omega\in S^{n-1}.$$

Let us first examine the orthogonal planes, that may appear in Lemma 2.1: If $(\tau, \xi)\in((1, \omega))^\perp$ then

$$\tau+\xi\cdot\omega=0, \tag{2.13}$$

which implies that $|\tau|\leqq|\xi\cdot\omega|\leqq|\xi|$. In other words, the orthogonal planes $((1, \omega))^\perp$, with $\omega\in S^{n-1}$, are disjoint from the open dual light cone: $|\tau|>|\xi|$.

If $|\tau|\leqq|\xi|$, $|\xi|>0$, then (τ, ξ) is contained in at least one plane of the form $((1, \omega))^\perp$, $\omega\in S^{n-1}$. In fact, we just have to choose ω such that $\frac{\xi}{|\xi|}\cdot\omega=-\frac{\tau}{|\xi|}$, which is possible, since the right hand side is of absolute value $\leqq 1$. If $|\tau|=|\xi|$ there is only one possible plane, namely the tangent plane to the dual light cone, passing through (τ, ω). If $|\tau|<|\xi|$, the set of possible planes can be parametrized by S^{n-2}.

When dealing with light-ray transforms, we write

$$\mathscr{R}u(1, \omega, 0, x)=\mathscr{R}u(1, \omega, x)=\int_{-\infty}^{+\infty} u(s, x+s\omega)\,ds. \tag{2.14}$$

The distribution kernel is $K(\omega, x; s, y)=\delta(x+s\omega-y)$, and we get

$$WF'(\mathscr{R})\subset\{(\omega, x, s\Pi_\omega(\theta), \theta; s, y, -\omega\cdot\theta, \theta);\ \theta\neq 0, y=x+s\omega\}, \tag{2.15}$$

where Π_ω denotes the natural projection $T^*_\omega\mathbb{R}^n\to T^*_\omega S^{n-1}$. Again we conclude that if $u\in\mathscr{D}'(\mathbb{R}^{n+1})$ has its support in a cylinder: $|x|\leqq R$, then $\mathscr{R}u(1, \omega, x)$ is well defined as a distribution on $S^{n-1}\times\mathbb{R}^n$, and we can even define $\mathscr{R}u(1, \omega, \cdot)$ as a distribution for every fixed ω.

We shall next establish a uniqueness result, and as a warm up we start with a simple special case. Let $u\in C^\infty_0(\mathbb{R}^{1+n})$ and assume that $\mathscr{R}u(1, \omega, x)=0$ for $(\omega, x)\in V\times\mathbb{R}^n$, where $V\subset S^{n-1}$ is open and non-empty. Formula (2.11) now takes the form

$$\mathscr{R}u(1, \omega, x)=C\int\hat{u}(-\xi\cdot\omega, \xi)\,e^{ix\cdot\xi}d\xi \tag{2.16}$$

for a suitable constant $C>0$ (depending on n only). Hence $\hat{u}(-\xi\cdot\omega, \xi)$ is equal to a constant times the Fourier transform of $x\mapsto\mathscr{R}u(1, \omega, x)$. The last constant is equal to $(2\pi)^{-n}$ in view of the more direct calculation:

$$\hat{u}(-\xi\cdot\omega, \xi)=\iint e^{-i(-t\xi\cdot\omega+x\cdot\xi)}u(t, x)\,dt\,dx$$

[change of variables: $(t, x)=(t, y+t\omega)$: $dt\,dx=dt\,dy$].

$$\hat{u}(-\xi\cdot\omega,\xi)=\int e^{-iy\cdot\xi}\Big(\int u(t,y+t\omega)\,dt\Big)\,dy \tag{2.17}$$
$$=\int e^{-iy\cdot\xi}\mathscr{R}u(1,\omega,y)\,dy.$$

From this, we conclude that $\hat{u}$ vanishes on some open set in $\mathbb{R}^{n+1}$. Since $\hat{u}$ extends to an entire function, it follows that $\hat{u}=0$ and hence that $u=0$.

Now take $u\in\mathscr{S}'(\mathbb{R}^{n+1})$ with support in a cylinder: $|x|\leqq R$. Then the map: $\mathbb{C}^n\ni\theta\mapsto e^{ix\cdot\theta}u\in\mathscr{S}'(\mathbb{R}^{n+1})$ is entire, so we know that the map $\mathbb{R}^n\ni\theta\mapsto\hat{u}(\tau,\ \xi-\theta)\in\mathscr{S}'(\mathbb{R}^{n+1}_{\tau,\xi})$ extends to an entire function with values in $\mathscr{S}'(\mathbb{R}^{n+1}_{\tau,\xi})$. This means that $\iint\hat{u}(\tau,\ \xi-\theta)\,\varphi(\tau,\ \xi)\,d\tau\,d\xi$ is entire in θ, for every $\varphi\in\mathscr{S}(\mathbb{R}^{n+1})$. Suppose in addition that $\hat{u}=0$ in some open ball $B((\tau_0,\ \xi_0),\ r)$ (of center $(\tau_0,\ \xi_0)$ and radius $r>0$). If $\varphi\in CF_0^\infty(B(\tau_0,\ \xi_0),\ r)$ we then know that $\iint\hat{u}(\tau,\ \xi-\theta)\,\varphi(\tau,\ \xi)\,d\tau\,d\xi$ vanishes when θ is real and small enough. By unique continuation, this integral vanishes for all θ, and we conclude that $\hat{u}=0$ in every ball $B((\tau_0,\ \xi_0+\theta),\ r)$ and hence in the whole slice: $|\tau-\tau_0|<r$.

Let $\mathscr{S}'_R=\{u\in\mathscr{S}'(\mathbb{R}^{n+1});\ \operatorname{supp}u\subset\{(t,\ x):\ |x|\leqq R\}\}$, equipped with the weak topology of $\mathscr{S}'$. Then it is easy to see that $u\mapsto\mathscr{R}u(1,\ \omega,\ x)$ is well defined and continuous: $\mathscr{S}'_R\to C^\infty(S^{n-1};\ \mathscr{S}'(\mathbb{R}^n))$, so (2.17) and a density argument show that $u\mapsto\hat{u}(-\xi\cdot\omega,\ \xi)$ is well defined and continuous from $\mathscr{S}'_R$ to $C^\infty(S^{n-1};\ \mathscr{S}'(\mathbb{R}^n))$. Moreover we still have (2.17) in the form

$$\hat{u}(-\xi\cdot\omega,\xi)=\mathscr{F}(\mathscr{R}u(1,\omega,\cdot))(\xi). \tag{2.17}$$

Now let $u\in\mathscr{S}'_R$ and assume that $\mathscr{R}u(1,\ \omega,\ \cdot)=0$ for ω in some open non-empty set $V\subset S^{n-1}$. We conclude that $\hat{u}|_{((1,\omega))^\perp}$ vanishes for every $\omega\in V$, where $\hat{u}|_{((1,\omega))^\perp}$ is the distribution $\hat{u}(-\xi\cdot\omega,\ \xi)$ just defined.

Let $\omega_0\in V$ and consider a point $(\tau_0,\ \xi_0)\in((1,\ \omega_0))^\perp$ with $|\tau_0|<|\xi_0|$, so that $\tau_0=-\xi_0\cdot\omega_0$ and ξ_0 is not parallel to ω_0. For $t\in]-\varepsilon,\ \varepsilon[$, $|\eta-\xi_0|<\varepsilon$, with $\varepsilon>0$ small, we map $(t,\ \eta)$ to $(\tau,\ \xi)=(-\eta\cdot\omega_t,\ \eta)$ with $\omega_t=(\omega_0+t\xi_0)/\|\omega_0+t\xi_0\|$. Since $\partial_t(-\eta\cdot\omega_t)\neq0$, we get a diffeomorphism onto a small neighborhood of $(\tau_0,\ \xi_0)$ if $\varepsilon>0$ is small enough. Then for φ with support in that neighborhood:

$$\iint\hat{u}(\tau,\xi)\,\varphi(\tau,\xi)\,d\tau\,d\xi=\int\Big(\int\hat{u}(-\eta\cdot\omega_t,\eta)\,\varphi(-\eta\omega_t,\eta)(d(\tau,\xi)/d(t,\eta))\,d\eta\Big)\,dt=0.$$

Here the integrals except the one in t are defined in the distribution sense, and the first equality is justified as above by approximating u by C_0^∞-functions. Hence $\hat{u}$ vanishes in a neighborhood of $(\tau_0,\ \xi_0)$. For every $\tau_0\in\mathbb{R}$, we can find ξ_0 and ω_0 with the above properties. As we have just seen, this implies that $\hat{u}$ vanishes in a slize $|\tau-\tau_0|<r$ for some $r>0$, and since we can choose τ_0 arbitrary, we deduce that $\hat{u}$ and hence u vanish identically.

Summing up, we have proved:

Proposition 2.2. *Let $u\in\mathscr{S}'(\mathbb{R}_t\times\mathbb{R}^n_x)$ have its support in a cylinder: $|x|\leqq R$. Suppose that $V\subset S^{n-1}$ is open and non-empty, and that $\mathscr{R}u(1,\ \omega,\ \cdot)=0$ (in $\mathscr{S}'(\mathbb{R}^n)$) for every $\omega\in V$. Then $u=0$.*

We end this section with a remark about products of exponential solutions of the wave equation. We have seen above, that the light-ray transform gives the Fourier transform outside the dual light-cone, and if we want to prove a uniqueness theorem using the light ray transform, then we have to use also unique continuation for entire functions. One might ask whether scalar products

of u and products of pure real exponential solutions of the wave equations with frequencies which are approximately opposite, would not give a direct approach. The following observations indicate that this is not the case: Let $e^{i(\tau t+\xi\cdot x)}$ and $e^{i(\sigma t+\eta\cdot x)}$ be two exponential solutions of $\Box u=0$ so that $\tau^2-\xi^2=\sigma^2-\eta^2=0$. Since there will also be lower order terms in the operators, we want $|(\tau, \xi)|$, and $|(\sigma, \eta)|$ to be large, but we also want the frequency of the product: $(\tau+\sigma, \xi+\eta)$ to stay in some given neighborhood of (0, 0). This means that τ, σ will have opposite signs: $\tau\sigma<0$. We then get:

$$(\xi+\eta)^2-(\tau+\sigma)^2=2(\xi\cdot\eta-\tau\cdot\sigma)=2(|\tau|\cdot|\sigma|+\xi\cdot\eta)\geqq 2(|\tau|\cdot|\sigma|-|\xi|\cdot|\eta|)\geqq 0,$$

so the vectors $(\tau+\sigma, \xi+\eta)$ with $\tau^2-\xi^2=\sigma^2-\eta^2=0$, $\tau\sigma\leqq 0$, are outside the open dual light-cone. Conversely every (ρ, ζ) outside the open dual light cone (to $|\zeta|\geqq|\rho|$) can be written as such a sum, with $|(\tau, \xi)|$, $|(\sigma, \eta)|$ arbitrarily large: Assume for instance that $\rho>0$. Then take $\tau\gg 1$ arbitrarily large, $\sigma=\rho-\tau\ll -1$. Then the set of $\xi+\eta$, with $|\xi|=\tau$, $|\eta|=|\sigma|=\tau-\rho$, in the shell: $\{\tilde{\zeta};\ \rho\leqq|\tilde{\zeta}|\leqq 2\tau-\rho\}$, which contains ζ, if τ is large enough.

Acknowledgement. AGR thanks NSF and ONR for support.

Bibliography

1. Buhgeim, A.: Multidimensional inverse problems of spectral analysis. Sov. Math., Dokl. **32**, 353–355 (1985)
2. Buhgeim, A., Klibanov, M.: Global uniqueness of a class of multidimensional inverse problems. Sov. Math., Dokl. **24**, 244–247 (1981)
3. Hörmander, L.: The analysis of linear differential operator I. Berlin Heidelberg New York: Springer 1983
4. Klibanov, M.: On a class of inverse problems. Sov. Math., Dokl. **26**, 248–251 (1982)
5. Nachman, A., Sylvester, J., Uhlmann, G.: An n-dimensional Borg-Levinson Theorem. Commun. Math. Phys. **115**, 595–605 (1988)
6. Rakesh, Symes, W.: Uniqueness for an inverse problem for the wave equation. Commun. Partial Differ. Equations **13**, 87–96 (1988)
7. Ramm, A.G.: A uniqueness theorem for a boundary inverse problem. Inverse Probl. **4**, L 1–5 (1988)
8. Ramm, A.G.: Multidimensional inverse problems and completeness of the products of solutions to PDE. J. Math. Anal. Appl. **134**, 211–253 (1988); **136**, 568–574 (1988); **139**, 302 (1989); **136**, 352–356 (1989)
9. Ramm, A.G.: Multidimensional inverse scattering problems and completeness of the products of solutions to homogeneous PDE. Z. Angew. Math. Mech. **69**, 113–122 (1989)
10. Ramm, A.G.: A simple proof of uniqueness theorem in impedance tomography. Appl. Math. Lett. **1**, 287–290 (1988), **3**, 57–60 (1990)
11. Ramm, A.G.: Completeness of the set of products of solutions to PDE and uniqueness theorems in inverse scattering. Inverse Prob. **3**, 77–82 (1987), **4**, 877–886 (1988), **6**, 643–664 (1990)
12. Stefanov, P.: Uniqueness of the multidimensional inverse scattering problem for time dependent potentials. Math. Z. **201**, 541–560 (1989)
13. Sylvester, J., Uhlmann, G.: A global uniqueness theorem for an inverse boundary value problem. Ann. Math. **125**, 153–169 (1987)
14. Sakamoto, R.: Mixed problems for hyperbolic equations I, II. J. Math. Kyoto Univ. **10**, 349–373, 403–417 (1970)
15. Sylvester, J., Uhlmann, G.: The Dirichlet to Neumann map and applications. Proceedings of meeting in Arcata, California, on inverse problems, July 29–Aug. 4, 1989
16. Ramm, A.G., Rakesh: Property C and an inverse problem for a hyperbolic equation. J. Math. Anal. Appl. (1991) (to appear)

Math. Z. 206, 131–144 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Riemann Roch for equivariant $\mathfrak{D}$-modules-I

Roy Joshua *

Department of Mathematics, The Johns Hopkins University, Baltimore, MD 21218, USA

Received March 3, 1989; in final form January 24, 1990

In this paper we prove a Riemann-Roch theorem for weakly equivariant algebraic $\mathfrak{D}$-modules for the action of complex diagonalisable groups on smooth complex quasi-projective algebraic varieties. Let X denote such a variety provided with the action of a complex diagonalisable group D. Let $\mathcal{O}_X(\mathfrak{D}_X)$ denote the structure sheaf of the variety X (the sheaf of algebraic $\mathbb{C}$-linear differential operators on X, respectively). We begin by considering actions by complex linear algebraic groups in general. After defining the category of equivariant (quasi-coherent) $\mathcal{O}_X$-modules, we recall several basic facts about $\mathfrak{D}_X$-modules and define the category of (weakly-) equivariant $\mathfrak{D}_X$-modules. It is shown that every equivariant coherent algebraic $\mathfrak{D}_X$-module admits a good invariant global filtration. Next we obtain an equivariant 'first exact sequence of Spencer'.

Through the rest of the paper we restrict to actions by a complex diagonalisable group D. The equivariant first exact sequence of Spencer enables us to prove the Riemann-Roch theorem for D equivariant $\mathfrak{D}_X$-modules at the level of the Grothendieck groups as is done in ([Mal] and also [Ang-Lej]) for the non-equivariant case. A similar argument proves a Riemann-Roch theorem for D-equivariant $\mathfrak{D}_X$-modules whose supports are contained in a fixed D-invariant and possibly-singular closed subvariety.

The following papers ([J-2] through [J-4]) discuss various extensions and applications of the results of this paper. In [J-2] we apply the Riemann-Roch theorem of this paper to provide a trace formula for equivariant $\mathfrak{D}$-modules and perverse sheaves. In [J-3] the Riemann-Roch theorem is extended to actions by arbitrary algebraic group actions and we provide very general character formulae for equivariant $\mathfrak{D}$-modules and perverse sheaves. In [J-4] we study the higher algebraic K-theory as well as the derived category of the category of weakly equivariant $\mathfrak{D}$-modules.

(1.0) Let X denote a complex smooth quasi-projective algebraic variety provided with the action of a complex linear algebraic group G. (We will restrict to *algebraic* varieties throught this paper.) We first consider G-equivariant $\mathcal{O}_X$-modules as in ([T-1] 1.2 and also [Mum] p. 30). Let $\mu(\mathrm{pr})_2\colon G\times X\to X$ denote the group action (projection to the second factor, respectively).

* *Current address:* Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA

(1.1) **Definition.** A *G-equivariant $\mathcal{O}_X$-module* is an $\mathcal{O}_X$-module M for which there exists an isomorphism φ: $\mu^* M \to \mathrm{pr}_2^* M$ (of $\mathcal{O}_G \otimes_{\mathbb{C}} \mathcal{O}_X$-modules) satisfying the following conditions: (i). If $d_0(d_1, d_2)$: $G \times G \times X \to G \times X$ denotes the map $(g_1, g_2, x) \to (g_2, x)$ (the map $(g_1, g_2, x) \to (g_1 \cdot g_2, x)$, the map $(g_1, g_2, x) \to (g_1, g_2 \cdot x)$, respectively) then there is an obvious cocycle condition between the three pull-backs $d_0^*(\varphi)$, $d_1^*(\varphi)$ and $d_2^*(\varphi)$. (See [Mum], p. 30.) (ii). If s_0: $X \to G \times X$ denotes the map $x \to (e, x)$ where e is the identity element of G, then $s_0^*(\varphi) =$ the identity of X.

(1.1.1) *Remark.* Observe that $\mathrm{pr}_2^*(M) \cong \mathcal{O}_G \otimes_{\mathbb{C}} M$ and that the isomorphism φ in (1.1) corresponds (by adjunction) to the co-action $M \to \mu_*(\mathcal{O}_G \otimes_{\mathbb{C}} M)$. The conditions ensure that this map is compatible (in the obvious sense) with the co-action $\mathcal{O}_X \to \mu_*(\mathcal{O}_G \otimes_{\mathbb{C}} \mathcal{O}_X)$. One may also observe that the above condition is equivalent to the structure map $\mathcal{O}_X \otimes_{\mathbb{C}} M \to M$ being a map of $\mathcal{O}_G$-comodules where $\mathcal{O}_G$ co-acts co-diagonally on $\mathcal{O}_X \otimes_{\mathbb{C}} M$.

(1.1.2) A map f: $M \to M'$ of G-equivariant $\mathcal{O}_X$-modules is said to be *G-equivariant* if the obvious square made up of the isomorphisms in (1.1) and the maps induced by f commutes. The fact that pr_2^* and μ^* are exact functors ($\mathcal{O}_X$-modules) $\to$ ($\mathcal{O}_G \otimes_{\mathbb{C}} \mathcal{O}_X$-modules) shows that the category of G-equivariant $\mathcal{O}_X$-modules is abelian. (See [T-1] p. 543.) This category will be denoted $\mathrm{Mod}^G(\mathcal{O}_X)$ from now on and $\mathrm{Mod}^G_{q\mathrm{coh}}(\mathcal{O}_X)$ ($\mathrm{Mod}^G_{\mathrm{coh}}(\mathcal{O}_X)$) will denote the full subcategory of G-equivariant quasi-coherent (coherent, respectively) $\mathcal{O}_X$-modules.

(1.2) Next we recall some basic facts about $\mathfrak{D}$-modules from ([EGA] IV, 16). Let X be a smooth quasi-projective complex variety and let Δ: $X \to X \times X$ denote the diagonal immersion. Let J denote the sheaf of ideals defining X in $X \times X$. Then recall that the n-th infinitesimal neighborhood of X in $X \times X$ (denoted $X^{(n)}$) is the closed subscheme of $X \times X$ defined by $J^{(n+1)}$. Let $h^{(n)}$: $X^{(n)} \to X \times X$ be the closed immersion of $X^{(n)}$ in $X \times X$; let

$$\mathrm{pr}_1^{(n)}\colon\ X^{(n)} \xrightarrow{h^{(n)}} X \times X \xrightarrow{\mathrm{pr}_1} X, \qquad \mathrm{pr}_2^{(n)}\colon\ X^{(n)} \xrightarrow{h^{(n)}} X \times X \xrightarrow{\mathrm{pr}_2} X$$

where pr_1 (pr_2) denotes projection to the first (second, respectively) factor. We let $P_X^n = (\mathrm{pr}_1^{(n)})_* \circ (\mathrm{pr}_2^{(n)})^*(\mathcal{O}_X)$. The maps $\mathrm{pr}_1^{(n)}$ and $\mathrm{pr}_2^{(n)}$ induce maps of sheaves of rings $\mathcal{O}_X \to P_X^n$ which are both flat and provide P_X^n with two distinct $\mathcal{O}_X$-module structures; P_X^n is a coherent and locally free $\mathcal{O}_X$-module with respect to either of these structures. (The map $\mathcal{O}_X \to P_X^n$ induced by $\mathrm{pr}_2^{(n)}$ is denoted d_X^n and the corresponding $\mathcal{O}_X$-module structure is called the right module structure; the $\mathcal{O}_X$-module structure induced by $\mathrm{pr}_1^{(n)}$ is called the left module structure.) (See [EGA] IV, 16.7, 16.10.)

(1.3) *Differential operators.* (See [EGA] IV, 16.8.) Let X be as before and let F, G be two $\mathcal{O}_X$-modules and n a non-negative integer. A homomorphism of sheaves of additive groups D: $F \to G$ is said to be a differential operator of order $\leqq n$ if there exists a homomorphism u: $P_X^n(F) \to G$ of $\mathcal{O}_X$-modules such that $D = u \circ d^n_{X,F}$. (Here $P_X^n(F) = P_X^n \otimes_{\mathcal{O}_X} F$ and has the $\mathcal{O}_X$-module structure induced from the left $\mathcal{O}_X$-module structure on P_X^n; $d^n_{X,F}$ is the obvious map $F \to P_X^n(F)$ induced by d_X^n as in (1.2).) The sheaf $\mathrm{Diff}^n_X(F, G)$ of differential operators of order $\leqq n$ (from F to G) is now defined to be $\mathrm{Diff}^n_X(F, G) = \mathrm{Hom}_{\mathcal{O}_X}(P_X^n(F),$

G). We let Diff_X^n denote $\mathrm{Diff}_X^n(\mathcal{O}_X, \mathcal{O}_X)$ and may identify it with the dual of P_X^n. The left (right) $\mathcal{O}_X$-module structure on P_X^n induces the $\mathcal{O}_X$-module structure on Diff_X^n where $\mathcal{O}_X$ acts by left (right, respectively) multiplication. It is clear Diff_X^n is a coherent locally free $\mathcal{O}_X$-module with respect to either of these structures. (See [EGA] 16.8.5.) From now on we will let $\mathfrak{D}_X$ denote the ring of differential operators filtered by $\{\mathfrak{D}_X^n = \mathrm{Diff}_X^n\}$.

(1.4) Next we assume the situation of (1.0). It is now clear that each P_X^n and hence $\mathfrak{D}_X^n$ is a G-equivariant $\mathcal{O}_X$-module. We next show that the composition $\psi\colon \mathfrak{D}_X^n \otimes \mathfrak{D}_X^{n'} \to \mathfrak{D}_X^{n+n'}$ is also G-equivariant as in (1.1.2). (Since $\mathfrak{D}_X$ is *not commutative*, the usual arguments (as in [D-G] Chap. 1, Sect. 3) using 'representing schemes' do not readily carry over to show ψ is equivariant; hence we adopt the following alternate arguments.) Let $D\colon \mathcal{O}_X \to \mathcal{O}_X$ ($D'\colon \mathcal{O}_X \to \mathcal{O}_X$) be a differential operator of order $\leqq n$ ($\leqq n'$, respectively). Now we may write $D = u \circ d_X^n$ and $D' = v \circ d_X^{n'}$, where $u\colon P_X^n \to \mathcal{O}_X$, $v\colon P_X^{n'} \to \mathcal{O}_X$ are maps of $\mathcal{O}_X$-modules. The composite map $\mathcal{O}_X \xrightarrow{d_X^n} P_X^n \xrightarrow{u} \mathcal{O}_X \xrightarrow{d_X^{n'}} P_X^{n'}$ factors as $O_X \xrightarrow{d_X^{n+n'}} P_X^{n+n'} \xrightarrow{\omega} P_X^{n'}$ where ω is a map of $\mathcal{O}_X$-modules defined as follows. There exists a unique map

$$\delta\colon P_X^{n+n'} \to P_X^{n'}(P_X^n) \cong P_X^{n'} \otimes_{\mathcal{O}_X} P_X^n \tag{1.4.1}$$

of $\mathcal{O}_X$-modules so that the diagram

$$\begin{array}{ccc} \mathcal{O}_X & \xrightarrow{d_X^{n+n'}} & P_X^{n+n'} \\ \downarrow{\scriptstyle d_X^n} & & \downarrow{\scriptstyle \delta} \\ \mathscr{P}_X^n & \xrightarrow{d_X^{n'} \otimes 1} & P_X^{n'}(P_X^n) \end{array}$$

commutes. Now ω can be factored as $P_X^{n+n'} \xrightarrow{\delta} P_X^{n'}(P_X^n) \xrightarrow{1 \otimes u} P_X^{n'}$ and hence $D' \circ D\colon \mathcal{O}_X \to \mathcal{O}_X$ is a differential operator of order $\leqq n+n'$. This provides the ring structure of $\mathfrak{D}_X$. One may now reduce the equivariance of ψ to the equivariance of the map δ. An examination of the definition of the map δ from ([EGA] IV, 16.8) shows it is indeed equivariant. (Alternatively, use the following facts to conclude δ is equivariant: all the other three maps in the above square are clearly equivariant, the image of the map $d_X^n(d_X^{n+n'})$ generates $P_X^n(P_X^{n+n'}$, respectively) as an $\mathcal{O}_X$-module and δ is a map of $\mathcal{O}_X$-modules. (See [EGA] IV, 16.3.8.).) In view of (1.1.2) the G-equivariance of ψ provides the commutative square

$$\begin{array}{ccc} \mu^* \mathrm{Diff}_X^n \otimes \mu^* \mathrm{Diff}_X^m & \longrightarrow & \mu^* \mathrm{Diff}_X^{n+m} \\ \downarrow & & \downarrow \\ \mathrm{pr}_2^* \mathrm{Diff}_X^n \otimes \mathrm{pr}_2^* \mathrm{Diff}_X^m & \longrightarrow & \mathrm{pr}_2^* \mathrm{Diff}_X^{n+m} \end{array}$$

This shows that the map

$$\mu^* \mathfrak{D}_X \xrightarrow{\simeq} \mathrm{pr}_2^* \mathfrak{D}_X \tag{1.4.2}$$

is an isomorphism of sheaves of rings.

(1.5) **Definition.** A $\mathfrak{D}_X$-module M will mean an $\mathcal{O}_X$-module which is a left module (unless we say to the contrary) over the ring $\mathfrak{D}_X$. We will say a $\mathfrak{D}_X$-module M is *quasicoherent* (*coherent*) if M is quasi-coherent as an $\mathcal{O}_X$-module (if each $x \in X$ has an affine open neighborhood U_x and a short exact sequence

$$\bigoplus_{i=1}^{n} \mathfrak{D}_{U_x} \to \bigoplus_{i=1}^{m} \mathfrak{D}_{U_x} \to M_{|U_x} \to 0,$$

respectively).

Next we consider G-equivariant $\mathfrak{D}_X$-modules. We will again assume the situation of (1.0). Recall that $\mu\colon G \times X \to X$ ($\mathrm{pr}_2\colon G \times X \to X$) denotes the group action (the projection to the second factor, respectively). Observe that if M is a $\mathfrak{D}_X$-module $\mu^* M (\mathrm{pr}_2^* M)$ is a $\mu^* \mathfrak{D}_X$-module ($\mathrm{pr}_2^* \mathfrak{D}_X$-module, respectively).

(1.6) **Definition.** A $\mathfrak{D}_X$-module M will be called a *G-equivariant* (or a *weakly-G-equivariant*) $\mathfrak{D}_X$-module provided it is a G-equivariant $\mathcal{O}_X$-module and there exists an isomorphism $\varphi\colon \mu^* M \to \mathrm{pr}_2^* M$ compatible with the isomorphisms $\mu^* \mathfrak{D}_X \to \mathrm{pr}_2^* \mathfrak{D}_X$ and $\mu^* \mathcal{O}_X \to \mathrm{pr}_2^* \mathcal{O}_X$. (Observe that, as M is a G-equivariant $\mathcal{O}_X$-module, the isomorphism φ staisfies the conditions in (1.1).)

(1.6.1) The category of G-equivariant $\mathfrak{D}_X$-modules (which are also quasi-coherent, coherent) and G-equivariant maps will be denoted $\mathrm{Mod}^G(\mathfrak{D}_X)$ ($\mathrm{Mod}^G_{q\,\mathrm{coh}}(\mathfrak{D}_X)$, $\mathrm{Mod}^G_{\mathrm{coh}}(\mathfrak{D}_X)$, respectively). The exactness of pr_2^* and μ^* shows these are again abelian categories.

(1.7) (i) We observe as in (1.1.1) that the isomorphism φ corresponds to a coaction $M \to \mu_*(\mathcal{O}_G \otimes_{\mathbb{C}} M)$ which is compatible with the coaction $\mathfrak{D}_X \to \mu_*(\mathcal{O}_G \otimes_{\mathcal{O}_X} \mathfrak{D}_X)$. This is equivalent to the structure map $\mathfrak{D}_X \otimes_{\mathbb{C}} M \to M$ being a map of $\mathcal{O}_G$-comodules where $\mathcal{O}_G$ acts co-diagonally on $\mathfrak{D}_X \otimes_{\mathbb{C}} M$. As $\mu^*(\mathfrak{D}_X) \to \mathrm{pr}_2^*(\mathfrak{D}_X)$ is an isomorphism of sheaves of rings one may view the isomorphism in (1.6) as an isomorphism of $\mathrm{pr}_2^*(\mathfrak{D}_X)$-modules. Recall from ([Be-Ber], Sect. 3 or [B-B], p. 16) that there exists a stronger notion of equivariant $\mathfrak{D}_X$-modules where the isomorphism in (1.6) is required to be an isomorphism of $\mathfrak{D}_{G \times X}$-modules. The weaker notion in (1.6) is more appropriate for our purposes.

(1.7) (ii) Recall from ([Bo], p. 210) that a $\mathfrak{D}_X$-module M may be regarded as a quasicoherent $\mathcal{O}_X$-module with a flat connexion. (Observe that a connexion on M is a map $\nabla\colon M \to \Omega^1_X \otimes_{\mathcal{O}_X} M$. This corresponds to an action of the tangent sheaf $T^1_X \otimes_{\mathcal{O}_X} M \to M$ which extends to an action $\mathfrak{D}_X \otimes_{\mathcal{O}_X} M \to M$ since ∇ is flat.) Next observe that under the assumptions of (1.0) Ω^1_X, T^1_X and $\mathfrak{D}_X$ are G-equivariant $\mathcal{O}_X$-modules. Let M be a G-equivariant quasi-coherent $\mathcal{O}_X$-module. We will say a connexion ∇ on M is G-equivariant if ∇ is a G-equivariant map for the diagonal coaction of $\mathcal{O}_G$ on $\Omega^1_X \otimes_{\mathcal{O}_X} M$. Observe that ∇ being equivariant

is equivalent to the map $\mathfrak{D}_X \otimes_{\mathcal{O}_X} M \to M$ being equivariant for the diagonal coaction of $\mathcal{O}_G$ on $\mathfrak{D}_X \otimes M$. *Thus a weakly-G-equivariant $\mathfrak{D}_X$-module is a G-equivariant $\mathcal{O}_X$-module provided with a flat G-equivariant connexion ∇.*

(1.7) (iii) Now assume in addition to (1.0) that X is a smooth affine variety with co-ordinate ring $k[X]$. We recall (see [Bo], p. 238) the equivalence of categories

$$(\text{quasi-coherent } \mathfrak{D}_X\text{-modules}) \simeq (\text{modules over } \Gamma(X, \mathfrak{D}_X)).$$

Under the above assumptions this induces an equivalence of categories

$$(\text{quasi-coherent } G\text{-equivariant } \mathfrak{D}_X\text{-modules}) \simeq (G\text{-equivariant modules over } \Gamma(X, \mathfrak{D}_X)).$$

(The category on the right is defined as follows. Let $\mu(\mathrm{pr}_2)\colon k[X] \to k[G] \otimes k[X]$ denote the co-action (the projection to the second factor, respectively). We will let $\mu^*(\mathrm{pr}_2^*)$ denote the functor $(k[X]\text{-modules}) \to (k[G] \otimes k[X]\text{-modules})$ induced by μ (pr_2, respectively). Now one observes that the ring multiplication $\Gamma(X, \mathfrak{D}_X) \otimes \Gamma(X, \mathfrak{D}_X) \to \Gamma(X, \mathfrak{D}_X)$ is equivariant and hence there is an isomorphism of rings $\mu^*\Gamma(X, \mathfrak{D}_X) \to \mathrm{pr}_2^*\Gamma(X, \mathfrak{D}_X)$. A module M over $\Gamma(X, \mathfrak{D}_X)$ will be called G-equivariant if there exists an isomorphism $\varphi\colon \mu^* M \to \mathrm{pr}_2^* M$ compatible with $\mu^*\Gamma(X, \mathfrak{D}_X) \xrightarrow{\simeq} \mathrm{pr}_2^*\Gamma(X, \mathfrak{D}_X)$. The morphisms in this category are defined as in (1.1.2).)

(1.8.0) Assume in addition to the situation of (1.0) that Y is a smooth complex variety provided with a G-action and $f\colon X \to Y$ is a G-equivariant map. In this case we first of all observe that $f^{-1}(\mathcal{O}_Y)$ is also an equivariant sheaf of rings. (i.e. Let $\mu\colon G \times Y \to Y$ ($\bar{\mu}\colon G \times X \to X$) denote the group action. Then there exist co-actions $\mathcal{O}_Y \to \mu_*(\mathcal{O}_G \otimes_{\mathbb{C}} \mathcal{O}_Y)(f^{-1}\mathcal{O}_Y \to \bar{\mu}_*(\mathcal{O}_G \otimes_{\mathbb{C}} f^{-1}\mathcal{O}_Y)$ and the multiplication $\mathcal{O}_Y \otimes_{\mathbb{C}} \mathcal{O}_Y \to \mathcal{O}_Y$ ($f^{-1}(\mathcal{O}_Y) \otimes_{\mathbb{C}} f^{-1}(\mathcal{O}_Y) \to f^{-1}(\mathcal{O}_Y)$ is compatible with the above co-actions.) Similarly $f^{-1}(\mathfrak{D}_Y)$ is an equivariant sheaf of rings of on Y. The category of equivariant modules over $f^{-1}(\mathcal{O}_Y)$ and $f^{-1}(\mathfrak{D}_Y)$ may now be defined as in (1.1). These will be denoted $\mathrm{Mod}^G(f^{-1}(\mathcal{O}_Y))$ and $\mathrm{Mod}^G(f^{-1}(\mathfrak{D}_Y))$ and may readily be shown to be abelian categories.

(1.8.1) It is often convenient to consider the following more general situation. Let A denote a sheaf of not-necessarily-commutative algebras (with a two sided unit) over $\mathcal{O}_X$ or $f^{-1}(\mathcal{O}_Y)$. Now the category of G-equivariant left-modules on (X, A) may be defined in the obvious manner as above. (The main examples are clearly $A = \mathcal{O}_X, f^{-1}(\mathcal{O}_Y), \mathfrak{D}_X$ or $f^{-1}(\mathfrak{D}_Y)$.) Given sheaves of algebras A and B on X as above, a map $\alpha\colon A \to B$ is *equivariant* if it is compatible in the obvious sense with the co-action by $\mathcal{O}_G$.

(1.8.2) Assume the above situation. Let $EG \times_G X$ denote the simplicial scheme given by $(EG \times_G X)_n = G^n \times X$ with the degeneracies and face maps given in the usual manner. A sheaf of rings R on $EG \times_G X$ is a collection $R = \{R_n\}$, where R_n is a sheaf of rings on $(EG \times_G X)_n$ so that for each structure map $\psi\colon (EG \times_G X)_m \to (EG \times_G X)_n$, there exist maps $\Psi\colon \psi^* R_n \to R_m$ which are isomorphisms and so that the maps Ψ satisfy a compatibility condition as in ([Fr], p. 14). $(EG \times_G X, R)$ will be called a *simplicial ringed space.* A module M on the above simplicial ringed space consists of a collection $\{M_n\}$ satisfying the conditions: (i) M_n is a sheaf of left modules over the ringed space $((EG \times_G X)_n, R_n)$. (ii)

For each structure map $\psi: (EG \times_G X)_m \to (EG \times_G X)_n$, the map $\Psi: \psi^* M_n \to M_m$ is an isomorphism of R_m-modules and the maps Ψ satisfy a compatibility condition as mentioned above. Now one may readily verify the following assertions:

(i) Let A be a sheaf of algebras as in (1.8.1). Then we obtain a simplicial ringed space $(EG \times_G X, A_{EG \times_G X})$ by letting $A_{(EG \times_G X)_n} = \bigotimes^n \mathcal{O}_G \otimes A$.

(ii) Let $(EG \times_G X, A_{EG \times_G X})_{\leqq 2}$ denote the truncated simplicial ringed space given in dimension $i \leqq 2$ by $(EG \times_G X, A_{EG \times_G X})_i$. Then there exists an equivalence of categories.

(1.8.2.*) $\quad$ (G-equivariant A-modules on X) $\xrightarrow{\simeq}$ (modules on the truncated simplicial ringed space $(EG \times_G X, A_{EG \times_G X})_{\leqq 2}$).

(iii) Assume the situation of (1.8.0) and let $\mathcal{O}_X^{\sim}$ denote either the structure sheaf $\mathcal{O}_X$ of X or $f^{-1}(\mathcal{O}_Y)$. (f^* will denote the usual functor: (modules on $(Y, \mathcal{O}_Y)$) $\to$ (modules on $(X, \mathcal{O}_X)$) in the first case and will denote f^{-1}: (modules on $(Y, \mathcal{O}_Y)$) $\to$ (modules on $(X, f^{-1}(\mathcal{O}_Y))$ in the second case.) Let A be an equivariant sheaf of algebras over $\mathcal{O}_X^{\sim}$ (B an equivariant sheaf of algebras over $\mathcal{O}_Y$) as in (1.8.1). Assume in addition that f induces a map $f: (X, A) \to (Y, B)$ of ringed spaces which is also equivariant (i.e. the map $f^*(B) \to A$ is equivariant; the main example is with $B = \mathfrak{D}_Y$ and $A = f^{-1}(\mathfrak{D}_Y)$). Now one may readily verify that $f^*(f_*)$ induces a functor

f^*: (G-equivariant modules on (Y, B)) $\to$ (G-equivariant modules on (X, A))

f_*: (G-equivariant modules on (X, A)) $\to$ (G-equivariant modules on (Y, B)).

2 *In this section we prove the Riemann-Roch theorem*

(2.0) We will assume that G is a complex linear algebraic group acting on a smooth quasi-projective complex variety X in (2.1) through (2.4). First we prove the following result.

(2.1) **Proposition.** *Assume the situation of* (2.0). *Then for every coherent G-equivariant $\mathfrak{D}_X$-module M, there exists a G-equivariant surjection $\varepsilon: \mathfrak{D}_X \otimes E \otimes \omega_X^{\vee} \twoheadrightarrow M$ where E is a locally free G-equivariant $\mathcal{O}_X$-module of finite rank, ω_X is the sheaf of top-dimensional differential forms and the group G acts diagonally on the terms on the left.*

Proof. Let M be a G-equivariant coherent $\mathfrak{D}_X$-module. Now M is a quasicoherent G-equivariant $\mathcal{O}_X$-module and hence is the filtered colimit of its G-equivariant $\mathcal{O}_X$-coherent submodules $\{M_\alpha\}$. (See [T-1], p. 544.) By ([T-1], Lemma 5.6) each M_α is the G-equivariant quotient of a G-vector bundle Q_α. Let $\varepsilon_\alpha: Q_\alpha \twoheadrightarrow M_\alpha$ denote this G-equivariant surjection. Next observe that since M is an algebraic $\mathfrak{D}_X$-module, it has a good global filtration $\{N^k\}$ (which may not however be G-invariant). Recall that this implies there exists an integer k_0 so that if $k \geqq k_0$, then $\mathfrak{D}_X^i \cdot N^k = N^{k+i}$, for all i. Moreover each N^k is a coherent $\mathcal{O}_X$-module. Therefore N^k is contained in some $M_{\alpha(k_0)}$, for some index $\alpha(k_0)$. It follows that $\mathfrak{D}_X \cdot (M_{\alpha(k_0)}) \to M$ is a G-equivariant surjection of $\mathfrak{D}_X$-modules. Thus $\mathfrak{D}_X \otimes_{\mathcal{O}_X} Q_{\alpha(k_0)} \twoheadrightarrow \mathfrak{D}_X \otimes_{\mathcal{O}_X} M_{\alpha(k_0)} \twoheadrightarrow M$ is a G-equivariant surjection and clearly

$Q_{\alpha(k_0)}$ is a finitely generated locally free O_X-module. Let $E = Q_{\alpha(k_0)} \otimes_{\mathcal{O}_X} \omega_X$. Since $\omega_X \otimes_{\mathcal{O}_X} \omega_X^{\vee} \cong \mathcal{O}_X$, this E satisfies the conclusion of the proposition.

(2.2) In view of (2.1) observe that any equivariant coherent $\mathfrak{D}_X$-module has an equivariant resolution by a complex of locally free and coherent G-equivariant $\mathfrak{D}_X$-modules as in (2.1).

(2.3) **Corollary 1.** *Assume the situation of* (2.0). *Let M be a G-equivariant coherent $\mathfrak{D}_X$-module. Then M has a* good *G-invariant filtration. i.e. M has a nondecreasing filtration by $\mathcal{O}_X$-coherent G-equivariant submodules M^k so that i) $M^k = 0$ if $k \ll 0$ ii) $\mathfrak{D}_X^i \cdot M^k \subseteq M^{k+i}$ for all k, i and iii) there exists a k_0 so that for all $k \geqq k_0$, $\mathfrak{D}_X^i \cdot M^k = M^{k+i}$ for all i and (iv). $\bigcup_i M^i = M$.*

Proof. Let ε: $\mathfrak{D}_X \otimes E \otimes \omega_X^{\vee} \twoheadrightarrow M$ be the G-equivariant surjection as in (2.1). Observe that $\{D_X^k \otimes_{\mathcal{O}_X} E \otimes_{\mathcal{O}_X} \omega_X^{\vee} \mid k \geqq 0\}$ forms a filtration of $\mathfrak{D}_X \otimes_{\mathcal{O}_X} E \otimes_{\mathcal{O}_X} \omega_X^{\vee}$ satisfying the above conditions. Let $M^k = \varepsilon(\mathfrak{D}_X^k \otimes_{\mathcal{O}_X} E \otimes_{\mathcal{O}_X} \omega_X^{\vee})$. One may now verify that $\{M^k\}$ forms a filtration of M as in the corollary.

Next we will show that the above corollary provides a G-equivariant 'first-exact sequence of Spencer'. First observe that in the notation of (1.3), the tangent sheaf $T_X^1 = (\Omega_X^1)^{\vee} \subseteq \mathrm{Diff}_X^1 = \mathfrak{D}_X^1$. Let the dimension of X be n and let M be a G-equivariant coherent $\mathfrak{D}_X$-module provided with a good G-invariant filtration $\{M^k\}$. Then the first-exact sequence of Spencer of order $k+n$ (see [Ang-Lej], p. 133) (denoted $\mathrm{Sp}^{k+n}(M)$) is:

$$
\begin{aligned}
(2.4) \qquad & 0 \to \mathfrak{D}_X \otimes_{\mathcal{O}_X} \overset{n}{\bigwedge} T_X \otimes_{\mathcal{O}_X} M^k \to \mathfrak{D}_X \otimes_{\mathcal{O}_X} \overset{(n-1)}{\bigwedge} T_X \otimes_{\mathcal{O}_X} M^{k+1} \\
& \to \mathfrak{D}_X \otimes_{\mathcal{O}_X} T_X \otimes_{\mathcal{O}_X} M^{k+n-1} \to \mathfrak{D}_X \otimes_{\mathcal{O}_X} M^{k+n}
\end{aligned}
$$

where the differential d: $\mathfrak{D}_X \otimes \overset{p}{\bigwedge} T_X \otimes M^j \to \mathfrak{D}_X \otimes \overset{p-1}{\bigwedge} T_X \otimes M^{j+1}$ is given by

$$
\begin{aligned}
d(P \otimes (t_1 \wedge \ldots \wedge t_p) \otimes m) = \sum_{i=1}^{p} (-1)^{i-1} [& P t_i \otimes (t_1 \wedge \ldots \wedge \hat{t}_i \wedge \ldots \wedge t_p) \\
& \otimes m - P \otimes (t_1 \wedge \ldots \wedge \hat{t}_i \wedge \ldots \wedge t_p) \otimes t_i m] \\
& + \sum_{i<j} (-1)^{i+j} P \otimes ([t_i, t_j] \\
& \wedge t_1 \wedge \ldots \wedge \hat{t}_i \wedge \ldots \wedge \hat{t}_j \wedge \ldots \wedge t_p) \otimes m.
\end{aligned}
$$

Since each of the M^k is a G-equivariant $\mathcal{O}_X$-module, we observe that all the terms in (2.4) are G-equivariant $\mathfrak{D}_X$-modules (where $\mathfrak{D}_X$ acts on the left, on the factor $\mathfrak{D}_X$). Now the maps appearing in the first summation are induced by the left multiplication by $\mathfrak{D}_X$ on $\mathfrak{D}_X$ and on M, along with the natural isomorphisms of $\overset{(n)}{\bigwedge} T_X$ permuting the various factors. Hence they are G-equivariant. (See (1.7) (i).) Similarly one sees that the maps appearing in the second summation are equivariant. Hence (2.4) is a complex in $\mathrm{Mod}_{\mathrm{coh}}^G(\mathfrak{D}_X)$. One

observes as in ([Mal], pp. 54–55) that if $k \gg 0$, the complex (2.4) is acyclic and the obvious (G-equivariant) augmentation ε: $\mathfrak{D}_X \otimes_{\mathcal{O}_X} M \to M$ makes it a resolution of M. Moreover the obvious equivariant map $\mathrm{Sp}^{k+n}(M) \to \mathrm{Sp}^{k+n+1}(M)$ is a quasi-isomorphism again for $k \gg 0$.

We proceed to prove the Riemann-Roch theorem for equivariant $\mathfrak{D}$-modules at the level of K_0. For the rest of this section we will only consider actions by diagonalisable groups.

(2.5) Let D be a complex diagonalisable group acting on a smooth complex quasi-projective variety Z. Recall $\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z)$ ($\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_Z)$) denotes the category of D-equivariant and coherent (D-equivariant and quasi-coherent, respectively) $\mathfrak{D}_Z$-modules; $\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z)$ ($\mathrm{Mod}^D_{q\mathrm{coh}}(\mathcal{O}_Z)$) denotes the corresponding categories of D-equivariant $\mathcal{O}_Z$-modules. Let $D^+(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_Z))$ ($D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z))$, $D^+(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathcal{O}_Z))$, $D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z))$) be the derived category of complexes in $\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_Z)$ ($\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z)$, $\mathrm{Mod}^D_{q\mathrm{coh}}(\mathcal{O}_Z)$, $\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z)$, respectively) that are bounded below. $D^+_{b,\mathrm{coh}}(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_Z))$ ($D^+_{b,\mathrm{coh}}(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathcal{O}_Z))$ will denote the derived category of complexes in $\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_Z)$ ($\mathrm{Mod}^D_{q\mathrm{coh}}(\mathcal{O}_Z)$, respectively) that are bounded below and have bounded coherent cohomology. Finally we will let $D^+_b(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z))$ ($D^+_b(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z))$ denote the derived category of bounded below complexes in $\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z)$ ($\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z)$, respectively) with bounded cohomology sheaves.

(2.6) We next define a functor d_Z: $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z))$ as in [Ang-Lej], p. 132). (Here $\mathcal{O}_Z$ denotes the structure sheaf of Z and K_0 the Grothendieck group of the appropriate category.) Let π: $T^*Z \to Z$ (i: $Z \to T^*Z$) denote the obvious projection from the cotangent bundle of Z (the zero section imbedding of Z in T^*Z, respectively). Observe that $\mathrm{grd}(\mathfrak{D}_Z) = \pi_*(O_{T^*Z})$. Let $\tilde{}$: $\mathrm{Mod}_{q\mathrm{coh}}(\mathrm{grd}(\mathfrak{D}_Z)) \to \mathrm{Mod}_{q\mathrm{coh}}(\mathcal{O}_Z)$ denote the equivalence of categories as in ([EGA] 9.2.5). One observes readily that $\tilde{}$ preserves coherence, is an exact functor and sends D-equivariant $\mathrm{grd}(\mathfrak{D}_Z)$-modules to D-equivariant $\mathcal{O}_{T^*Z}$-modules. Clearly $\tilde{}$ induces a functor

$$\tilde{}: K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathrm{grd}(\mathfrak{D}_Z)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_{T^*Z})).$$

Let M be a D-equivariant and coherent $\mathfrak{D}_Z$-module provided with a good invariant filtration $\{M^k\}$. Now $\mathrm{grd}(M)$ is a D-equivariant and coherent $\mathrm{grd}(\mathfrak{D}_Z)$-module; one may show as in ([Lau], Lemma 6.1.2 and [Bj], Chap. 2, proof of 6.2) that the class of $\mathrm{grd}(M)$ in $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathrm{grd}(\mathfrak{D}_Z))$ is independent of the good D-invariant filtration chosen. (Recall (2.3) shows the existence of such filtrations.) We let $\mathrm{grd}(M)^{\tilde{}}$ denote the image of $\mathrm{grd}(M)$ in $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_{T^*Z}))$ under the map $\tilde{}$. Let d_Z be defined by $d_Z(M) =$ the class of $\left[\mathcal{O}_Z \overset{L}{\underset{i^{-1}(\mathcal{O}_{T^*Z})}{\otimes}} i^{-1}(\mathrm{grd}(M)^{\tilde{}})\right] \otimes_{\mathcal{O}_Z} \omega_Z$, using a good D-invariant filtration of M. (One may use (2.2) to conclude that every coherent D-equivariant $\mathcal{O}_{T^*Z}$-module of the form $\mathrm{grd}(M)^{\tilde{}}$ has a resolution by D-equivariant locally free $\mathcal{O}_{T^*Z}$-modules; this defines $\overset{L}{\otimes}$ as above.) Clearly $d_Z(M)$ defines a class in $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z))$ independent of the choice of a good D-invariant filtration. If M is a $\mathfrak{D}_Z$-module of the form $\mathfrak{D}_Z \otimes E \otimes \omega_Z^{\vee}$, (where E is a D-equivariant coherent $\mathcal{O}_Z$-module)

we observe that $E \simeq d_Z(\mathfrak{D}_Z \otimes E \otimes \omega_Z^{\vee})$. (See [T-1], p. 557 to conclude that the D-equivariant $\mathcal{O}_Z$-coherent sheaf E has an equivariant resolution by D-equivariant vector bundles. This provides a D-equivariant resolution of the $\mathfrak{D}_Z$-module $\mathfrak{D}_Z \otimes E \otimes \omega_Z^{\vee}$.)

(2.7) Assuming the situation of (2.5) we may define an exact functor E: $\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Z) \to \mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z)$ by $F \to \mathfrak{D}_Z \otimes_{\mathcal{O}_Z} F \otimes_{\mathcal{O}_Z} \omega_Z^{\vee}$. Evidently this induces a map of the Grothendieck groups which will be denoted E_*. Clearly it is a section to the map d_Z in (2.6). E_* is hence injective; the discussion following (2.4) shows it is also surjective. Therefore both E_* and d_Z are *isomorphisms.*

(2.8) Let X, Y be two complex quasi-projective smooth varieties provided with the action of a complex diagonalisable group D and let $f\colon X \to Y$ be a D-equivariant map. We will assume that the D-actions are linear. (Observe from [Su], p. 8 that we may always make this assumption, without any loss of generality, if D is connected.) We proceed to define two functors:

$$(2.8^*) \qquad Rf_*\colon D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)) \to D^+(\mathrm{Mod}^D_{q\,\mathrm{coh}}(\mathcal{O}_Y))$$

$$\int_f\colon D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) \to D^+(\mathrm{Mod}^D_{q\,\mathrm{coh}}(\mathfrak{D}_Y))$$

First of all we obtain the D-equivariant $(f^{-1}(\mathfrak{D}_Y), \mathfrak{D}_X)$-bimodule $\mathfrak{D}_{Y \leftarrow X} = f^{-1}(\mathfrak{D}_Y \otimes_{\mathcal{O}_Y} \omega_Y) \otimes_{f^{-1}(\mathcal{O}_Y)} \omega_X^{\vee}$ where the group acts diagonally on the terms of the right. One may verify (using [Bj], p. 230) that ω_Y (ω_X) is an equivariant right $\mathfrak{D}_Y$- ($\mathfrak{D}_X$-) module; the equivariant left $f^{-1}(\mathfrak{D}_Y)$ module structure on $\mathfrak{D}_{Y \leftarrow X}$ is clear.)

Next recall from ([Su], pp. 8–9) that if a diagonalisable group acts on a smooth quasi-projective variety, under the assumptions of (2.8) each point has an affine and D-invariant Zariski open neighborhood. Therefore we first consider the case when both X and Y are affine. If $F^{\cdot} \in D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X))$, $(M^{\cdot} \in D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)))$ we will let $Rf_* F^{\cdot} = f_* F^{\cdot}\left(\int_f M^{\cdot} = f_*\left(\mathfrak{D}_{Y \leftarrow X} \overset{L}{\underset{\mathfrak{D}_X}{\otimes}} M^{\cdot}\right)\right.$, respectively$\left.\right)$. (1.8.2) (iii) shows that $Rf_* F^{\cdot}(\int_f M^{\cdot})$ is an equivariant complex of equivariant $\mathcal{O}_Y$- ($\mathfrak{D}_Y$-, respectively) modules. Under the equivalence of categories in (1.7) (iii), we observe that the complex $\int_f M^{\cdot}$ corresponds to $\Gamma(X, \mathfrak{D}_{Y \leftarrow X}) \underset{\Gamma(X, \mathfrak{D}_X)}{\otimes} \Gamma(X, M^{\cdot})$. Hence the above functor preserves quasi-isomorphisms and defines a functor between the derived categories in (2.8*). (It is clear that Rf_* also defines a functor between the corresponding derived categories as in (2.8*).)

Next we consider the general case. Let $U = \{U_i | i\}$ be an affine Zariski open invariant cover of X. Let $C(U)$ denote the associated simplicial scheme defined by the Cech-construction. (Recall $C(U)_0 = \coprod_i U_i$, $C(U)_1 = \coprod_{i_1 < i_2} (U_{i_1} \cap U_{i_2})$, ..., with the obvious face maps and degeneracies.) Let $\varepsilon_n\colon C(U)_n \to X$ denote the structure map. If F is a D-equivariant coherent $\mathcal{O}_X$-module (M is a coherent D-equivariant and locally free $\mathfrak{D}_X$-module) we let $C^{\cdot}(F)$ ($C^{\cdot}(\mathfrak{D}_{Y \leftarrow X} \otimes_{\mathfrak{D}_X} M)$ denote the Cech-

complex given in dimension n by $\varepsilon_n * \varepsilon_n^{-1}(F)$ ($\varepsilon_n * \varepsilon_n^{-1}(\mathfrak{D}_{Y \leftarrow X} \otimes_{\mathfrak{D}_X} M)$, respectively.) Observe this is an equivariant complex of $f^{-1}(\mathfrak{D}_Y)$-modules. (See (1.8.2).) If $F^{\cdot} \in D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X))$ ($M^{\cdot} \in D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X))$, where each M^i is a locally free $\mathfrak{D}_X$-module) we let

$$Rf_*(F^{\cdot}) = \mathrm{Tot}(f_* C^{\cdot}(F^{\cdot})) \quad (\textstyle\int_f M^{\cdot} = \mathrm{Tot}(f_*(C^{\cdot}(\mathfrak{D}_{Y \leftarrow X} \otimes_{\mathfrak{D}_X} M^{\cdot}), \text{ respectively})$$

where Tot denotes the total complex. The discussion about the affine case above, (2.2) and (1.8.2) show that these define functors as in (2.8*).

(2.9) Next we will assume that $f\colon X \to Y$ is a *proper*, D-equivariant map of smooth quasi-projective complex varieties provided with linear actions by the complex diagonalisable group D as in (2.8). In this case we will show that the functors in (2.8*) induce functors

$$\text{(2.9*)} \qquad \begin{aligned} Rf_*&\colon D_b^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)) \to D^+_{b,\mathrm{coh}}(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathcal{O}_Y)) \\ \textstyle\int_f&\colon D_b^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) \to D^+_{b,\mathrm{coh}}(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_Y)). \end{aligned}$$

The assertion about Rf_* above follows from the fact that in the non-equivariant case Rf_* may again be computed using Cech-resolutions. The statement about $\int_f$ may be verified as follows. In view of the equivariant first exact sequence of Spencer (see (2.4)) it suffices to verify this for a single $\mathfrak{D}_X$-module of the form $\mathfrak{D}_X \otimes F \otimes \omega_X^{\vee}$, where F is a D-equivariant coherent $\mathcal{O}_X$-module. The projection formula (which clearly holds in the equivariant case) provides a D-equivariant quasi-isomorphism

$$\text{(2.9.1)} \qquad \textstyle\int_f(\mathfrak{D}_X \otimes E \otimes \omega_X^{\vee}) = \mathfrak{D}_Y \otimes (Rf_* E) \otimes \omega_Y^{\vee},\ E \in \mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)$$

as Rf_* also may be computed using a Cech-resolution and each such E has an equivariant resolution by a complex of equivariant locally free coherent $\mathcal{O}_X$-modules – (see [T-1], p. 557). (2.9*) follows.

(Under the assumptions of (2.8) we may also define à functor

$$\text{(2.9)}' \qquad f^!\colon D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y)) \to D^+(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_X)).$$

First one verifies using ((1.7 (ii) and ([Bj], p. 233) that $\mathfrak{D}_{X \to Y}$ is an equivariant left $\mathfrak{D}_X$ (right $f^{-1}\mathfrak{D}_Y$) bimodule. Let $c =$ the relative dimension of X over Y and let $N^{\cdot} \in D^+(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y))$. We let $f^!(N^{\cdot}) = \left(\mathfrak{D}_{X \to Y} \overset{L}{\underset{\mathfrak{D}_Y}{\otimes}} f^{-1}(N^{\cdot})\right)[c]$. The left derived functor $\overset{L}{\underset{\mathfrak{D}_Y}{\otimes}}$ is taken using the observation in (2.2). Let $D^+_{b,\mathrm{hol}}(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y))$ ($D^+_{b,\mathrm{hol}}(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_X))$) denote the derived category of bounded-below complexes

in $\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y)$ ($\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_X)$, respectively) with bounded holonomic cohomology. One observes that $f^!$ induces a functor

$$f^!\colon D^+_{b,\mathrm{hol}}(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y)) \to D^+_{b,\mathrm{hol}}(\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_X)).$$

We will make use of this fact in ([J-2] (2.4)).

Remark. In what follows we will denote the map $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y))$ ($K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Y))$) induced by $\int_f$ (Rf_*, respectively) by the same symbol.

(2.10) **Theorem.** *(Equivariant-Riemann-Roch for K_0: the smooth case).*

Assume the hypotheses of (2.9). *Let* $\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)$ ($\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)$) *denote the category of D-equivariant coherent* $\mathfrak{D}_X$*-modules (*$\mathcal{O}_X$*-modules, respectively) and similarly for Y. Then we obtain the commutative diagram*

(2.10*)
$$\begin{array}{ccc} K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) & \xrightarrow{\int_f} & K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y)) \\ \Big\downarrow{\scriptstyle d_X} & & \Big\downarrow{\scriptstyle d_Y} \\ K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)) & \xrightarrow{Rf_*} & K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Y)). \end{array}$$

Proof. Let $A(X)$ ($A(X)_{\mathrm{coh}}$) denote the full subcategory of $\mathrm{Mod}^D_{q\mathrm{coh}}(\mathfrak{D}_X)$ consisting of D-equivariant $\mathfrak{D}_X$-modules of the form $\mathfrak{D}_X \otimes E \otimes \omega_X^\vee$, where E is a quasi-coherent (coherent, respectively) D-equivariant $\mathcal{O}_X$-module. $A(Y)$ and $A(Y)_{\mathrm{coh}}$ will denote the corresponding categories for Y.

Let M be a D-equivariant coherent $\mathfrak{D}_X$-module and let $\mathrm{Sp}^{k+n}(M)^\cdot$ denote the 'first-exact-sequence of Spencer of order $n+k$' (See (2.4).) It is a D-equivariant resolution of M by $\mathfrak{D}_X$-modules in $A(X)_{\mathrm{coh}}$ provided k is large enough. We therefore obtain the surjections

(2.10.1) $K_0(A(X)_{\mathrm{coh}}) \twoheadrightarrow K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X))$ and $K_0(A(Y)_{\mathrm{coh}}) \twoheadrightarrow K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y))$

of the Grothendieck groups. (The Grothendieck groups on the left are defined in the obvious manner.)

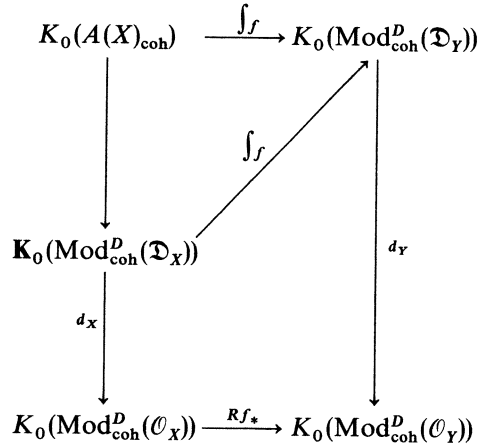
Next recall the definition of the D-module $\mathfrak{D}_{Y\leftarrow X}$ in (2.8). As in (2.9.1) we obtain a D-equivariant quasi-isomorphism

(2.10.2) $$\int_f (\mathfrak{D}_X \otimes E \otimes \omega_X^\vee) \simeq \mathfrak{D}_Y \otimes (Rf_* E) \otimes \omega_Y^\vee,\ E \in \mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)$$

since Rf_* also may be computed using a Cech-resolution. In order to prove the theorem, now it suffices to prove the commutativity of the diagram obtained by replacing the top row of (2.10*) with the map $\int_f$: $K_0(A(X_{\mathrm{coh}}) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y))$. Recall d_Y is an inverse to the isomorphism $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Y)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y))$ in (2.7). (2.10.2) now shows that the com-

position of the left vertical map, the bottom row and the inverse to the right vertical map ($=d_Y$) in

(2.10.3)

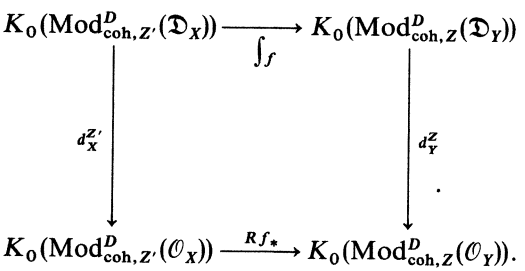


is the same as the top row. This proves the theorem.

(2.10′) **Theorem.** *(Equivariant-Riemann-Roch for K_0: the singular case). Assume in addition to the hypotheses of* (2.10) *that Z is a closed and possibly singular D-invariant subvariety of Y; let Z' be a closed D-invariant and possibly singular subvariety of $f^{-1}(Z)$.*

Let $\mathrm{Mod}^D_{\mathrm{coh},Z'}(\mathfrak{D}_X)$ $(\mathrm{Mod}^D_{\mathrm{coh},Z'}(\mathcal{O}_X))$ *denote the category of D-equivariant coherent $\mathfrak{D}_X$-modules ($\mathcal{O}_X$-modules, respectively) with supports contained in Z'. Let* $\mathrm{Mod}^D_{\mathrm{coh},Z}(\mathfrak{D}_Y)$ $(\mathrm{Mod}^D_{\mathrm{coh},Z}(\mathcal{O}_Y))$ *denote the corresponding categories for Y. Then we obtain the commutative diagram*

(2.10′*)



Proof. First observe that (2.4) provides an equivariant resolution of any $\mathfrak{D}_X$-module with support in Z' by a complex of equivariant modules of the form $\mathfrak{D}_X \otimes E^{\cdot} \otimes \omega_X^{\vee}$, with support of each E^i contained in Z'. The map d_X: $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X))$ (d_Y: $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Y)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_Y))$ induces a map $d_X^{Z'}$ (d_Y^{Z}, respectively) as in (2.10′*). It follows as in (2.7) that the two vertical maps in (2.10′*) are isomorphisms. As in the proof of (2.10) we may now replace the top row of (2.10′*) by $K_0(A(X)_{\mathrm{coh},Z'}) \to K_0(\mathrm{Mod}^D_{\mathrm{coh},Z}(\mathfrak{D}_Y))$ where $A(X)_{\mathrm{coh},Z'}$ is the full subcategory of $A(X)_{\mathrm{coh}}$ consisting of modules with supports in Z'. Finally observe that the two horizontal maps take sheaves with supports in Z' to sheaves with supports in Z. In view

of (2.10.2) it now suffices to prove this for Rf_* applied to a coherent $\mathcal{O}_X$-module with supports contained in Z'. This follows readily from 'formal proper-base-change' as in ([EGA] III, Prop. 4.2) and the fact that the completion of a noetherian local ring at the maximal ideal is faithfully flat over the original local ring.

Assume the hypotheses of (2.10'). We proceed to compose the maps d_X, d_Y with the corresponding natural transformations from the algebraic K-theory of equivariant coherent sheaves to equivariant topological K-homology. For this purpose we restrict to the mod-l^v version of D-equivariant K-homology defined as in ([T-2], p. 519) where l is any prime. The algebraic K-theory spectrum of the category $\mathrm{Mod}^D_{\mathrm{coh},Z'}(\mathcal{O}_X)$ ($\mathrm{Mod}^D_{\mathrm{coh},Z}(\mathcal{O}_Y)$) will be denoted $G^D(Z')$ ($G^D(Z)$, respectively) from now on. The corresponding mod-l^v versions are obtained by smashing these spectra with the Moore spectrum $M(l^v)$ which is the homology cofiber of the map $\Sigma^0 \xrightarrow{l^v} \Sigma^0$. These will be denoted $G^D_{l^v}(Z')$ and $G^D_{l^v}(Z)$ respectively. Let $G^{D,\mathrm{top}}_{l^v}(Z')$ and $G^{D,\mathrm{top}}_{l^v}(Z)$ denote the corresponding equivariant mod-l^v topological K-homology spectra defined as in ([T-2], p. 520). Under the hypotheses of (2.10'), let $\bar{f}\colon Z' \to Z$ denote the (equivariant) map induced by f. Observe (see [T-2], pp. 518–520) that the induced maps $G^D_{l^v}(Z') \to G^D_{l^v}(Z))$ and $G^{D,\mathrm{top}}_{l^v}(Z') \to G^{D,\mathrm{top}}_{l^v}(Z)$ are independent of the imbeddings $Z' \to X$ and $Z \to Y$.

(2.11) **Theorem.** *(Equivariant Riemann-Roch for Grothendieck groups.) Assuming the hypotheses of* (2.10') *we obtain the commutative diagram*

$$
\begin{array}{ccc}
K_0(\mathrm{Mod}^D_{\mathrm{coh},Z}(\mathfrak{D}_X)) & \xrightarrow[\int_f]{} & K_0(\mathrm{Mod}^D_{\mathrm{coh},Z}(\mathfrak{D}_Y)) \\
\downarrow{\scriptstyle d_X^{Z'}} & & \downarrow{\scriptstyle d_Y^{Z}} \\
\pi_0(G^D_{l^v}(Z')) & \xrightarrow[R\bar{f}_*]{} & \pi_0(G^D_{l^v}(Z)) \\
\downarrow & & \downarrow \\
\pi_0(G^{D,\mathrm{top}}_{l^v}(Z')) & \xrightarrow{\bar{f}_*} & \pi_0(G^{D,\mathrm{top}}_{l^v}(Z))
\end{array}
\tag{2.11*}
$$

where π_0 denotes the 0-th homotopy group of the corresponding spectrum.

Proof. The proof follows readily by combining the commutative square (2.10'*) with the commutative square given by the Equivariant-Riemann-Roch theorem in ([T-2], p. 524, (3.1)).

Remark. Let M denote a maximal compact subgroup of the Lie group $D(\mathbb{C})$. It is shown in ([T-3], Sect. 5) that $G^{D,\mathrm{top}}_{l^v}(Z')$ is naturally weakly-equivalent to the familiar mod-l^v M-equivariant topological K-homology of $Z'(\mathbb{C})$. This may be defined under the assumptions of (2.10') to be the M-equivariant topological

K-theory spectrum of $X(\mathbb{C})$ with supports in $Z'(\mathbb{C})$ smashed with the Moore spectrum $M(l^v)$. Applications of the above theorems will be discussed in [J-2] and in other forthcoming papers.

Acknowledgement. The author wishes to thank Robert Thomason, Masaki Kashiwara, J. Sekiguchi and Y. Shimuzu for helpful discussions.

References

[Ang-Lej] Angeniol, B., Lejeune-Jalabert, M.: Le theoreme de Riemann-Roch singulier pour les $\mathfrak{D}$-modules. Asterisque **130**, 130–159 (1985)

[Be-Ber] Beilinson, A., Bernstein, J.: Localisation de g-modules. C.R. Acad. Sci., Paris. **292**, January 1981

[Bj] Bjork, J.-E.: Rings of differential operators. New York: North Holland 1979

[Bo] Borel, A. et al.: Algebraic $\mathfrak{D}$-modules. Perspectives in mathematics, vol 2. Boston: Birkhäuser 1986

[B-B] Borho, W., Brylinski, J.-L.: Differential operators on homogeneous spaces-III. Invent. Math. 1–6 (1985)

[D-G] Demazure, M., Gabriel, P.: Groupes algebriques. Tome I. New York: North Holland 1970

[EGA] Dieudonne, J.A., Grothendieck, A.: Elements of geometrie algebrique. (Grundlehren, Chap. III, IV) Berlin Heidelberg New York: Springer 1971. Publ. Math. vols. 11, 17, 20, 24, 28

[Fr] Friedlander, E.: Etale homotopy of simplicial schemes. Ann. Math. Stud. **104**, (1983)

[Hart] Hartshorne, R.: Algebraic geometry. (Graduate Texts in Math., vol. 52) Berlin Heidelberg New York: Springer 1977

[J-1] Joshua, R.: Vanishing of odd dimensional intersection cohomology. Math. Z. **195**, 239–253 (1987)

[J-2] Joshua, R.: The trace formula for equivariant $\mathfrak{D}$-modules and perverse sheaves. Math. Z. (in press)

[J-3] Joshua, R.: Riemann-Roch for equivariant $\mathfrak{D}$-modules-II (to appear)

[J-4] Joshua, R.: Higher K-theory of weakly equivariant $\mathfrak{D}$-modules (to appear)

[Kash] Kashiwara, M.: B-functions and holonomic systems. Invent. Math. **38**, 33–53 (1978)

[Lau] Laumon, G.: Sur la categorie des $\mathfrak{D}$-modules filtres. (Lect. Notes Math., vol. 1016, pp. 150–237) Berlin Heidelberg New York: Springer 1983

[Mal] Malgrange, B.: Sur les image direct de $\mathfrak{D}$-modules. Manus. Math. **50**, 49–71 (1985)

[Mum] Mumford, D.: Geometric invariant theory. (Erg. Math., vol. 34) Berlin Heidelberg New York: Springer 1965

[Pop] Popescu, N.: Abelian categories with applications to rings and modules. New York London: Academic Press 1973

[Qui] Quillen, D.: Higher algebraic K-theory-I. (Lect. Notes Math., vol. 341) Berlin Heidelberg New York: Springer 1974

[SGA3] Demazure, M., Grothendieck, A.: Schemas en groupes-I. (Lect. Notes Math., vol. 151) Berlin Heidelberg New York: Springer 1970

[SGA6] Berthelot, P., Grothendieck, A., Illusie, L.: Theorie de intersection et theorem de Riemann-Roch. (Lect. Notes Math., vol. 225) Berlin Heidelberg New York: Springer 1971

[Sum] Sumihiro, H.: Equivariant completion. J. Math. Kyoto Univ. **14**, 1–28 (1974)

[T-1] Thomason, R.: Algebraic K-theory of group scheme actions. Ann. Math. Stud. **113**, 539–562 (1987)

[T-2] Thomason, R.: Lefschtez-Riemann-Roch theorem and coherent trace formula. Invent. Math. **85**, 515–543 (1986)

[T-3] Thomason, R.: Equivariant algebraic vs. topological K-homology. Atiyah-Segal style (preprint, 1986)

Math. Z. 206, 145–151 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On preordered fields related to Hilbert's 17th problem

Zeng Guangxin

Department of Mathematics, Fuzhou Teachers College, Jiangxi Province, People's Republic of China

Received August 8, 1988; in final form April 7, 1989

1. Introduction

In reference [4], McKenna introduced the concept of ordered fields with the Weak Hilbert Property to mean the class of ordered fields for which Hilbert's 17th problem is affirmative. In reference [7], we generalized the Weak Hilbert Property to preordered fields as follows:

A preordered field (K, S) is said to *have the Weak Hilbert Property*, if any rational function $f(X_1, \ldots, X_m)$ over K, which is positive definite on K for every ordering of (K, S), can be represented in the form

$$f(X_1, \ldots, X_m) = \sum_{i=1}^{n} s_i h_i^2(X_1, \ldots, X_m),$$

where $s_i \in S$, $h_i(X_1, \ldots, X_m) \in K(X_1, \ldots, X_m)$ for $i = 1, \ldots, n$.

Moreover, in [7], we have given some characterizations of these preordered fields.

The present paper is a continuation of reference [7]. Here, we shall make use of the terminology and results in [7], and shall reconsider preordered fields with the Weak Hilbert Property.

In this paper, we introduce another new notion, the so-called strongly local density, and show the equivalence between this notion and the Weak Hilbert Property. As a succedent result, we give again a characterization of preordered fields with the Weak Hilbert Property, which just generalizes the crucial implication in the definition of Weak Hilbert Property from rational to arbitrary function fields. Furthermore we discuss such a question: Let a preordered field (K^*, S^*) be an extension of (K, S), i.e., $K \subseteq K^*$ and $S \subseteq S^*$, has (K^*, S^*) the Weak Hilbert Property if (K, S) has the Weak Hilbert Property? For this, we may obtain an interesting result by using the mentioned above fact. Finally, we point out the fact that the family of preordered fields with the Weak Hilbert Property is not inductive, in other words, it is not closed under union of chains, although, as is pointed out by McKenna in [4], the family of ordered fields with the Weak Hilbert Property is inductive.

2. Main results

Before we establish the main results, the following definitions are needed:

Definition 1. Let (K, S) be a preordered field. A finitely generated extension K^* of K is said to be *a finitely generated real extension* of (K, S), if K^* is formally real with respect to S, that is, -1 can not be expressed in the form $\sum_{i=1}^{n} s_i b_i^2$, where $s_i \in S$, $b_i \in K^*$ for $i = 1, \ldots, n$.

Definition 2. A preordered field (K, S) is said to be *strongly locally dense*, if, for every finitely generated real extension K^* of (K, S), $\alpha, \beta \in K^*$ and $\alpha \neq \beta$, there exists always some $a \in K$, such that $a \in (\alpha, \beta)_{P^*}$ for some ordering P^* of K^* with $S \subseteq P^*$, where the meaning of $(\alpha, \beta)_{P^*}$ is as in Definition 2 of [7].

Now we are able to show the following result:

Theorem 1. *A preordered field (K, S) has the Weak Hilbert Property if and only if it is strongly locally dense.*

Proof. "$\Leftarrow$": Since every finite real extension of (K, S) is always a finitely generated real extension, the hypothesis "strongly local density" implies that (K, S) is locally dense. By Theorem 3 of [7], (K, S) has the Weak Hilbert Property.

"$\Rightarrow$": Let K^* be a finitely generated real extension of (K, S), $\alpha, \beta \in K^*$ and $\alpha \neq \beta$. We will show that K^* has an ordering P^* such that $S \subseteq P^*$ and $a \in (\alpha, \beta)_{P^*}$ for some $a \in K$.

Since K^* is finitely generated over K, we may assume $K^* = K(t_1, \ldots, t_m, \theta)$, where $t_1, \ldots, t_m$ is a transcendence base of K^* over K, θ is algebraic over the field $K(t_1, \ldots, t_m)$. Then we have $\alpha = u(t_1, \ldots, t_m, \theta)$, $\beta = v(t_1, \ldots, t_m, \theta)$, where $u(X_1, \ldots, X_m, X_{m+1})$, $v(X_1, \ldots, X_m, X_{m+1})$ are rational functions over K in $m+1$ variables $X_1, \ldots, X_m, X_{m+1}$.

Let the minimum polynomial of θ over $K(t_1, \ldots, t_m)$ be

$$G(\bar{t}; Y) = Y^d + e_1(\bar{t}) Y^{d-1} + \ldots + e_d(\bar{t})$$

where $e_i(\bar{t}) \in K(t_1, \ldots, t_m)$, $i = 1, \ldots, d$.

Since K^* is formally real with respect to S, K^* has an ordering P_0^* such that $S \subseteq P_0^*$. So the polynomial $G(\bar{t}; Y)$ has the root θ in the real closure of the ordered field (K^*, P_0^*).

Let $G_0(\bar{t}; Y) = G(\bar{t}; Y)$, $G_1(\bar{t}; Y), \ldots, G_s(\bar{t}; Y)$ be the Sturm sequence of $G(\bar{t}; Y)$, where $G_i(\bar{t}; Y)$ is a polynomial over $K(t_1, \ldots, t_m)$ in the indeterminate Y, $i = 0, 1, \ldots, s$. Moreover, we denote the set of all non-zero coefficients of these polynomials $G_i(\bar{t}; Y)$ by $\{c_j(\bar{t}) | j = 1, \ldots, r\}$. Obviously, $c_j(\bar{t}) \in K(t_1, \ldots, t_m)$, $j = 1, \ldots, r$.

Let $(\bar{K}, \bar{P}_0)$ be the real closure of the ordered field (K, P_0), where $P_0 = P_0^* \cap K$. Then, by Lang's Homomorphism Theorem, there is an K-algebra homomorphism τ from $K[t_1, \ldots, t_m, \theta, \alpha, \beta]$ to $\bar{K}$ satisfying the following conditions:

(i) $\tau\alpha \neq \tau\beta$, i.e., $u(\tau t_1, \ldots, \tau t_m, \tau\theta) \neq v(\tau t_1, \ldots, \tau t_m, \tau\theta)$;

(ii) $c_j(\tau t_1, \ldots, \tau t_m) \neq 0, j = 1, \ldots, r$.

and (iii) $G_Y'(\tau t_1, \ldots, \tau t_m, \tau\theta) \neq 0$.

Denote $K_1 = K(\tau t_1, \ldots, \tau t_m, \tau\alpha, \tau\beta, \tau\theta)$, we have that K_1 is a finite real extension of (K, S) because K_1 has the ordering $\bar{P}_0 \cap K_1$ containing S. By Theorem

3 of [7], (K, S) is locally dense since (K, S) has the Weak Hilbert Property. Hence K_1 has an ordering P_1 such that $S \subseteq P_1$ and $a \in (\tau\alpha, \tau\beta)_{P_1}$ for some $a \in K$.

Let $(\bar{K}_1, \bar{P}_1)$ be the real closure of the ordered field (K_1, P_1), then the polynomial $G(\overline{\tau t}; Y) = Y^d + e_1(\overline{\tau t})Y^{d-1} + \ldots + e_d(\overline{\tau t})$ has the root $\tau\theta$ in $\bar{K}_1$, and $G_0(\overline{\tau t}; Y) = G(\overline{\tau t}; Y), \ldots, G_s(\overline{\tau t}; Y)$ is the Sturm sequence of $G(\overline{\tau t}; Y)$. Choose a sufficiently small element $\delta \in \bar{P}_1$, we may assume that each of the polynomials $G_i(\overline{\tau t}; Y)$ has no roots different from $\tau\theta$ in the close interval $[\tau\theta - \delta, \tau\theta + \delta]_{\bar{P}_1}$ of $\bar{K}_1$. By the condition (iii), $\tau\theta$ is a simple zero of the polynomial $G(\overline{\tau t}; Y)$. Then, by the well-known Sturm's theorem, $\bigvee^{\bar{P}_1}_{\tau\theta+\delta} - \bigvee^{\bar{P}_1}_{\tau\theta-\delta} = 1$, where $\bigvee^{\bar{P}_1}_{\tau\theta\pm\delta}$ denote the numbers of variations in sign of $\{G_0(\overline{\tau t}; \tau\theta \pm \delta), G_1(\overline{\tau t}; \tau\theta \pm \delta), \ldots, G_s(\overline{\tau t}; \tau\theta \pm \delta)\}$ with respect to the ordering $\bar{P}_1$, respectively.

Let $\eta_1, \ldots, \eta_m$ be m indeterminates over $\bar{K}_1$, then we easily construct an ordering P_L of the field $L = \bar{K}_1(\eta_1, \ldots, \eta_m)$ such that all the η's are infinitesimal over $\bar{K}_1$ with respect to P_L, and we denote the real closure of the ordered field (L, P_L) by $(\bar{L}, \bar{P}_L)$. In this case it is obvious that $P_1 \subseteq \bar{P}_L$.

In view of infinitesimality of the η's, we easily see that $G_i(\overline{\tau t + \eta}, \tau\theta \pm \delta)$ has the same sign with respect to $\bar{P}_L$ as $G_i(\overline{\tau t}; \tau\theta \pm \delta)$ with respect to $\bar{P}_1$ for $i = 0, 1, \ldots, s$. Hence $\bigvee^{\bar{P}_L}_{\tau\theta+\delta} - \bigvee^{\bar{P}_L}_{\tau\theta-\delta} = \bigvee^{\bar{P}_1}_{\tau\theta+\delta} - \bigvee^{\bar{P}_1}_{\tau\theta-\delta} = 1$, where $\bigvee^{\bar{P}_L}_{\tau\theta\pm\delta}$ denote the numbers of variations in sign of $\{G_0(\overline{\tau t + \eta}; \tau\theta \pm \delta), G_1(\overline{\tau t + \eta}; \tau\theta \pm \delta), \ldots, G_s(\overline{\tau t + \eta}; \tau\theta \pm \delta)\}$ with respect to the ordering $\bar{P}_L$, respectively.

Notice that the Sturm sequence of $G(\overline{\tau t + \eta}; Y)$ is exactly $G_0(\overline{\tau t + \eta}; Y)$, $G_1(\overline{\tau t + \eta}; Y), \ldots, G_s(\overline{\tau t + \eta}; Y)$. Then, by Sturm's theorem, the polynomial $G(\overline{\tau t + \eta}; Y)$ has only a single root θ^* in the open interval $(\tau\theta - \delta, \tau\theta + \delta)_{\bar{P}_L}$ of $\bar{L}$. Hence $|\theta^* - \tau\theta| <_{\bar{P}_L} \delta$. Since δ can arbitrarily be chosen as small as possible in $\bar{P}_1$, $\theta^* - \tau\theta$ is infinitesimal over $\bar{K}_1$ with respect to $\bar{P}_L$.

Since $\eta_1, \ldots, \eta_m$, $\theta^* - \tau\theta$ are infinitesimal over $\bar{K}_1$ with respect to $\bar{P}_L$, and $a \in (\tau\alpha, \tau\beta)_{P_1}$, i.e., either $u(\tau t_1, \ldots, \tau t_m, \tau\theta) <_{\bar{P}_1} a <_{\bar{P}_1} v(\tau t_1, \ldots, \tau t_m, \tau\theta)$ or $v(\tau t_1, \ldots, \tau t_m, \tau\theta) <_{\bar{P}_1} a <_{\bar{P}_1} u(\tau t_1, \ldots, \tau t_m, \tau\theta)$, we have either $u(\tau t_1 + \eta_1, \ldots, \tau t_m + \eta_m, \theta^*) <_{\bar{P}_L} a <_{\bar{P}_L} v(\tau t_1 + \eta_1, \ldots, \tau t_m + \eta_m, \theta^*)$ or $v(\tau t_1 + \eta_1, \ldots, \tau t_m + \eta_m, \theta^*) <_{\bar{P}_L} a <_{\bar{P}_L} u(\tau t_1 + \eta_1, \ldots, \tau t_m + \eta_m, \theta^*)$, i.e., $a \in (u(\overline{\tau t + \eta}; \theta^*), v(\overline{\tau t + \eta}; \theta^*))_{\bar{P}_L}$.

Furthermore, we may obtain the K-injection π of $K^* = K(t_1, \ldots, t_m, \theta)$ to $\bar{L}$ such that $\pi t_i = \tau t_i + \eta_i$, $i = 1, \ldots, m$, and $\pi\theta = \theta^*$. Then we may define an ordering P^* of K^* as follows:

$$\text{for } a^* \in K^*,\ a^* \in P^* \text{ iff } \pi a^* \in \bar{P}_L .$$

Obviously, $S \subseteq P^*$, and $a \in (\alpha, \beta)_{P^*}$ since $\pi a = a \in (u(\overline{\tau t + \eta}; \theta^*), v(\overline{\tau t + \eta}; \theta^*))_{\bar{P}_L} = (\pi\alpha, \pi\beta)_{\bar{P}_L}$. Therefore (K, S) is strongly locally dense. □

By Theorem 1 and its proof, we can again give a characterization of preordered fields with the Weak Hilbert Property via special real places. To prove this result, we need the following lemma, which is actually a strengthened version of Definition 2.

Lemma. *If a preordered field (K, S) is strongly locally dense, then, for every finitely generated real extension K^* of (K, S), $\alpha_1, \ldots, \alpha_m \in K^*$ and $\delta \in \Sigma K^{*2}$, there exist $a_1, \ldots, a_m \in K$ such that $|\alpha_i - a_i| <_{P^*} \delta$, $i = 1, \ldots, m$, for some ordering P^* of K^* with $S \subseteq P^*$.*

Proof. We proceed by induction on m. The result is trivial for $m = 1$. Now suppose there exist $a_1, \ldots, a_{m-1} \in K$ such that $|\alpha_i - a_i| <_{P_1^*} \delta$, $i = 1, \ldots, m-1$, for some ordering P_1^* of K^* with $S \subseteq P_1^*$. Then the field $L = K^*(\sqrt{\delta^2 - (\alpha_1 - a_1)^2}, \ldots, \sqrt{\delta^2 - (\alpha_{m-1} - a_{m-1})^2})$ is a finitely generated real extension of (K, S), since $L \subseteq \bar{K}^*$, the real closure of the ordered field (K^*, P_1^*). Again, by strongly local density of (K, S), there exists some $a_m \in K$ such that $|\alpha_m - a_m| <_{P_L} \delta$ for some ordering P_L of L with $S \subseteq P_L$. Therefore, for the ordering P^* of K^* with $P^* = P_L \cap K^*$, $S \subseteq P^*$ and $|\alpha_i - a_i| <_{P^*} \delta$, $i = 1, \ldots, m$. This completes the proof. □

Now, we prove the following result:

Theorem 2. *A preordered field (K, S) has the Weak Hilbert Property if and only if, for every finitely generated real extension K^* of (K, S) with a transcendency base $t_1, \ldots, t_m$ over K and every $f \in K^*$, we have $f \in \Sigma SK^{*2}$ whenever $\sigma(f)$ is nonnegative for all places σ from K^* into some real closure $(\bar{K}, \bar{P})$, $S \subseteq P$, at which $\sigma(f) \neq \infty$ and $\sigma(t_i) \in K$ for $i = 1, \ldots, m$.*

Proof. "⇐": Let $f(X)$ be a positive definite polynomial over K in one variable X, we will show that $f(X) \in \Sigma S(K(X))^2$, i.e., $f(X)$ is strongly positive definite. So, by Theorem 3 of [7], (K, S) has the Weak Hilbert Property. Indeed, $K(X)$ is a finitely generated real extension of (K, S) with transcendency base X. For every place σ from $K(X)$ into some real closure $(\bar{K}, \bar{P})$, $S \subseteq P$, at which $\sigma(f(X)) \neq \infty$ and $\sigma(X) \in K$, we have $\sigma(f(X)) = f(\sigma(X))$. Since $f(X)$ is positive definite, we have $\sigma(f(X)) \geqq_{\bar{P}} 0$. Hence, by the hypothesis, $f(X) \in \Sigma S(K(X))^2$, and the proof of the "if" part is completed.

"⇒": Suppose that (K, S) has the Weak Hilbert Property, and that $f \notin \Sigma SK^{*2}$ for a finitely generated real extension K^* of (K, S) with transcendency base $t_1, \ldots, t_m$ over K and some $f \in K^*$. Then K^* has an ordering P^* such that $S \subseteq P^*$ and $f <_{P^*} 0$. Assume that $K^* = K(t_1, \ldots, t_m, \theta)$ where θ is algebraic over the field $K(t_1, \ldots, t_m)$, then $f = f(t_1, \ldots, t_m, \theta) = f(\bar{t}; \theta)$, where $f(\bar{X}; Y)$ is a rational function over K in $m+1$ variables $X_1, \ldots, X_m, Y$. For the rational function $f(\bar{X}; Y)$ and $t_1, \ldots, t_m, \theta$, there exists some $\delta \in \Sigma K^{*2}$ satisfying the condition in Lemma 3 of [7].[1]

Let notations $G(\bar{t}; Y)$, $G_i(\bar{t}; Y)$, $i = 0, 1, \ldots, s$, and $c_j(\bar{t})$, $j = 1, \ldots, r$, be as in the proof of the "only if" part of Theorem 1. Without loss of generality, we may assume that neither $\theta + \delta$ nor $\theta - \delta$ is a root for every polynomial $G_i(\bar{t}; Y)$, $i = 0, 1, \ldots, s$. Now the polynomial $G(\bar{t}; Y)$ has the root θ in (K^*, P^*), by Sturm's theorem we have $\bigvee_{\theta+\delta}^{P^*} - \bigvee_{\theta-\delta}^{P^*} \neq 0$, where $\bigvee_{\theta\pm\delta}^{P^*}$ denote the numbers of variations in sign of $\{G_0(\bar{t}; \theta \pm \delta), G_1(\bar{t}; \theta \pm \delta), \ldots, G_s(\bar{t}; \theta \pm \delta)\}$ with respect to the ordering P^*, respectively.

Let δ_i, δ_i' and $\delta_j'' \in \Sigma K^{*2}$ satisfy the condition in Lemma 3 of [7], respectively for $G_i(\bar{X}; Y)$ and $t_1, \ldots, t_m, \theta + \delta$, $G_i(\bar{X}; Y)$ and $t_1, \ldots, t_m, \theta - \delta$, and $c_j(\bar{X})$ and $t_1, \ldots, t_m$, $i = 0, 1, \ldots, s$; $j = 1, \ldots, r$. Put $u_i = \operatorname{sign}_{P^*}(G_i(\bar{t}; \theta + \delta))G_i(\bar{t}; \theta + \delta)$, $v_i = \operatorname{sign}_{P^*}(G_i(\bar{t}; \theta - \delta))G_i(\bar{t}; \theta - \delta)$, $i = 0, 1, \ldots, s$, where $\operatorname{sign}_{P^*}$ is the sign function, i.e., $\operatorname{sign}_{P^*}(x) = 1$ if $x >_{P^*} 0$ or $\operatorname{sign}_{P^*}(x) = -1$ if $x <_{P^*} 0$. Then

[1] In Lemma 3 of [7], the considered object $f(\bar{X})$ is a polynomial. In fact, the lemma still holds when $f(\bar{X})$ is rational. If $f(\bar{X}) = f_1(\bar{X})/f_2(\bar{X})$, where both $f_1(\bar{X})$ and $f_2(\bar{X})$ are polynomials, then we consider the polynomial $f_1(\bar{X})f_2(\bar{X})$ instead of $f(\bar{X})$, and obtain the corresponding result for rational functions

$K^{**} = K^*(\sqrt{u_0}, \ldots, \sqrt{u_s}, \sqrt{v_0}, \ldots, \sqrt{v_s}, \sqrt{-f})$ is a finitely generated real extension of (K, S), since $K^{**} \subseteq \bar{K}^*$, the real closure of the ordered field (K^*, P^*). By Theorem 1, (K, S) is strongly locally dense. Hence, by the lemma above, there exist $a_1, \ldots, a_m \in K$ such that $|t_i - a_i| <_{P^{**}} \varepsilon$ for some ordering P^{**} of K^{**} with $S \subseteq P^{**}$, where $\varepsilon = \frac{\delta}{1+\delta} \cdot \prod_{i=0}^{s} \left(\frac{\delta_i}{1+\delta_i} \cdot \frac{\delta_i'}{1+\delta_i'} \right) \cdot \prod_{j=1}^{r} \frac{\delta_j''}{1+\delta_j''} \in \Sigma K^{*2}$. So, we have $-f = (\sqrt{-f})^2 \in P^{**}$, and $\operatorname{sign}_{P^*}(G_i(\bar{t}; \theta \pm \delta)) G_i(\bar{t}; \theta \pm \delta) = (\sqrt{\operatorname{sign}_{P^*}(G_i(\bar{t}; \theta \pm \delta)) \cdot G_i(\bar{t}; \theta \pm \delta)})^2 \in P^{**}$, $i = 0, 1, \ldots, s$. Thereby $f(\bar{t}; \theta) <_{P^{**}} 0$, and $\operatorname{sign}_{P^*}(G_i(\bar{t}; \theta \pm \delta)) = \operatorname{sign}_{P^{**}}(G_i(\bar{t}; \theta \pm \delta))$, $i = 0, 1, \ldots, s$.

Notice that $|t_i - a_i| <_{P^{**}} \varepsilon <_{P^{**}} \min\{\delta, \delta_0, \delta_1, \ldots, \delta_s, \delta_0', \delta_1', \ldots, \delta_s', \delta_1'', \ldots, \delta_r''\}$. By Lemma 3 of [7], we have $\operatorname{sign}_{P^{**}}(G_i(\bar{a}; \theta \pm \delta)) = \operatorname{sign}_{P^{**}}(G_i(\bar{t}; \theta \pm \delta)) = \operatorname{sign}_{P^*}(G_i(\bar{t}; \theta \pm \delta))$, $i = 0, 1, \ldots, s$; and $c_j(\bar{a}) \neq 0$, $j = 1, \ldots, r$, since $\operatorname{sign}_{P^{**}}(c_j(\bar{a})) = \operatorname{sign}_{P^{**}}(c_j(\bar{t}))$. Hence $\bigvee_{\theta+\delta}^{P^{**}} - \bigvee_{\theta-\delta}^{P^{**}} = \bigvee_{\theta+\delta}^{P^*} - \bigvee_{\theta-\delta}^{P^*} \neq 0$, where $\bigvee_{\theta\pm\delta}^{P^{**}}$ denote the numbers of variations in sign of $\{G_0(\bar{a}; \theta \pm \delta), G_1(\bar{a}; \theta \pm \delta), \ldots, G_s(\bar{a}; \theta \pm \delta)\}$, and the Sturm sequence of $G(\bar{a}; Y)$ is precisely $G_0(\bar{a}; Y), G_1(\bar{a}; Y), \ldots, G_s(\bar{a}; Y)$. By Sturm's theorem, the polynomial $G(\bar{a}; Y)$ has a root ξ in (K^{**}, P^{**}), the real closure of the ordered field (K^{**}, P^{**}), such that $\xi \in (\theta - \delta, \theta + \delta)_{\overline{P^{**}}}$, i.e., $|\theta - \xi| <_{\overline{P^{**}}} \delta$. Hence, by Lemma 3 of [7], $\operatorname{sign}_{\overline{P^{**}}}(f(\bar{a}; \xi)) = \operatorname{sign}_{\overline{P^{**}}}(f(\bar{t}; \theta)) = -1$, i.e., $f(\bar{a}; \xi) <_{\overline{P^{**}}} 0$.

Denote the algebraic closure of K in K^{**} by $\bar{K}$. Then, by Lemma 1 of [3], $\bar{K}$ is real closed. So $(\bar{K}, \bar{P})$ is the real closure of the ordered field (K, P), where $\bar{P} = P^{**} \cap \bar{K}$, and $P = P^{**} \cap K$. It is obvious that $S \subseteq P$ and $\xi \in \bar{K}$.

Let $\eta_1, \ldots, \eta_m$ be m indeterminates over $\bar{K}$, and let P_1^* be an ordering of the field $\bar{K}(\eta_1, \ldots, \eta_m)$ such that all the η's are infinitesimal over $\bar{K}$. As in the proof of the "only if" part of Theorem 1, we can show that the polynomial $G(a + \eta; Y)$ has a root θ^* in $(\bar{L}, \bar{P}_L)$, the real closure of the ordered field $(\bar{K}(\eta_1, \ldots, \eta_m), P_1^*)$, and that $\theta^* - \xi$ is infinitesimal over $\bar{K}$ with respect to $\bar{P}_L$.

Likewise, we may obtain the K-injection π of $K^* = K(t_1, \ldots, t_m, \theta)$ to $\bar{L}$ such that $\pi t_i = a_i + \eta_i$, $i = 1, \ldots, m$; and $\pi\theta = \theta^*$.

Let σ_1 be the canonical place of $\bar{L}$ with respect to $\bar{K}$ (cf. [3]). Then the composition $\sigma = \sigma_1 \cdot \pi$ is obviously a place from K^* into $(\bar{K}, \bar{P})$. Since $\theta^* - \xi$ and η_i, $i = 1, \ldots, m$, are infinitesimal over $\bar{K}$, we have $\sigma_1(\theta^* - \xi) = 0$ and $\sigma_1(\eta_i) = 0$ for $i = 1, \ldots, m$. Hence $\sigma(t_i) = \sigma_1(\pi(t_i)) = \sigma_1(a_i + \eta_i) = \sigma_1(a_i) + \sigma_1(\eta_i) = a_i \in K$, $i = 1, \ldots, m$, and $\sigma(\theta) = \sigma_1(\pi(\theta)) = \sigma_1(\theta^*) = \sigma_1(\xi + \theta^* - \xi) = \sigma_1(\xi) + \sigma_1(\theta^* - \xi) = \xi$. Thereby $\sigma(f) = f(\sigma(t); \sigma(\theta)) = f(\bar{a}; \xi) <_{\bar{P}} 0$. This completes the proof of the "only if" part. □

Furthermore, Theorem 1 will be applied to show the next result. Here, for ease of representation we give again a definition as follows:

Definition 3. Let both (K, S) and (K^*, S^*) be preordered fields. (K^*, S^*) is called *a finitely generated extension of* (K, S), if K^* as a field is finitely generated over K, and S^* as a preordering is also finitely generated over S, that is, there exist finitely many elements $\alpha_1, \ldots, \alpha_m$ in K^* such that $S^* = \left\{ \sum_{i=1}^{n} \alpha_1^{k_{i1}} \ldots \alpha_m^{k_{im}} s_i b_i^{*2} \mid n \in \mathbb{N}, s_i \in S, b_i^* \in K^*, k_{ij} = 0 \text{ or } 1, i = 1, \ldots, n, j = 1, \ldots, m. \right\}$.

Theorem 3. *Let both* (K, S) *and* (K^*, S^*) *be preordered fields, and let* (K^*, S^*) *be a*

finitely generated extension of (K, S). *If* (K, S) *has the Weak Hilbert Property, then* (K^*, S^*) *has also the Weak Hilbert Property.*

Proof. By Theorem 1, it is sufficient to prove that (K^*, S^*) is strongly locally dense. Let K^{**} be a finitely generated real extension of (K^*, S^*), $\alpha, \beta \in K^{**}$ and $\alpha \neq \beta$. By the hypothesis, we may assume that S^* is generated over S by $\alpha_1, \ldots, \alpha_m$. Since K^{**} is formally real with respect to S^*, K^{**} has an ordering P_0^{**} such that $S^* \subseteq P_0^{**}$. Let (L, P_L) be the real closure of the ordered field (K^{**}, P_0^{**}). Then $K^{**}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_m}) \subseteq L$, and $K^{**}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_m})$ has an ordering $P_L \cap K^{**}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_m}) \supseteq S^* \supseteq S$. Hence $K^{**}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_m})$ is a finitely generated real extension of (K, S). By Theorem 1, (K, S) is strongly locally dense since (K, S) has the Weak Hilbert Property. So $K^{**}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_m})$ has an ordering P^* with $S \subseteq P^*$ such that $a \in (\alpha, \beta)_{P^*}$ for some $a \in K \subseteq K^*$. Putting $P^{**} = P^* \cap K^{**}$, we have $S \subseteq P^{**}$ and $\alpha_i = (\sqrt{\alpha_i})^2 \in P^* \cap K^{**} = P^{**}$ for $i = 1, \ldots, m$. Therefore K^{**} has the ordering P^{**} such that $S^* \subseteq P^{**}$, and for some $a \in K^*$ $a \in (\alpha, \beta)_{P^{**}}$. □

As a corollary of Theorem 5 of [4], McKenna has pointed out that the family of ordered fields with the Weak Hilbert Property is inductive. However, the analogous statement about preordered fields is not true. Let Ω be the family of preordered fields with the Weak Hilbert Property, and let $\leqq$ be the partial ordering of Ω such that for $(K_1, S_1), (K_2, S_2) \in \Omega$, $(K_1, S_1) \leqq (K_2, S_2)$ means (K_2, S_2) is an extension of (K_1, S_1), i.e., $K_1 \subseteq K_2$ and $S_1 \subseteq S_2$. If $\{(K_i, S_i) | i \in I\}$ is a chain in Ω, then it is easy to prove that $\left(\bigcup_i K_i, \bigcup_i S_i\right)$ is a preordered field, which is called the union of the chain $\{(K_i, S_i) | i \in I\}$. Then we have the following

Proposition. Ω *is not inductive with respect to* $\leqq$, *that is, the union of any chain in* Ω *is not necessarily in* Ω.

Proof. Suppose the statement is not true. Then the union of any chain in Ω has also the Weak Hilbert Property. From this, we will deduce a contradiction.

Following the example of Dubois [1], let F be the quadratic closure of the ordered field $(\mathbb{Q}(t), P)$ in its real closure, in which the indeterminate t is positive and infinitesimal over the field $\mathbb{Q}$ of rational numbers. Dubois has pointed out that F possesses the only ordering ΣF^2, but F has no the Weak Hilbert Property.

Consider the set $\Gamma = \{(K, \Sigma K^2) | K \subseteq F$ and $(K, \Sigma K^2)$ is a preordered field with the Weak Hilbert Property$\}$, we have $\Gamma \neq \varnothing$ since $(\mathbb{Q}, \mathbb{Q}^+) \in \Gamma$. Let $\{(K_i, S_i) | i \in I\}$ be a chain in Γ. Then it is also a chain in Ω. By the supposition, its union has the Weak Hilbert Property, and $\left(\bigcup_i K_i, \bigcup_i S_i\right) \in \Gamma$. Hence the set Γ is inductive with respect to $\leqq$. By Zorn's lemma, Γ contains a maximal element. However, we may assert that there are no maximal elements in the partially ordered set Γ. Indeed, for any $(K, \Sigma K^2) \in \Gamma$, necessarily $K \neq F$. Then there exists some $\alpha \in F$ such that $\alpha \notin K$. Obviously, $(K(\alpha), \Sigma K(\alpha)^2)$ is a finitely generated extension of $(K, \Sigma K^2)$. By Theorem 3, $(K(\alpha), \Sigma K(\alpha)^2)$ has the Weak Hilbert Property. Therefore $(K(\alpha), \Sigma K(\alpha)^2) \in \Gamma$, and $(K, \Sigma K^2) < (K(\alpha), \Sigma K(\alpha)^2)$. This is a contradiction. □

Let us finish this paper by the following problem, which may be regarded as the converse of Theorem 3.

Problem. Let (K, S) and (K^*, S^*) be as in Theorem 3. Has (K, S) necessarily the Weak Hilbert Property if (K^*, S^*) has the Weak Hilbert Property?

References

1. Dubois, D.W.: Note on Artin's solution of Hilbert's 17th problem. Bull. Am. Math. Soc. **73**, 540–541 (1967)
2. Lam, T.Y.: The theory of ordered fields. Lect. Notes in Pure and Appl. Math. **55**, 1–152 (1980)
3. Lang, S.: The theory of real places. Ann. Math. **57**, 371–390 (1953)
4. McKenna, K.: New facts about Hilbert's 17th problem. Lect. Notes Math. **498**, 220–230 (1975)
5. Nagata, M.: Field Theory. New York Basel: Dekker 1977
6. Prestel, A.: Lectures on formally real fields. IMPA Lect. Notes, no. 22. Rio de Janeiro, 1975
7. Zeng Guangxin: A characterization of preordered fields with the Weak Hilbert Property. Proc. Am. Math. Soc. **104**, 335–342 (1988)

Math. Z. 206, 153–168 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Shapiro's lemma and its consequences in the cohomology theory of modular Lie algebras

Rolf Farnsteiner[1] **and Helmut Strade**[2]

[1] Department of Mathematics, University of Wisconsin, Milwaukee, WI 53201, USA
[2] Mathematisches Seminar der Universität Hamburg, Bundesstrasse 55, D-2000 Hamburg 13, Federal Republic of Germany

Received May 3, 1989; in final form October 1, 1989

Introduction

The cohomology theory of modular Lie algebras has received considerable attention in the course of the last decade, partly due to the many obstacles encountered as compared to the non-modular theory. In this paper, we introduce new methods which furnish a unifying conceptual background for various other results (cf. [6, 10, 11, 24]). Our approach affords a simultaneous treatment of ordinary and restricted cohomology groups – a fact, which in view of the fundamental differences found between these two branches of the general theory of supplemented algebras, may be somewhat surprising.

Recent work by Friedlander and Parshall (cf. [15, 16]) has elicited interesting analogues between certain facets of the cohomology theories of groups and restricted Lie algebras, respectively. Parts of this article give further evidence for this affinity by providing a Lie theoretic version of Shapiro's lemma.

In the first section we introduce a principle of induction relative to certain subalgebras of the universal enveloping algebra $\mathscr{U}(L)$ of a finite dimensional modular Lie algebra L. The particular choice of these subalgebras allows to successively exploit features of the theory of Frobenius extensions (cf. Theorem 1.4) as well as the $\vee$-product theory of extension functors. Our approach results in a general cohomology reduction theorem for a class of modules that encompasses various types of Verma modules. Among the many consequences are a proof of a conjecture by G.B. Seligman concerning the existence of finite dimensional L-modules for which $H^n(L, M) \neq (0)$ if $0 \leqq n \leqq \dim_F L$, as well as a substantial extension and conceptualization of the results of [6] and [24]. Selected applications are given in the concluding section.

1. A natural equivalence

Throughout this note F is always a field of positive characteristic p and L a finite dimensional Lie algebra over F.

The notion of an induced module plays a central role in the representation theory of modular Lie algebras. In particular, all irreducible representations of solvable Lie algebras can be obtained through this process (cf. [26]). The situation for simple Lie algebras is inherently more complicated, yet many irreducible representations still arise as Verma modules (cf. [3, 21]). In this section we study modules which generalize restricted Verma modules.

In 1964 Berkson proved, that if $(L, [p])$ denotes a restricted finite dimensional Lie algebra, then its restricted universal enveloping algebra $u(L)$ is a Frobenius algebra (cf. [1]). In [27, Sect. 5.4] we have generalized this fact to algebras $\mathcal{U}(L)/I\mathcal{U}(L)$, where I is a certain subspace of the center of $\mathcal{U}(L)$. The methods developed there allow further generalizations which eventually will yield Shapiro's lemma for interesting classes of algebras.

In adopting the notation of [27], we define a partial ordering on $\mathbb{N}_0^k$ via $a \leqq b :\Leftrightarrow a_i \leqq b_i$ $1 \leqq i \leqq k$. Furthermore, we put $|a| := \Sigma_{1 \leqq i \leqq k} a_i$. Let $L_0 \subset L$ be a subalgebra with finite cobasis $\{e_1, \ldots, e_k\}$. Then there exist $m_1, \ldots, m_k$ and $v_i \in \mathcal{U}(L)_{(p^{m_i}-1)}$ such that $z_i := e_i^{p^{m_i}} + v_i$ belong to the center of $\mathcal{U}(L)$. We consider the subalgebra

$$\mathcal{O}(L, L_0) := \operatorname{alg}_F(L_0 \cup \{z_1, \ldots, z_k\})$$

in $\mathcal{U}(L)$.

Given $a \in \mathbb{N}_0^k$ we put $e^a := e_1^{a_1} \ldots e_k^{a_k}$ and define $\tau := (p^{m_1} - 1, \ldots, p^{m_k} - 1)$. Jacobson's refinement of the Poincaré-Birkhoff-Witt theorem then shows that $\mathcal{U}(L)$ is a free left (and right) $\mathcal{O}(L, L_0)$-module with basis $\{e^a;\ 0 \leqq a \leqq \tau\}$.

Lemma 1.1. *The following statements hold:*

(1) *If* $a \in \mathbb{N}_0^k$, *and* $a \nleqq \tau$, then $e^a \in \mathcal{O}(L, L_0)\mathcal{U}(L)_{(|a|-1)}$.

(2) $\mathcal{O}(L, L_0)\mathcal{U}(L)_{(|\tau|-1)} = \Sigma_{0 \leqq a < \tau} \mathcal{O}(L, L_0) e^a =: X$.

(3) *Let* $a, b \in \mathbb{N}_0^k$ *be given, such that* $|a + b| \leqq |\tau|$. *Then*

$$e^a e^b \equiv \delta(a + b, \tau) e^\tau \bmod (X) .$$

Proof. (1) By assumption there is $l \in \{1, \ldots, k\}$ such that $a_l \geqq p^{m_l}$. Recall that $z_l = e_l^{p^{m_l}} + v_l$, where $v_l \in \mathcal{U}(L)_{(p^{m_l}-1)}$. Consequently, $e^a \equiv z_l e^{a - p^{m_l}\varepsilon_l} \bmod (\mathcal{U}(L)_{(|a|-1)})$, and our assertion follows.

(2) We only have to verify the inclusion "$\subset$". We shall prove inductively that $\mathcal{O}(L, L_0)\mathcal{U}(L)_{(l)} \subset X$ for $l < |\tau|$, the assertion being trivial for $l = 0$. Let $l > 0$ and take any $u \in \mathcal{U}(L)_{(l)}$. The Poincaré-Birkhoff-Witt theorem ensures that u can be written as

$$u = \sum_{|a| \leqq l} \alpha(a) e^a, \qquad \alpha(a) \in \mathcal{O}(L, L_0) .$$

If $a \leqq \tau$, then $a < \tau$ and $e^a \in X$. Otherwise, (1) in conjunction with the inductive hypothesis yields $e^a \in \mathcal{O}(L, L_0)\mathcal{U}(L)_{(l-1)} \subset X$.

(3) Since $|a + b| \leqq |\tau|$ we obtain $e^a e^b \equiv e^{a+b} \bmod (\mathcal{U}(L)_{(|\tau|-1)})$. Owing to (2) this implies $e^a e^b \equiv e^{a+b} \bmod (X)$. If $a + b \leqq \tau$, the assertion follows. Otherwise, (1) shows that $e^{a+b} \equiv 0 \bmod (X)$. □

Lemma 1.2. *Let* $x \in L$ *and suppose that* $[x, e_j] \equiv \Sigma_{1 \leqq i \leqq k} \alpha_{ij}(x) e_i \bmod (L_0)$,

$1 \leqq j \leqq k$. *Let* $a \leqq \tau$ *and* $\Gamma(a) := \{b \leqq \tau;\ b \neq a, |b| = |a|\}$. *Then there exists a function* $\chi\colon \Gamma(a) \to F$ *such that*

$$[x, e^a] \equiv \Big(\sum_{1 \leqq i \leqq k} a_i \alpha_{ii}(x)\Big) e^a + \sum_{b \in \Gamma(a)} \chi(b) e^b \bmod (\mathcal{O}(L, L_\sigma)\mathcal{U}(L)_{(|a|-1)}) .$$

In particular,

$$[L_0, X] \subset X,\ [x, e^\tau] \equiv -\operatorname{tr}(\operatorname{ad}_{L/L_0} x) e^\tau \bmod (X)\ \forall x \in L_0$$

Proof. Let $\Omega(a) := \{b \geqq 0;\ b \neq a, |b| = |a|\}$. According to [27, (4.1), p. 215] there is a mapping $f\colon \Omega(a) \to F$ such that

$$[x, e^a] \equiv \Big(\sum_{1 \leqq i \leqq k} a_i \alpha_{ii}(x)\Big) e^a + \sum_{b \in \Omega(a)} f(b) e^b \bmod (\mathcal{O}(L, L_0)\mathcal{U}(L)_{(|a|-1)}) .$$

It readily follows from (1) of (1.1) that $e^b \in \mathcal{O}(L, L_0)\mathcal{U}(L)_{(|a|-1)}\ \forall b \in \Omega(a) - \Gamma(a)$. Consequently,

$$\sum_{b \in \Omega(a)} f(b) e^b \equiv \sum_{b \in \Gamma(a)} f(b) e^b \bmod (\mathcal{O}(L, L_0)\mathcal{U}(L)_{(|a|-1)})$$

and the asserted result follows. We apply this congruence to particular situations: If $|a| < |\tau|$, then $[x, e^a] \in \mathcal{O}(L, L_0)\mathcal{U}(L)_{(|\tau|-1)} \subset X$ for all $x \in L_0$. If $a = \tau$, then $[x, e^\tau] \equiv -\operatorname{tr}(\operatorname{ad}_{L/L_0} x) e^\tau + \Sigma_{b \in \Gamma(\tau)} \gamma(b) e^b \bmod (X)$. $\Gamma(\tau)$, however, is void. □

Let $\sigma\colon L_0 \to F$ be the Lie algebra homomorphism given by $\sigma(x) := \operatorname{tr}(\operatorname{ad}_{L/L_0} x)$ $\forall x \in L_0$. Since the correspondence $x \mapsto x + \sigma(x)1$ is a homomorphism from L_0 to $\mathcal{U}(L_0)^-$ there exists exactly one homomorphism $\psi\colon \mathcal{U}(L_0) \to \mathcal{U}(L_0)$ that extends the above mapping. Evidently, $x \mapsto x - \sigma(x)1$ defines the inverse of ψ. According to Jacobson's refinement of the Poincaré-Birkhoff-Witt theorem $\mathcal{O}(L, L_0)$ is isomorphic to $F[z_1, \ldots, z_k] \otimes_F \mathcal{U}(L_0)$, the first factor being a polynomial algebra in k indeterminates. Consequently, $\rho := \mathrm{id} \otimes \psi$ defines an algebra automorphism of $\mathcal{O}(L, L_0)$. We decompose

$$\mathcal{U}(L) = \oplus_{0 \leqq a < \tau} \mathcal{O}(L, L_0) e^a \oplus \mathcal{O}(L, L_0) e^\tau = X \oplus \mathcal{O}(L, L_0) e^\tau$$

as $\mathcal{O}(L, L_0)$-modules and let μ be the projection onto the second summand

$$\mu\colon \mathcal{U}(L) \to \mathcal{O}(L, L_0), \qquad \mu(x + \alpha e^\tau) = \alpha \qquad \forall \alpha \in \mathcal{O}(L, L_0),\quad x \in X .$$

Theorem 1.3. *With the definitions of* μ *and* ρ *as above, we have*

(1) $\mu(\alpha u) = \alpha\mu(u)$; $\mu(u\alpha) = \mu(u)\rho(\alpha)$ $\forall \alpha \in \mathcal{O}(L, L_0)$, $u \in \mathcal{U}(L)$.

(2) *There are uniquely determined elements* $(u_a)_{0 \leqq a \leqq \tau}$ *and* $(v_a)_{0 \leqq a \leqq \tau}$, *such that*

$$\mu(e^b u_a) = \delta(b, a) = \mu(v_a e^b) .$$

(3) *Let* I *be a right ideal of* $\mathcal{O}(L, L_0)$. *Then*

$$I\mathcal{U}(L) = u \in \mathcal{U}(L) | \mu(u\mathcal{U}(L)) \subset I .$$

(4) *Let* I *be a left ideal of* $\mathcal{O}(L, L_0)$. *Then*

$$\mathcal{U}(L)I = u \in \mathcal{U}(L) | \mu(\mathcal{U}(L)u) \subset \rho(I) .$$

Proof. (1) It obviously suffices to verify the second identity for the generating set $L_0 \cup \{z_1, \ldots, z_k\}$ of $\mathcal{O}(L, L_0)$. Since the z_i clearly have the requisite property we

consider $x \in L_0$ and $u = v + \beta e^\tau$, $v \in X$, $\beta \in \mathcal{O}(L, L_0)$. Then $\mu(u) = \beta$ and (1.2) yields

$$ux = vx + \beta e^\tau x \equiv xv + \beta(x + \sigma(x)1)e^\tau \equiv \beta\rho(x)e^\tau \bmod(X) .$$

(2) Give $(e^a)_{0 \leqq a \leqq \tau}$ a linear ordering by defining

$$a \ll b := |a| < |b|, \text{ or } |a| = |b| \text{ and } a \text{ precedes } b \text{ lexicographically} .$$

Put $f_a := e^{\tau - a}$ and consider tuples $a \ll b$:

If $|a| < |b|$ then $e^a f_b \equiv e^{\tau - b + a} \equiv 0 \bmod(\mathcal{U}(L)_{(|\tau|-1)})$. If $|a| = |b|$ then $|\tau + a - b| = |\tau|$, but $\tau + a - b \nleqq \tau$. Again we obtain $e^a f_b \equiv 0 \bmod(X)$. Similarly, $f_b e^a \equiv 0 \bmod(X)$ if $a \ll b$. Hence $\mu(e^a f_b) = 0 = \mu(f_b e^a)$ if $a \ll b$, $\mu(e^a f_a) = 1 = \mu(f_a e^a)$. It is then a straightforward routine to construct the unique u_a, v_a.

(3), (4) We just prove (4), (3) being similar. The inclusion $\mathcal{U}(L)I \subset \{u \in \mathcal{U}(L) \mid \mu(\mathcal{U}(L)u) \subset \rho(I)\}$ is true, since

$$\mu(\mathcal{U}(L)\mathcal{U}(L)I) \subset \mu(\mathcal{U}(L))\rho(I) \subset \mathcal{O}(L, L_0)\rho(I) \subset \rho(I) .$$

Assume that $u \in \mathcal{U}(L)$, $\mu(\mathcal{U}(L)u) \subset \rho(I)$. Observing that $(e^a)_{0 \leqq a \leqq \tau}$ is a basis of $\mathcal{U}(L)$ as a right $\mathcal{O}(L, L_0)$-module we obtain,

$$u = \sum_{0 \leqq a \leqq \tau} e^a \alpha(a), \alpha(a) \in \mathcal{O}(L, L_0) .$$

An application of (2) shows that

$$\mu(v_a u) = \mu(v_a e^a)\rho(\alpha(a)) = \rho(\alpha(a)) .$$

Hence $\alpha(a) \in \rho^{-1}(\rho(I)) = I$ for all $a \leqq \tau$. □

Theorem 1.3 can be restated by saying that $\mathcal{U}(L)$ is a Frobenius extension of $\mathcal{O}(L, L_0)$ (cf. [22]).

We proceed by introducing certain induced modules, which, as will be seen later, naturally generalize the concept of a restricted Verma module. Let V be an L_0-module. The action of $\mathcal{U}(L_0)$ on V can be extended to $\mathcal{O}(L, L_0)$ by letting the polynomial algebra $F[z_1, \ldots, z_k]$ operate via its canonical supplementation. *Henceforth all L_0-modules will be considered $\mathcal{O}(L, L_0)$-modules in this fashion.* We introduce a twisted action on V by setting $x \circ v := xv + \sigma(x)v$. The new L_0-module will be called V_σ and is easily seen to be isomorphic to $V \otimes_F F_\sigma$. Note that the resulting $\mathcal{O}(L, L_0)$-module structure on V_σ coincides with the one on V defined via $\alpha \circ v := \rho(\alpha)v \quad \forall \alpha \in \mathcal{O}(L, L_0), v \in V$.

Consider the *coinduced module* $\mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), V)$ with $\mathcal{U}(L)$-action given via $(u \cdot f)(x) := f(xu) \; \forall u, x \in \mathcal{U}(L)$.

Theorem 1.4. *Let V be an L_0-module. Then the mapping*

$\varphi_V : \mathcal{U}(L) \otimes_{\mathcal{O}(L, L_0)} V_\sigma \to \mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), V)$; $\varphi_V(u \otimes v)(u') := \mu(u'u)v$ *defines a natural isomorphism of $\mathcal{U}(L)$-modules.*

Proof. Consider the bilinear mapping $f : \mathcal{U}(L) \times V_\sigma \to \mathrm{Hom}_F(\mathcal{U}(L), V)$, $f((u, v))(u') = \mu(u'u)v$. Let $\alpha \in \mathcal{O}(L, L_0)$, $u' \in \mathcal{U}(L)$. According to (1) of (1.3) we obtain, observing the $\mathcal{O}(L, L_0)$-structure of V_σ,

$$f((u\alpha, v))(u') = \mu(u'u\alpha)v = \mu(u'u)\rho(\alpha)v = f((u, \alpha \circ v))(u') .$$

Consequently, f is $\mathcal{O}(L, L_0)$-balanced, and thereby induces a mapping

$$\varphi_V\colon \mathcal{U}(L) \otimes_{\mathcal{O}(L, L_0)} V_\sigma \to \mathrm{Hom}_F(\mathcal{U}(L), V) .$$

The verification of the inclusion $\mathrm{im}\varphi_V \subset \mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), V)$ as well as the $\mathcal{U}(L)$-linearity of φ_V is routine.

Assume that $y = \Sigma_{0 \leqq a \leqq \tau} e^a \otimes w_a \in \ker\varphi_V$. By applying (1.3 (2)) we obtain

$$0 = \varphi_V(y)(v_b) = \mu\Big(\sum_a v_b e^a\Big) w_a = w_b \qquad \text{for all } b .$$

Thus $y = 0$, proving that φ_V is injective.

In order to verify surjectivity we observe, that every element $f \in \mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), V)$ can be written as

$$f = \sum_{0 \leqq a \leqq \tau} \varphi_V(v_a \otimes f(e^a)) .$$

It remains to verify the naturality. To that end let $f\colon V \to W$ be a homomorphism of L_0-modules. Clearly, f is also a homomorphism between V_σ and W_σ. We claim that the diagram

$$\begin{array}{ccc} \mathcal{U}(L) \otimes_{\mathcal{O}(L, L_0)} V_\sigma & \xrightarrow{\varphi_V} & \mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), V) \\ \downarrow \mathrm{id} \otimes f & & \downarrow f^* \\ \mathcal{U}(L) \otimes_{\mathcal{O}(L, L_0)} W_\sigma & \xrightarrow{\varphi_W} & \mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), W) \end{array}$$

is commutative. Noting that f^* and $\mathrm{id} \otimes f$ are homomorphisms of $\mathcal{U}(L)$-modules, the assertion follows from the ensuing calculation:

$$f^* \circ \varphi_V(1 \otimes v)(u') = \mu(u') f(v) = (\varphi_W \circ (\mathrm{id} \otimes f))(1 \otimes v)(u') \; \forall u' \in \mathcal{U}(L). \quad \square$$

Remarks. (1) If one applies the above result to the module $V_{-\sigma}$ one obtains, observing $(V_{-\sigma})_\sigma = V$, natural isomorphisms

$$\mathcal{U}(L) \otimes_{\mathcal{O}(L, L_0)} V \cong \mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), V_{-\sigma}) .$$

(2) Suppose that L_0 acts nilpotently on L/L_0 or that $(\mathrm{ad}_{L/L_0} L_0)^{(1)} = (\mathrm{ad}_{L/L_0} L_0)$. Then $\sigma = 0$ and every L_0-module V gives rise to an isomorphism $\mathcal{U}(L) \otimes_{\mathcal{O}(L, L_0)} V \cong \mathrm{Hom}_{\mathcal{O}(L, L_0)}(\mathcal{U}(L), V)$.

We now specialize our considerations to restricted Lie algebras $(L, [p])$. Adopting the above general assumptions, we let L_0 be a p-subalgebra of L and fix a linear form $S \in L^*$. In that case we can define

$$z_i := e_i^p - e_i^{[p]} - S(e_i)^p 1 \qquad 1 \leqq i \leqq k$$

so that $m_i = 1$ for all i. Let V be an L_0-module with character $S|_{L_0}$. Then V_σ also has character $S|_{L_0}$. We let $u(L, S)$ denote the factor algebra of $\mathcal{U}(L)$ by the ideal generated by the z_i (cf. [27] for basic properties).

Proposition 1.5. *Let V be an L_0-module with character $S|_{L_0}$. Then there exist natural isomorphisms of $\mathcal{U}(L)$-modules*

(1) $\gamma_V\colon \mathcal{U}(L) \otimes_{\mathcal{O}(L, L_0)} V \to u(L, S) \otimes_{u(L_a S|L_a)} V$

(2) $\lambda_V\colon \mathrm{Hom}_{u(L_a, S|L_a)}(u(L, S), V) \to \mathrm{Hom}_{\mathcal{O}(L, L_a)}(\mathcal{U}(L), V)$.

Proof. (1) Let $\pi\colon \mathscr{U}(L) \to u(L, S)$ denote the canonical projection and observe that π maps $\mathscr{U}(L_0)$ onto $u(L_0, S|_{L_0})$. Consider the bilinear mapping $q\colon \mathscr{U}(L) \times V \to u(L, S) \otimes_{u(L_0, S|_{L_0})} V$ which is given by

$$q(u, v) = \pi(u) \otimes v \qquad \forall u \in \mathscr{U}(L),\ v \in V\,.$$

Since $q(uz_i, v) = \pi(uz_i) \otimes v = \pi(u)\pi(z_i) \otimes v = 0$ and $z_i v = 0$ $(1 \leqq i \leqq k)$, as well as

$$q(ux, v) = \pi(u)\pi(x) \otimes v = \pi(u) \otimes \pi(x)v = \pi(u) \otimes xv = q(u, xv) \qquad \forall x \in L_0$$

it follows that q is $\mathscr{O}(L, L_0)$-balanced. Hence there results an F-linear mapping $\gamma_V\colon \mathscr{U}(L) \otimes_{\mathscr{O}(L, L_0)} V \to u(L, S) \otimes_{u(L_0, S|_{L_0})} V$ such that $\gamma_V(u \otimes v) = \pi(u) \otimes v$. Clearly, γ_V is a homomorphism of $\mathscr{U}(L)$-modules. Let $(v_j)_{j \in J}$ be an F-basis of V. Then $\mathscr{B} := \{e^a \otimes v_j | j \in J,\ 0 \leqq a \leqq \tau\}$ and $\mathscr{C} := \{\pi(e^a) \otimes v_j | j \in J,\ 0 \leqq a \leqq \tau\}$ are bases of $\mathscr{U}(L) \otimes_{\mathscr{O}(L, L_0)} V$ and $u(L, S) \otimes_{u(L_0, S|_{L_0})} V$, respectively. Since γ_V maps $\mathscr{B}$ bijectively onto $\mathscr{C}$, the invertibility of γ_V follows.

To verify the naturality let $f\colon V \to W$ be a homomorphism between two L_0-modules with character S. Then $\gamma_W \circ (\mathrm{id} \otimes f)(u \otimes v) = \gamma_W(u \otimes f(v)) = \pi(u) \otimes f(v) = (\mathrm{id} \otimes f)(\pi(u) \otimes v) = (\mathrm{id} \otimes f) \circ \gamma_V(u \otimes v)$, as desired.

(2) Consider $\lambda_V\colon \mathrm{Hom}_{u(L_0, S|_{L_0})}(u(L, S), V) \to \mathrm{Hom}_{\mathscr{O}(L, L_0)}(\mathscr{U}(L), V)$; $\lambda_V(f) := f \circ \pi$. For $x \in L_0$, we have $\lambda_V(f)(xu) = f(\pi(x)\pi(u)) = \pi(x) f(\pi(u)) = x\lambda_V(f)(u)$ while $\lambda_V(f)(z_i u) = f(\pi(z_i)\pi(u)) = 0 = z_i \lambda_V(f)(u)$ $1 \leqq i \leqq k$.

Consequently, λ_V is well defined. Let u, u' be two elements of $\mathscr{U}(L)$. Then $(u \cdot \lambda_V(f))(u') = \lambda_V(f)(u'u) = f(\pi(u')\pi(u)) = (\pi(u) \cdot f)(\pi(u')) = (u \cdot f)(\pi(u')) = \lambda_V(u \cdot f)(u')$. As a result λ_V is a homomorphism of $\mathscr{U}(L)$-modules. The injectivity of λ_V is a direct consequence of its definition. We let $\Phi_v^{(a)}\colon u(L, S) \to V$ denote the $u(L_0, S|_{L_0})$-linear mapping sending $\pi(e^b)$ into $\delta(a, b)v$ $(0 \leqq b \leqq \tau)$. The analogous $\mathscr{O}(L, L_0)$-linear mapping $\mathscr{U}(L) \to V$ has $\Phi_v^{(a)}$ as a preimage under λ_V. Hence λ_V is an isomorphism. Finally, let $\alpha\colon V \to W$ be a homomorphism of L_0-modules. Given $f \in \mathrm{Hom}_{u(L_0, S|_{L_0})}(u(L, S), V)$ we obtain

$$\begin{aligned}\lambda_W \circ \mathrm{Hom}(\mathrm{id}, \alpha)(f) &= \mathrm{Hom}(\mathrm{id}, \alpha)(f) \circ \pi = \alpha \circ f \circ \pi \\ &= \mathrm{Hom}(\mathrm{id}, \alpha)(f \circ \pi) = \mathrm{Hom}(\mathrm{id}, \alpha) \circ \lambda_V(f).\end{aligned}$$ □

We continue to assume that L_0 is a p-subalgebra of the restricted Lie algebra $(L, [p])$ and specialize $S = 0$. The set $J := \{x^p - x^{[p]} | x \in L\}$ is contained in $\mathscr{O}(L, L_0)$. Let $I := \mathscr{O}(L, L_0)J$ be the two sided ideal of $\mathscr{O}(L, L_0)$ which is generated by J. Then $\mathscr{U}(L)/\mathscr{U}(L)I = u(L, 0) =: u(L)$ and the canonical projection $\pi\colon \mathscr{U}(L) \to u(L)$ induces an isomorphism $\mathscr{O}(L, L_0)/I \cong u(L_0)$. According to (1.3) there exists a $u(L_0)$-linear mapping $\nu\colon u(L) \to u(L_0)$ such that

$$\nu(\pi(u)) = \pi(\mu(u)) \qquad \forall u \in \mathscr{U}(L)\,.$$

If $\gamma\colon u(L_0) \to u(L_0)$ denotes the unique automorphism of $u(L_0)$ that maps x onto $x + \sigma(x)$ $(x \in L_0)$, then we obtain, in analogy with (1.3), the following facts:

(1) $\nu(\alpha u) = \alpha \nu(u) \qquad \forall u \in u(L),\ \alpha \in u(L_0)$
(2) $\nu(u\alpha) = \nu(u)\gamma(\alpha) \qquad \forall u \in u(L),\ \alpha \in u(L_0)$
(3) there exist uniquely determined families $(u_a)_{0 \leqq a \leqq \tau}$ and $(v_a)_{0 \leqq a \leqq \tau}$ such that $\nu(e^b u_a) = \delta(b, a) = \nu(v_a e^b) \qquad \forall a, b \leqq \tau$.

The following result can now be obtained by adopting the arguments of (1.4) mutatis mutandis.

Corollary 1.6. *Let V be a restricted L_0-module. Then the mapping ξ_V: $u(L) \otimes_{u(L_0)} V_\sigma \to \mathrm{Hom}_{u(L_0)}(u(L), V)$ which is given by $\xi_V(u \otimes v)(u') = v(u'u)v$ $\forall\, u, u' \in u(L)$, $\forall\, v \in V$ defines a natural equivalence.* □

2. Reduction theorems

Adopting the general assumptions and the notation of the preceding section we now turn to the study of cohomology groups. Let ε: $\mathscr{U}(L) \to F$ denote the natural supplementation. In the sequel we shall assume that the elements $z_1, \ldots, z_k$ are contained in $\ker \varepsilon$. Then the restriction of ε to $F[z_1, \ldots, z_k]$ is the natural supplementation of the polynomial ring.

Theorem 2.1. *Let V be an L_0-module. Then there exist isomorphisms*

$$H^n(L, \mathscr{U}(L) \otimes_{\mathscr{C}(L, L_0)} V) \cong \oplus_{p+q=n} \Lambda^p(L/L_0) \otimes_F H^q(L_0\, V_{-\sigma}) \qquad \forall\, n \geqq 0\,.$$

Proof. From (1.4) and its succeeding remark we obtain the following isomorphisms:

$$H^n(L, \mathscr{U}(L) \otimes_{\mathscr{C}(L, L_0)} V) \cong H^n(\mathscr{U}(L), \mathrm{Hom}_{\mathscr{C}(L, L_0)}(\mathscr{U}(L), V_{-\sigma})) \cong$$

$$\mathrm{Ext}^n_{\mathscr{U}(L)}(F, \mathrm{Hom}_{\mathscr{C}(L, L_0)}(\mathscr{U}(L), V_{-\sigma}))\,.$$

Since $\mathscr{U}(L)$ is a free left $\mathscr{C}(L, L_0)$-module, Proposition 7.2 [2, p. 196] ensures the existence of isomorphisms

$$\mathrm{Ext}^n_{\mathscr{U}(L)}(F, \mathrm{Hom}_{\mathscr{C}(L, L_0)}(\mathscr{U}(L), V_{-\sigma})) \cong \mathrm{Ext}^n_{\mathscr{C}(L, L_0)}(F, V_{-\sigma})\,.$$

In order to compute the right-hand groups we recall that $\mathscr{C}(L, L_0)$ is isomorphic to $F[z_1, \ldots, z_k] \otimes_F \mathscr{U}(L_0)$. As F is a field the $\vee$-product is well-defined (cf. [2, p. 204 f]):

$$\vee : \mathrm{Ext}_{\mathscr{G}}(F, F) \otimes_F \mathrm{Ext}_{\mathscr{U}(L_0)}(F, V_{-\sigma}) \to \mathrm{Ext}_{\mathscr{C}(L, L_0)}(F \otimes_F F, F \otimes_F V_{-\sigma})\,,$$

where $\mathscr{G} := F[z_1, \ldots, z_k]$. Since the algebras $\mathscr{G}$ and $\mathscr{U}(L_0)$ are left noetherian, Theorem 3.1. [2, p. 209] applies and $\vee$ is an isomorphism of degree 0. We thus obtain, by considering degree n constituents, isomorphisms

$$\oplus_{p+q=n} \mathrm{Ext}^p_{\mathscr{G}}(F, F) \otimes_F \mathrm{Ext}^q_{\mathscr{U}(L_0)}(F, V_{-\sigma}) \cong \mathrm{Ext}^n_{\mathscr{C}(L, L_0)}(F \otimes_F F, F \otimes_F V_{-\sigma})$$

$$\cong \mathrm{Ext}^n_{\mathscr{C}(L, L_0)}(F, V_{-\sigma})\,.$$

Since $\mathscr{G}$ is the universal enveloping algebra of the k-dimensional abelian Lie algebra with underlying vector space L/L_0 the groups $\mathrm{Ext}^p_{\mathscr{G}}(F, F)$ can be identified with the wedge products $\Lambda^p(L_0)$. □

The following Corollary proves a conjecture by G.B. Seligman [23, p. 102]. We like to thank G.M. Benkart for pointing out the applicability of the Reduction Theorem to this problem.

Corollary 2.2. *Let L be a finite dimensional Lie algebra. Then there exists a finite dimensional L-module M such that $H^n(L, M) \neq 0$ for $0 \leqq n \leqq \dim_F L$.*

Proof. Put $L_0 = (0)$. Then $H^q(L_0, F) = F$ for $q = 0$ and (0) otherwise. Hence (2.1) yields

$$H^n(L, \mathscr{U}(L) \otimes_{\mathscr{U}(L, L_0)} F_\sigma) \cong \Lambda^n(L) . \quad \square$$

It is well known that the cohomology groups of L with coefficient in an L-module M vanish in case one of the z_i operates invertibly on M (cf. [12]). The following direct consequence of (2.1) shows that there are $\mathscr{U}(L)$-modules Q for which $H^*(L, Q) = (0)$ while $z_i Q = (0)$ $1 \leqq i \leqq k$.

Corollary 2.3. *Suppose that $L_0 \neq (0)$ is properly contained in L. There exists an L-module $Q \neq (0)$ such that $H^n(L, Q) = (0)$ $\forall\ n \geqq 0$ and $z_i Q = (0)$ $1 \leqq i \leqq k$. In particular, Q is not $\mathscr{U}(L)$-injective.*

Proof. We let L_0 act on $\mathscr{U}(L_0)$ via left multiplication and consider an injective hull I of $\mathscr{U}(L_0)$. Since I is an essential extension of $\mathscr{U}(L_0)$ and $\mathscr{U}(L_0)^{L_0} = (0)$, it follows that $H^0(L_0, I) = (0)$. Owing to the injectivity of I we thus obtain $H^q(L_0, I) = (0)$ $\forall\, q \geqq 0$. According to Theorem 2.1 the cohomology groups of L with coefficients in $Q := \mathscr{U}(L) \otimes_{\mathscr{U}(L, L_0)} I_\sigma$ are trivial. Since the z_i annihilate Q, Q is not divisible hence not injective. $\square$

Let $(L, [p])$ be a restricted Lie algebra with restricted universal enveloping algebra $u(L) = u(L, 0)$. Following Hochschild [20] we define the restricted cohomology groups of L with coefficients in a restricted L-module M via

$$H^n_*(L, M) := H^n(u(L), M) \qquad \forall\, n \geqq 0 .$$

The cohomology groups on the right hand side are those of the supplemented F-algebra $(u(L), \varepsilon)$. In analogy to (2.1) we obtain the following result which can also be derived from the cohomology theory of algebraic groups.

Theorem 2.4. *Let L be a restricted Lie algebra, L_0 a p-subalgebra and V a restricted L_0-module. Then there exist isomorphisms*

$$H^n_*(L, u(L) \otimes_{u(L_0)} V) \cong H^n_*(L_0, V_{-\sigma}) \qquad \forall\, n \geqq 0 .$$

Proof. Since $u(L)$ is a free left $u(L_0)$-module, we obtain $\operatorname{Ext}^n_{u(L_0)}(u(L), V_{-\sigma}) = (0)$ $\forall\, n > 0$. According to (4.1.4) of [2, p. 118] there result isomorphisms

$$\operatorname{Ext}^n_{u(L)}(F, \operatorname{Hom}_{u(L_0)}(u(L), V_{-\sigma})) \cong \operatorname{Ext}^n_{u(L_0)}(F, V_{-\sigma}) \qquad \forall\, n \geqq 0 .$$

We now apply (1.6) in order to obtain

$$H^n_*(L, u(L) \otimes_{u(L_0)} V) \cong \operatorname{Ext}^n_{u(L)}(F, u(L) \otimes_{u(L_0)} V) \cong$$

$$\operatorname{Ext}^n_{u(L)}(F, \operatorname{Hom}_{u(L_0)}(u(L), V_{-\sigma})) \cong \operatorname{Ext}^n_{u(L_0)}(F, V_{-\sigma}) \cong H^n_*(L_0, V_{-\sigma})\ \forall\, n \geqq 0 . \quad \square$$

As an easy application we obtain the following non-vanishing criterion for the second restricted cohomology groups.

Corollary 2.5. *Suppose that $L_0 \subset L$ is a non-toral p-subalgebra of L. Let V be a restricted finite dimensional L_0-module such that $xv = \sigma(x)v$ for all $x \in L_0$ and $v \in V$. Then*

$$\dim_F H^2_*(L, u(L) \otimes_{u(L_0)} V) \geqq \dim_F V .$$

Proof. According to (2.4) the second restricted cohomology group of L with coefficients in $u(L) \otimes_{u(L_0)} V$ is isomorphic to $H^2_*(L_0, V_{-\sigma})$. Since the latter coefficient module is trivial its second cohomology group is isomorphic to the $(\dim_F V)$-fold direct sum of $H^2_*(L_0, F)$. As L_0 is not a torus, this space is, owing to (3.1) of [9], at least one-dimensional. □

Examples. Let $\underline{n} := (n_1, \ldots, n_m)$ be an m-tuple with positive integers as entries and consider $L = X(m; \underline{n})^{(2)}$ $(X = W, S, H)$ simple Lie algebras of Cartan type. For the contact algebras $K(m; \underline{n})^{(1)}$ one can argue similarly but has to handle the details a little bit more carefully. These are subalgebras of the derivation algebra of the divided power series algebra $A(m; \underline{n})$. Let L_0 denote the unique maximal subalgebra of L, which is invariant under all automorphisms. Then $L = L_0 \oplus \Sigma_{0 \leqq i \leqq m} FD_i$, where the D_i are the analogues of the partial derivatives. Put $z_i := D_i^{p^{n_i}} \in \mathscr{U}(L)$. If V is a L_0-module, the computation of the cohomology groups H^r of L with coefficients in $\mathscr{U}(L) \otimes_{\mathscr{U}(L_0)} V$ is reduced by (2.1) to the computation of $H^q (q \leqq r)$ of L_0 with coefficients in $V_{-\sigma}$. For $L = W(1; \underline{n})$, and $L = W(m; \underline{n})$, $L = H(2; \underline{n})^{(2)}$ these specializations of (2.1) were verified by more computational techniques in [6] and [24], respectively. (We mention here, that the modules investigated in [24] are coinduced modules, as will be shown in a forthcoming paper by R. Farnsteiner). By the same token (2.4) traces the restricted cohomology of the restricted Cartan type Lie algebras with coefficients in $u(L) \otimes_{u(L_0)} V$ to that of the subalgebra L_0. Since many of the irreducible restricted modules for the Cartan type Lie algebras are restricted Verma modules, our results provide an efficient method to reduce the computations of cohomology groups of these Lie algebras with coefficients in irreducible modules to the analogous problem for classical Lie algebras $\mathfrak{gl}(m)$, $\mathfrak{sl}(m)$, $\mathfrak{sp}(m)$ and $\mathfrak{sp}(m) \oplus F$.

Let $(L, [p])$ be restricted and suppose that $L_0 \subset L$ is a p-subalgebra satisfying $\sigma(L_0) = (0)$ (Note that p-ideals have this property). Given a restricted L-module M, (1.6) furnishes a surjective and natural p-homomorphism

$$\Theta_M \colon \mathrm{Hom}_{u(L_0)}(u(L), M) \to M$$

of L-modules. In analogy with the cohomology theory of finite groups (cf. [17, p. 226] we thus obtain *corestriction mappings*

$$\mathrm{Cor}_n \colon H^n_*(L_0, M) \to H^n_*(L, M) \qquad \forall n \geqq 0 .$$

Note that these mappings define a natural transformation between the relevant functors. The ensuing result investigates the composite of Cor_n and the canonical restriction transformations $\mathrm{Res}_n \colon H^n_*(L, M) \to H^n_*(L_0, M)$. Recall that u_0 is the unique element of $u(L)$ subject to $v(e^a u_0) = \delta(a, 0)$ $0 \leqq a \leqq \tau$.

Theorem 2.6. *Let M be a restricted L-module. Then*

$$\mathrm{Cor}_n \circ \mathrm{Res}_n = \varepsilon(u_0)\, \mathrm{id}_{H^n_*(L, M)} \qquad \forall n \geqq 0 .$$

Proof. We proceed by induction on n. For $n = 0$, Res_0 simply embeds M^L into M^{L_0} while the corestriction map first sends an element $m \in M^{L_0}$ onto $f_m \in \mathrm{Hom}_{u(L_0)}(u(L), M)^L$, where $f_m(u) := \varepsilon(u)m \; \forall u \in u(L)$. The isomorphism given in (1.6) maps f_m onto $u_0 \otimes m$. Hence $\mathrm{Cor}_0(m) = u_0 m$. These deliberations then prove, that $\mathrm{Cor}_0 \circ \mathrm{Res}_0(x) = u_0 x = \varepsilon(u_0)x$, for all $x \in M^L$. Now assume n to be positive and let $(0) \to M \to I \to Q \to (0)$ be an injective presentation of M. The long exact

cohomology sequence along with the naturality of the restriction and corestriction maps give rise to a commutative diagram

$$\begin{array}{ccccc} H_*^{n-1}(L, Q) & \longrightarrow & H_*^n(L, M) & \longrightarrow & (0) \\ \downarrow & & \downarrow & & \\ H_*^{n-1}(L, Q) & \longrightarrow & H_*^n(L, M) & \longrightarrow & (0). \end{array}$$

with exact rows. The inductive hypothesis now yields the desired result. □

Examples. (1) Suppose that $L = Fe \oplus L_0$, where L_0 is a p-ideal of L and $e^{[p]} \in L_0$. It readily follows that $v(e^\tau e^a) = \delta(a, 0)\ 0 \leqq a \leqq \tau$, whence $u_0 = e^\tau$. As a result $\mathrm{Cor}_n \circ \mathrm{Res}_n = 0$.

(2) Let L_0 be a maximal torus of a classical semisimple Lie algebra L. Then $\sigma(L_0) = (0)$ and $H_*^n(L_0, M) = (0)\ \forall n > 0$ for every L_0-module M. Consequently, $\mathrm{Cor}_n \circ \mathrm{Res}_n = 0$.

Theorem 2.7. *Let $L_0 \subset L_1$ be two p-subalgebras of L such that*

(*a*) $\mathrm{tr}(\mathrm{ad}_{L/L_0} x) = 0 \quad \forall x \in L_0$

(*b*) $\mathrm{tr}(\mathrm{ad}_{L_1/L_0} x) = 0 \quad \forall x \in L_0$

(*c*) $\mathrm{tr}(\mathrm{ad}_{L/L_1} x) = 0 \quad \forall x \in L_1$.

Given a restricted L-module M, let $\mathrm{Cor}_n^{(0)}: H_*^n(L_0, M) \to H_*^n(L_1, M)$; $\mathrm{Cor}_n^{(1)}: H_*^n(L_1, M) \to H_*^n(L, M)$ and $\mathrm{Cor}_n: H_*^n(L_0, M) \to H_*^n(L, M)$ the corresponding corestriction maps. Then $\mathrm{Cor}_n = \mathrm{Cor}_n^{(1)} \circ \mathrm{Cor}_n^{(0)}\ \forall n \geqq 0$.

Proof. Let $\{e_1, \ldots, e_t\}$ and $\{e_{t+1}, \ldots, e_k\}$ be cobases of L_0 in L_1 and L_1 in L, respectively. There result decompositions

$$u(L) = \bigoplus_{0 \leqq a'' \leqq \tau''} u(L_1) e^{a''} = \bigoplus_{0 \leqq a'' \leqq \tau''} \bigoplus_{0 \leqq a' \leqq \tau'} u(L_0) e^{(a', a'')},$$

where $a'' \in \mathbb{N}_0^{k-t}$ and $a' \in \mathbb{N}_0^t$. If $v: u(L) \to u(L_0)$, $v_0: u(L_1) \to u(L_0)$ and $v_1: u(L) \to u(L_1)$ denote the corresponding projections, then the above decomposition readily yields $v = v_0 \circ v_1$. We let $u_0 \in u(L_1)$ and $u_1 \in u(L)$ denote the elements satisfying $v_0(e^{a'} u_0) = \delta(a', 0)\ 0 \leqq a' \leqq \tau'$ and $v_1(e^{a''} u_1) = \delta(a'', 0)\ 0 \leqq a'' \leqq \tau''$, respectively. For $0 \leqq (a', a'') \leqq \tau := (\tau', \tau'')$ we obtain

$$v(e^{(a', a'')} u_1 u_0) = v_0(e^{a'} v_1(e^{a''} u_1) u_0) = \delta(a'', 0) v_0(e^{a'} u_0) = \delta((a', a''), 0).$$

We now prove the asserted identity by induction on n. For $n = 0$ and $m \in M^{L_0}$ the proof of (2.6) in conjunction with the above entails

$$\mathrm{Cor}_0(m) = u_1 u_0 m = (\mathrm{Cor}_0^{(1)} \circ \mathrm{Cor}_0^{(0)})(m).$$

Now let n be positive and consider an injective presentation

$$(0) \to M \to I \to Q \to (0)$$

of M. Since $u(L)$, $u(L_0)$ and $u(L_1)$ are, owing to [1], Frobenius algebras we may apply [7, (5.3)] in order to see that I is injective over $u(L_0)$ and $u(L_1)$. There results

a commutative diagram

$$\begin{array}{ccccc} H_*^{n-1}(L_0, Q) & \longrightarrow & H_*^n(L_0, M) & \longrightarrow & (0) \\ \downarrow & & \downarrow & & \\ H_*^{n-1}(L_1, Q) & \longrightarrow & H_*^n(L_1, M) & \longrightarrow & (0) \\ \downarrow & & \downarrow & & \\ H_*^{n-1}(L, Q) & \longrightarrow & H_*^n(L, M) & \longrightarrow & (0) \end{array}$$

whose rows are exact and whose vertical arrows are given by the relevant corestriction mappings. The inductive hypothesis, applied to Q, now yields the desired result. □

Corollary 2.8. *Suppose that F is algebraically closed and let L_0 be a p-ideal of L. Then the following statements are equivalent:*

(1) *L/L_0 is a torus*

(2) *$\mathrm{Cor}_n \circ \mathrm{Res}_n$ is invertible for every restricted L-module M and $n \geqq 0$. In either case, we have $\mathrm{Cor}_n \circ \mathrm{Res}_n = (-1)^k \,\mathrm{id}_{H_*^n(L, M)}\, \forall\, n \geqq 0$, where $k = \dim_F L/L_0$.*

Proof. (1) $\Rightarrow$ (2): We proceed by induction on k. Owing to [27, (3.6), p. 82] L_0 possesses a cobasis $\{e_1, \ldots, e_k\}$ such that $e_i^{[p]} \equiv e_i \bmod(L_0)$. If $k = 1$, we obtain for $e := e_1$ the identities

$$v(e^{p-1+i}) = v(e^i) + (e^p - e)v(e^{i-1}) = \delta(i, p-1) \qquad 1 \leqq i \leqq p-1\,,$$

whence $u_0 = e^{p-1} - 1$. Thus $\varepsilon(u_0) = -1$ and (2.7) yields the desired result. In the general case, we consider the p-ideal $L_1 := \bigoplus_{i=1}^{k-1} Fe_i \oplus L_0$. Note that L/L_1 and L_0/L_1 are tori. Since $L_0 \subset L_1$ satisfy the requirements of (2.7) the inductive hypothesis yields

$$\begin{aligned} \mathrm{Cor}_n \circ \mathrm{Res}_n &= \mathrm{Cor}_n^{(1)} \circ \mathrm{Cor}_n^{(0)} \circ \mathrm{Res}_n = \mathrm{Cor}_n^{(1)} \circ \mathrm{Cor}_n^{(0)} \circ \mathrm{Res}_n^{(0)} \circ \mathrm{Res}_n^{(1)} \\ &= (-1)^{k-1}\, \mathrm{Cor}_n^{(1)} \circ \mathrm{Res}_n^{(1)} = (-1)^k\, \mathrm{id}_{H_*^n(L, M)}\,, \end{aligned}$$

for every restricted L-module M.

(2) $\Rightarrow$ (1): Let $e \in L - L_0$ be given such that $e^{[p]} \in L_0$. We consider the p-subalgebra $L_1 := Fe \oplus L_0$ and observe that $\mathrm{tr}(\mathrm{ad}_{L/L_1} x) = 0\ \forall\, x \in L_1$. Consequently, (2.7) may be applied to the extension $L_0 \subset L_1 \subset L$. Let M be a restricted L-module. According to example (1) the composition $\mathrm{Cor}_n^{(0)} \circ \mathrm{Res}_n^{(0)}$ vanishes. By virtue of (2.7) the same applies to $\mathrm{Cor}_n \circ \mathrm{Res}_n$ and our present assumption forces $H_*^n(L, M) = (0)$ $\forall\, n \geqq 0$. The choice $M = F$ and $n = 0$ confutes this conclusion. As a result, the p-mapping is nonsingular on L/L_0 and the assertion now follows from a consecutive application of [5] and [27, (3.7), p. 83]. □

Remark. The foregoing result applies in particular to Borel subalgebras of classical semisimple Lie algebras. For these algebras the restriction mappings to the nilpotent radical L_0 are always injective.

3. Low dimensional cohomology groups and applications

By selecting a few illustrative examples we will demonstrate how our methods can be utilized in order to effectively compute cohomology groups of modular Lie algebras.

Theorem 3.1. *Suppose that F is an algebraically closed field of characteristic $p > 2$. Let $(L, [p])$ be a finite dimensional solvable restricted Lie algebra, V an irreducible restricted L-module. Then there exists a p-subalgebra $P \subset L$ and a linear form $\lambda \in P^*$ such that*

(1) $H^n(L, V) \cong \bigoplus_{p+q=n} \Lambda^p(L/P) \otimes_F H^q(P, F_{\lambda-\sigma})$

(2) $H^n_*(L, V) \cong H^n_*(P, F_{\lambda-\sigma}) \qquad \forall n \geqq 0.$

Proof. According to [26, Satz 3], there exists a p-subalgebra $P \subset L$, and a one-dimensional P-module F_λ such that $V \cong u(L) \otimes_{u(P)} F_\lambda$. Statement (1) now follows by applying (1.5) and (2.1) successively. Using (2.4) instead of (2.1) in the preceding argument we readily obtain (2). □

Let $(L, [p])$ be a classical semisimple Lie algebra with Borel subalgebra $\mathscr{B}$. For any restricted homomorphism $\lambda\colon \mathscr{B} \to F$ we let $V(\lambda) := u(L) \otimes_{u(\mathscr{B})} F_\lambda$ denote the restricted Verma module of L with highest weight λ. It is well known that any irreducible restricted L-module is isomorphic to the unique irreducible quotient of some $V(\lambda)$. The methods employed in the verification of the foregoing result now yield proofs for the first two statements of the ensuing result.

Theorem 3.2. *Let $(L, [p])$ be a finite dimensional classical semisimple restricted Lie algebra. Then the following statements hold:*

(1) $H^n(L, V(\lambda)) \cong \bigoplus_{p+q=n} \Lambda^p(L/\mathscr{B}) \otimes_F H^q(\mathscr{B}, F_{\lambda-\sigma})$

(2) $H^n_*(L, V(\lambda)) \cong H^n_*(\mathscr{B}, F_{\lambda-\sigma}) \qquad \forall n \geqq 0.$

(3) $H^2_*(L, V(\sigma)) = (0)$.

Proof. We only need to verify (3). Since $\mathscr{B}$ is non-abelian our assertion follows immediately from (2) and (2.5). □

The preceding deliberations reduce questions regarding the cohomology groups to the study of $H^n(P, F_\lambda)$ for suitable subalgebras P and $\lambda \in P^*$. We derive some further potentially useful propositions for an attack on the latter problem. The cohomology groups of the Borel subalgebras can be computed by means of the Hochschild-Serre spectral sequence.

Proposition 3.3. *Let $(L, [p])$ and $\mathscr{B}$ be as in (3.2) and assume that $T \subset \mathscr{B}$ is a maximal torus of L. Then:*

(1) *If $\mathscr{B} = T \oplus N$, $N \subset \mathscr{B}$ the nilradical of $\mathscr{B}$, then the Hochschild-Serre spectral sequence converging to $H^*(\mathscr{B}, F_\lambda)$ has initial term*

$$E_2^{p,q} \cong \Lambda^p(T) \otimes_F H^q(N, F)_{-\lambda}.$$

(2) *If $\lambda | T$ is not a sum of positive roots, then $H^n(\mathscr{B}, F_\lambda) = (0)$ $\forall n \geqq 0$.*

Proof. The Hochschild-Serre spectral sequence has initial term

$$E_2^{p,q} \cong H^p(T, H^q(N, F_\lambda)).$$

Since N acts trivially on F_λ, the T-modules $H^q(N, F_\lambda)$ and $H^q(N, F) \otimes_F F_\lambda$ are canonically isomorphic. According to [13, (6.3)] the cohomology groups of T with coefficients in $H^q(N, F) \otimes_F F_\lambda$ coincide with those of its Fitting-0-component $(H^q(N, F) \otimes_F F_\lambda)^T$. Consequently,

$$E_2^{p,q} \cong H^p(T, (H^q(N, F) \otimes_F F_\lambda)^T) \cong \Lambda^p(T) \otimes_F (H^q(N, F) \otimes_F F_\lambda)^T,$$

the second factor in the tensor product being obviously isomorphic to $H^q(N, F)_{-\lambda}$.

(2) If λ is not a sum of positive roots, then $H^q(N, F)_{-\lambda} = 0 \; \forall q \geqq 0$. Thus the spectral sequence collapses and the vanishing of the cohomology groups $H^n(\mathscr{B}, F_\lambda)$ follows from [2, (5.6), p. 326]. □

We finally compute low-dimensional cohomology groups with coefficients in a one-dimensional module for arbitrary Lie algebras. Let λ: $P \to F$ be a nonzero homomorphism of Lie algebras. Let $I := \ker\lambda$. Put $P_\lambda^{(1)} := \{[x, y] - \lambda(x)y + \lambda(y)x | x, y \in P\}$. Observe that $P_\lambda^{(1)} \subset I$.

Proposition 3.4. $H^1(P, F_\lambda) \cong \mathrm{Hom}_F(I/P_\lambda^{(1)}, F_\lambda) \cong I/P_\lambda^{(1)}$.

Proof. A straightforward computation shows, that there is an isomorphism γ: $\mathrm{Der}_F(P, F_\lambda) \to \mathrm{Hom}_F(P/P_\lambda^{(1)}, F_\lambda)$ that maps the space of inner derivations onto the image of $\mathrm{Hom}_F(P/I, F_\lambda) \to \mathrm{Hom}_F(P/P_\lambda^{(1)}, F_\lambda)$. The desired isomorphism can now be obtained from the exact sequence $(0) \to \mathrm{Hom}_F(P/I, F_\lambda) \to \mathrm{Hom}_F(P/P_\lambda^{(1)}, F_\lambda) \to \mathrm{Hom}_F(I/P_\lambda^{(1)}, F_\lambda) \to (0)$. □

Let V be an L-module with representation $\rho: L \to \mathrm{gl}(V)$ and annihilator $I := \ker\rho$. For $n \geqq 1$ we consider the mappings Γ_n: $C^n(L, V) \to C^{n-1}(L, \mathrm{Hom}_F(I, V))$ which are given by

$$\Gamma_n(f)(x_1, \ldots, x_{n-1})(x_n) = f(x_1, \ldots, x_n); x_i \in L \qquad 1 \leqq i \leqq n-1, x_n \in I \,.$$

Put $\Gamma_0 := 0$ and $\Gamma := \oplus_{n \geqq 0} \Gamma_n$.

Proposition 3.5. *The following statements hold*:

(1) *Γ is a mapping of complexes and hence induces mappings*

$$\Omega_n: H^n(L, V) \to H^{n-1}(L, \mathrm{Hom}_F(I, V)) \qquad \forall n \geqq 0 \,.$$

(2) *If* $\mathrm{codim}_F I \leqq 1$, *then Ω_2 is injective.*

(3) *If* $\mathrm{codim}_F I \leqq 1$, *then*

$$\mathrm{im}\,\Omega_2 = \{[\varphi]; \varphi \in Z^1(L, \mathrm{Hom}_F(I, V)), \varphi(x)(x) = 0 \; \forall x \in I\} \,.$$

Proof. (1) Let δ and ∂ denote the coboundary operators of $C(L, V)$ and $C(L, \mathrm{Hom}_F(I, V))$, respectively. Given $f \in C^n(L, V)$ we obtain for $x_0, \ldots, x_{n-1} \in L$, $x_n \in I$, observing $x_n \cdot V = (0)$,

$$\begin{aligned}
&\Gamma_{n+1}(\delta(f))(x_0, \ldots, x_{n-1})(x_n) = \delta(f)(x_0, \ldots, x_n) \\
&= \sum_{i=0}^{n-1} (-1)^i x_i \cdot f(x_0, \ldots, \hat{x}_i, \ldots, x_n) \\
&\quad + \sum_{0 \leqq p < q \leqq n-1} (-1)^{p+q} f([x_p, x_q], x_0, \ldots, \hat{x}_p, \ldots, \hat{x}_q, \ldots, x_n) \\
&\quad + \sum_{i=0}^{n-1} (-1)^{i+n} f([x_i, x_n], x_0, \ldots, \hat{x}_i, \ldots, x_{n-1}) \\
&= \sum_{0 \leqq p < q \leqq n-1} (-1)^{p+q} \Gamma_n(f)([x_p, x_q], x_0, \ldots, \hat{x}_p, \ldots, \hat{x}_q, \ldots, x_{n-1})(x_n) + \\
&\quad \sum_{i=0}^{n-1} (-1)^i x_i \cdot (\Gamma_n(f)(x_0, \ldots, \hat{x}_i, \ldots, x_{n-1})(x_n)) \\
&\quad - \sum_{i=0}^{n-1} (-1)^i f(x_0, \ldots, \hat{x}_i, \ldots, x_{n-1}, [x_i, x_n])
\end{aligned}$$

$$= \sum_{0 \leq p < q \leq n-1} (-1)^{p+q} \Gamma_n(f)([x_p, x_q], x_0, \ldots, \hat{x}_p, \ldots, \hat{x}_q, \ldots, x_{n-1})(x_n) +$$

$$\sum_{i=0}^{n-1} (-1)^i (x_i \cdot \Gamma_n(f)(x_0, \ldots, \hat{x}_i, \ldots, x_{n-1}))(x_n)$$

$$= \partial(\Gamma_n(f))(x_0, \ldots, x_{n-1})(x_n) .$$

(2) Let f be an element of $Z^2(L, V)$ such that $\Gamma_2(f)$ is inner. Then there is $g \in \mathrm{Hom}_F(I, V)$ such that $f(x, y) = (x \cdot g)(y) = x \cdot g(y) - g([x, y]) = -g([x, y]) + x \cdot g(y) - y \cdot g(x)\ \forall x \in L\ \forall y \in I$. If $I = L$ we obtain $f = \delta(g)$ and the cohomology class of f vanishes. Otherwise we write $L = I \oplus Fx_0$ and extend g to L by setting $g(x_0) = 0$. Then $f(x, y) = \delta(g)(x, y)\ \forall x \in L\ \forall y \in I$. Since both maps are alternating it remains to consider the case $x \in I$, $y = x_0$:

$$f(x, x_0) = -f(x_0, x) = -\delta(g)(x_0, x) = \delta(g)(x, x_0) .$$

Again, the cohomology class of f is zero, proving the injectivity of Ω_2.

(3) The result is well-known in case of $I = L$ (cf. [10, (1.3)]). We therefore assume that $L = I \oplus Fx_0$. One inclusion follows directly from the definitions. Let $\psi \in Z^1(L, \mathrm{Hom}_F(I, V))$ be a derivation satisfying $\psi(x)(x) = 0\ \forall x \in I$. Define $\varphi \in C^2(L, V)$ by means of

$$\varphi(x + \alpha x_0, y + \beta x_0) := \psi(x)(y) + \alpha\psi(x_0)(y) - \beta\psi(x_0)(x) \qquad \text{for } x, y \in I .$$

Let $a, b, c \in L$. As $\delta(\varphi)$, given by

$$\delta(\varphi)(a, b, c) = -\varphi([a, b], c) + \varphi([a, c], b) - \varphi([b, c], a) + a \cdot \varphi(b, c) - b \cdot \varphi(a, c) + c \cdot \varphi(a, b)$$

is alternating we only have to consider the ensuing cases:

(a) $a, b, c \in I$

In this case the last three terms vanish and we obtain, observing the fact that ψ is a derivation,

$$\begin{aligned} \delta(\varphi)(a, b, c) &= -\psi([a, b])(c) + \psi([a, c])(b) - \psi([b, c])(a) \\ &= -(a \cdot \psi(b))(c) + (b \cdot \psi(a))(c) + \psi([a, c])(b) - \psi([b, c])(a) \\ &= \psi(b)([a, c]) - \psi(a)([b, c]) + \psi([a, c])(b) - \psi([b, c])(a) = 0 . \end{aligned}$$

(b) $a, b \in I$, $c = x_0$.

We obtain

$$\begin{aligned} \delta(\varphi)(a, b, x_0) &= -\varphi([a, b], x_0) + \varphi([a, x_0], b) - \varphi([b, x_0], a) + x_0 \cdot \varphi(a, b) \\ &= \psi(x_0)([a, b]) + \psi([a, x_0])(b) - \psi([b, x_0])(a) + x_0 \cdot \psi(a)(b) \\ &= \psi(x_0)([a, b]) + (a \cdot \psi(x_0))(b) - (x_0 \cdot \psi(a))(b) - (b \cdot \psi(x_0))(a) + \\ &\quad (x_0 \cdot \psi(b))(a) + x_0 \cdot \psi(a)(b) \\ &= \psi(x_0)([a, b]) - \psi(x_0)([a, b]) - x_0 \cdot \psi(a)(b) + \psi(a)([x_0, b]) + \\ &\quad \psi(x_0)([b, a]) + x_0 \cdot \psi(b)(a) - \psi(b)([x_0, a]) + x_0 \cdot \psi(a)(b) \end{aligned}$$

$$= \psi(a)([x_0, b]) + \psi(x_0)([b, a]) + x_0 \cdot \psi(b)(a) - \psi(b)([x_0, a])$$
$$= \psi([b, x_0])(a) - (b \cdot \psi(x_0))(a) + (x_0 \cdot \psi(b))(a) = 0 .$$

As a result, φ is a derivation whose cohomology class is evidently a preimage of the class defined by ψ. □

We summarize some of our results for $n \leqq 2$.

Theorem 3.6. *Let $(L, [p])$ be a restricted Lie algebra, P a p-subalgebra $\lambda: P \to F$ a homomorphism of restricted Lie algebras, $I := \ker(\lambda - \sigma)$, and $V = u(L) \otimes_{u(P)} F_\lambda$ an induced L-module. Then*

1) *If $\lambda \neq \sigma$:*
$$H^0(L, V) = (0)$$
$$H^1(L, V) \cong I/P^{(1)}_{\lambda-\sigma}$$
$$H^2(L, V) \cong H^2(P, F_{\lambda-\sigma}) \oplus (L/P) \otimes_F I/P^{(1)}_{\lambda-\sigma}$$
2) *If $\lambda = \sigma$:*
$$H^0(L, V) \cong F$$
$$H^1(L, V) \cong L/P^{(1)}$$
$$H^2(L, V) \cong H^2(P, F) \oplus (L/P) \otimes_F (P/P^{(1)}) \oplus \Lambda^2(L/P).$$ □

Acknowledgement: The authors would like to thank G.P. Hochschild for commenting on an earlier version of this paper.

References

1. Berkson, A.J.: The u-algebra of a restricted Lie algebra is Frobenius. Proc. Am. Math. Soc. **15**, 14–15 (1964)
2. Cartan, H., Eilenberg, S.: Homological algebra. Princeton: Princeton University Press 1956
3. Chang, H.J.: Über Wittsche Lie Ringe. Abh. Math. Semin. Univ. Hamb. **14**, 151–184 (1941)
4. Chevalley, C., Eilenberg, S.: Cohomology theory of Lie groups and Lie algebras. Trans. Am. Math. Soc. **63**, 85–124 (1948)
5. Chwe, B.S.: On the commutativity of restricted Lie algebras. Proc. Am. Math. Soc. **16**, 547 (1965)
6. Dzhumadil'daev, A.S.: Cohomology of modular Lie algebras. Math. USSR, Sb. **47**, 127–143 (1984)
7. Faith, C., Walker, E.A.: Direct sum decompositions of injective modules. J. Algebra **5**, 203–221 (1967)
8. Farnsteiner, R.: On ad-semisimple Lie algebras. J. Algebra **83**, 510–519 (1983)
9. Farnsteiner, R.: Conditions for the commutativity of restricted Lie algebras. Commun. Algebra **13**, 1475–1489 (1985)
10. Farnsteiner, R.: Central extensions and invariant forms of graded Lie algebras. Algebras Groups Geom. **3**, 431–455 (1986)
11. Farnsteiner, R.: Dual space derivations and $H^2(L, F)$ of modular Lie algebras. Can. J. Math. **39**, 1078–1106 (1987)
12. Farnsteiner, R.: On the cohomology of associative algebras and Lie algebras. Proc. Am. Math. Soc. **99**, 415–420 (1987)
13. Farnsteiner, R.: On the vanishing of homology and cohomology groups of associative algebras. Trans. Am. Math. Soc. **306**, 651–665 (1988)
14. Farnsteiner, R.: Cohomology groups of infinite dimensional algebras. Math. Z. **199**, 407–423 (1988)
15. Friedlander, E.M., Parshall, B.J.: Geometry of p-unipotent Lie algebras. J. Algebra **109**, 25–45 (1987)
16. Friedlander, E.M., Parshall, B.J.: Support varieties for restricted Lie algebras. Invent. Math. **86**, 553–562 (1986)

17. Hilton, P.J., Stammbach, U.: A course in homological algebra. (Graduate Texts, vol. 4) Berlin Heidelberg New York: Springer 1970
18. Hochschild, G.P., Serre, J.P.: Cohomology of Lie algebras. Ann. Math. **57**, 591–603 (1953)
19. Hochschild, G.P.: On the cohomology groups of an associative algebra. Ann. Math. **46**, 58–67 (1945)
20. Hochschild, G.P.: Cohomology of restricted Lie algebras. Am. J. Math. **76**, 555–580 (1954)
21. Humphreys, J.E.: Restricted Lie algebras (and beyond). Contemp. Math. **13**, 91–98 (1982)
22. Pareigis, B.: Einige Bemerkungen über Frobenius-Erweiterungen. Math. Ann. **153**, 1–13 (1964)
23. Seligman, G.B.: Modular Lie algebras. (Ergeb. Math. Grenzgeb. vol. 40), Berlin Heidelberg New York: Springer 1967
24. Sen, C., Shen, G.: Cohomology of graded Lie algebras of Cartan type of characteristic p. Abh. Math. Semin. Univ. Hamb. **57**, 139–156 (1987)
25. Strade, H.: Representations of the Witt algebra. J. Algebra **49**, 595–605 (1977)
26. Strade, H.: Darstellungen Auflösbarer Lie Algebren. Math. Ann. **232**, 15–32 (1978)
27. Strade, H., Farnsteiner, R.: Modular Lie algebras and their representations. (Textbooks and Monographs, vol. 116) New York: Dekker 1988

Math. Z. 206, 169–192 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

Quasi-invariant measures on commutative Lie groups of infinite product type

Akihito Hora *
Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606, Japan

Received March 3, 1989; in final form January 24, 1990

Contents

Introduction and main results 169
1 Preliminary results 175
2 Fourier transforms 177
3 Quasi-invariance of power probabilities 178
4 Quasi-invariance of exchangeable probabilities 184
5 Quasi-invariant diagonal transformations on sequence spaces 188

Introduction and main results

Two measures μ and ν are said to be equivalent (denoted by $\mu \sim \nu$) if the μ-null sets coincide with the ν-null sets. The first significant result in this direction was obtained by Kakutani [K]. He established the famous criterion for the equivalence of product measures which is now known as Kakutani's dichotomy theorem. This result inspired many mathematicians and has been generalized and extended by several authors; for example, Ritter [R], Janssen [J], Sato [Sa] etc.

In the study of functional or stochastic analysis, we often need to find out whether the measure transformed by a certain mapping is equivalent to the original one. If the two are equivalent, we call this mapping a quasi-invariant transformation of the measure. In particular, the quasi-invariance with respect to translations and rotations has been investigated in great detail.

In general, Sudakov showed that there exist no measures which are quasi-invariant with respect to all the translations in infinite dimensional vector spaces and ever since systematical studies on the subset of quasi-invariant translations of measures have been done. We refer to Shepp [S], Umemura (=Yamasaki) [U], Xia [X], Shimomura [Sh 1] and Yamasaki [Y 1, Y 2]. Above all, the books [X], [Y 1] and [Y 2] are good references in this field.

Here we recall some results about product measures on $\mathbf{R}^\infty$. Shepp characterized the l_2-quasi-invariance of these measures ([S]). Shimomura and Yamasaki studied the structure of the quasi-invariant translations of these measures and

* *Current adress*: Department of Mathematics, Kyushu University 33, Fukuoka 812, Japan

obtained many results about l_p-quasi-invariance with interesting concrete examples ([Sh 1], [Y 1], [Y 2]).

In this paper we extend the above theory due to Shepp, Shimomura and Yamasaki in the setting of commutative Lie groups. Since compact subgroups appear in contrast with the case of vector spaces, the situation becomes somewhat different. We deal with the quasi-invariance with respect to "modified l_p-spaces" of power probabilities and exchangeable probabilities. Let S be a set with a measurable structure. We call a probability μ on $S^{\mathbf{N}}$ a power probability (or a stationary product measure) if μ is a direct product of identical probabilities on S and μ an exchangeable probability if μ is invariant with respect to the actions of all the finite permutations of $\mathbf{N}$ to $S^{\mathbf{N}}$. As for this terminology and their properties, we refer to Dubins and Freedman [D-F]. Finally, we will apply the obtained results to quasi-invariance of measures on $\mathbf{R}^{\mathbf{N}}$ and $\mathbf{C}^{\mathbf{N}}$ with respect to linear transformations by diagonal matrices.

Our methods used in this paper are based on the convergence of products of independent random variables and on the Fourier transforms on finitely generated commutative Lie groups. If μ is a power probability on $G^{\mathbf{N}}$, countable product of an abelian group G, of the form: $\mu=\lambda^{\infty}$, $d\lambda=\psi(g)^2\,dg$, $\psi(g)>0$ (a.e.) and $(a_k)_{k\in\mathbf{N}}$ is an element of $G^{\mathbf{N}}$, then it follows by the well-known discussions (see §1) that (a_k)-quasi-invariance of μ is characterized by the $L_2(\mu)$-convergence of $\prod_{k=1}^{\infty}\psi(g_k-a_k)/\psi(g_k)$. In order to estimate the convergence of this random variable, we carry the G-translations into multiplication by characters of G using the Fourier transforms. We consider the probability defined as $d\tilde{\lambda}(\gamma)=|\hat{\psi}(\gamma)|^2\,d\gamma$ on $\hat{G}$ by the Fourier transform of ψ and call $\tilde{\lambda}$ the adjoint measure of λ. It will turn out that the moment properties of the adjoint measure $\tilde{\lambda}$ are crucial to determine the subgroup consisting of all the quasi-invariant translations as in the case of vector spaces.

Now we present our main results in detail and explain the organization of this paper. Let G be a finitely generated commutative Lie group, namely, a finite direct product of $\mathbf{R}$ (the real numbers), $\mathbf{T}(=\mathbf{R}/2\pi\mathbf{Z})$, $\mathbf{Z}$ (the integers) and $\mathbf{Z}_p$'s (cyclic groups of oder p, p: power of a prime number). If G has the representation: $G=\mathbf{R}^n\times\mathbf{T}^m\times\mathbf{Z}^l\times\prod_i\mathbf{Z}_{p_i}$, then $R=\mathbf{R}^n$, $T=\mathbf{T}^m$, $Z=\mathbf{Z}^l$ and $F=\prod_i\mathbf{Z}_{p_i}$ are called the real part, the torus part, the integral part and the finite part of G respectively. Of course we regard each part as a subgroup of G. We denote by $G^{\mathbf{N}}$ the countable product group of G with the standard Borel measurable structure. Our concerns throughout this paper are the quasi-invariance of measures on $G^{\mathbf{N}}$. In the sequel we deal only with probability measures. This restriction, however, is not essential because we can immediately interpret the obtained results in the case of σ-finite measures. Let $\tau_a\mu$ denote the translated measure of μ by a; i.e. $\tau_a\mu(B)=\mu(B-a)$ for each Borel set B. For a probability μ on $G^{\mathbf{N}}$, we put

$$Y_\mu=\{a\in G^{\mathbf{N}};\ \tau_a\mu\sim\mu\} \tag{0.1}$$

where '$\sim$' means the equivalence of two probabilities. And for a probability λ on G, we put

$$I_\lambda=\{g\in G;\ \tau_g\lambda=\lambda\}. \tag{0.2}$$

Note that I_λ is closed and must be contained in the compact part K of G since λ is a finite measure. Y_μ and I_λ are called the subgroup of the quasi-invariant translations of μ and the subgroup of the invariant translations of λ respectively.

We define a distance d on G as the sum of the natural distances on $\mathbf{R}$, $\mathbf{T}$, $\mathbf{Z}$ and $\mathbf{Z}_p$. The exact value of the distance is not essential in our discussions. We can replace the distance by another equivalent one. For a subgroup H of K and a positive number p, we put

$$|g|_H = \inf_{h\in H} d(g,h), \qquad g\in G, \tag{0.3}$$

$$l_p(G,H) = \left\{ a=(a_k)_{k\in\mathbf{N}}\in G^{\mathbf{N}};\ \sum_{k=1}^{\infty} |a_k|_H^p < \infty \right\}. \tag{0.4}$$

If a probability μ on $G^{\mathbf{N}}$ is expressed as an infinite product of the identical probability λ on G (denoted by $\mu=\lambda^\infty$), we call μ a *power probability* or a stationary product measure. Our first result is the following

Theorem 1 *For an arbitrary probability λ on G, its power probability $\mu=\lambda^\infty$ on $G^{\mathbf{N}}$ satisfies $Y_\mu \subset l_2(G, I_\lambda)$.*

The proof of this theorem will be given in §3.

Next, we consider the converse inclusion: $l_p(G,H)\subset Y_\mu$ where $p\leqq 2$ (from Theor. 1). As a necessary condition for this inclusion, Y_μ must contain $G_0^{\mathbf{N}}(=\{a=(a_k)\in G^{\mathbf{N}};\ a_k=0$ except a finite number of k's$\})$, which holds if and only if λ is equivalent to the Haar measure on G. Then we put $d\lambda(g)=\psi(g)^2\,dg\,(\psi>0$ a.e.), namely ψ^2 is the Radon-Nikodym derivative of λ. Let $\hat{G}$ denote the character group of G. We naturally identify $(\mathbf{R}^n\times\mathbf{T}^m\times\mathbf{Z}^l\times\prod_i \mathbf{Z}_{p_i})^\wedge$ with $\mathbf{R}^n\times\mathbf{Z}^m\times\mathbf{T}^l\times\prod_i \mathbf{Z}_{p_i}$. We denote the pairing of $g\in G$ and $\gamma\in\hat{G}$ by $\langle g,\gamma\rangle$. For example, $\langle x,\xi\rangle = e^{ix\xi}(x\in\mathbf{R},\ \xi\in\mathbf{R})$, $\langle t,n\rangle = e^{itn}(t\in\mathbf{T},\ n\in\mathbf{Z})$, $\langle k,\kappa\rangle = e^{i\frac{2\pi}{p}k\kappa}(k\in\mathbf{Z}_p,\ \kappa\in\mathbf{Z}_p)$ etc. and $\langle (g_1,g_2),(\gamma_1,\gamma_2)\rangle = \langle g_1,\gamma_1\rangle\langle g_2,\gamma_2\rangle$. This is a notation appearing in [P], Chapter IV. Let $\hat{f}$ denote the Fourier transform of f:

$$\hat{f}(\gamma) = \int_G \langle g,\gamma\rangle f(g)\,dg \tag{0.5}$$

where $f\in L_2(G)$. We consider the following moment conditions for finite measures (or L_1-functions) on $\hat{G}$:

$$[\mathcal{M}_p, \nu]:\ \int_{\hat{G}} |\gamma|^p\,d\nu < \infty \tag{0.6}$$

where $0<p<\infty$, ν is a finite measure on $\hat{G}$ and $|\cdot|$ denotes a natural distance from 0 on $\hat{G}$. If ν has the Radon-Nikodym derivative ρ with respect to the Haar measure on $\hat{G}$, we often write $[\mathcal{M}_p, \rho]$ instead of $[\mathcal{M}_p, \nu]$. Our second result is the following

Theorem 2 *Let μ be a power probability on $G^{\mathbf{N}}$ satisfying $\mu=\lambda^\infty$, $d\lambda(g)=\psi(g)^2\,dg$ and $\psi>0$ a.e. We put*

$$\phi(g) = \int_{\hat{G}} \mathrm{Re}\langle g,\gamma\rangle |\hat{\psi}(\gamma)|^2\,d\gamma. \tag{0.7}$$

(1) *Let* $0<p\leqq 2$. *Then we have*

$$l_p(G, I_\lambda) \underset{\text{(resp.}=)}{\subset} Y_\mu \Leftrightarrow 1-\phi(g) \underset{\text{(resp.}\sim)}{\leqq} C|g|_{I_\lambda}^p$$

where C *is a constant independent of* g.

(2) $l_2(G, I_\lambda)=Y_\mu \Leftrightarrow [\mathscr{M}_2, |\hat{\psi}|^2] \Leftrightarrow \dfrac{\partial \psi}{\partial x_i}, \dfrac{\partial \psi}{\partial t_j} \in L_2(G)$ *where* $\{x_i, t_j\}$ *denotes a system of local coordinates of the continuous part* $R\times T$ *of* G *and* $\dfrac{\partial}{\partial x_i}, \dfrac{\partial}{\partial t_j}$ *are in distribution sense.*

(3) *Let* $0<p<2$.

$$[\mathscr{M}_p, |\hat{\psi}|^2] \Rightarrow l_p(G, I_\lambda) \subsetneqq Y_\mu.$$

$$l_p(G, I_\lambda) \subset Y_\mu \Rightarrow [\mathscr{M}_q, |\hat{\psi}|^2] \quad \text{for } 0<\forall q<p.$$

For $a=(a_k)_{k\in\mathbf{N}} \in Y_\mu$, *the density of* $\tau_a \mu$ *with respect to* μ *is given as*

$$\frac{d\tau_a \mu}{d\mu}(g_1, g_2, \ldots) = \prod_{k=1}^{\infty} \frac{\psi(g_k-a_k)^2}{\psi(g_k)^2}. \tag{0.8}$$

This theorem will be proved in §3 with Theorem 1.

Now we consider exchangeable probabilities on $G^{\mathbf{N}}$, which are substantial generalizations of power probabilities. Let $\mathscr{S}_\infty$ denote the finite permutations of $\mathbf{N}$. $\sigma\in\mathscr{S}_\infty$ acts on $G^{\mathbf{N}}$ as $\sigma(g_1, g_2, \ldots)=(g_{\sigma(1)}, g_{\sigma(2)}, \ldots)$. An *exchangeable probability* on $G^{\mathbf{N}}$ is by definition an $\mathscr{S}_\infty$-invariant probability (i.e. $\sigma\circ\mu=\mu$ for $\forall\sigma\in\mathscr{S}_\infty$). Since G is a standard space, we can represent an exchangeable probability on $G^{\mathbf{N}}$ as a superposition of power probabilities (see §4). The possibility of such representations in general cases was studied by Dubins and Freedman [D-F]. Let μ be an exchangeable probability on $G^{\mathbf{N}}$. Thus μ enjoys the following ergodic decomposition:

$$\mu = \int_\Lambda \theta_\alpha^\infty \, d\nu(\alpha) \tag{0.9}$$

where Λ: parameter space, ν: probability on Λ, θ_α: probability on G parametrized by $\alpha\in\Lambda$, θ_α^∞: power probability on $G^{\mathbf{N}}$ and $\theta_\alpha^\infty \perp \theta_\beta^\infty$ (singular) if $\alpha\neq\beta$. ν is often called a mixing measure.

Theorem 3 *Let* μ *be an exchangeable probability on* $G^{\mathbf{N}}$ *enjoying the expression* (0.9). *Denote by* $\{K_k\}_{k\in\mathbf{N}}$ *the family of the subgroups of the compact part* $T\times F$ *(see the following remark) and by* p *the canonical projection of* $G^{\mathbf{N}}$ *onto the first factor* G. *Put*

$$K_0 = \bigcap \{K_k; \nu(\{\alpha\in\Lambda; I_{\theta_\alpha}=K_k\})>0\} \tag{0.10}$$

and $\Delta=\{(g, g, \ldots)\in G^{\mathbf{N}};\ g\in G\}$ *(i.e. the diagonal subgroup of* $G^{\mathbf{N}}$*). Then we have* $Y_\mu \subset l_2(G, K_0)+\Delta$, *in particular* $Y_\mu \subset l_2(G, I_{p\circ\mu})+\Delta$. *Moreover, if* $a=(a_k)_{k\in\mathbf{N}}\in Y_\mu$, *its corresponding element of* Δ *can be taken as the sequence of a limit point of* $\{a_k\}_{k\in\mathbf{N}}$.

Remark. $T\times F$ contains only a countable number of subgroups. We omit giving the proof since it is elementary and rather tedious.

This theorem will be proved in §4 with the ergodic decomposition (0.9) of μ. We will also note that the intersection in (0.10) is actually taken finite times and that Δ is not negligible.

Lastly, we treat the quasi-invariance of probabilities on $\mathbf{R}^{\mathbf{N}}$ and $\mathbf{C}^{\mathbf{N}}$ with respect to diagonal transformations:

$$[a_n]=\begin{bmatrix} a_1 & & \\ & a_2 & \\ & & \ddots \end{bmatrix}, \quad a_n\in\dot{\mathbf{R}} \quad \text{or} \quad \dot{\mathbf{C}}$$

where $\dot{\mathbf{R}}=\mathbf{R}\setminus\{0\}$ and $\dot{\mathbf{C}}=\mathbf{C}\setminus\{0\}$. Let $\mathbf{R}_+$ denote the multiplicative group of the positive real numbers. We identify $\dot{\mathbf{R}}$ with $\mathbf{R}_+\times\mathbf{Z}_2$ by the correspondence: $x\in\dot{\mathbf{R}}\mapsto(|x|, \operatorname{sgn} x)\in\mathbf{R}_+\times\mathbf{Z}_2$ where $\operatorname{sgn} x=0$ if $x>0,=1$ if $x<0$. Similarly we regard $\mathbf{C}$ as $\mathbf{R}_+\times\mathbf{T}$ through the representation by the polar coordinates. Further $\mathbf{R}_+$ is isomorphic to $\mathbf{R}$ by 'log'. By virtue of these identifications, we can interpret the actions of diagonal transformations onto $\mathbf{R}^{\mathbf{N}}$ and $\mathbf{C}^{\mathbf{N}}$ as the translations on $(\mathbf{R}\times\mathbf{Z}_2)^{\mathbf{N}}$ and $(\mathbf{R}\times\mathbf{T})^{\mathbf{N}}$ respectively. This enables us to apply the above theorems to the estimation of the quasi-invariant diagonal transformations of probabilities on $\dot{\mathbf{R}}^{\mathbf{N}}$ and $\dot{\mathbf{C}}^{\mathbf{N}}$. Excluding the Dirac measure δ_0 supported by $\{0\}$, we can treat the quasi-invariance of probabilities on $\mathbf{R}^{\mathbf{N}}$ and $\mathbf{C}^{\mathbf{N}}$ by this method, that is to say, we may replace $\dot{\mathbf{R}}^{\mathbf{N}}$ and $\dot{\mathbf{C}}^{\mathbf{N}}$ by $\mathbf{R}^{\mathbf{N}}$ and $\mathbf{C}^{\mathbf{N}}$ respectively (see §5). Since δ_0 is invariant with respect to all the linear transformations, we need not take it into account. We put for a probability μ on $\mathbf{R}^{\mathbf{N}}$ (or $\mathbf{C}^{\mathbf{N}}$)

$$L_\mu=\{A=[a_n];\ a_n\in\dot{\mathbf{R}}(\text{or }\dot{\mathbf{C}}),\ A\circ\mu\sim\mu\} \tag{0.11}$$

where $(A\circ\mu)(B)=\mu(A^{-1}B)$ for each Borel set B and

$$l_p^D=\left\{A=[a_n];\ a_n\in\mathbf{R}(\text{or }\mathbf{C}),\ \sum_{n=1}^{\infty}|a_n|^p<\infty\right\}. \tag{0.12}$$

Let us begin with the case of $\mathbf{R}^{\mathbf{N}}$. For a function f on $\mathbf{R}$, we put

$$(M_\pm f)(\xi)=\int_0^\infty r^{i\xi-\frac{1}{2}}f(\pm r)\,dr, \qquad \xi\in\mathbf{R}. \tag{0.13}$$

If f is symmetric, we have $M_+f=M_-f$. $(M_+f)(\xi)(\text{resp.}(M_-f)(\xi))$ is nothing but the value of the Mellin transform of $f(r)$ $(\text{resp.}f(-r))$ at $\frac{1}{2}+i\xi$. It seems to be natural that the Mellin transforms should appear since $\{r^{i\xi};\ \xi\in\mathbf{R}\}$ forms the character group of $\mathbf{R}_+$. We summarize the results for the case of $\mathbf{R}^{\mathbf{N}}$.

Theorem 4 (I) *Let $\mu=\lambda^\infty$ be a power probability on $\mathbf{R}^{\mathbf{N}}$ where $\lambda(\neq\delta_0)$ is a probability on $\mathbf{R}$. We put*

$$E=\{[e_n];\ e_n\in\mathbf{R},\ |e_n|=1\}.$$

and denote the identity matrix on $\mathbf{R}^{\mathbf{N}}$ by Id. The following hold if λ is symmetric (resp. *not symmetric*).

(1) $L_\mu\subset E+l_2^D(\text{resp. } L_\mu\subset Id+l_2^D)$.

(2) *Let λ take the form of $d\lambda(u)=\psi(u)^2\,du(\psi>0.$ a.e., du: the Lebesgue measure on $\mathbf{R}$). We use the notation* (0.6).

(2.a) $[\mathcal{M}_2, |M_+\psi|^2]$ (resp. $|M_\pm\psi|^2$) $\Leftrightarrow E+l_2^D$(resp. $Id+l_2^D$)$=L_\mu$.
(2.b) *Let* $0<p<2$.

$$[\mathcal{M}_p, |M_+\psi|^2](\text{resp.}|M_\pm\psi|^2) \Rightarrow E+l_p^D(\text{resp. } Id+l_p^D)\subsetneqq L_\mu.$$
$$E+l_p^D(\text{resp. } Id+l_p^D)\subset L_\mu \Rightarrow [\mathcal{M}_q, |M_+\psi|^2](\text{resp.}|M_\pm\psi|^2) \quad \textit{for } 0<\forall q<p.$$

For $A=[a_n]\in L_\mu$, *the density of* $A\circ\mu$ *with respect to* μ *is given as*

$$\frac{dA\circ\mu}{d\mu}(u_1,u_2,\ldots)=\prod_{n=1}^{\infty}\frac{1}{|a_n|}\frac{\psi(u_n a_n^{-1})^2}{\psi(u_n)^2}, \quad (u_1,u_2,\ldots)\in\mathbf{R}^{\mathbf{N}}. \tag{0.14}$$

(II) *Let* μ *be an exchangeable probability on* $\mathbf{R}^{\mathbf{N}}$ *and* p *denote the projection of* $\mathbf{R}^{\mathbf{N}}$ *onto the first axis* $\mathbf{R}$. *If* $p\circ\mu$ *is symmetric* (resp. *not symmetric*), *then we have* $L_\mu\subset\mathbf{R}E+l_2^D$ (resp. $L_\mu\subset\mathbf{R}Id+l_2^D$).

Let us turn to the case of $\mathbf{C}^{\mathbf{N}}$. For a function $f(z)(=f(re^{i\theta}))$ on $\mathbf{C}$, we put

$$(Mf)(\xi,m)=\int_{\mathbf{R}_+\times\mathbf{T}} r^{i\xi}e^{im\theta}f(re^{i\theta})\,dr\,d\theta, \quad (\xi,m)\in\mathbf{R}\times\mathbf{Z}. \tag{0.15}$$

For a probability $\lambda(\neq\delta_0)$ on $\mathbf{R}$ or $\mathbf{C}$, we put

$$I_\lambda=\{y\in\mathbf{R}\text{ or }\mathbf{C};\sigma_y\lambda=\lambda\} \tag{0.16}$$

where $\sigma_y\lambda(B)=\lambda(By^{-1})$ for each Borel set B of $\mathbf{R}$ or $\mathbf{C}$. Using the same notation with (0.2) will not confuse us. I_λ is a finite subgroup of the unit circle U of $\mathbf{C}$ unless it coincides with the whole U. Our results for the case of $\mathbf{C}^{\mathbf{N}}$ is summarized as

Theorem 5 (I) *Let* $\mu=\lambda^\infty$ *be a power probability on* $\mathbf{C}^{\mathbf{N}}$ *where* $\lambda(\neq\delta_0)$ *is a probability on* $\mathbf{C}$. *We put*

$$E_{I_\lambda}=\{[e_n];e_n\in I_\lambda\}.$$

(1) $L_\mu\subset E_{I_\lambda}+l_2^D$.
(2) *Let* λ *take the form of* $d\lambda(u,v)=\psi(u+iv)^2\,du\,dv$ *where* $\psi>0$(a.e.), $\int_{\mathbf{R}^2}\psi(u+iv)^2\,du\,dv=1$.
(2.a) $[\mathcal{M}_2,|M\psi|^2]\Leftrightarrow E_{I_\lambda}+l_2^D=L_\mu$.
(2.b) *Let* $0<p<2$.

$$[\mathcal{M}_p,|M\psi|^2]\Rightarrow E_{I_\lambda}+l_p^D\subsetneqq L_\mu.$$
$$E_{I_\lambda}+l_p^D\subset L_\mu \Rightarrow [\mathcal{M}_q,|M\psi|^2] \quad \textit{for } 0<\forall q<p.$$

For $A=[a_n]\in L_\mu$, *the density of* $A\circ\mu$ *with respect to* μ *is given as*

$$\frac{dA\circ\mu}{d\mu}(z_1,z_2,\ldots)=\prod_{n=1}^{\infty}\frac{1}{|a_n|^2}\frac{\psi(z_n a_n^{-1})^2}{\psi(z_n)^2}, \quad (z_1,z_2,\ldots)\in\mathbf{C}^{\mathbf{N}}. \tag{0.17}$$

(II) *Let* μ *be an exchangeable probability on* $\mathbf{C}^{\mathbf{N}}$ *enjoying the expression* (0.9). *Denote by* $\{U_k\}_{k\in\mathbf{N}}$ *the family of the subgroups of the unit circle* U *of* $\mathbf{C}$ *and by* p *the canonical projection of* $\mathbf{C}^{\mathbf{N}}$ *onto the first factor* $\mathbf{C}$. *Put*

$$U_0=\cap\{U_k;\nu(\{\alpha\in\Lambda;I_{\theta_\alpha}=U_k\})>0\}. \tag{0.18}$$

Then we have $L_\mu\subset\mathbf{C}E_{U_0}+l_2^D$, *in particular* $L_\mu\subset\mathbf{C}E_{I_{p\circ\mu}}+l_2^D$.

The intersection in (0.18) is actually taken finite times as in (0.10). Theorem 4 and Theorem 5 will be proved in §5.

In §1, we prepare several preliminary results on the quasi-invariant translations of measures. Although most of them are derived from the well-known results (c.f. [Y]) through simple modifications, we give their proofs (or outlines of proofs) for the sake of self-containedness. §2 is devoted to the proof of a proposition about the Fourier transform on finitely generated commutative Lie groups, which will play an important role in the following discussions. In the rest of this paper, we prove Theorem 1–5 stated above. We mark the end of *proof* with '□'.

1 Preliminary results

In this section we prove two preliminary propositions. One is a criterion for quasi-invariant translations in terms of the convergence of the associated series and the other claims the continuity of the embedding of l_p into Y_μ. The following proofs are parallel to the ones for the case of vector spaces due to Yamasaki.

Proposition 1.1 *Let μ be a power probability on $G^{\mathbf{N}}$ satisfying $\mu=\lambda^\infty$, $d\lambda(g)=\psi(g)^2 dg$ and $\psi>0$* a.e. *For $a=(a_k)_{k\in\mathbf{N}}\in G^{\mathbf{N}}$, we get $a\in Y_\mu$ if and only if*

$$\sum_{k=1}^{\infty}\{1-\int_G \psi(g-a_k)\psi(g)\,dg\}<\infty. \tag{1.1}$$

The density is given as

$$\frac{d\tau_a\mu}{d\mu}(g_1,g_2,\ldots)=\lim_{n\to\infty}\frac{dp_n\circ\tau_a\mu}{dp_n\circ\mu}(g_1,\ldots,g_n)=\prod_{k=1}^{\infty}\frac{\psi(g_k-a_k)^2}{\psi(g_k)^2}. \tag{1.2}$$

Proof (*cf.* [Y 2], Part B, Theor. 7.1, 20.2). Let p_n denote the projection of $G^{\mathbf{N}}$ onto G^n such that $p_n(g_1,g_2,\ldots)=(g_1,\ldots,g_n)$. Using the martingale convergence theorem and the equality

$$\frac{dp_n\circ\tau_a\mu}{dp_n\circ\mu}(g_1,\ldots,g_n)=\prod_{k=1}^{n}\frac{\psi(g_k-a_k)^2}{\psi(g_k)^2},$$

we see that

$$a\in Y_\mu\Leftrightarrow\sqrt{\frac{dp_n\circ\tau_a\mu}{dp_n\circ\mu}}\quad\text{converges in}\quad L_2(\mu)$$

$$\Leftrightarrow\left\|\prod_{k=1}^{n}\frac{\psi(g_k-a_k)}{\psi(g_k)}-\prod_{k=1}^{m}\frac{\psi(g_k-a_k)}{\psi(g_k)}\right\|_{L_2(\mu)}\xrightarrow{m,n\to\infty}0$$

$$\Leftrightarrow\int_{G^n}\prod_{k=1}^{m}\frac{\psi(g_k-a_k)^2}{\psi(g_k)^2}\prod_{k=m+1}^{n}\frac{\psi(g_k-a_k)}{\psi(g_k)}\,d\lambda^n\xrightarrow{m,n\to\infty}1$$

$$\Leftrightarrow\prod_{k=m+1}^{n}\int_G\psi(g-a_k)\psi(g)\,dg\xrightarrow{m,n\to\infty}1\Leftrightarrow(1.1).\quad\square$$

Let Y_μ be endowed with the Kakutani topology defined by the Kakutani distance, which is equivalent to the distance: $\|\tau_x \mu - \tau_y \mu\|$. The Kakutani topology is an important tool for the evaluation of Y_μ. For details, we refer to [Y 1] and [Y 2]. The topology on $l_p(G, I_\lambda)$ (defined in (0.4)) is the one induced by the seminorm: $\sum_{k=1}^{\infty} |a_k|_{I_\lambda}^p$.

Proposition 1.2 *Let $0<p<\infty$ and $\mu=\lambda^\infty$ be a power probability on $G^{\mathbf{N}}$. The inclusion $l_p(G, I_\lambda)\subset Y_\mu$ implies necessarily the continuity of the embedding.*

Proof. (c.f. [Y 2], Part B, Theor. 18.5). Noting that $l_p(G, I_\lambda)$ is a complete metric space, we give a proof based on Baire's category theorem. Let R and T denote the real part and the torus part of G respectively. Put $I'=I_\lambda\cap T$. It suffices to show that the embedding: $l_p(R\times T, I')\subset Y_\mu$ is continuous. For $\varepsilon>0$, put

$$U_\varepsilon=\{x\in l_p(R\times T, I'); \|\tau_x \mu-\mu\|\leqq\varepsilon\}.$$

To apply Baire's category theorem, we show the following two properties.

(i) U_ε is a closed set in $l_p(R\times T, I')$.

(ii) $l_p(R\times T, I')=\bigcup_{n=1}^{\infty} nU_\varepsilon$.

Admitting (i) and (ii), we see by Baire's theorem that for some $n_0\in\mathbf{N}$, $x_0\in n_0\, U_\varepsilon$ and $\delta>0$

$$\|x\|_{l_p}\leqq\delta\Rightarrow x_0+x\in n_0\, U_\varepsilon\Rightarrow x\in n_0\, U_\varepsilon-n_0\, U_\varepsilon\subset n_0\, U_{2\varepsilon},$$

i.e. $\|x\|_{l_p}\leqq\delta/n_0\Rightarrow\|\tau_x \mu-\mu\|\leqq 2\varepsilon$. This implies the continuity of $l_p(R\times T, I')\subset Y_\mu$.

Proof of (i) (*c.f.* [Y 2], Part B, Theor. 18.2). Put

$$C=\{\varphi(g_1, \ldots, g_n); \varphi\in C_0(G^n), n\in\mathbf{N}\}.$$

For $x\in l_p(R\times T, I')$, we have

$$\|\tau_x \mu-\mu\|=\sup_{\varphi\in C,\, \|\varphi\|_\infty\leqq 1}\int_{G^{\mathbf{N}}}(\varphi(y+x)-\varphi(y))\, d\mu(y).$$

Since $\int_{G^{\mathbf{N}}}(\varphi(y+x)-\varphi(y))\, d\mu(y)$ is continuous in x, their supremum is lower semi-continuous in x, from which follows (i).

Proof of (*ii*) (*c.f.* [Y 2], Part B, *Theor.* 18.4). Let $x=(x_k)\in l_p(R\times T, I')$, x_k^R: real part of x_k, x_k^T: torus part of x_k. Identifying $\mathbf{T}$ with the unit circle of $\mathbf{C}$, we put $y_t=(tx_k^R, e^{it x_k^T})_{k\in\mathbf{N}}\in(R\times T)^{\mathbf{N}}$ for $t\in\mathbf{R}$. Then $Y=\{y_t; t\in\mathbf{R}\}$ is a 1-parameter subgroup of Y_μ. Let m denote the Haar measure on Y and $\mu*m$ the convolution of μ with m. $\mu*m$ is Y-invariant and equivalent to μ. Put $\rho=d\mu/d\mu*m$ (the Radon-Nikodym derivative). The equality

$$\|\tau_{y_t} \mu-\mu\|=\int_{G^{\mathbf{N}}}|\rho(y+y_t)-\rho(y)|\, d\mu*m(y)$$

and the Y-invariance of $\mu*m$ imply that $\|\tau_{y_t} \mu-\mu\|\to 0$ as $t\to 0$. Noting that $x=y_1$, we thus see that $x\in nU_\varepsilon$ for a sufficiently large n. □

Remark. Let Y_μ^0 and C_μ denote the maximal connected subgroup of Y_μ and the subgroup of the continuous translations (c.f. [Y 1], §25; [Y 2], §18; [Y-H], §1). The above proof of (ii) shows the inclusion: $Y_\mu^0 \subset C_\mu$.

2 Fourier transforms

It turned out that the elements of Y_μ are characterized by (1.1). In order to clarify the role of G-translations, we consider the Fourier transform of (1.1). Noting that

$$\int_G \psi(g-a_k)\psi(g)\,dg = \int_{\hat{G}} \langle a_k, \gamma\rangle |\hat{\psi}(\gamma)|^2 d\gamma = \int_{\hat{G}} \mathrm{Re}\langle a_k, \gamma\rangle |\hat{\psi}(\gamma)|^2\,d\gamma,$$

we rewrite (1.1) as

$$\sum_{k=1}^{\infty} (1-\phi(a_k)) < \infty \tag{2.1}$$

(see (0.7)). Hence $\lim_{k\to\infty} \phi(a_k) = 1$ is a necessary condition for $a \in Y_\mu$. The purpose of this short section is to prove the following

Proposition 2.1 *Let $d\lambda(g) = \psi(g)^2 dg$ be a probability on G and $a = (a_k)_{k\in\mathbf{N}} \in G^{\mathbf{N}}$. Under the notations in* (0.3) *and* (0.7), *we have*

$$\phi(a_k) \xrightarrow{k\to\infty} 1 \Leftrightarrow |a_k|_{I_\lambda} \xrightarrow{k\to\infty} 0.$$

Lemma $\phi(g) = 1 \Leftrightarrow g \in I_\lambda$.

Proof. Let $\tilde{\lambda}$ denote the adjoint probability of λ i.e. $d\tilde{\lambda}(\gamma) = |\hat{\psi}(\gamma)|^2 d\gamma$. ϕ is the characteristic function of $\tilde{\lambda}$. Hence

$$\phi(g) = 1 \Leftrightarrow \langle g, \gamma\rangle = 1 \quad \text{for } \tilde{\lambda}\text{-a.e. } \gamma.$$

Let g^L denote the non-compact part of g (where $L = R \times Z$). Then $\phi(g) = 1$ implies $g^L = 0$. Indeed, if $g^L \neq 0$, $\{\gamma \in \hat{G}; \langle g, \gamma\rangle = 1\}$ is a null set with respect to the Haar measure on $\hat{G}$ therefore also to $\tilde{\lambda}$, which contradicts $\phi(g) = 1$. We can thus assume that g belongs to the compact part $K = T \times F$ of G. Put

$$E_g = \{\gamma \in \hat{K}; \langle g, \gamma\rangle = 1\}.$$

Then we see that

$$\begin{aligned} \langle g, \gamma\rangle = 1 \text{ for } \tilde{\lambda}\text{-a.e.} &\Leftrightarrow \operatorname{supp}\hat{\psi} \subset \hat{L} \times E_g \Leftrightarrow \langle g, \gamma\rangle\, \hat{\psi}(\gamma) = \hat{\psi}(\gamma) \text{(a.e.)} \\ &\Leftrightarrow \tau_g \psi = \psi \text{(a.e.)} \Leftrightarrow \tau_g \lambda = \lambda \Leftrightarrow g \in I_\lambda. \end{aligned} \tag{2.2}$$

This completes the proof of lemma. □

Proof of Proposition 2.1. Being the Fourier transform of an $L_1(\hat{G})$-function, ϕ is continuous on G. Further (2.2) implies that $\phi(h+g) = \phi(h)$ for $g \in I_\lambda$. Hence, $\phi(a_k) \to \phi(0) = 1$ if $|a_k|_{I_\lambda} \to 0$. Let us show the converse implication. By the Riemann-Lebesgue theorem, $\phi(g) \to 0$ if $g \to \infty$. We can hence assume that $\{a_k\}$ is a precompact set. An arbitrary subsequence of (a_k) contains a converging

sequence (b_j). Take a limit point b of (b_j), i.e. $|b_j - b|_{I_\lambda} \xrightarrow{j\to\infty} 0$. By the continuity of ϕ, we have

$$\phi(b) = \lim_{j\to\infty} \phi(b_j) = \lim_{k\to\infty} \phi(a_k) = 1,$$

which is equivalent to $b \in I_\lambda$ by lemma. This implies that $|a_k|_{I_\lambda} \to 0$. □

If μ is a power probability on $G^{\mathbf{N}}$ satisfying $\mu = \lambda^\infty$, $d\lambda(g) = \psi(g)^2 dg$ and $\psi > 0$ a.e., Proposition 1.1 and 2.1 show that $|a_k|_{I_\lambda} \to 0$ is necessary for $a = (a_k)_{k\in\mathbf{N}} \in Y_\mu$. In particular, denoting the integral part of a_k by a_k^Z and the finite part by a_k^F, we have $a_k^Z = 0$ and $a_k^F \in I_\lambda$ for sufficiently large k's.

3 Quasi-invariance of power probabilities

This section is devoted to the proofs of Theorem 1 and Theorem 2. We first prove Theorem 1 in the case where λ is equivalent to the Haar measure on G (3.1) and secondly complete the proof using this result (3.2). Theorem 2 is proved in 3.3 and 3.4.

3.1 Proof of Theorem 1

Let $\mu = \lambda^\infty$ be a power probability on $G^{\mathbf{N}}$ satisfying $d\lambda(g) = \psi(g)^2 dg$ and $\psi(g) > 0$ (a.e.). We show that $a = (a_k)_{k\in\mathbf{N}} \in Y_\mu$ implies $\sum_{k=1}^{\infty} |a_k|_{I_\lambda}^2 < \infty$. It follows from Proposition 1.1, (2.1) and Proposition 2.1 that $|a_k|_{I_\lambda} \to 0$ as $k \to \infty$ if $a \in Y_\mu$. Hence we can assume without loss of generality that $a_k \in R \times T$ (i.e. the continuous part of G) for every k. Let us rewrite (2.1) as

$$\sum_{k=1}^{\infty} \int_{\hat{G}} (1 - \mathrm{Re}\langle a_k, \gamma\rangle)\, d\tilde{\lambda}(\gamma) < \infty \tag{3.1}$$

where $d\tilde{\lambda}(\gamma) = |\hat{\psi}(\gamma)|^2 d\gamma$ is the adjoint probability of λ. For $a_k = (a_k^R, a_k^T) \in R \times T \subset G$ and $\gamma = (\xi, n) \in R \times Z (\simeq \hat{R}^x \hat{T}) \subset \hat{G}$, their pairing is expressed as $\langle a_k, \gamma\rangle = \exp i(a_k^R \cdot \xi + a_k^T \cdot n)$ where each component of a_k^T belongs to $[-\pi, \pi)$ and '$\cdot$' denotes the canonical inner product (i.e. the sum of component-wise products). Considering an appropriate translation of each a_k^T by an element of I_λ in condition (3.1), we can assume that $a_k^T \to 0$ as $k \to \infty$ since $\langle g, \gamma\rangle = 1$ for $\tilde{\lambda}$-a.e. γ if $g \in I_\lambda$. Let λ' be a probability on $R \times Z (\subset \hat{G})$ defined by $\lambda'(B) = \tilde{\lambda}(B \times T \times F)$ for each Borel set B of $R \times Z$. Thus we have from (3.1)

$$\sum_{k=1}^{\infty} \int_{R\times Z} \{1 - \cos(a_k^R \cdot \xi + a_k^T \cdot n)\}\, d\lambda'(\xi, n) < \infty. \tag{3.2}$$

Put $Z_0 = \{n \in Z;\ \lambda'(R \times \{n\}) > 0\}$. Condition (3.2) implies that

$$\sum_{k=1}^{\infty} \{1 - \cos(a_k^R \cdot \xi + a_k^T \cdot n)\} < \infty \quad \text{for } n \in Z_0 \text{ and } \lambda'|_{R\times\{n\}}\text{-a.e. } \xi. \tag{3.3}$$

First we show that $(a_k^R)_{k\in\mathbf{N}}\in l_2(R)$. For a fixed $n_0\in Z_0$, let R_1 denote the subset of R consisting of the ξ's for which (3.3) holds. From condition (3.3) and the assumption: a_k^R, $a_k^T\to 0$ as $k\to\infty$, we have $\sum_{k=1}^{\infty}|a_k^R\cdot\xi+a_k^T\cdot n_0|^2<\infty$ for $\xi\in R_1$ hence

$$(a_k^R\cdot(\xi-\xi'))_{k\in\mathbf{N}}\in l_2 \quad \text{for } \xi,\xi'\in R_1. \tag{3.4}$$

Since a positive value is assigned to R_1 with respect to the Lebesgue measure on R, $\{\xi-\xi';\ \xi,\ \xi'\in R_1\}$ contains 0 as an interior point. Hence (3.4) implies that $(a_k^R\cdot\xi)_{k\in\mathbf{N}}\in l_2$ for $\xi\in R$ i.e. $(a_k^R)_{k\in\mathbf{N}}\in l_2(R)$.

Secondly we show that $\sum_{k=1}^{\infty}|a_k^T|_{I_\lambda}^2<\infty$. It follows from (3.3) and (3.4) that

$$(a_k^T\cdot n)_{k\in\mathbf{N}}\in l_2 \quad \text{for } n\in Z_0. \tag{3.5}$$

We embed both $T=[-\pi,\pi)^m$ and $Z=\mathbf{Z}^m$ canonically in $\mathbf{R}^m$. Let R_0 denote the linear hull of Z_0 (with $\mathbf{R}$-coefficients) in $\mathbf{R}^m$. If $\dim R_0=m$, (3.5) immediately implies that $(a_k^T)_{k\in\mathbf{N}}\in l_2(\mathbf{R}^m)$. If $\dim R_0<m$, put $R_0^\perp=\{\eta\in\mathbf{R}^m;\ \xi\cdot n=0 \text{ for } \xi\in R_0\}$. For $\eta\in R_0^\perp$, taking a sufficiently small $\alpha\in\mathbf{R}$, we can regard $\alpha\eta$ as an element of $[-\pi,\pi)^m(=T)$. Then $\alpha\eta\in I_\lambda$ holds. Indeed, using the inverse Fourier transform, we have

$$\begin{aligned}\psi(x,t-\alpha\eta,y)&=\int_{\hat G}\exp -i\{x\cdot\xi+(t-\alpha\eta)\cdot n+y\cdot\tau\}\,\hat\psi(\xi,n,\tau)\,d\xi\,dn\,d\tau\\&=\int_{R\times Z_0\times T\times F}\exp -i(x\cdot\xi+t\cdot n+y\cdot\tau)\,\hat\psi(\xi,n,\tau)\,d\xi\,dn\,d\tau=\psi(x,t,y).\end{aligned}$$

Now we decompose a_k^T as $a_k^T=b_k+c_k$ where $b_k\in R_0$, $c_k\in R_0^\perp$. Then (3.5) implies $(b_k)_{k\in\mathbf{N}}\in l_2(\mathbf{R}^m)$. Since $a_k^T\to 0$ therefore $c_k\to 0$ as $k\to\infty$, we can assume that $c_k\in I_\lambda$. Hence the required condition: $\sum_{k=1}^{\infty}|a_k^T|_{I_\lambda}^2<\infty$ holds. We have thus proved Theorem 1 under the restriction that λ is equivalent to the Haar measure on G.

3.2 Proof of Theorem 1 (continued)

Let $\mu=\lambda^\infty$ be a power probability on $G^{\mathbf{N}}$. Considering the convolution of λ with a probability ν on G which is equivalent to the Haar measure on G, we see that $I_\lambda\subset I_{\lambda*\nu}$ holds and $\lambda*\nu$ is equivalent to the Haar measure.

Lemma *If $\hat\nu$ has no zeros on $\hat G$, then the equality: $I_\lambda=I_{\lambda*\nu}$ holds.*

Proof.

$$\begin{aligned}g\in I_{\lambda*\nu}&\Leftrightarrow\tau_g(\lambda*\nu)=\lambda*\nu\Leftrightarrow\langle g,\gamma\rangle\,\hat\lambda(\gamma)\,\hat\nu(\gamma)=\hat\lambda(\gamma)\,\hat\nu(\gamma)\quad\text{for }\gamma\in\hat G\\&\Leftrightarrow\langle g,\gamma\rangle\,\hat\lambda(\gamma)=\hat\lambda(\gamma)\ \text{for }\gamma\in\hat G\Leftrightarrow\tau_g\lambda=\lambda\Leftrightarrow g\in I_\lambda.\quad\square\end{aligned}$$

Now we give an example of a probability which satisfies the condition of lemma.

Example. In order to construct such an example, it suffices to take probabilities ν_R, ν_T, ν_Z and ν_F on $\mathbf{R}$, $\mathbf{T}$, $\mathbf{Z}$ and $\mathbf{Z}_p$ respectively which satisfy the required condition and to consider their direct product. We denote by N an appropriate normalizing constant and by dx, dt, dm, dk the Haar measures on $\mathbf{R}$, $\mathbf{T}$, $\mathbf{Z}$, $\mathbf{Z}_p$ respectively.

ν_R: standard Gaussian measure on $\mathbf{R}$ i.e. $d\nu_R(x)=\frac{1}{N}\exp-\frac{1}{2}|x|^2\,dx$.

ν_T: Put $\theta(t)=\sum_{m\in\mathbf{Z}}\exp-imt\exp-\frac{1}{2}|m|^2$. From an elementary property of theta functions we see that $\theta(t)$ is positive on $\mathbf{T}$. Put $d\nu_T(t)=\frac{1}{N}\theta(t)\,dt$. Then $\hat{\nu}_T(m)=\frac{1}{N}\exp-\frac{1}{2}|m|^2$ has no zeros on $\mathbf{Z}$.

ν_Z: Put $d\nu_Z(m)=\frac{1}{N}\exp-\frac{1}{2}|m|^2 dm$. The same discussion as above shows that $\hat{\nu}_Z(t)=\frac{1}{N}\theta(-t)$ has no zeros on $\mathbf{T}$.

ν_F: Let f be a positive function on $\mathbf{Z}_p$ satisfying $f(k)=1$ for $k\in\mathbf{Z}_p\setminus\{0\}$. Taking $f(0)$ sufficiently large and putting $d\nu_F(k)=\frac{1}{N}f(k)\,dk$, we can obtain $\hat{f}(\kappa)=\sum_{k=0}^{p-1}f(k)\exp i\frac{2\pi}{p}k\kappa\neq 0$ for $\kappa\in\mathbf{Z}_p$.

We continue the proof of Theorem 1. Let ν be a probability on G satisfying the condition of lemma. Put $\mu'=(\lambda*\nu)^\infty$. Since we can apply the result of 3.1 to μ', we have $Y_{\mu'}\subset l_2(G, I_{\lambda*\nu})=l_2(G, I_\lambda)$. On the other hand, we easily see $Y_\mu\subset Y_{\mu*\nu^\infty}=Y_{\mu'}$. Combining these two inclusions completes the proof of Theorem 1.

3.3 Proof of Theorem 2-(1)

Note that (see (2.1))

(3.6)
$$a\in l_p(G, I_\lambda)\Leftrightarrow\sum_{k=1}^{\infty}|a_k|_{I_\lambda}^p<\infty$$

$$a\in Y_\mu\Leftrightarrow\sum_{k=1}^{\infty}(1-\phi(a_k))<\infty.$$

Hence the implication '$\Leftarrow$' in (1) is obvious. We show the converse implication. From the observation that for $a=(a_k)_{k\in\mathbf{N}}\in Y_\mu$

$$\|\tau_a\mu-\mu\|\sim\left\|1-\sqrt{\frac{d\tau_a\mu}{d\mu}}\right\|_{L_2(\mu)}$$

$$=\left\{\int_{G^{\mathbf{N}}}\left|1-\prod_{k=1}^{\infty}\frac{\psi(g_k-a_k)}{\psi(g_k)}\right|^2 d\mu(g_1, g_2, \ldots)\right\}^{\frac{1}{2}}$$

$$=\sqrt{2}\left\{1-\prod_{k=1}^{\infty}\int_G\psi(g-a_k)\,\psi(g)\,dg\right\}^{\frac{1}{2}}$$

and the continuity of the embedding: $l_p(G, I_\lambda) \subset Y_\mu$ proved in Proposition 1.2, we see that for any $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\|a\|_{l_p(G,I_\lambda)} \leqq \delta \Rightarrow \sum_{k=1}^{\infty} \{1 - \int_G \psi(g-a_k)\psi(g)\,dg\} \leqq \varepsilon. \tag{3.7}$$

Let $x \in R \times T, |x|_{I_\lambda} \leqq 1$ and x_λ denote the nearest point to x in I_λ (i.e. $|x-x_\lambda| = |x|_{I_\lambda}$). Put

$$a = \Big(\overbrace{\frac{\delta}{n^{\frac{1}{p}}}(x-x_\lambda)+x_\lambda, \ldots, \frac{\delta}{n^{\frac{1}{p}}}(x-x_\lambda)+x_\lambda}^{n \text{ terms}}, 0, \ldots\Big) \in (R \times T)^{\mathbf{N}}.$$

Then $\|a\|_{l_p(G,I_\lambda)} \leqq \delta |x-x_\lambda| \leqq \delta$. Hence, from (3.7),

$$\sum_{k=1}^{n} \left\{1 - \int_G \psi\left(g - \frac{\delta}{n^{\frac{1}{p}}}(x-x_\lambda) - x_\lambda\right)\psi(g)\,dg\right\} \leqq \varepsilon \quad \text{i.e.}$$

$$1 - \int_G \psi\left(g - \frac{\delta}{n^{\frac{1}{p}}}(x-x_\lambda) - x_\lambda\right)\psi(g)\,dg \leqq \frac{\varepsilon}{n} \quad \text{for } x \in R \times T, |x|_{I_\lambda} \leqq 1. \tag{3.8}$$

Now let $y \in R \times T$ be arbitrarily given so as to satisfy $|y|_{I_\lambda} \leqq \delta$. Take $n \in \mathbf{N}$ such that

$$\frac{\delta}{(n+1)^{\frac{1}{p}}} < |y|_{I_\lambda} \leqq \frac{\delta}{n^{\frac{1}{p}}}.$$

Then there exists $x \in R \times T$ which satisfies $|x|_{I_\lambda} \leqq 1$, $x_\lambda = y_\lambda$ and $y = \frac{\delta}{n^{\frac{1}{p}}}(x-x_\lambda) + x_\lambda$. Applying (3.8) to y, we have

$$1 - \int_G \psi(g-y)\psi(g)\,dg \leqq \frac{\varepsilon}{n} = \frac{n+1}{n} \frac{\varepsilon}{(n+1)^{\frac{1}{p}p}} \leqq 2\varepsilon \frac{|y|^p_{I_\lambda}}{\delta^p}. \tag{3.9}$$

In the case where $y \in R \times T$ and $|y|_{I_\lambda} > \delta$,

$$1 - \int_G \psi(g-y)\psi(g)\,dg \leqq 2 < \frac{2}{\delta^p}|y|^p_{I_\lambda}. \tag{3.10}$$

Considering the Fourier transform of the left hand sides of (3.9) and (3.10), we obtain

$$1 - \phi(y) \leqq \frac{2}{\delta^p}|y|^p_{I_\lambda} \quad \text{for } y \in R \times T. \tag{3.11}$$

Next let $z \in Z \times F$. Since $Z \times F$ is a discrete group, there exists a positive number c such that $z \notin I_\lambda \Rightarrow |z|_{I_\lambda} \geqq c$. Hence,

$$1 - \phi(z) \leqq 2 \leqq \frac{2}{c^p}|z|^p_{I_\lambda} \quad \text{for } z \in Z \times F. \tag{3.12}$$

(Note that $|z|_{I_\lambda}=0 \Leftrightarrow z \in I_\lambda \Leftrightarrow \phi(z)=1$ by lemma in §2.) For general $g \in G$, decomposing g as $g=y+z$, $y \in R \times T$, $z \in Z \times F$ and using the inequality:

$$1-\mathrm{Re}\langle y+z, \gamma\rangle \leqq 2(1-\mathrm{Re}\langle y, \gamma\rangle + 1-\mathrm{Re}\langle z, \gamma\rangle),$$

we see from (3.11) and (3.12) that

$$\begin{aligned} 1-\phi(g) &\leqq 2(1-\phi(y)+1-\phi(z)) \\ &\leqq \frac{4}{\delta^p}|y|_{I_\lambda}^p + \frac{4}{c^p}|z|_{I_\lambda}^p \leqq C|g|_{I_\lambda}^p \end{aligned} \tag{3.13}$$

where C is a positive constant depending only on G, λ, p. The rest of statement (1) is obvious from (3.6).

3.4 Proof of Theorem 2-(2)(3)

First we prove (2). Let condition $[\mathscr{M}_2, |\hat{\psi}|^2]$ $(=[\mathscr{M}_2, \tilde{\lambda}])$ hold (see (0.6) for the definition). Since $\langle g, \gamma\rangle = 1$ for $\tilde{\lambda}$-a.e. γ if $g \in I_\lambda$,

$$1-\mathrm{Re}\langle g, \gamma\rangle = 1-\mathrm{Re}\langle g-g_\lambda, \gamma\rangle \lesssim |g|_{I_\lambda}^2 |\gamma|^2 \quad \tilde{\lambda}\text{-a.e.} \tag{3.14}$$

where g_λ denotes an element of I_λ satisfying $|g-g_\lambda|=|g|_{I_\lambda}$. Hence, the inequality

$$\begin{aligned} 1-\phi(g) &= \int_{\hat{G}} (1-\mathrm{Re}\langle g, \gamma\rangle)\, d\tilde{\lambda}(\gamma) \\ &\lesssim |g|_{I_\lambda}^2 \int_{\hat{G}} |\gamma|^2 d\tilde{\lambda}(\gamma) \quad \text{for } g \in G \end{aligned} \tag{3.15}$$

and $[\mathscr{M}_2, \tilde{\lambda}]$ shows the inclusion: $l_2(G, I_\lambda) \subset Y_\mu$ by virtue of (1). The converse inclusion necessarily holds from Theorem 1. Conversely let $l_2(G, I_\lambda) = Y_\mu$ hold. By (1) there exists $C>0$ such that $1-\phi(g) \leqq C|g|_{I_\lambda}^2$ for $g \in G$. Let $\xi \in \mathbf{R}$ denote the first component of the real part of $\gamma \in \hat{G}$. Then,

$$\int_{\hat{G}} (1-\cos u\xi)\, d\tilde{\lambda}(\gamma) \leqq Cu^2 \quad \text{for } u \in \mathbf{R}.$$

Using Fatou's lemma, we have

$$\int_{\hat{G}} |\xi|^2 d\tilde{\lambda}(\gamma) \leqq 2 \liminf_{u \to 0} \int_{\hat{G}} \frac{1-\cos u\xi}{u^2}\, d\tilde{\lambda}(\gamma) \leqq 2C. \tag{3.16}$$

For the other components of the real part and for the integral part, the same estimations as (3.16) hold. The torus and finite parts make no matter since $|\gamma|^2$ is bounded on $T \times F$. Thus we see that $[M_2, \tilde{\lambda}]$ holds. The second equivalence immediately follows from an elementary property of the Fourier transforms.

Secondly we prove (3). Let $0<p<2$. Using $1-\cos t \lesssim |t|^p$ for $t \in \mathbf{R}$, we have, similarly to (3.14), $1-\mathrm{Re}\langle g, \gamma\rangle \lesssim |g|_{I_\lambda}^p |\gamma|^p$ $\tilde{\lambda}$-a.e. Then,

$$1-\phi(g) \lesssim |g|_{I_\lambda}^p \int_{\hat{G}} |\gamma|^p d\tilde{\lambda}(\gamma). \tag{3.17}$$

Hence $[\mathscr{M}_p, \tilde{\lambda}]$ implies $l_p(G, I_\lambda) \subset Y_\mu$ by virtue of (1). Now assume that $l_p(G, I_\lambda) \subset Y_\mu$ holds. Let ξ denote the first component of the real part of $\gamma \in \hat{G}$ as above. By (1) there exists $C>0$ such that

$$\int_{\hat{G}} (1-\cos u|\xi|)\, d\tilde{\lambda}(\gamma) \leqq C(u^p \wedge 1) \quad \text{for } u>0. \tag{3.18}$$

Let ε be an arbitrary number satisfying $0<\varepsilon<p$. From (3.18),

$$\int_{\hat{G}} \frac{1-\cos u|\xi|}{u^{p-\varepsilon+1}}\, d\tilde{\lambda}(\gamma) \leqq C\, \frac{u^p \wedge 1}{u^{p-\varepsilon+1}}$$

where $1<p-\varepsilon+1<3$. Integrating the both sides with respect to u on $[0, \infty)$ and using the formula:

$$\int_0^\infty \frac{1-\cos au}{u^r}\, du = \frac{\pi a^{r-1}}{2\Gamma(r)\sin\frac{\pi}{2}(r-1)} \quad \text{for } a \geqq 0, \quad 1<r<3, \tag{3.19}$$

we have

$$\int_{\hat{G}} \frac{\pi|\xi|^{p-\varepsilon}}{2\Gamma(p-\varepsilon+1)\sin\frac{\pi}{2}(p-\varepsilon)}\, d\tilde{\lambda}(\gamma) \leqq C\left(\frac{1}{\varepsilon}+\frac{1}{p-\varepsilon}\right) \quad \text{i.e.}$$

$$\int_{\hat{G}} |\xi|^{p-\varepsilon}\, d\tilde{\lambda}(\gamma) \leqq \frac{2C}{\pi}\left(\frac{1}{\varepsilon}+\frac{1}{p-\varepsilon}\right)\Gamma(p-\varepsilon+1)\sin\frac{\pi}{2}(p-\varepsilon). \tag{3.20}$$

The same estimations as (3.20) hold for the other components of the real part and for the integral part of $\gamma \in \hat{G}$. Thus we see that $[\mathscr{M}_q, \tilde{\lambda}]$ holds for $0<\forall q<p$. It remains to show that $[\mathscr{M}_p, \tilde{\lambda}]$ does not hold if $l_p(G, I_\lambda) = Y_\mu$. By (1), $l_p(G, I_\lambda) = Y_\mu$ implies that

$$\int_{\hat{G}} (1-\cos u|\xi|)\, d\tilde{\lambda}(\gamma) \sim u^p \quad (u \to 0, u>0).$$

Then, using (3.19) again (where $a=|\xi|$, $r-1=p$), we have

$$\int_{\hat{G}} |\xi|^p\, d\tilde{\lambda}(\gamma) = \int_{\hat{G}} \frac{2\Gamma(p+1)}{\pi}\sin\frac{\pi p}{2}\left(\int_0^\infty \frac{1-\cos|\xi|u}{u^{p+1}}\, du\right) d\tilde{\lambda}(\gamma)$$

$$\sim \frac{2\Gamma(p+1)}{\pi}\sin\frac{\pi p}{2}\int_0^\infty \frac{u^p \wedge 1}{u^{p+1}}\, du = +\infty.$$

The density (0.8) is nothing but (1.2). This completes the proof of (3).

Remark. Assertion (2) is a generalization of Theorem 3.1 in [Sh1] to the case of commutative Lie groups, while assertion (3) is of the same form as Theorem 3.5 in [Sh2] concerning differentiable measures on $\mathbf{R}^{\mathbf{N}}$.

4 Quasi-invariance of exchangeable probabilities

As stated in Introduction, a probability on $G^{\mathbf{N}}$ is said to be exchangeable if it is invariant with respect to all the actions of $\mathscr{S}_\infty$. First we prove the ergodic decomposition (0.9) of exchangeable probabilities on $G^{\mathbf{N}}$. This enables us to derive Theorem 3 from Theorem 1. If G did not have a good measurable structure, all exchangeable probabilities would not enjoy such decompositions. For details, we refer to [D-F].

4.1 Ergodic decomposition of exchangeable probabilities

Let us begin with a well-known lemma.

Lemma ([Y 2], Part B, § 7, Example) *Let $\mu=\lambda^\infty$ and $\mu'=\lambda'^\infty$ be power probabilities on $G^{\mathbf{N}}$. Then $\lambda\neq\lambda'\Rightarrow\mu\perp\mu'$.*

Proof. Since μ and μ' are $\mathscr{S}_\infty$-ergodic, the equivalent-singular dichotomy holds for them. Hence it suffices to show $\mu\sim\mu'$ in order to establish $\mu\perp\mu'$. It is obvious that $\lambda\sim\lambda'$ implies $\mu\sim\mu'$. Let $\lambda\sim\lambda'$ holds i.e. $d\lambda'=f\,d\lambda$, $f>0$ (λ-a.e.). From the same discussion as proof of Proposition 1.1, we see

$$\mu\sim\mu'\Leftrightarrow\sum_{k=1}^{\infty}\left(1-\int_G\sqrt{f(g)}\,d\lambda(g)\right)<\infty\Leftrightarrow\int_G\sqrt{f(g)}\,d\lambda(g)=1$$

$$\Leftrightarrow\sqrt{f}=1\,(\lambda\text{-a.e.})\Leftrightarrow\lambda=\lambda'$$

Thus $\lambda\neq\lambda'\Rightarrow\mu\sim\mu'$. □

Proposition 4.1 *An arbitrary exchangeable probability μ on $G^{\mathbf{N}}$ enjoys the ergodic decomposition expressed as* (0.9): $\mu=\int_A\theta_\alpha^\infty\,d\nu(\alpha)$.

Proof. Let $\mathscr{F}$ and $\mathscr{E}$ denote the Borel σ-field and the exchangeable σ-field on $G^{\mathbf{N}}$ respectively, i.e. $\mathscr{E}=\{B\in\mathscr{F};\ \sigma B=B \text{ for } \sigma\in\mathscr{S}_\infty\}$. Since G is standard, there exists the conditional probability with respect to $\mathscr{E}$, which is denoted by μ^x at $x\in G^{\mathbf{N}}$. $\mu^x(B)$ is $\mathscr{E}$-measurable for every $B\in\mathscr{F}$ and satisfies

$$\mu(E\cap B)=\int_E\mu^x(B)\,d\mu(x)\quad\text{for } B\in\mathscr{F},\quad E\in\mathscr{E}. \tag{4.1}$$

Let p denote the projection of $G^{\mathbf{N}}$ onto the first factor G. From de Finetti's theorem (see [M], VIII, Théor. 5) we see that

$$\mu^x=(p\circ\mu^x)^\infty\quad\text{for } \mu\text{-a.e. } x. \tag{4.2}$$

Equalities (4.1) and (4.2) imply the decomposition:

$$\mu=\int_{G^{\mathbf{N}}}(p\circ\mu^x)^\infty\,d\mu(x).$$

We modify the parameter space so that the integrands may be mutually singular. Let $\{A_k\}_{k\in\mathbf{N}}$ generate the Borel σ-field on G. We define an equivalence relation between the elements of $G^{\mathbf{N}}$ by

$$x\sim y \Leftrightarrow p\circ\mu^x(A_k)=p\circ\mu^y(A_k) \quad \text{for } k\in\mathbf{N}.$$

Relation ' $\sim$ ' induces a partition of $G^{\mathbf{N}}$ into $\mathscr{E}$-measurable sets. Put $\Lambda=G^{\mathbf{N}}/\sim$(i.e. all the equivalence classes) and denote by $\tilde{x}$ the equivalence class corresponding to $x\in G^{\mathbf{N}}$. Let us define a probability ν on Λ as the image measure of $\mu|_E$ by this canonical projection of $G^{\mathbf{N}}$ onto Λ. Put $\theta_\alpha=p\circ\mu^x$ for $\alpha=\tilde{x}\in\Lambda$. θ_α is well-defined and the correspondence: $\alpha\leftrightarrow\theta_\alpha$ is one-to-one since $\tilde{x}=\tilde{y}\Leftrightarrow x\sim y\Leftrightarrow p\circ\mu^x=p\circ\mu^y$. Moreover $\alpha\neq\beta$ implies $\theta_\alpha\neq\theta_\beta$ therefore $\theta_\alpha^\infty\perp\theta_\beta^\infty$ by lemma. Thus we obtain the desired decomposition (0.9). □

4.2 Proof of Theorem 3

Let μ be an exchangeable probability on $G^{\mathbf{N}}$ expressed as (0.9). First we show that every θ_α can be assumed to be equivalent to the Haar measure on G without loss of generality. Let ρ be a probability on G which is equivalent to the Haar measure and whose characteristic function $\hat{\rho}$ has no zeros on $\hat{G}$. Then $\theta_\alpha*\rho$ is also equivalent to the Haar measure. Moreover we see $I_{\theta_\alpha*\rho}=I_{\theta_\alpha}$ by lemma in 3.2. $\mu*\rho^\infty$ is an exchangeable probability expressed as

$$\mu*\rho^\infty=(\int_\Lambda \theta_\alpha^\infty\, d\nu(\alpha))*\rho^\infty=\int_\Lambda(\theta_\alpha*\rho)^\infty\, d\nu(\alpha)$$

and satisfies $Y_\mu\subset Y_{\mu*\rho^\infty}$. Hence we can justify the additional assumption by considering $\mu*\rho^\infty$ instead of μ if necessary.

Now we prove Theorem 3 assuming that θ_α is of the form: $d\theta_\alpha(g)=\psi_\alpha(g)^2\,dg$, $\psi_\alpha>0$ (a.e.). Let $a=(a_k)_{k\in\mathbf{N}}\in Y_\mu$. Then

$$\int_\Lambda \tau_a\theta_\alpha^\infty\, d\nu(\alpha)\sim\int_\Lambda\theta_\alpha^\infty\, d\nu(\alpha). \tag{4.3}$$

Since the equivalent-singular dichotomy holds for the family: $\{\theta_\alpha^\infty,\ \tau_a\theta_\alpha^\infty\}_{\alpha\in\Lambda}$, (4.3) implies that, for ν-a.e. $\alpha\in\Lambda$ (denoted by $\alpha\in\Lambda_a$ where $\nu(\Lambda_a)=1$), there exists $\beta\in\Lambda$ such that $\tau_a\theta_\alpha^\infty\sim\theta_\beta^\infty$. From the same discussion as Proposition 1.1, we see that for $\alpha\in\Lambda_a$

$$\sum_{k=1}^\infty\{1-\int_G\psi_\alpha(g-a_k)\,\psi_\beta(g)\,dg\}<\infty$$

and by the Fourier transform

$$\sum_{k=1}^\infty\{1-\int_{\hat{G}}\langle a_k,\gamma\rangle\,\hat{\psi}_\alpha(\gamma)\,\overline{\hat{\psi}_\beta(\gamma)}\,d\gamma\}<\infty. \tag{4.4}$$

Since $\hat{\psi}_\alpha\overline{\hat{\psi}_\beta}$ is an L_1-function, $\varphi(g)=\int_{\hat{G}}\langle a_k,\gamma\rangle\,\hat{\psi}_\alpha(\gamma)\,\overline{\hat{\psi}_\beta(\gamma)}\,d\gamma$ is continuous and satisfies $\varphi(g)\to0$ as $g\to\infty$ (by the Riemann-Lebesgue theorem). Hence (4.4)

implies that $\{a_k\}_{k\in\mathbf{N}}$ is bounded therefore precompact. Let c_0 be an arbitrary limit point of $\{a_k\}_{k\in\mathbf{N}}$. By (4.4) and the continuity of φ, we have $\varphi(c_0)=1$. Hence $\int_G \psi_\alpha(g-c_0)\psi_\beta(g)\,dg=1$, i.e. $\psi_\alpha(g-c_0)=\psi_\beta(g)$ for a.e.g. Put $c=(c_0, c_0, \ldots)\in\Delta$. Then

$$\tau_c\,\theta_\alpha^\infty=\theta_\beta^\infty\sim\tau_a\,\theta_\alpha^\infty \tag{4.5}$$

hence $a-c\in Y_{\theta_\beta^\infty}$. Since c is independent of the choice of α, we have $a-c\in\bigcap_{\alpha\in\Lambda_a} Y_{\theta_\beta^\infty}$. Thus, using Theorem 1, we obtain

$$a-c\in\bigcap_{\alpha\in\Lambda_a} l_2(G, I_{\theta_\alpha}). \tag{4.6}$$

Here we show a simple lemma.

Lemma *Let H and H' be closed subgroups of $K=T\times F$ (i.e. the compact part of G). Then $l_2(G, H)\cap l_2(G, H')=l_2(G, H\cap H')$.*

Proof. The inclusion '$\supset$' is obvious. We show the converse inclusion. Let $x=(x_k)_{k\in\mathbf{N}}\in l_2(G, H)\cap l_2(G, H')$. Then $|x_k|_H$, $|x_k|_{H'}\to 0$ therefore $|x_k|_{H\cap H'}\to 0$ as $k\to\infty$. Hence we can assume that $x_k\in R\times T$ and $x_k\to 0$ as $k\to\infty$ by considering an appropriate translation of x_k by an element of $H\cap H'\cap T$ if necessary. Since all the involved elements are thus contained in a neighborhood of 0 in $R\times T$, we can regard G and H, H' as a Euclidean space and its subspaces. In this case we easily see that there exists a positive constant C (depending on H, H') such that

$$|v|_{H\cap H'}\leqq C(|v|_H+|v|_{H'}) \quad \text{for } v\in G.$$

This implies $x\in l_2(G, H\cap H')$. □

Let us continue the proof of Theorem 3. We note that the number of the subgroups of $T\times F$ is countable. The proof (which is done by induction on the dimension of T) is omitted here. If $H_1, H_2, \ldots$ are decreasing (resp. increasing) subgroups of $T\times F$, there exists $n_0\in\mathbf{N}$ such that $H_n=H_{n_0}$ holds for $n\geqq n_0$ since the dimension of H_n as a Lie group is finite and decreasing (resp. increasing). Combining this observation with Lemma, we see that the intersection in (4.6) is actually taken finite times and

$$a-c\in l_2(G, \bigcap_{\alpha\in\tilde{\Lambda}_a} I_{\theta_\alpha}) \tag{4.7}$$

where $\tilde{\Lambda}_a$ is a finite subset of Λ_a such that $\bigcap_{\alpha\in\tilde{\Lambda}_a} I_{\theta_\alpha}=\bigcap_{\alpha\in\Lambda_a} I_{\theta_\alpha}$. Noting that $K_0=\max\{\bigcap_{\alpha\in\tilde{\Omega}} I_{\theta_\alpha};\ v(\Omega)=1, \Omega\subset\Lambda\}$ (see (0.10)), we obtain

$$Y_\mu\subset l_2(G, K_0)+\Delta \tag{4.8}$$

from (4.7). Further $K_0\subset I_{p\circ\mu}$ since $p\circ\mu=\int_\Lambda \theta_\alpha\,dv(\alpha)$. Hence $Y_\mu\subset l_2(G, I_{p\circ\mu})+\Delta$ in particular. We have already seen that the elements of Δ corresponding to $a=(a_k)_{k\in\mathbf{N}}\in Y_\mu$ is given as a sequence of an arbitrary limit point of $\{a_k\}_{k\in\mathbf{N}}$. In

fact, if c_0 and c_0' are two limit points of $\{a_k\}_{k\in\mathbf{N}}$ such that $(a_k)_{k\in\mathbf{N}}\in Y_\mu$, we see that $c_0-c_0'\in\bigcap_{\alpha\in\Lambda_a} I_{\theta_\alpha}\subset K_0$ from (4.5).

We terminate this section with a simple example which assures that Δ is not negligible in Theorem 3.

Example. Let λ be a probability on $\mathbf{R}$, $c_n>0$ and $\sum_{n\in\mathbf{Z}} c_n=1$, that is to say, we regard $(c_n)_{n\in\mathbf{Z}}$ as a mixing measure on the parameter space $\mathbf{Z}$. Put

$$\mu=\sum_{n\in\mathbf{Z}} c_n(\tau_n\lambda)^\infty. \tag{4.9}$$

Then $(1, 1, \ldots)\in Y_\mu$ holds since

$$\tau_{(1,1,\ldots)}\mu(B)=0\Leftrightarrow\sum_{n\in\mathbf{Z}} c_n(\tau_{n+1}\lambda)^\infty(B)=0$$

$$\Leftrightarrow(\tau_n\lambda)^\infty(B)=0 \text{ for } \forall n\in\mathbf{Z}\Leftrightarrow\mu(B)=0.$$

On the other hand, since $p\circ\mu=\sum_{n\in\mathbf{Z}} c_n\,\tau_n\lambda$,

$$1\in I_{p\circ\mu}\Leftrightarrow\sum_{n\in\mathbf{Z}}(c_n-c_{n-1})\lambda(B-n)=0 \quad \text{for every Borel set } B \text{ of } \mathbf{R}. \tag{4.10}$$

For example, if λ is uniformly distributed on $[-\frac{1}{2},\frac{1}{2}]$, we have $\lambda([-\frac{1}{2},\frac{1}{2}]-n)=0$ for $n\neq 0$. Then (4.10)$\Rightarrow c_0=c_{-1}$. Hence $c_0\neq c_{-1}\Rightarrow|1|_{I_{p\circ\mu}}>0\Rightarrow(1, 1, \ldots)\notin l_2(G, I_{p\circ\mu})$. We thus see that $Y_\mu\not\subset l_2(G, I_{p\circ\mu})$.

5 Quasi-invariant diagonal transformations on sequence spaces

In this section we apply the preceding results to the investigation of the quasi-invariance of probabilities on real or complex sequence spaces with respect to the linear transformations by diagonal matrices. After some preliminary remarks (5.1), we prove Theorem 4 and Theorem 5 (5.2).

5.1 Preliminaries

Our concerns are the quasi-invariance of probabilities which differ from δ_0 on $\mathbf{R}^\mathbf{N}$ or $\mathbf{C}^\mathbf{N}$. Let us show, however, that it suffices to deal with probabilities on $\dot{\mathbf{R}}^\mathbf{N}$ or $\dot{\mathbf{C}}^\mathbf{N}$ (where $\dot{\mathbf{R}}=\mathbf{R}\setminus\{0\}$, $\dot{\mathbf{C}}=\mathbf{C}\setminus\{0\}$). Since it makes no difference whether the space is $\mathbf{R}^\mathbf{N}$ or $\mathbf{C}^\mathbf{N}$, we consider just the complex case. In the case of majorizing L_μ (i.e. (I)-(1) and (II) of Theor. 4, 5), it is no matter since $L_\mu\subset L_{\mu|_{\mathcal{C}^\mathbf{N}}}$ holds. In the case of minorizing L_μ by $E_\lambda+l_p^D$ (i.e. (I)-(2) of Theor. 4, 5), we have the necessary condition that $\sigma_z\lambda\sim\lambda$ holds for $\forall z\in\dot{\mathbf{C}}$ where $\mu=\lambda^\infty$ and $\sigma_z\lambda=\lambda(\cdot z^{-1})$. We can hence express λ as $\lambda=\alpha\delta_0+(1-\alpha)\lambda_0$ for $0\leqq\alpha<1$ and for

a probability λ_0 which is equivalent to the Lesbegue measure on $\mathbf{C}(\simeq \mathbf{R}^2)$. By the same discussion as Proposition 1.1,

$$A=[a_n]\in L_\mu \Leftrightarrow \sum_{n=1}^{\infty}\left(1-\int_{\mathbf{C}}\sqrt{\frac{d\sigma_{a_n}\lambda}{d\lambda}}\,d\lambda\right)<\infty.$$

Combining this with the equality

$$1-\int_{\mathbf{C}}\sqrt{\frac{d\sigma_z\lambda}{d\lambda}}\,d\lambda=1-\alpha\int_{\mathbf{C}}\sqrt{\frac{d\sigma_z\lambda}{d\lambda}}\,d\delta_0-(1-\alpha)\int_{\mathbf{C}}\sqrt{\frac{d\sigma_z\lambda}{d\lambda}}\,d\lambda_0$$

$$=1-\alpha-(1-\alpha)\int_{\dot{\mathbf{C}}}\sqrt{\frac{d\sigma_z\lambda_0}{d\lambda_0}}\,d\lambda_0=(1-\alpha)\left(1-\int_{\dot{\mathbf{C}}}\sqrt{\frac{d\sigma_z\lambda_0}{d\lambda_0}}\,d\lambda_0\right)$$

and putting $\mu_0=\lambda_0^{\infty}$ (regarded as a probability on $\dot{\mathbf{C}}^{\mathbf{N}}$), we see $A\in L_\mu \Leftrightarrow A\in L_{\mu_0}$. Thus it suffices to consider λ_0 instead of λ, that is to say, to restrict λ on $\dot{\mathbf{C}}$ and to normalize it.

As stated in Introduction, the actions of diagonal transformations onto $\dot{\mathbf{R}}^{\mathbf{N}}$ and $\dot{\mathbf{C}}^{\mathbf{N}}$ are transferred to the translations on $(\mathbf{R}\times\mathbf{Z}_2)^{\mathbf{N}}$ and $(\mathbf{R}\times\mathbf{T})^{\mathbf{N}}$ respectively in the following way. Let us define Φ: $\dot{\mathbf{R}}\to\mathbf{R}\times\mathbf{Z}_2$ or $\dot{\mathbf{C}}\to\mathbf{R}\times\mathbf{T}$ such that

$$\Phi^{-1}(x,k)=e^x(-1)^k \quad \text{for } (x,k)\in\mathbf{R}\times\mathbf{Z}_2 \tag{5.1}$$
$$\Phi^{-1}(x,\theta)=e^{x+i\theta} \quad \text{for } (x,\theta)\in\mathbf{R}\times\mathbf{T}.$$

Φ is an isomorphism of multiplicative groups $\dot{\mathbf{R}}$, $\dot{\mathbf{C}}$ onto additive groups $\mathbf{R}\times\mathbf{Z}_2$, $\mathbf{R}\times\mathbf{T}$. For a diagonal matrix $A=[a_n](a_n\in\dot{\mathbf{R}}$ or $\dot{\mathbf{C}})$, put

$$\Phi A=(\Phi a_1,\Phi a_2,\ldots)\in(\mathbf{R}\times\mathbf{Z}_2)^{\mathbf{N}} \quad \text{or} \quad (\mathbf{R}\times\mathbf{T})^{\mathbf{N}}. \tag{5.2}$$

Then we easily see

$$\Phi\sigma_a=\tau_{\Phi a}\Phi, \qquad \Phi\circ A=\tau_{\Phi A}\Phi \tag{5.3}$$

$$a\in I_\lambda \Leftrightarrow \Phi a\in I_{\Phi\circ\lambda} \tag{5.4}$$

$$A\in L_\mu \Leftrightarrow \Phi A\in Y_{\Phi\circ\mu}\ . \tag{5.5}$$

where λ is a probability on $\dot{\mathbf{R}}$ or $\dot{\mathbf{C}}$, μ is a probability on $\dot{\mathbf{R}}^{\mathbf{N}}$ or $\dot{\mathbf{C}}^{\mathbf{N}}$ and $\Phi\circ\mu$ is the image measure of μ by $\Phi(a_1,a_2,\ldots)=(\Phi a_1,\Phi a_2,\ldots)$. Moreover, under (5.5), we have

$$\frac{dA\circ\mu}{d\mu}(y_1,y_2,\ldots)=\frac{d\tau_{\Phi A}\Phi\circ\mu}{d\Phi\circ\mu}(\Phi y_1,\Phi y_2,\ldots), \quad (y_1,y_2,\ldots)\in\dot{\mathbf{R}}^{\mathbf{N}} \text{ or } \dot{\mathbf{C}}^{\mathbf{N}}. \tag{5.6}$$

5.2 Proof of Theorem 4 and 5

First we consider the real case. Theorem 4 can be derived from Theorem 1, 2 and 3 through (5.5) and the following correspondence.

Proposition 5.1 *We use the notations in Theorem* 4.

(1) *For* $A=[a_n]$ *and* $\Phi A=(\Phi a_n)$,

$$A\in E+l_p^D \Leftrightarrow \Phi A\in l_p(\mathbf{R}\times\mathbf{Z}_2,\mathbf{Z}_2)$$
$$A\in Id+l_p^D\Leftrightarrow\Phi A\in l_p(\mathbf{R}\times\mathbf{Z}_2,\{0\}).$$

(2) *Let* λ *take the form of* $d\lambda(u)=\psi(u)^2\,du$, $\psi>0$ (a.e.) *and* $h(x,k)$ *denote the square root of the density of* $\Phi\circ\lambda$. *Then,*

$$[\mathcal{M}_p,|M_{\pm}\psi|^2]\Leftrightarrow[\mathcal{M}_p,|\hat{h}|^2].$$

Proof. (1) Let $a_n\in\dot{\mathbf{R}}$ and $\Phi a_n=(x_n,k_n)\in\mathbf{R}\times\mathbf{Z}_2$, i.e. $a_n=e^{x_n}(-1)^{k_n}$. Then,

$$(\Phi a_n)\in l_p(\mathbf{R}\times\mathbf{Z}_2,\mathbf{Z}_2)\Leftrightarrow\sum_{n=1}^{\infty}|x_n|^p<\infty\Leftrightarrow\sum_{n=1}^{\infty}(\log|a_n|)^p<\infty$$
$$\Leftrightarrow\sum_{n=1}^{\infty}(1-|a_n|)^p<\infty\Leftrightarrow[a_n]\in E+l_p^D,$$

$$(\Phi a_n)\in l_p(\mathbf{R}\times\mathbf{Z}_2,\{0\})\Leftrightarrow\sum_{n=1}^{\infty}|x_n|^p<\infty\quad\text{and}\quad\sum_{n=1}^{\infty}|k_n|^p<\infty$$
$$\Leftrightarrow a_n\geqq 0\quad\text{with finite exceptions and}\quad\sum_{n=1}^{\infty}(\log a_n)^p<\infty$$
$$\Leftrightarrow\sum_{n=1}^{\infty}(1-a_n)^p<\infty\Leftrightarrow[a_n]\in Id+l_p^D.$$

(2) We have

$$d\Phi\circ\lambda(x,k)=\psi(e^x(-1)^k)^2\,e^x\,dx\,dk \tag{5.7}$$

since, for a Borel set B of $\mathbf{R}$ and $l\in\mathbf{Z}_2$,

$$\Phi\circ\lambda(B\times\{l\})=\lambda((-1)^l\exp B)=\int_{(-1)^l\exp B}\psi(u)^2\,du$$
$$=\int_B\psi(e^x(-1)^l)^2\,e^x\,dx=\int_{B\times\{l\}}\psi(e^x(-1)^k)^2\,e^x\,dx\,dk.$$

Hence

$$\begin{aligned}\hat{h}(\xi,\kappa)&=\int_{\mathbf{R}\times\mathbf{Z}_2}e^{ix\xi}(-1)^{k\kappa}\psi(e^x(-1)^k)\,e^{\frac{x}{2}}\,dx\,dk\\&=\int_{\mathbf{R}}e^{ix\xi}\,e^{\frac{x}{2}}\psi(e^x)\,dx+(-1)^{\kappa}\int_{\mathbf{R}}e^{ix\xi}\,e^{\frac{x}{2}}\psi(-e^x)\,dx\\&=\int_0^{\infty}r^{i\xi-\frac{1}{2}}\psi(r)\,dr+(-1)^{\kappa}\int_0^{\infty}r^{i\xi-\frac{1}{2}}\psi(-r)\,dr\\&=M_+\psi(\xi)+(-1)^{\kappa}M_-\psi(\xi),\end{aligned} \tag{5.8}$$

from which we see

$$(5.9)\quad [\mathcal{M}_p, |\hat{h}|^2] \Leftrightarrow \int_{\mathbf{R}\times\mathbf{Z}_2} |(\xi,\kappa)|^p |M_+\psi(\xi)+(-1)^\kappa M_-\psi(\xi)|^2\, d\xi d\kappa < \infty$$
$$\Leftrightarrow \int_{\mathbf{R}} |\xi|^p |M_+\psi(\xi)+M_-\psi(\xi)|^2\, d\xi + \int_{\mathbf{R}} (|\xi|+1)^p |M_+\psi(\xi)$$
$$-M_-\psi(\xi)|^2\, d\xi < \infty \Leftrightarrow [\mathcal{M}_p, |M_+\psi|^2] \text{ and } [\mathcal{M}_p, |M_-\psi|^2]. \quad \square$$

Now it is easy to verify Theorem 4. We see from (5.4) that

$$\lambda\text{: symmetric} \Leftrightarrow I_\lambda = \{-1, 1\} \Leftrightarrow I_{\Phi\circ\lambda} = \mathbf{Z}_2,$$
$$\lambda\text{: not symmetric} \Leftrightarrow I_\lambda = \{1\} \Leftrightarrow I_{\Phi\circ\lambda} = \{0\}.$$

(I) follows immediately from (5.5) and Proposition 5.1. As for the density, since

$$(5.10)\quad \frac{d\tau_{\Phi A}\,\Phi\circ\mu}{d\Phi\circ\mu}(x_1, k_1, x_2, k_2, \ldots) = \prod_{n=1}^{\infty} \frac{\psi(\Phi^{-1}((x_n, k_n)-\Phi a_n))^2\, e^{x_n - \log|a_n|}}{\psi(\Phi^{-1}(x_n, k_n))^2\, e^{x_n}}$$

holds by (5.7) and (0.8), formula (5.6) implies (0.14). Let μ be an exchangeable probability on $\mathbf{R}^{\mathbf{N}}$ expressed as $\mu = \int_\Lambda \theta_\alpha^\infty\, d\nu(\alpha)$. If $p\circ\mu = \int_\Lambda \theta_\alpha\, d\nu(\alpha)$ is not symmetric, we have $\nu(\{\alpha\in\Lambda;\ I_{\Phi\circ\theta_\alpha} = \{0\}\}) = \nu(\{\alpha\in\Lambda;\ \theta_\alpha \text{ is not symmetric}\}) > 0$ therefore $K_0 = \{0\}$ in (0.10) (replacing θ_α and $T\times F$ by $\Phi\circ\theta_\alpha$ and $\mathbf{Z}_2$ respectively). Hence, from Proposition 5.1-(1) and Theorem 3, letting $A = [a_n]\in L_\mu$ and c_0 be a limit point of $\{a_n\}_{n\in\mathbf{N}}$, we see

$$A\in L_\mu \Leftrightarrow (\Phi a_n)\in Y_{\Phi\circ\mu} \Leftrightarrow (\Phi a_n)-(\Phi c_0)\in l_2(\mathbf{R}\times\mathbf{Z}_2, \{0\})$$
$$\Leftrightarrow [a_n c_0^{-1}]\in Id + l_2^D \Leftrightarrow [a_n]\in[c_0] + l_2^D (\subset \mathbf{R}Id + l_2^D).$$

If c_0' is another limit point of $\{a_n\}_{n\in\mathbf{N}}$, it follows from (4.5) that $\Phi c_0 - \Phi c_0' \in K_0 = \{0\}$. Hence $\{a_n\}_{n\in\mathbf{N}}$ has a unique limit point if $[a_n]\in L_\mu$ and $p\circ\mu$ is not symmetric. In the case where $p\circ\mu$ is symmetric, we see similarly that

$$A = [a_n]\in L_\mu \Rightarrow (\Phi a_n)-(\Phi c_0)\in l_2(\mathbf{R}\times\mathbf{Z}_2, K_0) \subset l_2(\mathbf{R}\times\mathbf{Z}_2, \mathbf{Z}_2)$$
$$\Rightarrow [a_n c_0^{-1}]\in E + l_2^D \Leftrightarrow [a_n]\in c_0 E + l_2^D (\subset \mathbf{R}E + l_2^D).$$

The limit point c_0 of $\{a_n\}_{n\in\mathbf{N}}$ is unique modulo the sign. We have thus proved Theorem 4.

Let us turn to the complex case.

Proposition 5.2 *We use the notations in Theorem 5. Denote by U the unit circle of $\mathbf{C}$. There is one-to-one correspondence between the closed subgroups of $\mathbf{T}$ and U through $H \leftrightarrow \exp iH$.*

(1) $A = [a_n]\in E_{\exp iH} + l_p^D \Leftrightarrow \Phi A = (\Phi a_n)\in l_p(\mathbf{R}\times\mathbf{T}, H)$.
(2) *Let λ take the form of $d\lambda(u, v) = \psi(u+iv)^2\, du dv$, $\psi > 0$ (a.e.) and $h(x, \theta)$ denote the square root of the density of $\Phi\circ\lambda$. Then,*

$$[\mathcal{M}_p, |M\psi|^2] \Leftrightarrow [\mathcal{M}_p, |\hat{h}|^2].$$

Proof. (1) Put $\Phi a_n=(x_n, \theta_n)$ i.e. $a_n=e^{x_n+i\theta_n}$. Then,

$$(\Phi a_n)\in l_p(\mathbf{R}\times\mathbf{T}, H)\Leftrightarrow \sum_{n=1}^{\infty}|x_n|^p<\infty \quad \text{and} \quad \sum_{n=1}^{\infty}|\theta_n|_H^p<\infty$$

$$\Leftrightarrow \sum_{n=1}^{\infty}(\log|a_n|)^p<\infty \quad \text{and} \quad \sum_{n=1}^{\infty}|\theta_n|_H^p<\infty$$

$$\Leftrightarrow \sum_{n=1}^{\infty}(1-|a_n|)^p<\infty \quad \text{and} \quad \sum_{n=1}^{\infty}|\arg a_n|_H^p<\infty$$

$$\Leftrightarrow \sum_{n=1}^{\infty}|a_n|_{\exp iH}^p<\infty \Leftrightarrow [a_n]\in E_{\exp iH}+l_p^D.$$

(2) Noting that

$$d\Phi\circ\lambda(x,\theta)=\psi(e^{x+i\theta})^2\,e^{2x}\,dx\,d\theta \tag{5.11}$$

since, for a Borel set $B\times C$ of $\mathbf{R}\times\mathbf{T}$,

$$\Phi\circ\lambda(B\times C)=\int_{\Phi^{-1}(B\times E)}\psi(u+iv)^2\,du\,dv=\int_{(\exp B)\times E}\psi(re^{i\theta})^2\,r\,dr\,d\theta$$

$$=\int_{B\times E}\psi(e^{x+i\theta})^2\,e^{2x}\,dx\,d\theta,$$

we have

$$\hat{h}(\xi,m)=\int_{\mathbf{R}\times\mathbf{T}}e^{i\xi x+im\theta}\,\psi(e^{x+i\theta})\,e^{x}\,dx\,d\theta \tag{5.12}$$

$$=\int_{\mathbf{R}_+\times\mathbf{T}}r^{i\xi}\,e^{im\theta}\,\psi(re^{i\theta})\,dr\,d\theta=M\psi(\xi,m). \quad \square$$

Now let us verify Theorem 5. (I) is an immediate consequence of Proposition 5.2. We obtain the formula of density (0.17) from (0.8), (5.6) and (5.11) similarly to (0.14). Let μ be an exchangeable probability on $\mathbf{C}^{\mathbf{N}}$ expressed as $\mu=\int_A \theta_\alpha^\infty\,d\nu(\alpha)$. Letting $A=[a_n]\in L_\mu$ and denoting by c_0 a limit point of $\{a_n\}_{n\in\mathbf{N}}$, we see from Proposition 5.2-(1) and Theorem 3 that

$$A\in L_\mu\Leftrightarrow(\Phi a_n)\in Y_{\Phi\circ\mu}\Rightarrow(\Phi a_n)-(\Phi c_0)\in l_2(\mathbf{R}\times\mathbf{T}, K_0)$$

$$\Leftrightarrow\sum_{n=1}^{\infty}|\Phi(a_n c_0^{-1})|_{K_0}^2<\infty\Leftrightarrow\sum_{n=1}^{\infty}|a_n c_0^{-1}|_{\exp iK_0}^2<\infty$$

$$\Leftrightarrow\sum_{n=1}^{\infty}|a_n-c_0\exp iK_0|^2<\infty\Leftrightarrow[a_n]\in c_0\exp iK_0+l_2^D$$

where K_0 is the one in (0.10) under the replacement of θ_α and $T\times F$ by $\Phi\circ\theta_\alpha$ and $\mathbf{T}$ respectively. We note that $U_0=\exp iK_0$ holds by (5.4). If c_0' is another limit point of $\{a_n\}_{n\in\mathbf{N}}$, we have $c_0\,c_0'^{-1}\in U_0$ through the similar discussion to the real case. The proof of Theorem 5 is thus completed.

Acknowledgement. I wish to express my gratitude to Prof. Yasuo Yamasaki. A series of his works and lectures introduced me to the present studies. Furthermore, he gave me hearty encouragement and invaluable comments during my preparation for this paper. Especially Theorem 3 was fairly improved by him.

References

[D-F] Dubins, L.E., Freedman, D.A.: Exchangeable processes need not be mixtures of independent identically distributed random variables. Probab. Theory Relat. Fields **48**, 115–132 (1979)

[J] Janssen, A.: Absolute continuity and singularity of distributions of dependent observations: Gaussian and exchangeable measures. In: Heyer, H. (ed.) Probability measures on groups VIII. Proceedings, Oberwolfach 1985. (Lect. Notes Math. vol. 1210, pp 108–124) Berlin Heidelberg New York: Springer 1986

[K] Kakutani, S.: On equivalence of infinite product measures. Ann. Math. **49**, 214–224 (1948)

[M] Meyer, P.: Probabilités et potentiel. Paris: Hermann 1966

[P] Parthasarathy, K.R.: Probability measures on metric spaces. New York: Academic Press 1967

[R] Ritter, G.: On Kakutani's dichotomy theorem for infinite products of not necessarily independent functions. Math. Ann. **239**, 35–53 (1979)

[Sa] Sato, H.: On the convergence of the product of independent random variables. J. Math. Kyoto Univ. **27**, 381–385 (1987)

[S] Shepp, L.A.: Distinguishing a sequence of random variables from a translate of itself. Ann. Math. Stat. **36**, 1107–1112 (1965)

[Sh1] Shimomura, H.: An aspect of quasi-invariant measures on $\mathbf{R}^\infty$. Publ. Res. Inst. Math. Sci. **11**, 749–773 (1976)

[Sh2] Shimomura, H.: An aspect of differentiable measures on $\mathbf{R}^\infty$. Publ. Res. Inst. Math. Sci. **23**, 791–811 (1987)

[U] Umemura(=Yamasaki), Y.: Measures on infinite dimensional vector spaces. Publ. Res. Inst. Math. Sci. **1**, 1–47 (1966)

[X] Xia, D.X.: Measure and integration theory on infinite-dimensional spaces. New York: Academic Press 1972

[Y1] Yamasaki, Y.: Measures on infinite dimensional spaces, vol. 2, Invariant measures (in Japanese). Tokyo: Kinokuniya 1978

[Y2] Yamasaki, Y.: Measures on infinite dimensional spaces. Singapore Philadelphia: World Scientific 1985

[Y-H] Yamasaki, Y., Hora, A.: Differentiable shifts for measures on infinite dimensional spaces. Publ. Res. Inst. Math. Sci. **23**, 275–296 (1987)

Math. Z. 206, 193–210 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

On Schläfli's reduction formula

Ruth Kellerhals

Max-Planck-Institut für Mathematik, Gottfried-Claren-Straße 26, W-5300 Bonn 3, FRG

Received December 7, 1989; in final form February 16, 1990

Introduction

In polyhedral geometry and in particular for the problem of calculating volumes of non-euclidean polytopes, orthoschemes are the most basic objects.

Let X denote the n-dimensional sphere S^n or the n-dimensional hyperbolic space H^n. An orthoscheme in X is a simplex bounded by hyperplanes $H_0, \ldots, H_n$ such that $H_i \perp H_j$ for $|i-j|>1$. Hence, a planar orthoscheme is a right-angled triangle, whose area formula can be expressed by the well-known defect formula.

Ludwig Schläfli generalized this formula to spherical orthoschemes of even dimensions; Schläfli's reduction formula represents the volume of an even dimensional spherical orthoscheme in terms of the volumes of certain lower dimensional ones (see [S]). This formula can be easily extended to the hyperbolic case by analytic continuation (see [H], p. 134ff). In this paper, we consider a more general class of hyperbolic polytopes, the so-called (complete) orthoschemes of degree d, $0 \leqq d \leqq 2$, (see [IH] and [K 2]). These polytopes arise e.g. as a particular class of fundamental polytopes in the classification problem for hyperbolic Coxeter groups. Our aim is to show that a generalized reduction formula holds for even dimensional complete orthoschemes; we shall see that this reduction law simplifies with increasing degree (of truncation) d.

For the following, it is most convenient to describe a polytope by means of its scheme. The scheme of a polytope $P \subset X$ is a weighted graph (characterizing $P \subset X$ up to congruence) in which the nodes correspond to the bounding hyperplanes of P. Two nodes are joined by an edge if the corresponding hyperplanes are not orthogonal; the weight on an edge equals either the cosine of the dihedral angle between the corresponding hyperplanes, or, for diverging hyperplanes in H^n, the hyperbolic cosine of their distance. A pair of non-adjacent nodes is characterized by the weight zero. To every scheme $\Sigma=(n_1 \ldots n_m)$ with nodes $n_1, \ldots, n_m$ corresponds a symmetric matrix (a_{ij}) of order m wherein $a_{ii}=1$ and, for $i \neq j$, a_{ij} equals the negative of the weight between n_i, n_j. A scheme is called elliptic resp. hyperbolic if its matrix is positive definite resp. of index of inertia 1. For more details, see Chap. 1 of this article.

Hence, a spherical orthoscheme R of dimension n is represented by a linear elliptic scheme Σ of order $n+1$. Denote by $f_n(\Sigma)$ or f_n Schläfli's normalized volume function for R given by

$$(1)\quad f_n(\Sigma)=f_n:=c_n\,\mathrm{Vol}_n(R) \quad \text{with} \quad c_n=\frac{2^{n+1}}{\mathrm{Vol}_n(S^n)}=\frac{2^n}{\pi^{(n+1)/2}}\,\Gamma\left(\frac{n+1}{2}\right), \qquad f_0:=1.$$

If the scheme Σ consists of two disjoint components σ_1, σ_2 of orders n_1+1, $n_2+1\geqq 1$, then $f_n(\Sigma)=f_{n_1}(\sigma_1)\cdot f_{n_2}(\sigma_2)$.

With these preliminaries, Schläfli's reduction formula for spherical orthoschemes can be stated as follows:

$$(2)\qquad f_{2n}(\Sigma)=\sum_{k=0}^{n}\frac{(-1)^k}{k+1}\binom{2k}{k}\sum_{\sigma} f_{2n-(2k+1)}(\sigma), \qquad \sum f_{-1}:=1,$$

where σ runs through all subschemes of Σ of order $2(n-k)$ all of whose components have even order.

For example, we obtain

$$f(012)=f(01)+f(12)-1,$$

$$f(01234)=f(0123)+f(01)f(34)+f(1234)$$
$$-\{f(01)+f(12)+f(23)+f(34)\}+2.$$

Now, let R_d, $0\leqq d\leqq 2$, denote a (complete) orthoscheme of degree d in H^n, i.e., R_d is a d-times truncated orthoscheme bounded by hyperplanes $H_0, \ldots, H_{n+d}$ such that $H_i \perp H_j$ for $j\neq i-1, i, i+1$ (for $d=2$, indices are taken modulo $n+3$). They are described by a linear hyperbolic scheme of order $n+1$ or $n+2$ for $d=0$ or $d=1$, or, by a cyclic hyperbolic scheme of order $n+3$ for $d=2$.

Modifying Schläfli's function for complete orthoschemes $R_d\subset H^n$ with graph Σ_d by

$$(3)\qquad F_n(\Sigma_d):=i^n c_n\,\mathrm{Vol}_n(R_d) \quad \text{with} \quad F_0:=1,\ i^2=-1,$$

we show that the following reduction formula holds for $d=0, 1, 2$:

$$(4)\qquad F_{2n}(\Sigma_d)=\sum_{k=0}^{n}\frac{(-1)^k}{k+1}\binom{2k}{k}\sum_{\sigma} f_{2n-(2k+1)}(\sigma), \qquad \sum f_{-1}:=1,$$

where σ runs through all *elliptic* subschemes of Σ_d of order $2(n-k)$ all of whose components are of even order.

Since the Schläfli function takes only rational values on the set of spherical Coxeter orthoschemes $\left(\text{all dihedral angles are of the form } \frac{\pi}{p},\ p\in\mathbf{N},\ p\geqq 2\right)$, the volume of a complete hyperbolic Coxeter orthoscheme of dimension $2n$ is a rational multiple of π^n. The complete Coxeter orthoschemes were classified by Im Hof in 1983 (see [IH]). He showed that they exist only for dimensions $\leqq 9$; in even dimensions $\geqq 4$, there are only finitely many examples, whose volumes are determined explicitly in an appendix.

1 Complete orthoschemes

1.1 Let X^n denote either the n-dimensional euclidean space E^n, the n-sphere S^n or the n-dimensional hyperbolic space H^n. Let S^n be embedded in E^{n+1}, and use for H^n the model in the Lorentz space $E^{1,n}$ of signature $(1, n)$, i.e.: If $E^{1,n}$ denotes the real vector space R^{n+1} equipped with the bilinear form of signature $(1, n)$

$$\langle x, y\rangle = -x_0 y_0 + x_1 y_1 + \ldots + x_n y_n,$$

$$\forall x = (x_0, \ldots, x_n) \in R^{n+1}, \quad \forall y = (y_0, \ldots, y_n) \in R^{n+1},$$

then H^n can be interpreted as

$$H^n = \{x \in E^{1,n} \mid \langle x, x\rangle = -1,\ x_0 > 0\}.$$

Or, in the projective model, H^n is the interior of real projective space P^n with respect to the quadric $Q_{1,n} = \{[x] \in P^n \mid \langle x, x\rangle = 0\} =: \partial H^n$.

1.2 Let $P \subset H^n$ denote a convex polytope bounded by finitely many hyperplanes H_i, $i \in I$, which are characterized by unit normal vectors $e_i \in E^{1,n}$ directed inwards with respect to P, say, i.e. (for basic notations and properties, see [V 1], § 1):

$$H_i = e_i^{\perp} := \{x \in H^n \mid \langle x, e_i\rangle = 0\} \quad \text{with} \quad \langle e_i, e_i\rangle = 1.$$

We always assume that P is acute-angled $\left(\text{i.e., all dihedral angles } \neq \frac{\pi}{2} \text{ are of measure strictly less than } \frac{\pi}{2}\right)$ and of finite volume. Then, every face $F \subset P$ of dimension k, $1 \leqq k \leqq n-1$, is contained in exactly $n-k$ of the bounding hyperplanes of P (see [A 2], Lemma 1, p. 762), and F is itself an acute-angled polytope of finite volume (see [A 2], Lemma 2, p. 762).

The Gram matrix $G(P) := (\langle e_i, e_j\rangle)_{i,j \in I}$ of the vectors e_i, $i \in I$, associated to P is an indecomposable symmetric matrix of signature $(1, n)$ with entries $\langle e_i, e_i\rangle = 1$ and $\langle e_i, e_j\rangle \leqq 0$ for $i \neq j$, having the following geometrical meaning (see [V 1], § 1):

$$\langle e_i, e_j\rangle = \begin{cases} 0 & \text{if } H_i \perp H_j, \\ -\cos\alpha_{ij} & \text{if } H_i, H_j \text{ intersect on } P \text{ at the angle } \alpha_{ij} = \angle(H_i, H_j), \\ -1 & \text{if } H_i, H_j \text{ are parallel,} \\ -\cosh l_{ij} & \text{if } H_i, H_j \text{ admit a common perpendicular of length } l_{ij}. \end{cases}$$

On the other hand, if $G = (g_{ij})$ is an indecomposable symmetric $m \times m$ matrix of rank $n+1$ with $g_{ii} = 1$ and $g_{ij} \leqq 0$, for $i \neq j$, then G can be realized as Gram matrix $G(P)$ of an acute-angled polytope $P \subset X^n$ of finite volume in the following way (see [V 1], § 2):

1. If G is positive definite (G is *elliptic*), then $m = n+1$, and G is the Gram matrix of a simplex in S^n uniquely defined up to a motion.

2. If G is positive semidefinite (G is *parabolic*), then $m = n+2$, and G is the Gram matrix of a simplex in E^{n+1} uniquely defined up to a similarity.

3. If G is of signature $(1, n)$ (G is *hyperbolic*), then G is the Gram matrix of a convex polytope with m facets (faces of codimension 1) in H^n uniquely defined up to a motion.

In terms of the Gram matrix $G(P)$, the combinatorial structure of an acute-angled polytope $P \subset H^n$ can be described as follows (see [V 1], § 3):

If P is compact, then the positive definite principal submatrices $G_J := (g_{ij})_{i,j \in J}$ of $G(P)$, $J \subset I$ with $1 \leqq |J| \leqq n$, are in one-to-one-correspondence with the non-empty faces

$$P^J := P \cap (\bigcap_{j \in J} H_j),$$

and P^J has codimension $|J|$ (see [V 1], Theor. 3.1). In particular, a vertex $p \in P$ is characterized by a positive definite principal submatrix of $G(P)$ of order n describing the spherical vertex polytope P_p (intersection of P with the surface of a sufficiently small ball around p) of dimension $n-1$ associated to p.

If P is not compact, but of finite volume, then a point $q \in \partial H^n$ is an infinite vertex of P if and only if for $J_q := \{i \in I \mid H_i \ni q\}$ the principal submatrix $G_{J_q} = (g_{ij})_{i,j \in J_q}$ is parabolic of rank $n-1$ (or equivalently, the vertex polytope P_q is a euclidean polytope of dimension $n-1$) (see [V 1], Theor. 3.2).

1.3 In practice, however, the language of schemes is much more convenient for the geometrical description of certain classes of polytopes (see [V 2], § 3). A scheme Σ is a weighted graph (see [V 2], § 2) whose nodes n_i, n_j are joined by an edge with positive weight σ_{ij} or are not joined at all; the last fact will be indicated by $\sigma_{ij} = 0$. A subscheme of Σ is a subgraph of Σ with each pair of nodes connected by the same weighted edge as in Σ. The number $|\Sigma|$ of nodes is called the order of Σ. To every scheme Σ of order m corresponds a symmetric matrix $A(\Sigma) = (a_{ij})$ of order m with $a_{ii} = 1$ on the diagonal and non-positive entries $a_{ij} = -\sigma_{ij} \leqq 0$, $i \neq j$, off it. Σ is connected if and only if $A(\Sigma)$ is indecomposable. Rank, determinant and character of definiteness of Σ are defined to be the corresponding ones of $A(\Sigma)$. Furthermore, Σ is said to be either elliptic, or parabolic, or hyperbolic if either all its components are elliptic, or – apart from elliptic components – there is at least one parabolic component, or exactly one component is hyperbolic.

Now, the scheme $\Sigma(P)$ of an acute-angled polytope $P \subset X^n$ is the scheme whose matrix $A(\Sigma)$ coincides with the Gram matrix $G(P)$, i.e., whose nodes i correspond to the bounding hyperplanes $H_i = e_i^{\perp}$ (or equivalently to their normal vectors e_i) of P and whose weights equal $-\langle e_i, e_j \rangle_{X^n}$, $i, j \in I$. The scheme of a face $F \subset P$ is denoted by $\Sigma(F)$ and is called the *face scheme* of F in P; $\Sigma(F)$ is not a subscheme of $\Sigma(P)$, since, in general, it does not inherit the weights of $\Sigma(P)$.

Two acute-angled polytopes $P_1, P_2 \subset H^n$ are said to be *of the same schematic type* if their schemes $\Sigma(P_1)$, $\Sigma(P_2)$ are of the same graphical type (i.e., their underlying graphs as one-dimensional simplicial complexes are homeomorphic) and if corresponding weights σ_{ij}^1 of $\Sigma(P_1)$ and σ_{ij}^2 of $\Sigma(P_2)$ satisfy:

$$\sigma_{ij}^1 \begin{cases} > \\ = 1 \\ < \end{cases} \Leftrightarrow \sigma_{ij}^2 \begin{cases} > \\ = 1. \\ < \end{cases}$$

It follows that polytopes of the same schematic type are of the same combinatorial type (see *1.2* and [A 1]).

For the schemes of Coxeter polytopes $P_C \subset X^n$ $\left(\text{all dihedral angles are of the form } \frac{\pi}{p},\ p \in \mathbf{N},\ p \geqq 2\right)$ we adopt the usual conventions and – for convenience – use them sometimes even in the non-Coxeter case: If two nodes are related by the weight $\cos\frac{\pi}{p}$, then they are joined by a $(p-2)$-fold line for $p=3, 4$ and by a single line marked p $\left(\text{or } \alpha=\frac{\pi}{p}\right)$ for $p \geqq 5$. If two bounding hyperplanes of $P_C \subset X^n$, $X^n \neq S^n$, are parallel, then the corresponding nodes are joined by a line marked ∞; if they are divergent (occurring at most in the hyperbolic case), then their nodes are joined by a dotted line and the weight $\geqq 1$ is dropped.

The elliptic and parabolic Coxeter schemes were classified by Coxeter in 1934 (see [C]). Hyperbolic Coxeter schemes, however, are only partially classified (see e.g. [V 1], Chap. II).

1.4 The simplest examples of schemes are the linear and cyclic ones. One class of acute-angled hyperbolic polytopes with linear and cyclic schemes is the following (see [IH] and [K 2]): An *n-dimensional complete orthoscheme of degree d*, $0 \leqq d \leqq 2$, or, for short, an *n-orthoscheme of degree d* is a convex polytope in H^n, $n \geqq 2$, denoted by R_d such that its scheme $\Sigma(R_d)$ is connected and linear of length $n+d+1$ for $d=0$, 1 or cyclic of order $n+3$ for $d=2$.

Hence, orthoschemes of degree d in H^n are bounded by $n+d+1$ hyperplanes $H_0, \ldots, H_{n+d}$ such that

$$H_i \perp H_j \quad \text{for } j \neq i-1, i, i+1, \tag{5}$$

where, for $d=2$, indices are taken modulo $n+3$.

Geometrically, orthoschemes of degree d can be described as follows:

For degree $d=0$, they coincide with the class of (ordinary) orthoschemes introduced by Schläfli (see [S], [BH]): An orthoscheme in $X^n (n \geqq 1)$ is an n-simplex bounded by $n+1$ hyperplanes $H_0, \ldots, H_n$ such that $H_i \perp H_j$ for $|i-j|>1$. This is equivalent to say that it has vertices $P_0, \ldots, P_n$ numbered in such a way that $\operatorname{span}(P_0, \ldots, P_i) \perp \operatorname{span}(P_i, \ldots, P_n)$ for $1 \leqq i \leqq n-1$. The initial and final vertices P_0, P_n of the orthogonal edge-path $P_i P_{i+1}$, $i=0, \ldots, n-1$, are called principal vertices, since they are distinguished in several ways. E.g. in H^n, at most the principal vertices may be points at infinity (see [BH], Satz 15, p. 188).

Using the projective model for H^n (see *1.1*), we can derive orthoschemes of degree $d=1$ or $d=2$ from an ordinary one by allowing one or both of its principal vertices (and with them possibly further vertices) to lie outside the quadric $Q_{1,n}$, and then by cutting off the ideal vertices by means of the polar hyperplanes, say H_{n+1} resp. H_{n+2}, corresponding to P_n resp. P_0 (inasmuch as they lie outside $Q_{1,n}$). Hence, orthoschemes of degree $d=0, 1, 2$ are d-times truncated (or "polarily completed") orthoschemes bounded by hyperplanes $H_0, \ldots, H_{n+d}$ with the property (5).

Complete orthoschemes $R_d \subset H^n$, $0 \leqq d \leqq 2$, have at most $n+3$ non-right dihedral angles (or *essential* angles) $\alpha_1, \ldots, \alpha_m$, $m \leqq n+3$, and all of them are

acute, i.e., $\alpha_i < \frac{\pi}{2}$ for $i=1, \ldots, m$ (see [BH], §4.8, Hilfssatz 2, and the definition). Furthermore, by construction, complete orthoschemes are of finite volume (see [V1], Theor. 4.1). Hence, a face of R_d is also an acute-angled polytope of finite volume (see *1.2*); in fact, it is itself a complete orthoscheme (see [K2], *1.3*). If Σ_d is the scheme of R_d, we denote by $\Sigma_d(l)$ the face scheme of the apex $R_d(l)=R_d\cap H_{l-1}\cap H_l$ in R_d associated to the essential angle $\alpha_l = \angle(H_{l-1}, H_l)$, $1\leqq l\leqq m$, of R_d. $\Sigma_d(l)$ is not a subscheme of Σ_d. However, the scheme of a vertex polytope of R_d is a subscheme of Σ_d of order n (see *1.2*) and is therefore a $(n-1)$-orthoscheme with essential angles of the same measure as certain essential angles of R_d. Since the vertex polytope of a face $F\subset R_d$, $2\leqq\dim F\leqq n-1$, associated to a vertex $p\in F$ is the (non-empty) intersection of F with the vertex polytope of R_d associated to p, we conclude by iteration that every subscheme of order k of the scheme $\Sigma(F)$, $2\leqq k\leqq|\Sigma(F)|-1$, is the face scheme of order k of a subscheme of Σ_d, and vice versa. In particular, the set of subschemes of order k of $\Sigma_d(l)$ $(2\leqq k\leqq n+d-2)$ is identical with the set of face schemes of order k of subschemes of Σ_d describing the apex of the dihedral angle of measure α_l in R_d.

2 The volume function of Schläfli

2.1 For $n\geqq 1$, let Σ denote the elliptic scheme of order $n+1\geqq 2$ of a spherical n-othoscheme R. Then, the normalized volume function

$$f_n(\Sigma)=f_n:=c_n\,\mathrm{Vol}_n(R) \tag{6}$$

with

$$c_n=\frac{2^{n+1}}{\mathrm{Vol}_n(S^n)}=\frac{2^n}{\pi^{(n+1)/2}}\,\Gamma\left(\frac{n+1}{2}\right), \qquad f_0:=1,$$

is called the function of Schläfli (see Introduction, (1), and [S], No. 23, p. 238). The function f_n is proportional to $\mathrm{Vol}_n(R)$ such that $f_n=1$ for the orthoscheme with all dihedral angles equal to $\frac{\pi}{2}$. The function of Schläfli satisfies the following factorization property (see [S], No. 23, p. 238, or [BH], Hilfssatz 1, p. 213):

Lemma. *Let Σ denote a linear elliptic scheme of order $n+1$ consisting of disjoint components $\sigma_1, \ldots, \sigma_r$ of orders $n_1+1, \ldots, n_r+1\geqq 1$. Then*

$$f_n(\Sigma)=f_{n_1}(\sigma_1)\ldots f_{n_r}(\sigma_r). \tag{7}$$

For spherical Coxeter orthoschemes (see *1.3*), Schläfli determined explicitly all possible values of f_n (see [S], No. 30, p. 268ff). Using the standard notations for schemes of spherical Coxeter orthoschemes (see [V1], §5, Table 1), his results read as:

$$f_1(G_2^p)=\frac{2}{p},\quad p\geqq 2; \qquad f_2(H_3)=\frac{1}{15}; \tag{8}$$

$$f_3(F_4)=\frac{1}{72}; \qquad f_3(H_4)=\frac{1}{900}; \tag{9}$$

$$f_n(A_{n+1}) = \frac{2^{n+1}}{(n+2)!}, \quad n \geqq 0; \tag{10}$$

$$f_n(B_{n+1}) = \frac{1}{(n+1)!}, \quad n \geqq 1. \tag{11}$$

By means of the trigonometric principle, or equivalently, by interpreting hyperbolic n-space H^n as upper half of the pseudosphere of radius $i = \sqrt{-1}$ in R^{n+1} (see [BH], p. 20–21 or p. 210), it is obvious how to generalize the notion of Schläfli's function to orthoschemes $R_d \subset H^n$ of degree d, $0 \leqq d \leqq 2$, and with graph Σ_d (see also [BH], p. 212):

The function

$$F_n(\Sigma_d) := i^n c_n \operatorname{Vol}_n(R_d) \quad \text{with} \quad i^2 = -1, \quad F_0 := 1, \tag{12}$$

where the constant c_n is defined as in (6), is called the *Schläfli function of the complete orthoscheme* R_d.

Hence, for even dimensions,

$$F_{2n}(\Sigma_d) = (-1)^n \left(\frac{2}{\pi}\right)^n \cdot \prod_{p=1}^{n} (2p-1) \cdot \operatorname{Vol}_{2n}(R_d), \quad n \geqq 1,$$

is a real-valued function.

2.2 Let $\mathscr{R}_\kappa$ be the set of compact complete orthoschemes in H^n of combinatorial type κ (see [A1], § 1). Since every element of $\mathscr{R}_\kappa$ is a polytope with dihedral angles not exceeding $\frac{\pi}{2}$, its congruence class is uniquely determined by its dihedral angles (see [A1], § 3, Uniqueness Theor.). Hence, Schläfli's volume function $F_n = F_n | \mathscr{R}_\kappa$ restricted on $\mathscr{R}_\kappa$ may be regarded as a function of the dihedral angles. The differential of F_n depending on the dihedral angles can be represented by Schläfli's formula in the following way (see [K2], § 2, and 1.4):

Theorem (Schläfli's differential formula). *Let F_n, $n \geqq 2$, be the Schläfli function on the set $\mathscr{R}_\kappa$ of compact complete orthoschemes in H^n of combinatorial type κ with essential angles $\alpha_1, \ldots, \alpha_{m(\kappa)}$ and with scheme Σ. Denote by $F_{n-2}(k)$ the Schläfli function of the apex of codimension 2 associated to the essential angle α_k of measure $f_1(k) := f_1(\alpha_k)$, $1 \leqq k \leqq m(\kappa)$. Then*

$$dF_n(\Sigma) = \sum_{k=1}^{m(\kappa)} F_{n-2}(k)\, df_1(k). \tag{13}$$

This formula was established by Schläfli for spherical simplexes, and separately for the more basic orthoschemes. Much later, H. Kneser gave a second, very elegant proof (see [Kn]) for both, the spherical and the hyperbolic case. As Schläfli already pointed out (see [S], No. 25, p. 246ff, No. 32, p. 72ff, and [V1], Coroll., p. 48), the differential formula for orthoschemes can be extended to arbitrary acute-angled polytopes. For $\tilde{R} \in \mathscr{R}_\kappa$, this can be seen analogously by subdivision into orthoschemes and application of Schläfli's formula for each of the dissecting orthoschemes. Then, collecting all differential expressions in

$d\,\mathrm{Vol}_n(\tilde{R})$ suitably, one obtains the generalized Schläfli formula for complete hyperbolic orthoschemes in terms of the dihedral angles.

Moreover, the Theorem remains true for the Schläfli function on the set of acute-angled polytopes of fixed schematic type (see *1.3*) in H^n, $n \geqq 2$, which are non-compact but of finite volume (for $n=3$, we have to cut off a horospherical neighborhood around each vertex at infinity before evaluating Schläfli's function on the apex edges). By dissection, it suffices again to consider asymptotic orthoschemes, and among them only those of dimension $n \geqq 3$. Since an orthoscheme in H^n has at most the two principal vertices at infinity (see *1.4*), there are only d-asymptotic orthoschemes, $0 \leqq d \leqq 2$, and their congruence classes are described by n essential angles $\alpha_1, \ldots, \alpha_n$ ($n-d$ of them form a system of independent parameters). Furthermore, each d-asymptotic orthoscheme may be interpreted as limiting polytope of a sequence of compact orthoschemes. Hence, by analyticity of the volume function, Schläfli's differential formula (13) holds for a family of d-asymptotic orthoschemes in H^n. We formulate this result only partially in the following sense:

Corollary. *Let F_n, $n \geqq 4$, be the Schläfli function on the set of complete orthoschemes of schematic type ζ and of finite volume in H^n with essential angles $\alpha_1, \ldots, \alpha_{m(\zeta)}$ and with scheme Σ_ζ. Denote by $F_{n-2}(k)$ the Schläfli function of the apex of codimension 2 associated to the essential angle α_k of measure $f_1(k) := f_1(\alpha_k)$, $1 \leqq k \leqq m(\zeta)$. Then*

$$dF_n(\Sigma_\zeta) = \sum_{k=1}^{m(\zeta)} F_{n-2}(k)\, df_1(k). \tag{14}$$

3 The reduction formula

3.1 Generalizing Schläfli's method for spherical orthoschemes, we prove the following reduction formula for even dimensional hyperbolic orthoschemes of degree d, $0 \leqq d \leqq 2$, in terms of the modified Schläfli function (for complete orthoschemes of dimension four, this is already proved in [K 1], §4, by a different method):

Theorem (Reduction formula). *Denote by $R_d \subset H^{2n}$, $0 \leqq d \leqq 2$, $n \geqq 1$, a $2n$-dimensional orthoscheme of degree d with scheme Σ_d. Then*

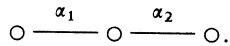
$$F_{2n}(\Sigma_d) = \sum_{k=0}^{n} \frac{(-1)^k}{k+1} \binom{2k}{k} \sum_{\sigma} f_{2n-(2k+1)}(\sigma), \qquad \sum f_{-1} := 1, \tag{15}$$

where σ runs through all elliptic subschemes of order $2(n-k)$ of Σ_d all of whose components are of even order.

Proof. For $0 \leqq d \leqq 2$ and $n \geqq 1$, let R_d be a compact orthoscheme of degree d and dimension $2n$. Proving the formula (15) we proceed by induction on the dimension.

For $n=1$, we have by definition (see (8) and (12)) that $F_2(\Sigma_d) = -\frac{2}{\pi}\mathrm{Vol}_2(R_d)$ and $f_1(\alpha) := f_1(\sigma) = \frac{2}{\pi}\alpha$, if σ is of weight $\cos\alpha$. By varying the degree d of R_d, we obtain the following cases (see *1.4*):

0. For $d=0$, $R_0=R_0(\alpha_1, \alpha_2)$ is a right-angled triangle with essential angles $0 \leqq \alpha_1, \alpha_2 < \frac{\pi}{2}$ and with area $\mathrm{Vol}_2(R_0)=\frac{\pi}{2}-(\alpha_1+\alpha_2)$. On the other hand, its scheme Σ_0 is given by (using the notation of *1.3*)

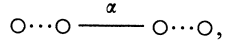


Thus, the formula (15) reads:

$$F_2(\Sigma_0)=f_1(\alpha_1)+f_1(\alpha_2)-1,$$

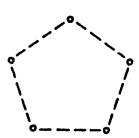
which is the above area formula for a 2-orthoscheme in terms of the Schläfli functions.

1. For $d=1$, $R_1=R_1(\alpha)$ is a Lambert quadrilateral $\left(\text{quadrilateral with three right angles and one acute angle } \alpha,\ 0 \leqq \alpha < \frac{\pi}{2}\right)$ of area $\mathrm{Vol}_2(R_1)=\frac{\pi}{2}-\alpha$. Since its scheme Σ_1 is of the form (see *1.3*)



the formula (15) evaluates as $F_2(\Sigma_1)=f_1(\alpha)-1$.

2. For $d=2$, R_2 is a totally right-angled pentagon of constant area $\mathrm{Vol}_2(R_2)=\frac{\pi}{2}$. The scheme Σ_2 of R_2 is given by a cycle (see *1.3*)



Obviously, Σ_2 contains no elliptic subscheme of order >1. Hence, the formula (15) yields $F_2(\Sigma_2)=-1$ as required.

Now, assume that the assertion holds for complete orthoschemes of even dimensions $<2n$. To show that the volume of a $2n$-dimensional orthoscheme R_d of degree d, $0 \leqq d \leqq 2$, satiesfies the formula (15), we interpret the essential angles $\alpha_1, \ldots, \alpha_m$ of R_d as independent variables (keeping all notations as before) and differentiate (15) with respect to the measure $f_1(l) := f_1(\alpha_l)$, $1 \leqq l \leqq m$. Schläfli's differential formula for hyperbolic complete orthoschemes (see *2.2*, Coroll.) has the following effect on the left hand side of (15):

$$\frac{\partial}{\partial f_1(l)} F_{2n}(\Sigma_d)=F_{2n-2}(l), \tag{16}$$

where, by respecting the dependence among $\alpha_1, \ldots, \alpha_m$, $F_{2n-2}(l)$ denotes the normalized volume of the apex $R_d(l)$ corresponding to the angle α_l of measure $f_1(l)$ (see *1.4*). Since $R_d(l)$ is itself a complete orthoscheme of dimension $2(n-1)$ (see *1.4*), we have by induction for the Schläfli function of its scheme $\Sigma_d(l)$:

$$F_{2n-2}(\Sigma_d(l)) = \sum_{k=0}^{n-1} (-1)^k a_k \sum_{\sigma'(l)} f_{2n-2-(2k+1)}(\sigma'(l)), \tag{17}$$

where

$$a_k := \frac{1}{k+1}\binom{2k}{k},$$

and where $\sigma'(l)$ runs through all elliptic subschemes of order $2(n-1-k)$ of $\Sigma_d(l)$ all of whose components are of even order. The differentiation of the right hand side of (15) with respect to $f_1(l)$ leads together with Schläfli's differential formula for spherical orthoschemes to

$$\frac{\partial}{\partial f_1(l)} \sum_{k=0}^{n} (-1)^k a_k \sum_{\sigma} f_{2n-(2k+1)}(\sigma) = \sum_{k=0}^{n-1} (-1)^k a_k \sum_{\sigma''(l)} f_{2n-2-(2k+1)}(\sigma''(l)), \tag{18}$$

where $\sigma''(l)$ corresponds to the apex of the angle α_l in the orthoscheme with scheme σ in the summation (15). Hence, each $\sigma''(l)$ is an elliptic scheme of even order $|\sigma|-2$. Moreover, all components $\sigma_1, \ldots, \sigma_r$, $r \geqq 1$, of $\sigma''(l)$ are of even order: In fact, the differentiation of $f_{2n-(2k+1)}(\sigma)$ with respect to $f_1(l)$ affects only one, say σ_ν, of the components of σ all of them being of even order; its associated spherical orthoscheme R_ν has therefore no essential angle equal to $\frac{\pi}{2}$, and the same holds for each of its faces of codimension 2 (see [BH], §4.2, Satz 3, and §4.3, Satz 1). Hence, the apex orthoscheme of R_ν associated to the angle α_l is described by a connected elliptic scheme of even order $|\sigma_\nu|-2$.

Now, every subscheme $\sigma'(l)$ of order $2i$, $1 \leqq i \leqq n-1$, in the summation (17) occurs as a subscheme $\sigma''(l)$ of order $2i$ in the summation (18), and vice versa (see *1.4*). Hence, by induction hypothesis, we proved (15) up to the value of

$$\frac{(-1)^n}{n+1}\binom{2n}{n} \cdot \sum f_{-1}.$$

It remains to show that this constant of integration, written in the form $(-1)^n a_n$ (see (15)), is given by

$$(-1)^n a_n = \frac{(-1)^n}{n+1}\binom{2n}{n}. \tag{19}$$

We check (19) first for the case $d=0$ considering the following scheme of order $2n+1$

$$\Sigma_{2n}(\varepsilon, \varepsilon') : \overset{\varepsilon'}{\circ\!\!-\!\!\!-\!\!\circ}\overset{\frac{\pi}{2}-\varepsilon}{-\!\!\circ\!\!-\!\!\cdots\!\!-\!\!\circ\!\!-\!\!}\overset{\frac{\pi}{2}-\varepsilon}{\circ\!\!-\!\!\!-\!\!\circ}, \qquad \varepsilon, \varepsilon' > 0,$$

with

$$\Sigma_k(\varepsilon) : \overset{\frac{\pi}{2}-\varepsilon}{\circ\!\!-\!\!\!-\!\!\circ\!\!-\!\!\cdots\!\!-\!\!\circ\!\!-\!\!\!-\!\!\circ}, \qquad k = 1, \ldots, 2n-1,$$

among the subschemes of order $k+1$. The determinants of these schemes satisfy the following recursion formulas (see [S], No. 27, p. 257):

$$\det \Sigma_{2n}(\varepsilon, \varepsilon') = \det \Sigma_{2n-1}(\varepsilon) - \cos^2 \varepsilon' \det \Sigma_{2n-2}(\varepsilon) \tag{20}$$

$$\det \Sigma_k(\varepsilon) = \det \Sigma_{k-1}(\varepsilon) - \sin^2 \varepsilon \det \Sigma_{k-2}(\varepsilon), \quad k=2, \ldots, 2n-1, \tag{21}$$

where $\det \Sigma_0 := 1$ and $\det \Sigma_1 = \cos^2 \varepsilon$. From (21) we derive that (see [S], No. 28, p. 265)

$$\det \Sigma_k(\varepsilon) = \frac{1}{2^{k+2}} \frac{(1+\sqrt{1-4\sin^2\varepsilon})^{k+2} - (1-\sqrt{1-4\sin^2\varepsilon})^{k+2}}{\sqrt{1-4\sin^2\varepsilon}}, \quad k \geqq 1,$$

i.e., $\det \Sigma_k(\varepsilon) > 0$, and by (21), $\det \Sigma_k(\varepsilon) < \det \Sigma_{k-1}(\varepsilon)$ for $\varepsilon < \frac{\pi}{6}$ and $k \geqq 1$. Furthermore, we can represent

$$\frac{\det \Sigma_k(\varepsilon)}{\det \Sigma_{k-1}(\varepsilon)}, \quad k=1, \ldots, 2n-1,$$

as a finite continued fraction with k partial quotients (see [S], No. 27, p. 258) using the classical notation of Pringsheim:

$$\frac{\det \Sigma_k(\varepsilon)}{\det \Sigma_{k-1}(\varepsilon)} = 1 - \cfrac{\sin^2\varepsilon}{1 - \cfrac{\sin^2\varepsilon}{\ddots - \cfrac{\sin^2\varepsilon}{1-\sin^2\varepsilon}}} = 1 - \underbrace{\frac{\sin^2\varepsilon|}{|1} - \cdots - \frac{\sin^2\varepsilon|}{|1}}_{k \text{ times}}.$$

Hence, by (20), we can choose ε' in terms of $\varepsilon \in \left(0, \frac{\pi}{6}\right)$ in such a way that $\det \Sigma_{2n}(\varepsilon, \varepsilon') < 0$, $\det \Sigma_{2n-1}(\varepsilon, \varepsilon') > 0$, and that $\varepsilon'(\varepsilon) \to 0$ for $\varepsilon \to 0$; e.g., let $\varepsilon' \in \left(0, \frac{\pi}{2}\right)$ with

$$\cos^2 \varepsilon' = 1 - \underbrace{\frac{\sin^2\varepsilon|}{|1} - \cdots - \frac{\sin^2\varepsilon|}{|1}}_{2n-2 \text{ times}} - \frac{\sin^2\varepsilon|}{|2}, \quad 0 < \varepsilon < \frac{\pi}{6}\,!$$

Thus, $\Sigma_{2n}(\varepsilon, \varepsilon')$ describes a compact hyperbolic $2n$-orthoscheme $R_{2n}(\varepsilon)$ (see *1.2*) which, for $\varepsilon \to 0$, converges to a degenerate orthoscheme R_e with angles $0, \frac{\pi}{2}, \ldots, \frac{\pi}{2}$ and with scheme

$$\Sigma_e : \overset{\infty}{\circ\!\!-\!\!-\!\!-\!\!\circ}\ \circ\ \circ\ \cdots\ \circ\ \circ.$$

Since $\det \Sigma_e = 0$, the $n+1$ normal vectors in $E^{1,n}$ associated to the bounding hyperplanes of R_e are linearly dependent. Therefore, we have in the limit a decrease of dimension, i.e., $\mathrm{Vol}_{2n}(R_e) = 0$. Geometrically, this can be seen by observing that, for $\varepsilon \to 0$, the vertex orthoscheme of $R_{2n}(\varepsilon)$, described by $\Sigma_{2n-1}(\varepsilon)$, converges to a totally rectangular spherical orthoscheme with edge lengths $\frac{\pi}{2}$

(see [BH], Folgerung, p. 82); this implies that $R_{2n}(\varepsilon)$ has a triangular face of area $\Delta(\varepsilon)=\frac{\pi}{2}-(h(\varepsilon)+\varepsilon')>0$ (see [BH], Satz 1, p. 109), where h is a continuous positive function with $h(\varepsilon)\to\frac{\pi}{2}$ for $\varepsilon\to 0$. Hence, we have $\Delta(\varepsilon)\to 0$ for $\varepsilon\to 0$ implying that at least two vertices of the limiting polytope coincide. Hence, (15) yields:

$$0=F_{2n}(\Sigma_e)=\sum_{k=0}^{n-1}\frac{(-1)^k}{k+1}\binom{2k}{k}\sum_{\sigma} f_{2n-(2k+1)}(\sigma)+(-1)^n a_n,$$

where σ runs through all elliptic subschemes of Σ_e of order $2(n-k)$ all of whose components are of even order. This condition implies the following identity (for a tricky proof, see [S], p. 255–256, and *2.1*):

$$0=\sum_{k=0}^{n-1}\frac{(-1)^k}{k+1}\binom{2k}{k}\binom{n+k}{2k}+(-1)^n a_n.$$

Since (see [S], p. 256)

$$\sum_{k=0}^{n}\frac{(-1)^k}{k+1}\binom{2k}{k}\binom{n+k}{2k}=0,$$

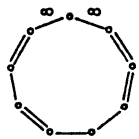
we deduce (19).

For $d>0$, we remark first that there are orthoschemes $R_d\subset H^n$, $n\geqq 3$, of degree d of different schematic type; although the graphical type and the order of the associated schemes Σ_d are constant, the weights or, more precisely, the number of subschemes of one character of definiteness may differ. Geometrically, this means that orthoschemes of degree d of different schematic type are bounded by the same number of hyperplanes, but in view of their mutual position in H^n, the number of (finite or infinite) vertices may differ. However, the sets of orthoschemes of degree d of different schematic type are not disjoint. Indeed, their intersections consist of the asymptotic complete orthoschemes. Secondly, we observe that a doubly asymptotic orthoscheme of degree 0 (both principal vertices are points at infinity) can be interpreted as asymptotic limiting case of an orthoscheme of degree d (the polar hyperplanes associated to the principal vertices touch the absolute quadric and truncate therefore without effect). Hence, by comparison in the appropriate asymptotic limiting cases, one immediately sees that the constant of integration always equals

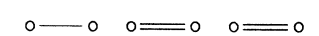
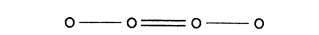
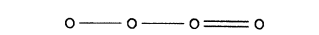
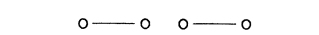
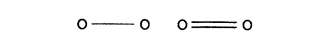
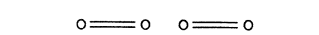
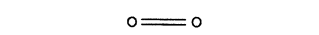
$$\frac{(-1)^n}{n+1}\binom{2n}{n}.$$

Q.E.D.

3.2 The Reduction formula (15) can now be applied to all complete Coxeter orthoschemes (for dimensions $2n\geqq 4$, see Appendix). We want to give one example in detail: Consider the following cyclic scheme Σ of order 9 describing a 6-dimensional non-compact Coxeter orthoscheme R of degree 2:



Determining the volume $\mathrm{Vol}_6(R)$ according to (15), we have to pick out all different elliptic subschemes σ of Σ of even order $6-2k$, $k=0, 1, 2$, together with their multiplicities $\mu(\sigma)$, and evaluate $f_{6-(2k+1)}$ on σ (see 2.1, Lemma and (8)–(11)):

k	$\mu(\sigma)$	σ	$f_{6-(2k+1)}(\sigma)$
0	1		$\frac{1}{144}$
	1		$\frac{1}{48}$
	1		$\frac{1}{6}$
1	1		$\frac{1}{72}$
	2		$\frac{1}{24}$
	1		$\frac{4}{9}$
	5		$\frac{1}{3}$
	4		$\frac{1}{4}$
2	3		$\frac{2}{3}$
	4		$\frac{1}{2}$

Hence, according to (15), we obtain:

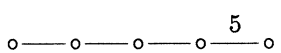
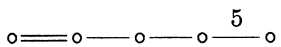
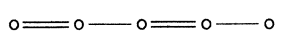
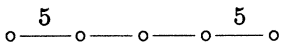
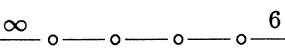
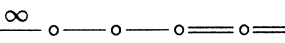
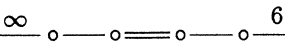
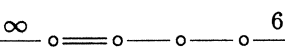
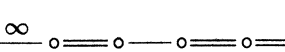
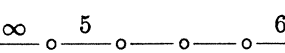
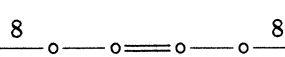
$$\begin{aligned} F_6(\Sigma)=(-1)^3 c_6 \,\mathrm{Vol}_6(R) &= \sum_{k=0}^{3} \frac{(-1)^k}{k+1}\binom{2k}{k} \sum f_{6-(2k+1)} \\ &= \sum f_5 - \sum f_3 + 2\cdot\sum f_1 - 5\cdot\sum f_{-1} \\ &= \frac{7}{36} - \frac{77}{24} + 2\cdot 4 - 5 \\ &= -\frac{1}{72}. \end{aligned}$$

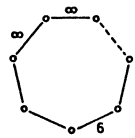
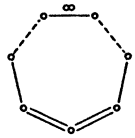
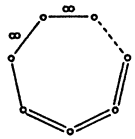
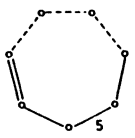
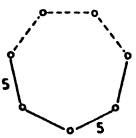
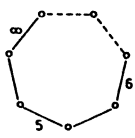
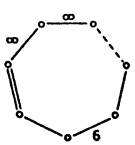
Since $c_6 = \dfrac{120}{\pi^3}$, we get:

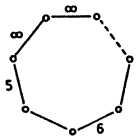
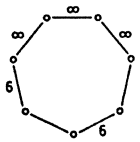
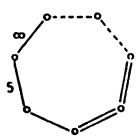
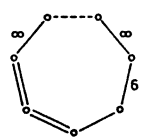
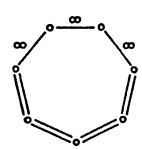
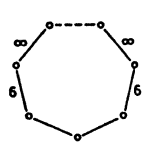
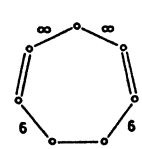
$$\mathrm{Vol}_6(R) = -\frac{1}{c_6}\cdot F_6(\Sigma) = \frac{\pi^3}{8'640} \simeq 0.0036.$$

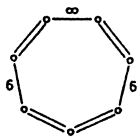
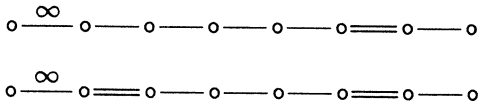
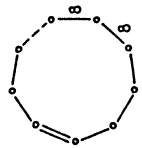
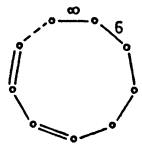
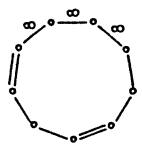
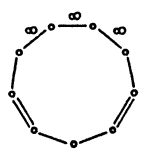
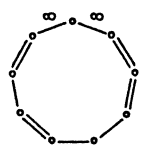
Appendix

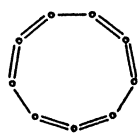
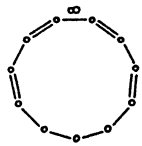
The complete Coxeter orthoschemes $R_{d,C}$ of degree d, $0 \leqq d \leqq 2$, were classified by Im Hof in 1983 (see [IH]). They exist only up to dimension 9; in each dimension $\geqq 4$, there are finitely many examples. We now list the graphs Σ_d and volumes $\mathrm{Vol}_{2n}(R_{d,C})$ of all complete Coxeter orthoschemes of degree d and of even dimensions $2n \geqq 4$:

$2n$	d	Σ_d	$\mathrm{Vol}_{2n}(R_{d,C})$
4	0		$\frac{\pi^2}{10'800} \simeq 0.0009$
			$\frac{17\pi^2}{21'600} \simeq 0.0078$
			$\frac{\pi^2}{864} \simeq 0.0114$
			$\frac{13\pi^2}{5'400} \simeq 0.0238$
4	1		$\frac{\pi^2}{540} \simeq 0.0183$
			$\frac{\pi^2}{288} \simeq 0.0343$
			$\frac{41\pi^2}{10800} \simeq 0.0375$
			$\frac{17\pi^2}{4'320} \simeq 0.0388$
			$\frac{5\pi^2}{864} \simeq 0.0571$
			$\frac{\pi^2}{288} \simeq 0.0343$
			$\frac{\pi^2}{144} \simeq 0.0685$
			$\frac{61\pi^2}{900} \simeq 0.6689$
			$\frac{11\pi^2}{1'728} \simeq 0.0628$

$2n$	d	Σ_d	$\mathrm{Vol}_{2n}(R_{d,C})$
4	2		$\frac{\pi^2}{108} \simeq 0.0914$
			$\frac{\pi^2}{108} \simeq 0.0914$
			$\frac{\pi^2}{72} \simeq 0.1371$
			$\frac{\pi^2}{120} \simeq 0.0823$
			$\frac{\pi^2}{90} \simeq 0.1097$
			$\frac{7\pi^2}{540} \simeq 0.1279$
			$\frac{\pi^2}{72} \simeq 0.1371$

$2n$	d	Σ_d	$\mathrm{Vol}_{2n}(R_{d,C})$
4	2		$\frac{\pi^2}{60} \simeq 0.1645$
			$\frac{\pi^2}{54} \simeq 0.1828$
			$\frac{\pi^2}{90} \simeq 0.1097$
			$\frac{\pi^2}{72} \simeq 0.1371$
			$\frac{\pi^2}{48} \simeq 0.2056$
			$\frac{\pi^2}{108} \simeq 0.0914$
			$\frac{\pi^2}{48} \simeq 0.2056$

$2n$	d	Σ_d	$\mathrm{Vol}_{2n}(R_{d,C})$
4	2		$\frac{5\pi^2}{216} \simeq 0.2285$
6	1		$\frac{11\pi^3}{86'400} \simeq 0.0039$ $\frac{\pi^3}{86'400} \simeq 0.0004$
6	2		$\frac{\pi^3}{43'200} \simeq 0.0007$
			$\frac{7\pi^3}{129'600} \simeq 0.0017$
			$\frac{\pi^3}{17'280} \simeq 0.0018$
			$\frac{389\pi^3}{25'290} \simeq 0.4653$
			$\frac{\pi^3}{8'640} \simeq 0.0036$

$2n$	d	Σ_d	$\mathrm{Vol}_{2n}(R_{d,C})$
6	2		$\frac{\pi^3}{3'240} \simeq 0.0096$
8	2		$\frac{569\pi^4}{43'545'600} \simeq 0.0013$

Acknowledgements. The author would like to express her gratitude to J. Böhm and H. Debrunner for helpful suggestions and remarks.

References

[A1] Andreev, E.M.: On Convex Polyhedra in Lobačevskii Spaces. Math. USSR Sb. **10** No. 3, 413–440 (1970)
[A2] Andreev, E.M.: The Intersection of the Plane Boundaries of a Polytope with Acute Angles. Math. Notes **8**, 761–764 (1970)
[BH] Böhm, J., Hertel, E.: Polyedergeometrie in n-dimensionalen Räumen konstanter Krümmung. Basel: Birkhäuser 1981
[C] Coxeter, H.S.M.: Discrete Groups Generated by Reflections. Ann. Math. II. Ser. **35**, 588–621 (1934)
[IH] Im Hof, H.-C.: A Class of Hyperbolic Coxeter Groups. Expo. Math. **3**, 179–186 (1985)
[H] Hopf, H.: Die Curvatura integra Clifford-Kleinscher Raumformen. Nachr. Ges. Wiss. Gött., Math.-Phys. Kl., 1928, 131–141 (1925)
[K1] Kellerhals, R.: Über den Inhalt hyperbolischer Polyeder in den Dimensionen drei und vier. Dissertation, Basel 1988
[K2] Kellerhals, R.: On the Volume of Hyperbolic Polyhedra. Math. Ann. **285**, 541–569 (1989)
[Kn] Kneser, H.: Der Simplexinhalt in der nichteuklidischen Geometrie. Dt. Math. **1**, 337–340 (1936)
[S] Schläfli, L.: Theorie der vielfachen Kontinuität. In: Gesammelte Mathematische Abhandlungen **1**. Basel: Birkhäuser 1950
[V1] Vinberg, È.B.: Hyperbolic Reflection Groups. Russ. Math. Surv. **40**, 31–75 (1985)
[V2] Vinberg, È.B.: The Absence of Crystallographic Groups of Reflections in Lobachevsky Spaces of Large Dimension. Trans. Mosc. Math. Soc. **47**, 75–112 (1985)

Math. Z. 206, 211–217 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Äquivariante Kompaktifizierungen des $\mathbb{C}^n$

Christoph Gellhaus

Mathematisches Institut, Ruhr-Universität, Universitätsstraße 150, W-4630 Bochum 1, Federal Republic of Germany

Received August 2, 1989; in final form February 2, 1990

Eine kompakte komplexe Mannigfaltigkeit X heißt *fast-homogen*, wenn eine abgeschlossene komplexe Untergruppe G der Gruppe der holomorphen Automorphismen $\mathrm{Aut}_{\mathcal{O}}(X)$ eine offene Bahn $\Omega := G/H$ in X hat. Somit ist Ω eine dichte, offene und zusammenhängende Untermannigfaltigkeit von X, so daß das Komplement $E := X \setminus \Omega$ eine analytische Teilmenge von X ist. Eine vollständige Klassifikation fast-homogener Mannigfaltigkeiten X ist weitgehend aussichtslos, sogar in speziellen Fällen wie z.B. dem der komplex symmetrischen Räumen im Sinne von A. Borel [B]. Unter zusätzlichen Annahmen, z.B. bei Bedingungen an die Dimension von X, an die analytische Menge E oder an Ω, gewinnt man Resultate (siehe z.B. [HO1, HO2, O1 und O2]).

In dieser Arbeit befassen wir uns mit der Frage nach der Existenz einer maximalen Anzahl analytisch = algebraisch unabhängiger meromorpher Funktionen auf fast-homogenen kompakten komplexen Mannigfaltigkeiten und untersuchen dies unter topologischen Bedingungen an die offene Bahn Ω:

Wenn alle Homotopiegruppen der offenen Bahn verschwinden, d.h. wenn Ω "asphärisch" ist, ist die offene Bahn schon biholomorph zu $\mathbb{C}^n$ äquivalent [HO2]. Solche X nennen wir *äquivariante Kompaktifizierung* von $\mathbb{C}^n$. Wir zeigen:

Satz. *Es sei G eine komplexe Liegruppe, die auf einer kompakten komplexen Mannigfaltigkeit X fast-homogen wirkt, so daß die offene Bahn G/H asphärisch ist.*

Dann ist X eine rationale Moišezon-Mannigfaltigkeit, also bimeromorph zu $\mathbb{P}_n(\mathbb{C})$ äquivalent. Die offene Bahn $G/H \cong \mathbb{C}^n$ ist dann affin-algebraisch in X, und es existiert eine G-äquivariante bimeromorphe Abbildung $\varphi_L \colon X \dashrightarrow Y \subset \mathbb{P}(H^0(X, L)^*)$ für ein G-stabiles holomorphes Geradenbündel L auf X.

1 Meromorphe Funktionen auf fast-homogenen Mannigfaltigkeiten

Es sei X eine kompakte komplexe Mannigfaltigkeit, die bezüglich einer zusammenhängenden komplexen Liegruppe G fast-homogen ist. Es sei $\tilde{X}$ eine algebraische Reduktion von X, so daß die Abbildung p holomorph ist:

$$\begin{array}{ccc} \tilde{X} & & \text{mit} \quad a(\tilde{X}) = a(X) \\ \downarrow p & & \\ V & & \text{mit} \quad a(V) = a(\tilde{X}) = a(X) \end{array}$$

Es sei $x_0 \in X$ ein allgemeiner Punkt in der offenen Bahn $\Omega := G/H$ mit $H = \{g \in G \mid g(x_0) = x_0\}$ und

$$A(x_0) := \bigcap \{D \subset X \mid D \text{ ist Hyperfläche in } X, \text{die } x_0 \text{ enthält}\}.$$

Dann bezeichne F die x_0 enthaltende irreduzible Komponente von $A(x_0)$. Es ist X zu $\tilde{X}$ bimeromorph äquivalent, und die allgemeine Faser $\tilde{F}$ von p ist eine zusammenhängende Mannigfaltigkeit, die in X als J-Kompaktifizierung F realisiert werden kann mit

$$J := \{g \in G \mid g(F) = F\}.$$

In der Tat gilt für Punkte $x, y \in F \cap G/H$ mit $g(x) = y$, daß g die analytische Menge F stabilisiert. Somit gilt $H < J$ und J wirkt fast-transitiv auf F. Der Schnitt von F mit G/H liefert durch Translation mit Automorphismen aus G eine Zerlegung von G/H, die durch ein G-äquivariantes holomorphes Faserbündel π: $G/H \to G/J$ gegeben wird. Die Universalität dieser Faserung nutzen wir in einer speziellen Situation:

Die $\mathcal{G}$-antikanonische Faserung [HO1] der offenen Bahn G/H ist meromorph und G-äquivariant fortsetzbar; denn die Abbildung wird durch die Schnitte im antikanonischen Bündel auf X gegeben, die durch die Liealgebra $\mathcal{G}$ induziert werden:

$$\begin{array}{ccccc} G/H & \longrightarrow & G/N \subset \mathbb{P}_m(\mathbb{C}) & \text{mit} & N := N_G(H^0) \\ \cap & & \cap & & \\ X & \longrightarrow & Y \subset & & \mathbb{P}_m(\mathbb{C}). \end{array}$$

Da die Hyperflächen in G/N dort die Punkte trennen, erhält man auf der offenen Bahn das folgende kommutative Diagramm:

$$\begin{array}{ccc} X & \longrightarrow & Y \\ \cup & & \cup \\ G/H & \longrightarrow & G/N \\ & \searrow \quad \nearrow & \\ & G/J & \end{array}.$$

Dann ist F eine J-äquivariante Kompaktifizierung, so daß J/H parallelisierbar ist, da $J < N$ gilt: $J/H = (J/H^0)/(H/H^0) = L/\Gamma$ wobei L eine komplexe Liegruppe mit diskreter Untergruppe Γ ist.

Nun beginnen wir mit dem einfachsten Fall: $H = H^0$, d.h. F ist eine äquivariante Gruppenkompaktifizierung von J/H. Dabei ist zu erwähnen, daß äquivariante Kompaktifizierungen von L/Γ mit diskretem Γ schon im dreidimensionalen Fall für "fast keine" Γ existieren [M]. Für die nicht notwendigerweise glatte J-äquivariante Kompaktifizierung F gilt:

Lemma 1.1. *Es sei* $\tau\colon \hat{F} \to F$ *eine J-äquivariante Desingularisierung* [Hi] *von F mit Randdivisor* $\hat{E} := \tau^*(E|_F)$, *wobei E der Randdivisor in X ist. Dann gilt für die Iitaka-Dimension:* $\mathcal{K}(\hat{F}, \hat{E}) \leqq 0$.

Beweis. Es sei nun $p\colon \tilde{X} \to V$ eine algebraische Reduktion von X, so daß die Abbildung p holomorph ist. Dabei bezeichne $\Psi\colon \tilde{X} \to X$ eine eigentliche Modifikation, die eingeschränkt auf eine generische glatte zusammenhängende Faser $\tilde{F}$ von p eine eigentliche Modifikation auf eine J-äquivariante Kompaktifizierung F liefert: $\psi := (\Psi|_{\tilde{F}})\colon \tilde{F} \to F$. Nun ist $\hat{F}$ eine glatte J-äquivariante Gruppenkompaktifizierung mit Randdivisor $\hat{E} = \tau^*(E|_F)$. Aber $\tilde{E} := \psi^*(E|_F) = (\Psi^*(E))|_{\tilde{F}}$. Somit folgt mit Theorem 12.1 aus [U]: $\mathcal{K}(\tilde{F}, \tilde{E}) \leqq 0$.

Die eigentlichen Modifikationen $\tau\colon \hat{F} \to F$ und $\psi\colon \tilde{F} \to F$ ergeben eine bimeromorphe Abbildung $\sigma\colon \hat{F} \dashrightarrow \tilde{F}$. Es sei Z der Graph dieser bimeromorphen Abbildung zwischen den Mannigfaltigkeiten $\hat{F}$ und $\tilde{F}$, so daß σ durch folgende eigentliche Modifikationen

$$\hat{\tau}\colon Z \to \hat{F} \quad \text{und} \quad \tilde{\psi}\colon Z \to \tilde{F}$$

gegeben wird. Es gilt somit: $\tau \circ \hat{\tau} = \psi \circ \tilde{\psi}$, d.h.

$$\hat{\tau}^*(\hat{E}) = \hat{\tau}^*(\tau^*(E|_F)) = \tilde{\psi}^*(\psi^*(E|_F)) = \tilde{\psi}^*(\tilde{E}).$$

Nun folgt mit Lemma 5.3 aus [U]:

$$\mathcal{K}(\hat{F}, \hat{E}) = \mathcal{K}(Z, \hat{\tau}^*(\hat{E})) = \mathcal{K}(Z, \tilde{\psi}^*(\tilde{E})) = \mathcal{K}(\tilde{F}, \tilde{E}) \leqq 0. \quad \square$$

Im halbeinfachen Fall gibt es Beispiele, bei denen F nicht Moišezon ist. Wir betrachten komplex auflösbare Gruppen, die bei den äquivarianten Kompaktifizierungen des $\mathbb{C}^n$ auftauchen [HO2].

2 Äquivariante Gruppenkompaktifizierungen

Nun untersuchen wir äquivariante Kompaktifizierungen des $\mathbb{C}^n$.

Wir können im folgenden immer annehmen, daß G eine zusammenhängende einfach-zusammenhängende komplex auflösbare Liegruppe ist [HO2]. Da H zusammenhängend ist, folgt mit langer exakter Homotopiesequenz, daß auch H einfach-zusammenhängend ist. Dann gilt:

$$G \cong \mathbb{C}^{n+k}, \quad H \cong \mathbb{C}^k \text{ und } G/H \cong \mathbb{C}^n$$

als komplexe Mannigfaltigkeiten.

Zuerst untersuchen wir G-äquivariante *Gruppenkompaktifizierungen* von $\mathbb{C}^n$, d.h. $H = \{e\}$. Dann existiert eine Basis

$$\Theta_1, \ldots, \Theta_n \in H^0(G, T^{1,0}G)$$

der rechts-invarianten Vektorfelder auf G, die sich wegen der Links-Wirkung von G auf X als holomorphe Vektorfelder auf X fortsetzen lassen, so daß die dazu dualen rechts-invarianten holomorphen 1-Formen

$$\omega_1, \ldots, \omega_n \in H^0(G, \Omega^1)$$

für auflösbare Gruppenstruktur den folgenden Bedingungen [N] genügen:

(1) $\quad d\omega_1 = \ldots = d\omega_r = 0 \qquad$ für $1 \leqq r \leqq n$

(2) $\quad d\omega_i = \sum_{1 \leqq j < k \leqq i} c_{jk}^i \cdot \omega_j \wedge \omega_k \qquad$ für $r+1 \leqq i \leqq n.$

Wenn G eine nilpotente Gruppenstruktur besitzt, gilt entsprechend:

(1') $\quad d\omega_1 = \ldots = d\omega_r = 0 \qquad$ für $2 \leqq r \leqq n$

(2') $\quad d\omega_i = \sum_{1 \leqq j < k < i} c_{jk}^i \cdot \omega_j \wedge \omega_k \qquad$ für $r+1 \leqq i \leqq n.$

Alles gilt für geeignete Strukturkonstanten $c_{jk}^i \in \mathbb{C}$.

Lemma 2.1. *Es sei X eine nilpotente G-äquivariante Gruppenkompaktifizierung von $G \cong \mathbb{C}^n$. Dann gilt für die Iitaka-Dimension des antikanonischen Bündels auf X:*

$$\varkappa(X, K_X^{-1}) \geqq r \geqq 2.$$

Beweis. Für eine G-äquivariante Gruppenkompaktifizierung X sind auch die rechts-invarianten holomorphen 1-Formen $\omega_1, \ldots, \omega_n$ als meromorphe 1-Formen auf X fortsetzbar, so daß wir in der Notation nicht zwischen den Θ_i bzw. den ω_i und deren Fortsetzungen unterscheiden werden: Es definiert $\Theta_1 \wedge \ldots \wedge \Theta_n \in H^0(X, K_X^{-1})$ einen holomorphen Schnitt im antikanonischen Bündel auf X, der auf der offenen Bahn G nicht verschwindet, aber in einem Punkt des Randdivisors E. Definiere nun die Fortsetzung der ω_i auf einem $\Theta \in H^0(X, T^{1,0}X)$ wie folgt

$$\omega_i(\Theta) := (\Theta_1 \wedge \ldots \wedge \Theta_{i-1} \wedge \Theta \wedge \Theta_{i+1} \wedge \ldots \wedge \Theta_n) \cdot (\Theta_1 \wedge \ldots \wedge \Theta_n)^{-1}.$$

Dies liefert als Produkt eines meromorphen Schnittes im kanonischen Bündel mit einem holomorphen Schnitt im antikanonischen Bündel eine meromorphe Funktion auf X, die eingeschränkt auf G holomorph ist. Auf G gilt natürlich: $\omega_i(\Theta_j) = \delta_{ij}$.

Für die $r \geqq 2$ geschlossenen meromorphen 1-Formen $\omega_1, \ldots, \omega_r$ auf X existieren meromorphe Funktionen Φ_i auf X mit $d\Phi_i = \omega_i$: Wähle z.B. $z_0 := 0 \in \mathbb{C}^n$ und wähle einen differenzierbaren Weg γ von z_0 nach $z \in X$ und setze:

$$\Phi_i(z) := \int_\gamma \omega_i.$$

Dies liefert erst einmal eine wohldefinierte holomorphe Funktion auf $\mathbb{C}^n$, die sich auch meromorph auf X fortsetzen läßt, da X und $\mathbb{C}^n$ einfach-zusammenhängend sind.

Nach Konstruktion der Φ_i ergibt sich, daß $\Phi_1, \ldots, \Phi_r$ direkt aus den geschlossenen meromorphen 1-Formen $\omega_1, \ldots, \omega_r$ durch Integration gewonnen werden können. Im antikanonischen Bündel auf X haben wir einen holomorphen Schnitt $s_0 := \Theta_1 \wedge \ldots \wedge \Theta_n \in H^0(X, K_X^{-1})$.

Der Träger des Nullstellendivisors von s_0 ist gleich dem Träger des Randdivisors E der Kompaktifizierung X. Definiere nun auf X meromorphe Schnitte im antikanonischen Bündel: $s_i := \Phi_i \cdot (\Theta_1 \wedge \ldots \wedge \Theta_n) = \Phi_i \cdot s_0$ für $i = 1, \ldots, r$. Wegen

$s_0 \cdot \omega_i \in H^0(X, \Omega^1 \otimes K_X^{-1})$ und der Konstruktion der Φ_i durch Integration der ω_i gilt: $s_i = s_0 \cdot \Phi_i \in H^0(X, K_X^{-1})$. Somit sind $s_1, \ldots, s_r$ in der Tat holomorphe Schnitte im antikanonischen Bündel auf X, und es gilt:

$$s_0, s_1, \ldots, s_r \in H^0(X, K_X^{-1}) \quad \text{mit} \quad \Phi_1 = s_1/s_0, \ldots, \Phi_r = s_r/s_0. \quad \square$$

Proposition 2.2. *Es sei X eine auflösbare G-äquivariante Gruppenkompaktifizierung von $G \cong \mathbb{C}^n$. Dann ist X Moišezon and für die Iitaka-Dimension des antikanonischen Bündels auf X gilt:*

$$\mathcal{K}(X, K_X^{-1}) \geqq r \geqq 1.$$

Beweis. Man hat im auflösbaren Fall mindestens $r \geqq 1$ geschlossene meromorphe 1-Formen $\omega_1, \ldots, w_r$, die r analytisch unabhängige meromorphe Funktionen $\Phi_1, \ldots, \Phi_r$ auf X liefern, d.h. für die algebraische Dimension von X gilt: $a(X) \geqq r \geqq 1$. Analog zu dem nilpotenten Fall folgt auch:

$$\mathcal{K}(X, K_X^{-1}) \geqq r \geqq 1.$$

Es gilt $\{\Phi_1 = \ldots = \Phi_r = 0\} \cap G \cong G'$, d.h. für die in Kapitel 1 betrachtete J-äquivariante Gruppenkompaktifizierung F gilt: J ist eine positiv-dimensionale Untergruppe von G', d.h. J ist nilpotent, und F ist eine nilpotente Gruppenkompaktifizierung von J.

Nun betrachten wir wie in Lemma 1.1 zu F eine glatte J-äquivariante Gruppenkompaktifizierung $\hat{F}$ mit dem Randdivisor $\hat{E} = \hat{E}_1 + \ldots + \hat{E}_m$ (Zerlegung in irreduzible Komponenten). Da die Vektorfelder, die durch die Gruppenwirkung auf $\hat{F}$ induziert werden, auf dem Randdivisor einen dort verschwindenden antikanonischen Schnitt liefern, gilt mit positiven ganzen Zahlen n_i, für $i = 1, \ldots, m$:

$$K_{\hat{F}}^{-1} = n_1 \cdot \hat{E}_1 + \ldots + n_m \cdot \hat{E}_m.$$

Nach Lemma 2.1 gilt dann: $1 \leqq \mathcal{K}(\hat{F}, K_{\hat{F}}^{-1})$, d.h. es existiert eine positive ganze Zahl k, so daß für zwei holomorphe Schnitte $s_1, s_0 \in H^0(\hat{F}, K_{\hat{F}}^{-k})$ der Quotient s_1/s_0 eine nicht-konstante meromorphe Funktion auf $\hat{F}$ liefert. Es seien $\hat{e}_1, \ldots, \hat{e}_m$ definierende Schnitte für $\hat{E}_1, \ldots, \hat{E}_m$. Für $i = 0, 1$ betrachte man folgende holomorphe Schnitte

$$t_i := s_i \otimes \hat{e}_1^{k \cdot (n_2 + \ldots + n_m)} \otimes \ldots \otimes \hat{e}_m^{k \cdot (n_1 + \ldots + n_{m-1})}$$

in $H^0(\hat{F}, k \cdot (n_1 + \ldots + n_m) \cdot \hat{E})$. Nun liefert auch $t_1/t_0 = s_1/s_0$ eine nicht-konstante meromorphe Funktion auf $\hat{F}$. Somit gilt $1 \leqq \mathcal{K}(\hat{F}, \hat{E})$ im Widerspruch zu Lemma 1.1. $\square$

3 Äquivariante Kompaktifizierungen des $\mathbb{C}^n$

Satz 3.1. *Es sei G eine komplexe Liegruppe, die auf einer kompakten komplexen Mannigfaltigkeit X fast-homogen wirkt, so daß die offene Bahn G/H asphärisch ist.*

Dann ist X eine rationale Moišezon-Mannigfaltigkeit, also bimeromorph zu $\mathbb{P}_n(\mathbb{C})$ äquivalent. Die offene Bahn $G/H \cong \mathbb{C}^n$ ist dann affin-algebraisch in X, und es existiert eine G-äquivariante bimeromorphe Abbildung φ_L: $X \dashrightarrow Y \subset \mathbb{P}(H^0(X, L)^*)$ für ein G-stabiles holomorphes Geradenbündel L auf X.

Beweis. Weil schon das Radikal transitiv auf G/H wirkt, dürfen wir annehmen, daß G auflösbar ist und $H=H^0$ gilt. Für die in Kapitel 1 betrachtete J-äquivariante Kompaktifizierung F gilt: J ist eine positiv-dimensionale Untergruppe von $N:=N_G(H^0)=N_G(H)$, d.h. J/H ist eine auflösbare Untergruppe von G, und F ist eine auflösbare Gruppenkompaktifizierung von J/H. Nun betrachten wir wie im Beweis von Prop. 2.2 zu F eine glatte J/H-äquivariante auflösbare Gruppenkompaktifizierung $\hat{F}$ mit dem Randdivisor $\hat{E}$. Wieder liefern die Vektorfelder, die durch die Gruppenwirkung auf $\hat{F}$ induziert werden, einen auf dem Randdivisor verschwindenden antikanonischen Schnitt. Nach Prop. 2.2 gilt dann: $1 \leqq \mathscr{K}(\hat{F}, K_{\hat{F}}^{-1})$, und mit den gleichen Argumenten wie im Beweis von Prop. 2.2 gilt daher: $1 \leqq \mathscr{K}(\hat{F}, \hat{E})$ im Widerspruch zu Lemma 1.1.

Also ist X eine Moišezon-Mannigfaltigkeit, und es gilt: $H^1(X, \mathcal{O})=0$ [RV], d.h. alle holomorphen Geradenbündel sind G-stabil. Somit existiert ein G-stabiles holomorphes Geradenbündel L auf X, das eine G-äquivariante meromorphe Abbildung

$$\varphi_L\colon\ X \dashrightarrow Y \subset \mathbb{P}(H^0(X, L)^*) \simeq \mathbb{P}_N(\mathbb{C})$$

liefert. Diese Abbildung hat die Eigenschaft, daß sie auf der offenen Bahn biholomorph auf das Bild ist. Daher ist X bimeromorph zu einer G-äquivarianten Kähler-Kompaktifizierung Y äquivalent, auf der der algebraische Abschluß von G in $\mathrm{Aut}_{\mathcal{O}}(\mathbb{P}_N(\mathbb{C}))$ wirkt. Nach einem Satz von Rosenlicht-Matsumura [HO2] ist Y aber schon projektiv-rational, so daß $\mathbb{C}^n$ affin-algebraisch in X ist; denn die Stabilisatoruntergruppe des Randdivisors in Y ist eine algebraische Untergruppe von $\mathrm{Aut}_{\mathcal{O}}(\mathbb{P}_N(\mathbb{C}))$, die $\mathbb{C}^n$ als offene Bahn hat. □

Als eine unmittelbare Konsequenz von [F] ergibt sich:

Korollar 3.2. *Das Radikal R von G hat einen Fixpunkt in X. Ferner hat jedes holomorphe Vektorfeld $\Theta \in H^0(X, T^{1,0}X)$ auf jeder Θ-stabilen abgeschlossenen analytischen Menge $Y \subset X$ einen Fixpunkt.*

Korollar 3.3. *Es sei G eine komplexe Liegruppe, die auf einer kompakten komplexen Mannigfaltigkeit X mit $\dim_{\mathbb{C}} X=3$ und $b_2(X)=1$ fast-homogen wirkt, so daß die offene Bahn G/H asphärisch ist. Dann gibt es für X nur die folgenden Möglichkeiten:*

1) $X \cong \mathbb{P}_3(\mathbb{C})$
2) $X \cong Q_3 \subset \mathbb{P}_4(\mathbb{C})$
3) *X ist der Schnitt der Grassmann-Mannigfaltigkeit* Gr(2, 5) *der 2-Ebenen in $\mathbb{C}^5$, die durch Plücker-Koordinaten in $\mathbb{P}_9(\mathbb{C})$ eingebettet wird, mit einem allgmeinen $\mathbb{P}_6(\mathbb{C})$.*

Beweis. Nach Satz 3.1 ist X Moišezon, und es gilt mit [PS]: $c_1(X)>0$. Die Zahl r mit $K_X^{-1}=L^r$ für den positiven Erzeuger L von $\mathrm{Pic}(X)$ heißt der Index von X. Nun zeigen wir: $r \geqq 2$. Dies beweist nach [PS] das Korollar.

Wenn X den Index 1 hätte, dann wäre X eine äquivariante Gruppen-Kompaktifizierung des $\mathbb{C}^3$. Die Integration der Form ω_1 liefert aber auf $\mathbb{C}^3$ eine wohldefinierte holomorphe Funktion Φ_1 im Widerspruch zu der Annahme über den Index. □

Teile dieser Arbeit sind in meiner Bochumer Dissertation „Beiträge zur Theorie der kompakten komplexen Mannigfaltigkeiten mit holomorphen Vektorfeldern" enthalten. An dieser Stelle möchte ich Herrn Professor Dr. Alan T. Huckleberry für seine Unterstützung und für die Anregung zu dieser Arbeit danken.

Literatur

[B] Borel, A.: Symmetric compact complex spaces. Arch. Math. **33**, 49–56 (1979)

[F] Fujiki, A.: On automorphism groups of compact Kähler manifolds. Invent. Math. **44**, 225–258 (1978)

[Hi] Hirnonaka, H.: Bimeromorphic smoothing of a complex-analytic space. Math. Inst. Warwick Univ., UK 1971

[HO1] Huckleberry, A.T.; Oeljeklaus, E.: Classification theorems for almost homogeneous spaces. Revue de l'Institut Elie Cartan, Janvier 1984

[HO2] Huckleberry, A.T.; Oeljeklaus, E.: Homogeneous spaces from a complex analytic viewpoint. Manifolds and Lie groups. (Papers in honor of Y. Matsushima, Progress in Math) Boston: Birkhäuser 1981

[M] Massmann, B.: Äquivariante Kählersche Kompaktifizierungen von drei-dimensionalen auflösbaren komplexen Mannigfaltigkeiten. Diplomarbeit, Ruhr-Universität Bochum

[N] Nakamura, I.: Complex parallelisable manifolds and their small deformations. J. Differ. Geom. **10**, 85–112 (1975)

[O1] Oeljeklaus, E.: Über fast-homogene kompakte komplexe Mannigfaltigkeiten. Schriftenr. Math. Inst. Univ. Münster, 2. Ser. **1** (1970)

[O2] Oeljeklaus, E.: Fast-homogene Kählermannigfaltigkeiten mit verschwindender erster Bettizahl. Manuscr. Math. **7**, 175–183 (1972)

[PS] Peternell, Th.; Schneider, M.: Compactifications of $\mathbb{C}^3$. I. Math. Ann. **280**, 129–146 (1988)

[RV] Remmert, R.; Ven, A. van de: Zwei Sätze über die komplex-projektive Ebene. Nieuw. Arch. Wiskd., III. Ser. **8**, 147–157 (1960)

[U] Ueno, K.: Classification theory of algebraic varietes and compact complex spaces. (Lect. Notes Math., vol. 439) Berlin Heidelberg New York: Springer 1975

Math. Z. 206, 219–224 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

Cheeger-Goreski-MacPherson's conjecture for the varieties with isolated singularities

Takeo Ohsawa

Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606, Japan

Received July 6, 1990

1 Introduction

In [2] Cheeger-Goreski-MacPherson asked the following question. Let $V \subset \mathbb{P}^N(\mathbb{C})$ be a projective variety and let V' be the set of regular points of V. Is the L^2 de Rham cohomology group of V' with respect to the Fubini-Study metric canonically isomorphic to the intersection cohomology group of V? In virtue of the works of Cheeger [1], Hsian-Pati [3] and Nagase [4], we know that the conjecture is true if $\dim V \leqq 2$. Our proof will be based on the previous work [5] which complements a result of L. Saper asserting that there exists a complete Kähler metric ds^2 on V' such that the L^2 de Rham cohomology group of V' with respect to ds^2 is canonically isomorphic to the intersection cohomology group of V. To relate Saper's L^2 cohomology group to the original one, the two metrics ds^2 and say ds_V^2 shall be connected by a path $ds_V^2 + \varepsilon ds^2$, $\varepsilon \in [0, 1]$. Then we shall see that a sharpened version of Hörmander-Donnelly-Fefferman's L^2 estimate is true for this family (cf. Prop. 10), which enables us to compare the space of harmonic forms with respect to ds_V^2 and ds^2. In short, it will be shown that the ordinary L^2 cohomology is canonically isomorphic to certain L^2 cohomology which is already known to be isomorphic to the intersection cohomology.

The manuscript of the present work was written during the author's stay at Bergische Universität Gesamthochschule Wuppertal on the occasion of Aktivität in Komplex Analysis in MPI at Bonn. He would like to express deep gratitude to the hospitality of these institutions.

2 Preliminaries

Let G_1 and G_2 be two pre-Hilbert spaces and let $\bar{G}_1$ and $\bar{G}_2$ be their completions, respectively. A linear map from G_1 to G_2 is *closable* if it has a closed extension as a linear operator from $\bar{G}_1$ to $\bar{G}_2$. Given a closable linear map $T: G_1 \to G_2$, the minimal closed extension of T is denoted by $T_{\min}$. Recall that $T_{\min} = (T^*)^*$,

where T^* denotes the adjoint of T, and that the maximal closed extension of T, say $T_{\max}$, is given by $((T^*|G_2)_{\min})^*$.

Let (X, ds_X^2) be a Hermitian manifold of dimension n and let d be the exterior derivative acting on the set of smooth differential forms on X, say $C(X)$. Let $C_0(X)$ be the subset of $C(X)$ consisting of compactly supported forms and let d_0 be the restriction of d to $C_0(X)$. With respect to the canonical pre-Hilbert space structure of $C_0(X)$, d_0 is clearly a closable linear map. We shall denote the inner product of $\overline{C_0(X)}$ by $(,)$ and the L^2 norm by $\| \|$. By an abuse of notation we put $d_{\max}=(d_0)_{\max}$ and $d_{\min}=(d_0)_{\min}$. The formal adjoint of d will be denoted by $d^\vee$. By Lebesgue's theory we may canonically identify $\overline{C_0(X)}$ as a subset of the space of locally square integrable forms on X, say $L_{\mathrm{loc}}(X)$. For any relatively compact open subset $K\subset X$ and $f\in L_{\mathrm{loc}}(X)$, f_K will denote the trivial extension of $f|_K$ to X. For the sake of brevity we set $\|f\|_K:=\|f_K\|$ and $(f, g)_K:=(f_K, g_K)$ for $f, g\in L_{\mathrm{loc}}(X)$. The pointwise norm of a form u will be denoted by $|u|$.

The L^2 cohomology group of X with respect to ds_X^2 is defined by

$$H_{(2)}(X)(=H_{(2)}(X, ds_X^2)):=\mathrm{Ker}\, d_{\max}/\mathrm{Im}\, d_{\max}.$$

The r-th component of $H_{(2)}(X)$ will be denoted by $H^r_{(2)}(X)$. By $H^r(X)$ (resp. $H^r_0(X)$) we denote the ordinary r-th de Rham cohomology group of X (resp. that with compact support). Then we have the following canonical long exact sequence:

$$(1)\qquad \dots\to H^r_0(X)\to H^r_{(2)}(X)\to \varprojlim_{K\subset X} H^r_{(2)}(X\setminus K)\to H^{r+1}_0(X)\to\dots$$

where K runs through the compact subsets of X.

Recall that there is an orthogonal decomposition

$$\overline{C_0(X)}=\mathrm{Ker}\, D\oplus\overline{\mathrm{Im}\, D}$$

if $D=d_{\max}+d_{\max}{}^*$ or $=d_{\min}+d_{\min}{}^*$. We put $\mathscr{H}(=\mathscr{H}(X)):=\mathrm{Ker}(d_{\max}+d_{\max}{}^*)$ and $\mathscr{H}_{(*)}(=\mathscr{H}_{(*)}(X)):=\mathrm{Ker}(d_{\min}+d_{\min}{}^*)$. Clearly Hodge's star operator gives an isomorphism between $\mathscr{H}$ and $\mathscr{H}_{(*)}$.

Let $\mathrm{dist}(K_1, K_2)$ denote the distance between two subsets K_1 and K_2 of X. Given a real valued function φ on X we shall denote by X^φ_ε the ε-sublevel set of φ, i.e. we set $X^\varphi_\varepsilon=\{x\in X;\ \varphi(x)<\varepsilon\}$.

Proposition 1. *Let $D=d$ or $d^\vee$, and let $\varphi: K\to(0,\infty)$ be a continuous function such that X^φ_ε are relatively compact for all ε and that $\int_t^1 \inf_{X^\varphi_\varepsilon}|d\varphi^{-1}|\,d\varepsilon$ is non-integrable on $(0, 1)$. Then*

$$\{u\in\mathrm{Dom}\, D_{\max};\ \|\varphi^{-1}u\|<\infty\}\subset\mathrm{Dom}\, D_{\min}.$$

Proof. By hypothesis there exist an increasing family of compact subsets of X, say $X_m (m=1, 2, \dots)$, and C^∞ functions $r_m: X\to[0, 1]$ such that $r_m(x)=1$ on X_m, $r_m(x)=0$ on $X\setminus X_{m+1}$ and $m|dr_m(x)|<\varphi(x)^{-1}$ everywhere. Hence the conclusion follows immediately from the inequality

$$\|D_{\max}((r_m-r_{m'})u)\|\leqq(m^{-1}+m'^{-1})\|\varphi^{-1}u\|+\|(r_m-r_{m'})D_{\max}u\|.$$

Let $C^r(X)$ be the r-th component of $C(X)$. For each $u \in C^r(X)$, $e(u)$ will denote the exterior multiplication by u from the left hand side, and the adjoint of $e(u)$ will be denoted by $e(u)^\vee$. Furthermore, let $C = \sum \sqrt{-1}^{p-q} \prod_{p,q}$ where $\prod_{p,q}$ denotes the projection to the (p, q)-component of $C(X)$.

Let ω be the fundamental form of ds_X^2. The adjoint of $e(\omega)$ will be denoted by Λ according to the custom. Given two homogeneous operators S and T on $C(X)$, we define the graded commutator $[S, T]$ by $S \circ T - (-1)^{\deg S \deg T} T \circ S$.

Proposition 2. *If the metric ds_X^2 is Kählerian, then*

$$[d, C^{-1} e(u)^\vee C] - [C^{-1} d^\vee C, e(u)] = [e(du), \Lambda]$$

for any $u \in C^1(X)$.

Proof is immediate from $[e(u), \Lambda] = -C^{-1} e(u)^\vee C$, $[d, \Lambda] = C^{-1} d^\vee C$ and the graded Jacobi identity $[[S, T], U] - [S, [T, U]] = (-1)^{\deg T \deg U} [[S, U], T]$.

3 A Sharp L^2 Estimate

Let $\mathbb{C}^N$ be the complex number space of dimension N, let $z = (z_1, \ldots, z_N)$ be the coordinate of $\mathbb{C}^N$, and let $|z| = \sqrt{\sum_{i=1}^N |z_i|^2}$. Then we put $\mathbb{B}_*^N = \{z \in \mathbb{C}^N; 0 < |z| < 1/4\}$.

Proposition 3. *Let $u \in C^r(\mathbb{B}_*^N) \cap C_0(\mathbb{B}_*^N)$. Then with respect to the euclidean metric*

$$\| (|z| \log |z|)^{-1} u \| \leqq 3(\|du\| + \|d^\vee u\|) \tag{2}$$

if $r \neq N$.

Proof. In Proposition 2 we set $X = \mathbb{B}_*^N$ and $u = C^{-1} d \log \log |z|^{-1}$. Since the eigenvalues of du with respect to the euclidean metric, say $\lambda_1 \leqq \lambda_2 \leqq \ldots \leqq \lambda_N$ are given by

$$\lambda_1 = |z|^{-2} (\log |z|)^{-2}$$

and

$$\lambda_2 = \ldots = \lambda_N = |z|^{-2} (-\log |z|)^{-1}$$

because $-\log |z| > 1$ on the punctured ball $\mathbb{B}_*^N$, the operator $[e(du), \Lambda]: C^r(\mathbb{B}_*^N) \hookleftarrow$ is positive (resp. negative) with eigenvalues not less than λ_1 (resp. not more than $-\lambda_1$) if $r \neq N$. Taking the Kähler identity into account, we obtain (2) immediately.

In what follows let V be a compact irreducible complex space of dimension n with isolated singularities and let V' be the set of regular points of V.

Definition. A Hermitian metric of V is a C^∞ Hermitian metric ds^2 on V' such that, for each point $x \in V \setminus V'$ there exist a neighbourhood $U \ni x$, a holomorphic embedding ι of U into a domain say $\Omega \subset \mathbb{C}^N$, and a C^∞ Hermitian metric dH^2 defined on Ω which satisfies $(\iota|_{U \setminus \{x\}})^* dH^2 = ds^2|_{U \setminus \{x\}}$.

We say a Hermitian metric ds^2 of V is distinguished if the above U and ι can be so chosen that $ds^2|_{U \setminus \{x\}}$ is the pull-back of the euclidean metric for all $x \in V \setminus V'$.

In virtue of Courant's mini-max principle for the eigenvalues of positive definite Hermitian forms, the same argument as in Proposition 3 works for arbitrary complex submanifolds of $\mathbb{B}_*^N$ and yields the following estimate.

Proposition 4. *Let ds_V^2 be a distinguished Hermitian metric of V. Then for each $x \in V \setminus V'$ there exists a neighbourhood $U \ni x$ such that, with respect to the distance function $\delta_x(y) = \operatorname{dist}(x, y)$,*

$$\|(\delta_x \log \delta_x)^{-1} u\| \leqq 3(\|du\| + \|d^\vee u\|) \tag{3}$$

for all $u \in C^r(V') \cap C_0(U \setminus \{x\})$ with $r \neq n$.

Let us fix any positive continuous function φ_V on V' such that $\varphi_V = -\delta_x \log \delta_x$ on $U \setminus \{x\}$ for some neighbourhood $U \ni x$, for any $x \in V \setminus V'$. Then we infer from Proposition 4 the following.

Proposition 5. *Let ds_V^2 be a distinguished Hermitian metric of V. Then there exist a constant C_2 and a relatively compact open subset $K_0 \subset V'$ such that*

$$\|\varphi_V^{-1} u\| \leqq C_2(\|u\|_{K_0} + \|du\| + \|d^\vee u\|) \tag{4}$$

for all $u \in C^r(V') \cap C_0(V')$ with $r \neq n$.

Here let us recall our previous result in [5]. An L^2 estimate on complete Kähler manifolds has been applied and yielded

Theorem 6. *Under the above situation*

$$\varinjlim_{K \subset V'} H^r_{(2)}(V' \setminus K) = \{0\} \quad \text{for } r > n.$$

Taking the corresponding L^2 cohomology vanishing for the adjoint complex $\cdot \xleftarrow{d^\vee} \cdot \xleftarrow{d^\vee} \cdot$ into account*, we summarize what we have proved in [5] as follows.

Theorem 7. *Under the above situation, $H^r_{(2)}(V') \cong H^r_0(V')$ (resp. $\cong H^r(V')$) if $r > n+1$ (resp. $r < n-1$), $\dim H_0^{n+1}(V') \geqq \dim H^{n+1}_{(2)}(V')$, $\dim H^{n-1}(V') \geqq \dim H^{n-1}_{(2)}(V')$ and $\dim H^n_{(2)}(V') \geqq \dim \operatorname{Im}(H^n_0(V') \to H^n(V'))$. Moreover $\operatorname{Im} d_{\max}$ is a closed subspace of $\overline{C_0(V')}$.*

Corollary 8. $H_{(2)}(V') \cong \mathscr{H} \cong \mathscr{H}_{(*)}$.

We also refer to a theorem due to L. Saper [6].

Theorem 9. *Let ds_V^2 be as above. Then there exists a C^∞ function $\psi: V' \to (-\infty, 0)$ such that $ds_V^2 + \partial\bar{\partial}\psi$ is a complete Hermitian metric on V' satisfying*

$$H^r_{(2)}(V', ds_V^2 + \partial\bar{\partial}\psi) \cong \begin{cases} H^r(V') & \text{if } r < n \\ \operatorname{Im}(H^n_0(V') \to H^n(V')) & \text{if } r = n \\ H^r_0(V') & \text{if } r > n. \end{cases}$$

Here $\partial\bar{\partial}\psi$ denotes the complex Hessian of ψ.

* See "Note added in proof"

Evaluating the eigenvalues of $dC^{-1}d(\log\log|z|^{-1}+\varepsilon\psi)$ with respect to $ds_V^2+\varepsilon\nabla\bar{\nabla}\psi$, where $|z|$ denotes the holomorphic coordinate associated to the local holomorphic embedding of a neighbourhood of a singular point, by the same reasoning as before we obtain

Proposition 10. *In the statement of Theorem 9, the function ψ can be chosen so that, with respect to the L^2 norms $\|\ \|_\varepsilon$ associated to $ds_\varepsilon^2:=ds_V^2+\varepsilon\nabla\bar{\nabla}\psi$ for $\varepsilon\in(0,1]$, there exist a constant C_3 and a family of continuous functions $\varphi_\varepsilon: V'\to(0,\infty)$ which converges uniformly on compact subsets of V' to φ_V, satisfying*

$$\|\varphi_\varepsilon^{-1}u\|_\varepsilon\leqq C_3(\|u\|_{K_0}+\|du\|_\varepsilon+\|d^\vee u\|_\varepsilon) \tag{5}$$

for all $u\in C^r(V')\cap C_0(V')$ with $r\neq n$, where K_0 is as in Proposition 5.

Now we are ready to prove the following statement.

Main theorem. *Let V be a compact irreducible complex space with isolated singularities, and let ds_V^2 be a Hermitian metric of V. Then there exist canonical isomorphisms*

$$H^r_{(2)}(V')\cong\begin{cases}H^r(V') & \text{if } r<n\\ \mathrm{Im}(H^n_0(V')\to H^n(V')) & \text{if } r=n\\ H^r_0(V') & \text{if } r>n.\end{cases}$$

Proof. Since $H_{(2)}(V')$ is independent of the choice of metrics within the quasi-isometry class of ds_V^2, we may assume that ds_V^2 is a distinguished Hermitian metric of V.

Let ψ, ds_ε^2, K_0 and C_3 be as in Proposition 10, and let $\mathscr{H}^r_\varepsilon$ be the space of L^2 harmonic r-forms on V' with respect to ds_ε^2. Then, for all $\varepsilon\in(0,1]$ we have

$$\dim\mathscr{H}^{n\pm1}\leqq\dim\mathscr{H}_1^{n\pm1}=\dim\mathscr{H}_\varepsilon^{n\pm1}$$

and

$$\dim\mathscr{H}^n\geqq\dim\mathscr{H}_1^n=\dim\mathscr{H}_\varepsilon^n.$$

First we shall prove that $\dim\mathscr{H}^{n\pm1}=\dim\mathscr{H}_1^{n\pm1}$. Suppose on the contrary that $\dim\mathscr{H}^{n+1}<\dim\mathscr{H}_1^{n+1}$ for instance. Then $\dim\mathscr{H}^{n+1}<\dim\mathscr{H}_\varepsilon^{n+1}$ for all $\varepsilon\in(0,1]$. Let $f_\varepsilon\in\mathscr{H}_\varepsilon^{n+1}$ be chosen so that $\|\varphi_\varepsilon^{-1}f_\varepsilon\|_\varepsilon=1$ and $(f_\varepsilon,u)_{K_0}=0$ for all $u\in\mathscr{H}^{n+1}$. By (5) there exist $f\in\mathrm{Ker}(d_{\max}+d_{\max}^\vee)$ satisfying $\|\varphi_V^{-1}f\|\leqq1$ and $\|f\|_{K_0}\geqq C_3^{-1}$, which is the weak limit of a subsequence of $\{f_\varepsilon\}_{\varepsilon\searrow0}$. By Proposition 1 we have $f\in\mathscr{H}$. Hence $(f_\varepsilon,f)_{K_0}=0$ for all ε. But it implies that $\|f\|_{K_0}=0$, which is a contradiction. Thus we have proved that $\dim\mathscr{H}^{n+1}=\dim\mathscr{H}_\varepsilon^{n+1}$ for sufficiently small ε, so that $\mathscr{H}^{n+1}\cong\mathscr{H}_1^{n+1}$, too. Similarly we obtain $\mathscr{H}^{n+1}_{(*)}\cong\mathscr{H}_1^{n+1}$, so that $\mathscr{H}^{n+1}\cong\mathscr{H}_1^{n+1}$. Since $\dim\mathscr{H}^{n+1}=\dim\mathscr{H}_1^{n+1}<\infty$, the same reasoning leads to

$$\{f\in\mathscr{H}^{n+1};\ \|f\|=1\quad\text{and}\quad|(f,u)_{K_0}|<C_3^{-1}\quad\text{for all } u\in\mathscr{H}_\varepsilon^{n+1}\ \text{with}\ \|\varphi_\varepsilon^{-1}u\|_\varepsilon=1\}=\emptyset\quad\text{if } 0<\varepsilon\ll1.$$

Hence there must exist a constant C_4 independent of ε such that

$$\|\varphi_\varepsilon^{-1}f\|_\varepsilon\leqq C_4(\|df\|_\varepsilon+\|d^*f\|_\varepsilon)$$

for all $f\in\mathrm{Dom}\,d\cap\mathrm{Dom}\,d^*\cap\overline{C^{n+1}(V')\cap C_0(V')}^{\varepsilon}$ satisfying $(f,u)_{K_0}=0$ for any $u\in\mathscr{H}_{\varepsilon}^{n+1}$. Here $^{-\varepsilon}$ denotes the completion with respect to $\|\ \|_{\varepsilon}$, and we are not distinguishing the minimal and the maximal closed extensions of d since they coinside by the completeness of the metric ds_{ε}^2. Similarly there must exist a constant C_5 such that

$$\|\varphi_{\varepsilon}^{-1}f\|_{\varepsilon}\leqq C_5(\|df\|_{\varepsilon}+\|d^*f\|_{\varepsilon})$$

for all $f\in\mathrm{Dom}\,d\cap\mathrm{Dom}\,d^*\cap\overline{C^{n-1}(V')\cap C_0(V')}^{\varepsilon}$ satisfying $(f,u)_{K_0}=0$ for any $u\in\mathscr{H}_{\varepsilon}^{n-1}$.

Therefore, for every $f\in\overline{C^n(V')\cap C_0(V')}$ which satisfies $(f,u)_{K_0}=0$ for all $u\in\mathscr{H}_{\varepsilon}^n$, one can find $g_{\varepsilon}\in\mathrm{Dom}\,d^*$ and $h_{\varepsilon}\in\mathrm{Dom}\,d$ with respect to ds_{ε}^2, such that

$$f_{K_0}=d^*g_{\varepsilon}+dh_{\varepsilon}$$

$$\|\varphi_{\varepsilon}^{-1}g_{\varepsilon}\|_{\varepsilon}\leqq C_4\|f_{K_0}\|_{\varepsilon}$$

and

$$\|\varphi_{\varepsilon}^{-1}h_{\varepsilon}\|_{\varepsilon}\leqq C_5\|f_{K_0}\|_{\varepsilon}.$$

Now suppose that $\dim\mathscr{H}^n>\dim\mathscr{H}_1^n=k$. Then, by Aronszajn's unique continuation theorem there must exist $f^1,\ldots,f^{k+1}\in\mathscr{H}^n$ such that $f_{K_0}^1,\ldots,f_{K_0}^{k+1}$ are linearly independent. Let $L\subset\mathscr{H}^n$ be the linear subspace spanned by $\{f^1,\ldots,f^{k+1}\}$. Then we can choose a sequence $\{l_m;\ m=1,2,\ldots\}\subset L$ such that $\|l_m\|=1$ and $l_{mK_0}\perp\mathscr{H}_{1/m}^n$ for all m. Thus, using what we have mentioned above, we can find $g_m\in\mathrm{Dom}\,d^*$ and $h_m\in\mathrm{Dom}\,d$ satisfying

$$l_{mK_0}=d^*g_m+dh_m$$

$$\|\varphi_{1/g}^{-1}g_m\|_{1/m}\leqq C_4\|l_{mK_0}\|_{1/m}$$

$$\|\varphi_{1/m}^{-1}h_m\|_{1/m}\leqq C_5\|l_{mK_0}\|_{1/m}.$$

Letting $m\to\infty$, we choose a weakly convergent subsequence of l_m, g_m and h_m and finally obtain $l\in L$, $g\in\mathrm{Dom}\,d_{\max}^{\vee}$ and $h\in\mathrm{Dom}\,d_{\max}$ satisfying $l_{K_0}=d_{\max}^{\vee}g+d_{\max}h$, $\|\varphi_V^{-1}g\|\leqq C_4$ and $\|\varphi_V^{-1}h\|\leqq C_5$. By Proposition 1 we have $g\in\mathrm{Dom}\,d_{\min}^{\vee}$ and $h\in\mathrm{Dom}\,d_{\min}$. Therefore $\|l_{K_0}\|^2=(l_{K_0},l\|=(d_{\min}^{\vee}g+d_{\min}h,l)=0$. On the other hand $l\neq0$ since $\dim L<\infty$ and $\|l_m\|=1$. But it is a contradiction to Aronszajn's theorem or that $f_{K_0}^1,\ldots,f_{K_0}^{k+1}$ were linearly independent. Q.E.D.

References

1. Cheeger, J.: On the Hodge theory of Riemannian pseudomanifolds. Proc. Symp. Pure Math. **36**, 91–146 (1980)
2. Cheeger, J., Goreski, M., MacPherson, R.: L^2-cohomology and intersection homology for singular varieties. Ann. Math. Stud. **102**, Seminar on Differential Geometry, 303–340 (1982)
3. Hsiang, W.C., Pati, V.: L^2-cohomology of normal algebraic surfaces I. Invent. Math. **81**, 395–412 (1985)
4. Nagase, M.: Remarks on the L^2-cohomology of singular algebraic surfaces. J. Math. Soc. Japan **41**, 97–116 (1989)
5. Ohsawa, T.: Hodge spectral sequence on compact Kähler spaces. Publ. RIMS, **23**, 613–615 (1987) Suppl.: Publ. RIMS, **26** (to appear)
6. Saper, L.: L^2-cohomology and intersection homology of certain algebraic varieties with isolated singularities. Invent. Math. **82**, 207–255 (1985). L^2-cohomology of Kähler varieties with isolated singularities. Preprint

Note added in proof

The proof of Theorem 7 is incomplete. Corrected argument is given in [5], Supplement. The author is grateful to colleagues including G. Marinescu and M. Stern who have pointed out this gap.

Math. Z. 206, 225–232 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Complex borel structure in *JC*-algebras

L.J. Bunce and J.D. Maitland Wright
Analysis and Combinations Research Centre, Mathematics Department, P.O. Box 220, University of Reading, Reading, R662AX, United Kingdom

Received May 16, 1989, in final form February 15, 1990

Given an *JC*-algebra A let $\hat{A}$ denote the Jordan equivalence classes of the Type I factor (or dense) representations of A [2, § 2] and let $\mathrm{Prim}(A)$ be the space of primitive ideals of A with the hull-kernel topology. If $\rho \in P(A)$, the pure states of A, let $\tau_\rho\colon A \to A^{**}c(\rho)$ $(a \to a \cdot c(\rho))$ be the canonical Type I factor representation of A associated with ρ, where $c(\rho)$ is the central support of ρ in A^{**}. There are natural surjections

$$P(A) \xrightarrow{\nu} \hat{A} \xrightarrow{\mu} \mathrm{Prim}(A) \qquad (\rho \mapsto [\tau_\rho] \to \mathrm{Ker}\,\tau_\rho)$$

and $\hat{A}$ is given the inverse topology induced by μ. If A is *separable* then, with the relative weak $*$ topology on $P(A)$, the *Mackey-Borel* structure of $\hat{A}$ is defined to be the quotient Borel structure of $P(A)$ obtained from ν.

This is, of course, precisely the situation which obtains for C^*-algebras but with $\hat{A}$ (for a C^*-algebra A) representing the equivalence classes of irreducible $*$-representations of A and with τ_ρ replaced by π_ρ, the G.N.S. representation associated with $\rho \in P(A)$. If A is a *JC*-algebra or a C^*-algebra then the open subsets of $P(A)$, $\hat{A}$ and $\mathrm{Prim}(A)$ corresponding to all representations of dimension >1 are denoted by $P_1(A)$, $\hat{A}_1$ and $\mathrm{Prim}_1(A)$ respectively.

Let A be a *JC*-algebra identified with its image inside the universal enveloping C^*-algebra $C^*(A)$, and let ϕ be the canonical involutory $*$-antiautomorphism of $C^*(A)$ [4; 6, § 7]. If A is of *complex type* [2, § 5], that is, if $A^{**}c(\rho) \simeq B(H)_{sa}$, for some complex Hilbert space H, for all $\rho \in P(A)$, then the restriction function

$$r\colon\ P_1(C^*(A)) \to P_1(A)$$

is a well-defined continuous surjection with two elements in each fibre and, via the natural $\mathbb{Z}_2$-actions induced by ϕ, the maps

$$\hat{r}\colon\ C^*(A)\hat{}_1 \to \hat{A}_1 \qquad ([\pi_\rho] \mapsto [\tau_{r(\rho)}])$$

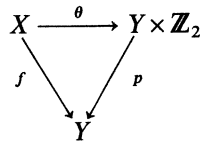
and

$$r\colon\ \mathrm{Prim}_1(C^*(A)) \to \mathrm{Prim}_1(A) \qquad (P \mapsto P \cap A)$$

constitute continuous $\mathbb{Z}_2$-bundles with $\hat{r}$ having fibre $\mathbb{Z}_2$. Neither bundle need be locally trivial, but is trivial if and only if A is isomorphic to the self-adjoint part of a C^*-algebra. The details can be found in [1, § 8]. The last mentioned property is an algebraic phenomenon essentially (c.f. Lemma 3), given that for any JC-algebra or C^*-algebra B the topology of $\hat{B}$ is a faithful reflection of ideal structure. This is not the case with the Mackey-Borel structure, however, as $\hat{B}$ can have a rich Mackey-Borel structure and trivial topology simultaneously.

Henceforth A will represent a separable JC-algebra. In view of the above remarks it is natural to wonder what insights can be gained from investigation of the relative behaviour of the Mackey-Borel structures of $C^*(A)^{\wedge}$ and $\hat{A}$. The question is considered in this paper and the main result, approached by a sequence of lemmas, relates Mackey-Borel structure to primitive ideals, to ideal structures and to enveloping Borel $*$-algebras.

By a Borel $\mathbb{Z}_2$-bundle we mean a surjective Borel map $f: X \to Y$ between Borel spaces X and Y with a $\mathbb{Z}_2$-action on X by fibre preserving Borel isomorphisms such that the quotient Borel space $X/\mathbb{Z}_2$ is Borel isomorphic to Y. Such a bundle is said to be *trivial* if there is a commutative diagram



where θ is a Borel isomorphism of X onto $Y \times \mathbb{Z}_2$ and p is the natural projection.

In the foregoing the $\mathbb{Z}_2$-actions on $P(C^*(A))$ and $C^*(A)^{\wedge}$ are considered to be the natural actions induced by the canonical involutory $*$-antiautomorphism ϕ of $C^*(A)$ (i.e. $\rho \mapsto \phi^*(\rho)$ and $[\pi_\rho] \mapsto [\pi_{\phi^*(\rho)}]$). Recall that A is said to be *universally reversible* if A is reversible in $C^*(A)$. Equivalently, A has no spin factor representations of dimension greater than 4 [4]. Consequently every JC-algebra of complex type is universally reversible (see also [2, Theor. 4.6]).

Lemma 1. *Let A be of complex type. Then*

(i) *$r: P_1(C^*(A)) \to P_1(A)$ is a trivial Borel $\mathbb{Z}_2$-bundle.*

(ii) *$\hat{r}: C^*(A)^{\wedge}_1 \to \hat{A}_1$ is a locally trivial Mackey-Borel $\mathbb{Z}_2$-bundle.*

Proof. (i) By [2, Coroll. 5.6] r is continuous with, for each ρ in $P_1(C^*(A))$,

$$r^{-1}(r(\rho)) = \{\rho, \phi^*(\rho)\}, \qquad \rho \neq \phi^*(\rho).$$

In particular, the equivalence relation induced by r is $\mathbb{Z}_2$-orbit equivalence. Obviously each resulting equivalence class is a closed subset and the saturation of an open subset E of $P_1(C^*(A))$ is $E \cup \phi^*(E)$ and so is open, hence Borel. Therefore since $P_1(C^*(A))$ is a Polish space ([9, 4.3.2], for example), it follows from [10, Theor. A.15, p. 383] that there is a Borel subset S of $P_1(C^*(A))$ which contains one and only one of ρ and $\phi^*(\rho)$ for each ρ. Thus $S \cup \phi^*(S) = P_1(C^*(A))$ and $S \cap \phi^*(S) = \emptyset$. Therefore the restriction $r: S \to P_1(A)$ is a Borel bijection between standard Borel spaces ($P_1(A)$ is Polish) and hence a Borel isomorphism [9, 4.6.12]. It follows that if E is a Borel subset of $P_1(C^*(A))$ then

$$r(E) = r(E \cap S) \cup r(E \cap \phi^*(S))$$

is a Borel subset of $P_1(A)$ and this, together with the preceeding remarks, proves (i).

(ii) The diagram

$$\begin{array}{ccc} P_1(C^*(A)) & \xrightarrow{\alpha} & C^*(A)\hat{}_1 \\ \downarrow{\scriptstyle r} & & \downarrow{\scriptstyle \hat{r}} \\ P_1(A) & \xrightarrow[\beta]{} & \hat{A}_1 \end{array}$$

is commutative, where α, β are the canonical (continuous) surjections. Thus if E is a Mackey-Borel subset of $\hat{A}_1$ then

$$\alpha^{-1}\hat{r}^{-1}(E)=r^{-1}\beta^{-1}(E)$$

is a Borel subset of $P_1(C^*(A))$ so that $\hat{r}^{-1}(E)$ is a Mackey-Borel subset of $C^*(A)\hat{}_1$.

On the other hand, from the results of [2, § 5], we see that

$$\begin{aligned}\beta^{-1}[\tau_{r(\rho)}] &= \{r(\sigma): [\pi_\rho]=[\pi_\sigma] \quad \text{or} \quad [\pi_{\phi^*(\sigma)}], \quad \sigma\in P_1(C^*(A))\} \\ &= r(\alpha^{-1}([\pi_\rho]))\end{aligned}$$

for each ρ in $P_1(C^*(A))$ which implies that

$$r(\alpha^{-1}(E))=\beta^{-1}(\hat{r}(E))$$

for every subset E of $P_1(C^*(A))$. Together with (i) this implies that $\hat{r}$ sends Mackey-Borel sets to Mackey-Borel sets and therefore defines a Mackey-Borel $\mathbb{Z}_2$-bundle. Local triviality is immediate from the facts that $\hat{r}$ has fibre $\mathbb{Z}_2$ and that the points of $C^*(A)\hat{}$ (and hence of $\hat{A}$) are Mackey-Borel sets (e.g. [9, 4.7.4].

Lemma 2. *Let the Mackey-Borel $\mathbb{Z}_2$-bundle $\hat{r}$: $C^*(A)\hat{} \to \hat{A}_1$ be trivial, where A is of complex type. Then $\phi(P)\neq P$ for all $P\in \mathrm{Prim}_1(C^*(A))$ (i.e. $\check{r}$: $\mathrm{Prim}_1(C^*(A)) \to \mathrm{Prim}_1(A)$ has fibre $\mathbb{Z}_2$).*

Proof. Since both the hypothesis and the claimed outcome hold for A if and only if they are true for $\tilde{A}$ (the algebra formed from A and an adjoined unit) we can suppose without loss that A has a unit.

By hypothesis there is a Mackey-Borel subset E of $C^*(A)\hat{}_1$ satisfying

$$E\cup\hat{\phi}(E)=C^*(A)\hat{}_1, \qquad E\cap\hat{\phi}(E)=\emptyset$$

where $\hat{\phi}([\pi_\rho])=[\pi_{\phi^*(\rho)}]$ for all $\rho\in P_1(C^*(A))$. In order to obtain a contradiction suppose that $P\in\mathrm{Prim}_1(C^*(A))$ and $\phi(P)=P$. Then with $Q=P\cap A$ we get

$$P=C^*(Q), \qquad C^*(A)/P=C^*(A/Q),$$

by using [4, Theor. 4.1, Lemma 4.3]. Thus, passing to A/Q and $E\cap(C^*(A)\hat{}_1\setminus C^*(Q))\hat{}$ we can suppose that $C^*(A)$ is primitive.

In that case we can choose $\rho\in P_1(C^*(A))$ (the Lemma is vacuous otherwise) such that π_ρ is injective. Writing F_ρ for the minimal split face of the state space K of $C^*(A)$ which contains ρ, this means that $\overline{F}_\rho$ (weak$*$closure)$=K$ (see e.g. [5, § 4]). By a standard argument using Milman's Theorem this implies that

$\partial_e F_\rho$ is weak$*$dense in $\partial_e K$ and hence in $P_1(C^*(A))$, wherein it lies. Note that in the notation of Lemma 1 (ii), $\partial_e F_\rho = \alpha^{-1}([\pi_\rho])$. Choose now a countable uniformly dense subgroup G of the unitary group U of $C^*(A)$ and consider the action of U on $P(C^*(A))$ defined by

$$u \cdot \sigma(x) = \sigma(uxu^*), \quad \sigma \in P(C^*(A)), \quad x \in C^*(A), \quad u \in U.$$

By [3, 2.8.6],

$$U \cdot \rho = \alpha^{-1}([\pi_\rho]),$$

and it follows that $G \cdot \rho$ is weak$*$dense in $P_1(C^*(A))$. But, using [3, 2.8.6] once more, $S = \alpha^{-1}(E)$ is a G-invariant Borel subset of $P_1(C^*(A))$ with complement $\phi^*(S)$. Therefore, by [12, Lemma 1.1] one of S and $\phi^*(S)$ is meagre and hence both are, implying that $P_1(C^*(A))$ is meagre, which is impossible and the proof is complete.

The next Lemma is virtually [1, Theor. 8.6] but a short proof is possible which does not use state spaces or orientations.

Lemma 3. *Let A be universally reversible. Then A contains a non-associative non-zero norm closed Jordan ideal isomorphic to the self-adjoint part of a C^*-algebra if and only if $C^*(A)$ contains a non-zero norm closed ideal I such that $I \cap \phi(I) = 0$.*

Proof. Suppose that I is a non-zero norm closed ideal of $C^*(A)$ with $I \cap \phi(I) = 0$. It is easy to check that $R = \{x \in C^*(A) : \phi(x) = x^*\}$, is a real C^*-algebra with $R \cap iR = 0$ and $C^*(A) = R + iR$. Reversibility implies that $A = R_{sa}$ [4, Lemma 4.2]. Simple calculations now show that the map $\pi : I_{sa} \to A$ defined by $\pi(x) = x + \phi(x)$, for each $x \in I_{sa}$, is a Jordan isomorphism onto $(I + \phi(I)) \cap A$.

The converse is immediate from [4, Coroll. 4.6] given the fact that every non-associative JC-algebra contains a subalgebra without one-dimensional representations.

Separability is clearly an unnecessary condition in Lemma 3 but we repeat our standing hypothesis that A is a separable JC-algebra.

A norm closed Jordan ideal of A is said to be *essential* if it has non-zero intersection with every norm closed Jordan ideal of A.

Lemma 4. *Every primitive quotient of A contains a non-zero norm closed Jordan ideal isomorphic to the self-adjoint part of a C^*-algebra if and only if A is of complex type and $\check{r}$: $\mathrm{Prim}_1(C^*(A)) \to \mathrm{Prim}_1(A)$ has fibre $\mathbb{Z}_2$. In which case A contains an essential norm closed Jordan ideal which is isomorphic to the self-adjoint part of a C^*-algebra.*

Proof. Note that either of the conditions claimed to be equivalent in the first statement imply that A is universally reversable. Given, then, $P \in \mathrm{Prim}_1(C^*(A))$ we have $C^*(A \cap P) = P \cap \phi(P)$ and $C^*(A/A \cap P) = C^*(A)/P \cap \phi(P)$, using [4, Lemma 4.3], and it is therefore immediate from Lemma 3 that $A/P \cap A$ contains a non-zero ideal isomorphic to the self-adjoint part of a C^*-algebra if and only if $P \cap \phi(P)$ is not primitive, equivalently, $P \neq \phi(P)$. Since $\check{r}$: $\mathrm{Prim}_1(C^*(A)) \to \mathrm{Prim}(A)$ is surjective, the first statement is proved.

Now suppose that A satisfies the conditions of the first statement. We show first that A contains a non-zero ideal isomorphic to the self-adjoint part of a C^*-algebra. Passing to the ideal of A corresponding to $\mathrm{Prim}_1(A)$, we may suppose that A has no one dimensional representations.

Then $\phi(P) \neq P$ for all $P \in \mathrm{Prim}(C^*(A))$. For each $x \in C^*(A)$ consider the function

$$f_x\colon \mathrm{Prim}(C^*(A)) \to \mathbb{R}$$

defined by $f_x(P) = \|x + P\|$. Since $C^*(A)$ is separable there is a dense G_δ subset S of $\mathrm{Prim}(C^*(A))$ such that for all $x \in C^*(A)$, f_x is continuous on S (see [3, 3.9.4]) and hence continuous on the ϕ-invariant dense G_δ subset $S_1 = S \cap \phi(S)$. Choose $P_0 \in S_1$ and let $x_0 \in \phi(P_0) \setminus P_0$. Then $f_{x_0}(\phi(P_0)) = 0 < f_{x_0}(P_0)$ so that continuity at $\phi(P_0)$ and P_0 implies that these points have disjoint open neighbourhoods which, by structure theory, means that there exist ideals I, J of $C^*(A)$ with $I \not\subset P$, $J \not\subset \phi(P)$ and $I \cap J = 0$. It follows that $K = I \cap \phi(J) \neq 0$ and $K \cap \phi(K) = 0$, and the desired conclusion results from Lemma 3.

Finally, since the hypothesis is inherited by norm closed Jordan ideals, we see that the restricted sum of a maximal orthogonal family of norm closed Jordan ideals which are isomorphic to the self-adjoint part of a C^*-algebra is an essential ideal of A of the required type.

The remaining results, including the main theorem, require a knowledge of Borel $*$-algebras. We refer to Pedersen's systematic treatise in [9] for Borel C^*-algebras and to [8] for Jordan Borel $*$-algebras (there called J. Baire $*$-algebras). In order to facilitate the discussion we will make some general clarifying remarks.

A JC-algebra B is said to be a Jordan-Borel $*$-algebra if every norm bounded increasing sequence in B has a least upper bound in B and B has a separating family of σ-normal states. Equivalently, B has a faithful (σ-normal) realisation as a monotone sequentially closed Jordan algebra of self-adjoint operators on a complex Hilbert space.

It is implicit in Kadison's article [7], as observed in [8], (see also the proof of [9, 4.5.3]) that given a JC-algebra $B \subset B(H)_{sa}$, H a complex Hilbert space, then the smallest class of elements in $B(H)_{sa}$ which contains B and is closed under strong monotone sequential limits (the monotone sequential closure of B) is a JC-algebra, and hence a Jordan-Borel $*$-algebra. If B is the self-adjoint part of a C^*-algebra then so is its monotone sequential closure, [9, 4.5.3].

If B is a JC-algebra (resp. a C^*-algebra) then the Jordan-Borel $*$-algebra (resp. Borel C^*-algebra) generated by B in B^{**} is called the *enveloping Jordan-Borel $*$-algebra* (resp.: enveloping Borel C^*-algebra) of B and is written, $\beta(B)$.

Given any JC-algebra B we have the usual commutative diagram of canonical inclusions

$$\begin{array}{ccc} B & \hookrightarrow & B^{**} \\ \downarrow & & \downarrow \\ C^*(B) & \hookrightarrow & C^*(B)^{**} \end{array}$$

so that $B \subset \beta(C^*(B))$ and hence $\beta(B) \subset \beta(C^*(B))$. Also the canonical involutory $*$-antiautomorphism ϕ of $C^*(B)$ extends (uniquely) to a normal involutory $*$-antiautomorphism of $C^*(B)^{**}$ which in turn restricts to a σ-normal involutory $*$-antiautomorphism of $\beta(C^*(B))$. We continue to denote by ϕ both this extension and restriction. Note that ϕ is the identity on B^{**} and hence on $\beta(B)$.

In a similar vein using extension properties of $C^*(B)$ and of enveloping Borel $*$-algebras, [9, 4.5.9] and its Jordan analogue [8, p. 62], we obtain the following universal property.

Lemma 5. *Let B be a JC-algebra and let $\pi\colon \beta(B)\to B(H)_{sa}$ be a σ-normal Jordan representation. Then π extends uniquely to a σ-normal $*$-representation $\bar{\pi}\colon \beta(C^*(B))\to B(H)$.*

Notice that if B is any JC-algebra and $e_1\,B^{**}$, $e_2\,C^*(B)^{**}$ are the atomic parts of B^{**} and $C^*(B)^{**}$ respectively then $e_2\leqq e_1$, which may be seen from the fact that every irreducible Jordan representation $B\to B(H)$ extends to an irreducible $*$-representation of $C^*(B)$. Since $\beta(C^*(B))\to e_2\,\beta(C^*(B))$ $(x\mapsto xe_2)$ is faithful [9, 4.5.13], it follows that $\beta(B)\to e_i\,\beta(B)$, $i=1,2$, are faithful.

Returning now to our standing hypothesis that A is a separable JC-algebra, following Pedersen [9] we say that A is *Borel isomorphic to the self-adjoint part of a C^*-algebra* if $\beta(A)\simeq\beta(C_{sa})$ for some separable C^*-algebra C.

The next Lemma is the Jordan analogue of [9, 4.6.2] the proof of which is similar to the original. The details are left to the reader.

Lemma 6. *Let I be a norm closed Jordan ideal of A with associated open central projection e. Then $\beta(I)=e\beta(A)$ and $\beta(A/I)\simeq(1-e)\,\beta(A)$.*

Lemma 7. *Let $\beta(A)$ be isomorphic to the self-adjoint part of a C^*-algebra. Then A contains a norm closed Jordan ideal isomorphic to the self-adjoint part of a C^*-algebra.*

Proof. If A is non unital then $\tilde{A}$, the algebra formed by adjoining a unit to A, certainly satisfies the hypothesis and as A is also an essential ideal of $\tilde{A}$ we may suppose without loss that A has a unit. We may also suppose that A is not associative.

Let $\beta(A)\simeq C_{sa}$ where C is a C^*-algebra. Then C is a countably generated Borel C^*-algebra so that [9, 4.5.11] implies that there exists a separable C^*-algebra B and a surjective σ-normal homomorphism $\pi\colon \beta(B_{sa})\to\beta(A)$. From Lemma 5 and the remarks prior to it we see that π induces a commutative diagram

$$\begin{array}{ccccc}
\beta(B_{sa}) & \hookrightarrow & \beta(C^*(B_{sa})) & \xrightarrow{\phi'} & \beta(C^*(B_{sa})) \\
\downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi} \\
\beta(A) & \hookrightarrow & \beta(C^*(A)) & \xrightarrow[\phi]{} & \beta(C^*(A))
\end{array}$$

where $\bar{\pi}$ is surjective and ϕ', ϕ are the canonical (extended) σ-normal involutory $*$-antiautomorphisms. The commutator ideal I of B is non-zero and so using [4, Coroll. 4.5] we find that

$$\beta(C^*(I_{sa}))=\beta(J)\oplus\phi'\,\beta(J)$$

for some non-zero ideal J of I. Put $I_0=\bar{\pi}(\beta(J))$, then $I_0\cap\phi(I_0)=0$.

Consider now the σ-ideal M of $\beta(C^*(A))$ consisting of all those elements x of $\beta(C^*(A))$ for which $\{\rho\in P(C^*(A));\rho(x)\neq 0\}$ is a meagre subset of $P(C^*(A))$, [11, Prop. 2.1]. Observe that $\phi(M)=M$ because

$$\phi^*(\{\rho\in P(C^*(A));\rho(x)\neq 0\})=\{\rho\in P(C^*(A));\rho(\phi(x))\neq 0\}.$$

Therefore if I_0 is contained in M then so is $\phi(I_0)$ and consequently $\beta(C^*(A))/M$, must be a homomorphic image of the commutative C^*-algebra $\beta(C^*(A_{sa}/I_{sa}))$.

Since the natural map $q\colon C^*(A)\to\beta(C^*(A))/M$ is injective [11, Theor. 2.3] this is impossible. Hence $I_0\not\subset M$. Let $\bar{I}_0$ be the image of I_0 in $\beta(C^*(A))/M$. Then these remarks together with [11, Theor. 3.1], imply that $q^{-1}(\bar{I}_0)\neq 0$. But $\bar{I}_0\cap\overline{\phi(I_0)}=0$, and so,

$$q^{-1}(\bar{I}_0)\cap\phi(q^{-1}(\bar{I}_0))=q^{-1}(\bar{I}_0)\cap q^{-1}(\overline{\phi(I_0)})=0,$$

and the result follows from Lemma 3.

We can now state the main result.

Theorem. *The following are equivalent for a separable JC-algebra A of complex type.*

(i) $\hat{r}\colon C^*(A)\hat{}_1\to\hat{A}_1$ *is a trivial Mackey-Borel* $\mathbb{Z}_2$-bundle.

(ii) $\check{r}\colon \mathrm{Prim}_1(C^*(A))\to\mathrm{Prim}_1(A)$ *has fibre* $\mathbb{Z}_2$.

(iii) A/Q *contains a non-zero norm closed Jordan ideal isomorphic to the self-adjoint part of a C^*-algebra for all* $Q\in\mathrm{Prim}(A)$.

(iv) A *has a countable composition series* $(I_\alpha)_{0\leqq\alpha\leqq\lambda}$ *of norm closed Jordan ideals such that* $I_{\alpha+1}/I_\alpha$ *is isomorphic to the self-adjoint part of a C^*-algebra for all* $\alpha<\lambda$.

(v) A *is Borel isomorphic to the self-adjoint part of a C^*-algebra.*

(vi) $\beta(A)\simeq C_{sa}$ *for some C^*-algebra C.*

Proof. (i)$\Rightarrow$(ii)$\Leftrightarrow$(iii) by Lemmas 2, 3 and 4, and (iii)$\Rightarrow$(iv) by Lemma 4 together with transfinite induction and separability. Using Lemmas 6 and 7 the implication (vi)$\Rightarrow$(iv) follow similarly, whereas (v)$\Rightarrow$(vi) is trivial. It remains to prove (iv)$\Rightarrow$(v), (i).

Given a countable composition series satisfying (iv), let ρ_α be the open central projection in A^{**} of I_α for each α. Then $\beta(I_{\alpha+1}/I_\alpha)\simeq(e_{\alpha+1}-e_\alpha)\,\beta(A)$ for each α, by Lemma 6. By assumption there exists a separable C^*-algebra B such that

$$B_{sa}\simeq(\Sigma(I_{\alpha+1}/I_\alpha))_0 \qquad \text{(restricted direct sum)}$$

and so

$$\beta(B_{sa})\simeq\Sigma(e_{\alpha+1}-e_\alpha)\,\beta(A)=\beta(A),$$

proving (v).

In order to prove (iv)$\Rightarrow$(i), by passing to the ideal of A corresponding to $\hat{A}_1$ it is clearly enough to suppose that A has no one-dimensional representations and hence neither does the C^*-algebra B constructed above. Applying [4, Coroll. 4.5] and Lemma 5, we therefore obtain,

$$\beta(C^*(B_{sa}))\simeq\beta(C^*(A))=J\oplus\phi(J)$$

for some ideal J, which implies a non-zero central projection e of $\beta(C^*(A))$ with $\phi(e)=1-e$. Since e is Borel measurable considered as an affine function on the state space of $C^*(A)$ the set $S=\{\rho\in P(C^*(A));\rho(e)=0\}$ is clearly a saturated subset for the Mackey-Borel structure with $S\cup\phi^*(S)=P(C^*(A))$ and $S\cap\phi^*(S)=\emptyset$. This together with Lemma 1 (i) implies (i) and the proof is complete.

References

1. Alfsen, E.M., Shultz, F.W.: State Spaces of C^*-algebras. Oslo University. Preprint 1978
2. Alfsen, E.M., Hanche-Olsen, H., Shultz, F.W.: State Spaces of C^*-algebras. Acta Math. **144**, 267–305 (1980)
3. Dixmier, J.: C^*-algebras. New York: North-Holland 1977
4. Hanche-Olsen, H.: On the structure and tensor products of JC-algebras. Can. J. Math. **35**, 1059–1074 (1983)
5. Hanche-Olsen, H.: Split faces and ideal structure of operator algebras. Math. Scand. **48**, 137–144 (1981)
6. Hanche-Olsen, H., Stormer, E.: Jordan Operator Algebra. New York: Pitman 1984
7. Kadison, R.: Unitary invariants for representations of operator algebras. Ann. Math. **66**, 304–379 (1957)
8. Kehlet, E.: On the monotone sequential closure of a C^*-algebra. Math. Scand. **25**, 59–70 (1969)
9. Pedersen, G.K.: C^*-algebra and their automorphism groups. New York: Academic Press 1979
10. Takesaki, M.: Theory of operator algebras I. Berlin Heidelberg New York: Springer 1979
11. Wright, J.D.M.: Regular σ-completions of C^*-algebras. J. Lond. Math. Soc. II. Ser. **12**, 299–309 (1976)
12. Sullivan, D., Weiss, B., Wright, J.D.M.: Generic Dynamics and monotone complete C^*-algebras. Trans. Am. Soc. **295**, 795–809 (1986)

Math. Z. 206, 233–240 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Singular integrals supported on the image of an analytic function *

Linda Saal and Marta Urciuolo

Facultad de Matemática, Astronomía y Física, Av. Valparaíso y R. Martínez, Ciudad Universitaria, 5000 Córdoba, Argentina

Received August 24, 1989; in final form May 2, 1990

1 Introduction

Ricci and Stein studied singular integral operators whose kernels are certain homogeneous distributions of critical degree supported on a homogeneous analytic submanifold of a simply connected nilpotent Lie group [6]. In this paper we study singular integral operators supported on the image of an analytic function $g: \mathbb{R}^m \to \mathbb{R}^n$ whose kernels $k \in C^\infty(\mathbb{R}^m - \{0\})$ have mean value zero and are homogeneous of critical degree with respect to a group of dilations $\{d_r\}$ of $\mathbb{R}^m$. More precisely, if $|\ |$ is a homogeneous norm associated to d_r then the operator T given by

$$Tf(x) = \text{p.v.} \int_B f(x - g(t))\, k(t)\, dt$$

is bounded on $L^p(\mathbb{R}^n)$, $1 < p < \infty$ where B denotes the unit ball with respect to $|\ |$ (Sects. 2 and 3).

In the case that k is a homogeneous kernel of degree $-m$ with respect to the usual dilations of $\mathbb{R}^m$, this result is contained in [2].

We also observe that the theorem still holds if we replace $\mathbb{R}^n$ by a simply connected nilpotent Lie group and k by a kernel of the form $k(t) = \sum_{j=1}^{\infty} 2^{jb} \Lambda_j(2^j t)$ for an adequate choice of $\{\Lambda_j\}_{j=1}^{\infty}$.

2 Some preliminaries

Let us consider $\mathbb{R}^n$ with the dilations given by $d_r(t_1, \ldots, t_m) = r \cdot t = (r^{b_1} t_1, \ldots, r^{b_m} t_m)\ \forall r > 0$, where $1 \leqq b_1 \leqq \ldots \leqq b_m$, $b_i \in \mathbb{R}\ \forall i$. Let $b = \sum_{i=1}^{m} b_i$ be the

* Supported in part by CONICET and CONICOR, Argentine.

homogeneous degree of $\mathbb{R}^m$ and let $|\ |$ be a homogeneous norm with respect to $\{d_r\}$ [4]. Let $B=\{t\in\mathbb{R}^m \,|\, |t|\leqq 1\}$ and $\sum=\{t\in\mathbb{R}^m \,|\, |t|=1\}$.

Let g be an analytic function from B to $\mathbb{R}^n$, $g(0)=0$, and such that the image of g generates $\mathbb{R}^n$, in the sense that it is not contained in any proper subspace of $\mathbb{R}^n$. We observe that since g is analytic, this is equivalent to the fact that the derivatives of g in the origin generate $\mathbb{R}^n$. First of all we want to associate to g an analytic function $g_0\colon B\to\mathbb{R}^n$ which will be homogeneous with respect to certain dilations on $\mathbb{R}^n$.

For a given multi-index $\mathscr{I}=(i_1, \ldots, i_m)$, we denote $|\mathscr{I}|=i_1+\ldots+i_m$, $d(\mathscr{I})=i_1b_1+\ldots+i_mb_m$ and $D^{\mathscr{I}}g=\partial^{|\mathscr{I}|}g/\partial t_1^{i_1}\ldots\partial t_m^{i_m}$. If $\mathscr{I}=(i_1, \ldots, i_m)$, $\mathscr{J}=(j_1, \ldots, j_m)$ we say $\mathscr{I}<\mathscr{J}$ if either $\sum_{k=1}^{m} i_kb_k<\sum_{k=1}^{m} j_kb_k$ or $\sum i_kb_k=\sum j_kb_k$ and $i_{k_0}<j_{k_0}$ where k_0 is the first index such that $i_{k_0}\neq j_{k_0}$. This is a well ordering for the set of multi-index. Following an idea in [8], let $\mathscr{I}_1$ be the first multi-index such that $D^{\mathscr{I}}g(0)\neq 0$ and set $e_1=D^{\mathscr{I}_1}g(0)$ and $a_1=d(\mathscr{I}_1)$. Chosen $e_1, \ldots, e_{k-1}$, let $\mathscr{I}_k$ be the first multi-index such that $\{e_1, \ldots, e_{k-1}, D^{\mathscr{I}_k}g(0)\}$ are linearly independent and set $e_k=D^{\mathscr{I}_k}g(0)$, $a_k=d(\mathscr{I}_k)$. Since g generates $\mathbb{R}^n$, we obtain a basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$. From now on we fix this basis.

Let $\{D_r\}_{r>0}$ be the group of dilations on $\mathbb{R}^n$ given by $D_r(x_1, \ldots, x_n)=(r^{a_1}x_1, \ldots, r^{a_n}x_n)$; we define $g_0(t)=\lim_{r\to 0} D_r^{-1}g(r\cdot t)\ \forall t\in B$. Then $g_0(r\cdot t)=D_rg_0(t)$ and $D^{\mathscr{I}_k}g_0(0)=e_k$, $k=1, \ldots, n$. More explicity, $g_0(t)=(P_{a_1}(t), \ldots, P_{a_n}(t))$, where P_{a_i} is a homogeneous polynomial with respect to $\{d_r\}$ of homogeneous degree a_i and $g(t)=(P_{a_1}(t)+\varphi_1(t), \ldots, P_{a_n}(t)+\varphi_n(t))$ where $\varphi_i(t)$ is an analytic function and

$$\lim_{t\to 0}\frac{|\varphi_i(t)|}{|t|^{a_i}}=0. \tag{2.1}$$

This implies that the Taylor expansion of φ_i must be $\sum_{\langle\mathscr{J}, b\rangle>a_i}\alpha_{\mathscr{J}}t^{\mathscr{J}}$ where $\langle\mathscr{J}, b\rangle=j_1b_1+\ldots+j_mb_m$ and thus if $\xi\in\mathbb{C}^m$ and $r\cdot\xi$ denotes the usual extension of d_r to $\mathbb{C}^m$ then $\frac{\varphi_i(r\cdot\xi)}{r^{a_i}}\xrightarrow[r\to 0]{}0$. Now we prepare several results which shall be used in the next section.

Note 2.2. There exists $N\leqq n$ such that the map given by $\phi(t^1, \ldots, t^N)=g(t^1)+\ldots+g(t^N)$ has rank n for a.e. $(t^1, \ldots, t^N)\in B^N$. This follows from the fact that g is an analytic function whose image generates $\mathbb{R}^n$. The same holds for

$$\phi_0(t^1, \ldots, t^N)=g_0(t^1)+\ldots+g_0(t^N).$$

In fact: There exist

$$j_1, \ldots, j_n\in\{1, \ldots, m\},\ s\cdot t.$$

(*) $\det(\partial_{j_1}g(t^1), \ldots, \partial_{j_n}g(t^n))\neq 0$ for a.e. $(t^1, \ldots, t^n)\in B^n$: Otherwise, since an analytic function which vanishes on a set of positive measure is identically zero, $\det(\partial_{i_1}g(t^1), \ldots, \partial_{i_n}g(t^n))=0$ for all $i_1, \ldots, i_n$ and $t^1, \ldots, t^n\in B^n$. This implies $\det(D^{(\alpha_1)}g(0), \ldots, D^{(\alpha_n)}g(0))=0$ for all multiindices $\alpha_1, \ldots, \alpha_n$ with $|\alpha_j|\geqq 1$, which contradicts the fact that $g(\mathbb{R}^m)$ generates $\mathbb{R}^n$. But (*) implies that (rank ϕ) $(t^1, \ldots, t^n)=n$ for a.e. $(t^1, \ldots, t^n)\in B^N$.

Note 2.3. Let $\mathscr{J}(t^1, \ldots, t^N) = (\sum_i |A^i(t^1, \ldots, t^N)|^2)^{1/2}$ where A^i are the minors of order n of the Jacobian matrix of ϕ. Define $\phi_k(t^1, \ldots, t^N) = D_{2^k}\phi(2^{-k}\cdot t^1, \ldots, 2^{-k}\cdot t^N)$ and let $\mathscr{J}_k$, A_k^i and $\mathscr{J}_0$, A_0^i be the analogous functions for ϕ_k and ϕ_0, respectively. Then

$$\mathscr{J}_k(t^1, \ldots, t^N) \xrightarrow[k\to\infty]{} \mathscr{J}_0(t^1, \ldots, t^N)$$

uniformly on B^N. Indeed,

$$\phi_k(t^1, \ldots, t^N) = \sum_{j=1}^{N} (P_{a_1}(t^j), \ldots, P_{a_n}(t^j)) + \sum_{j=1}^{N} (2^{ka_1}\varphi_1(2^{-k}\cdot t^j), \ldots, 2^{ka_n}\varphi_n(2^{-k}\cdot t^j))$$

and

$$\frac{\partial}{\partial t_i}(2^{ka_l}\varphi_l(2^{-k}\cdot t)) = 2^{ka_l}2^{-kb_i}\frac{\partial \varphi_l}{\partial t_i}(2^{-k}\cdot t), \qquad l = 1, \ldots, n.$$

By 2.1 the Taylor expansion of $\dfrac{\partial \varphi_l}{\partial t_i}$ is a sum of homogeneous polynomials of homogeneous degree higher than $a_l - b_i$ and the observation follows.

Note 2.4. Since ϕ has rank n a.e. Proposition 2.1 in [6] states that if ψ is a C^1-function supported on B^N then the transported measured by ϕ of ψ is absolutely continuous and its density ρ satisfies an L^1-Hölder condition

$$\int_{\mathbb{R}^n} |\rho(x+y) - \rho(x)|\, dx < C\|y\|^\sigma,$$

with $\sigma > 0$ depending on ϕ, and $C > 0$ depending also on ψ. Furthermore we can see that

$$\int_{\mathbb{R}^n} |\rho(x+y) - \rho(x)|\, dx \leqq \tilde{C}\left(\int |\nabla \psi| + \int |\psi|\right)^\sigma \left(\int |\psi(t)|\, |\mathscr{J}(t)|^{\frac{2\sigma}{1-\sigma}} dt\right)^{1-\sigma}$$

for some $\sigma > 0$, with $\tilde{C}$ depending only on ϕ. This estimation is in [7].

Note 2.5. We will make use of lemma 2.1 in [5].

Let $t^0 \in \mathbb{R}^k$ and let $\mathscr{A} = \mathscr{A}(t^0, M)$ be the Banach space of analytic functions on the closed cube $t^0 + [-M, M]^k$ in $\mathbb{R}^k$ having an absolutely convergent power series expansion $f = \sum a_J (t - t^0)^J$ on this cube, where the norm on $\mathscr{A}$ is defined by

$$\|f\|_{\mathscr{A}} = \sum_J |a_J|\, M^{|J|}.$$

If $f \in \mathscr{A}$ and $\xi \in \mathbb{R}^k$, $\|\xi\| = 1$, let $f_\xi(t) = f(t^0 + t\xi)$, for $t \in \mathbb{R}$, and let $\deg(\xi) = \min\{j \in \mathbb{N}_0 : f_\xi^{(j)}(0) \neq 0\}$ if $f_\xi \neq 0$ and $\deg(\xi) = \infty$ if $f_\xi = 0$. Finally, let $\gamma(f) = \min_{\|\xi\| = 1} \{\deg(\xi)\}$. If $f \neq 0$ we clearly have $\gamma(f) < \infty$.

Lemma [5]. *Let $f\in\mathscr{A}$, $f\neq 0$ and let $0<\alpha<\frac{1}{\gamma(f)}$. Then there exist constants $\beta>0$, $\delta>0$ and $C>0$ such that*

$$\int_{\|t-t^0\|<\beta} |f(t)+h(t)|^{-\alpha}dt\leqq C$$

for every function h in $\mathscr{A}$ with $\|h\|_{\mathscr{A}}\leqq\delta$.

Lemma 2.6. *Let $r\cdot t=(r^{b_1}t_1,\ldots,r^{b_k}t_k)$ a group of dilations in $\mathbb{R}^k$, $b=b_1+\ldots+b_k$. Let $p>0$, $A_r(t)=A_0(t)+r^{-p}R(r\cdot t)$ where A_0 and R are analytic functions in the unit ball B of $\mathbb{R}^k$ and such that A_0 is homogeneous of degree p and $r^{-p}R(r\cdot t)\xrightarrow[r\to 0]{}0$ uniformly on B.*

Then there exist $\alpha>0$, $C>0$ such that $\int_B |A_r(t)|^{-\alpha}dt\leqq C$ independent of r.

Proof. Since $r^{-p}R(r\cdot t)\to 0$ uniformly on B, there exists M such that

$$R(t)=\sum_{\langle J,b\rangle>p} a_J t^J \quad \forall\,\|t\|\leqq M.$$

Let $G_r(t)=r^{-p}R(r\cdot t)=r^{-p}\sum_{\langle J,b\rangle>p} a_J r^{\langle J,b\rangle}t^J\ \forall\,\|t\|\leqq M$ and then $\|G_r\|_{\mathscr{A}}\xrightarrow[r\to 0]{}0$.

Then there exist $\beta>0$, $\alpha>0$, $C>0$ and $r_0>0$ such that

$$\int_{\|t\|<\beta} |A_0(t)+G_r(t)|^{-\alpha}dt<c\ \forall r\leqq r_0.$$

We now choose r_1 such that $r_1\cdot B\subset\{t/\|t\|<\beta\}$ and by a change of variable

$$\int_B |A_0(t)+r^{-p}R(r\cdot t)|^{-\alpha}dt\leqq r_1^{p\alpha-b}\int_{\|s\|<\beta}|A_0(s)+G_{r/r_1}(s)|^{-\alpha}ds\leqq c r_1^{p\alpha-b}\quad\square$$

3 The main result

We denote by $\|\ \|_n$ a homogeneous norm in $\mathbb{R}^n$ associated with the group of dilations D_r.

Lemma 3.1. *Let $\{f_j\}_{j>0}$ be a family of functions defined on $\mathbb{R}^n$ satisfying*

a) $\int f_j=0$

b) $\int|f_j(x+y)-f_j(x)|\,dx\leqq c\,2^{j\varepsilon}|y|_n^{\varepsilon}$ *for some* $0<\varepsilon<1$

c) $\operatorname{supp} f_j\subset\{x/|x|_n\leqq c\,2^{-j}\}$

d) $\|f_j\|_1\leqq c$.

If v is the singular kernel given by $v=\sum_{j>0} f_j$ then the operator of convolution by v is bounded on $L^p(\mathbb{R}^n)$, $1<p<\infty$ with a norm that only depends on p, ε and c.

Sketch of the proof. The boundedness on $L^2(\mathbb{R}^n)$ follows by using Cotlar's lemma. To obtain the weak type $1-1$ we check that $\int_{|x|_n \geq c'|y|_n} |v(x+y)-v(x)|\,dx \leq c_\varepsilon$ [3]; the lemma follows by interpolation. □

For a more detailed proof see [2].

Theorem 3.2. *Let k be a $C^\infty(\mathbb{R}^m-\{0\})$ function, homogeneous with respect to $\{d_r\}_{r>0}$ of degree $-b$ and such that $\forall 0<\varepsilon<\varepsilon'<1 \int_{\varepsilon<|t|<\varepsilon'} k(t)\,dt=0$.*

Let $g: B \to \mathbb{R}^n$ be an analytic function such that the image of g generates $\mathbb{R}^n$ and $g(0)=0$. Then the operator defined by

$$Tf(x)=p\cdot v \int_{|t|<1} f(x-g(t))\,k(t)\,dt \tag{3.3}$$

is bounded on $L^p(\mathbb{R}^n)$ $1<p<\infty$.

Proof. We decompose T as in [6]. We take $\theta \in C_0^\infty(1/2, 2)$ such that

$$\sum_{l=1}^{\infty} k(t)\,\theta(2^l|t|)=\sum_{l=1}^{\infty} k_l(t)=k(t), \qquad t\in B.$$

Let μ_l be the measure defined by $\mu_l(f)=\int_{|t|<1} f(g(t))\,k_l(t)\,dt$. We also define $\theta_0(x)=\theta(|x|_n)$, $\theta_l(x)=2^{la}\theta_0(D_{2^l}x)=2^{la}\theta(2^l|x|_n)$, where $a=a_1+\ldots+a_n$ is the homogeneous degree of $\mathbb{R}^n$. If $c=\int\theta(y)\,dy$ and $\eta_l(x)=c^{-1}(\theta_{l+1}(x)-\theta_l(x))$ then for each fixed j, we have that $\delta=\sum_{l=j}^{\infty}\eta_l+c^{-1}\theta_j$ in the sense of distributions and thus

$$Tf=\sum_{j=1}^{\infty}\mu_j * f=\sum_{j=1}^{\infty}\sum_{l=j}^{\infty}\eta_l*\mu_j*f+c^{-1}\sum_{j=1}^{\infty}\theta_j*\mu_j*f$$

$$=\sum_{k=0}^{\infty}\sum_{j=1}^{\infty}\eta_{k+j}*\mu_j*f+c^{-1}\sum_j\theta_j*\mu_j*f=\sum_{k\geq 0}M_k f+Lf.$$

If we check that

$$\forall 0<\varepsilon<1,\ \|M_k\|_{p,p}\leq C_{\varepsilon,p}2^{\varepsilon k} \quad \text{and} \quad \|L\|_{p,p}\leq C \tag{3.4}$$

and

$$\exists \sigma>0 \quad \text{such that} \quad \|M_k\|_{2,2}\leq C2^{-\sigma k} \tag{3.5}$$

then the theorem follows by interpolation and duality arguments.

3.4 holds by lemma 3.1 since

$$\int|\eta_{k+j}*\mu_j(x+y)-\eta_{k+j}*\mu_j(x)|\,dx\leq c2^{(k+j)\varepsilon}|y|_n^{\varepsilon}\ \forall 0<\varepsilon<1. \quad \text{See [6]}$$

The proof of 3.5 is based on Cotlar's lemma and the iteration argument used in [1].

$$\|M_k\|_{2,2}\leq C2^{-\sigma k} \quad \text{for some } \sigma>0$$

provided that

$$\|\eta_{k+j} * \mu_j * \mu_l^* * \eta_{k+l}^*\|_{2,2} \leqq c\, 2^{-\sigma k} 2^{-|j-l|}.$$

We will check this for $j > l$

$$\|\eta_{k+j} * \mu_j * \mu_l^* * \eta_{k+l}^*\|_{2,2} \leqq c\, \|\eta_{k+j} * \mu_j * \mu_l^*\|_{2,2}$$

and

$$\begin{aligned}\|\eta_{k+j} * \mu_j * \mu_l^*\|_{2,2} &= \|\eta_{k+j} * \mu_j * \mu_l^* * \mu_l * \mu_j^* * \eta_{k+j}^*\|_{2,2}^{1/2} \\ &\leqq \|\eta_{k+j} * \mu_j\|_{2,2}^{1/2}\, \|\mu_l^* * \mu_l * \mu_j^* * \eta_{k+j}^*\|_{2,2}^{1/2} \leqq \dots \\ &\leqq \|\eta_{k+j} * \mu_j\|_{2,2}^{1/2^{-q}}\, \|(\mu_l^* * \mu_l)^{2^q-1} * \mu_j^* * \eta_{k+j}\|_{2,2}^{2^{-q}}.\end{aligned}$$

Let N and ϕ be as in 2.2 and $\psi_l(t^1, \dots, t^N) = \prod_{i=1}^{N} k_l(t^i)$.

Since $\mu_l * \dots * \mu_l$ is the transported measure by ϕ of ψ_l, we choose q such that $(\mu_l^* * \mu_l)^{2^q-1}$ be absolutely continuous with a density ρ_l satisfying the inequality in 2.4. Now we set $N = 2^q - 1$.

Thus we must check

$$\|\rho_l * \mu_j^* * \eta_{k+j}^*\|_1 \leqq c\, 2^{-\sigma k} 2^{(l-j)\sigma} \quad \text{for some } \sigma > 0 \tag{3.6}$$

We recall that $\phi_l(t^1, \dots, t^N) = D_{2^l}\phi(2^{-l}\cdot t^1, \dots, 2^{-l}\cdot t^N)$ as in 2.3. We define $\tilde{\psi}(t^1, \dots, t^N) = 2^{-lNb}\psi_l(2^{-l}\cdot t^1, \dots, 2^{-l}\cdot t^N)$ so that $\tilde{\rho}_l(y) = 2^{-la}\rho_l(D_{2^{-l}}y)$ is the density of the transported measure of $\tilde{\psi}$ by ϕ_l and $\operatorname{supp}\tilde{\psi} \subset \{(t^1, \dots, t^N)/\tfrac{1}{2} \leqq |t^i| \leqq 2\}$.

Notice that $\tilde{\psi}(t^1, \dots, t^N) = \prod_{1}^{N} k_0(t^j)$ is independent of l.

If we prove

$$\int_{R^n} |\tilde{\rho}_l(x+y) - \tilde{\rho}_l(x)|\, dx \leqq c\, |y|_n^{\sigma} \quad \text{for some } c, \sigma > 0 \tag{3.7}$$

independent of l, then $\int |\rho_l(x+y) - \rho_l(x)|\, dx \leqq c\, 2^{l\sigma} |y|_n^{\sigma}$ and 3.6 follows since $\rho_l * \mu_j^*$ satisfies an L^1-Hölder condition with constant $c\, 2^{l\sigma}$ and $\operatorname{supp}\eta_{k+j} \subset \{x/|x|_n < c\, 2^{-(k+j)}\}$.

By 2.4

$$\int |\tilde{\rho}_l(x+y) - \tilde{\rho}(x)|\, dx \leqq c\, \|y\|^{\sigma} \Big(\int_{\operatorname{supp}\tilde{\psi}} |\tilde{\psi}| + |\nabla\tilde{\psi}| \Big)^{\sigma} \Big(\int_{\operatorname{supp}\tilde{\psi}} \frac{|\tilde{\psi}|}{|\mathscr{J}_l|^{\frac{2\sigma}{1-\sigma}}} \Big)^{1-\sigma}$$

with c independent of l, for some $\sigma > 0$.

Thus 3.7 follows if we prove

$$\int_{\operatorname{supp}\tilde{\psi}} |\mathscr{J}_l(t)|^{-\alpha}\, dt \leqq c \quad \text{independent of } l, \quad \text{for some } \alpha > 0. \tag{3.8}$$

Let A_0 be a minor of $\mathscr{J}_0$ such that $A_0 \not\equiv 0$ a.e. and let A_l, A be the corresponding minors for $\mathscr{J}_l$ and $\mathscr{J}$ respectively. A_0 is a homogeneous polynomial of homogeneous degree $q = a - i_1 b_1 - \dots - i_m b_m$, $i_1 + \dots + i_m = n$ where i_j is the number of columns of A_0 corresponding to derivate with respect to t_j.

Since $A_l(t^1, \ldots, t^N) = 2^{lq} A(2^{-l} \cdot t^1, \ldots, 2^{-l} \cdot t^N)$ and $A_l(t^1, \ldots, t^N) \to A_0(t^1, \ldots, t^N)$ uniformly on B^N by 2.3, it is easy to see that $A(t) = A_0(t) + R(t)$ where R is an analytic function such that $r^{-q} R(r \cdot t^1, \ldots, r \cdot t^N) \xrightarrow[r \to 0]{} 0$ uniformly on B^N.

$$\begin{aligned} \int_{\operatorname{supp} \tilde{\psi}} |\mathscr{J}_l(t^1, \ldots, t^N)|^{-\alpha}{}_{dt} &\leqq \int_{\operatorname{supp} \tilde{\psi}} |A_l(t^1, \ldots, t^N)|^{-\alpha} dt \\ &= \int_{\operatorname{supp} \tilde{\psi}} |2^{lq}(A(2^{-l} \cdot t^1, \ldots, 2^{-l} \cdot t^N))|^{-\alpha} dt \\ &= \int_{\operatorname{supp} \tilde{\psi}} |A_0(t) + r^{-q} R(r \cdot t)|^{-\alpha} dt \end{aligned}$$

where $t = (t^1, \ldots, t^N)$, $r = 2^{-l}$ and $r \cdot t = (r \cdot t^1, \ldots, r \cdot t^N)$. Now the theorem follows from lemma 2.6. □

Remark 1. If g generates a proper subspace of $\mathbb{R}^n$, the theorem still holds. ([6], pag. 31).

Remark 2. Let $k(t) = \sum_{j \in \mathbb{N}} 2^{jb} \Lambda_j(2^j \cdot t)$ with Λ_j satisfying

1) $\int \Lambda_j = 0$.
2) There exist $c_1 > 0$, $c_2 > 0$ and a compact set K such that for every j, $\Lambda_j \in C_0^1(K)$, $|\Lambda_j| \leqq c_1$, and $|\nabla \Lambda_j| \leqq c_2$.

Then if g is like in theorem 3.2, the operator T defined by $Tf(x) = \text{p.v.} \int_B f(x - g(t)) k(t) dt$ is bounded on $L^p(\mathbb{R}^n)$, $1 < p < \infty$. In the case that k is homogeneous of degree $-b$ with respect to d_r, $\Lambda_j = \Lambda_0 = k\theta \; \forall j$.

Remark 3. Assume now that G is a simply connected, nilpotent Lie group. Let $g: B \subset \mathbb{R}^m \to G$ be a real analytic function such that the image of g generates G in the sense that it is not contained in any proper closed subgroup of G.

Let k be in $\mathscr{C}^\infty(\mathbb{R}^m - \{0\})$, with mean value zero and homogeneous with respect to $\{d_r\}$ of degree $-b$.

Theorem 3.9. *The operator T given by $Tf(x) = \text{p.v.} \int_B f(x g(t)^{-1}) k(t) dt$ is bounded on $L^p(G)$, $1 < p < \infty$.*

Sketch of the proof. To prove the analogue of 2.2, we assume that g has maximal rank K at t_0. Thus there exists a neighborhood of t_0 such that its image is a submanifold W of dim K. An analytic continuation argument implies that W generates G. Then we can apply proposition 1.1 in [6] which states that if W is a connected analytic submanifold of G that generates G then there exists N such that the map $\phi(w_1, \ldots, w_N) = w_1 \ldots w_N$ has rank n at some point $(w_1, \ldots, w_N) \in W^N$. Thus $\phi(t^1, \ldots, t^N) = g(t^1) \ldots g(t^N)$ has rank n for $a \cdot e(t_1, \ldots, t^N) \in B^N$.

Let $\mathscr{G}$ be the Lie algebra of G. We may identify $\mathscr{G}$ and G as manifolds and assume that exp = id. Define dilations $\{D_r\}_{r>0}$ on the Euclidean space $\mathscr{G}$ which are induced by the dilations on $\mathbb{R}^m$ via the mapping g as in Sect. 2. By passing to a suitable free nilpotent group which has G as a quotient we may actually assume that the dilations $\{D_r\}$ are group-resp. Lie algebra automorphisms (cf. [6] for details about this method). By introducing coordinates on $\mathscr{G}$ as in Sect. 2 we may identify $\mathscr{G}$ and G with $\mathbb{R}^n$. Then a multiple

product $x^1 \dots x^N$ in G is given by $x^1 \dots x^N = Q(x^1, \dots, x^N)$, where Q is a polynomial mapping $Q=(Q_1, \dots, Q_n)$ and Q_k is homogeneous of degree a_k, i.e. $Q_k(D_r x^1, \dots, D_r x^N) = r^{a_k} Q_k(x^1, \dots, x^N)$. Moreover, as in Sect. 2 $g_i = P_{a_i} + \varphi_i$, where $P_{a_i}(r \cdot t) = r^{a_i} P_{a_i}(t)$ and where φ_i has a Taylor expansion of the form $\sum_{\langle J, b \rangle > a_i} \alpha_J t^J$. Together this implies

$$\begin{aligned} \phi(t^1, \dots, t^N) &= g(t^1) \dots g(t^N) \\ &= P(t^1) \dots P(t^N) + R(t^1, \dots, t^N), \end{aligned}$$

where $P=(P_{a_1}, \dots, P_{a_n})$, and where $R_k(t)$ is a finite linear combination of the form

$$\prod_{i,j} P_{a_i}(t^j)^{u_{ij}} \varphi_i(t^j)^{v_{ij}},$$

where $\sum_{i,j} a_i(u_{ij} + v_{ij}) = d_k$ and where at least one $v_{ij} \neq 0$. Thus,

$$\begin{aligned} \phi_k(t^1, \dots, t^N) &= P(t^1) \dots P(t^N) + D_{2^k} R(2^{-k} t^1, \dots, 2^{-k} t^N) \\ &= \phi_0(t^1, \dots, t^N) + R_k(t^1, \dots, t^N), \end{aligned}$$

where R_k is a polynomial in $(t^1, \dots, t^N)$ whose coefficients are $O(2^{-k})$. This easily implies the analogue of 2.3, at least in the case where the $\{D_r\}$ are automorphisms.

The rest of the proof of Theorem 3.9 follows along the lines in [6], with the modifications explained in the previous section, where the case of non-automorphic dilations can be handled by transference as in [6] or [7]. □

Acknowledgements. We wish to express our thanks to Dr. Fulvio Ricci for his invaluable guidance. We are also indebted to Dr. Tomás Godoy and Dr. Detlef Müller who generously helped us. We also thank the referee for his suggestions which led us to an improvement of this manuscript.

References

1. Christ, M.: Hilbert transform along curves I. Nilpotent groups. Ann. Math., II. Ser. **122**, 575–596 (1985)
2. Christ, M., Nagel, A., Stein, E., Wainger, S.: Singular and Maximal Radon Transforms. (Preprint)
3. Coifman, R.R., Weiss, G.: Analyse harmonique non-commutative sur certains espaces homogenes. (Lect. Notes Math., vol. 242). Berlin Heidelberg New York: Springer 1971
4. Folland, G., Stein, E.: Hardy Spaces on Homogeneous Groups. Princeton: Princeton University Press 1982
5. Müller, D.: Singular kernels supported by homogeneous submanifolds. J. Reine Angew. Math. **356**, 90–118 (1985)
6. Ricci, F., Stein, E.: Harmonic Analysis on Nilpotent Groups and Singular Integrals II. Singular Kernels supported on Submanifolds. J. Funct. Anal. **78**, 56–84 (1988)
7. Saal, L.: A result about the Hilbert transform along curves. Proceedings of the A.M.S. (to appear)
8. Stein, E., Wainger, S.: Problems in harmonic analysis related to curvature. Bull. Am. Math. Soc. **84**, 1239–1295 (1978)

Math. Z. 206, 241–254 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

The existence of algebraic surfaces with preassigned Chern numbers

Zhijie Chen *

Department of Mathematics, East China Normal University, 200062 Shanghai, People's Republic of China

Received October 20, 1989; in final form March 10, 1990

To Cao Xi-Hua on his 70th Birthday

Introduction

Is is well-known that there is a unique minimal smooth model in each birational class of complex algebraic surfaces of general type. Any smooth algebraic surface of general type can be obtained by successive blowing-ups from a certain minimal surface. The Chern numbers $c_1^2(S)$ and $c_2(S)$ are two fundamental discrete invariants of a smooth surface S. The minimal surfaces with same Chern numbers may be divided into a finite number of families, or more precisely, there exists a quasi-projective coarse moduli scheme for the minimal surfaces of general type with fixed Chern numbers. By Noether's formula

$$c_1^2(S)+c_2(S)=12\chi(\mathcal{O}_S),$$

where $\chi(\mathcal{O}_S)=h^0(\mathcal{O}_S)-h^1(\mathcal{O}_S)+h^2(\mathcal{O}_S)$ is the Euler-Poincaré characteristic of $\mathcal{O}_S$, and by the equality

$$c_1^2(S)=K_S^2,$$

where K_S is the canonical divisor of S, the Chern numbers $(c_1^2(S), c_2(S))$ are equivalent to the discrete invariants $(K_S^2, \chi(\mathcal{O}_S))$. Since sometimes it is more convenient to use K_S^2 and $\chi(\mathcal{O}_S)$ than $c_1^2(S)$ and $c_2(S)$, by abuse of language, we will refer $(K_S^2, \chi(\mathcal{O}_S))$ as the Chern numbers of the surface S in this paper.

The Chern numbers of a minimal surface of general type are restricted by the following inequalities:

$$2\chi(\mathcal{O}_S)-6\leqq K_S^2\leqq 9\chi(\mathcal{O}_S),$$
$$K_S^2,\ \chi(\mathcal{O}_S)>0,$$

It is natural to ask the following question of existence: For any pair of positive integers (a, b) satisfying the inequality $2a-6\leqq b\leqq 9a$, can we find a minimal surface S of general type with $K_S^2=b$, $\chi(\mathcal{O}_S)=a$?

* Supported by National Natural Science Foundation of China

The aim of this article is to establish the following statement:

Theorem 1. *For every pair of positive integers* (a, b) *satisfying*

$$2a-6 \leqq b \leqq 9a-C,$$

where C *is a constant less than 9327, there exists a minimal surface* S *of general type with* $K_S^2=b$, $\chi(\mathcal{O}_S)=a$.

By a careful computer check this result can be improved as follows.

Theorem 2. (1) *For every pair of positive integers* (a, b) *satisfying*

$$2a-6 \leqq b \leqq 9a-347,$$

or

$$9a-300 \leqq b \leqq 9a-345, \quad b \neq 9a-k,$$
$$k=301, 304, 305, 307, 312, 316, 319, 331,$$

there exists a minimal surface S *of general type with* $K_S^2=b$, $\chi(\mathcal{O}_S)=a$.

(2) *For every positive integer* k *such that*

$$k=0, 15, 30, 34, 39, 45, 47, 49, 54, 58, 60, 62, 64, 65, 66, 68, 69, 73, 75,$$

and

$$k \geqq 77, \quad k \neq 82, 87, 91, 106, 121,$$

there exist infinitely many minimal surfaces S *of general type with* $K_S^2=9\chi(\mathcal{O}_S)-k$.

Thus there remains only the border region to be filled in.
Historically it is Persson [5] who first raised this question and proved in 1981 the following theorem.

Theorem (Persson). *Let* a, b *be positive integers satisfying*

$$2a-6 \leqq b \leqq 8a,$$
$$b \neq 8a-k,$$

where $k=2$, 19 *or* k *is odd and* $1 \leqq k \leqq 15$, *then there exists a minimal surface of general type* S *with* $K_S^2=b$, $\chi(\mathcal{O}_S)=a$.

The strategy of Persson is simple but effective. First he established a lattice of invariants, then by imposing singularities he perturbed these surfaces to fill in the gaps. In Persson's original work, he dealt with double coverings and induced singularities via singularities of branch curves. Thus Persson's examples are all genus two fibrations. He had conjectured and later Xiao [6] proved that for any minimal surface S with a genus two fibration its Chern numbers must satisfy $K_S^2 \leqq 8\chi(\mathcal{O}_S)$. That is why the upper bound of the slope appeared in Persson's theorem is equal to 8. By the same strategy, Chen [2] used hyperelliptic fibrations to raise the upper bound of the slope to 8.757 asymptotically.

Theorem (Chen). *For every positive integers* a *and* b *satisfying*

$$2a-6 \leqq b < 8.757a-865.7a^{2/3},$$
$$a > C,$$

where C is a large constant, there exists a minimal surface S of general type with $K_S^2=b$, $\chi(\mathcal{O}_S)=a$.

Recently, Xiao set forth an idea to fill in the unknown region $b>8a$. He suggested that beginning with a fibered surface with $K^2=9\chi$, it would be possible to construct a series of new surfaces by making base extensions and twisting general and special fibres (see [7]). Basically this is still the same strategy. But instead of inducing singularities by coverings, we will impose singularities by quotients. To this end, first we make a cyclic base extension, then take the quotient with respect to a cyclic group of automorphisms so that the quotient surface has Hirzebruch-Jung singularities. In this construction which is described in Sect. 1, a general fibre of the desingularization of the quotient surface is isomorphic to the corresponding one in the original surface. Only a finite number of fibres are changed, they are said to be twisted. Hence the Chern numbers are perturbed. The next thing is to construct suitable cyclic coverings of the base curve such that the resulted quotient singularities could be manipulated by changing the branch divisor of the covering. This is stated in a technical lemma of Sect. 2. Sect. 3 is devoted to describe a Horikawa's surface with extreme Chern numbers (i.e. $K_S^2=9\chi(\mathcal{O}_S)$) which appeared in an unpublished paper of Inoue [4, § 6]. This surface is the starting point of the construction. The main theorems will be shown in Sect. 4.

Finally, the author would like to thank the referee for his valuable suggestions.

1. Construction of new fibered surfaces by twisting fibres

In this section we shall describe the technique of constructing new fibered surfaces by twisting fibres of an old one.

At first we fix our notations. Let X be a smooth complex projective surface with a fibration $f: X\to C$ (i.e. a surjective morphism to a smooth curve C with connected fibres) and X has a C-automorphism $\gamma\in \mathrm{Aut}_C(X)=\{\gamma\in \mathrm{Aut}(X)\mid f\circ\gamma=f\}$. Denote by F a general fibre of f and set

$$g:=g(F),$$
$$b:=g(C),$$
$$m:=\#(\langle\gamma\rangle).$$

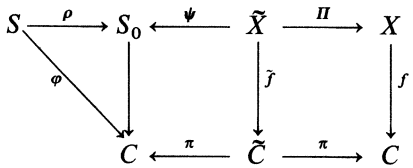
We construct an m-fold cyclic covering $\pi: \tilde{C}\to C$ with totally ramified points $P_1,\ldots,P_s\in\tilde{C}$. For simplicity, we will suppose that all the fibres $f^{-1}(\pi(P_i))$ are smooth, it is sufficient for our purpose. Hence the pull-back $\tilde{X}=X\times_C\tilde{C}$ is a smooth surface with a fibration $\tilde{f}: \tilde{X}\to\tilde{C}$ and we have a Cartesian square:

$$\begin{array}{ccc} \tilde{X}=X\times_C\tilde{C} & \xrightarrow{\Pi} & X \\ \downarrow{\scriptstyle \tilde{f}} & & \downarrow{\scriptstyle f} \\ \tilde{C} & \xrightarrow{\pi} & C \end{array}$$

Now we try to look for an automorphism σ of the surface $\tilde{X}$ such that the quotient $\tilde{X}/\langle\sigma\rangle$ has Hirzebruch-Jung singularities (please refer to [1, Ch. III,

§5] for notions and about cyclic quotients and Hirzebruch-Jung singularities). H. Cartan's theorem claims that every action of a finite group on a manifold can locally be linearized, that is, for a cyclic group $\langle\sigma\rangle\cong\mathbb{Z}_n$ acting on a surface, locally the action of σ on a fixed point P can be expressed as $\sigma^*(x, y)=(\zeta^a x, \zeta^b y)$ where ζ is an n-th root of unity. If either a or b vanishes, then P is not isolated and the quotient will be smooth at P. Hence only the isolated fixed points can induce Hirzebruch-Jung singularities.

Let us examine some automorphisms of $\tilde{X}$. The C-automorphism γ has a pull-back $\tilde{\gamma}\in\mathrm{Aut}_{\tilde{C}}(\tilde{X})\subseteq\mathrm{Aut}_C(\tilde{X})$. On the other hand, the group $\mathrm{Aut}_C(\tilde{C})=\langle\beta\rangle$ is cyclic of order m. Its generator β has a pull-back $\tilde{\beta}\in\mathrm{Aut}_X(\tilde{X})\subseteq\mathrm{Aut}_C(\tilde{X})$ which commutes with $\tilde{\gamma}$, i.e. $\tilde{\beta}\tilde{\gamma}=\tilde{\gamma}\tilde{\beta}$. Although automorphisms $\tilde{\beta}$ and $\tilde{\gamma}$ have no isolated fixed points, their product $\tilde{\beta}\tilde{\gamma}$ has only isolated fixed points. Therefore the automorphism $\tilde{\beta}\tilde{\gamma}$ is just what we wanted. Let $G=\langle\tilde{\beta}\tilde{\gamma}\rangle$ which is a cyclic group of order m. Taking the normalization S_0 of the quotient $\tilde{X}/G$, after minimal desingularization, we can obtain a smooth surface S with a fibration $\varphi: S\to C$. The fibre $\varphi^{-1}(P)$ are naturally isomorphic to $f^{-1}(P)$ for any $P\in C-\{\pi(P_1), \ldots, \pi(P_s)\}$.



As our object is to construct surfaces with different Chern numbers, we will calculate the Chern numbers of the fibered surface S. We may denote by $\Sigma=\{Q_1, \ldots, Q_t\}$ the set of singularities on the normal surface S_0. To simplify the calculation, for our purpose it is sufficient to assume that m is a prime. Hence every singularity Q_i is of type $A_{m,q}$.

At first we calculate the Euler number $e(S)$ as follows:

$$e(\tilde{X})=me(X)-(m-1)se(F),$$

$$e(\tilde{X})=me(S-\rho^{-1}(\Sigma))+e(\Sigma)=me(S)+t-m\sum_{i=1}^{t}e(\rho^{-1}(Q_i)),$$

$$e(S)=e(X)+\frac{2}{m}s(m-1)(g-1)+\sum_{i=1}^{t}\left(e(\rho^{-1}(Q_i))-\frac{1}{m}\right).$$

To calculate K_S^2, we introduce a rational divisor $E_Q\in\mathrm{Div}(S)\otimes\mathbb{Q}$ for a singularity $Q\in\Sigma$ which satisfies two conditions: (1) $\mathrm{supp}\,E_Q\subseteq\rho^{-1}(Q)$; (2) $(K_S+E_Q)\cdot D=0$ for any divisor D with $\mathrm{supp}\,D\subseteq\rho^{-1}(Q)$. Later we will show the existence and uniqueness of this rational divisor E_Q. Now we have

$$K_{\tilde{X}}^2=m\left(K_X+\frac{(m-1)s}{m}F\right)^2=mK_X^2+4s(m-1)(g-1),$$

$$K_{\tilde{X}}^2=m\left(K_S+\sum_{i=1}^{t}E_{Q_i}\right)^2=m\left(K_S^2+\sum_{i=1}^{t}K_S E_{Q_i}\right),$$

$$K_S^2=K_X^2+\frac{4}{m}s(m-1)(g-1)-\sum_{i=1}^{t}K_S E_{Q_i}.$$

Then

$$\chi(\mathcal{O}_S)=\frac{1}{12}(K_S^2+e(S))$$

$$=\chi(\mathcal{O}_X)+\frac{s}{2m}(m-1)(g-1)+\frac{1}{12}\sum_{i=1}^{t}\left(e(\rho^{-1}(Q_i))-K_S E_{Q_i}-\frac{1}{m}\right).$$

Now we will investigate the singularity $Q\in\Sigma$ to determine the values of $e(\rho^{-1}(Q))$ and $K_S E_Q$.

If Q is a singularity of type $A_{m,q}$ such that

$$\frac{m}{q}=[|e_1|, |e_2|, \ldots, |e_r|],$$

then

$$\rho^{-1}(Q)=C_1\cup C_2\cup\ldots\cup C_r,$$

where C_i are rational curves with intersection numbers

$$\begin{aligned} C_i^2&=e_i\leqq -2,\\ C_i C_j&=\delta_{i,j-1}+\delta_{i-1,j}, \end{aligned}\qquad i,j=1,\ldots,r.$$

Hence

$$e(\rho^{-1}(Q))=r+1.$$

To determine the rational divisor $E_Q\in\mathrm{Div}(S)\otimes\mathbb{Q}$, we assume

$$E_Q=\sum_{i=1}^{r}a_i C_i,\qquad a_i\in\mathbb{Q},$$

$$M=\begin{pmatrix} e_1 & 1 & & 0\\ 1 & e_2 & & 0\\ & & \ddots & \\ 0 & 0 & & e_r \end{pmatrix}.$$

The condition

$$(K_S+E_Q)\cdot C_j=0,\qquad j=1,\ldots,r,$$

implies

$$E_Q\cdot C_j=\sum_{i=1}^{r}a_i C_i C_j=-K_S C_j=e_j+2,\qquad j=1,\ldots,r,$$

that is,

$$M\begin{pmatrix}a_1\\ \vdots\\ a_r\end{pmatrix}=\begin{pmatrix}e_1+2\\ \vdots\\ e_r+2\end{pmatrix},$$

$$\begin{pmatrix}a_1\\ \vdots\\ a_r\end{pmatrix}=M^{-1}\begin{pmatrix}e_1+2\\ \vdots\\ e_r+2\end{pmatrix}.$$

This implies the existence and uniqueness of E_Q.

What we need is the following intersection number

$$K_S E_Q = \sum_{i=1}^{r} a_i K_S C_i = \sum_{i=1}^{r} a_i(-e_i-2)$$

$$= -(e_1+2, \ldots, e_r+2) M^{-1} \begin{pmatrix} e_1+2 \\ \vdots \\ e_r+2 \end{pmatrix}.$$

2. Cyclic coverings of a curve

In this section, we will prove a lemma about cyclic coverings of a curve. Our object is to construct a cyclic covering $\tilde{C}$ of C such that there exists a C-automorphism we expect.

Assume that an automorphism σ of order m acts on a surface $\tilde{C}$ with fixed points Q_i. Then locally at Q_i we have $\sigma^*(z_i)=\zeta_i z_i$ where z_i is a local coordinate centered at Q_i and ζ_i is an m-th root of unity. Furthermore we have some compatibility conditions on ζ_i. This observation provided the motivation for the following lemma.

Lemma 1. *Let C be a smooth complex projective curve, $P_1, \ldots, P_s \in C$ be distinct points, m a positive integer. If r integers $a_1, \ldots, a_r$ satisfying*

(1) $1 \leqq a_i \leqq m-1$ *for* $i=1, \ldots, r$,

(2) $(a_i, m)=1$ *for* $i=1, \ldots, r$,

(3) $m \Big| \sum_{i=1}^{r} a_i$,

are given, then there exists an m-fold cyclic covering $\pi: \tilde{C} \to C$ with totally ramified branch points $P_1, \ldots, P_s$. Furthermore, for every C-automorphism $\beta \in \mathrm{Aut}_C(\tilde{C})$, there exist analytic local coordinates z_k centered at $\pi^{-1}(P_k)$ $(1 \leqq k \leqq s)$ such that $\gamma^* z_k = \zeta^{b_k} z_k$, where ζ is an m-th root of unity and $a_k b_k \equiv 1 \pmod m$.

Proof. To prove this lemma, what we wanted to do is to construct an $\mathcal{O}_C$-algebra $\mathcal{A}$ such that locally $\mathcal{A}|_{U_\alpha} \cong \mathcal{O}_{U_\alpha}[T]/(T^m - f_\alpha)$. It is well-known that $\tilde{U}_\alpha := \mathrm{Spec}\, \mathcal{O}_{U_\alpha}[T]/(T^m - f_\alpha) \to U_\alpha$ is an m-fold cyclic covering with branch points at the zeros of f_α. $\tilde{U}_\alpha$ is non-singular if and only if f_α has only simple zeros. Furthermore, for any automorphism $\sigma \in \mathrm{Aut}_{U_\alpha}(\tilde{U}_\alpha)$ and for every ramification point $Q_k \in \tilde{U}_\alpha$, locally we have $\sigma^*(z_k)=\zeta z_k$, where ζ is an m-th root of unity and does not depend on the choice of Q_k. Hence the key of the proof is to choose suitably the f_α's and to find a way to glue these local $\mathcal{O}_{U_\alpha}$-algebras properly to form a global $\mathcal{O}_C$-algebra.

There exists an invertible sheaf $\mathcal{L}$ on C, such that

$$\mathcal{L}^m = \mathcal{O}_C\left(\sum_{i=1}^{r} a_i P_i\right).$$

Let

$$\mathcal{L}_k = \mathcal{L}^{-k} \otimes \mathcal{O}_C\left(\sum_{i=1}^{r} \left[\frac{k a_i}{m}\right] P_i\right), \qquad k=1, \ldots, m-1,$$

$$\mathcal{A} = \mathcal{O}_C \oplus \mathcal{L}_1 \oplus \cdots \oplus \mathcal{L}_{m-1},$$

Then $\mathscr{A}$ is a locally free sheaf of rank m on the curve C. Now we will define a multiplication of $\mathscr{A}$ such that $\mathscr{A}$ becomes an $\mathcal{O}_C$-algebra. To this end we take

$$\begin{array}{l} s_i \in \Gamma(C, \mathcal{O}_C(P_i)), \\ \operatorname{div} s_i = P_i, \end{array} \quad \text{for } i=1, \ldots, r.$$

For any open subset $U \subseteq C$ and for sections $\alpha \in \Gamma(U, \mathscr{L}_k)$, $\beta \in \Gamma(U, \mathscr{L}_l)$, $0 \leqq k, l \leqq m-1$, define the product $\alpha \cdot \beta$ as

$$\alpha \cdot \beta = \alpha \otimes \beta \otimes \prod_{i=1}^{r} s_i^{\left[\frac{(k+l)a_i}{m}\right]-\left[\frac{ka_i}{m}\right]-\left[\frac{la_i}{m}\right]},$$

hence

$$\alpha \cdot \beta \in \begin{cases} \Gamma(U, \mathscr{L}_{k+l}) & \text{if } k+l<m, \\ \Gamma(U, \mathcal{O}_C) & \text{if } k+l=m, \\ \Gamma(U, \mathscr{L}_{k+l-m}) & \text{if } k+l>m. \end{cases}$$

It is not difficult to verify that the associative law holds for so defined multiplication, thus $\mathscr{A}$ becomes an $\mathcal{O}_C$-algebra. Let $\tilde{C} = \operatorname{Spec} \mathscr{A}$, then $\pi: \tilde{C} \to C$ is a finite morphism of degree m. We shall prove that π is a cyclic covering with $P_1, \ldots, P_r$ as totally ramified branch points.

Take an open cover $\{U_\alpha\}$ of C which have the following properties: (1) Each U_α contains at most one P_i; (2) Restricted to U_α, all the above defined invertible sheaves are isomorphic to $\mathcal{O}_{U_\alpha}$.

Under the isomorphism $\mathcal{O}_C(P_i)|_{U_\alpha} \xrightarrow{\sim} \mathcal{O}_C|_{U_\alpha}$, the image of the restriction $s_i|_{U_\alpha}$ of the global section s_i is denoted by $s_{i,\alpha} \in \Gamma(U_\alpha, \mathcal{O}_C)$. If $P_k \in U_\alpha$, then $P_i \notin U_\alpha$ for $i \neq k$, i.e. $s_{i,\alpha} \in \Gamma(U_\alpha, \mathcal{O}_C^*)$.

Let $0 \leqq b_k < m$ be the integer satisfying $b_k a_k \equiv 1 \pmod{m}$, then we have

$$\left[\frac{b_k a_k}{m}\right] = \frac{b_k a_k}{m} - \frac{1}{m}.$$

Denote by $t \in \Gamma(U_\alpha, \mathscr{L}_{b_k})$ the image of $1 \in \Gamma(U_\alpha, \mathcal{O}_C)$ under the isomorphism $\mathcal{O}_C|_{U_\alpha} \xrightarrow{\sim} \mathscr{L}_{b_k}|_{U_\alpha}$. By the multiplication rule of $\mathscr{A}$, we have

$$t^m = t \cdot t \cdot \ldots \cdot t \in \Gamma(U_\alpha, \mathcal{O}_C),$$

that is,

$$\begin{aligned} t^m &= \varepsilon_\alpha \prod_{i=1}^{r} s_i^{\left[\frac{mb_k a_i}{m}\right]-m\left[\frac{b_k a_i}{m}\right]} \\ &= \varepsilon_\alpha s_k \prod_{i=1, i\neq k}^{r} s_i^{b_k a_i - m\left[\frac{b_k a_i}{m}\right]}, \end{aligned}$$

where $\varepsilon_\alpha \in \Gamma(U_\alpha, \mathcal{O}_C^*)$. Since

$$\varepsilon_\alpha \prod_{i \neq k} s_i^{b_k a_i - m\left[\frac{b_k a_i}{m}\right]} \in \Gamma(U_\alpha, \mathcal{O}_C^*),$$

and

$$t^j \in \Gamma(U_\alpha, \mathcal{L}_{r_j}), \quad \text{here } r_j \equiv b_k j (\mathrm{mod}\, m), \quad j = 1, \ldots, m-1,$$

as an $\mathcal{O}_{U_\alpha}$-algebra $\mathcal{A}|_{U_\alpha} \cong \mathcal{O}_{U_\alpha}[t] \cong \mathcal{O}_{U_\alpha}[T]/(T^m - f_\alpha)$ where f_α differs from s_k by a unit factor. Therefore $\tilde{C}$ is non-singular and π is an m-fold cyclic covering with $P_i, \ldots, P_r$ as totally ramified branch points.

$\gamma \in \mathrm{Aut}_C(\tilde{C})$ implies $\gamma^* \in \mathrm{Aut}_{\mathcal{O}_C}(\mathcal{A})$, so $\gamma^* t = \zeta_\alpha t$ where ζ_α is some m-th root of unity. Moreover we have

$$\gamma^*(\mathcal{L}_i|_{U_\alpha}) = \mathcal{L}_i|_{U_\alpha} \quad \text{for } i = 1, \ldots, m-1,$$

in particular, for any

$$u \in \Gamma(V_\alpha, \mathcal{L}_1), \quad V_\alpha \subseteq U_\alpha,$$

we have

$$\gamma^* u = \zeta_\alpha^{a_k} u.$$

Similarly, assuming $P_1 \in U_\beta$, by the same argument we conclude that for any

$$v \in \Gamma(V_\beta, \mathcal{L}_1), \quad V_\beta \subseteq U_\beta,$$

one has

$$\gamma^* v = \zeta_\beta^{a_1} v = \zeta v,$$

where $\zeta := \zeta_\beta^{a_1}$ is an m-th root of unity. If we take an open subset $V \subseteq V_\alpha \cap V_\beta$, then for any $v \in \Gamma(V, \mathcal{L}_1)$, we have

$$\gamma^* v = \zeta_\alpha^{a_k} v = \zeta v,$$

hence

$$\zeta_\alpha = \zeta^{b_k}.$$

Now let U_α be a coordinate neighborhood of P_k with an analytic local coordinate z centered at P_k. Then the analytic local rings are

$$\mathcal{O}_{C, P_k}^{\mathrm{hol}} \cong \mathbb{C}\{z\},$$
$$\mathcal{O}_{\tilde{C}, \pi^{-1}(P_k)}^{\mathrm{hol}} \cong \mathbb{C}\{z\}[w],$$

where $w^m = z$. By the preceding argument we have

$$\gamma^* w = \zeta^{b_k} w.$$

3. A Horikawa's surface

In this section we shall recall a Horikawa's surface H which appeared in Inoue's preprint [4, § 6]. This surface has extreme Chern numbers. It is the starting point of our construction of a series of examples.

Let $\phi: B \to \mathbb{P}^1_{\mathbb{C}}$ be a 5-fold cyclic covering with 3 ramification points $Q_1, Q_2, Q_3 \in B$ and $\mathrm{Aut}_{\mathbb{P}^1}(B) = \langle v \rangle \cong \mathbb{Z}_5$. Set

$$\Gamma_i = \{(P, v^i(P)) \in B \times B\}, \quad i = 0, 1, 2, 3, 4,$$
$$\Gamma = \Gamma_0 + \Gamma_1 + \ldots + \Gamma_4.$$

Let Δ be the diagonal of $\mathbb{P}^1 \times \mathbb{P}^1$. Then

$$\Gamma = (\phi, \phi)^* \Delta,$$

Since $\Delta \equiv \mathbb{P}^1 \times \phi(Q_1) + \phi(Q_1) \times \mathbb{P}^1$, we have

$$\Gamma \equiv 5(B \times Q_1 + Q_1 \times B).$$

Let $\sigma: (B \times B)^* \to B \times B$ be a blowing-up with (Q_1, Q_1), (Q_2, Q_2), (Q_3, Q_3) as its centers. Denote

$$E_k = \sigma^{-1}(Q_k, Q_k), \quad k = 1, 2, 3,$$
$$\Gamma_i^* = \sigma^* \Gamma_i - E_1 - E_2 - E_3, \quad i = 0, 1, \ldots, 4,$$
$$\Gamma^* = \Gamma_0^* + \ldots + \Gamma_4^* = \sigma^* \Gamma - 5(E_1 + E_2 + E_3).$$

Since

$$\Gamma^* \equiv 5\sigma^*(B \times Q_1 + Q_1 \times B) - 5(E_1 + E_2 + E_3),$$

we can find an invertible sheaf $\mathcal{M}$ on $(B \times B)^*$ such that

$$\mathcal{M}^{\otimes 5} = \mathcal{O}_{(B \times B)^*}(\Gamma^*).$$

Thus the invertible sheaf $\mathcal{M}$ determines a 5-fold cyclic covering

$$\Phi: H \to (B \times B)^*.$$

We have

$$\mathrm{Aut}_{(B \times B)^*}(H) = \langle \alpha \rangle \cong \mathbb{Z}_5.$$

The smooth surface H has a fibration

$$h: H \to B$$

with $g(B)=2$. Let F_H be a general fibre of h, then we have

$$\begin{aligned} e(H)&=75,\\ K_H^2&=225,\\ \chi(\mathcal{O}_H)&=25,\\ g(F_H)&=16. \end{aligned}$$

4. The proof of theorems

As the first step, we use base change to construct a series of minimal surfaces such that their Chern numbers form a lattice in the coordinate plane.

It is not difficult to carry out this task. Just make an n-fold covering $\tau: C\to B$ such that there are $2d$ ramified general points with ramification index 2 on C. Then $X=H\times_B C$ is a smooth surface and we have a Cartesian square:

$$\begin{array}{ccc} X & \xrightarrow{T} & H \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle h} \\ C & \xrightarrow{\tau} & B \end{array}$$

An easy computation shows that

$$\begin{aligned} e(X)&=ne(H)+4d(g(F_H)-1)=75n+60d,\\ K_X^2&=nK_H^2+8d(g(F_H)-1)=225n+120d,\\ \chi(\mathcal{O}_X)&=n\chi(\mathcal{O}_H)+d(g(F_H)-1)=25n+15d. \end{aligned}$$

In particular,

$$\begin{aligned} 9\chi(\mathcal{O}_X)-K_X^2&=15d,\\ K_X^2-8\chi(\mathcal{O}_X)&=25n. \end{aligned}$$

Note that the linear transformation from $(\chi(\mathcal{O}_X), K_X^2)$ to $(9\chi(\mathcal{O}_X)-K_X^2,\ K_X^2-8\chi(\mathcal{O}_X))$ is unimodular. Hence it is harmless to take $9\chi(\mathcal{O}_X)-K_X^2$ as abscissa and $K_X^2-8\chi(\mathcal{O}_X)$ as ordinate. Now we have obtained a lattice of index 15×25 on the coordinate plane.

The second step is to fill in the integral points inside this lattice by twisting fibres. The automorphism $\alpha\in \mathrm{Aut}_{(B\times B)^*}(H)$ has a pull-back $\gamma\in \mathrm{Aut}_C(X)$. Locally for any general point Q on the ramification divisor $T^{-1}\Phi^{-1}\Gamma^*$ we can choose analytic local coordinates x, y centered at Q such that $\gamma^* x=x$, $\gamma^* y=\zeta y$ where $\zeta^5=1$. By Lemma 1, this fifth root of unity is uniquely determined. From here on we shall fix this ζ. By choosing suitably a 5-fold cyclic covering $\pi: \tilde{C}\to C$

and $\beta \in \mathrm{Aut}_C(\tilde{C})$, we can make a series of fibered surfaces by twisting fibres of old ones. These surfaces are minimal by construction. Note that in this case $m=5$, $g=16$ and $t=5s$. We have

$$K_S^2 = K_X^2 + \sum_{i=1}^{t} \left(\frac{48}{5} - K_S E_{Q_i} \right),$$

$$\chi(\mathcal{O}_S) = \chi(\mathcal{O}_X) + \frac{1}{12} \sum_{i=1}^{t} \left(e(\rho^{-1}(Q_i)) - K_S E_{Q_i} + \frac{71}{5} \right),$$

$$9\chi(\mathcal{O}_S) - K_S^2 = 15d + \frac{1}{4} \sum_{i=1}^{t} \left(3e(\rho^{-1}(Q_i)) + K_S E_{Q_i} + \frac{21}{5} \right),$$

$$K_S^2 - 8\chi(\mathcal{O}_S) = 25n - \frac{1}{3} \sum_{i=1}^{t} \left(2e(\rho^{-1}(Q_i)) + K_S E_{Q_i} - \frac{2}{5} \right).$$

The next thing is to calculate the $e(\rho^{-1}(Q_i))$ and $K_S E_{Q_i}$. For any singularity $Q \in \Sigma \subset S_0$, $P = \tilde{f}(\psi^{-1}(Q)) \in \tilde{C}$ must be a ramification point. We can choose an analytic local coordinate w centered at P such that $\beta^* w = \zeta^b w$. Then there are analytic local coordinates x, y centered at $\psi^{-1}(Q)$ such that $\tilde{f}^* w = x$, $\tilde{\gamma}^* y = \zeta y$. Hence $(\tilde{\beta}\tilde{\gamma})^* x = \zeta^b x$, $(\tilde{\beta}\tilde{\gamma})^* y = \zeta y$. This implies that Q is a Hirzebruch-Jung singularity of type $A_{5,b}$. Only 3 different types of singularities can occur, they are $A_{5,1}$, $A_{5,2} = A_{5,3}$ and $A_{5,4}$. In the following table we list the corresponding $e(\rho^{-1}(Q))$, $K_S E_Q$ and the specialization vector $(\delta_{K^2}, \delta_\chi)$ defined as

$$\delta_{K^2}|_Q := \tfrac{48}{5} - K_S E_Q,$$
$$\delta_X|_Q := \tfrac{1}{12}(e(\rho^{-1}(Q)) - K_S E_Q + \tfrac{71}{5}).$$

Type of Q	$[\lvert e_1\rvert, \ldots, \lvert e_r\rvert]$	$e(\rho^{-1}(Q))$	$K_S E_Q$	$\delta_{K^2}\lvert_Q$	$\delta_\chi\lvert_Q$
$A_{5,1}$	[5]	2	9/5	39/5	6/5
$A_{5,2}$	[3, 2]	3	2/5	46/5	7/5
$A_{5,4}$	[2, 2, 2, 2]	5	0	48/5	8/5

To apply Lemma 1, we should determine integers $a_1, \ldots, a_r$ satisfying the three conditions of the lemma. Note that there are 13 fundamental combinations, we list them in a table. In this table $v_i = (v_{ix}, v_{iy})$ represent the increments to the coordinates $(x, y) = (9\chi - K^2, K^2 - 8\chi)$ by a fundamental combination N° i, i.e.

$$v_{ix} = \tfrac{1}{4} \sum (3e(\rho^{-1}(Q_i)) + K_S E_{Q_i} + \tfrac{21}{5}),$$
$$v_{iy} = \tfrac{1}{3} \sum (2e(\rho^{-1}(Q_i)) + K_S E_{Q_i} - \tfrac{2}{5}).$$

	$(a_1, \ldots, a_r)$	$(A_{5,1}, A_{5,2}, A_{5,4})$	v_{ix}	v_{iy}
1.	(1, 4)	(1, 0, 1)	39	−25
2.	(2, 3)	(0, 2, 0)	34	−20
3.	(1, 1, 3)	(2, 1, 0)	47	−28
4.	(1, 2, 2)	(1, 2, 0)	49	−29
5.	(3, 3, 4)	(0, 2, 1)	58	−36
6.	(2, 4, 4)	(0, 1, 2)	65	−42
7.	(1, 1, 1, 2)	(3, 1, 0)	62	−37
8.	(2, 2, 2, 4)	(0, 3, 1)	75	−46
9.	(1, 3, 3, 3)	(1, 3, 0)	66	−39
10.	(3, 4, 4, 4)	(0, 1, 3)	89	−58
11.	(1, 1, 1, 1, 1)	(5, 0, 0)	75	−45
12.	(2, 2, 2, 2, 2)	(0, 5, 0)	85	−50
13.	(4, 4, 4, 4, 4)	(0, 0, 5)	120	−80

In the preceding table, the reader may notice that every a_i gives rise to an $A_{5,a}$ singularity.

If we take p copies of combination N° 4 and q copies of combination N° 6, that is, let

$$a_i=\begin{cases}1 & 1\leqq i\leqq p\\ 2 & p<i\leqq 3p+q\\ 4 & 3p+q<i\leqq 3p+3q\end{cases}$$

Then

$$9\chi(\mathcal{O}_S)-K_S^2=15d+49p+65q,$$
$$K_S^2-8\chi(\mathcal{O}_S)=25n-29p-42q.$$

For any integer $x\geqq 326$, there exist integers $d_0\geqq 0$, $0\leqq p_0\leqq 4$, $0\leqq q_0\leqq 2$ such that

$$5(3d_0+13q_0)+49p_0=x.$$

Take

$$\begin{aligned}p&=p_0+5u,\\ q&=q_0+2u, \quad u\geqq 0,\\ d&=d_0-25u.\end{aligned}$$

We have

$$5d+49p+65q=x, \tag{1}$$
$$y=25n-29p-42q=-29p_0-42q_0+25n-229u. \tag{2}$$

Hence for any integer $y\geqq 0$ there exist integers $n\geqq 2$, $0\leqq u\leqq 24$ satisfying equation (2). This implies that for any integers x, y satisfying

$$x\geqq 15\times 25\times 24+49\times 4+65\times 2=9326,$$
$$y\geqq 0,$$

we can always find integers $n\geqq 2$, d, p, $q\geqq 0$ such that the equations (1) and (2) hold. In other words, for any positive integers a, b satisfying

$8a \leqq b \leqq 9a-9326$, there exists a minimal projective surface S of general type with $K_S^2=b$, $\chi(\mathcal{O}_S)=a$. Combined with a result of Chen [2] which claims that for every integers a and b satisfying $2a-6 \leqq b \leqq 8a$ and $a \geqq 246$, such surface exists as well, the theorem 1 can be thus proved.

The constant C may be approved by computer check. This is the scheme of computation: starting from a point (x, y), $0 \leqq x \leqq 14$, $0 \leqq y \leqq 24$, the computer will search for a minimal distance $d(x, y)$ from this point to the nearest lattice point (in the direction of x-axis), i.e.

$$d(x, y)=\min\left\{ \sum_{i=1}^{13} c_i v_{ix} \;\middle|\; \begin{array}{l} \sum_{i=1}^{13} c_i v_{ix} \equiv x \pmod{15}, \\ \sum_{i=1}^{13} c_i v_{iy} \equiv y \pmod{25}, \\ c_i \geqq 0, \quad i=1, \ldots, 13, \\ c_i \in \mathbb{Z} \end{array} \right\}.$$

Then for each $x=0, \ldots, 14$, find the maximum and minimum of the $d(x, y)$:

$$\begin{aligned} d_{\max}(x)&=\max_{0 \leqq y \leqq 24} d(x, y), \\ d_{\min}(x)&=\min_{0 \leqq y \leqq 24} d(x, y), \end{aligned} \qquad x=0, \ldots, 14.$$

Hence for any integers x, $y \geqq 0$ satisfying

$$\begin{aligned} &x \equiv r \pmod{15}, \quad 0 \leqq r \leqq 14, \\ &x \geqq d_{\max}(r), \end{aligned}$$

we can always construct a fibered surface S with

$$\begin{aligned} \chi(\mathcal{O}_S)&=x+y, \\ 9\chi(\mathcal{O}_S)-K_S^2&=x. \end{aligned}$$

Furthermore, for any integers $x \geqq 0$ satisfying

$$\begin{aligned} &x \equiv r \pmod{15}, \quad 0 \leqq r \leqq 14, \\ &x \geqq d_{\min}(r), \end{aligned}$$

there exist infinitely many integers $y \geqq 0$ and fibered surfaces S such that

$$\begin{aligned} \chi(\mathcal{O}_S)&=x+y, \\ 9\chi(\mathcal{O}_S)-K_S^2&=x. \end{aligned}$$

We list the final result in the following table:

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$d_{\max}(x)$	300	361	302	303	334	320	276	322	293	294	310	311	327	283	314
$d_{\min}(x)$	0	136	47	78	34	65	66	97	68	39	85	86	102	58	89

It should be mentioned that surfaces at the lattice points $(15d, 25)$ could not be constructed in the first step, but on the one hand by taking suitable copies of combinations N° 11 and N° 13 the lattice points $(15d, 25)$ $(d \geqq 13)$ can be filled in from another lattice points, on the other hand, it is not too difficult to verify that this fact does not affect our final result. Hence the conclusion may be derived immediately. The Theorem 2 is shown.

As a reference, we list the values of $d(x, y)$ with respect to $x = 9\chi - K^2$ and $y = K^2 - 8\chi$.

y \ x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0	211	287	78	124	170	156	202	248	39	85	296	117	163	209
1	285	241	122	168	124	200	246	202	83	294	250	161	207	163	209
2	120	166	122	198	244	200	81	157	248	159	205	161	207	283	239
3	240	286	242	288	244	155	201	322	203	249	205	116	327	283	164
4	75	286	242	153	199	245	201	247	203	114	160	281	162	238	284
5	75	151	197	153	34	245	231	112	158	114	160	236	192	73	119
6	195	241	197	243	199	275	156	202	158	234	280	236	117	238	284
7	195	241	197	273	154	200	156	232	278	234	115	191	282	193	239
8	150	361	152	198	274	65	276	322	113	189	235	191	237	283	104
9	150	196	272	228	109	320	111	187	233	189	235	281	237	148	194
10	150	226	107	153	109	185	231	187	68	189	235	146	192	148	194
11	105	151	272	183	229	275	66	277	233	144	190	236	192	268	314
12	225	316	227	273	229	140	231	307	188	234	190	266	312	268	149
13	225	271	62	138	184	140	186	232	188	99	310	101	147	223	179
14	225	136	182	228	184	230	186	97	143	264	145	221	267	58	269
15	180	226	182	228	184	260	141	187	143	219	265	221	102	223	269
16	180	226	302	258	139	185	141	217	263	219	100	311	267	178	224
17	300	346	137	183	259	215	261	307	98	174	265	176	222	268	89
18	135	181	137	213	259	215	96	172	218	174	220	176	222	133	254
19	255	211	92	303	94	170	216	172	218	264	220	131	177	133	179
20	120	136	257	198	214	260	216	262	218	159	175	221	177	253	299
21	255	166	212	258	49	260	216	127	173	219	175	251	297	88	134
22	210	256	47	123	334	125	171	217	173	249	295	86	132	208	164
23	210	256	167	288	169	215	171	247	293	249	130	206	252	208	254
24	165	211	167	213	169	245	126	247	128	204	250	206	252	208	254

References

1. Barth, W., Peters, C., Van de Ven, A.: Compact Complex Surfaces. Berlin Heidelberg New York: Springer 1984
2. Chen, Z.: On the geography of surfaces – simply connected minimal surfaces with positive index. Math. Ann. **277**, 141–164 (1987)
3. Hirzebruch, F.: Über vierdimensionale Riemannsche Flächen mehrdeutiger analytischer Funktionen von zwei Veränderlichen. Math. Ann. **126**, 1–22 (1953)
4. Inoue, M.: Some surfaces of general type with positive indices. Preprint (1981)
5. Persson, U.: Chern invariants of surfaces of general type. Compos. Math. **43**, 3–58 (1981)
6. Xiao, G.: Surfaces fibrées en courbes de genre deux. (Lect. Notes Math., vol. 1137). Berlin Heidelberg New York: Springer 1985
7. Xiao, G.: Problem list. In: Birational Geometry of Algebraic Varieties – Open Problems. Katata 1988

Math. Z. 206, 255–264 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On manifolds of positive Ricci curvature with large diameter

Yukio Otsu

Kyushu University, Department of Mathematics, Faculty of Science, Fukuoka 812, Japan

Received March 22, 1990; in final form May 9, 1990

0 Introduction

It is well known that if M is an n-dimensional complete Riemannian manifold with Ricci curvature $\mathrm{Ric}_M \geqq n-1$, then the diameter $d(M) \leqq \pi$ and that if $\mathrm{Ric}_M \geqq n-1$ and $d(M)=\pi$, M is isometric to the unit sphere $S^n(1)$ (Mayers [M], Cheng [C]). Since the sectional curvature $K_M \geqq 1$ implies $\mathrm{Ric}_M \geqq n-1$, we have the same results under this condition. Furthermore if $K_M \geqq 1$ and $d(M) > \pi/2$, then M is homeomorphic to S^n (Grove and Shiohama [G-S]). Thus we are naturally led to a question: If $\mathrm{Ric}_M \geqq n-1$ and $d(M)$ is close to π, is M homeomorphic to S^n? Recently Grove and Petersen [G-P] proved that under the assumption that $K_m \geqq -k^2$ and the volume $\mathrm{vol}(M) \geqq v_0$ for some k, $v_0 > 0$ the answer is affirmative (see also Wu [W], Eschenburg [E]). In the first part of this paper we prove that if there is no restriction on K_M the answer is negative, that is,

Theorem 1. *For any integer $n \geqq 5$ and $l = 2, \ldots, n-2$, there exists a sequence of C^∞ metrics $\{g_i\}_{i=1,2,\ldots}$ on $S^{n-l} \times S^l$ such that*

$$\mathrm{Ric}_{M_i} \geqq n-1, \quad \mathrm{vol}(M_i) \geqq c_{n,l}\,\mathrm{vol}(S^n(1)),$$
$$d(M_i) \to \pi \quad \text{as} \quad i \to \infty,$$

where $M_i := (S^{n-l} \times S^l, g_i)$ and $c_{n,l}$ is a positive constant depending on n and l.

Remark. (1) If $n=2$, since $\mathrm{Ric}_M \geqq 1$ implies $K_M \geqq 1$, $d(M) > \pi/2$ implies M is diffeomorphic to S^2. If $n=3$, it follows from Hamilton [H] that metrics on M with $\mathrm{Ric}_M \geqq 2$ are deformed to $K_M = 1$. However, since it is difficult to chase the change of diameter, we cannot say that M is homeomorphic to S^3. If $n=4$, it seems difficult to answer the question[1].

[1] After the preparation of this paper the author learned that Anderson [An2] independently constructed similar metrics on $\mathbb{C}P^n$ for any n by similar technique. Hence this case is solved.

(2) Numerically $c_{5,2} \geqq 1/4$. In general we will show that

$$c_{n,l} \geqq \frac{\omega_{n-l-1}\,\omega_l}{\omega_{n-1}} \left(\frac{1}{2}\,\frac{n-l-1}{n-1}\right)^{\frac{n-l-1}{2}} \left(\frac{1}{2}\,\frac{l-1}{n-1}\right)^{\frac{l}{2}}$$

for $l \leqq [n/2]$, where $\omega_l := \mathrm{vol}(S^l(1))$ and $[\cdot]$ is Gauss's symbol.

(3) In place of diameter, we can consider a similar problem about volume: If $\mathrm{Ric}_M \geqq n-1$ and $\mathrm{vol}(M)$ is close to $\mathrm{vol}(S^n(1))$, is M homeomorphic to S^n? In this case, it is known that M is diffeomorphic to S^n under the assumption $K_M \geqq -k^2$ for some $k>0$ (Yamaguchi [Y], see also [I], [K], [S]). If there is a counterexample to the problem, it may also provide another example to our problem since $\mathrm{vol}(M)$ being close to $\mathrm{vol}(S^n(1))$ implies $d(M)$ being close to π. However the author does not know whether there exists such an example.

In the second part of this paper we consider examples of complete open Riemannian manifolds. In the previous paper [O] we defined an invariant $V_0(M)$ for an open complete Riemannian manifold M with $\mathrm{Ric}_M \geqq 0$ by

$$V_0(M) := \lim_{r \to \infty} \frac{\mathrm{vol}(B(p,r))}{b_0(r)},$$

where $p \in M$, $B(p, r)$ is the r-ball at p and $b_0(r)$ is the volume of an r-ball in $\mathbb{R}^n$. We say that M is an asymptotically nonnegatively curved manifold with C_0 if there are a point $p \in M$ and a nonnegative nonincreasing function λ: $[0\ \infty) \to \mathbb{R}$ such that

$$C_0 := \int_0^\infty r\,\lambda(r)\,dr < \infty,$$

$$K_M|_q \geqq -\lambda(d(p,q))$$

for any $q \in M$ (Abresch [Ab]). Then we proved the following:

Theorem 2 (Otsu [O]). *There exists $\varepsilon = \varepsilon(n, C_0) > 0$ such that if M is an asymptotically nonnegatively curved manifold with C_0 such that*

$$\mathrm{Ric}_M \geqq 0 \quad \textit{and} \quad V_0(M) \geqq 1-\varepsilon,$$

then there is no critical point of distance function from p. In particular M is diffeomorphic to $\mathbb{R}^n$.

In the paper it remained to show that the condition on $V_0(M)$ is essentially necessary to restrict the topology of asymptotically nonnegatively curved manifolds with $\mathrm{Ric}_M \geqq 0$. As an analogue of Theorem 1 we obtain examples which show the condition on $V_0(M)$ is inevitable:

Theorem 3. *For any integer $n \geqq 5$ and $l = 2, \ldots, [n/2]$, there exists a C^∞ metric g such that $M := (\mathbb{R}^{n-l} \times S^l, g)$ is an asymptotically nonnegatively curved manifold with C_0 such that*

$$\tfrac{1}{8} < C_0 < 8, \qquad \mathrm{Ric}_M \geqq 0, \qquad V_0(M) \geqq c_{n,l},$$

where $c_{n,l}$ is the constant in Theorem 1.

1 Proof of Theorem 1

It is enough to consider $S^{m+1}\times S^l$ for integers m, l with $m+l=n-1$ and $2\leqq l\leqq[n/2]$ for the proof of Theorem 1. For any $a>0$ let $S^n(\bar{a})$ denote the n-sphere with radius $\bar{a}=a/\pi$ in $\mathbb{R}^{n+1}$. Let $(r, x)\in[0, a]\times S^m\to S^{m+1}$ be the polar coordinates of $S^{m+1}(\bar{a})$. First we examine $S^{m+1}(\sqrt{m/(n-1)})\times S^l(\sqrt{(l-1)/(n-1)})$ for $m\geqq 1$ and $l\geqq 2$. The metric is written as

$$ds^2=dr^2+\frac{m}{n-1}\sin^2\sqrt{\frac{n-1}{m}}r\,ds_m^2+\frac{l-1}{n-1}ds_l^2,$$

where ds_m^2 is the metric of $S^m(1)$. Then $\mathrm{Ric}\geqq n-1$ and the diameter is $\pi\sqrt{(n-2)/(n-1)}$.

In order to find a sequence of metrics on $S^{m+1}\times S^l$ whose diameters converge to π, we investigate warped product metrics as follows:

$$g=dr^2+f(r)^2\,ds_m^2+h(r)^2\,ds_l^2, \tag{1.1}$$

where $f, h:[0, a]\to\mathbb{R}$ are C^∞ functions which need to satisfy

$$f(r)>0 \quad \text{on} \quad (0, a), \tag{1.2.1}$$
$$f(0)=f(a)=0, \quad f'(0)=1, \quad f'(a)=-1,$$
$$h(r)>0 \quad \text{on} \quad [0, a], \tag{1.2.2}$$
$$h'(0)=h'(a)=0.$$

Let v, w be unit tangent vectors of $S^m(1)$, $S^l(1)$ respectively and let $\tilde{v}$, $\tilde{w}$ be the tangent vectors of $S^{m+1}\times S^l$ such that $\tilde{v}:=(0, v, 0)$, $\tilde{w}:=(0, 0, w)$ in the coordinates $(0, a)\times S^m\times S^l\to S^{m+1}\times S^l$. Then $U:=\dfrac{\partial}{\partial r}$, $V:=\dfrac{1}{f}\tilde{v}$ and $W:=\dfrac{1}{h}\tilde{w}$ are respectively orthonormal vectors of $M:=(S^{m+1}\times S^l, g)$. A straightforward calculation gives us

$$\mathrm{Ric}(U, V)=\mathrm{Ric}(V, W)=\mathrm{Ric}(W, U)=0, \tag{1.3}$$
$$\mathrm{Ric}(U, U)=-m\frac{f''(r)}{f(r)}-l\frac{h''(r)}{h(r)},$$
$$\mathrm{Ric}(V, V)=-\frac{f''(r)}{f(r)}+(m-1)\frac{1-f'(r)^2}{f(r)^2}-l\frac{f'(r)}{f(r)}\frac{h'(r)}{h(r)},$$
$$\mathrm{Ric}(W, W)=-\frac{h''(r)}{h(r)}+(l-1)\frac{1-h'(r)^2}{h(r)^2}-m\frac{f'(r)}{f(r)}\frac{h'(r)}{h(r)}.$$

Note that if we can choose $f=b\sin r$, $h=b'\sin r$ for constants

$$0<b\leqq\sqrt{\frac{m-1}{n-2}}, \tag{1.4.1}$$

$$0<b'\leqq\sqrt{\frac{l-1}{n-2}}, \tag{1.4.2}$$

then $\mathrm{Ric}_M \geqq n-1$ and $d(M)=\pi$. Unfortunately this metric is singular at $r=0$, π since f and h do not satisfy the condition (1.2) at $r=0, \pi$. Thus we modify these functions near $r=0$ and a so that they satisfy (1.2), $\mathrm{Ric}_M \geqq n-1$ and $a \to \pi$ as we change constants in their constructions.

In the construction we need the following lemma:

Lemma 1.5 (Compare [Lemma 3, G-Y]). *For $\alpha<\gamma$ let $\varphi, \psi: [\alpha, \gamma] \to \mathbb{R}$ be C^∞ functions such that $\varphi(\beta)=\psi(\beta)>0$ and $\varphi'(\beta)>\psi'(\beta)$ for $\beta \in (\alpha, \gamma)$. Then for small $\varepsilon>0$ there exist $\varepsilon'>0$ and a C^∞ function $\phi: [\alpha, \gamma] \to \mathbb{R}$ such that*

$$
\begin{aligned}
(1-\varepsilon)\max(\varphi(r), \psi(r)) \leqq \phi(r) &\leqq \min(\varphi(r), \psi(r)), \\
\psi'(r) \leqq \phi'(r) &\leqq \varphi'(r), \\
\phi''(r) \leqq \max(\varphi''(r), \psi''(r)) \quad &\text{on } [\beta-\varepsilon', \beta+\varepsilon'], \\
\phi(r)=\varphi(r) \quad &\text{on } [\alpha, \beta-\varepsilon'], \\
\phi(r)=\psi(r) \quad &\text{on } [\beta+\varepsilon', \gamma].
\end{aligned}
$$

Proof. For $c:=(\varphi'(\beta)+\psi'(\beta))/2$ let $\rho(r):=c(r-\beta)+\varphi(\beta)$, $\tilde{\varphi}(r):=\varphi(r)-\rho(r)$ and $\tilde{\psi}(r):=\psi(r)-\rho(r)$. We may assume that $\tilde{\varphi}'(r)>0$ and $\tilde{\psi}'(r)<0$ on $[\beta-\varepsilon', \beta+\varepsilon']$. Then we can choose C^∞ functions $\xi, \zeta: [\alpha, \gamma] \to \mathbb{R}$ such that

$$
\begin{aligned}
&\xi(r)=\tilde{\varphi}'(r) && \zeta(r)=0 && \text{on } [\alpha, \beta-\varepsilon'], \\
&\xi(r)=0 && \zeta(r)=\tilde{\psi}'(r) && \text{on } [\beta+\varepsilon', \gamma], \\
&0 \leqq \xi(r) \leqq \tilde{\varphi}'(r) && \tilde{\psi}'(r) \leqq \zeta(r) \leqq 0 && \text{on } [\beta-\varepsilon', \beta+\varepsilon'], \\
&\xi'(r) \leqq \tilde{\varphi}''(r) && \zeta'(r) \leqq 0 && \text{on } [\beta-\varepsilon', \beta], \\
&\xi'(r) \leqq 0 && \zeta'(r) \leqq \tilde{\psi}''(r) && \text{on } [\beta, \beta+\varepsilon'],
\end{aligned}
$$

$$
\int_{\beta-\varepsilon'}^{\beta+\varepsilon'} \xi(r)\,dr = -\tilde{\varphi}(\beta-\varepsilon'), \qquad \int_{\beta-\varepsilon'}^{\beta+\varepsilon'} \zeta(r)\,dr = \tilde{\psi}(\beta+\varepsilon').
$$

Set $\phi(r):=\int_\alpha^r \{\xi(r)+\zeta(r)\}\,dr+\rho(r)+\tilde{\varphi}(\alpha)$. Computations show that ϕ is the desired function if we take sufficiently small $\varepsilon'>0$.

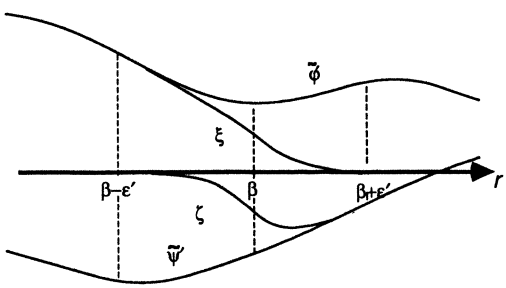


Fig. 1.

We demand that f and h are symmetric functions with respect to $r=a/2$ so that it is enough to manipulate them on $[0, a/2]$. Let $k>0$ and

$$k' := k\sqrt{\frac{m}{l}\left(1-\frac{1}{k}\sqrt{\frac{n-1}{m}}\right)}.$$

For $b, b'>0$ and $\delta_0>0$ define functions

$$f_0(r) := \frac{1}{k}\sin kr, \qquad f_1(r) := b\sin(r+\delta_0),$$

$$h_0(r) := \frac{1}{k'}\cosh k'r, \qquad h_1(r) := b'\sin(r+\delta_0).$$

For any $\varepsilon''>0$ we can take $0<t_1<t_2<t_3<\pi/2, 0<t_0, b, b'>0$ such that

(1.6)
$$\tan t_2 \geqq \sqrt{\frac{l}{m}}\coth\sqrt{\frac{m}{l}}\,t_2,$$

$$\tan t_3 - t_3 < t_0 < (1+\varepsilon'')\tan t_2 - t_2,$$

$$\frac{1}{1+\varepsilon''}\cos t_1 \leqq \frac{\sin t_2}{t_2+t_0}, \qquad \sqrt{\frac{l}{m}}\coth\sqrt{\frac{m}{l}}\,t_1 - t_1 \leqq \tan t_3 - t_3,$$

$$\frac{\sin t_2}{t_2+t_0} < b < \frac{\sin t_3}{t_3+t_0},$$

$$\sqrt{\frac{l}{m}}\frac{\cosh\sqrt{\frac{m}{l}}\,t_1}{t_0+t_1} < b' < \min\left(\sqrt{\frac{l}{m}}\frac{\cosh\sqrt{\frac{m}{l}}\,t_2}{t_0+t_2}, \sinh\sqrt{\frac{m}{l}}\,t_1\right).$$

Let $\delta_i = t_i/k$ for $i=0, 1, 2, 3$. Then it follows from the above that

(1.7)
$$h_0(\delta_1) < h_1(\delta_1), \qquad h_1(\delta_2) < h_0(\delta_2),$$
$$f_0(\delta_2) < f_1(\delta_2), \qquad f_1(\delta_3) < h_0(\delta_3),$$
$$h_1'(r) < h_0'(r) \quad \text{on } [\delta_1, \delta_2],$$
$$f_1'(r) < f_0'(r) < (1+\varepsilon'')f_1'(r) \quad \text{on } [\delta_1, \delta_3]$$

for large $k>0$. The following claim shows that b also satisfies (1.4.1).

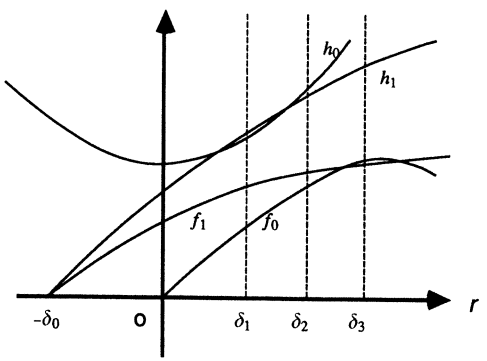


Fig. 2.

Claim 1.8. *For any* $n \geqq 5$, $[n/2] \geqq l \geqq 2$ *and small* $\varepsilon'' > 0$ *we can take* t_0, t_1, t_2, t_3, b *and* b' *such that*

$$0.8\sqrt{\frac{l}{m}} < t_1 < t_2 < t_3 < 1.2\sqrt{\frac{l}{m}}$$

and

$$\sqrt{\frac{m}{2(n-1)}} \leqq b \leqq \sqrt{\frac{m-1}{n-1.96}}$$

for large $k > 0$.

Proof. If $(m, l) \neq (2, 2)$, we assume

$$\tan t_2 = \sqrt{\frac{l}{m}} \coth \sqrt{\frac{m}{l}}\, t_2. \tag{1.9}$$

The first inequalities follow from

$$\tan 1.2\sqrt{\frac{l}{m}} > 1.2\sqrt{\frac{l}{m}} > \sqrt{\frac{l}{m}} \coth 1.2,$$

$$\tan 0.8\sqrt{\frac{l}{m}} < \sqrt{\frac{l}{m}} \frac{\tan 0.8 \times 1.3}{1.3} < \sqrt{\frac{l}{m}} \coth 0.8$$

for any $m \geqq 2$ because $\sqrt{l/m} \leqq \sqrt{3/2}$. Since coth is a decreasing function, we also have

$$\tan t_3 \geqq \sqrt{\frac{l}{m}} \coth 1.2 > 1.01\sqrt{\frac{l}{m-1}}$$

for any $m \geqq 4$, which gives

$$b \leqq \cos t_3 < \sqrt{\frac{m-1}{n-1.96}}.$$

If $m = 3$, since $l/m \geqq 2/3$,

$$\tan 1.1\sqrt{\frac{l}{m}} > 1.1\sqrt{\frac{l}{m}}\left(1 + \frac{1}{3}(1.1)^2 \frac{2}{3}\right) > \sqrt{\frac{l}{m}} \coth 1.1,$$

hence $t_3 < 1.1\sqrt{\frac{l}{m}}$ and

$$\tan t_3 \geqq \sqrt{\frac{l}{3}} \coth 1.1 > 1.01\sqrt{\frac{l}{2}}.$$

The left side inequality is similarly proved. If $m = 2$, then l is 2 or 3. If $(m, l) = (2, 3)$, we can take $t_2 = 1.053\ldots$ as the solution of (1.9). Then b satisfies the inequalities.

If $(m, l)=(2, 2)$, we take $t_2=1$. Then the inequalities are satisfied for sufficiently small $\varepsilon''>0$.

Remark. When $(m, l)=(2, 2)$, we have $t_2=0.937\ldots$ as the solution of (1.9). Then b dose not satisfy the later inequalities for sufficiently small $\varepsilon''>0$.

Let us take $\varepsilon=1/(1+k\sqrt{m/(n-1)})$ in Lemma (1.5), $\varepsilon''=1/k$ in (1.6) and t_0, t_1, t_2, t_3, b and b' in Claim (1.8). From Lemma (1.5) we get C^∞ functions $f, \tilde{h}$: $[0, a/2]\to\mathbb{R}$ and $\varepsilon'>0$ such that

(1.10)
$$f(r)=f_0(r) \qquad \tilde{h}(r)=h_0(r) \qquad \text{on } [0, \beta-\varepsilon'],$$

$$f(r)=f_0(r) \qquad \begin{cases} (1-\varepsilon)h_0(r)\leqq\tilde{h}(r)\leqq h_0(r) \\ \tilde{h}'(r)\geqq h_0'(r) \\ -\dfrac{\tilde{h}''(r)}{\tilde{h}(r)}\geqq -\dfrac{1}{1-\varepsilon}k'^2 \end{cases} \qquad \text{on } [\beta-\varepsilon', \beta+\varepsilon'],$$

$$\begin{cases} (1-\varepsilon'')f_1(r)\leqq f(r)\leqq f_1(r) \\ f_1'(r)\leqq f'(r)\leqq(1+\varepsilon'')f_1'(r) \\ \qquad -\dfrac{f_1''(r)}{f(r)}\geqq 1 \end{cases} \qquad \tilde{h}(r)=h_1(r) \quad \text{on } [\beta+\varepsilon', \delta_3],$$

$$f(r)=f_1(r) \qquad \tilde{h}(r)=h_1(r) \quad \text{on } [\delta_3, a/2],$$

where $r=\beta\in(\delta_3, \delta_3)$ is the solution of $h_0(r)=h_1(r)$. Although b satisfies (1.4.1) as shown in (1.8), b' dose not satisfy (1.4.2). However, in this case the situation is different from that of b: A multiplication of positive constant to $f(r)$ dose not preserve the condition (1.2.1), but the operation to $h(r)$ keeps (1.2.2). Thus put $h(r):=c\tilde{h}(r)$ for

(1.11)
$$c:=\frac{1}{\sinh\sqrt{\frac{m}{l}}t_1}\sqrt{0.8\frac{l-1}{n-1}}.$$

Then $h(r)=cb'\sin(r+\delta_0)$ for $r\in[\beta+\varepsilon', a/2]$ and

(1.12)
$$cb'\leqq\sqrt{0.8\frac{l-1}{n-1}}.$$

By the functions f and h we clearly obtain a C^∞ metric g_k on $S^{m+1}\times S^l$ for large $k>0$. Let $M_k:=(S^{m+1}\times S^l, g_k)$. The computation to show $\mathrm{Ric}_{M_k}\geqq n-1$ for large $k>0$ is completed as follows: on $[\beta+\varepsilon', a/2]$ we estimate $f(r)$ and $h(r)$ in (1.3) by $f_1(r)=b\sin(r+\delta_0)$ and $ch_1(r)=cb'\sin(r+\delta_0)$ respectively by using (1.10). Then $\mathrm{Ric}_{M_k}\geqq n-1$ follows from (1.8) and (1.12). On $[0, \beta+\varepsilon']$ similarly we replace them by $f_0(r)=\frac{1}{k}\sin kr$ and $ch_0(r)=\frac{c}{k'}\cosh k'r$ respectively. Then $\mathrm{Ric}_{M_k}\geqq n-1$ follows from (1.11) and easy computations (e.g., $(\tanh k'r)/k'\leqq(\tan kr)/k$).

Next we estimate the volume $\mathrm{vol}(M_k)$ for large $k>0$. It follows from (1.8) and (1.11) that

$$\begin{aligned}\mathrm{vol}(M_k) &= \omega_m\,\omega_l \int_{\delta_3}^{\pi-2\delta_0-\delta_3} f^m(r)\,h^l(r)\,dr \\ &> \omega_m\,\omega_l\left(\frac{1}{2}\frac{m}{n-1}\right)^{\frac{m}{2}}\left(\frac{1}{2}\frac{l-1}{n-1}\right)^{\frac{l}{2}}\int_0^{\pi}\sin^{n-1}r\,dr \\ &= \frac{\omega_m\,\omega_l}{\omega_{n-1}}\left(\frac{1}{2}\frac{m}{n-1}\right)^{\frac{m}{2}}\left(\frac{1}{2}\frac{l-1}{n-1}\right)^{\frac{l}{2}}\mathrm{vol}(S^n(1)).\end{aligned}$$

(Numerical estimates give more accurate estimate of the volume.) Because $d(M_k)\geqq\pi-2\delta_0$, the proof is completed.

Remark. (1) In these examples the subsets $\{r=0\}$ are totally geodesic manifolds, which are isometric to $S^l(c/k)$ for c in (1.11). Hence they contain small closed geodesics, which means that there is no uniform estimate on their injectivity radii.

(2) If $\mathrm{Ric}_M\geqq n-1$ and $\mathrm{vol}(M)>\frac{1}{2}\,\mathrm{vol}(S^n(1))$, then M is simply connected. If $K_M\geqq 1$, then $d(M)>\pi/2$ implies that M is simply connected ([G-S]). However, we cannot replace volume by diameter in case of Ricci curvature for $n\geqq 5$ because we can replace S^l in $(S^{m+1}\times S^l, g_k)$ by $\mathbb{R}P^l$.

(3) Theorem 1 states that for any $n\geqq 5$ there are at least $[n/2]$ different homotopy classes in the class of n-dimensional manifolds M which have a sequence of C^∞ metrics $\{g_i\}$ such that

$$\mathrm{Ric}_{M_i}\geqq n-1, \qquad \mathrm{vol}(M_i)\geqq c_n\,\mathrm{vol}(S^n(1))$$

and

$$d(M_i)\to\pi \quad \text{as} \quad i\to\infty$$

for $M_i:=(M, g_i)$ and $c_n:=\min_{2\leqq l\leqq[n/2]} c_{n,l}>0$. If there is no restriction on the volume, we have:

Corollary. *For any $n\geqq 6$ the class of n-dimensional manifolds M which have a sequence of C^∞ metrics $\{g_i\}$ such that*

$$\mathrm{Ric}_{M_i}\geqq n-1, \qquad d(M_i)\to\pi \quad \text{as} \quad i\to\infty$$

for $M_i:=(M, g_i)$ contains infinitely many homotopy classes.

Proof. In the construction of the example let $l=3$. Then replace S^3 by the lens spaces $L(p,q)$.

If there is a lower bound of the volume, the author does not know whether there exists a bound of the number of topological types.

2 Proof of Theorem 3

For g_k in §1 let $M'_k:=(S^{m+1}\times S^l, k^2 g_k)$. Then the examples of Theorem 3 are obtained as the pointed Hausdorff limit of (M'_k, p) for some $p\in\{r=0\}$ as $k\to\infty$. We will give more constructive description of them. We consider that the metric

(1.1) is defined on $\mathbb{R}^{m+1}\times S^l$, that is, with the following conditions instead of the condition (1.2).

(1.2)′ $$f(r)>0, \quad h(r)>0 \quad \text{on } (0,\infty),$$
$$f(0)=0, \quad f'(0)=1, \quad h(0)>0, \quad h'(0)=0.$$

For $b, b'>0$ and $t_0>0$ in (1.6) let

$$\bar{f}_0(r):=\sin r, \quad \bar{f}_1(r):=b(r+t_0),$$
$$\bar{h}_0(r):=\sqrt{\frac{l}{m}}\cosh\sqrt{\frac{m}{l}}\,r, \quad \bar{h}_1(r):=b'(r+t_0).$$

Then for $0<t_1<t_2<t_3$, $\varepsilon''>0$ in (1.6) we have

(1.6)′ $$\bar{h}_0(t_1)<\bar{h}_1(t_1), \quad \bar{h}_1(t_2)<\bar{h}_0(t_2),$$
$$\bar{f}_0(t_2)<\bar{f}_1(t_2), \quad \bar{f}_1(t_3)<\bar{f}_0(t_3),$$
$$\bar{h}_1'(r)<\bar{h}_0'(r) \quad \text{on } [t_1,t_2],$$
$$\bar{f}_1'(r)<\bar{f}_0'(r)<(1+\varepsilon'')\bar{f}_1'(r) \quad \text{on } [t_1,t_3].$$

Thus we can construct C^∞ functions $\bar{f}$, $\bar{h}\colon[0,\infty)\to\mathbb{R}$ and the metric $\bar{g}$ as in the proof of Theorem 1. A similar calculation shows that

$$\mathrm{Ric}_{\bar{M}}\geqq 0, \quad V_0(\bar{M})\geqq c_{n,l}$$

for $\bar{M}:=(\mathbb{R}^{n-l}\times S^l, \bar{g})$. We can easily prove that $K_{\bar{M}}\geqq 0$ for $r\geqq t_3$ and $K_{\bar{M}}\geqq -m/l$ for $r<t_3$. Let $p\in\{r=0\}$ and $\lambda\colon[0,\infty)\to\mathbb{R}$ be any nonnegative nonincreasing function such that

$$K_{\bar{M}}|_q\geqq -\lambda(d(p,q))$$

for any $q\in\bar{M}$. Since the diameter of the totally geodesic submanifold $\{r=0\}$ is less than $2.5\sqrt{l/m}$, we may assume that

$$\lambda(r)\leqq\frac{m}{l} \quad \text{for } r<t_3+2.5\sqrt{\frac{l}{m}},$$
$$\lambda(r)=0 \quad \text{for } r\geqq t_3+2.5\sqrt{\frac{l}{m}}.$$

Thus

$$C_0:=\int_0^\infty r\,\lambda(r)\,dr\leqq\int_0^{4\sqrt{l/m}} r\frac{m}{l}\,dr=8.$$

Since λ needs to satisfy

$$\lambda(r)\geqq\frac{m}{l} \quad \text{for } r<t_1,$$
$$\lambda(r)\geqq 0 \quad \text{for } r\geqq t_1,$$

it follows from (1.8) that

$$C_0 \geqq \int_0^{\sqrt{l/m}/2} r \frac{m}{l} dr = \frac{1}{8}.$$

Acknowledgement. Recently there were many interesting examples of manifolds with restricted Ricci curvature ([An1], [G-Y], [S-Y], etc.). The author was stimulated very much from these works. He thanks Professor K. Shiohama for his encouragement and help. He also indebted to K. Cho, J. Itho and T. Yamaguchi.

References

[Ab] Abresch, U.: Lower curvature bounds, Toponogov's theorem and bounded topology I. Ann. Sci. Éc. Norm. Supér., **18**, 651–670 (1985)

[An1] Anderson, M.: Short geodesic and gravitational instantons. J. Differ. Geom. **31**, 265–275 (1990)

[An2] Anderson, M.: Metrics of positive Ricci curvature with large diameter. Preprint

[C] Cheng, S.Y.: Eigenvalue comparison theorem and its applications. Math. Z. **143**, 289–297 (1975)

[E] Eschenburg, J.: Diameter, volume and topology for positive Ricci curvature. Preprint

[G-Y] Gao, L.Z., Yau, S.T.: The existence of negatively Ricci curved metrics on three manifolds. Invent. Math. **85**, 637–652 (1986)

[G-P] Grove, K., Petersen, P.: On the excess of metric space and manifolds. Preprint (1989)

[G-S] Grove, K., Shiohama, K.: A generalized sphere theorem. Ann. Math., II. Ser. **106**, 201–211 (1977)

[H] Hamillton, R.S.: Three manifolds with positive Ricci curvature. J. Differ. Geom. **17**, 255–306 (1982)

[I] Itokawa, Y.: The topology of certain Riemannian manifolds with positive Ricci curvature. J. Differ. Geom. **18**, 151–155 (1983)

[K] Katsuda, A.: Gromov's convergence theorem and its application. Nagoya Math. J. **100**, 11–48 (1985)

[M] Myers, S.B.: Riemannian manifolds with positive mean curvature. Duke Math. J. **8**, 401–404 (1941)

[O] Otsu, Y.: Topology of complete open manifolds with non-negative Ricci curvature. In: Shiohama, K. (ed.) Geometry of manifolds. (Perspect. Math., vol. 8, pp. 295–302) San Diego: Academic Press 1989

[S-Y] Sha, J.P., Yang, D.G.: Examples of manifolds of positive Ricci curvature. J. Differ. Geom. **29**, 95–103 (1989)

[S] Shiohama, K.: A sphere theorem for manifolds of positive Ricci curvature. Trans. Am. Math. Soc. **275**, 811–819 (1983)

[W] Wu, J.Y.: A diameter pinching sphere theorem for positive Ricci curvature. Proc. Am. Math. Soc. **107**, 797–802 (1989)

[Y] Yamaguchi, T.: Lipschitz convergence of manifolds of positive Ricci curvature with large volume. Math. Ann. **284**, 423–436 (1989)

Math. Z. 206, 265–276 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Remarks on the existence and uniqueness of global decaying solutions of the nonlinear dissipative wave equations

Mitsuhiro Nakao

Department of Mathematics, College of General Education, Kyushu University, Fukuoka 810, Japan

Received February 16, 1990; in final form May 29, 1990

0 Introduction

Let Ω be a bounded domain in N dimensional Euclidean space R^N with smooth boundary $\partial\Omega$ (C^2 class is sufficient) and consider the initial-boundary value Problem:

$$(1) \qquad \begin{cases} u_{tt}-\Delta u+\sigma(u_t)+f(u)=0 & \text{on } \Omega\times R^+ \\ u(x,0)=u_0, \quad u_t(x,0)=u_1 & \text{and } u|_{\partial\Omega}=0, \end{cases}$$

where $\sigma(v)$ and $f(u)$ are functions like $\sigma(v)=|v|^r v$, $r>-1$, and $f(u)=|u|^\alpha u$, $\alpha>0$, respectively.

For simplicity we take $\sigma(v)=|v|^r v$ and $f(u)=|u|^\alpha u$ for a moment. Then, concerning existence and uniqueness of global solutions the following theorem is known as a standard one.

Theorem 0 (Cf. Lions [2], Strauss [10]). (i) *Assume that* $0\leqq\alpha<4/(N-2)$ ($0\leqq\alpha<\infty$ *if* $N=1, 2$). *Then, for each* $(u_0, u_1)\in H_1^0\cap L^{\max(r+2,2)}$ is $\times L^2$, $r>-1$, *the problem* (1) *admits a solution* u *in the class*

$$C^1([0,\infty); L^2)\cap C([0,\infty); H_1^0)\cap W^{1,r+2}([0,\infty); L^{r+2}).$$

(ii) *Moreover, if* $0\leqq\alpha\leqq 2/(N-2)$ ($0\leqq\alpha<\infty$ *if* $N=1, 2$), *the solution is unique.*
(iii) *Assume further* $(u_0, u_1)\in H_2\cap H_1^0\times H_1^0\cap L^{r+2}$ *and* $0\leqq\alpha\leqq 2/(N-2)$ ($0\leqq\alpha<\infty$ *if* $N=1, 2$). *Then, the solution belongs to*

$$W^{2,\infty}([0,\infty); L^2)\cap W^{1,\infty}([0,\infty); H_1^0)\cap L^\infty([0,\infty); H_2\cap H_1^0)\cap W^{1,r+2}([0,\infty); L^{r+2}).$$

For the assertions in the above Theorem the dissipative term $\sigma(u_t)$ does not work profitably and hence they are still valid for the case $\sigma(v)\equiv 0$ (we take $r=0$ in this case).

In the present paper we are interested in the existence and uniqueness of global solutions for the case $\alpha > 2/(N-2)$ which is excluded in the standard results. For such a case we will see that the dissipative term $\sigma(u_t)$ works in an essential manner and by this some global existence and uniqueness results are derived for certain small initial data. Such an assertion is known more or less for the linear dissipative case $\sigma(u_t) = v u_t$, $v = \text{const.} > 0$, or more generally for the definitely dissipative case $\sigma'(u_t) > \varepsilon_0 > 0$ (cf. [7, 9]). But, for our typical example $\sigma(v) = |v|^r v$ we see $\sigma'(0) = 0$ if $r > 0$ and $\lim_{v \to \infty} \sigma'(v) = 0$ if $-1 < r < 0$, and hence, the argument required here would become necessarily very delicate and difficult.

Fortunately, the decay property of the solutions of (1) with $f(u) \equiv 0$ or $f(u) = |u|^\alpha u$, $0 \leqq \alpha \leqq 2/(N-2)$, has been investigated in a series of our papers [3–6, 8] and the techniques or results in these works can be applied to our purpose. See also Haraux and Zuazua [1] and Zuazua [11] for related works on the decay of solutions.

1 Preliminaries and result

We use only familiar function spaces (real valued) $L^p(\Omega)$, $W^{k,p}(\Omega)$, $H_2(\Omega)$, $C([0, \infty); H_1^0)$ etc. ('Ω' in notation will be dropped in what follows) and the definitions of them are omitted. We shall denote by $\|\cdot\|_p$ the L^p norm and in particular we use $\|\cdot\|$ for $\|\cdot\|_2$.

The following two lemmas are essential for our arguments.

Lemma 1.1 (Gagliardo-Nirenberg). *Let $1 \leqq r < q \leqq \infty$ and $p \leqq q$. Then, the inequality*

$$\|u\|_{W^{k,q}} \leqq C \|u\|_{W^{m,p}}^{\theta} \|u\|_r^{1-\theta} \quad \text{for } u \in W^{m,p} \tag{2}$$

holds with some $C > 0$ and

$$\theta = \left(\frac{k}{N} + \frac{1}{r} - \frac{1}{q}\right)\left(\frac{m}{N} + \frac{1}{r} - \frac{1}{p}\right)^{-1}$$

provided that $0 < \theta \leqq 1$ (we assume $0 < \theta < 1$ if $q = \infty$).

Lemma 1.2 (Cf. [3]). *Let $\phi(t)$ be a nonnegative nonincreasing function on $[0, T]$, $T > 1$, such that*

$$\phi(t)^{1+\gamma} \leqq C_0(\phi(t) - \phi(t+1)) \quad \text{for } 0 \leqq t \leqq T-1 \tag{3}$$

with some $C_0 > 0$ and $\gamma > 0$. Then, we have

$$\phi(t) \leqq \{\phi(0)^{-\gamma} + \gamma C_0^{-1}(t-1)^+\}^{-1/\gamma} \quad \text{for } 0 \leqq t \leqq T \tag{4}$$

($(t-1)^+ \equiv \max(t-1, 0)$). (*If $\gamma = 0$ in (3) we get $\phi(t) \leqq \lambda e^{-\lambda t} \phi(0)$ with $\lambda = C_0(C_0 - 1)^{-1} > 0$.*)

Let us state our precise hypotheses on $\sigma(u)$ and $f(u)$.

Hyp. I σ belongs to $C^1(R-\{0\})\cap C(R)$ and satisfies

$$k_0 \min(|v|^2, |v|^{2+r}) \leqq \sigma(v)v \quad \text{for } v\in R, \tag{5}$$

$$|\sigma(v)| \leqq k_1\{(\sigma(v)v)^{(1+r)/(2+r)} + |v|\} \quad \text{for } v\in R \tag{6}$$

and

$$\sigma'(v) \geqq 0 \quad \text{for } v\in R(v\neq 0) \tag{7}$$

with some $k_0>0$, $k_1>0$ and $-1<r<\infty$.

Hyp. II f belongs to $C^1(R)$ and there exists $L>0$ (possibly $L=\infty$) such that

$$f(0)=0, \quad |f'(u)| \leqq k_2 |u|^\alpha \tag{8}$$

for some $0\leqq\alpha<\infty$, and

$$\|\nabla u\|^2 + \int_\Omega f(u)u\,dx \geqq k_3\{\|\nabla u\|^2 + \int_\Omega F(u)\,dx\} \geqq k_4 \|\nabla u\|^2 \tag{9}$$

for $u\in H_2\cap H_1^0$ with $\|\Delta u\| \leqq L$, where $k_2=k_2(L)$, $k_3=k_3(L)$ and $k_4=k_4(L)$ are some positive constants and we set $F(u)\equiv \int_0^u f(\eta)\,d\eta$.

It is easy to see that if $f(u)=|u|^\alpha u$, $\alpha>0$, Hyp. II is satisfied with $L=\infty$. Also if $f(u)=-|u|^\alpha u$ with $0<\alpha\leqq 4/(N-4)$ ($0<\alpha<\infty$ if $N=3, 4$), then Hyp. II is satisfied with a small $L>0$, which is observed by use of the inequality

$$\|u\|_{\alpha+2}^{\alpha+2} \leqq C\|\nabla u\|^2 \|\Delta u\|^\alpha \leqq CL^\alpha \|\nabla u\|^2 \tag{10}$$

if $0<\alpha\leqq 4/(N-4)$ ($0<\alpha<\infty$ if $N=3, 4$).

Our results read as follows.

Theorem 2.1. *Suppose that the Hypotheses I, II are satisfied with* r *and* α *such that*

(a) $0<r\leqq 2/(N-4)$ ($0<r<\infty$ *if* $N=3, 4$),

(b) $2/(N-2)<\alpha<2/(N-4)$ ($2/(N-2)<\alpha<\infty$ *if* $N=3, 4$)

(c) $(4-N)\alpha+2>2r$.

Then, there exists an open set $S_1\subset H_2\cap H_1^0\times H_1^0$ *including* $(0, 0)$ *such that if* $(u_0, u_1)\in S_1$, *the problem* (1) *admits a global unique solution* u *belonging to*

$$W^{2,\infty}([0,\infty); L^2)\cap W^{1,\infty}([0,\infty); H_1^0)\cap L^\infty([0,\infty); H_2)\cap W^{1,r+2}([0,\infty); L^{r+2})$$

and moreover, the solution u *meets the following decay or boundedness properties:*

$$\|u_t(t)\| + \|\nabla u(t)\| \leqq C_1(1+t)^{-1/r}, \tag{11}$$

and

$$\operatorname*{ess.\,sup}_{t\geqq 0}\{\|u_{tt}(t)\| + \|\nabla u_t(t)\| + \|\Delta u(t)\|\} < L. \tag{12}$$

Theorem 2.2. *Suppose that the Hypotheses I, II are satisfied with* α *and* r *such that*

(a) $-1<r<0$,

(b) $2/(N-2)<\alpha<2/(N-4)$ ($2/(N-2)<\alpha<\infty$ *if* $N=3, 4$)

and

(c) $(4-N)\alpha+2>-r\max\{N-2, 2(1+r)^{-1}\}$.

Then, there exists an open set $S_2\subset H_2\cap H_1^0\times H_1^0$ *including* $(0, 0)$ *such that if* $(u_0, u_1)\in S_2$, *the problem* (1) *admits a unique solution* u *belonging to*

$$W^{2,\infty}([0,\infty); L^2)\cap W^{1,\infty}([0,\infty); H_1^0)\cap L^\infty([0,\infty); H_2)$$

and moreover, the solution u *satisfies*

$$\|u_t(t)\|+\|\nabla u(t)\|\leqq C_1(1+t)^{(1+r)/r} \tag{13}$$

and

$$\operatorname*{ess.\,sup}_{t\geqq 0}\{\|u_{tt}(t)\|+\|\nabla u_t(t)\|+\|\Delta u(t)\|\}<L. \tag{14}$$

Remarks. (i) The precise definitions of S_i, $i=1, 2$, will be given in the course of the proofs.

(ii) For the most typical case: $N=3$, $\sigma(u_t)=|u_t|^r u_t$ and $f(u)=|u|^\alpha u$, the conditions in Theorem 2.1 are satisfied if

$$r>0, \quad 2<\alpha<\infty \quad \text{and} \quad \alpha+2>2r$$

and the conditions of Theorem 2.2 are satisfied if

$$-1<r<0, \quad 2<\alpha<\infty \quad \text{and} \quad \alpha+2>-2r/(1+r).$$

(Note that in this case we can take $L=\infty$.)

(iii) For the case $r=0$ the assertions of Theorem 2.1 is valid under the condition (b) with the estimate (11) replaced by $\|u_t(t)\|+\|\nabla u(t)\|\leqq C_1 e^{-\lambda t}$ for some $\lambda>0$. The proof is much easier than the case $r\neq 0$ and it is left to the readers.

3 Proof of the uniqueness

Here we shall prove the uniqueness part of Theorems 2.1 and 2.2.

Let u and v be two solutions in the sense of Theorems 2.1 or 2.2. Then, we have easily

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\{\|u_t(t)-v_t(t)\|^2+\|\nabla(u(t)-v(t))\|^2 \\
&\qquad+\int_\Omega\{\sigma(u_t)-\sigma(v_t)\}(u_t-v_t)\,dx \\
&=\int_\Omega\{f(u)-f(v)\}(u_t-v_t)\,dx \\
&\leqq C\int_\Omega(|u|^\alpha+|v|^\alpha)|u-v||u_t-v_t|\,dx \\
&\leqq C\{\int_\Omega(|u|+|v|)^{N\alpha}\,dx\}^{1/N}\|u-v\|_{2N/(N-2)}\|u_t-v_t\| \\
&\leqq C(\|u(t)\|_{H_2}+\|v(t)\|_{H_2})^\alpha\|u(t)-v(t)\|_{H_1}\|u_t-v_t\| \\
&\leqq C(T)\{\|\nabla(u(t)-v(t))\|^2+\|u_t(t)-v_t(t)\|^2\}, \quad 0\leqq t<T,
\end{aligned} \tag{15}$$

where we have used the facts

$$N\alpha<2N/(N-4) \quad \text{and} \quad \sup_{0\leqq t\leqq T}(\|u(t)\|_{H_2}+\|v(t)\|_{H_2})\leqq C(T)<\infty.$$

The inequality (15) together with the monotonicity of σ implies immediately $u(t)=v(t)$.

4 Proof of Theorem 2.1 (Existence)

We shall give the proof of existence part of Theorem 2.1. We can utilize the Galerkin's approximation method and a standard compactness argument for our purpose and for this it suffices to derive a priori estimates (11) and (12) by energy methods for an assumed smooth solution $u(t)$ (the results should be in fact applied to approximate solutions) (cf. Lions [2]). Throughout this section we assume that the conditions made in Theorem 2.1 are satisfied.

Proposition 4.1. *Let $u(t)$ be a (smooth) solution on $[0, T)$, $T>0$. Assume that*

$$\|\Delta u(t)\|\leqq K \quad \text{on } [0, T) \tag{16}$$

for some $K<L$. Then, it holds that

$$E(u(t))=\tfrac{1}{2}\{\|u_t(t)\|^2+\|\nabla u(t)\|^2\}+\int_\Omega F(u(t))\,dx\leqq g_K(t) \tag{17}$$

on $[0, T)$, where $g_K(t)$ is defined by

$$g_K(t)=\left[I_0^{-r}+\frac{r}{2}C^{-1}\{1+I_0^r+I_0^{2(r-\theta)/(1+\theta)}K^{(r+2)\theta/(1+\theta)}\}^{-1}(t-1)^+\right]^{-2/r} \tag{18}$$

with

$$\theta=\max\left\{\frac{r(N-2)-4}{2(r+2)},\,0\right\}, \qquad I_0^2\equiv E(u(0))=\tfrac{1}{2}\{\|u_1\|^2+\|\nabla u_0\|^2\}+\int_\Omega F(u_0)\,dx$$

and a constant $C>0$ independent of (u_0, u_1).

Proof. The proof is essentially included in [5] and we sketch it briefly.

Multiplying the Eq. (1) by u_t and integrating we have

$$\begin{aligned} k_0\int_t^{t+1}\int_\Omega \min(|u_t|^2, |u_t|^{2+r})\,dx\,ds \\ \leqq \int_t^{t+1}\int_\Omega \sigma(u_t)u_t\,dx\,ds=E(u(t))-E(u(t+1))\equiv D(t)^{r+2} \end{aligned} \tag{19}$$

for $0 \leqq t < T-1$. From this we see

$$\int_t^{t+1} \int_{\{|u_t(x,s)| \leqq 1\}} |u_t|^{2+r} \, dx \, ds \leqq k_0^{-1} D(t)^{r+2}$$

and

$$\int_t^{t+1} \int_{\{|u_t(x,s)| \leqq 1\}} |u_t|^2 \, dx \, ds \leqq C \left\{ \int_t^{t+1} \int_{\{|u_t(x,s)| \leqq 1\}} |u_t|^{2+r} \, dx \, ds \right\}^{2/(r+2)} \leqq C D(t)^2.$$

Thus, we have

$$\int_t^{t+1} \int_\Omega |u_t|^2 \, dx \, ds \leqq C \{ D(t)^2 + D(t)^{r+2} \} \tag{20}$$

and hence, there exist $t_1 \in [t, t+\frac{1}{4}]$ and $t_2 \in [t+\frac{3}{4}, t+1]$ such that

$$\|u_t(t_i)\|^2 \leqq C \{ D(t)^2 + D(t)^{r+2} \}, \qquad i = 1, 2. \tag{21}$$

Next, multiplying the Eq. (1) by $u(t)$ and integrating over $\Omega \times [t_1, t_2]$ we have

$$\begin{aligned} &\int_{t_1}^{t_2} \{ \|\nabla u(t)\|^2 + \int_\Omega f(u) u \, dx \} \, ds \\ &= \int_{t_1}^{t_2} \|u_t(s)\|^2 \, ds + (u_t(t_1), u(t_1)) - (u_t(t_2), u(t_2)) \\ &\quad + \int_{t_1}^{t_2} \int_\Omega \sigma(u_t) u \, dx \, ds. \end{aligned} \tag{22}$$

Here, by use of the assumption on σ and Hölder's inequality,

$$\begin{aligned} &\int_{t_1}^{t_2} \int_\Omega \sigma(u_t) u \, dx \, ds \\ &\leqq C \int_{t_1}^{t_2} \int_\Omega \{ (\sigma(u_t) u_t)^{(1+r)/(2+r)} + |u_t| \} |u| \, dx \, ds \\ &\leqq C \left\{ \int_{t_1}^{t_2} \int_\Omega \sigma(u_t) u_t \, dx \, ds \right\}^{(1+r)/(2+r)} \left\{ \int_{t_1}^{t_2} \|u\|_{r+2}^{r+2} \, ds \right\}^{1/(2+r)} \\ &\quad + C \left\{ \int_{t_1}^{t_2} \|u_t(s)\|^2 \, ds \right\}^{1/2} \sup_{t \leqq s \leqq t+1} \|u(s)\|. \end{aligned} \tag{23}$$

Moreover, by Lemma 2.1,

$$\|u(s)\|_{r+2} \leqq C \|u(s)\|_{2N/(N-2)}^{1-\theta} \|\Delta u\|^{\theta} \leqq C \|u(s)\|_{H_1^0}^{1-\theta} K^{\theta} \tag{24}$$

with

$$\theta = \max\left\{\frac{r(N-2)-4}{2(r+2)}, 0\right\} \quad (<1).$$

Thus, from (20)–(24) and (8) we get

$$\begin{aligned} &\int_{t_1}^{t_2} E(u(s))\, ds \\ &\leqq C\{D(t)^2 + D(t)^{r+2} + \sqrt{D(t)^2 + D(t)^{r+2}} \sup_{t \leqq s \leqq t+1} \sqrt{E(u(s))} \\ &\quad + D(t)^{r+1} K^{\theta} \sup_{t<s<t+1} E(u(s))^{(1-\theta)/2}\} \end{aligned} \tag{25}$$

and hence,

$$E(u(t_2)) \leqq 2\{\text{the right hand side of (25)}\}.$$

This and the inequality

$$E(u(t)) = \sup_{t \leqq s \leqq t+1} E(u(s)) \leqq E(u(t_2)) + \int_t^{t+1} \int_{\Omega} \sigma(u_t)\, u_t\, dx\, ds$$

imply

$$E(u(t)) \leqq C\{D(t)^{r+2} + D(t)^2 + D(t)^{2(r+1)/(1+\theta)} K^{2\theta/(1+\theta)}\}. \tag{26}$$

Now, we observe that $r > \theta$ under the condition (a) $0 < r < 2/(N-4)$ ($0 < r < \infty$ if $N = 3, 4$). Also, note that $D(t)^{r+2} \leqq E(u(t)) \leqq E(u(0)) \equiv I_0^2$. From these facts and (26) we obtain

$$E(u(t)) \leqq C\{I_0^{2r/(r+2)} + 1 + I_0^{4(r-\theta)/(1+\theta)(r+2)} K^{2\theta/(1+\theta)}\} D(t)^2$$

or

$$\begin{aligned} E(u(t))^{1+r/2} &< C\{1 + I_0^r + I_0^{2(r-\theta)/(1+\theta)} K^{\theta(r+2)/(1+\theta)}\} \\ &\quad \cdot \{E(u(t)) - E(u(t+1)\} \end{aligned} \tag{27}$$

which implies the desired estimate (17) by Lemma 2.2. q.e.d.

Proposition 4.2. *Let $u(t)$ be a (smooth) local solution on $[0, T)$, $T > 0$. Assume that*

$$\|\Delta u(t)\| \leqq K \quad \text{on } [0, T)$$

for some $0 < K (< L)$. Then, we have

$$E(\nabla u(t)) \equiv \tfrac{1}{2}\{\|\nabla u_t(t)\|^2 + \|\Delta u(t)\|^2\} \leqq Q_1(I_0, I_1, K) \quad \text{on } [0, T) \tag{28}$$

provided that the right hand side is meaningful, where we set

$$Q_1(I_0, I_1, K) \equiv \{I_1^{-(\alpha(N-2)-2)/2} - q_1(I_0, K)\}^{-4/(\alpha(N-2)-2)}. \tag{29}$$

Here,

$$q_1(I_0, K) \equiv C(\alpha(N-2)-2)\left\{I_0^{(1-\theta_0)} + \frac{2}{(1-\theta_0)-r} I_0^{\alpha(1-\theta_0)-r} \cdot (1 + I_0^r + I_0^{2(r-\theta)/(1-\theta)} K^{(r+2)\theta/(1+\theta)})\right\} \tag{29}$$

with

$$\theta_0 = \frac{N-2}{2} - \frac{1}{\alpha} \quad (<1), \qquad \theta = \max\left\{\frac{r(N-2)-4}{2(r+2)}, 0\right\}$$

and

$$I_1^2 \equiv E(\nabla u(0)) \equiv \tfrac{1}{2}\{\|\nabla u_1\|^2 + \|\Delta u_0\|^2\}.$$

Before proceeding to the proof we note that $\alpha(N-2)-2>0$ and $\alpha(1-\theta_0) - r = \{2r-(4-N)\alpha-2\}/2>0$ by the conditions (a) and (c), respectively. $Q_1(I_0, I_1, K)$ is meaningful if $I_1^{-(\alpha(N-2)-2)/2} - q_1(I_0, K)>0$, that is,

$$I_1^{\alpha(N-2)-2}\, q_1^2(I_0, K) < 1. \tag{30}$$

Proof of Proposition 4.2. Multiplying the Eq. (1) by $-\Delta u_t$ and integrating we have

$$\frac{d}{dt} E(\nabla u(t)) + \int_\Omega \sigma'(u_t) |\nabla u_t|^2 \, dx + \int_\Omega f'(u) \nabla u \nabla u_t \, dx = 0. \tag{31}$$

Using the assumptions on σ and f it follows from (31) that

$$\begin{aligned} \frac{d}{dt} E(\nabla u(t)) &\leqq C \int_\Omega |u|^\alpha |\nabla u| |\nabla u_t| \, dx \\ &\leqq C\{\int_\Omega |u|^{N\alpha} \, dx\}^{1/N} \\ &\quad \cdot \{\int_\Omega |\nabla u|^{2N/(N-2)} \, dx\}^{(N-2)/2N} \{\int_\Omega |\nabla u_t|^2 \, dx\}^{1/2} \\ &\leqq C \|u(t)\|_{N\alpha}^\alpha \|\Delta u(t)\| \, \|\nabla u_t(t)\| \\ &\leqq C \|u(t)\|_{N\alpha}^\alpha E(\nabla u(t)). \end{aligned} \tag{32}$$

Further, we observe that

$$\begin{aligned} \|u(t)\|_{N\alpha}^\alpha &\leqq C \|u(t)\|_{2N/(N-2)}^{\alpha(1-\theta_0)} \|\Delta u(t)\|^{\alpha\theta_0} \text{ (by Lemma 2.1)} \\ &\leqq C \|\nabla u(t)\|^{\alpha(1-\theta_0)} E(\nabla u(t))^{\alpha\theta_0/2} \end{aligned} \tag{33}$$

with $\theta_0 = \dfrac{N-2}{2} - \dfrac{1}{\alpha} (\leqq 1)$.

From (32), (33) and Proposition 4.1 we have

$$\frac{d}{dt}E(\nabla u(t)) \leqq C g_K(t)^{\alpha(1-\theta_0)/2} E(\nabla u(t))^{\alpha(N-2)/4+1/2}, \tag{34}$$

which implies

$$E(\nabla u(t)) < \left\{ I_1^{-(\alpha(N-2)-2)/2} - \frac{C}{4}(\alpha(N-2)-2) \cdot \int_0^t g_K(s)^{\alpha(1-\theta_0)/2}\, ds \right\}^{-4/(\alpha(N-2)-2)} \tag{35}$$

as long as the right hand side is meaningful.

Here, setting

$$P_K = \frac{r}{2}(1 + I_0^r + I_0^{2(r-\theta)/(1+\theta)} K^{(r+2)\theta/(1+\theta)})^{-1}$$

we see

$$\int_0^\infty g_K(s)^{\alpha(1-\theta_0)/2}\, ds = \int_0^\infty \{I_0^{-r} + P_K(s-1)^+\}^{-\alpha(1-\theta_0)/r}\, ds$$
$$= I_0^{\alpha(1-\theta_0)} + P_K^{-1}(\alpha(1-\theta_0)-r)^{-1} r I_0^{\alpha(1-\theta_0)-r},$$

where we have used the assumption (c) to ensure $\alpha(1-\theta_0)/r > 1$.

By (35) and (36) we arrive at the desired estimate (28). q.e.d.

Let $0 < K < L$ and define $S_K \subset H_2 \cap H_1^0 \times H_1^0$ by

$$S_K = \{(u_0, u_1) \in H_2 \cap H_1^0 \times H_1^0 \mid I_1^{(\alpha(N-2)-2)}\, q_1^2(I_0, K) < 1 \text{ and } Q_1(I_0, I_1, K) < 2^{-1} K^2\}. \tag{37}$$

Note that if I_0 and I_1 are sufficiently small, then $(u_0, u_1) \in S_K$. If $(u_0, u_1) \in S_K$ for some K, then the smooth (approximate) solution $u(t)$ of the problem (1) exists on $[0, T)$ and satisfies $\|\Delta u(t)\| < K$ on $[0, T)$ for some $T > 0$. Thus, we can apply Propositions 4.1 and 4.2 to get

$$E(\nabla u(t)) \leqq Q_1(I_0, I_1, K) < \tfrac{1}{2} K^2 \quad \text{on } [0, T)$$

and in particular,

$$\|\Delta u(t)\| \leqq \sqrt{2Q_1} < K \quad \text{on } [0, T).$$

Thus, we conclude that the smooth (approximate) solution exists on $[0, \infty)$ and the estimates (17) and (28) hold on $[0, \infty)$.

In conclusion, setting

$$S_1 = \bigcup_{0 < K < L} S_K \tag{38}$$

the problem (1) admits a desired solution on $[0, \infty)$ for the initial-datum $(u_0, u_1) \in S_1$. The proof of Theorem 2.1 is now complete.

5 Proof of Theorem 2.2 (Existence part)

In this section we assume $-1<r<0$ and give the proof of existence part of Theorem 2.2. It suffices for our purpose to derive a priori estimatea (13) and (14) for an assumed smooth solution $u(t)$.

Proposition 5.1. *Let $u(t)$ be a (smooth) solution of* (1) *such that*

$$E(\nabla u(t)) \leqq \tfrac{1}{2}K^2 \quad \text{on } [0, T),\ T>0,\ (0<K<L). \tag{39}$$

Then, under the conditions (a), (b) and (c) in Theorem 2.2 it holds that

$$E(u(t)) \leqq \left\{ I_0^{-(2\nu+r)/2(1-\nu)} + \frac{2\nu+r}{2(1-\nu)} G_K^{-1}(t-1)^+ \right\}^{-2(1-\nu)/(2\nu+r)} \equiv h_K(t) \tag{40}$$

for $0 \leqq t < T$, where we set

$$\nu = \max\left\{-r, \frac{-Nr}{4-Nr+r}\right\}, \qquad \theta = \frac{-Nr}{4-Nr+2r}$$

and

$$G_K \equiv C\{I_0^{4(\nu+r)/(2+r)} + K^{2\theta} I_0^{4(\nu-\theta)/(2+r)} + I_0^{2(2\nu+r)/(2+r)}\}. \tag{41}$$

Proof. It is proved in [8] (see also [6]) that the solution $u(t)$ satisfies the difference inequality

$$E(u(t)) \leqq C\{D(t)^{2+2r} + D(t)^{2(1-\theta)} \sup_{t \leqq s \leqq t+1} E(\nabla u(s))^{\theta} + D(t)^{2+r}\} \tag{42}$$

where $D(t)^{2+r} \equiv E(u(t)) - E(u(t+1))$. Therefore, under the assumption $E(\nabla u(t)) \leqq \frac{1}{2}K^2$ on $[0, T)$, we have

$$E(u(t))^{1+(2\nu+r)/2(1-r)} \leqq G_K\{E(u(t)) - E(u(t+1)), \tag{43}$$

which implies (40) by Lemma 2.2. q.e.d.

Proposition 5.2. *Let $u(t)$ be a smooth solution of* (1) *satisfying* (39). *Then, under the assumptions of Theorem 2.2 we have the estimate*

$$\begin{aligned} E(\nabla u(t)) &\leqq Q_2(I_0, I_1, K) \\ &\equiv \left\{ I_1^{-(\alpha(N-2)-2)/4} - \frac{C}{4}(\alpha(N-2)-2)\, q_2(I_0, K) \right\}^{-4/(\alpha(N-2)-2)} \end{aligned} \tag{44}$$

on $[0, T)$ provided that the right hand side is meaningful, where we set

$$q_2 = I_0^{((4-N)\alpha+2)/2} + \frac{4(1-\nu)}{2(2\nu+r)-(1-\nu)((4-N)\alpha+2)} G_K I_0^{((4-N)\alpha+2)/2-(2\nu+r)/2(1-\nu)}.$$

Proof. Similarly as in the proof of Proposition 4.2 we obtain again (34) with g_K replaced by h_K, i.e.,

$$\frac{d}{dt} E(\nabla u(t)) \leqq C h_K^{(1-\theta_0)/2} E(\nabla u(t))^{\alpha(N-2)/4+1/2}. \tag{45}$$

(The derivation of (45) is somewhat formal because $\sigma'(0)$ may not exist, but it is valid for an appropriate approximate solution.)

Now, we see

$$\begin{aligned}
&\int_0^\infty h_K(s)^{(1-\theta_0)/2}\, ds \quad \left(\theta_0 = \frac{N-2}{2} - \frac{1}{\alpha}\right) \\
&\quad = \int_0^\infty \left\{ I_0^{-(2\nu+r)/2(1-\nu)} + \frac{2\nu+r}{2(1-\nu)} G_K^{-1}(s-1)^+ \right\}^{-(1-\nu)((4-N)\alpha+2)/2(2\nu+r)} ds \\
&\quad = q_2(I_0, K)
\end{aligned}$$

providee that $(1-\nu)((4-N)\alpha+2)/2(2\nu+r)>1$, i.e., the condition (c) $(4-N)\alpha+2>-r\max\{N-2, 2(1+r)^{-1}\}$ holds. q.e.d.

We define

$$S_K = \left\{ (u_0, u_1) \in H_2 \cap H_1^0 \times H_1^0 \,\middle|\, \frac{C}{4}(\alpha(N-2)-2)\, q_2(I_0, K)\, I_1^{(\alpha(N-2)-2)/4} < 1 \right.$$
$$\left. \text{and } Q_2(I_0, I_1, K) < \frac{K^2}{2} \right\}$$

and set

$$S_2 = \bigcup_{0<K<L} S_K.$$

Then, it is clear that if $(u_0, u_1) \in S_2$, the corresponding solution of the problem (1) in the sence of Theorem 2.2 exists globally and satisfies the estimates

$$E(u(t)) \leqq h_K(t) \quad \text{and} \quad E(\nabla u(t)) \leqq Q_2(I_0, I_1, K) < K^2/2$$

for all $t \geqq 0$. The proof of Theorem 2.2 is now finished.

A final remark. As is seen from the proofs of Theorems 2.1 and 2.2, we can weaken the latter assumption in (8) of Hyp. I as follows:

$$|f'(u)| \leqq k_2(|u|^\alpha + |u|^\beta)$$

for some $0<\alpha\leqq\beta<\infty$. That is, our theorems are still valid if we replace the condition (b) by

(b′) $$2/(N-2) < \alpha \leqq \beta < 2/(N-4)$$

$$(2/(N-2) < \alpha \leqq \beta < \infty \text{ if } N = 3, 4).$$

References

1. Haraux, A., Zuazua, E.: Decay estimates for some semilinear damped hyperbolic problems. Arch. Ration. Mech. Anal. **100**, 191–206 (1987)
2. Lions, J.L.: Quelque Méthodes de Resolution de Problèmes aux Limites Non-Linèares. Paris: Dunod Gautier Villars 1969
3. Nakao, M.: Asymptotic stability of the bounded or almost periodic solution of the wave equation with nonlinear dissipative term. J. Math. Anal. Appl. **58**, 336–343 (1977)
4. Nakao, M.: Periodic solution and decay for some nonlinear wave equations with sublinear dissipative terms. Nonlinear Anal. Theory. Methods Appl. **10**, 587–602 (1986)
5. Nakao, M.: On the decay of solutions of some nonlinear dissipative wave equations in higher dimensions. Math. Z. **193**, 227–234 (1986)
6. Nakao, M.: On solutions of the wave equation with a sublinear dissipative term. J. Differ. Equations **69**, 204–215 (1987)
7. Nakao, M.: Existence of global classical solutions of the initial-boundary value problem for some nonlinear wave equations. J. Math. Anal. Appl. **146**, 217–240 (1990)
8. Nakao, M.: On solutions of the wave equation with a sublinear dissipative term II. Funkc. Ekvacioj, Ser. Int. **33**, 307–316 (1990)
9. Shibata, Y.: On the global existence of classical solutions of mixed problem for some second order nonlinear hyperbolic operators with dissipative term in the interior domain. Funkc. Ekvacioj, Ser. Int. **25**, 303–345 (1982)
10. Strauss, W.: Energy Methods in Nonlinear Partial Differential Equations. Notas de Matematics, Rio de Janeiro: 1969
11. Zuazua, E.: Stability and decay for a class of nonlinear hyperbolic problems. Asymptotic Anal. **1**, 161–185 (1988)

Math. Z. 206, 277–291 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Elliptic genera, modular forms over KO_*, and the Brown-Kervaire invariant

Serge Ochanine*
CNRS, Université de Paris-Sud, Mathématique, bât 425, F-91405 Orsay, France

Received December 7, 1989; in final form April 21, 1990

Let Ω_*^{SO} be the oriented cobordism ring and Λ any commutative $\mathbb{Q}$-algebra. An *elliptic genus* over Λ, as originally defined in [14], is a ring homomorphism

$$\varphi\colon \Omega_*^{SO} \to \Lambda$$

satisfying

$$\sum_{i \geq 0} \varphi[\mathbb{C}P_{2i}]\, u^{2i} = (1 - 2\delta u^2 + \varepsilon u^4)^{-1/2}$$

Here

$$\delta = \varphi[\mathbb{C}P_2] \quad \text{and} \quad \varepsilon = \varphi[\mathbb{H}P_2]$$

are two parameters in Λ which determine φ completely.

In the most interesting *universal* examples, Λ is the ring $\mathbb{Q}[[q]]$ of formal power series over $\mathbb{Q}$, and for any oriented manifold V, $\varphi[V]$ is the q-expansion of a level 2 modular form whose values at the two cusps are, up to an inessential factor, the $\hat{A}$-genus $\hat{A}[V]$ and the signature $\sigma(V)$ (cf. [9, 5, 10, 23, 8]).

Though defined for oriented manifolds, the elliptic genera reveal their most striking properties, such as rigidity (constancy) under compact Lie group actions ([3, 15]) or integrality [6], on spin manifolds. Both rigidity and integrality rely on the fact noticed by E. Witten [22] that in the universal examples, the coefficients of $\varphi[V]$ are indices of twisted Dirac operators, therefore KO-characteristic numbers.

In this paper we consider a refined elliptic genus

$$\beta_q\colon \Omega_*^{\text{spin}} \to KO_*[[q]]$$

whose values are q-expansions of level 2 modular forms over the coefficient ring KO_* of the real K-theory. In dimensions divisible by 4, $\beta_q[V]$ is essentially the above universal genus $\varphi[V]$. On the other hand, in dimensions $8m+1$,

* *Present address:* Max-Planck-Institut für Mathematik, Gottfried-Claren-Str. 26, W-5300 Bonn 3, Federal Republic of Germany

$8m+2$, $\beta_q[V]$ is a modular form over $\mathbb{F}_2$ (in the sense of J.-P. Serre [18]), and can be expressed as a polynomial in the basic form $\bar{\varepsilon}=\sum_{n\geq 1} q^{(2n-1)^2}$:

$$\beta_q[V]=a_0+a_1\,\bar{\varepsilon}+\ldots+a_m\,\bar{\varepsilon}^m.$$

It turns out that a_0 is the Atiyah invariant while a_m is the KO-part of the Brown-Kervaire invariant of V.

Being a refinement of an elliptic genus, β_q retains at least a few of the properties of the latter. For example, M. Bendersky ([2]) recently proved that $\beta_q[V]=0$ for a spin manifold V admitting an odd type semi-free circle action, which implies the vanishing of both the Atiyah invariant and the KO-part of the Brown Kervaire invariant*.

1 Definition of β_q

Let E be a real vector bundle over X. Writing $\Lambda^i(E)$ and $S^i(E)$ respectively for the exterior and the symmetric powers of E, and

$$\Lambda_t(E)=\sum_{i\geq 0}\Lambda^i(E)\,t^i,$$
$$S_t(E)=\sum_{i\geq 0}S^i(E)\,t^i,$$

one defines the *Witten characteristic class* Θ_q ([22], cf. [10]) by

$$\Theta_q(E)=\bigotimes_{n\geq 1}(\Lambda_{-q^{2n-1}}(E)\otimes S_{q^{2n}}(E)).$$

For any E, $\Theta_q(E)$ is a formal power series in q whose coefficients are virtual vector bundles over X. Moreover, one has

$$\Theta_q(E)=1-E\cdot q+\ldots$$

and

$$\Theta_q(E\oplus F)=\Theta_q(E)\cdot\Theta_q(F).$$

Therefore Θ_q canonically extends to $KO(X)$:

$$\Theta_q\colon\ KO(X)\to KO(X)[[q]].$$

Let $\beta_q(E)$ be defined by

$$\beta_q(E)=\Theta_q(E-\dim E).$$

Then

$$\beta_q(E)=b_0(E)+b_1(E)\,q+\ldots$$

where

$$b_0(E)=1$$
$$b_i(E)\in\widetilde{KO}(X)\qquad(i>0)$$

and

$$\beta_q(E\oplus F)=\beta_q(E)\cdot\beta_q(F).$$

* For a proof valid for all odd type actions see [16].

It is easy to see that $b_i (i \geqq 0)$ are stable KO-characteristic classes and can be expressed as polynomials in the Pontrjagin classes π_i defined by (cf. [21]):

$$\sum \pi_i(E)\, u^i = \Lambda_t(E - \dim E),$$

where

$$u = \frac{t}{(1+t)^2}.$$

For example

$$b_1 = -\pi_1$$
$$b_2 = \pi_2 - \pi_1$$
$$b_3 = -\pi_3 + 4\pi_2 - \pi_1^2 - 4\pi_1$$

and, more generally,

$$b_i = (-1)^i \pi_i + \text{lower terms}.$$

Let now V^n be a closed spin manifold, and $[V^n] \in KO_n(V^n)$ be the fundamental class of V^n in real K-theory.

Definition.

$$\beta_q[V^n] = \beta_q(TV)[V^n] = \sum_{i \geqq 0} b_i[V^n]\, q^i,$$

where TV is the tangent bundle of V^n and

$$b_i[V^n] = b_i(TV)[V^n] \in KO_n = KO_n(\text{point})$$

is the KO-characteristic number corresponding to b_i.

One can easily see that β_q defines a ring homomorphism (genus)

$$\beta_q \colon \Omega_*^{\text{spin}} \to KO_*[[q]].$$

Considered as $\mathbb{Z}/8$-graded, the ring KO_* is generated by two elements η and ω of degree 1 and 4 respectively subject to the relations

$$2\eta = \eta^3 = \eta\omega = 0, \qquad \omega^2 = 4.$$

Clearly, β_q preserves the degree mod 8.

Let

$$ph \colon KO^*(X) \to H^{**}(X; \mathbb{Q})$$

be the Pontrjagin character defined as the composition

$$KO^*(X) \xrightarrow{\otimes \mathbb{C}} K^*(X) \xrightarrow{\text{Chern char.}} H^{**}(X; \mathbb{Q}).$$

For $X = \text{point}$ one has $KO^*(X) \cong KO_*$ and $H^{**}(X; \mathbb{Q}) \cong \mathbb{Q}$, and ph is entirely determined by

$$ph(\eta) = 0, \qquad ph(\omega) = 2.$$

In particular, ph is integral:

$$ph \colon KO_* \to \mathbb{Z}.$$

Composing β_q with ph leads to a genus

$$\varphi_q = ph \circ \beta_q \colon \Omega_*^{\text{spin}} \to \mathbb{Z}[[q]]$$

such that

$$\varphi_q[V^n] = \sum_{i \geq 0} ph(b_i[V^n])\, q^i = \sum_{i \geq 0} ph(b_i(TV))\, \hat{\mathfrak{A}}(TV)[V^n]\, q^i$$

where $\hat{\mathfrak{A}}(TV)$ is the total $\hat{\mathfrak{A}}$-class of V^n. In particular, the constant term of $\varphi_q[V^n]$ is the $\hat{A}$-genus $\hat{A}[V^n]$.

Theorem 1 ([10, 23]) *φ_q is the restriction to Ω_*^{spin} of an elliptic genus*

$$\varphi_q \colon \Omega_*^{SO} \to \mathbb{Q}[[q]]$$

with parameters

$$\delta = -\tfrac{1}{8} - 3 \sum_{n \geq 1} \Big(\sum_{\substack{d \mid n \\ d \text{ odd}}} d \Big) q^n$$

$$\varepsilon = \sum_{n \geq 1} \Big(\sum_{\substack{d \mid n \\ n/d \text{ odd}}} d^3 \Big) q^n \quad \square$$

2 Modular forms over graded rings

It turns out that $\beta_q[V^n]$ can be interpreted as a modular form of degree n over the graded ring KO_*.

If Γ is a subgroup of $SL_2(\mathbb{Z})$ of finite index, let $M_*^\Gamma(\mathbb{C})$ be the graded ring of modular forms over $\mathbb{C}$ for Γ. Thus $M_w^\Gamma(\mathbb{C})$ is the group of forms of weight w. We will always identify a modular form from $M_*^\Gamma(\mathbb{C})$ with its q-expansion. This way $M_*^\Gamma(\mathbb{C})$ becomes a subring in $\mathbb{C}[[q^{1/h}]]$, where h is the smallest positive integer such that $\begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix}$ belongs to Γ.

Let now $M_*^\Gamma(\mathbb{Z})$ be the graded subring of $M_*^\Gamma(\mathbb{C})$ of forms having integral q-expansions

$$M_*^\Gamma(\mathbb{Z}) = \bigoplus_n M_n^\Gamma(\mathbb{Z})$$

$$M_n^\Gamma(\mathbb{Z}) = M_n^\Gamma(\mathbb{C}) \cap \mathbb{Z}[[q^{1/h}]].$$

For any graded commutative ring with unit

$$R_* = \bigoplus_n R_n,$$

the canonical injection

$$M_*^\Gamma(\mathbb{Z}) \to \mathbb{Z}[[q^{1/h}]]$$

extends to a ring homomorphism

$$R_* \otimes_{\mathbb{Z}} M_*^\Gamma(\mathbb{Z}) \to R_*[[q^{1/h}]].$$

We define $M^\Gamma(R_*)$ to be the image of this homomorphism, and will call its elements *modular forms over R_* for Γ*.

Notice that $M^\Gamma(R_*)$ is canonically a graded R_*-algebra:

$$M^\Gamma(R_*)=\bigoplus_n M^\Gamma(R_n),$$

where $M^\Gamma(R_n)$ is the image of $R_n\otimes M^\Gamma_*(\mathbb{Z})$. We refer to the elements of $M^\Gamma(R_n)$ as forms of *degree n*.

If for a certain n, R_n has no torsion, then

$$R_n\otimes M^\Gamma_*(\mathbb{Z})\to M^\Gamma(R_n)$$

is an isomorphism. In this case,

$$M^\Gamma(R_n)=\bigoplus_w M^\Gamma_w(R_n),$$

where

$$M^\Gamma_w(R_n)\cong R_n\otimes M^\Gamma_w(\mathbb{Z}).$$

We will say that forms from $M^\Gamma_w(R_n)$ have *weight w*.

In the general situation, a form $F\in M^\Gamma(R_n)$ may come from integral forms of different weights, and the weight of F cannot be defined correctly. Instead, one defines an increasing *filtration* of $M^\Gamma(R_n)$ as follows: a form $F\in M^\Gamma(R_n)$ has filtration $\leqq f$ if F is the image of an element of

$$R_n\otimes\Big(\bigoplus_{w\leqq f} M^\Gamma_w(\mathbb{Z})\Big),$$

i.e. if

$$F=\sum r_j F_j,$$

where $F_j\in M^\Gamma_*(\mathbb{Z})$ are forms of weight $\leqq f$ and $r_j\in R_n$.

3 Modular forms over KO_*

From now on Γ will designate the group $\Gamma_0(2)$ of matrices

$$\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in SL_2(\mathbb{Z})$$

such that $c\equiv 0(\mathrm{mod}\ 2)$. The series δ and ε of theorem 1 are the basic examples of modular forms for $\Gamma_0(2)$. More precisely, let

$$\delta_0=-8\delta=1+24q+24q^2+96q^3+\ldots$$

Proposition 1 (cf. [8], Anhang I) (i) $\delta_0\in M^\Gamma_2(\mathbb{Z})$, $\varepsilon\in M^\Gamma_4(\mathbb{Z})$;
(ii) $M^\Gamma_*(\mathbb{Z})=\mathbb{Z}[\delta_0,\varepsilon]$.

Consider now $M^\Gamma(KO_*)$. For $n\equiv 0(\mathrm{mod}\ 4)$, one has $KO_n\cong\mathbb{Z}$. Thus

$$M^\Gamma(KO_n)\cong KO_n\otimes M^\Gamma_*(\mathbb{Z}).$$

It follows that:

(a) a modular form of degree $n=8m$ and weight w over KO_* can be written in a unique way as a polynomial $P(\delta_0, \varepsilon)$ of weight w with integer coefficients;
(b) a modular form of degree $n=8m+4$ and weight w over KO_* can be written in a unique way as $\omega P(\delta_0, \varepsilon)$, where $P(\delta_0, \varepsilon)$ is a polynomial of weight w with integer coefficients.

Notice now that one has $\delta_0 \equiv 1 \pmod 2$. Let $\bar{\varepsilon}$ be the reduction mod 2 of $\varepsilon \in \mathbb{Z}[[q]]$. It is easy to see that

$$\bar{\varepsilon} = \sum_{n \geqq 1} q^{(2n-1)^2} = q + q^9 + q^{25} + \ldots$$

For $n=8m+r (r=1, 2)$, one has $KO_n = \mathbb{F}_2 \eta^r$ and the map

$$KO_n \otimes M^{\Gamma}_*(\mathbb{Z}) \to KO_n[[q]]$$

is essentially the reduction mod 2:

$$\eta^r \otimes P(\delta_0, \varepsilon) \mapsto \eta^r \bar{P}(1, \bar{\varepsilon}),$$

where $P(\delta_0, \varepsilon)$ is a polynomial with integer coefficients and $\bar{P}$ is its reduction mod 2. As $\bar{\varepsilon} = q + \ldots$, the powers of $\bar{\varepsilon}$ are linearly independent over $\mathbb{F}_2$. Therefore:

(c) a modular form F of degree $n=8m+r (r=1, 2)$ and filtration $\leqq f$ over KO_* can be written in a unique way as $\eta^r Q(\bar{\varepsilon})$, where

$$Q(\bar{\varepsilon}) = a_0 + a_1 \bar{\varepsilon} + \ldots + a_s \bar{\varepsilon}^s (a_i \in \mathbb{F}_2)$$

and $4s \leqq f$. The filtration of F is exactly $4s$ if and only if $a_s \neq 0$.

The additive structure of $M^{\Gamma}(KO_*)$ is completely described by (a), (b), and (c). The ring structure is given by the following theorem.

Theorem 2 (i) *The kernel of*

$$KO_* \otimes M^{\Gamma}_*(\mathbb{Z}) \to M^{\Gamma}(KO_*)$$

is the principal ideal generated by $\eta \otimes (\delta_0 - 1)$.
(ii) *The commutative* KO_**-algebra* $M^{\Gamma}(KO_*)$ *is generated by* δ_0 *and* ε *subject to the single relation* $\eta \delta_0 = \eta$.

The proof is immediate from the above description of

$$KO_* \otimes M^{\Gamma}_*(\mathbb{Z}) \to KO_*[[q]].$$

4 $\beta_q[V^n]$ as a modular form

We will now see that $\beta_q[V^n]$ is a modular form of degree n over KO_*.

Theorem 3 (i) *If* $n=4s$, *then* $\beta_q(\Omega_n^{\text{spin}})$ *is the set of all modular forms of degree* n *and weight* $2s$ *over* KO_*.
(ii) *If* $n=8m+r (r=1, 2)$, *then* $\beta_q(\Omega_n^{\text{spin}})$ *is the set of all modular forms of degree* n *and filtration* $\leqq 4m$ *over* KO_*.
(iii) $\beta_q(\Omega_*^{\text{spin}})$ *is the subring of* $M^{\Gamma}(KO_*)$ *generated by* η, $\omega\delta_0$, δ_0^2 *and* ε.

Proof. Part (iii) clearly follows from (i), (ii) and the above description of $M^{\Gamma}(KO_*)$.

Part (i) is a simple consequence of the definition of φ_q, the description of ph and the following theorem:

Theorem 4 ([6], cf. [10]) *For any spin manifold V^{4s}, $\varphi_q[V^{4s}]$ is a modular form from $M^{\Gamma}_{2s}(\mathbb{Z})$. More precisely,*

$$\varphi_q(\Omega^{\text{spin}}_{8m}) = M^{\Gamma}_{4m}(\mathbb{Z})$$
$$\varphi_q(\Omega^{\text{spin}}_{8m+4}) = 2M^{\Gamma}_{4m+2}(\mathbb{Z})$$

The proof of the remaining part (ii) relies on the following construction due to R.E. Stong (cf. [21, p. 341] for the details):

Let $\bar{S}^1$ be the circle equipped with its non-trivial spin structure. $\bar{S}^1$ represents the non-zero element of $\Omega^{\text{spin}}_1 \cong \mathbb{F}_2$. If V is an $(8m+2)$-dimensional spin manifold, then $\bar{S}^1 \times V$ is the boundary of a compact spin manifold U. On the other hand, $2\bar{S}^1$ is the boundary of a compact spin manifold M^2. Therefore one can form a closed $(8m+4)$-dimensional spin manifold $T(V)$ by glueing together two copies of U and $-M^2 \times V$ along

$$\partial(2U) = 2\bar{S}^1 \times V = \partial(M^2 \times V).$$

Though involving arbitrary choices of M^2 and U, this construction induces a well-defined homomorphism

$$T\colon\ \Omega^{\text{spin}}_{8m+2} \to \Omega^{\text{spin}}_{8m+4} \otimes \mathbb{F}_2.$$

Let

$$t\colon\ KO_2 \to KO_4 \otimes \mathbb{F}_2$$

be the isomorphism which sends η^2 to $\omega \otimes 1$.

Proposition 2 (cf. [21, p. 343]) *If ξ is a polynomial in the Pontrjagin classes π_i, then one has in $KO_4 \otimes \mathbb{F}_2$:*

$$\xi[T(V)] \otimes 1 = t(\xi[V]). \quad \square$$

Roughly speaking, $\xi[V]$ is the reduction mod 2 of $\xi[T(V)]$.

Let $I_* \subset \Omega^{\text{spin}}_*$ be the ideal of classes with vanishing Pontrjagin KO-characteristic numbers. Proposition 2 implies that T induces a homomorphism

$$\tilde{T}\colon\ \Omega^{\text{spin}}_{8m+2}/I_{8m+2} \to (\Omega^{\text{spin}}_{8m+4}/I_{8m+4}) \otimes \mathbb{F}_2.$$

Proposition 3 (cf. [21, p. 344]) *$\tilde{T}$ is an isomorphism.* $\square$

The coefficients of $\beta_q[V]$ are Pontrjagin KO-characteristic numbers. Therefore one has:

$$\beta_q[T(V)] \otimes 1 = t(\beta_q[V])$$

in $(KO_4 \otimes \mathbb{F}_2)[[q]]$. By Theorem 3 (i),

$$\beta_q[T(V)] = \omega P(\delta_0, \varepsilon),$$

where $P(\delta_0, \varepsilon)$ is a polynomial of weight $4m+2$ in δ_0, ε with integer coefficients. Therefore

$$\beta_q[V] = \eta^2 \bar{P}(1, \bar{\varepsilon})$$

is a modular form of degree $8m+2$ and filtration $\leqq 4m$ over KO_*. Proposition 3 implies that all such forms can be obtained from spin manifolds V, and this settles the case of manifolds of dimension $8m+2$.

The proof in the case of $(8m+1)$-dimensional manifolds is similar. Instead of T one considers the multiplication by $\bar{S}^1$ homomorphism

$$S\colon\ \Omega^{\mathrm{spin}}_{8m} \to \Omega^{\mathrm{spin}}_{8m+1}.$$

If ξ is a polynomial in the classes π_i, then

$$\xi[\bar{S}^1 \times M] = \eta \cdot \xi[M]$$

for any spin manifold M. Thus S induces a homomorphism

$$\tilde{S}\colon\ \Omega^{\mathrm{spin}}_{8m}/I_{8m} \to \Omega^{\mathrm{spin}}_{8m+1}/I_{8m+1}.$$

Proposition 4 (cf. [21, p. 344]) *$\tilde{S}$ is into.* □

It follows that

$$\beta_q(\Omega^{\mathrm{spin}}_{8m+1}) = \eta \cdot \beta_q(\Omega^{\mathrm{spin}}_{8m})$$

and the result follows from (i) and the description of $M^r(KO_*)$.

5 Characteristic classes a_i

Let $h(q) = q + \ldots$ be any series from $\mathbb{Z}[[q]]$ whose reduction mod 2 is

$$\sum_{n \geqq 1} q^{(2n-1)^2} = q + q^9 + q^{25} + \ldots$$

For example, one can take $h(q) = \varepsilon(q)$. Another possible choice for $h(q)$ is the Ramanujan series

$$\Delta(q) = q \prod_{n \geqq 1} (1-q^n)^{24} = q - 24q^2 + 252q^3 - \ldots \star$$

For any real vector bundle E over X define

$$\alpha_t(E) \in KO(X)[[t]]$$

by

$$\alpha_t(E) = \beta_q(E),$$

where

$$t = h(q).$$

Since the leading term of $h(q)$ is q, this series is invertible in $\mathbb{Z}[[q]]$, therefore $\alpha_t(E)$ is well-defined. Clearly, one has

$$\alpha_t(E \oplus F) = \alpha_t(E)\,\alpha_t(F).$$

$\star$ It is an amusing exercise to show that $\Delta \equiv \varepsilon \pmod 2$ and even, as noticed by P. Landweber, $\Delta \equiv \varepsilon \pmod{16}$.

If

$$\alpha_t(E) = a_0(E) + a_i(E)\, t + a_2(E)\, t^2 + \ldots,$$

then $a_i(E)$ is a polynomial in the Pontrjagin classes $\pi_i(E)$ such that

$$a_0(E) = 1$$
$$a_i(E) \in \widetilde{KO}(X) \qquad (i > 0)$$

and

$$a_i(E) = (-1)^i\, \pi_i(E) + \text{lower terms.}$$

Notice that while $a_i(E)$ depends on the choice of $h(q)$, its reduction mod 2, that is its image in $KO(X) \otimes \mathbb{F}_2$ is independent of any choice.

By definition of a_i, for any spin manifold V^n one has:

$$\beta_q[V^n] = a_0[V^n] + a_1[V^n]\, t + a_2[V^n]\, t^2 + \ldots,$$

where

$$a_i[V^n] = a_i(TV)[V^n].$$

On the other hand, the reduction mod 2 of $\beta_q[V^n]$ is of the form (cf. Sect. 3):

$$a_0 + a_1\, \bar{\varepsilon} + \ldots + a_m\, \bar{\varepsilon}^m,$$

where $a_i \in KO_n \otimes \mathbb{F}_2$ and $m = [n/8]$. Comparing these two expressions leads to the following:

Theorem 5 (i) *For $i > [n/8]$, one has $a_i[V^n] \otimes 1 = 0$ in $KO_n \otimes \mathbb{F}_2$.*
(ii) *One has in $(KO_n \otimes \mathbb{F}_2)[[q]]$:*

$$\beta_q[V^n] \equiv a_0[V^n] + a_1[V^n]\, \bar{\varepsilon} + \ldots + a_m[V^n]\, \bar{\varepsilon}^m,$$

where $m = [n/8]$.

6 The Brown-Kervaire invariant

Notice that for $n = 8m+2$, the constant term $a_0[V^n] = 1[V^n]$ is the so-called Atiyah invariant ([1]). We will see now that $a_m[V^n]$ has an interpretation in terms of the Brown-Kervaire invariant of V^n.

Let V^n, $n = 8m+2$, be a spin manifold. As mentioned earlier, $S^1 \times V = \partial U$, where U is a compact spin manifold. It is shown in [13] that the signature $\sigma(U)$ is divisible by 8, and that

$$k(V) = \sigma(U)/8 \in \mathbb{F}_2$$

is a spin cobordism invariant satisfying

$$k(S^1 \times S^1 \times M) = \sigma(M) \bmod 2$$

for all $8m$-dimensional spin manifolds M. For a large class of manifolds, including all complex-spin manifolds [20], $k(V)$ agrees with the Brown-Kervaire invariant [4]. For a general spin manifold V, $k(V)$ can be thought of as the KO-part of the Brown-Kervaire invariant (cf. [13] for the details).

More generally, one defines an invariant $\kappa(V^n) \in KO_n \otimes \mathbb{F}_2$ by

$$\kappa(V^n) = \begin{cases} \sigma(V), & n \equiv 0 (\mathrm{mod}\ 8) \\ k(\bar{S}^1 \times V)\eta, & n \equiv 1 (\mathrm{mod}\ 8) \\ k(V)\eta^2, & n \equiv 2 (\mathrm{mod}\ 8) \\ (\sigma(V)/16)\omega, & n \equiv 4 (\mathrm{mod}\ 8) \end{cases}$$

The multiplicative properties of k are summarized by saying that κ defines a ring homomorphism

$$\kappa\colon\ \Omega_*^{\mathrm{spin}} \to KO_* \otimes \mathbb{F}_2.$$

A new proof of this will be given later.

Theorem 6 *Let V^n be a spin manifold. Then*

$$a_m[V^n] = \kappa(V^n)$$

in $KO_n \otimes \mathbb{F}_2$, where $m = [n/8]$.

Proof. Consider first the case when $n = 8m+4$. According to Theorem 3,

$$\beta_q[V^n] = \omega(a_0 \delta_0^{2m+1} + a_1 \delta_0^{2m-1} \varepsilon + \ldots + a_m \delta_0 \varepsilon^m),$$

where $a_i \in \mathbb{Z}$. Then

$$\varphi_q[V^n] = 2(a_0 \delta_0^{2m+1} + a_1 \delta_0^{2m-1} \varepsilon + \ldots + a_m \delta_0 \varepsilon^m).$$

If we consider φ_q as an elliptic genus over $\mathbb{Z}[\delta, \varepsilon]$, the signature $\sigma(V^n)$ is obtained by specializing $\delta = 1$, $\varepsilon = 1$, or $\delta_0 = -8$, $\varepsilon = 1$. Thus,

$$\begin{aligned} \sigma(V^n) &= 2(a_0(-8)^{2m+1} + a_1(-8)^{2m-1} + \ldots + a_m(-8)) \\ &\equiv 16 a_m (\mathrm{mod}\ 32), \end{aligned}$$

and

$$\kappa(V^n) = a_m \omega \bmod 2.$$

On the other hand, by Theorem 5,

$$a_m \omega = a_m[V^n] \bmod 2,$$

therefore

$$\kappa(V^n) = a_m[V^n] \bmod 2.$$

If $n = 8m+2$, Proposition 2 gives

$$\begin{aligned} t(a_m[V^n]) &= a_m[T(V)] \bmod 2 \\ &= (\sigma(T(V))/16)\,\omega \bmod 2 \end{aligned}$$

by the previous case.

By definition,

$$T(V) = (2U) \cup (-M^2 \times V),$$

where $\partial U = \bar{S}^1 \times V$. Thus

$$\sigma(T(V)) = 2\sigma(U).$$

On the other hand,

$$k(V)=\frac{\sigma(U)}{8}=\frac{\sigma(T(V))}{16} \bmod 2.$$

Comparing with the above expression for $t(a_m[V^n])$, we obtain:

$$a_m[V^n]=k(V^n)\,\eta^2=\kappa(V^n).$$

If $n=8m+1$,

$$a_m[V^n]\,\eta=a_m[\bar{S}^1\times V^n]=k(\bar{S}^1\times V^n)\,\eta^2,$$

therefore

$$a_m[V^n]=k(\bar{S}^1\times V^n)\,\eta=\kappa(V^n)$$

since the multiplication by η is an isomorphism $KO_1 \xrightarrow{\cong} KO_2$.

Finally, if $n=8m$, then

$$a_m[V^n]\,\eta^2=a_m[\bar{S}^1\times\bar{S}^1\times V^n]=k(\bar{S}^1\times\bar{S}^1\times V^n)\,\eta^2=\sigma(V^n)\,\eta^2,$$

and

$$a_m[V^n]\equiv\sigma(V^n)(\bmod 2). \quad \square$$

Corollary 1 *$\kappa\colon \Omega_*^{\mathrm{spin}}\to KO_*\otimes\mathbb{F}_2$ is a ring homomorphism.*

Proof. Let V_1 and V_2 be two spin manifolds of dimension n_1 and n_2 respectively, and let

$$m_1=[n_1/8],\qquad m_2=[n_2/8],\qquad m=[(n_1+n_2)/8].$$

By Theorem 6,

$$\kappa(V_1\times V_2)=a_m[V_1\times V_2]=\sum_{i_1+i_2=m} a_{i_1}[V_1]\,a_{i_2}[V_2].$$

Notice that $m\geqq m_1+m_2$. If $m=m_1+m_2$, then Theorem 5 (i) and theorem 6 imply:

$$\kappa(V_1\times V_2)=a_{m_1}[V_1]\,a_{m_2}[V_2]=\kappa(V_1)\,\kappa(V_2).$$

If $m>m_1+m_2$, then Theorem 5 (i) gives

$$\kappa(V_1\times V_2)=0$$

and one has to check that

$$\kappa(V_1)\,\kappa(V_2)=0.$$

But $m>m_1+m_2$ is possible only in one of the following cases:

(1) $$n_1\equiv n_2\equiv 4(\bmod 8).$$

In this case

$$\kappa(V_1)\,\kappa(V_2)=0$$

since $\omega^2\equiv 0(\bmod 2)$.

(2) $$n_1\equiv 5,6,7(\bmod 8)\quad\text{or}\quad n_2\equiv 5,6,7(\bmod 8).$$

In this case $\kappa(V_1)$ or $\kappa(V_2)$ is zero. $\square$

Corollary 2 *Let V^n, $n=8m+r(r=1, 2)$ be a spin manifold. The filtration of $\beta_q[V^n]$ is exactly $4m$ if and only if $\kappa(V^n)\neq 0$.*

Proof. This follows from Theorem 6 and the description of $M^\Gamma(KO_{8m+r})$ in Sect. 3. □

7 The *SU*-case

Theorem 3 describes the subring $M_*=\beta_q(\Omega_*^{\text{spin}})\subset M^\Gamma(KO_*)$. Using the results of [6] one can easily determine the image of the special unitary cobordism ring Ω_*^{SU} under β_q. We will focus on the dimensions $8m+1$, $8m+2$, leaving the easier remaining cases to the reader.

Theorem 7 (i) *If $n=8m+1$, then $\beta_q(\Omega_n^{SU})\subset\beta_q(\Omega_n^{\text{spin}})$ is the subgroup of forms of the form $\eta P(\varepsilon^2)$ where P is a polynomial of degree $\leqq m/2$ over $\mathbb{F}_2$.*
(ii) *If $n=8m+2$, then $\beta_q(\Omega_n^{SU})=\beta_q(\Omega_n^{\text{spin}})$.*

Corollary. *If M^n, $n=8m+1$, is an SU-manifold, then*

$$a_i[M^n]=0$$

for all odd *i. For instance,*

$$\pi_1[M^n]=0,$$
$$(\pi_3+\pi_1^2)[M^n]=0.$$

Proof. (i) According to [6], an element from $\varphi_q(\Omega_{8m}^{SU})$ can be written as

$$2P(\delta_0^2,\varepsilon)+Q(\delta_0^2,\varepsilon^2),$$

where P, Q are two polynomials with integer coefficients. On the other hand, one has

$$\Omega_{8m+1}^{SU}=[\bar{S}^1]\cdot\Omega_{8m}^{SU}$$

where $\bar{S}^1$ is the circle S^1 equipped with its non-trivial SU-structure (cf. [21, Chap. X]). Therefore,

$$\beta_q(\Omega_{8m+1}^{SU})=\eta\cdot\beta_q(\Omega_{8m}^{SU})$$

and the result follows.

Part (ii) is an immediate consequence of the following proposition.

Proposition 5 *The canonical map*

$$\Omega_{8m+2}^{SU}\to\Omega_{8m+2}^{\text{spin}}/I_{8m+2}$$

is onto. In other words, any spin manifold of dimension $8m+2$ has the same KO-characteristic numbers as an SU-manifold.

Proof. Notice first that the homomorphism T used in the proof of Theorem 4 can be defined using SU-manifolds: there is a homomorphism

$$T^c\colon\ \Omega_{8m+2}^{SU}\to\Omega_{8m+4}^{SU}\otimes\mathbb{F}_2$$

with preserves the mod 2 KO-characteristic numbers. Let $I^c_* \subset \Omega^{SU}_*$ be the ideal of classes with vanishing KO-characteristic numbers. Then T^c induces a homomorphism

$$\tilde{T}^c\colon\ \Omega^{SU}_{8m+2}/I^c_{8m+2} \to (\Omega^{SU}_{8m+4}/I^c_{8m+4}) \otimes \mathbb{F}_2,$$

and there is a commutative diagram

$$\begin{array}{ccc} \Omega^{SU}_{8m+2}/I^c_{8m+2} & \xrightarrow{\tilde{T}^c} & (\Omega^{SU}_{8m+4}/I^c_{8m+4}) \otimes \mathbb{F}_2 \\ \downarrow{\scriptstyle \lambda} & & \downarrow{\scriptstyle \mu} \\ \Omega^{\text{spin}}_{8m+2}/I_{8m+2} & \xrightarrow{\tilde{T}} & (\Omega^{\text{spin}}_{8m+4}/I_{8m+4}) \otimes \mathbb{F}_2 \end{array}$$

in which λ and μ are induced by the forgetful homomorphism. One has to show that λ is onto. It is well known (cf. [19]) that

$$\Omega^{SU}_{8m+4} \to \Omega^{\text{spin}}_{8m+4}/\text{Tors}$$

is onto. As $I_{8m+4} = \text{Tors}\, \Omega^{\text{spin}}_{8m+4}$, this implies that μ is onto. Thus to prove the proposition, it will suffice to show that $\tilde{T}^c$ is onto.

Let $B_* \subset \Omega^{SO}_*/\text{Tors}$ be the subring of classes represented by U-manifolds with spherical determinant. According to Stong ([21, p. 282]), B_* is a polynomial algebra and $\Omega^{SU}_{8m+4}/I^c_{8m+4} \subset B_{8m+4}$ is exactly the subgroup $2B_{8m+4}$.

Let M^{8m+4} be an SU-manifold, and let W^{8m+4} be a U-manifold with spherical determinant such that $[M] = 2[W]$ in B_{8m+4}. Dualizing the determinant of W gives an SU-manifold V^{8m+2} and we have

$$W = U \cup (-D^2 \times V)$$

where U is an SU-manifold with boundary $\bar{S}^1 \times V$, namely the complement of a tubular neighbourhood of V in W (cf. [13]).

By definition, $T^c([V])$ is represented by the manifold $Z = (2U) \cup (-M^2 \times V)$, where M^2 is an SU-manifold such that $\partial M^2 = 2\bar{S}^1$. It is easy to see that Z is cobordant to $2W$ as a U-manifold. Therefore Z and $2W$ have the same rational Pontrjagin numbers. Hence Z and M have the same KO-characteristic numbers, that is represent the same element in $\Omega^{SU}_{8m+4}/I^c_{8m+4}$. □

8 Final remarks

(1) According to Theorem 6, the reduction mod 2 of the class a_m measures the KO-part of the Brown-Kervaire invariant in dimension $8m+2$. For instance,

$$\begin{aligned} k(V^{10}) &= \pi_1[V^{10}] \\ k(V^{18}) &= (\pi_2 + \pi_1)[V^{18}] \\ k(V^{26}) &= (\pi_3 + \pi_1^2)[V^{26}]. \end{aligned}$$

Other sequences $a_0, a_1, \ldots$ having the same property have been constructed in [13]. For example,

$$a_m = L_{2m}(\pi_1, \ldots, \pi_{2m}) + (\pi_1^3 + \pi_1 \pi_2 + \pi_3) L_{2m-2}(\pi_1, \ldots, \pi_{2m-2}),$$

where L_{2m} is the reduced mod 2 Hirzebruch's polynomial, is such a sequence. A simple comparison of the first few terms shows that the new classes a_m have far fewer terms. Besides, they have better multiplicative properties. The classes a_m have been used in [17] to represent $k(V)$ as the index of a twisted Dirac operator on V.

Notice that the mod 2 reduction of $h(q)$ is of the form $q + o(q^8)$. Therefore on has

$$a_m \equiv b_m (\mathrm{mod}\ 2)$$

for $m \leqq 8$. Thus in dimensions $n \leqq 71$, $\kappa(V)$ is measured by the Witten class $b_{[n/8]}$

(2) The genus

$$\varphi\colon \Omega_*^{SO} \to M^{\Gamma}(\mathbb{Z}[1/2])$$

was used by Landweber, Ravenel and Stong [12] to construct an elliptic (co)homology theory $E\ell\ell_*$([10, 11]). Namely they showed that

$$E\ell\ell_*(\,) = \Omega_*^{SO}(\,) \otimes_{\varphi} M^{\Gamma}(\mathbb{Z}[1/2])[\varepsilon^{-1}]$$

is a homology theory. Here $M^{\Gamma}(\mathbb{Z}[1/2])$ is considered as an Ω_*^{SO}-module via φ.

By analogy with the Conner-Floyd isomorphism ([7])

$$KO_*(\,) \cong \Omega_*^{Sp}(\,) \otimes KO_*$$

one can ask whether the functor

$$\Omega_*^{Sp}(\,) \otimes_{\beta_q} M_*[\varepsilon^{-1}],$$

where $M_* \subset M^{\Gamma}(KO_*)$ is the image of β_q described in Theorem 3 (iii), is a homology theory. A positive answer to this question would provide a way of eliminating the undesirable 1/2 in the definition of $E\ell\ell_*(\,)$.

Acknowledgements. I would like to thank W. Hoyt, N. Katz and D. Rohrlich for helpful advice on modular forms. I am grateful to F. Hirzebruch and R. Stong for their interest in my work. I am particularly indebted to M. Bendersky, M. Kreck, P. Landweber, H. Miller and S. Stolz for their useful comments.

References

1. Atiyah, M.: Riemann Surfaces and Spin Structures. Ann. Scient. Éc. Norm. Super., IV. Ser. **4**, 47–62 (1971)
2. Bendersky, M.: Applications of the Ochanine Genus. Math. Z. **206**, (in press)
3. Bott, R., Taubes, C.: On the Rigidity Theorems of Witten. J. AMS, **2**, 137–186 (1989)
4. Brown, E.H.: Generalizations of the Kervaire Invariant. Ann. Math., II. Ser. **95**, 368–383 (1972)
5. Chudnovsky, D.V., Chudnovsky, G.V.: Elliptic Modular Functions and Elliptic Genera. Topology **27**, 163–170 (1988)

6. Chudnovsky, D.V., Chudnovsky, G.V., Landweber, P.S., Ochanine, S., Stong, R.E.: Integrality and Divisibility of Elliptic Genera (to appear)
7. Conner, P.E., Floyd, E.E.: The Relation of Cobordism to K-theories. (Lect. Notes Math., vol. 28) Berlin Heidelberg New York: Springer 1966
8. Hirzebruch, F.: Mannigfaltigkeiten und Modulformen. University of Bonn. (Lect. Notes by Th. Berger and R. Jung) 1988
9. Landweber, P.S., Stong, R.E.: Circle Actions on Spin Manifolds and Characteristic Numbers. Topology **27**, 145–162 (1988)
10. Landweber, P.S.: Elliptic Cohomology and Modular Forms. In: Landweber, P.S. (ed.) Elliptic Curves and Modular Forms in Topology. Proceedings, Princeton 1986. (Lect. Notes Math., vol. 1326, pp. 55–68) Berlin Heidelberg New York: Springer 1988
11. Landweber, P.S.: Supersingular Elliptic Curves and Congruences for Legendre Polynomials. In: Landweber, P.S. (ed.) Elliptic Curves and Modular Forms in Topology. Proceedings, Princeton 1986. (Lect. Notes Math., vol. 1326, pp. 69–93) Berlin Heidelberg New York: Springer 1988
12. Landweber, P.S., Ravenel, D.C., Stong, R.E.: Periodic cohomology theories defined by elliptic curves (to appear)
13. Ochanine, S.: Signature modulo 16, invariants de Kervaire généralisés et nombres caractéristiques dans la K-théorie réelle. Bull. SMF, Mémoire **5** (1981)
14. Ochanine, S.: Sur les genres multiplicatifs définis par des intégrales elliptiques. Topology **26**, 143–151 (1987)
15. Ochanine, S.: Genres elliptiques équivariants. In: Landweber, P.S. (ed.) Elliptic Curves and Modular Forms in Topology. Proceedings, Princeton 1986. (Lect. Notes Math., vol. 1326, pp. 107–122) Berlin Heidelberg New York: Springer 1988
16. Ochanine, S.: A vanishing theorem for the elliptic genus (preprint)
17. Rubinsztein, R.L.: An Analytic Formula for the Kervaire Invariant of some $(8n+2)$-dimensional Spin-manifolds. Uppsala 1988 (preprint)
18. Serre, J.-P.: Formes modulaires et fonctions zêta p-adiques. In: Kwyck, W., Serre, J.-P. (eds.) Modular Functions of One Variable III. Proceedings, Antwerp 1972. (Lect. Notes Math., vol. 350, pp. 191–268) Berlin Heidelberg New York: Springer 1973
19. Stong, R.E.: Relations among characteristic numbers, II. Topology **5**, 133–148 (1966)
20. Stong, R.E.: On Complex-Spin Manifolds. Ann. Math., II. Ser. **85**, 526–536 (1967)
21. Stong, R.E.: Notes on Cobordism Theory. Princeton: Princeton University Press 1968
22. Witten, E.: The Index of the Dirac Operator in Loop Space. In: Landweber, P.S. (ed.) Elliptic Curves and Modular Forms in Topology. Proceedings, Princeton, 1986. (Lect. Notes Math., vol. 1326, pp. 161–181) Berlin Heidelberg New York: Springer 1988
23. Zagier, D.: Note on the Landweber-Stong Elliptic Genus. In: Landweber, P.S. (ed.) Elliptic Curves and Modular Forms in Topology. Proceedings, Princeton 1986. (Lect. Notes Math., vol. 1326, p. 216–224) Berlin Heidelberg New York: Springer 1988

Note added in proof

An elliptic homology theory with coefficient ring M_* has been recently defined by M. Kreck and S. Stolz.

Math. Z. 206, 293–302 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

The trace formula for equivariant $\mathfrak{D}$-modules and perverse sheaves

Roy Joshua*

Department of Mathematics, The Johns Hopkins University, Baltimore, MD 21218, USA

Received March 23, 1989; in final form January 1, 1990

In this paper we apply the equivariant Riemann-Roch theorem (see [J-2]) to provide a trace formula for coherent algebraic $\mathfrak{D}$-modules on a smooth projective complex algebraic variety that are equivariant (in the sense of [J-2] (1.6)) for the action of a complex diagonalisable group. Let X denote such a variety acted on by the complex diagonalisable group D. We begin by providing a general trace formula for all equivariant coherent $\mathfrak{D}_X$-modules. The trace formula takes on a particularly interesting form in the case of holonomic $\mathfrak{D}_X$-modules; the Riemann-Hilbert correspondence then provides a trace formula for all equivariant perverse sheaves. In particular, if $Z \subseteq X$ is an invariant, possibly singular, closed subvariety of X and $D = T$ is a complex torus, we obtain a trace formula relating the intersection cohomology of Z with an appropriate hypercohomology of the fixed point scheme Z^T generalising the well known trace formula (see [Iver]) in the case Z is also smooth. In the forthcoming paper [J-3], we combine the trace formula along with a generalization of the above Riemann-Roch theorem to provide very general character formulae for equivariant $\mathfrak{D}$-modules and perverse sheaves.

(1.0) We adopt the terminology of ([J-2]). Let D denote a complex diagonalisable group acting on a smooth complex quasi-projective algebraic variety X. (Unless explicitly mentioned to the contrary, all varieties considered are assumed to be *algebraic*.) Let M denote the abelian group of characters of D and let $R(D) = \mathbb{Z}[M]$ denote the corresponding ring of characters. (Recall that M is finitely generated; it is a free abelian group if D is a torus and is isomorphic to $\mathbb{Z}/l^\nu$ if $D =$ the l^ν-th roots of unity.) For each $\chi \in M$, we let e^χ denote the corresponding element of $R(D)$. Let $S' \subseteq R(D)$ denote the multiplicative subset generated by the elements of the form $1 - e^\chi$, χ a nontrivial character of D and let p be a prime ideal in $\mathbb{Z}[M]$ so that $S' \subseteq Z[M] - p$. Let $S = \mathbb{Z}[M] - p$. This will be fixed throughout the following discussion.

* *Current address*: Department of Mathematics, The Ohio State University, Columbia, OH 431210, USA

Let $K_D^0(X)$ denote the Grothendieck group of D-equivariant locally free coherent $\mathcal{O}_X$-modules. If the action is trivial one obtains the isomorphism

$$K_D^0(X) \cong K^0(X) \underset{\mathbb{Z}}{\otimes} R(D)\,. \tag{1.0.1}$$

(See [SGA 3] I, Expose 1, 4.4.) In general $K_D^0(X)$ is an algebra over $R(D)$.

(1.1) Let $\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X)(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_X))$ denote the abelian category of coherent equivariant $\mathcal{O}_X$-modules as in ([J-2] (1.1)) (left $\mathfrak{D}_X$-modules as in ([J-2] (1.6)), respectively). Now observe that the tensor product makes $K_0(\mathrm{Mod}_{\mathrm{coh}}^D(O_X))$ a $K_D^0(X)$-module and hence an $R(D)$-module. Next observe that if M is a D-equivariant coherent left $\mathfrak{D}_X$-module and F is a D-equivariant locally free $\mathcal{O}_X$-module, then $M \otimes_{\mathcal{O}_X} F$ is also a coherent D-equivariant $\mathfrak{D}_X$-module. (Here we let $\mathfrak{D}_X$ act on the left on M and the group D diagonally on $M \otimes_{\mathcal{O}_X} F$.) It follows that $K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_X))$ is also a $K_D^0(X)$-module and hence an $R(D)$-module.

Let $D_b^+(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X))(D_b^+(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_X)))$ denote the derived category of bounded-below complexes in $\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X)(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_X)$, respectively) with bounded cohomology. Let $\mathrm{Mod}_{q\mathrm{coh}}^D(\mathcal{O}_X)(\mathrm{Mod}_{q\mathrm{coh}}^D(\mathfrak{D}_X))$ denote the corresponding category of quasi-coherent modules. (See [J-2] (1.6.1).) Let $D_{b,\mathrm{coh}}^+(\mathrm{Mod}_{q\mathrm{coh}}^D(\mathcal{O}_X))$ $(D_{b,\mathrm{coh}}^+(\mathrm{Mod}_{q\mathrm{coh}}^D(\mathfrak{D}_X)))$ denote the derived category of complexes in $\mathrm{Mod}_{q\mathrm{coh}}^D(\mathcal{O}_X))$, $(\mathrm{Mod}_{q\mathrm{coh}}^D(\mathfrak{D}_X)$, respectively) that are bounded below and have bounded coherent cohomology.

(1.2) Next let $f: X \to Y$ be a D-equivariant *proper* map of smooth quasi-projective complex varieties X and Y provided with an action by D, which we assume is linear. (Observe from [Su], p. 8 that if the group D is connected there is no loss of generality in this assumption.) Recall that $Rf_*(\int_f)$ induces maps

$$Rf_*: D_b^+(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X)) \to D_{b,\mathrm{coh}}^+(\mathrm{Mod}_{q\mathrm{coh}}^D(\mathcal{O}_Y))$$

$$\textstyle\int_f: D_b^+(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_X)) \to D_{b,\mathrm{coh}}^+(\mathrm{Mod}_{q\mathrm{coh}}^D(\mathfrak{D}_Y))$$

(See [J-2], (2.9)). The induced maps (which we denote by the same symbols)

$$Rf_*: K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X)) \to K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_Y)) \quad \text{and}$$

$$\textstyle\int_f: K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_X)) \to K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_Y))$$

are seen to be $R(D)$-linear by the projection formula. We also observe (from the definition in [J-2], (2.6)) that the map $d_X: K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathfrak{D}_X)) \to K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X))$ is also $R(D)$-linear. Similarly its inverse (defined in [J-2], (2.7)) is also seen to be $R(D)$-linear.

(1.3) From now on we assume X is a *projective* smooth variety acted on by a diagonalisable group D as in (1.0). Let X^D denote the fixed point scheme; observe that this is a disjoint union of its irreducible components each of which is smooth. (See [Iver], (1.3) and [Sh], p. 87.) Let N denote the conormal sheaf associated to the closed immersion $i: X^D \to X$. Let $i_*: K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_{X^D})) \to K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X))$ denote the Gysin map; let $\lambda_{-1}(N)$ = the class given by $\mathcal{O}_{X^D} - N + \lambda^2 N - \lambda^3 N + \ldots$ in $K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_{X^D}))$. Then one observes the following:

(i) As X, X^D are smooth we obtain the isomorphisms (see [T-1], p. 560)

$$K_D^0(X) \cong K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_X)), \quad K_D^0(X^D) \cong K_0(\mathrm{Mod}_{\mathrm{coh}}^D(\mathcal{O}_{X^D}))\,.$$

(ii) $\lambda_{-1}(N)$ is a unit in the localised ring $S^{-1}K_D^0(X^D)$ and i_* induces an isomorphism $S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_{X^D})) \cong S^{-1}K_D^0(X^D) \xrightarrow{\cong} S^{-1}K_D^0(X) \cong S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X))$. Its inverse is the map $i^*(\ ) \cap \lambda_{-1}(N)^{-1}$. (It is shown in ([Niel], p. 92) that (ii) is true if we replace S by the multiplicatively closed set S' as in (1.0). As $S' \subseteq S$, (ii) follows.)

(iii) Let $U = X - X^D$. Now observe that the maps in the localisation sequence

$$K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_{X^D})) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)) \to K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_U)) \to 0$$

(see [T-1], p. 548) are $R(D)$-linear. Hence it follows that

$$S^{-1}(K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_U)) \cong 0\ .$$

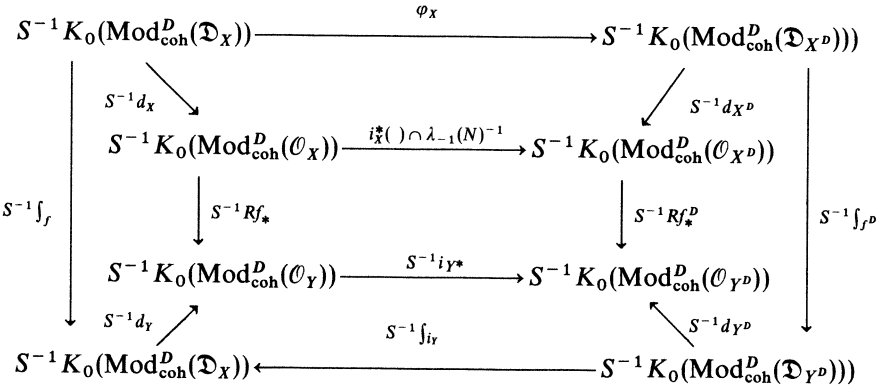
The Riemann-Roch-theorem for weakly equivariant $\mathfrak{D}$-modules (see [J-2], (2.10)) and the above discussion shows that the diagram

$$\begin{array}{ccc} S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) & \xleftarrow[S^{-1}\int_i]{} & S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_{X^D})) \\ \Big\downarrow{\scriptstyle S^{-1}d_X} & & \Big\downarrow{\scriptstyle S^{-1}d_{X^D}} \\ S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_X)) & \xleftarrow[S^{-1}i_*]{} & S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathcal{O}_{X^D})) \end{array} \tag{1.4}$$

is commutative. Since all the other three maps in (1.4) are isomorphisms so is the map in the top row. Thus an inverse to $S^{-1}\int_i$ in the top row is given by the composition of $S^{-1}d_X$, $i^*(\ ) \cap \lambda_{-1}(N)^{-1}$ and the inverse to $S^{-1}d_{X^D}$ given as in ([J-2], (2.6).)

(1.5) We will denote this map by φ_X from now on.

(1.6) Proposition. *(The general trace formula.) Assume in addition to the hypotheses of* (1.2) *that X and Y are projective. Let X^D, Y^D denote the fixed-point schemes. Let $i_X(i_Y, N)$ denote the closed immersion $X^D \to X$ ($Y^D \to Y$, the normal bundle associated i_X, respectively). Let $f^D: X^D \to Y^D$ be the map induced by f. Then we obtain the commutative diagram (where φ_X is the map in* (1.5))



Proof. It is clear in view of (1.4) and ([J-2], (2.10)).

2. (2.0) Assume the situation of (1.0). Let $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$ denote the full subcategory of the abelian category $\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)$ consisting of holonomic $\mathfrak{D}_X$-modules. Observe

(see [Bo], p. 216) that this is a Serre (or equivalently dense—see [Pop], p. 165) subcategory; hence $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$ is also abelian.

(2.1) Let $j: U \to X$ be the open immersion of a D-invariant open subvariety of X and let $i: Z \to X$ denote the closed immersion of its complement. Let $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)(\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X))$ denote the full subcategory of $\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_U)$ consisting of holonomic $\mathfrak{D}_U$-modules (the full subcategory of $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$ consisting of $\mathfrak{D}_X$-modules with support contained in Z, respectively). Observe that $\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X)$ is a Serre subcategory of the abelian category $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$. Let $j^*: \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X) \to \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)$ denote the obvious restriction functor.

(2.2) Lemma. *Assume the above situation. Then j^* induces an equivalence of categories*

$$j^*: \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)/\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X) \xrightarrow{\sim} \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)$$

Proof. We first show that j^* is surjective onto the objects of the category $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)$. Let $N \in \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)$. Then $j_*(N)$ is a holonomic D-equivariant $\mathfrak{D}_X$-module and $j^*j_*(N) \cong N$. ([J-2], (1.8.2) (iii)) shows $j_*(N)$ is an equivariant $\mathfrak{D}_X$-module; [Bo] p. 292 shows it is holonomic.) If $f: N \to N'$ is a map in $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)$, then $j_*(f): j_*N \to j_*N'$ is a map in $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$ and $j^*j_*(f) \cong f$. This shows that j^* is a full functor. Now it suffices to show j^* is also faithful. Let $f: M \to M'$ be a map in $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$ so that $j^*(f)$ is the zero-map in $\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)$. This can happen if and only if the image of f, $\mathrm{Im}(f)$, has supports contained in Z; observe that $\mathrm{Im}(f)$ is also holonomic since the category of holonomic $\mathfrak{D}$-modules is abelian. One observes readily that it is also equivariant and hence is an object of $\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X)$. Hence the class of f in the quotient category on the left in (2.2) is also zero. The lemma follows.

(2.3) *Localisation sequence.* Assume the situation of (2.1). Then we obtain a long exact sequence

$$\ldots \to K_i(\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X)) \to K_i(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)) \to K_i(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)) \to \ldots$$
$$\to K_0(\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X)) \to K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)) \to K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)) \to 0$$

where K_i denotes the i-th algebraic K-group of the corresponding category.

Proof. It is clear in view of (2.2) and ([Qu], Sect. 5, Theor. 5).

(2.4) Assume in addition to the hypotheses of (2.1) that Z is also a smooth scheme. (Observe that now Z is the disjoint union of its irreducible components each of which is also smooth. (See [Sh], p. 87.) We will from now on always assume that each such component is also invariant under the group action. The category $\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_Z)(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_Z))$ of D-equivariant coherent $\mathfrak{D}_Z$-modules (holonomic $\mathfrak{D}_Z$-modules, respectively) may now be defined in the obvious manner.

One observes as before that these are abelian categories and that the functor

$$\int_i: \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_Z) \to \mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X) \tag{2.4.1}$$

induces an equivalence of categories with the inverse given by the functor $M \to \underline{H}^0(i^!(M))$, $M \in \mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X)$. (See [J-2], (2.9′) and ((2.9)) for the definition of $i^!(\int_i$, respectively) in the equivariant context.) It follows that $\int_i$ induces an isomorphism $K_n(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_Z)) \xrightarrow{\sim} K_n(\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X))$ for all n. Let $\int_i$ also denote the composition of the map in (2.4.1) with the obvious inclusion $\mathrm{Mod}^D_{Z,\mathrm{hol}}(\mathfrak{D}_X) \to \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$.

(2.5) Lemma. *Under the above assumptions the induced map* $\int_i : K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_Z)) \to K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X))$ *is a split injection with a left-inverse provided by the map* $M \to \sum_n (-1)^n \underline{H}^n(i^! M)$, $M \in \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$.

Proof. Observe that if M has support contained in Z, then $\underline{H}^n(i^! M) = 0$ for all $n \neq 0$. Moreover $\underline{H}^0(i^! \int_i N) \cong N$, $N \in \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_Z)$. (See [Bo], p. 263 and p. 265.) The lemma follows.

(2.6) Let D denote a complex diagonalisable group acting on a smooth complex projective variety X. Let $\mathrm{Mod}^D_{\mathrm{locfree}}(\mathcal{O}_X)$ ($K^0_D(X)$) denote the category (the Grothendieck group, respectively) of D-equivariant locally free coherent $\mathcal{O}_X$-modules. If M is a holonomic left-$\mathfrak{D}_X$-module and $F \in \mathrm{Mod}^D_{\mathrm{locfree}}(\mathcal{O}_X)$ we observe that

(i) $M \otimes_{\mathcal{O}_X} F$ is a coherent left-$\mathfrak{D}_X$-module with the left-$\mathfrak{D}_X$-module structure induced from that of M

(ii) $M \otimes_{\mathcal{O}_X} F$ is holonomic. (To see this observe that the characteristic variety of $M \otimes_{\mathcal{O}_X} F$ is the same that of M.)

It follows that $K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X))$ is a module over the ring $K^0_D(X)$.

(2.7) Assume in addition to the hypotheses of (2.6) that $i : X^D \to X$ is the closed immersion of the fixed point scheme and $U = X - X^D$. Let $S \subseteq R(D)$ be as in (1.0). One observes from ((1.3 (iii)) that

$$S^{-1} K^0_D(U) \cong 0 . \tag{2.7.1}$$

As $K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U))$ is a module over $K^0_D(U)$, it follows that

$$S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)) \cong 0 . \tag{2.7.2}$$

(2.8) Proposition. *Assume the hypotheses of* (2.7). *Now* $\int_i$ *induces an isomorphism*

$$S^{-1} \int_i : S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)) \xrightarrow{\cong} S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X))$$

Proof. First observe (using the projection formula as in (1.4)) that $\int_i$ is $R(D)$-linear. One may readily verify that the maps $M \to j^*(M)$ (where j is $X - X^D \to X$) and $M \to \sum_n (-1)^n \underline{H}^n(i^! M)$, $M \in \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$ are $R(D)$-linear. Recall that $\int_i$ is a split injection (see (2.5)) and $S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_U)) \cong 0$ (see (2.7.2)). Now the localisation sequence (2.3) localised at S provides the required isomorphism.

(2.9) Proposition. *Assume the hypotheses of* (2.8). *Let*

$$\psi : S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)) \to S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_{X^D}))$$

$$(\varphi : S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) \to S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_{X^D})))$$

denote the map in lemma (2.5) *given by* $M \to \sum_n (-1)^n \underline{H}^n(i^! M)$, $M \in \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$ (*the map in* (1.5), *respectively.*) *Then the diagram*

$$\begin{array}{ccc} S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)) & \xrightarrow{\psi} & S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_{X^D})) \\ \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} \\ S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) & \xrightarrow{\varphi} & S^{-1} K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_{X^D})) \end{array} \tag{2.9.1}$$

commutes where the vertical maps are the obvious ones.

Proof. Let $\overline{\int}_i : S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_{X^D})) \to S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X))$ denote the map in (2.8) and let $\widetilde{\int}_i : S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_{X^D})) \to S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X))$ denote the map as in (1.4) for the time being. The diagram

$$\begin{array}{ccc} S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)) & \xleftarrow{\overline{\int}_i} & S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_{X^D})) \\ \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} \\ S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X)) & \xleftarrow{\widetilde{\int}_i} & S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_{X^D})) \end{array} \tag{2.9.2}$$

clearly commutes where the vertical maps α, β are the obvious ones. The commutativity of (2.9.2) is equivalent to the statement $\widetilde{\int}_i \circ \beta = \alpha \circ \overline{\int}_i$. Recall φ is an inverse to the map $\widetilde{\int}_i$; hence $\beta = \varphi \circ \widetilde{\int}_i \circ \beta = \varphi \circ \alpha \circ \overline{\int}_i$. Observe from (2.5) that $\psi \circ \overline{\int}_i =$ the identity map of $S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X))$. It follows that $\beta = \beta \circ (\psi \circ \overline{\int}_i) = (\beta \circ \psi) \circ \overline{\int}_i$. Thus $(\varphi \circ \alpha) \circ \overline{\int}_i = (\beta \circ \psi) \circ \overline{\int}_i$. As $\overline{\int}_i$ is a bijection, it follows that $\varphi \circ \alpha = \beta \circ \psi$. This proves (2.9).

(2.10) Assume the hypotheses of (2.8). Let $M \in \mathrm{Mod}^D_{\mathrm{hol}}(\mathfrak{D}_X)$. Now $i^!M$ is a bounded D-equivariant complex of $\mathfrak{D}_X$-modules with holonomic cohomology. The class $i^!M$ in $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X))$ is clearly represented by $\Sigma(-1)^n \underline{H}^n(i^!M)$. It follows therefore from (2.9) that $i^!M = \varphi(M)$ as classes in $S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathfrak{D}_X))$ where φ is as above.

(2.11) Theorem. (*Trace formula for equivariant holonomic $\mathfrak{D}_X$-modules.*) *Assume in addition to the hypotheses of* (1.6) *that M is a D-equivariant holonomic $\mathfrak{D}_X$-module. Then $d_Y(\int_f M)$, $Rf_*d_X(M)$ $Rf_*((i_X)_*d_{X^D}(i_X^!M))$, $(i_Y)_*Rf^D{}_*d_{X^D}(i_X^!M)$, $(i_Y)_*d_Y\int_{f^D}(i_X^!M)$ and $d_Y(\int_i \circ \int_{f^D}(i_X^!M))$ all represent the same class in $S^{-1}K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathscr{O}_Y))$.*

If $D = T$ is a torus acting trivially on Y, equality of these classes holds in the Grothendieck group $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathscr{O}_Y)) \cong K^0_D(Y)$.

Proof. The first statement follows readily from (1.6) using (2.10). Next assume $D = T$ is a torus acting trivially on Y. It follows now from (1.0.1) and ((1.3) (i)) that $K_0(\mathrm{Mod}^D_{\mathrm{coh}}(\mathscr{O}_Y)) \cong K^0_D(Y) \cong K_0(Y) \otimes_{\mathbb{Z}} R(D)$. Now the fact T is a torus implies that every element of S is a non-zero divisor in $K^0_D(Y)$. Therefore the equality of the above classes in $S^{-1}K^0_D(Y)$ implies their equality in the Grothendieck group $K^0_D(Y)$.

(2.12) Let Z denote a smooth *projective* complex variety acted on by a complex torus T and let $f: Z \to \mathrm{Spec}\,\mathbb{C}$ denote the structure map of Y. Let Z_{an} denote the corresponding analytic space with structure sheaf $\mathscr{O}_{Z_{\mathrm{an}}}$. Observe that there exists a map ε: $(Z_{\mathrm{an}}, \mathscr{O}_{Z_{\mathrm{an}}}) \to (Z, \mathscr{O}_Z)$ of ringed spaces. If M is an $\mathscr{O}_Z$-module, we let M_{an} denote the $\mathscr{O}_{Z_{\mathrm{an}}}$-module $\varepsilon^*(M) = (\mathscr{O}_{Z_{\mathrm{an}}}) \otimes_{\varepsilon^{-1}(\mathscr{O}_Z)} \varepsilon^{-1}(M)$. (Observe that ε^* is an exact functor as $\varepsilon^{-1}(\mathscr{O}_Z) \to \mathscr{O}_{Z_{\mathrm{an}}}$ is flat. Observe also that ε induces a map of the ringed spaces $(Z_{\mathrm{an}}, \mathfrak{D}_{Z_{\mathrm{an}}}) \to (Z, \mathfrak{D}_Z)$ where $\mathfrak{D}_{Z_{\mathrm{an}}} = \varepsilon^*(D_Z)$.) We proceed to define two additive homomorphisms

$$D^+_{b,\mathrm{coh}}(\mathrm{Mod}^T_{q\mathrm{coh}}(\mathfrak{D}_Z)) \to D^+_{b,\mathrm{coh}}(\mathrm{Mod}_{q\mathrm{coh}}(\mathfrak{D}_Z)) \to \mathbb{Z}\,.$$

If χ denotes the Euler characteristic the first (second) map is given by $M^\cdot \to \chi(\int_f M^\cdot) = \chi(\mathrm{Rhom}_{\mathfrak{D}_Z}(\mathscr{O}_Z, M^\cdot)$ ($M^\cdot \to \chi(R\Gamma(Z_{\mathrm{an}}, DR(M^\cdot)) = \chi(\mathrm{Rhom}_{\mathfrak{D}_{Z_{\mathrm{an}}}}(\mathscr{O}_{Z_{\mathrm{an}}}, M^\cdot_{\mathrm{an}})$, respectively).

(2.13) Lemma. *Assume the situation of* (2.12). *If* χ *denotes the Euler characteristic, and* $M^{\cdot} \in D^{+}_{b,\mathrm{coh}}(\mathrm{Mod}^{T}_{q\mathrm{coh}}(\mathfrak{D}_Z))$ *then* $\chi(\int_f M^{\cdot}) = \chi(R\Gamma(Z_{\mathrm{an}}, DR(M^{\cdot}))$.

Proof. Observe the existence of strongly convergent spectral sequences

$$E_2^{s,t} = R^s \mathrm{hom}_{\mathfrak{D}_Z}(\mathcal{O}_Z, \underline{H}^t(M^{\cdot})) \Rightarrow R^{s+t}\mathrm{hom}_{\mathfrak{D}_Z}(\mathcal{O}_Z, M^{\cdot}) \quad \text{and}$$

$$E_2^{s,t} = R^s \mathrm{hom}_{\mathfrak{D}_{Z_{\mathrm{an}}}}(\mathcal{O}_{Z_{\mathrm{an}}}, \underline{H}^t(M^{\cdot}_{\mathrm{an}})) \Rightarrow R^{s+t}\mathrm{hom}_{\mathfrak{D}_{Z_{\mathrm{an}}}}(\mathcal{O}_{Z_{\mathrm{an}}}, M^{\cdot}_{\mathrm{an}}))$$

Therefore it suffices to prove the lemma for a single module $M \in \mathrm{Mod}^{T}_{\mathrm{coh}}(\mathfrak{D}_Z)$. The equivariant first exact sequence of Spencer (see [J-2], (2.4)) shows every $M \in \mathrm{Mod}^{T}_{\mathrm{coh}}(\mathfrak{D}_Z)$ has a finite T-equivariant resolution by $\mathfrak{D}_Z$-modules of the form $\mathfrak{D}_Z \otimes F \otimes \omega_Z^{\vee}$, where F is a T-equivariant coherent $\mathcal{O}_Z$-module. Therefore it suffices to prove (2.13) only for $\mathfrak{D}_Z$-modules of the above form. Let F be a T-equivariant coherent $\mathcal{O}_Z$-module as above. Then observe that

$$\int_f (\mathfrak{D}_Z \otimes F \otimes \omega_Z^{\vee}) \cong R\Gamma(Z, F) \quad \text{and similarly}$$

$$R\Gamma(Z_{\mathrm{an}}, DR(\mathfrak{D}_Z \otimes F \otimes \omega_Z^{\vee})) \cong \int_{f_{\mathrm{an}}} (\mathfrak{D}_{Z_{\mathrm{an}}} \otimes F_{\mathrm{an}} \otimes \omega_{Z_{\mathrm{an}}}^{\vee}) \cong R\Gamma(Z_{\mathrm{an}}, F_{\mathrm{an}})$$

by the projection formula. (See for e.g. [J-2], (2.9.1).) As Z is projective, GAGA (see [GAGA]) provides a quasi-isomorphism of the above two complexes; hence they clearly have the same Euler-characteristic. This proves the lemma.

(2.14) Theorem. (*Trace formula for equivariant regular holonomic* $\mathfrak{D}$*-modules.*)
Let T *denote a complex torus acting on the projective smooth complex variety* X. *Let* $i: X^T \to X$ *denote the closed immersion of the fixed point scheme; let* $f: X \to \mathrm{Spec}\,\mathbb{C}$ ($f^T: X^T \to \mathrm{Spec}\,\mathbb{C}$) *denote the structure map of* X (X^T, *respectively.*) *Let* M *be a* T*-equivariant regular holonomic* $\mathfrak{D}_X$*-module. Then*

$$\chi(R\Gamma(X_{\mathrm{an}}, DR(M))) = \sum_n (-1)^n \dim H^n(X_{\mathrm{an}}, DR(M))$$

$$= \sum_n (-1)^n \dim H^n((X^T)_{\mathrm{an}}, DR(i^! M)) = \chi(R\Gamma(X^T_{\mathrm{an}}, DR(i^! M)))$$

Proof. It follows from (2.11) that $\int_f M$ and $\int_{fT}(i^! M)$ represent the same classes in the representation ring $R(T)$ and hence have the same Euler-characteristic. Now (2.13) applies to complete the proof.

(2.15) *Remark.* Observe using the Riemann-Hilbert correspondence (see [Bo], p. 326) that $DR(i^! M) \simeq Ri^!_{\mathrm{an}} DR(M)$ where $i^!_{\mathrm{an}}$ is the obvious functor defined for constructible sheaves on the analytic site.

3. In this section we apply the results of sections 1 and 2 to provide a trace formula for equivariant perverse sheaves.

(3.0) Assume the situation of (1.0). Let $\mu(pr_2): D \times X \to X$ denote the group action (projection to the second factor, respectively). Let $\sigma: X \to D \times X$ denote the map $x \to (e, x)$, e = the identity element of D; let $d_0(d_1, d_2) D \times D \times X \to D \times X$ denote the projection to the last two factors (the map $\eta \times \mathrm{id}$ where $\eta: D \times D \to D$ is the group multiplication, the map $\mathrm{id} \times \mu$, respectively). If α is any of the above structure maps, we will let α_{an} denote the induced map of analytic varieties. (We will always adopt the following convention: if $g: Y \to Z$ denotes a map of algebraic (or analytic) varieties with structure sheaves $\mathcal{O}_Y$ and $\mathcal{O}_Z$ respectively and F is an $\mathcal{O}_Z$-module, then $g^*(F) = \mathcal{O}_Y \otimes_{g^{-1}(\mathcal{O}_Z)} g^{-1}(F)$.

(3.1) Definition. Let M be a coherent $\mathfrak{D}_X$-module. Then M is a *strongly D-equivariant* $\mathfrak{D}_X$-module if there exists an isomorphism $\varphi: \mu^* M \to \mathrm{pr}_2^* M$ of $\mathfrak{D}_{(D \times X)} \cong \mathfrak{D}_D \otimes \mathfrak{D}_X$-modules so that (i) $\sigma^*(\varphi) =$ the identity and (ii) there is an obvious cocycle condition (see [J-2], (1.1)) between the 3 pull-backs $d_0^*(\varphi)$, $d_1^*(\varphi)$ and $d_2^*(\varphi)$.

(3.2) *Remark.* Observe that any strongly D-equivariant $\mathfrak{D}_X$-module is D-equivariant in the sense used in sections 1 and 2. (See [J-2], (1.6).) This follows from the fact that an isomorphism φ as in (3.1) is clearly an isomorphism of $\mathcal{O}_D \otimes \mathfrak{D}_X$-modules.

(3.3) Let X be a possibly-singular complex quasi-projective variety. A bounded complex of abelian sheaves $F^\cdot$ on X_{an} is constructible if there exists a filtration $X = X_n \supseteq X_{n-1} \supseteq X_{n-2} \supseteq \ldots \supseteq X_0 \supseteq X_{-1} = \phi$ by Zariski-closed subschemes X_i so that each $X_i - X_{i-1}$ is smooth and the cohomology sheaves $\underline{H}^i(F^\cdot)$ are locally constant on each $(X_i - X_{i-1})_{\mathrm{an}}$. A *perverse sheaf* on X_{an} is a bounded constructible complex $F^\cdot$ of abelian sheaves on X_{an} satisfying the following two conditions:

(a) the dimension of the support of $H^m(F^\cdot) \leqq -m$ for all integers m
(b) The Verdier dual of $F^\cdot$ also satisfies (a). (See [B-B-D], p. 102.)

(3.4) Definition. Let X be a complex, possibly-singular, quasi-projective variety acted on by a complex diagonalisable group D. A *D-equivariant perverse sheaf* $F^\cdot$ on X_{an} is a perverse sheaf on X_{an} so that there exists an isomorphism $\psi: \mu_{\mathrm{an}}^{-1} F^\cdot \to (\mathrm{pr}_2)_{\mathrm{an}}^{-1} F^\cdot$ in the category of perverse sheaves on $(D \times X)_{\mathrm{an}}$ satisfying the following conditions (in the appropriate category of perverse sheaves):

(i) $\sigma_{\mathrm{an}}^{-1}(\psi) =$ the identity and (ii) there is an obvious cocycle condition between the pull-backs $(d_0)_{\mathrm{an}}^{-1}(\psi)$, $(d_1)_{\mathrm{an}}^{-1}(\psi)$ and $(d_2)_{\mathrm{an}}^{-1}(\psi)$.

Assume the situation of (1.0). Let $\mathrm{Mod}_{r.h}^{s-D}(\mathfrak{D}_X)$, $\mathrm{Per}^D(X_{\mathrm{an}})$ denote the category of strongly D-equivariant regular holonomic $\mathfrak{D}_X$-modules (the category of D-equivariant perverse sheaves on X_{an}, respectively.)

(3.5) Proposition. *The De-Rham functor induces a surjection*

$$DR: \mathrm{Mod}_{r.h}^{s-D}(\mathfrak{D}_X) \to \mathrm{Per}^D(X_{\mathrm{an}})$$

where $\mathrm{Mod}_{r.h}^{s-D}(D_X)$ *is the category defined above.*

Proof. Let $F^\cdot \in \mathrm{Per}^D(X_{\mathrm{an}})$. The familiar Riemann-Hilbert correspondence shows that there exists a regular holonomic $\mathfrak{D}_X$-module M so that $DR(M) \simeq F^\cdot$. We next observe that $\mu_{\mathrm{an}}^{-1}(DR(M)) \simeq DR(\mu^*(M))$ and $(\mathrm{pr}_2)_{\mathrm{an}}^{-1}(DR(M)) \cong DR(\mathrm{pr}_2^*(M))$. Let $D_{r.h}^b(\mathrm{Mod}(\mathfrak{D}_{(D \times X)}))(D_c^b((D \times X)_{\mathrm{an}}))$ denote the derived category of bounded complexes of quasi-coherent $\mathfrak{D}_{(D \times X)}$-modules with regular holonomic cohomology (bounded complexes of abelian sheaves on $(D \times X)_{\mathrm{an}}$ with constructible cohomology, respectively.) As the De-Rham functor $DR: D_{r.h}^b(\mathrm{Mod}(\mathfrak{D}_{(D \times X)}) \to D_c^b((D \times X)_{\mathrm{an}})$ is an equivalence, it follows that the map ψ (as in (3.4)) comes from an isomorphism $\varphi: \mu^*(M) \to \mathrm{pr}_2^*(M)$ in the derived category $D_{r.h}^b(\mathrm{Mod}(\mathfrak{D}_{(D \times X)}))$. It follows that φ is an isomorphism of $\mathfrak{D}_{(D \times X)}$-modules. The cocycle condition on ψ implies a similar cocycle condition on φ; one may now readily verify that the isomorphism φ satisfies the necessary conditions in (3.1) for M to be a strongly D-equivariant $\mathfrak{D}_X$-module.

(3.6) Theorem. (*Trace formula for equivariant perverse sheaves.*) *Let X denote a smooth complex projective variety acted on by a complex torus T. Let $i: X^T \to X$*

denote the closed immersion of the fixed point scheme. Let $F^{\cdot}$ be a T-equivariant perverse sheaf on X_{an}. Then

$$\chi R\Gamma(X_{\mathrm{an}}, F^{\cdot}) = \sum_n (-1)^n \dim H^n(X_{\mathrm{an}}, F^{\cdot}) = \sum_n (-1)^n \dim H^n((X^T)_{\mathrm{an}}, Ri_{\mathrm{an}}^! F^{\cdot})$$
$$= \chi R\Gamma(X_{\mathrm{an}}, i_{\mathrm{an}*} Ri_{\mathrm{an}}^! F^{\cdot}) .$$

Proof follows readily from (3.5), (3.2), (2.14) and (2.15).

(3.7) Corollary. *Let Y denote a complex, possibly-singular, projective variety and let IC_Y denote its intersection cohomology complex with the middle perversity. Assume that a complex torus T acts on Y and let $i: Y^T \to Y$ denote the closed immersion of the fixed point scheme. Then*

$$\sum_n (-1)^n \dim IH^n(Y) = \sum_n (-1)^n \dim H^n((Y^T)_{\mathrm{an}}, Ri_{\mathrm{an}}^! IC_Y)$$

Proof. First of all we show that Y may be assumed to be normal without any loss of generality. Let $\pi: \tilde{Y} \to Y$ denote the normalisation of Y. Recall from ([J-1] (1)) the following facts: (i) the T-action may be lifted to $\tilde{Y}$ and $\pi^{-1}(Y^T) = (\tilde{Y})^T$ (ii) the intersection cohomology of Y (the pair $(Y, Y - Y^T)$) is isomorphic to the intersection cohomology of $\tilde{Y}$ (the pair $(\tilde{Y}, \tilde{Y} - \tilde{Y}^T)$, respectively.) Now a comparison of the long-exact sequences in intersection cohomology of the pairs $(Y, Y - Y^T)$ and $(\tilde{Y}, \tilde{Y} - \tilde{Y}^T)$ shows we may assume Y is normal.

Hence Y admits a T-equivariant closed immersion into a large projective space $\mathbb{P}^n$ onto which the T-action extends. (See [Su], Theor. 1.) Next let $i_0(i_1, i_2, i_3)$ denote the closed immersion $Y \to \mathbb{P}^n((\mathbb{P}^n)^T \to \mathbb{P}^n,\ Y^T \to (\mathbb{P}^n)^T,\ Y^T \to Y$, respectively) as well as the corresponding induced map of the associated analytic spaces. Next observe (see [J-1], p. 243) that the intersection cohomology complex IC_Y is a T-equivariant perverse sheaf on Y_{an}; hence $i_{0*}(IC_Y)$ is a T-equivariant perverse sheaf on $\mathbb{P}^n$. (To see this observe that the Verdier dual $D(i_{0*} IC_Y) \cong i_{0*} D(IC_Y)$. Now check the conditions (a) and (b) in (3.3).) We then obtain the pull-back square

$$\begin{array}{ccc} Y^T & \xrightarrow{i_2} & (\mathbb{P}^n)^T \\ {\scriptstyle i_3}\downarrow & & \downarrow{\scriptstyle i_1} \\ Y & \xrightarrow{i_0} & \mathbb{P}^n \end{array}$$

Hence we obtain the quasi-isomorphism $i_1^* \circ i_{0*} IC_Y \xrightarrow{\cong} i_{2*} \circ i_3^* IC_Y$. (Use proper-base-change.) Taking Verdier-duals and observing that IC_Y is self-dual, we obtain the quasi-isomorphism $Ri_1^! \circ i_{0*} IC_Y \xleftarrow{\cong} i_{2*} \circ Ri_3^! IC_Y$. Now the identifications

$$R\Gamma((\mathbb{P}^n)_{\mathrm{an}}, i_{0*}(IC_Y)) \simeq R\Gamma(Y_{\mathrm{an}}, IC_Y), \quad R\Gamma((\mathbb{P}^n)_{\mathrm{an}}^T, Ri_1^! \circ i_{0*} IC_Y)$$
$$\simeq R\Gamma((\mathbb{P}^n)_{\mathrm{an}}^T, i_{2*} \circ Ri_3^! IC_Y) \simeq R\Gamma((Y^T)_{\mathrm{an}}, Ri_3^! IC_Y)$$

along with (3.6) proves the corollary.

(3.8) *Examples.* (i) Assume in addition to the hypotheses of (3.7) that $Y^T = \{y\}$ is discrete. Then

$$\sum_n (-1)^n (\dim IH^n(Y)) = \sum_n (-1)^n \left(\bigoplus_{y \in Y^T} (\dim H^n(Ri_{y_{\mathrm{an}}}^! IC_Y)) \right),$$

where i_y is the closed immersion of y into Y. For example Y may be a Schubert

variety in G/B, where G is a complex reductive group, B a Borel subgroup and T the maximal torus.

(ii) Suppose in addition to the hypotheses of (3.7) that Y is also smooth. Now $IC_Y \simeq \mathbb{C}_Y$, the obvious constant sheaf on Y. As Y^T is also smooth (see [Iver], (1.3)), $Ri^!_{\text{an}}\underline{\mathbb{C}}_Y \simeq \underline{\mathbb{C}}_Y[-2c]$ where c = the codimension of Y^T in Y. We therefore obtain the well-known formula:

$$\sum_n (-1)^n(\dim H^n(Y;\mathbb{C})) = \sum_n (-1)^n(\dim H^n(Y^T;\mathbb{C}))\,.$$

(See [Iver], p. 229.)

(iii) Suppose in addition to the hypotheses of (3.7) that the inclusion of Y^T in Y is normally non-singular (as in [G-M], p. 115). Then $Ri^!_{\text{an}} IC_Y \simeq IC_{Y^T}$ and hence

$$\sum_n (-1)^n(\dim IH^n(Y)) = \sum_n (-1)^n(\dim IH^n(Y^T))\,.$$

References

[B-B-D] Bielinson, A., Bernstein, J., Deligne, P.: Faisceaux perverse, astérisque **100**, 1981

[Bo] Borel, A.: Algebraic D-modules. Perspectives in mathematics, vol. 2. Boston: Birkhäuser 1986

[GAGA] Serre, J.P.: Geometrie algebrique et geometrie analytique. Ann. Inst. Fourier, pp. 1–42, 1955

[G-M] Goresky, M., Macpherson, R.: Intersection cohomology–II. Invent. Math., **71**, 77–129 (1983)

[Iver] Iverson, B.: A fixed point formula for the action of tori on algebraic varieties. Invent. Math. **16**, 229–236 (1972)

[J-1] Joshua, R.: Vanishing of odd-dimensional intersection cohomology. Math. Z. **195**, 239–253 (1987)

[J-2] Joshua, R.: Riemann-Roch for equivariant D-modules–I. Math. Z. **206**, 131–144 (1991)

[J-3] Joshua, R.: Riemann-Roch for weakly-equivariant $\mathfrak{D}$-modules–II. (to appear)

[Mal] Malgrange, B.: Sur les images direct de D-modules, Manuscr. Math. **50**, 49–71 (1985)

[Niel] Nielson, H.A.: Diagonalizably linearised coherent sheaves. Bull. Soc. Math. Fr. **102**, 85–97 (1974)

[Pop] Popsecu, N.: Abelian categories with applications to rings and modules. New York London Academic Press 1973

[Quart] Quart, G.: Localisation theorem in K-theory for singular varieties, Acta. Math. **143**, 213–217 (1979)

[Qu] Quillen, D.: Higher algebraic K-theory–I. (Lect. Notes Math. vol. 341) Berlin Heidelberg New York: Springer 1974

[SGA 3] Demazure, M., Grothendieck, A.: Schemas en Groupes–I. (Lect. Notes Math., vol. 151) Berlin Heidelberg New York: Springer 1970

[SGA 6] Berthelot, P., Grothendieck, A., Illusie, L.: Theorie de intersection et theorem de Riemann-Roch. (Lect. Notes, Math., vol. 225) Berlin Heidelberg New York: Springer 1971

[Sh] Shafarevich, I.R.: Basic algebraic geometry. (Grundlehren Math., vol. 213) Berlin Heidelberg New York: Springer 1974

[Su] Sumihiro, H.: Equivariant completion. J. Math. Kyoto Univ. **14**, 1–28 (1974)

[T-1] Thomason, R.: Algebraic K-theory of group scheme actions. Ann. Math. Stud. **113**, 539–562 (1987)

[T-2] Thomason, R.: Lefschetz-Riemann-Roch and the coherent trace formula. Invent. Math. **85**, 515–543 (1986)

Math. Z. 206, 303–319 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

Picard groups of Witt rings

Robert W. Fitzgerald
Department of Mathematics, Southern Illinois University, Carbondale, Illinois 62901-4408, USA

Received February 22, 1989, in final form December 15, 1989

We are concerned here, as in [3, 4], with the question of when a regular ideal (or a power of one) of a Witt ring R is principal. In [3] a few partial results were found under the assumption that R had finite height and the set of orderings X_R was finite. The proofs mainly used primary decomposition. In [4] these results were strengthened and R was only required to have finite X_R. Results on Prüfer rings were the basic tool. Here we further strengthen previous results and extend them to arbitrary Witt rings. We use techniques from K-theory.

The first step is to show the Picard group, $\mathrm{Pic}(R)$, is isomorphic to the ideal class group, $\mathrm{Cl}(R)$. We show that each element $x \in \mathrm{Pic}(R)$ is the class of a finitely generated ideal containing an odd dimensional form, and as a result, that x has 2-power order. Working with $\mathrm{Pic}(R)$, instead of $\mathrm{Cl}(R)$ as in [4], allows the use of several exact sequences of K-theory. Analysis of one such sequence shows that if s is the reduced stability index of R and e is the exponent of $\mathrm{Pic}(R)$ then $e = 2^{s-2}$ ($e = 1$ if $s \leqq 1$) whenever e or s is finite. In particular, $\mathrm{Pic}(R) = 1$ (that is, all invertible ideals are principal) iff $s \leqq 2$. This is both a stronger result than previous versions ([3, 3.1], [4, 2.7]) and more general (R is arbitrary).

Analysis of a different exact sequence yields a complete computation of $\mathrm{Pic}(R)$ when X_R is finite. If R_{red} is the reduced Witt ring of R (R modulo the ideal of torsion elements) then $\mathrm{Pic}(R) \approx \mathrm{Pic}(R_{\mathrm{red}})$. We may thus assume R is reduced and hence Noetherian. Marshall [9] has shown such rings arise from copies of $\mathbb{Z}$ via a sequence of fiber products and group ring extensions. $\mathrm{Pic}(\mathbb{Z}) = 1$ and $\mathrm{Pic}(R_1 \times R_2) \approx \mathrm{Pic}(R_1) \oplus \mathrm{Pic}(R_2)$ while $\mathrm{Pic}(R[\Delta])$ can be computed in terms of $\mathrm{Pic}(R)$ and R. All Picard groups then can be found for reduced, Noetherian Witt rings. One consequence of the computation is that if $\{\alpha\}$ is open then the order in $\mathrm{Pic}(R)$ of the class of $P(\alpha, 5)$ is 2^{s-2} (or 1 if $s \leqq 1$), where $s = \mathrm{st}(\alpha, R)$ is the local stability index of Schwartz [11].

Throughout, R will denote an abstract Witt ring, as defined by Marshall [10]. G_R will be the group of one-dimensional forms associated to R and X_R will denote the set of orderings (equivalently, signatures) on R. Since, if $X_R = \varnothing$ then R is a local ring and $\mathrm{Pic}(R) = 1$ we will always assume $X_R \neq \varnothing$. The ideal of even dimensional forms is I_R and the ideal of torsion elements is R_t.

$U(R)$, $\mathrm{Cl}(R)$ and $\mathrm{Pic}(R)$ denote the group of units, the ideal class group and the

Picard group, respectively, of R. X_R has a topology given by a basis of clopen (=closed and open) subsets $H(a_1, \ldots, a_n) = \{\alpha \in X_R \mid a_i >_\alpha 0 \text{ for } 1 \leqq i \leqq n\}$, where $a_1, \ldots, a_n \in G_R$. For $\alpha \in X_R$ and $r \in R$, $\hat{r}(\alpha)$ denotes $\operatorname{sgn}_\alpha(r)$. Then $\hat{r}: X_R \to \mathbb{Z}$ is continuous, where $\mathbb{Z}$ has the discrete topology. We thus have a map from R to the ring of continuous functions $C(X_R, \mathbb{Z})$ given by $r \mapsto \hat{r}$. The *reduced stability index*, $\operatorname{st}(R)$, is k if the cokernel of this map has exponent 2^k and $\operatorname{st}(R) = \infty$ if the cokernel does not have finite exponent.

A *preordering* of R is a subgroup $T \subset G_R$, $T \neq G_R$, such that if $a, b \in T$ then the value set $D\langle a, b\rangle \subset T$. A preordering is a *fan* if every subgroup S, $T \subset S \subset G_R$, of index 2 in G_R such that $-1 \notin S$ is an ordering. If T is a preordering and I_T is the ideal of R generated by all $\langle 1, -t\rangle$, $t \in T$ then $R_T = R/I_T$ is a Witt ring. When T consists of all $x \in G_R$ represented by $n \cdot \langle 1\rangle$ for some n then $I_T = R_t$ and R_T is the reduced Witt ring of R, denoted R_{red}. For background information on $\operatorname{st}(R)$, R_T and fans see Lam [8].

We will denote the fiber product of two Witt rings by the usual product symbol. Thus:

$$R_1 \times R_2 = \{(r, s) \mid r \in R_1,\ s \in R_2 \text{ and } \dim(r) \equiv \dim(s) (\operatorname{mod} 2)\}\ .$$

Δ_n will denote the group of order 2^n and exponent 2, and $R[\Delta_n]$ will denote the extension of R by Δ_n. We will sometimes write Δ instead of Δ_1. We will also sometimes write $\mathbb{Z}_n$ for $\mathbb{Z}/n\mathbb{Z}$.

If $\alpha \in X_R$ and $n \in \mathbb{N}$ then $P(\alpha, n) = \{r \in R \mid \hat{r}(\alpha) \equiv 0 (\operatorname{mod} n)\}$. For an ideal $I \subset R$, $[I]$ will denote the class of I in $\operatorname{Cl}(R)$ (or $\operatorname{Pic}(R)$). If M and N are R-modules then the conductor $[N:M] = \{r \in R \mid rM \subset N\}$.

1 Arbitrary Witt rings

We begin with a very general lemma which is perhaps well-known.

Lemma 1.1 *Let A be a commutative ring such that every finitely generated projective A-module of constant rank is free. Let X be a compact topological space and endow A with the discrete topology. Let $R = C(X, A)$, be the ring of continuous functions. Then every finitely generated projective R-module of constant rank is free.*

Proof. Let M be a finitely generated projective R-module of constant rank k. We may assume $M \oplus N = R^k$ for some R-module N. Let $m_1, \ldots, m_t \in R^k$ generate M with $m_i = (f_{i1}, \ldots, f_{ik})$, $f_{ij} \in R$. Similarly, let $n_i = (g_{i1}, \ldots, g_{ik})$, $1 \leqq i \leqq s$, generate N. Since X is compact we may find a clopen cover $B_1, \ldots, B_p$ of X such that every f_{ij}, g_{ij} is constant on each B_r. Set $T = A^p$.

Define $\psi: T \to R$ by $\psi(a_1, \ldots, a_p)(x) = a_i$ iff $x \in B_i$. ψ is a ring homomorphism. Choose $c = (x_1, \ldots, x_p) \in B_1 \times \ldots \times B_p$ and define $\varphi: R \to T$ by $\varphi(g) = (g(x_1), \ldots, g(x_n))$. φ is also a ring homomorphism, $\varphi \circ \psi = \mathrm{id}_T$ and $\psi \circ \varphi$ fixes all m_i and n_i. Denote the extensions of ψ to T^k and φ to R^k again by ψ and φ

Claim 1 $\varphi(M)$ and $\varphi(N)$ are T-modules with $\varphi(M) \oplus \varphi(N) = T^k$.

For $t \in T$ and $m \in M$ define the T-module action $t\varphi(m)$ to be $\varphi(\psi(t)m)$. Note that $\psi\varphi(M)$ generates M over R (and $\psi\varphi(N)$ generates N) since $\psi\varphi$ fixes the generators m_i, n_i. From $M + N = R^k$ we get $\varphi(M) + \varphi(N) = T^k$ since φ is additive. If

$\alpha \in \varphi(M) \cap \varphi(N)$ then $\psi(\alpha) \in M \cap N = \{0\}$ and so $0 = \varphi\psi(\alpha) = \alpha$. Thus the sum $\varphi(M) + \varphi(N)$ is direct.

Claim 2 $\varphi(M)$ has constant rank.

$\varphi(M)$ is projective by Claim 1. Let $P \in \mathrm{Spec}(T)$ and set $Q = \varphi^{-1}(P) \in \mathrm{Spec}(R)$. Let $v = \mathrm{rank}(M)$ and let $L: M_Q \to R_Q^v$ be an isomorphism. Now $\psi(T \setminus P) \subset R \setminus Q$ since if $\psi(t) \in Q$ then $t = \varphi\psi(t) \in \varphi(Q) = P$. Thus ψ induces a map $\varphi(M)_p \to M_Q$ whose image generates M_Q over R_Q. φ induces a surjection $R_Q^v \to T_p^v$. Then the composition $\varphi \circ h \circ \psi$: $\varphi(M)_p \to T_p^v$ is a surjective T_p-homomorphism. In particular, $\mathrm{rank}_p(\varphi(M)) \geqq v$. If $w = \mathrm{rank}(N)$ then a similar argument shows $\mathrm{rank}_p(\varphi(N)) \geqq w$. Now $M \oplus N = R^k$ implies $v + w = k$ and $\varphi(M) \oplus \varphi(N) = T^k$ implies $\mathrm{rank}_p(\varphi(M)) + \mathrm{rank}_p(\varphi(N)) = k$. Thus for any $p \in \mathrm{Spec}(T)$, $\mathrm{rank}_p(\varphi(M)) = v$.

Claim 3 $\varphi(M)$ is a free T-module.

Recall that $T = A^p$ and projective modules over A of constant rank are assumed to be free. Claim 3 then follows from the general result: If projective R_i-modules ($i = 1, 2$) of constant (finite) rank are free then projective $R_1 \oplus R_2$-modules of constant finite rank are free. In the absence of a good reference, we sketch a proof of this.

Let M be a projective module of constant rank m over $S \equiv R_1 \oplus R_2$. If we set $e_1 = (1, 0)$ and $e_2 = (0, 1)$ then $M \approx e_1 M \oplus e_2 M$. Let $P \in \mathrm{Spec}(R_1)$ and set $Q = P \oplus R_2 \in \mathrm{Spec}(S)$. Note that $M_Q = (e_1 M)_Q$ and $S_Q = (R_1 \oplus 0)_Q = (R_1)_p \oplus 0$, since $e_1 \notin Q$ and $e_1(e_2 M) = 0$. Identify R_1 with $R_1 \oplus 0$ and let N_1 be $e_1 M$ viewed as an R_1-module under this identification. Then:

$$(N_1)_P = (e_1 M)_Q = M_Q \approx S_Q^m = (R_1)_P^m$$

Thus N_1 is locally free (hence projective) of constant rank m. By assumption, $N_1 \approx R_1^m$ and similarly if N_2 is $e_2 M$ viewed as an R_2-module then $N_2 \approx R_2^m$. Thus $M \approx e_1 M \oplus e_2 M \approx R_1^m \oplus R_2^m \approx S^m$ is free.

We now complete the proof of the lemma. For each choice of $c = (x_1, \ldots, x_p) \in B_1 \times \ldots \times B_p$ we have defined $\varphi_c: R \to A^p$ by $\varphi_c(g) = (g(x_1), \ldots, g(x_p))$. We have that $\varphi_c(M)$ is a free A^p-module. Note that if $d \in B_1 \times \ldots \times B_p$ is another choice then $\varphi_c(M) = \varphi_d(M)$ as $\varphi_c(m_i) = \varphi_d(m_i)$ and the m_i generate M. Let $z_1, \ldots, z_e$ be an A^p-basis for $\varphi(M)$ (which denotes $\varphi_c(M)$ for any c). We claim $\psi(z_1), \ldots, \psi(z_e)$ is an R-basis for M.

Each $\varphi(m_i)$ lies in $\varphi(M)$ and so in the A^p-span of the z_i. Hence each $m_i = \psi(\varphi(m_i))$ lies in the R-span of the $\psi(z_i)$. Thus $\psi(z_1), \ldots, \psi(z_e)$ generate M. Suppose $\Sigma r_i \psi(z_i) = 0$ for $r_i \in R$ with say $r_1 \neq 0$. There exists y_1, say $y_1 \in B_1$, with $r_1(y_1) \neq 0$. Extend y_1 to $d = \{y_1, \ldots, y_p\} \in B_1 \times \ldots \times B_p$. Then $0 = \Sigma \varphi_d(r_i)\varphi_d\psi(z_i) = \Sigma \varphi_d(r_i) z_i$ with $\psi_d(r_1) = (r_1(y_1), \ldots) \neq 0$. This is a contradiction and shows M is free. □

Throughout this section R will denote an arbitrary (i.e. possibly non-Noetherian) Witt ring of dimension 1. X_R is its (compact) space of orderings. Our first goal is to get a good representative for each class $x \in \mathrm{Pic}(R)$.

Lemma 1.2 *Let K be the total quotient ring of R.*

(1) *If R is not reduced then K is a local ring.*

(2) *If R is reduced then $K \approx C(X_R, \mathbb{Q})$.*

Proof. (1) Here $r \in R$ is regular iff $r \notin I_R$, the fundamental ideal of R. Thus $K = R_{I_R}$ is a local ring.

(2) Map $\alpha: K \to C(X_R, \mathbb{Q})$ by $\alpha(s^{-1}r) = (\hat{s})^{-1}(\hat{r})$. This is well-defined and injective. To show α is surjective choose any $f \in c(X_R, \mathbb{Q})$. X_R is compact so $f(X_R)$ is finite. X_R is thus the disjoint union of clopen sets $A_1, \ldots, A_n$ where $f|A_i$ is constant, say $f(A_i) = q_i \in \mathbb{Q}$. Choose $p_i \in R$ such that $\hat{p}_i \equiv 2^m$ on A_i and $\hat{p}_i \equiv 0$ on $X_R - A_i$ (this is possible by [2, 3.2]), for fixed m and all i. Set $\varphi = \Sigma(q_i p_i)/2^m \in K$. Then $f = \hat{\varphi}$ as desired. □

Theorem 1.3 $\mathrm{Cl}(R) \approx \mathrm{Pic}(R)$.

Proof. The sequence $0 \to \mathrm{Cl}(R) \to \mathrm{Pic}(R) \to \mathrm{Pic}(K)$ is exact [1, p. 503], where K is the total quotient ring of R. If R is not reduced then K is local (1.2) and so $\mathrm{Pic}(K) = 0$. If R is reduced then $K = C(X_R, \mathbb{Q})$ (1.2) and again $\mathrm{Pic}(K) = 0$, by (1.1). So $\mathrm{Cl}(R) \approx \mathrm{Pic}(R)$. □

We may thus assume each $x \in \mathrm{Pic}(R)$ is the class of an invertible ideal in R. We may refine this representation. Recall from [4] that an ideal $I \subset R$ is *regular* if it contains a non-zero-divisor and *strongly regular* if it contains an odd dimensional form. The following generalizes [4, 2.6].

Lemma 1.4 *If $I \subset R$ is an invertible ideal then there exists a strongly regular ideal J and $a, b \in R$ with $aI = bJ$.*

Proof. We may assume $I \subset I_R$ so that $I_{I_R} = (a)_{I_R}$, for some regular $a \in I$. Since I is invertible, I is finitely generated and we may write $I = (a, b_1, \ldots, b_n)$ for some $b_1, \ldots, b_n \in I$. For each b_i there exists $x_i \in R$ and $c_i \in R \setminus I_R$ with $c_i b_i = a x_i$, since $b \in (a)_{I_R}$. Set $c = \Pi c_i$ and $d_i = c/c_i$. Then $c \notin I_R$ and

$$\begin{aligned} cI &= (ac, c_1 b_1 d_1, \ldots, c_n b_n d_n) \\ &= (ac, a x_1 d_1, \ldots, a x_n d_n) \\ &= (a)(c, x_1 d_1, \ldots, x_n d_n)\,. \end{aligned}$$

Take $J = (c, x_1 d_1, \ldots, x_n d_n)$. □

The following extends [3, 2,9]. R_t denotes the torsion ideal of R.

Lemma 1.5 *Let $r \in R$ be odd dimensional.*

(1) $r^{2^k} \equiv 1 \pmod{2^k}$ *for all* $k \geqq 1$.

(2) $R_t \subset (r)$.

Proof. Write $r = \langle x \rangle + r_0$ with $r_0 \in I_R$. Then $r^2 = 1 + 2r_0 + r_0^2 \equiv 1 \pmod 2$. (1) follows by induction. Now suppose $2^k \omega = 0$. Then $(r^{2^k} - 1)\omega = 0$ and $\omega = r^{2^k}\omega \in (r)$. Hence $R_t \subset (r)$. □

Proposition 1.6 $\mathrm{Pic}(R) \approx \mathrm{Pic}(R_{\mathrm{red}})$, *where R_{red} is the reduced Witt ring R/R_t.*

Proof. We show $\mathrm{Cl}(R) \approx \mathrm{Cl}(R_{\mathrm{red}})$ and apply (1.3). Each element of $\mathrm{Cl}(R)$ may be represented as the class of a strongly regular ideal I (1.4) and $R_t \subset I$ by (1.5). We map $\mathrm{Cl}(R) \to \mathrm{Cl}(R_{\mathrm{red}})$ by sending $[I]$ to $[I/R_t]$. The map is injective since if I/R_t is principal, $I/R_t = (r + R_t)(R/R_t) = (r)/R_t$ (as r must be odd dimensional and so $R_t \subset (r)$). The map is also surjective. If $J \subset R_{\mathrm{red}}$ is strongly regular then $J = I/R_t$ for some strongly regular ideal $I \subset R$. J invertible implies I is invertible. So $[I] \in \mathrm{Cl}(R)$ and $[I] \to [J]$. □

We have that each $x \in \mathrm{Pic}(R)$ is the ideal class of an invertible, strongly regular ideal I. When R is Noetherian, every strongly regular ideal is invertible (cf. [4, 2.4]). Our next goal is to determine which strongly regular ideals are invertible when R is non-Noetherian.

Definition. Let $I \subset R$ be a strongly regular ideal. For each $\alpha \in X_R$, $\mathrm{sgn}_\alpha(I)$ is an ideal of $\mathbb{Z}$. Define $\hat{I}\colon X_R \to \mathbb{N}$ by $\hat{I}(\alpha) = k$ iff $\mathrm{sgn}_\alpha(I) = k\mathbb{Z}$. Set $K_n(I) = \hat{I}^{-1}(n)$.

Note that $\hat{I}(\alpha) = \min\{|\hat{r}(\alpha)|\colon r \in I,\ \hat{r}(\alpha) \neq 0\}$.

Lemma 1.7 *Let $r, s \in R$ with r odd dimensional. Suppose $\hat{r}(\alpha) \mid \hat{s}(\alpha)$ for all $\alpha \in X_R$. Then $r \mid s$ in R.*

Proof. If $|X_R| < \infty$ then $(r) = \bigcap P(\alpha, \hat{r}(\alpha))$ by primary decomposition [3, 2.3] and so $s \in (r)$. Suppose $|X_R| = \infty$ and let $T \subset X_R$ be a fan of finite index (cf. [8]). Set $n_\alpha = \hat{s}(\alpha)/\hat{r}(\alpha)$. Note that $r_T \in R_T$ is odd dimensional and that for $\alpha \in X_R/T$, $\hat{r}_T(\alpha) = \hat{r}(\alpha)$ and $\hat{s}_T(\alpha) = \hat{s}(\alpha)$. Hence there exists $w_T \in R_T$ with $s_T = w_T r_T$. In particular, $\sum_{X/T} n_\alpha \equiv 0 \pmod{|X_R/T|}$. By the representation theorem [8, 7.2], there exists $\omega \in R$ with $\hat{\omega}(\alpha) = n_\alpha$ for all $\alpha \in X_R$. Thus $s - \omega r \in R_t \subset (r)$, by (1.5), and so $s \in (r)$. □

Lemma 1.8 *Let $I \subset R$ be a strongly regular ideal and let $r \in R$ be odd dimensional. If $\hat{I} = \hat{r}$ then $I = (r)$.*

Proof. $\hat{I}(\alpha) = \hat{r}(\alpha)$ divides $\hat{s}(\alpha)$ for all $s \in I$, $\alpha \in X_R$. Hence $I \subset (r)$ by (1.7). Set $J = [I\colon (r)]$. We claim $\hat{J} \equiv 1$. Choose any $\alpha \in X_R$. Since $\hat{I}(\alpha) = \hat{r}(\alpha)$, there exists an $s \in I$ with $\hat{s}(\alpha) = \hat{r}(\alpha)$. Now $s = r\omega$, for some $\omega \in R$. Then $\omega \in J$ and $\hat{\omega}(\alpha) = 1$. Thus $\hat{J}(\alpha) = 1$.

The claim implies J is not contained in any maximal ideal of R (I_R or $P(\alpha, p)$). Hence $J = R$ and $I = (r)$. □

Lemma 1.9 *Let $I \subset R$ be a strongly regular ideal.*

(1) *$\hat{I}(X_R)$ is finite.*

(2) *If $n \in \hat{I}(X_R)$ does not divide any $m \in \hat{I}(X_R)$, $m \neq n$, then $K_n(I)$ is closed.*

Proof. (1) Let $r \in I$ be odd dimensional. For all $\alpha \in X_R$, $\hat{I}(\alpha)$ divides $\hat{r}(\alpha)$ and $\hat{r}(\alpha) \neq 0$. Now $\hat{r}$ is continuous and X_R compact, so there exists M with $|\hat{r}(\alpha)| \leqq M$ for all $\alpha \in X_R$. Thus $1 \leqq \hat{I}(\alpha) \leqq M$ for all $\alpha \in X_R$.

(2) Let $L_n = \bigcap_{r \in I} \left[\bigcup_{k=-\infty}^{\infty} \hat{r}^{-1}(kn) \right]$. If $\alpha \in L_n$ then $n \mid \hat{r}(\alpha)$ for all $r \in I$ and so $n \mid \hat{I}(\alpha)$. But no multiple of n (except n itself) lies in $\hat{I}(X_R)$, so we must have $\hat{I}(\alpha) = n$. Thus $L_n = K_n(I)$. The union of the $\hat{r}^{-1}(kn)$, in L_n, is a finite union (as $\hat{r}(X_R)$ is finite) and hence closed. So $L_n = K_n(I)$ is closed. □

Theorem 1.10 *Let $I \subset R$ be a strongly regular ideal. The following are equivalent:*

(1) *I is invertible,*

(2) *I is finitely generated,*

(3) *$\hat{I}$ is continuous,*

(4) *$K_n(I)$ is closed for all $n \in \mathbb{N}$,*

(5) *$K_n(I)$ is closed for all $n \in \hat{I}(X_R)$ which divide some $m \in \hat{I}(X_R)$, $m \neq n$.*

Proof. (1) → (2) and (3) → (4) → (5) are clear. We first prove (2) → (3). Say $I = (a_1, \ldots, a_m)$. Set $L_n = \bigcap_{i=1}^{m} \left(\bigcup_{k=-\infty}^{\infty} \hat{a}_i^{-1}(kn) \right)$. The union of the $\hat{a}^{-1}(kn)$ is finite,

hence open, and so L_n is open. Set $H_n = U_{r\in I} r^{-1}(n)$; H_n is open. Now $\alpha \in L_n \cap H_n$ iff $n \mid \hat{a}_i(\alpha)$ for $i = 1, \ldots, m$ and there exists $r \in I$ with $\hat{r}(\alpha) = n$. Thus $K_n(I) = L_n \cap H_n$ is open for every n. Since $K_n(I) = \hat{I}^{-1}(n)$ we have that $\hat{I}$ is continuous.

Lastly we prove (5) $\rightarrow$ (1). (5) combined with (1.9) shows that all $K_n(I)$ are closed. $K_n(I) = X_R \setminus \bigcup_{m \neq n} K_m(I)$ is also closed since $K_m(I) = \varnothing$ for all but finitely many m (1.9). Let $\hat{I}(X_R) = \{n_1, \ldots, n_k\}$. We may choose $m \in \mathbb{N}$ and $c_i \in R$ such that for all $i = 1, \ldots, k$, $\hat{c}_i \equiv 2^m$ on $K_{n_i}(I)$ and $\hat{c}_i \equiv 0$ on $X_R \setminus K_{n_i}(I)$ [2, 3.2]. Choose a 2-power t such that $n_i^t \equiv 1 \pmod{2^m}$ for all i and write $n_i^t - 1 = 2^m p_i$. Then for $c = 1 + \sum_{i=1}^{k} p_i c_i$ we have $\hat{c}(\alpha) = \hat{I}(\alpha)^t$ for all $\alpha \in X_R$. Hence $I^t = (c)$ by (1.8) and so I is invertible. □

By definition, in a Prüfer ring, an ideal is invertible iff it is finitely generated and regular. When R is reduced (and $R \neq \mathbb{Z}$), R is not a Prüfer ring but (1) $\leftrightarrow$ (2) in (1.10) shows a slightly weaker property holds. When R is not reduced then all regular ideals are strongly regular and so (1.10) proves R is Prüfer, as was earlier proven by Griffen [5].

The following generalizes [4, 2.9].

Corollary 1.11 *If $I \subset R$ is an invertible ideal then I^t is principal for some 2-power t. In particular, every element of* Pic(R) *has 2-power order.*

Proof. $[I] = [J] \in \mathrm{Cl}(R)$ for some invertible, strongly regular ideal J (1.4). The proof of (1.10) shows J^t is principal for some 2-power t, hence so is I^t. The second statement follows from (1.3). □

Corollary 1.12 *$P(\alpha, p)$ is invertible* iff *$\{\alpha\}$ is clopen.*

Proof. For any ordering $\beta \neq \alpha$ choose $a \in G$ with $a >_\beta 0$, $a <_\alpha 0$. Then $r = 1 + \frac{1}{2}(p-1)\langle 1, -a\rangle \in P(\alpha, p)$ and $\hat{r}(\beta) = 1$. Thus $K_1 = X_R \setminus \{\alpha\}$ and $K_p = \{\alpha\}$. The result then follows from (1.10). □

The last goal of this section is to relate the exponent of Pic(R) to the reduced stability index st(R).

Lemma 1.13 *Let R be reduced. The integral closure of R, $\bar{R}$, is isomorphic to $C(X_R, \mathbb{Z})$.*

Proof. We identify R with its image in $C(X_R, \mathbb{Z})$. Clearly $C(X_R, \mathbb{Z})$ is integrally closed in $C(X_R, \mathbb{Q})$, the total quotient ring of R (1.2). To show $C(X_R, \mathbb{Z})$ is integral over R choose any $f \in C(X_R, \mathbb{Z})$. We may find a finite disjoint cover of X_R consisting of clopen subsets A_i ($1 \leq i \leq n$) such that $f(A_i) = z_i \in \mathbb{Z}$. We may choose $m \in \mathbb{N}$ and $s_i \in R$ such that $\hat{s}_i \equiv 2^m$ on A_i and $\hat{s}_i \equiv 0$ on $X_R \setminus A_i$, for all i. Let f_i: $X_R \rightarrow \mathbb{Z}$ be the function equal to 1 on A_i and 0 on $X \setminus A_i$. Note that each f_i is continuous and $f = \Sigma z_i f_i$. Each f_i satisfies: $x^2 - (2^m + 1)x + \hat{s}_i = 0$. Thus each f_i, and so f, is integral over R. □

Lemma 1.14 *Let R be reduced, $\bar{R}$ its integral closure and $s =$* st(R). Set $B = [R : \bar{R}]$ and suppose $s < \infty$.

(1) *If $s \leq 1$ then $U(\bar{R}/B) = 1$.*

(2) *If $s \geq 2$ and $f + B \in U(\bar{R}/B)$ then $f^m + B = g + B$, for some $g \in U(\bar{R})$, where $m = 2^{s-2}$.*

Proof. We identify $\bar{R}$ with $C(X_R, \mathbb{Z})$. Let $f \in \bar{R}$ and $f + B \in U(\bar{R}/B)$. Let $\{A_1, \ldots, A_n\}$ be a disjoint clopen cover of X_R with $f(A_i) = z_i \in \mathbb{Z}$. We first show that each z_i is odd. If some z_i, say z_1, is even then let $g: X_R \to \mathbb{Z}$ be given by $g \equiv 1$ on A_1 and $g \equiv 0$ on $X \backslash A_1$. Note that g is continuous and so $g \in \bar{R}$. By definition of the reduced stability index, $2^s \bar{R} \in R$ and so $2^s \bar{R} \in B$. Now $gf^s = z_1^s g \in 2^s \bar{R} \in B$, so $f + B$ is a zero-divisor which is impossible.

If $s = 0$ then $\bar{R} = B = R$ and (1) holds vacuously. If $s = 1$ then $f \equiv 1 \pmod 2$ and so $f + B = 1 + B$ as $2\bar{R} \subset B$. Thus $U(\bar{R}/B) = 1$.

Now suppose $s \geqq 2$. $U(\mathbb{Z}/2^s\mathbb{Z}) \approx \{\pm 1\} \oplus C$, where C is a cyclic group of order $m = 2^{s-2}$. Thus for each i, $z_i^m \equiv \pm 1 \pmod{2^s}$. Let $\varepsilon_i = 1$ or -1 depending on whether $z_i^m \equiv 1$ or $-1 \pmod{2^s}$. Write $z_i^m - \varepsilon_i = 2^s x_i$ and define $g: X_R \to \mathbb{Z}$ and $h: X_R \to \mathbb{Z}$ by $g(\alpha) = \varepsilon_i$, $h(\alpha) = x_i$ for $\alpha \in A_i$. Clearly g and h are continuous. Then $f^m = g + 2^s h \in g + B$ and certainly g is a unit in $\bar{R}$. □

Lemma 1.15 *Let R be reduced and let $T \subset X_R$ be a preordering of finite index. Then the natural map $\mathrm{Pic}(R) \to \mathrm{Pic}(R_T)$ is surjective.*

Proof. Let $J \subset R_T$ be a strongly regular ideal. Let K_T be the ideal of R generated by all $\langle 1, -t\rangle$, with $t \in T$; $R_T \approx R/K_T$. We will construct an invertible ideal $L \subset R$ such that $(L + K_T)/K_T = J$, which is sufficient by (1.3) and (1.4). Let $\hat{J}(X_R/T) = \{n_1, \ldots, n_k\}$ and let $A_i = \hat{J}^{-1}(n_i)$.

Claim 1 There is a disjoint clopen cover $B_1, \ldots, B_k$ of X such that $B_i \cap (X_R/T) = A_i$.

Suppose $t \in \mathbb{N}$ and $p_i = \langle\langle a_{i1}T, \ldots, a_{it}T\rangle\rangle \in R_T$ $(1 \leqq i \leqq k)$ satisfy $\hat{p}_i \equiv 2^t$ on A_i and $\hat{p}_i \equiv 0$ on $(X_R/T) - A_i$. Set $\sigma_i = \langle\langle a_{i1}, \ldots, a_{it}\rangle\rangle$. Let $B_1 = \hat{\sigma}_1^{-1}(2^t)$. Then B_1 is clopen and $B_1 \cap (X_R/T) = A_1$. For $2 \leqq i < k$ set $B_i = \hat{\sigma}_i^{-1}(2^t) \backslash \left(\bigcup_{j=1}^{i-1} B_j \right)$. Then B_i is clopen and $B_i \cap (X/T) = A_i$, since $A_i = \hat{\sigma}_i^{-1}(2^t) \cap (X_R/T)$ and $\left(\bigcup_{j=1}^{k-1} B_j \right) \cap A_i = \varnothing$. Lastly, set $B_k = X_R \backslash \left(\bigcup_{j=1}^{k-1} B_j \right)$. Then B_k is clopen, $B_k \cap (X_R/T) = (X_R/T) \backslash \left(\bigcup_{j=1}^{k-1} A_j \right) = A_k$ and clearly $B_1, \ldots, B_k$ is a disjoint cover of X.

Set $L = \{r \in R \mid \hat{r}(\alpha) \in n_i\mathbb{Z} \text{ if } \alpha \in B_i\}$ and set $N = n_1 \cdot n_2 \cdot \ldots \cdot n_k$. Note that L is an ideal, N is odd and $N \in L$ so that L is strongly regular.

Claim 2 $\hat{L}(X_R) = \{n_1, \ldots, n_k\}$ and $\hat{L}^{-1}(n_i) = B_i$ for $1 \leqq i \leqq k$.

Let $\alpha \in B_i$ and note that $n_i \mid \hat{L}(\alpha)$ by construction. Now choose $p \in R$ and $d \in \mathbb{N}$ with $\hat{p} \equiv 2^d$ on B_i and $\hat{p} \equiv 0$ iff B_i. Then $n_i p \in L$ and $gcd(2^d n_i, N) = n_i$. Choose, $u, v \in \mathbb{Z}$ such that $u 2^d n_i + vN = n_i$ and set $s = u n_i p + vN \in L$. Then $\hat{s}(\alpha) = n_i$ and so $\hat{L}(\alpha) = n_i$. The same reasoning shows that if $\alpha \notin B_i$ then $\alpha \in B_j$ for some $j \neq i$ and $\hat{L}(\alpha) = n_j \neq n_i$. Thus $\hat{L}^{-1}(n_i) = B_i$.

Claim 2 shows L is invertible by (1.10). If say $n_1 \neq 1$ then for $\alpha \in A_1$ we have $\hat{L}(\alpha) = n_1$, $\hat{K}_r(\alpha) = 0$ and so $L + K_T$ is a proper ideal of R. Set $L_T = (L + K_T)/K_T \in R_T$. For any $\alpha \in X_R/T$, $\hat{L}_T(\alpha) = \hat{L}(\alpha) = n_i$ iff $\alpha \in A_i$ and so $\hat{L}_T = \hat{J}$. For any primary ideal $P(\alpha, p^i)$, p is a prime, we have $L_T \subset P(\alpha, p^i)$ iff $J \subset P(\alpha, p^i)$. Thus by primary decomposition [3, 2.3] we have $L_T = J$ as desired. □

Theorem 1.16 *Let R be reduced, and set $s = \mathrm{st}(R)$ and let e be the exponent of* $\mathrm{Pic}(R)$.

(1) *s is finite* iff *e is finite.*

(2) *If $s \leqq 1$ then $e = 1$.*

(3) *Suppose $s \geqq 2$ and both s and e are finite. Then $e = 2^{s-2}$.*

Proof. Let $\bar{R}$ be the integral closure of R and let $B = [R:\bar{R}]$. We have a Cartesian diagram (cf. [1, IX 5.6]):

$$\begin{array}{ccc} R & \longrightarrow & \bar{R} \\ \downarrow & & \downarrow \\ R/B & \longrightarrow & \bar{R}/B. \end{array}$$

Apply the Mayer–Vietoris sequence [1, IX5.3] to get the exact sequence:

$$U(R/B) \oplus U(\bar{R}) \xrightarrow{\beta} U(\bar{R}/B) \xrightarrow{\alpha} \mathrm{Pic}(R) \longrightarrow \mathrm{Pic}(R/B) \oplus \mathrm{Pic}(\bar{R}) .$$

Now $2^s \in B$ so that the only prime ideal of R containing B is I_R. Thus R/B is local and $\mathrm{Pic}(R/B) = 1$. Also, $\bar{R} \approx C(x, \mathbb{Z})$ by (1.13) and so $\mathrm{Pic}(\bar{R}) = 1$ by (1.1). Hence we have the exact sequence:

$$(*) \qquad U(R/B) \oplus U(\bar{R}) \xrightarrow{\beta} U(\bar{R}/B) \xrightarrow{\alpha} \mathrm{Pic}(R) \longrightarrow 1$$

Lemma 1.14 (1) now proves (2).

Suppose now that s is finite and $s \geqq 2$. We will show e is finite and $e \leqq 2^{s-2}$. Let $x \in \mathrm{Pic}(R)$ and let $x = \alpha(f + B)$ for some $f + B \in U(\bar{R}/B)$. By (1.14), if $m = 2^{s-2}$ then $f^m + B = g + B$ for some $g \in U(\bar{R})$. In particular, $f^m + B$ lies in the image of β. The exactness of $(*)$ gives $x^m = \alpha(f^m + B) = 1$ and so $e \leqq m = 2^{s-2}$.

Lastly, suppose e is finite. We will show s is finite and $2^{s-2} \leqq e$, thus completing the proof. Let T be a fan of finite index, say $|X_R/T| = 2^t$. Then $R_T = \mathbb{Z}[\Delta_t]$ by [8, 5.9] and the exponent of $\mathrm{Pic}(R_T)$ is 2^{t-2} ([4, 3.1], and also re-derived in the next section, independently of this theorem). From the surjection $\mathrm{Pic}\, R \to \mathrm{Pic}\, R_T$ (1.15) we see that $e \geqq 2^{t-2}$. Now $s = \sup\{\log_2 |X_R/T| : T \text{ is a fan of finite index}\}$ by [8, 13.7]. Thus $s \leqq 2 + \log_2 e$ and so s is finite and $2^{s-2} \leqq e$. □

The following generalizes [3, 3.1] and [4, 2.7]. This result was the major motivation for this work.

Corollary 1.17 *Let R be an arbitrary Witt ring and let* $\mathrm{st}(R)$ *denote the reduced stability index. Then* $\mathrm{Pic}(R) = 1$ *iff* $\mathrm{st}(R) \leqq 2$.

Proof. We may assume R is reduced by (1.6) and apply (1.16). □

We may rephrase part of (1.16) as follows:

Corollary 1.18 *Let R be an arbitrary Witt ring and let $s = \mathrm{st}(R)$. Let $I \in R$ be a strongly regular ideal such that $\hat{I}$ is continuous. Then I^m is principal, where $m = \max\{1, 2^{s-2}\}$.* □

2 Noetherian Witt rings

We compute $\mathrm{Pic}(R)$ for all Witt rings R with only finitely many orderings. We may assume that R is reduced (1.6) and hence Noetherian. Marshall ([9], see also [10]) has shown that all reduced, Noetherian Witt rings can be formed from copies of $\mathbb{Z}$ by a sequence of fiber products and group ring extensions. Of course, $\mathrm{Pic}(\mathbb{Z}) = 1$.

Lemma 2.1 *Let R_1 and R_2 by Witt rings and $R = R_1 \times R_2$. Then* $\mathrm{Pic}(R) = \mathrm{Pic}(R_1) \oplus \mathrm{Pic}(R_2)$.

Proof. The product here is the fiber product over $\mathbb{Z}_2$. The Mayer–Vietoris sequence gives:

$$U(\mathbb{Z}_2) \to \mathrm{Pic}(R) \to \mathrm{Pic}(R_1) \oplus \mathrm{Pic}(R_2) \to \mathrm{Pic}(\mathbb{Z}_2)$$

is exact. Clearly $U(\mathbb{Z}_2)$ and $\mathrm{Pic}(\mathbb{Z}_2)$ are trivial, so the desired isomorphism follows.

The rest of this section is thus concerned with the case of a group ring extension. We fix the notation. R is a reduced, Noetherian Witt ring, $\Delta = \{1, t\}$ is the group of order 2 and $S = R[\Delta]$. We have the following homomorphisms:

$$h_+ : S \to R \quad a + bt \mapsto a + b$$

$$h_- : S \to R \quad a + bt \mapsto a - b\,.$$

For an ideal $J \subset S$ we set $J_+ = h_+(J)$ and $J_- = h_-(J)$. J_+ and J_- are ideals of R.

Lemma 2.2 *Let $J \subset S$ and $I_1, I_2 \subset R$ be strongly regular ideals. Then:*

(1) $(\langle 1, -t\rangle R + I_1 S)(\langle 1, -t\rangle R + I_2 S) = \langle 1, -t\rangle R + I_1 I_2 S$

(2) $(\langle 1, t\rangle R + I_1 S)(\langle 1, t\rangle R + I_2 S) = \langle 1, t\rangle R + I_1 I_2 S$

(3) $J = (\langle 1, -t\rangle R + J_+ S)(\langle 1, t\rangle R + J_- S)$.

Proof. We prove (1) and note that (2) is similar. Now $(\langle 1, -t\rangle R + I_1 S)(\langle 1, -t\rangle R + I_2 S) = \langle\langle 1, -t\rangle\rangle R + \langle 1, -t\rangle (I_1 + I_2) + I_1 I_2 S = \langle 1, -t\rangle\ (2R + I_1 + I_2) + I_1 I_2 S = \langle 1, -t\rangle R + I_1 I_2 S$, since $2R + I_1 + I_2$ contains 2 and an odd dimensional form, hence equals R.

(3) Set $L = (\langle 1, -t\rangle R + J_+ S)(\langle 1, t\rangle R + J_- S)$. Then $h_+(L) = J_+(2R + J_-) = J_+ = h_+(J)$, since again $2R + J_-$ contains 2 and an odd dimensional form. Thus $J + (\langle 1, -t\rangle) = L + (\langle 1, -t\rangle)$. If $\alpha \in X_S$ satisfies $t >_\alpha 0$ then $\hat{J}(\alpha) = \hat{L}(\alpha)$. In particular, for any primary ideal $P(\alpha, p^i)$ we have $J \subset P(\alpha, p^i)$ iff $L \subset P(\alpha, p^i)$. Repeating the argument with h_- instead of h_+ gives that if $\beta \in X_S$ satisfies $t <_\beta 0$ then $J \subset P(\beta, p^i)$ iff $L \subset P(\beta, p^i)$. Thus, by primary decomposition [3, 2.3], we have $J = L$. □

Lemma 2.3 *The following sequence is exact:*

$$(*) \qquad 1 \to U(S) \to U(R) \oplus U(R) \xrightarrow{f} U(R/2R) \xrightarrow{g} \mathrm{Pic}(S)$$

$$\xrightarrow{h} \mathrm{Pic}(R) \oplus \mathrm{Pic}(R) \longrightarrow 1$$

where

$$f(u, v) = uv^{-1} + 2R$$

$$g(\omega + 2R) = [(\langle 1, -t\rangle, \omega)]$$

and

$$h([J]) = ([J_+], [J_-])\,.$$

Proof. Let $I_1 = (\langle 1, -t\rangle)$ and $I_2 = (\langle 1, t\rangle)$. Then $I_1 \cap I_2 = 0$, $S/I_1 \approx S/I_2 \approx R$ and $S/(I_1 + I_2) \approx S/(2, \langle 1, -t\rangle) \approx R/2R$. We thus have a Cartesian diagram (cf. [1, IX5.5]):

$$\begin{array}{ccc} S & \longrightarrow & R \\ \downarrow & & \downarrow \\ R & \longrightarrow & R/2R \end{array}$$

Note that $R/2R$ is a local ring (the only maximal ideal of R containing 2 is I_R) so that $\mathrm{Pic}(R/2R) = 1$. The Mayer–Vietoris sequence [1, IX5.3] yields (*).

The maps f and h are clear. The map g is the connecting homomorphism $K_1(R/2R) \to K_0(S)$ and for this we follow the discussion in [12]. For $\gamma + 2R \in U(R/2R)$, $g(\gamma + 2R) = [M]$ where $M = \{(r, r\gamma + 2v) | r, v \in R\} \subset R^2$. The S-module action on M is $(a + bt)\cdot(r, r\gamma + 2v) = ((a + b)r, (a - b)(\gamma r + 2v))$. Thus M is generated, as an S-module, by $m_1 = (1, \gamma)$ and $m_2 = (0, 2)$. The relations are $(a + bt)m_1 = (a + b)m_1 - b\gamma m_2 = (a - b)\langle 1, -t\rangle$ and $(a + bt)m_2 = (a - b)m_2$. Now $(a + bt)\langle 1, -t\rangle$ and $(a + bt)\gamma = (a + b)\gamma - b\gamma\langle 1, -t\rangle$. Hence, as S-module, M is isomorphic to the ideal $(\langle 1, -t\rangle, \gamma) \subset S$. □

We introduce more notation. Let A be a 2-primary abelian group (written additively). Let $A[2^n] = \{a \in A | 2^n a = 0\}$. The *i-th Ulm invariant* is:

$$u_i(A) = \dim_{\mathbb{Z}_2}[2^{i-1}A[2^i]/2^i A[2^{i+1}]] .$$

If A is a finite abelian group then A is isomorphic to $(\mathbb{Z}_2)^{u_1} \oplus (\mathbb{Z}_4)^{u_2} \oplus \cdots \oplus (\mathbb{Z}_{2^n})^{u_n}$ with $u_i = u_i(A)$.

The key to analyzing the sequence (*) is the following:

Lemma 2.4 *Let A, B and C be finite 2-primary abelian groups and suppose:*

$$0 \longrightarrow A \xrightarrow{g} B \xrightarrow{h} C \oplus C \longrightarrow 0$$

is exact. Further suppose:

(1) $2A = 0$,

(2) *If* $h(b) = (c, c)$ *and* $2c = 0$ *then* $2b = 0$,

and (3) *If* $h(b) = (c, 0)$ *and* $2b = 0$ *then* $c = 0$.

Then for $n \geqq 1$, $u_{n+1}(B) = u_{n+1}(C) + u_n(C)$.

Proof. For any group G let $G_n = 2^{n-1}G[2^n]$. Consider:

$$\varphi_1 \colon C_n \to B_{n+1} \qquad 2^{n-1}c \mapsto 2^n h^{-1}(c, 0)$$

$$\varphi_2 \colon C_{n+1} \to B_{n+1} \qquad 2^n c \mapsto 2^n h^{-1}(c, c)$$

Step 1 φ_1 and φ_2 are well-defined.

For any $c, d \in c$, if $b_1, b_2 \in h^{-1}(c, d)$ then $b_1 - b_2 \in g(A)$ and so $2^n b_1 = 2^n b_2$ (as $n \geqq 1$). Thus $2^n h^{-1}(c, 0)$ and $2^n h^{-1}(c, c)$ are well-defined elements of B. Let $c \in C[2^n]$ and $b \in h^{-1}(c, 0)$. Then $h(2^n b) = (2^n c, 0) = 0$, $2^n b \in g(A)$ and so $2^{n+1}b = 0$. Thus $\varphi_1(2^{n-1}c) = 2^n b \in B_{n+1}$ as required. Finally, suppose $c, d \in C[2^n]$ and $2^{n-1}c = 2^{n-1}d$. Let $b_1 \in h^{-1}(c, 0)$ and $b_2 \in h^{-1}(d, 0)$. Then $h(2^{n-1}(b_1 - b_2)) = 0$ implies $\varphi_1(2^{n-1}c) = 2^n b_1 = 2^n b_2 = \varphi_1(2^{n-1}d)$. Thus φ_1 is well-defined.

Let $c \in C[2^{n+1}]$ and $b \in h^{-1}(c, c)$. Then $h(2^n b) = (2^n c, 2^n c)$ and $2(2^n c) = 0$. By assumption (2), $2^{n+1} b = 0$ and so $\varphi_2(2^n c) = 2^n b \in B_{n+1}$ as required. Lastly, suppose $c, d \in C[2^{n+1}]$ and $2^n c = 2^n d$. Let $b_1 \in h^{-1}(c, c)$ and $b_2 \in h^{-1}(d, d)$. Then $2(2^{n-1} c - 2^{n-1} d) = 0$, so assumption (2) yields $\varphi_2(2^n c) = 2^n b_1 = 2^n b_2 = \varphi_2(2^n d)$. Thus φ_2 is well-defined.

Step 2 φ_1 and φ_2 are injective.

Let $c \in C[2^n]$ and suppose $\varphi_1(2^{n-1} c) = 0$. If $b \in h^{-1}(c, 0)$ then $h(2^{n-1} b) = (2^{n-1} c, 0)$ and $2^n b = \varphi_1(2^{n-1} c) = 0$. Assumption (3) then shows $2^{n-1} c = 0$. Next, let $c \in C[2^{n+1}]$ and suppose $\varphi_2(2^n c) = 0$. If $b \in h^{-1}(c, c)$ then $\varphi_2(2^n c) = 2^n b = 0$ and $0 = h(2^n b) = (2^n c, 2^n c)$ so that $2^n c = 0$.

We now define:

$$\psi = (\varphi_1, \varphi_2) \colon C_n \oplus C_{n+1} \to B_{n+1} .$$

Step 3 ψ is bijective.

Let $c \in C[2^n]$ and $d \in C[2^{n+1}]$ satisfy $\psi(2^{n-1} c, 2^n d) = 0$. Let $b_1 \in h_1^{-1}(c, 0)$ and $b_2 \in h^{-1}(d, d)$. Then $2^n b_1 + 2^n b_2 = 0$, $h(b_1 + b_2) = (c + d, d)$ and $0 = h(2^n(b_1 + b_2)) = (2^n d, 2^n d)$, as $2^n c = 0$. Thus $\varphi_2(2^n d) = 2^n h^{-1}(d, d) = 0$ and so $2^n d = 0$ by Step 2. Hence $0 = \psi(2^{n-1} c, 0) = \varphi_1(2^{n-1} c)$ and $2^{n-1} c = 0$ also by Step 2. ψ is then injective.

Let $b \in B[2^{n+1}]$ and let $h(b) = (c, d)$. Note that $2^{n+1} c = 0 = 2^{n+1} d$. Let $\varphi_2(2^n d) = 2^n b_0$ and $b_1 = b - b_0$. Then $2^{n+1} b_1 = 0$ and $h(2^n b_1) = (2^n c, 0)$. If $2^n c \neq 0$ then $2^n c \in C_{n+1}$ and $\varphi_1(2^n c) = 2^{n+1} h^{-1}(c, 0) = 2^{n+1} b_1 = 0$. But φ_1 is injective, so we must have $2^n c = 0$. Thus $\varphi_1(2^{n-1} c) = 2^n b_1$. So $2^n b = \varphi_1(2^{n-1} c) + \varphi_2(2^n d) = \psi(2^{n-1} c, 2^n d)$. This completes the proof of Step 3.

We now complete the proof of the lemma. ψ induces

$$\Psi \colon C_n / C_{n+1} \oplus C_{n+1} / C_{n+2} \to B_{n+1} / B_{n+2} .$$

and clearly Ψ is surjective. We check that Ψ is injective. Let $c \in C[2^n]$, $d \in C[2^{n+1}]$ and suppose $\Psi(2^{n-1} c + C_{n+1}, 2^n d + C_{n+2}) = 0 + B_{n+2}$. Then $\varphi_1(2^{n-1} c) + \varphi_2(2^n d) = 2^{n+1} b$ for some $b \in B[2^{n+2}]$. ψ is surjective (for each n) so that $2^{n+1} b = \varphi_1(2^n u) + \varphi_2(2^{n+1} v)$ where $u \in C[2^{n+1}]$ and $v \in C[2^{n+2}]$. Hence $\psi(2^{n-1}(c - 2u), 2^n(d - 2v)) = 0$. Then $2^{n-1}(c - 2u) = 0 = 2^n(d - 2v)$ as ψ is injective. Hence $2^{n-1} c \in C_{n+1}$ and $2^n d \in C_{n+2}$.

The isomorphism Ψ yields $u_{n+1}(B) = u_n(C) + u_{n+1}(C)$ (for $n \geqq 1$). □

Lemma 2.5 *Let W be a reduced Witt ring.*

(1) $U(W) = G_W$

(2) *If W is Noetherian then $U(W/2W) \approx (\mathbb{Z}_2)^{r-1}$, where $r = |X_W|$.*

Proof. (1) If $u \in U(W)$ then $\hat{u}(\alpha) = \pm 1$ for all $\alpha \in X_W$. Thus $\operatorname{sgn}_\alpha d(u)u > 0$ for all α [7, p. 153] implies $d(u)u = 1$ and $u = \langle d(u) \rangle$.

(2) Note that $u + 2W \in U(W/2W)$ iff $\dim u$ is odd (as $W/2W$ is a local ring with maximal ideal $I_W/2W$). Further, if $\dim u$ is odd then $u^2 \equiv 1 \pmod{2W}$ by (1.5). Thus $U(W/2W)$ has exponent 2 and it is enough to show $|U(W/2W)| = 2^{r-1}$.

We proceed by induction on $|G_W|$, using Marshall's classification of reduced, Noetherian Witt rings. First, $|U(\mathbb{Z}/2\mathbb{Z})| = 1$. $U(W_1 \times W_2/2(W_1 \times W_2))$ maps onto $U(W_1/2W_1) \oplus U(W_2/2W_2)$ by $(u, v) + 2(W_1 \times W_2) \mapsto (u + 2W_1, v + 2W_2)$. The kernel is $\{(1, 1) + 2(W_1 \times W_2), (1, -1) + 2(W_1 \times W_2)\}$. By induction $|U(W_1 \times W_2/2(W_1 \times W_2))| = 2 \cdot 2^{r_1 - 1} \cdot 2^{r_2 - 1} = 2^{r_1 + r_2 - 1}$, where $r_1 = |X_{W_1}|$, $r_2 = |X_{W_2}|$ and so $r_1 + r_2 = |X_{W_1 \times W_2}|$.

For $S = W[\Delta]$, $\Delta = \{1, t\}$, the map $(e, u + 2W, v + 2W) \mapsto t^e(u + (v+1)t) + 2S$ is easily seen to be a bijection from $\{\pm 1\} \oplus U(W/2W) \oplus U(W/2W)$ to $U(S/2S)$. Thus $|U(S/2S)| = 2 \cdot 2^{2(r-1)} = 2^{2r-1}$, where $r = |X_W|$ and hence $2r = |X_S|$. □

Corollary 2.6 *Let R be a reduced, Noetherian ring and let $S = R[\Delta]$. Then:*

$$|\mathrm{Pic}(S)| = \frac{2r}{g} |\mathrm{Pic}(R)|^2 ,$$

where $r = |X_R|$ and $g = |G_R|$.

Proof. The sequence (*) of (2.3) yields:

$$\begin{aligned} |\mathrm{Pic}(S)| &= \frac{|\mathrm{Pic}(R)|^2 |U(R/2R)||U(S)|}{|U(R)|^2} \\ &= \frac{|\mathrm{Pic}(R)|^2 \cdot 2^{r-1} \cdot (2g)}{|U(R)|^2} , \quad \text{by (2.5)} \\ &= \frac{2r}{g} |\mathrm{Pic}(R)|^2 . \end{aligned}$$ □

The next lemma is the key to verifying condition (3) of (2.4). Its lengthy proof will be postponed until after the main theorem.

Lemma 2.7 *Suppose $I \subset R$ is a strongly regular ideal and $I^2 = (1 + 2r)$ for some $r \in R$. Then I is principal.*

Theorem 2.8 *Let R be a reduced, Noetherian Witt ring and let $S = R[\Delta]$ with $\Delta = \{1, t\}$. Let $u_i = u_i(\mathrm{Pic}(R))$, $i \geqq 1$, be the Ulm invariants. Then:*

$$\begin{aligned} u_1(\mathrm{Pic}(S)) &= r - (u_2 + u_3 + \ldots) - \log_2 g \\ u_i(\mathrm{Pic}(S)) &= u_i + u_{i-1} \qquad (i \geqq 2) , \end{aligned}$$

where $r = |X_R|$ and $g = |G_R|$.

Proof. We first compute $u_i(\mathrm{Pic}(S))$ for $i \geqq 2$. Let f be the map in (*) of (2.3) and set $A = U(R/2R)/\mathrm{im}(f)$. Then (*) becomes:

$$1 \longrightarrow A \xrightarrow{g} \mathrm{Pic}(S) \xrightarrow{h} \mathrm{Pic}(R) \dot{\oplus} \mathrm{Pic}(R) \longrightarrow 1 .$$

A, $\mathrm{Pic}(S)$ and $\mathrm{Pic}(R)$ are all finite 2-primary abelian groups (cf. (1.11) and [4, 2.9]). We thus need only check the three conditions of (2.4). A has exponent 2 by (2.5), which gives condition (1).

We next check condition (2). Let $b \in \mathrm{Pic}(S)$ and $c \in \mathrm{Pic}(R)$ satisfy $h(b) = (c, c)$ and $c^2 = 1$. We will show $b^2 = 1$. We may assume $c = [I]$, with $I \subset R$ a strongly regular ideal, by (1.3), (1.4). I^2 is principal, say $I^2 = (r)$ for some odd dimensional $r \in R$. Set $J = (\langle 1, -t\rangle R + IS)(\langle 1, t\rangle R + IS)$. Then $J_+ = I = J_-$ and $h[J] = ([I], [I]) = (c, c)$. It suffices to show J^2 is principal since $b^{-1}[J] \in \ker h = g(U(S/2S))$ has exponent 2 by (2.5).

Now r being odd dimensional implies $2R + r \cdot R = R$ (the only maximal ideal of R containing 2 is I_R, while $r \notin I_R$). We apply (2.2): $J^2 = (\langle 1, -t\rangle R + I^2 S)(\langle 1, t\rangle R + I^2 S) = (\langle 1, -t\rangle r, \langle 1, t\rangle r, r^2) = (\langle 1, -t\rangle r, 2r, r^2) = (r)$. Thus $b^2 = [J]^2 = 1$.

Next we check condition (3). Let $b \in \mathrm{Pic}(S)$ and $c \in \mathrm{Pic}(R)$ satisfy $h(b) = (c, 1)$ and $b^2 = 1$. We will show $c = 1$. We may again assume $c = [I]$ with $I \subset R$

a strongly regular ideal. Set $J = \langle 1, -t\rangle R + IS$. Then $J_+ = I$, $J_- = R$ and $h[J] = ([I], [R]) = (c, 1)$. Then $b^{-1}[J] \in \mathrm{im}(g)$ and $[J^2] = 1$ by assumption. Further, $1 = h(b^2) = (c^2, 1)$ so that $I^2 = (r)$ for some odd dimensional $r \in R$. Now, by (2.2), $J^2 = \langle 1, -t\rangle R + (r)S = (\langle 1, -t\rangle, r)$. Thus $g(r + 2R) = [(\langle 1, -t\rangle, r)] = [J]^2 = 1$, by (2.3). The exactness of (*) gives $r + 2R = u + 2R$ for some $u \in U(R)$. Thus $I^2 = (r) = (1 + 2v)$ for some $v \in R$. Then (2.7) implies I is principal and so $c = 1$. This completes the proof that $u_i(\mathrm{Pic}(S)) = u_i + u_{i-1}$ for $i \geqq 2$.

To compute $u_1(\mathrm{Pic}(S))$ set $v_i = u_i(\mathrm{Pic}(S))$ $(i \geqq 1)$. Apply (2.6) to get:

$$\begin{aligned} r + 2\log_2|\mathrm{Pic}(R)| - \log_2 g = \log_2|\mathrm{Pic}(S)| &= \sum_{i\geqq 1} iv_i \\ &= v_1 + \sum_{i\geqq 2} i(u_{i-1} + u_i) \\ &= v_1 + 2\sum_{i\geqq 1} iu_i + \sum_{i\geqq 2} u_i \\ &= v_1 + 2\log_2|\mathrm{Pic}(R)| + \sum_{i\geqq 2} u_i\,. \end{aligned}$$

Hence $v_1 \equiv u_1(\mathrm{Pic}(S)) = r - \left[\sum_{i\geqq 2} u_i\right] - \log_2 g$. □

Remark 2.9 If $I \subset R$ is strongly regular, set $J_1 = \langle 1, -t\rangle R + IS$ and $J_2 = (\langle 1, -t\rangle S + IS)(\langle 1, t\rangle R + IS)$. Suppose $[I] \in \mathrm{Pic}(R)$ has order k. The proof of (2.8) shows $o[J_1] = 2k$ and $o[J_2] = k$.

We re-derive [4, 3.1] since it was used in Section 1.

Corollary 2.10 *Set* $R_n = \mathbb{Z}[\Delta_n]$.

(1) *If* $n \leqq 2$ *then* $\mathrm{Pic}(R_n) = 1$.

(2) *If* $n \geqq 3$ *then* $u_i(\mathrm{Pic}(R_n)) = \begin{cases} \binom{n}{i+2} & \text{if } 1 \leqq i \leqq n-2 \\ 0 & \text{if } i \geqq n-1. \end{cases}$

In particular, the exponent of $\mathrm{Pic}(R_n)$ *is* 2^{n-2}.

Proof. If $n \leqq 2$ then $\mathrm{st}(R_n) \leqq 2$ and (1) follows from (1.17). Let $n \geqq 2$; we prove (2) by induction. Now $r = |X_{R_n}| = 2^n$ and $g = |G_{R_n}| = 2^{n+1}$. Note that the case $n = 2$ is done by (1) and apply (2.8):

$$\begin{aligned} u_1(\mathrm{Pic}(R_{n+1})) &= 2^n - \sum_{i=1}^{n-2}\binom{n}{i+2} - (n+1) \\ &= 2^n - \left(2^n - \binom{n}{0} - \binom{n}{1} - \binom{n}{2} - \binom{n}{3}\right) - (n+1) \\ &= \binom{n}{2} + \binom{n}{3} = \binom{n+1}{3}. \end{aligned}$$

For $2 \leqq i \leqq n-2$ we have

$$u_i(\mathrm{Pic}(R_{n+1})) = u_{i-1}(\mathrm{Pic}(R_n)) + u_i(\mathrm{Pic}(R_n))$$
$$= \binom{n}{i+1} + \binom{n}{i+2} = \binom{n+1}{i+2}.$$

For $i = n-1$ we get $u_i(\mathrm{Pic}(R_{n+1})) = u_{i-1}(\mathrm{Pic}(R_n)) = 1 = \binom{n+1}{i+2}$. And for $i \geqq n$ we get $u_i(\mathrm{Pic}(R_{n+1})) = 0$, as desired. □

To illustrate the computation of $\mathrm{Pic}(R)$ we list the result for all reduced Witt rings with $|G_R| = 2^6$.

R	$\mathrm{Pic}(R)$	R	$\mathrm{Pic}(R)$
$\mathbb{Z}^6$	1	$\mathbb{Z}[\Delta_2] \times \mathbb{Z}[\Delta_2]$	1
$\mathbb{Z}^3 \times \mathbb{Z}[\Delta_2]$	1	$\mathbb{Z}^5[\Delta_1]$	1
$\mathbb{Z}^2 \times \mathbb{Z}^3[\Delta_1]$	1	$(\mathbb{Z}^2 \times \mathbb{Z}[\Delta_2])[\Delta_1]$	$\mathbb{Z}_2$
$\mathbb{Z}^2 \times \mathbb{Z}[\Delta_3]$	$\mathbb{Z}_2$	$(\mathbb{Z} \times \mathbb{Z}^3[\Delta_1])[\Delta_1]$	$(\mathbb{Z}_2)^2$
$\mathbb{Z} \times \mathbb{Z}^4[\Delta_1]$	1	$(\mathbb{Z} \times \mathbb{Z}[\Delta_3])[\Delta_1]$	$(\mathbb{Z}_2)^4 \oplus \mathbb{Z}_4$
$\mathbb{Z} \times (\mathbb{Z} \times \mathbb{Z}[\Delta_2])[\Delta_1]$	$\mathbb{Z}_2$	$\mathbb{Z}^4[\Delta_2]$	$(\mathbb{Z}_2)^3$
$\mathbb{Z} \times \mathbb{Z}^3[\Delta_2]$	$(\mathbb{Z}_2)^2$	$(\mathbb{Z} \times \mathbb{Z}[\Delta_2])[\Delta_2]$	$(\mathbb{Z}_2)^5 \oplus \mathbb{Z}_4$
$\mathbb{Z} \times \mathbb{Z}[\Delta_4]$	$(\mathbb{Z}_2)^4 \oplus \mathbb{Z}_4$	$\mathbb{Z}^3[\Delta_3]$	$(\mathbb{Z}_2)^7 \oplus (\mathbb{Z}_4)^2$
		$\mathbb{Z}[\Delta_5]$	$(\mathbb{Z}_2)^{10} \oplus (\mathbb{Z}_4)^5 \oplus \mathbb{Z}_8$

As a last corollary we give a local version of (1.16). Recall the *local stability index* of Schwartz [11]: if $\alpha \in X_R$ then $\mathrm{st}(\alpha, R) = \sup\{\log_2 |X_R/F| : F$ a fan, $\alpha \in X_R/F\}$.

Corollary 2.11 *Let R be an arbitrary Witt ring (not necessarily Noetherian). Let $\alpha \in X_R$ and set $I = P(\alpha, 5)$. If $\{\alpha\}$ is open in X_R and $s = \mathrm{st}(\alpha, R)$ is finite then the order of $[I] \in \mathrm{Pic}(R)$ is 2^{s-2} (unless $s \leqq 1$ in which case $[I] = 1$).*

Proof. We may assume R is reduced by (1.6). We first show the result holds if R is Noetherian, using induction on $|X_R|$ and Marshall's classification theorem. Clearly the result holds if $R = \mathbb{Z}(|X_R| = 1)$. If $|X_R| > 1$ then R is either a fiber product or a group ring.

First suppose $R = S \times T$. We may assume $\alpha(1 \times G_T) = 1$ and so $\alpha | G_S \times 1 = \alpha_0 \in X_S$. Set $I_0 = P(\alpha_0, 5) \subset S$. Every fan of R is of the form $F_1 \times G_T$ or $G_S \times F_2$ with F_1 and F_2 fans of S and T respectively. Hence:

$$s = \mathrm{st}(\alpha, R) = \sup\{\log_2 |X_R/F| : F \text{ a fan}, \alpha \in X_R/F\}$$
$$= \sup\{\log_2 |X_R/F_1 \times G_T| : F_1 \text{ a fan of } S, \alpha_0 \in X_S/F_1\}$$
$$= \mathrm{st}(\alpha_0, S).$$

Under the isomorphism $\mathrm{Pic}(R) \to \mathrm{Pic}(S) \oplus \mathrm{Pic}(T)$ of (2.1), $[I]$ is sent to $([I_0], 1)$. Thus $o[I] = o[I_0] = 2^{s-2}$, by induction.

Next suppose $R = S[\Delta]$ with $\Delta = \{1, t\}$. We may assume $\alpha(t) = 1$ and so $\alpha | G_S = \alpha_0 \in X_S$. Again set $I_0 = P_0(\alpha_0, 5) \subset S$; note that $I = I_0 R + \langle 1, -t\rangle R$. Each fan $F \subset G_S$ is again a fan in G_R and each fan $F \subset G_R$ represents to a fan $F \cap G_S$. Hence:

$$\begin{aligned} \mathrm{st}(\alpha, R) &= \sup\{\log_2 |X_R/F| : F \text{ a fan of } R,\ \alpha \in X_R/F\} \\ &= \sup\{\log_2 (2|X_S/F|) : F \text{ a fan of } S,\ \alpha_0 \in X_S/F\} \\ &= \mathrm{st}(\alpha_0, S) + 1 . \end{aligned}$$

Under the map $\mathrm{Pic}(R) \to \mathrm{Pic}(S) \oplus \mathrm{Pic}(S)$, $[I]$ is sent to $([I_0], 1)$ and so $o[I] = 2(o[I_0])$ by (2.9). Then by induction:

$$o[I] = 2(o[I_0]) = 2 \cdot 2^{\mathrm{st}(\alpha_0, S) - 2} = 2^{s-2} .$$

This completes the proof when R is Noetherian.

Now let R be an arbitrary (reduced) Witt ring. Choose a fan $T \subset G_R$ such that $s = \log_2 |X_R/T|$ and $\alpha \in X_R/T$. The image of I under $R \to R_T$ is $P_T(\alpha, 5)$. Now $\mathrm{st}(\alpha, R_T) = s$, R_T is Noetherian and $\mathrm{Pic}(R) \to \mathrm{Pic}(R_T)$ is surjective by (1.15). Hence $o[I] \geqq o[I_T] = 2^{s-2}$.

Further, $s = \inf\{n | \alpha \in H(a_1, \ldots, a_n)\}$ by [11]. Choose $a_1, \ldots, a_s \in G_R$ with $\alpha \in H(a_1, \ldots, a_s)$. We may write $5^{2^{s-2}} = 1 + 2^s k$ (that is, 5 generates $U(\mathbb{Z}/2^s\mathbb{Z})/\{\pm 1\}$, cf. [6]). We claim that $I^{2^{s-2}} = (1 + k\langle\langle a_1, \ldots, a_s\rangle\rangle)$; set $\omega = 1 + k\langle\langle a_1, \ldots, a_s\rangle\rangle$. Now $\omega \in P(\alpha, 1 + 2^s k) = I^{2^{s-2}}$. And if $\varphi \in P(\alpha, 1 + 2^s k)$ then $\hat{\omega} | \hat{\varphi}$ and so $\varphi \in (\omega)$ by (1.7). Thus $[I^{2^{s-2}}] = 1$ in $\mathrm{Pic}(R)$ and $o[I] \leqq 2^{s-2}$. Hence $o[I] = 2^{s-2}$. □

Remark. Corollary 2.11 can be slightly extended. Suppose $\{\alpha\}$ is open and $s = \mathrm{st}(\alpha, R)$ is finite. Then, as in [4, 2.8], $H(\alpha) = \{[P(\alpha, m)] | m \in \mathbb{Z}\} \subset \mathrm{Pic}(R)$ is a cyclic subgroup generated by $[P(\alpha, 5)]$. Indeed $H(\alpha)$ is isomorphic to $G \equiv U(\mathbb{Z}/2^s\mathbb{Z})/\{\pm 1\}$. If the order of m in G is k then $o[P(\alpha, m)] = k$.

To complete the proof of (2.8) we at last prove (2.7). The proof proceeds in four steps.

Lemma 2.12 *Let R be Noetherian and reduced. Suppose $\omega \in R$ and $y \in G_R$ satisfy $\langle 1\rangle + \langle y\rangle\omega^2 \in (4)$. Then $\langle x\rangle - \omega \in (2)$ for some $x \in G_R$.*

Proof. We induct on $|G_R|$, using Marshall's classification theorem. The result is clear for $R = \mathbb{Z}$ (ω must be odd, so take $x = 1$). Suppose $R = R_1 \times R_2$, $\omega = (u_1, u_2) \in R$ and $y = (y_1, y_2) \in G_{R_1} \times G_{R_2}$. Then $\langle 1\rangle + \langle y_i\rangle u_i \in 4R_i$ for $i = 1, 2$. By induction there exist $x_i \in G_{R_i}$ $(i = 1, 2)$ with $\langle x_i\rangle - u_i \in 2R_i$. Note that $\dim u_1 \equiv \dim u_2 \pmod 2$. Thus either $\dim u_1 \equiv \dim u_2 \pmod 4$ or $\dim u_1 \equiv 2 + \dim u_2 \pmod 4$. In the first case, $(\langle x_1\rangle - u_1, \langle x_2\rangle - u_2) \in 2R$. In the second case, $(\langle x_1\rangle - u_1, \langle -x_2\rangle - u_2) \in 2R$, since $\langle -x_2\rangle - u_2 = \langle x_2\rangle - u_2 - 2\langle x_2\rangle$ has the same dimension mod 4 as $\langle x_1\rangle - u_1$. Thus we take $x = (x_1, x_2)$ or $(x_1, -x_2)$.

Lastly, suppose $R = R_0[\Delta]$ with $\Delta = \{1, t\}$. Let $\omega = r + st$ with $r, s \in R_0$. Then $\langle 1\rangle + \langle y\rangle\omega^2 = \langle 1\rangle + \langle y\rangle(r^2 + s^2) + 2\langle yt\rangle rs$. If $y \in tG_{R_0}$ then $\langle 1\rangle + \langle y\rangle\omega^2 = 4(u + vt)$ implies $\langle 1\rangle + 2\langle yt\rangle rs = 4u$, which is impossible. Thus $y \in G_{R_0}$, $\langle 1\rangle + \langle y\rangle(r^2 + s^2) \in 4R_0$ and $rs \in 2R_0$. Now one of r, s is odd dimensional, say r. Then $s \in 2R_0$ and so $\langle 1\rangle + \langle y\rangle r^2 \in 4R_0$. By induction, $\langle x\rangle - r \in 2R_0$ for some $x \in G_{R_0}$. Hence $\langle x\rangle - \langle r + st\rangle \in 2R$ as desired. The case when s is odd dimensional is similar. □

Lemma 2.13 *Let R be reduced and Noetherian. Let $\omega_1, \omega_2 \in R$ be odd dimensional and let $y \in G_R$. Suppose $\omega_1^2 + \langle y\rangle\omega_2^2 \in (4)$. Then $\langle x\rangle\omega_1 - \omega_2 \in 2R$ for some $x \in G_R$.*

Proof. We have $\omega_1^2 - 1 \in 2R$, by (1.5), so $\omega_1^4 + \langle y\rangle\omega_1^2\omega_2^2 \in (4)$ implies $\langle 1\rangle + \langle y\rangle(\omega_1\omega_2)^2 \in (4)$. Thus $\langle x\rangle - \omega_1\omega_2 \in (2)$ for some $x \in G_R$ by (2.12). That is, $\langle x\rangle\omega_1 - \omega_1^2\omega_2 \in (2)$ and so $\langle x\rangle\omega_1 - \omega_2 \in (2)$. □

Lemma 2.14 *Let R be reduced and Noetherian and let $S = R[\Delta]$ with $\Delta = \{1, t\}$. Let $J \subset S$ be a strongly regular ideal. Suppose J_+ and J_- are principal ideals of R and that $J^2 = (1 + 2c)$ for some $c \in S$. Then J is principal.*

Proof. Let $J_+ = (\omega_1)$ and $J_- = (\omega_2)$, with $\omega_1, \omega_2 \in R$ odd dimensional. $J = (\langle 1, -t\rangle R + \omega_1 S)(\langle 1, t\rangle + \omega_2 S)$ and $J^2 = (\langle 1, t\rangle\omega_1^2, \langle 1, -t\rangle\omega_2^2, \omega_1^2\omega_2^2)$ by (2.2). Note that $\omega_1^2 \equiv \omega_2^2 \equiv 1 \pmod{2R}$, by (1.5). Write $\omega_1^2 - \omega_2^2 = 2r$ for some $r \in R$. Then we claim $J^2 = (\langle 1, t\rangle r + \omega_2^2)$. We first observe $(\langle 1, t\rangle r + \omega_2^2)_+ = (2r + \omega_2^2) = (\omega_1^2)$ and $(\langle 1, t\rangle r + \omega_2^2)_- = (\omega_2^2)$. Thus by (2.2), $(\langle 1, t\rangle r + \omega_2^2) = (\langle 1, -t\rangle R + \omega_1^2 S)(\langle 1, t\rangle R + \omega_2^2 S) = J^2$.

We have then that $(\langle 1, t\rangle r + \omega_2^2) = (1 + 2c)$ and so $\langle g\rangle(1 + 2c) = \langle 1, t\rangle r + \omega_2^2$ for some $g \in G_S$. Hence $2\langle g\rangle(1 + 2c) = \langle 1, t\rangle(\omega_1^2 - \omega_2^2) + 2\omega_2^2 = (\omega_1^2 + \omega_2^2) + t(\omega_1^2 - \omega_2^2)$. Let $c = c_0 + c_1 t$ with $c_0, c_1 \in R$. If $g \in G_R$ we obtain:

$$\langle g\rangle(2 + 4c_0) = \omega_1^2 + \omega_2^2 \quad \text{and} \quad 4\langle g\rangle c_1 = \omega_1^2 - \omega_2^2 .$$

If $g \in tG_R$, say $g = th$, then we obtain:

$$\langle h\rangle(2 + 4c_0) = \omega_1^2 - \omega_2^2 \quad \text{and} \quad 4\langle h\rangle c_1 = \omega_1^2 - \omega_2^2 .$$

In either case, $\omega_1^2 + \langle y\rangle\omega_2^2 \in (4)$ for some $y \in G_R$. Thus $\langle x\rangle\omega_1 - \omega_2 \in (2)$ for some $x \in G_R$ by (2.13).

We may thus replace ω_1 by $\langle x\rangle\omega_1$ in the representation of J. We have then $J = (\langle 1, -t\rangle R + \omega_1 S)(\langle 1, t\rangle R + \omega_2 S)$ with $\omega_1 - \omega_2 = 2v$ for some $v \in R$. Note that $(\langle 1, t\rangle v + \omega_2)_+ = (2v + \omega_2) = (\omega_1)$ and $(\langle 1, t\rangle v + \omega_2)_- = (\omega_2)$. Applying (2.2) once more shows $J = (\langle 1, t\rangle v + \omega_2)$ is principal. □

Proof of (2.7). Suppose the result is false and choose R for which the result fails with $|G_R|$ minimal. Clearly $R \neq \mathbb{Z}$, so R is a fiber product or a group ring. Suppose $R = R_1 \times R_2$. For a strongly regular ideal $I \subset R$ set $I_1 = \{a \in R_1 \mid (a, s) \in I, \text{ for some } s \in R_2\}$ and $I_2 = \{b \in R_2 \mid (r, b) \in I, \text{ for some } r \in R_1\}$. We are assuming that $I^2 = (1 + 2r)$ for some $r = (r_1, r_2) \in R$. Then I^2 is the fiber product over $\mathbb{Z}_2$ of $(1 + 2r_1)R_1$ and $(1 + 2r_2)R_2$. It is simple to verify that $I_1^2 = (1 + 2r_1)R_1$ and $I_2^2 = (1 + 2r_2)R_2$ (even though in general I is not $I_1 \times I_2$). By induction, $I_1 = c_1 R_1$ and $I_2 = c_2 R_2$. Then $I = (c_1, c_2)R$ is principal.

Suppose then $R = R_0[\Delta]$ with $\Delta = \{1, t\}$. Let $I \subset R$ be strongly regular and assume I is not principal while $I^2 = (1 + 2c)$. Write $I = \langle 1, -t\rangle I_- + \langle 1, t\rangle I_+ + I_- I_+$ as in (2.2). Then $(1 + 2c) = \langle 1, -t\rangle(I_-)^2 + \langle 1, t\rangle(I_+)^2 + I_-^2 I_+^2$. Let $c = c_0 + c_1 t$ with $c_0, c_1 \in R$. Then $(I^2)_- = (1 + 2(c_0 + c_1)) = (I_-)^2$ and $(1 + 2(c_0 - c_1)) = (I_+)^2$. By the minimality of $|G_R|$ we have that I_- and I_+ are principal. But then (2.14) implies I is principal, a contradiction.

Acknowledgement. I wish to thank the (anonymous) referee of [4] who not only suggested that K-theory was a more natural setting for those results but also gave several excellent ideas on how to implement the change of settings.

References

1. Bass, H.: Algebraic K-theory. New York: Benjamin 1968
2. Elman, R., Lam, T.Y.: Quadratic forms over formally real fields and Pythagorean fields. Am. J. Math. **94**, 1155–1194 (1972)
3. Fitzgerald, R.: Primary ideals in Witt rings. J. Algebra **96**, 368–385 (1985)
4. Fitzgerald, R.: Ideal class groups of Witt rings. J. Algebra (to appear)
5. Griffin, M.: Valuations and Prüfer rings. Can. J. Math. **26**, 412–429 (1974)
6. Ireland, K., Rosen, M.: A Classical Introduction to Modern Number Theory. Grad. Texts in Math., No. 84, Berlin Heidelberg New York Springer: 1982
7. Knebusch, M., Rosenberg, A., Ware, R.: Structure of Witt rings and quotients of abelian group rings. Am. J. Math. **94**, 119–155 (1972)
8. Lam, T.Y.: Orderings, Valuations and Quadratic Forms. Regional Conference Series in Math., No. **52**, Am. Math. Soc. 1983
9. Marshall, M.: Classification of finite spaces of orderings. Can. J. Math. **31**, 320–330 (1979)
10. Marshall, M.: Abstract Witt Rings. Queen's Pap. Pure Appl. Math., No. **57** (1980)
11. Schwartz, N.: Local stability and saturation in spaces of orderings. Can. J. Math. **35**, 454–477 (1983)
12. Silvester, J.R.: Introduction to Algebraic K-theory. London New York: Chapman and Hall 1981

Math. Z. 206, 321–331 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

On the existence of area-minimizing surfaces with free boundary*

Rugang Ye
Department of Mathematics, Stanford University, Stanford, CA 94305, USA

Received July 4, 1989, in final form January 22, 1990

0 Introduction

Let $\mathcal{N}$ be a complete Riemannian manifold of dimension $n \geqq 3$ and let $\mathcal{M} \subset \mathcal{N}$ be an embedded submanifold of dimension $m \geqq 2$ (this includes the case $\mathcal{M} = \partial\mathcal{N}$). A (2-dimensional) minimal surface $S \subset \mathcal{N}$ is said to have free boundary on $\mathcal{M}$, provided that $\partial S \subset \mathcal{M}$ and S meets $\mathcal{M}$ orthogonally along ∂S. $\mathcal{M}$ is called a "supporting submanifold" for S. A basic question is, given a free homotopy class of closed curves on $\mathcal{M}$, whether one can find a minimal surface with free boundary belonging to this class. Or one can just ask whether there exists a minimal surface whose free boundary is homotopically nontrivial on $\mathcal{M}$ (assuming necessary topological conditions, e.g. $\pi_1(\mathcal{M}) \neq 1$). In this paper we restrict to surfaces of disk type, although our methods apply to general types, too. We adopt the standard procedure of minimizing the Dirichlet integral among parametric surfaces. Choose a weakly convergent minimal sequence f_k for the Dirichlet integral. The basic difficulty is how to control the boundary values of f_k so that their homotopy condition is retained in the limit. In order to get around this difficulty, Courant [C] employed linking numbers to formulate topological conditions on boundary values and replaced competing surfaces by competing "surface sequences". (This idea has been further developed by Davids [D], Hildebrandt [Hi] and Harth [Ha].) The disadvantage of this approach is that linking numbers as homology invariants cannot fully capture homotopy. In a general ambient space, linking numbers can not even fully capture homology.

A second approach was introduced by Meeks and Yau [MY] in their proof of the *geometric loop theorem*. They approximate the above f_k's by surfaces with Jordan curves as boundary and then replace them by solutions of the Plateau problem for these Jordan curves. This approach yields very general existence results, but does not seem to work for the problem of prescribing boundary homotopy classes. On the other hand, this approach requires a geometric condition on the ambient manifold $\mathcal{N}$ which is much stronger than Morrey's homogeneous regularity.

* Research partially supported by NSF grant # DMS87-03537

In this paper we adopt a new approach introduced recently by Tolksdorf in [T]. The first step of this approach is to minimize the Dirichlet integral under the side condition that the boundary lengths of the competing surfaces do not exceed a bound L. The obtained minimizers may be called "L-minimizers". This step was done in [T] for euclidean ambient spaces; we carry the arguments over to general homogeneously regular manifolds. (This is the content of Lemma 6 below.) The crucial second step is how to get limits from L-minimizers as $L \to \infty$. Here Tolksdorf's arguments are incorrect. He tried to show that if L, L' are both large enough, then L-minimizers are also L'-minimizers. His arguments are based on an estimate for the Dirichlet integral of a comparison surface obtained by deforming an L-minimizer. But this estimate is completely wrong. We believe that Tolksdorf's idea for the second step in principle cannot work.

We establish the second step in the following way. We first observe that as $L \to \infty$, L-minimizers may only develop singularities at finitely many points along boundary. These singular points can be characterized fairly explicitly. Next we show that singularities lead to splitting of surfaces and decomposition of boundary homotopy classes. After enough steps of decompositions singularities will not occur and we can derive *a priori* Hölder estimates (independent of L) for L-minimizers, therewith we are done. (One notices that decompositions of homotopy classes also occur in the first step.) We are able to prove the following new theorem (already formulated in [T]): *if a free homotopy class satisfies the "Douglas condition", then there exists an area-minimizing surface with free boundary in it.* Interestingly, a complete existence theory for area-minimizing surfaces (of disk type) with free boundary can then be built upon this theorem. Some results of this theory are already contained in the work of Meeks–Yau, but we can cover all homogeneously regular manifolds. We also present *a priori* estimates for minimizing disks. We mostly focus on compact supporting submanifolds, but also indicate how to extend to noncompact ones.

The material of this paper is taken from the author's thesis [Y] written under the direction of Professor S. Hildebrandt. We are grateful to him for constant support, stimulation and encouragement.

1 Preliminaries

We assume $\mathcal{N}, \mathcal{M} \in \mathbb{C}^{k,\mu}$ for some $k \geqq 2$ and $\mu \in (0, 1)$, although the smoothness assumption on $\mathcal{M}$ can be much weakened by a simple approximation argument. To simplify presentation we only treat the case that $\mathcal{N}$ and $\mathcal{M}$ are both boundaryless. We further assume that the kernel $\ker i_*$ of the inclusion $i_*: \pi_1(\mathcal{M}) \to \pi_1(\mathcal{N})$ is nontrivial.

We define

$$H = \{f \in H^{1,2}(B, \mathcal{N}) \cap C^0(\partial B, \mathcal{N}) : f(\partial B) \subset \mathcal{M}\},$$

where B denotes the open unit disk in $\mathbb{R}^2$ and $H^{1,2}$ means the Sobolev space as defined in [M1], [M2]. If $G \neq \phi$ is a conjugacy invariant subset of $\ker i_*$ with $0 \notin G$, then we put $H_G = \{f \in H : [f|_{\partial B}] \in G\}$, where $[f|_{\partial B}]$ means the homotopy class of $f|_{\partial B}$. Since G is conjugacy invariant, the condition $[f|_{\partial B}] \in G$ does not depend on the choice of a base point for $\pi_1(\mathcal{M})$. Let $A(f) =$ the area of f, $D(f) =$ the Dirichlet integral of $f (= \frac{1}{2}\int_B \|df\|^2)$, $A_G = \inf\{A(f) : f \in H_G\}$,

$D_G = \inf\{D(f): f \in H_G\}$. Then $A_G = D_G[N]$. A mapping $f \in H_G$ with $D(f) = D_G$ (hence $A(f) = A_G$) will be called a minimizing disk or, more exactly, a minimizing G-disk.

Definition 1 $\mathcal{M}$ is called *uniform*, if there is some $\varepsilon_o > 0$ such that $\mathcal{M}$ is a retract of its open ε_0-neighborhood $\mathcal{O}_{\varepsilon_o}$ (i.e., there is a continuous surjection $\pi: \mathcal{O}_{\varepsilon_o} \to \mathcal{M}$ such that $\pi(p) = p$ for each $p \in \mathcal{M}$) and if $\mathcal{M}$ satisfies a chord-arc condition, i.e., there are $d_o > 0$, $A_o > 0$ such that if $\operatorname{dist}(p_1, p_2) \leqq d_o$, $p_1, p_2 \in \mathcal{M}$ (distance being in $\mathcal{N}$) then there is a rectifiable curve in $\mathcal{M}$ of length $\leqq A_0 \operatorname{dist}(p_1, p_2)$ connecting p_1, p_2.

In the sequel we shall fix ε_o, d_o, A_o and suppress the dependence of various estimates on ε_o, d_o, A_o and other geometrical quantities of $\mathcal{M}$ and $\mathcal{N}$. We note that a compact supporting submanifold is uniform.

Definition 2 $\mathcal{N}$ is called *uniformly Lipshitz* if there exists a constant $C > 0$ such that for each point $p \in \mathcal{N}$, there is a bi-Lipschitz homeomorphism of a neighborhood of p onto the unit ball in $\mathbb{R}^3$ with Lipschitz constants less than C and p mapped to o. The infinum of such C is called the *Lipschitz constant* of $\mathcal{N}$.

Remark. This is a reformulation of C.B. Morrey's "homogeneously regular" condition [M1], [M2]. We feel that "uniformly Lipschitz" is a more appropriate name.

Definition 3 $\mathcal{N}$ has *bounded metric*, iff there are constants $\lambda > 0$, $\Lambda > 0$ such that for each point $p \in \mathcal{N}$, there exists a C^2-coordinate chart (φ, U) around p such that $\varphi(U)$ is the unit ball in $\mathbb{R}^3$, $\varphi(p) = o$ and on this ball $\lambda|\xi|^2 \leqq g_{ij}\xi^i\xi^j \leqq \Lambda|\xi|^2$ ($\xi \in \mathbb{R}^3$ arbitrary), $|\partial g_{ij}/\partial x^k| \leqq \Lambda$, where g_{ij} are the metric coefficients on this chart.

Remark. This is the condition which Meeks–Yau [MY] used. Note that if $\mathcal{N}$ has bounded geometry, i.e., the injectivity radius of $\mathcal{N}$ is positive and the sectional curvature of $\mathcal{N}$ is bounded, then $\mathcal{N}$ has bounded metric, see [JK].

$B_r(w)$ will denote the open disk of radius r and center w in $\mathbb{R}^2 \simeq \mathbb{C}$. If $w \in \partial B$, then $C_r(\omega) = \partial B_r(w) \cap B$, $C_r^*(w) = \partial B_r(w) \setminus C_r(w)$, $I_r(w) = \partial B \cap B_r(w)$.

Lemma 1 *Suppose that $\mathcal{N}$ is uniformly Lipschitz and $\mathcal{M}$ is compact. Then there exists $a_o > 0$ such that $D(f) \geqq a_o$ whenever $f \in H^{1,2}(B, \mathcal{N}) \cap C^o(B)$, $f|_{\partial B}(\partial B) \subset \mathcal{O}_{\varepsilon_o/2}$, $f(w) \notin \mathcal{O}_{\varepsilon_o}$ for some $w \in B$ and f has the following minimizing property: $D(f) \leqq D(g)$ for any $g \in H(B, \mathcal{N}) \cap C^0(B)$ with $\{f \neq g\} \subset B$.*

Proof. Assume that there were $f_1, f_2, \ldots$ satisfying the above condition such that $D(f_k) \to 0$. We may assume $f_k(0) \notin \mathcal{O}_{\varepsilon_0}$ (performing a conformal reparametrization if necessary). Since $f_k(\partial B) \subset \mathcal{O}_{\varepsilon_o/2}$ and $\mathcal{O}_{\varepsilon_o/2}$ is compact, the assumption $D(f_k) \to 0$ implies that there is a subsequence $f_{k'}$ converging $H^{1,2}$-weakly to a constant map $f \equiv p \in \bar{\mathcal{O}}_{\varepsilon_o/2}$ [M1]. Since each f_k minimizes the Dirichlet integral, by Morrey's classical result [M2] we have a uniform estimate for the Hölder seminorms (of some fixed exponent) of f_k on $\overline{B_{1/2}(0)}$. Hence we may assume that on $\overline{B_{1/2}(0)}$, $f_{k'}$ converges uniformly to f. In particular, $p = \lim f_{k'}(0) \notin \bar{\mathcal{O}}_{\varepsilon_o/2}$, a contradiction.

QED

Lemma 2 *Let $\mathcal{N}$ have bounded metric. Then for each $\varepsilon > 0$ there exists an $a_\varepsilon > 0$ such that $D(f) \geqq a_\varepsilon$ holds for each $f \in C^0(\bar{\Omega}, \mathcal{N}) \cap C^2(\Omega)$ which is minimal (i.e. conformal and harmonic) and satisfies $dist(f(w), f(\partial\Omega)) \geqq \varepsilon$ for some $w \in \Omega$, where Ω denotes a plane domain.*

Proof. This follows from a monotonicity formula for the Dirichlet integral, see [MY] or [GHN] for details. QED

Lemma 3 *Assume either that $\mathcal{N}$ is uniformly Lipschitz, $\mathcal{M}$ is compact or that $\mathcal{N}$ has bounded metric, $\mathcal{M}$ is uniform. Then there is an $a_1 > 0$ such that for any G (as before) $A_G \geqq a_1$.*

Proof. In the first case, for any $g \in H_G$ we choose some $f \in H_G$ such that $f|_{\partial B} = g|_{\partial B}$ and $D(f) \leqq D(g')$ for all $g' \in H_G$ with $g'|_{\partial B} = g|_{\partial B}$. Then apply Lemma 1. In the second case, for any $g \in H_G$, we choose some f as in the first case. According to Theorem 9.4.5 in [M2], $f \in C^0(\bar{B}, \mathcal{N}) \cap C^2(B)$. Then for r near 1, $f|_{\partial B_r(0)}$ has image in $\mathcal{O}_{\varepsilon_0/2}$ and is freely homotopic to $f|_{\partial B}$ in $\mathcal{O}_{\varepsilon_0/2}$. We approximate $f|_{\partial B_r(0)}$ by C^2-Jordan loops $\tau: \partial B \to \mathcal{O}_{2\varepsilon_0/3}$ and then choose a solution $\tilde{f}$ of the Plateau problem for τ. If image $\tilde{f} \subset \mathcal{O}_{\varepsilon_0}$, then τ is contractable in $\mathcal{O}_{\varepsilon_0}$. Then $g|_{\partial B}$ would be contractable in $\mathcal{M}$, contradicting the assumption $0 \notin G$. Hence Lemma 2 implies that $D(\tilde{f}) \geqq a_{\varepsilon_0/3}$. QED

For later applications we now formulate the following.

Replacement procedure. Let $\mathcal{N}$ be uniformly Lipschitz and $\mathcal{M}$ be uniform. Consider a mapping $f \in H$ and a point $w \in \partial B$. For a given $r \in (0, 1)$, let w_1, w_2 denote the two ends of the arc $C_r(w)$. Let $\tau: C_r^*(w) \to \mathcal{M}$ be a rectifiable curve running from $f(w_1)$ to $f(w_2)$ as the parameter runs from w_1 to w_2. We assume that τ is parametrized proportional to arc-length (using polar coordinates for $\partial B_r(w)$) with length $L(\tau) \leqq A_0 \operatorname{dist}(f(w_1), f(w_2))$. We note the following fact ([M2] or [MY]): there exists $\bar{d}_1 > 0$ so that if $\operatorname{diam} f(C_r(w)) \leqq \bar{d}_1$ and $\int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta < \infty$, then there is a map $h \in H^{1,2}(B_r(w), \mathcal{N}) \cap C^0(\overline{B_r(w)}, \mathcal{N})$ with $h|_{C_r(w)} = f|_{C_r(w)}$, $h|_{C_r^*(w)} = \tau$ and

$$\int_{B_r(w)} \|dh\|^2 \leqq C \int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta \tag{1.1}$$

for some constant C depending only on the Lipschitz constant of $\mathcal{N}$. (We also notice that the inequality $\int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta < \infty$ is true for almost every r and implies that $f|_{\overline{C_r(w)}}$ is absolutely continuous.)

Assuming that such an h is present we construct a "big" map $f^1 \in H$ and a "small" map $f^2 \in H$. f^1 is first defined on $B \cup B_r(w)$ via $f^1|_{B \setminus B_r(w)} = f|_{B \setminus B_r(w)}$, $f^1|_{B_r(w)} = h$ and then pulled back to B by a conformal mapping from B onto $B \cup B_r(w)$. To define f^2, we consider the disjoint union of $\Omega_1 = \bar{B}_r(w) \cap B$ and $\Omega_2 = B_r \cup C_r$. Gluing them along $C_r(w)$ we get a Riemann surface of disk type. f^2 is first defined on this Riemann surface via $f^2|_{\Omega_1} = f|_{\Omega_1}, f^2|_{\Omega_2} = h$ and then pulled back to B by a conformal mapping. We have

$$D(f) = D(f^1) + D(f^2) - \frac{1}{2} \int_{B_r(w)} \|dh\|^2 \tag{1.2}$$

and

$$\alpha = \alpha_1 \alpha_2 \tag{1.3}$$

where α, α_1 and α_2 are conjugacy classes in $\pi_1(\mathscr{M})$ generated by $f|_{\partial B}, f^1|_{\partial B}$ and $f^2|_{\partial B}$ respectively.

We choose $d_1 < \min(\bar{d}_1, \varepsilon_0/2)$ small enough so that the following holds

1) a connecting curve τ as above exists and α_1, α_2 are independent of the choice of τ whenever $\operatorname{dist}(f(w_1), f(w_2)) \leqq d_1$,
2) the loop consisting of $f|_{C_r(w)}$ and the reverse of τ is contractible in O_{ε_o} (this is equivalent to saying that $f|_{\bar{C}_r(w)}$ is homotopic (with two ends fixed) to τ in O_{ε_o}) whenever $f|_{\bar{C}_r(w)}$ is continuous and $\operatorname{diam} f(C_r(w)) \leqq d_1$.

Proposition 1 1) *All minimizing disks of class $C^o(\bar{B})$ belong to $C^{k,\mu}(\bar{B})$ and to $C^\omega(\bar{B})$ if moreover $\mathscr{N}, \mathscr{M}$ are C^ω. They are minimal surfaces (i.e. conformal and harmonic) and are orthogonal to $\mathscr{M}$ along ∂B.*

2) *Suppose either that $\mathscr{N}$ is uniformly Lipschitz, $\mathscr{M}$ is compact or that $\mathscr{N}$ has bounded metric and $\mathscr{M}$ is uniform. Then all minimizing disks are of class $C^o(\bar{B})$.*

Proof. 1) is standard, we refer to [Jä], [HN], and [L]. To prove 2), let us fix G and a minimizing G-disk f. According to Morrey [M2], we have an interior Dirichlet growth condition: for any $w \in B$ and $B_r(w) \subset B$, there holds $\varphi(r) \leqq Cr^{2\sigma}$, where $\varphi(r) = \int_{B_r(w)} \|df\|^2$, and $C > 0, \sigma > 0$ depend only on A_G and the Lipschitz constant of $\mathscr{N}$. Now we choose $r_o \in (0, 1)$ such that $\int_{B \setminus B_r} \|df\|^2 < 2a_o$ in the first case and $< 2a_{\varepsilon_o/3}$ in the second case ($a_o, a_{\varepsilon_o/3}$ are given by Lemma 1 and Lemma 2 respectively). For an arbitrary $w \in \partial B$ and $r \in (0, 1 - r_0)$, we put $\varphi(r) = \int_{B \cap B_r(w)} \|df\|^2$. If $\int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta > \frac{d_1^2}{\pi}$, then $\varphi(r) \leqq \frac{2\pi D(f)}{d_1^2} \int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\| d\theta \leqq \frac{2\pi D(f)}{d_1^2} \varphi'(r) r$. If $\int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta \leqq \frac{d_1^2}{\pi}$, then $\operatorname{diam} f(C_r(w)) \leqq d_1$, hence there is a connecting curve τ as in Replacement Procedure homotopic (with two fixed ends) to $f|_{\bar{C}_r(w)}$ inside O_{ε_o}. On the other hand, $f|_{\bar{C}_r(w)}$ is homotopic (with two fixed ends) to $f|_{\bar{C}_r(w)}$ inside O_{ε_o}. Otherwise, by the arguments in the proof of Lemma 3 we would have $\int_{B \cap B_r(w)} \|df\|^2 \geqq 2a_0$ in the first case and $\geqq 2a_{\varepsilon_o/3}$ in the second case, a contradiction. Thus, the big map $f^1 \in H_G$. It follows that $\varphi(r) \leqq C \int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta \leqq C\varphi'(r) r$. Choosing $C_1 = \max\left(C, \frac{2\pi D(f)}{d_1^2} \right)$ and integrating the inequality $\varphi(r) \leqq C_1 \varphi'(r) r$ we get a Dirichlet growth condition at boundary. It follows from Morrey's Dirichlet growth theorem [M2] that f is Hölder continuous on $\bar{B}$. QED

Remark. Consider a continuous minimizing G-disk f such that $f(B \setminus B_{r_0}(0)) \subset \mathscr{O}_{\varepsilon_o}$ for some $r_0 > 0$. Then we have a Hölder seminorm estimate for f depending only on r_0 (and the data of $\mathscr{M}, \mathscr{N}$). This follows from the above argument, because now the above fact $f^1 \in H_G$ is a direct consequence of the assumption $f(B \setminus B_{r_0}(0)) \subset \mathscr{O}_{\varepsilon_o}$ and the continuity of f.

A special choice of G is the conjugacy class $[e] \subset \pi_1(S)$ generated by some $e \in \ker i_* \setminus \{0\}$. We put $K = \{[e] : e \in \ker i_* \setminus \{0\}\}$. We note that K corresponds to

the collection of all nontrivial free homotopy classes of loops in $\mathcal{M}$ which are contractible in $\mathcal{N}$.

Definition 4 $\alpha \in K$ satisfies the Douglas conditions iff $A_\alpha < A_\beta + A_\gamma$ for each pair (β, γ) with $\beta, \gamma \in K$ and $\alpha = \beta\gamma$.

Lemma 4 *If* $\alpha, \beta, \gamma \in K$, $\alpha = \beta\gamma$, *then* $A_\alpha \leqq A_\beta + A_\gamma$.

Proof. Connect $g \in H_\beta$, $g' \in H_\gamma$ by a narrow strip in $\mathcal{M}$ to get $f \in H_\alpha$ with $D(f) \leqq D(g) + D(g') + \varepsilon$ for arbitrarily small $\varepsilon > 0$.

2 The main theorems

Theorem 1 *Suppose that* $\mathcal{N}$ *is uniformly Lipschitz and* $\mathcal{M}$ *is compact. Then we have: if the Douglas condition holds for* $\alpha \in K$, *then there exists a minimizing* α*-disk and every minimizing* α*-disk* f *satisfies the following estimate*

$$|f|_{C^{k,\alpha}}(\bar{B}) \leqq C \tag{2.1}$$

for a constant C *depending only on* A_α *provided that* $f(0) \notin O_{\varepsilon_0}$.

Corollary 1 *For each* $\alpha \in K$ *there are* $\alpha_1, \ldots, \alpha_l \in K$ *such that* $\alpha = \alpha_1 \ldots \alpha_l$, $A_\alpha = \sum_1^l A_{\alpha_i}$ *and minimizing* α_i*-disks exist for* $i = 1, \ldots, l$. *In fact, each* α_i *satisfies the Douglas condition.*

Proof. Using Lemma 3 one readily shows that if some α does not allow such a decomposition, then $A_\alpha \geqq jA_1$ for arbitrarily large j. We note that, reversedly this corollary also implies the existence part of Theorem 1 in an abvious way.

Corollary 2 *For any* $C > 0$ *there are at most finitely many* $\alpha \in K$ *such that* $A_\alpha \leqq C$.

Proof. Otherwise, by Corollary 1 we obtain mutually different $\alpha_1, \alpha_2, \ldots \in K$ with $A_{\alpha_i} \leqq C$ for some $C > 0$ and a minimizing α_i-disk f_i for each i. We may choose $f_i(0) \notin \mathcal{O}_{\varepsilon_0}$. Then a subsequence $f_{i'}$ converges uniformly to some f, which implies that $f_{i'}|_{\partial B}$ is freely homotopic to $f|_{\partial B}$ for large i'. This is impossible, since $\alpha_{i'}$ are mutually different. QED

To prove Theorem 1 let us define for $\alpha \in K$ and $l > 0$ $H_{\alpha,l} = \{f \in H_\alpha$: the length (the total variation) $L(f|_{\partial B})$ of $f|_{\partial B} \leqq l\}$, $D_{\alpha,l} = \inf\{D(f) : f \in H_{\alpha,l}\}$.

Lemma 5 *For any* $\alpha \in K$, $H_{\alpha,l}$ *is nonvoid for large* l *and* $A_\alpha = \lim_{l \to \infty} D_{\alpha,l}$.

Proof. For each sufficiently small $\varepsilon > 0$, using the geodesic flow along the normal line of $\mathcal{M}$ which is defined with respect to a C^∞ metric approximating the original metric of $\mathcal{N}$, we can easily construct a C^1 mapping $\phi_\varepsilon : \mathcal{N} \to \mathcal{N}$ such that $\|d\phi_\varepsilon\| \leqq 1 + C_\varepsilon$ for some constant $C > 0$ and ϕ_ε is a deformation retract from $\mathcal{O}_\varepsilon$ onto $\mathcal{M}$. Now let $g \in H_\alpha$. As in the proof of Lemma 3 we consider $f \in H_\alpha \cap C^0(\bar{B}) \cap C^2(B)$ with $D(f) \leqq D(g)$. For each small $\varepsilon > 0$ we can find r near 1 such that $f(\bar{B} \setminus B_r(0)) \subset \mathcal{O}_\varepsilon$. Let $\varphi : \bar{B} \to \bar{B}_r(0)$ be a conformal transformation. Then $\tilde{f} := \phi_\varepsilon \circ f|_{\bar{B}_r} \circ \varphi$ is in H_α, has finite boundary length and satisfies $D(\tilde{f}) \leqq (1 + C\varepsilon) D(f)$. QED

Lemma 6 *Let $\alpha \in K$. Then for each large l, there is a decomposition $\alpha = \alpha_{l1} \dots \alpha_{lN_l}$, $\alpha_{li} \in K$ such that $D_{\alpha,l} \geqq \sum_{i=1}^{N_l} D_{\alpha_{li},l}$ and there exist $f_{li} \in H_{\alpha_{li},l}$ with $D(f_{li}) = D_{\alpha_{li},l}$.*

Proof. The argument is similar to [T]. Calling *normalized* if $L(f|_{\partial_i B}) = \frac{1}{3} L(f|_{\partial B})$ where $\partial_i B$ are the arcs of ∂B between $e^{\frac{\sqrt{-1}}{2}\pi(i-1)}$, $i = 1, 2, 3$, we set

$$l^* = \inf\left\{ \lim_{k\to\infty} L(f_k|_{\partial B}) : \{f_k\} \subset H_{\alpha,l}, \text{ each } f_k \text{ is normalized and } D(f_k) \to D_{\alpha,l} \right\}.$$

Using Arzela–Ascoli, reparametrization by arclength and the fact that α is nontrivial, one readily shows $l^* > 0$.

A sequence $\{f_k\}$ as in this definition with the property that $\lim L(f_k|_{\partial B}) = l^*$ will be called an (α, l)-*sequence*. Choose an (α, l)-sequence $\{f_k\}$. If $f_k|_{\partial B}$ are equicontinuous, then we can go to limit to get a minimizer of the Dirichlet integral in $H_{\alpha,l}$ and we are done. Otherwise, by the lemma of Courant–Lebesgue [C], [Jo] and the fact that f_k are normalized, we can find a subsequence of $\{f_k\}$, still denoted $\{f_k\}$ and $w_k \in \partial B$, $r_k \to 0$ such that

$$\int_{C_{r_k}(w_k)} \left\| \frac{\partial f_k}{\partial \theta} \right\| d\theta \to 0, \quad L(f_k|_{I_{r_k}}(w_k)) \geqq \delta \text{ for some } \delta > 0 \text{ and } L(f_k|_{\partial B \setminus I_{r_k}(w_k)}) > \frac{l^*}{3}.$$

We can use Schwartz' inequality $L(f_k|_{C_{r_k}(w_k)})^2 \leqq \pi \int_{C_{r_k}(w_k)} \left\| \frac{\partial f_k}{\partial \theta} \right\|^2 d\theta$ to estimate $\operatorname{diam} f(C_{r_k}(w_k))$. Applying Replacement Procuedure we then get big and small maps f_k^1, f_k^2 and the corresponding homotopy decompositions $\alpha = \alpha_k^1 \alpha_k^2$ such that $L(f_k^1|_{\partial B})$, $L(f_k^2|_{\partial B}) \leqq l^* - \frac{1}{2}\min\left(\delta, \frac{l^*}{3}\right)$ (for large k). Furthermore, $D(f_k) = D(f_k^1) + D(f_k^2) - o(1)$, $\lim o(1) = 0$. For large k, $\alpha_k^1, \alpha_k^2 \in K$. Otherwise we could get from f_k^1 or f_k^2 a normalized sequence $f_k^* \in H_{\alpha,l}$ such that $D(f_k^*) \to D_{\alpha,l}$, $\lim L(f_k^*|_{\partial B}) < l^*$, contradicting the definition of l^*. On the other hand, a simple argument using Arzela–Ascoli shows that for each $l > 0$ there are only finitely many $\beta \in K$ such that $H_{\beta,l} \neq \phi$. Hence, by choosing subsequences we may assume that all α_k^1 are equal to some α^1 and all α_k^2 are equal to some α^2. It follows that $D_{\alpha,l} \geqq D_{\alpha^1,l} + D_{\alpha^2,l}$. Now choose (α^1, l)- and (α^2, l)-sequences $\{\tilde{f}_k^1\}, \{\tilde{f}_k^2\}$. If both sequences are equicontinuous, then we are done. Otherwise we may again apply Replacement Procedure to decompose α^1 and (or) α^2. By Lemma 3 this decomposition procedure must stop after finitely many steps and we arrive at the desired decomposition. QED

To proceed, let us call an arc $I_r(w) \subset \partial B$ nontrivial for some $f \in H$, if $\operatorname{dist}(f(w_1), f(w_2)) \leqq d_1(w_1, w_2)$ are the ends of $I_r(w)$) and there is no connecting curve τ as in Replacement Procedure such that the loop consisting of $f|_{I_r(w)}$ and the reverse of τ is contractible in $\mathcal{M}$.

A point $z \in \partial B$ is called a singular point of a sequence $\{f_k\} \subset H$ if there is a subsequence f_k, a sequence $r_{k'} \in (0, \frac{1}{2})$, and a sequence $z_{k'} \in \partial B$ such that $z_{k'} \to z$,

$$\int_{C_{r_{k'}}(z_{k'})} \left\| \frac{\partial f_{k'}}{\partial \theta} \right\|^2 d\theta \to 0 \text{ and for each } k', I_{r_{k'}}(z_{k'}) \text{ is nontrivial for } f_{k'}.$$

If we can choose $\{k'\} = \{k\}$, then the singular point z is called uniform.

Lemma 7 *Let $\alpha_k \in K$ and $f_k \in H_{\alpha_k, l_k}$ satisfy $D(f_k) = D_{\alpha_k, l_k}$ where $\{l_k\}$ is a sequence of positive numbers. If $\sup_k D_{\alpha_k, l_k} < \infty$ and $\{f_k\}$ has no singular points, then there are $\sigma \in (0, 1)$ and $C > 0$ such that*

$$|f_k|_{C^{\sigma,\sigma}(\partial B)} \leqq C \text{ for each } k \text{ (actually, } |f_k|_{C^{\sigma,\sigma}(\bar{B})} \leqq C) .$$

Proof. We proceed as in the proof of Proposition 1. By Schwartz inequality and the assumption that $\{f_k\}$ has no singular points, there are $C_1 > 0, r_o \in (0, \frac{1}{2})$ such that for arbitrary k, $r \in (0, r_o]$ and $w \in \partial B$ the following holds: if $\int_{C_r(w)} \left\| \frac{\partial f_k}{\partial \theta} \right\|^2 d\theta \leqq C_1$ then $\operatorname{dist}(f_k(w_1), f_k(w_2)) \leqq d_1(w_1, w_2)$ are the ends of $C_r(w))$ and $I_r(w)$ is not nontrivial for f_k. Let $f = f_k, w \in \partial B$ and $r \in (0, r_o]$. if $\int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta > C_1$, then $\varphi(r) = \int_{B_r(w) \cap B} \|df\|^2 \leqq \frac{2D(f)}{C_1} \varphi'(r) r$. If $\int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta > C_1$, then we apply Replacement Procedure to get the big map f^1, choosing τ so that $L(\tau) = L(f|_{I_r(w)})$ if $L(f|_{I_r(w)}) \leqq A_o \operatorname{dist}(f(w_1), f(w_2))$. Then $f^1 \in H_{\alpha_k, l_k}$, whence $D(f) \leqq D(f^1)$. Consequently, $\varphi(r) \leqq C \int_{C_r(w)} \left\| \frac{\partial f}{\partial \theta} \right\|^2 d\theta \leqq Cr\varphi'(r)$ by (1.2), (1.1). It follows that $\varphi(r) \leqq C_1 C_2 r^{1/C_1}$ for $r \in (0, r_o]$, where $C_2 = \max\left(C, 2\sup_k \frac{D(f_k)}{C_1} \right)$. The desired estimate follows from (the argument of) Morrey's Dirichlet growth theorem and the fact that $\mathcal{M}$ is compact. QED

Proof of the existence part of Theorem 1. By Lemma 6 and Lemma 3 we obtain a sequence $l_k \to \infty$ and a sequence of decompositions $\alpha = \alpha_{k1} \ldots \alpha_{kN}$ for a fixed N such that

$$D_{\alpha, l_k} \geqq \sum_{i=1}^{N} D_{\alpha_{ki}, l_k}$$

and there exist minimizers $f_{ki} \in H_{\alpha_{ki}, l_k}$ with $f_{ki}(0) \notin \mathcal{O}_{\varepsilon_o}$. By Lemma 3 and Lemma 5 we have $A_\alpha \geqq N a_1$. If none of $\{f_{ki}\}$, $i = 1, \ldots, N$ has singular points, then by Lemma 7 and Lemma 5 we can go to limit to get $\alpha = \alpha_1 \ldots \alpha_N, f_i \in H_{\alpha_i}$, $A_\alpha \geqq \sum_{i=1}^{N} D(f_i)$. By the Douglas condition and Lemma 4 we must have $N = 1$, whence we are done.

If for some i, the seqeunce $f_k := f_{ki}$ has a singular point z, by selecting a subsequence we may assume that z is uniform. Let $z_k \to z$ be as in the definition of singular points. Applying Replacement Procedure to f_k at z_k we get big and small maps f_k^1, f_k^2 and the corresponding homotopy decompositions $(\alpha_k := \alpha_{ki})\alpha_k = \alpha_k^1 \alpha_k^2$ such that $D(f_k) = D(f_k^1) + D(f_k^2) - o(1)$. It is clear that $\alpha_k^2 \in K$, hence $D(f_k^2) \geqq a_1$ by Lemma 3. By Lemma 1 and the choice $f_k(o) \notin \mathcal{O}_{\varepsilon_o}$ we also have $D(f_k^1) \geqq a_1$ (we may assume $a_1 \leqq a_o$), hence $\alpha_k^1 \in K$ for large k, otherwise we would have, by Lemma 4,

$$A_\alpha \geqq \lim_{k \to \infty} \left(\sum_{j \neq i} D_{\alpha_{kj}, l_k} + D(f_k^2) \right) + a_1 \geqq A_\alpha + a_1 .$$

Writing $l_k^1 = L(f_k^1|_{\partial B})$, $l_k^2 = L(f_k^2|_{\partial B})$ we obtain by Lemma 5

$$\alpha_k^1 = \beta_k^1 \ldots \beta_k^{N_1}, \qquad \alpha_k^2 = \gamma_k^1 \ldots \gamma_k^{N_2}$$

and minimizers $g_k^i \in H_{\beta_k^i, l_k^i}$, $h_k^i \in H_{\gamma_k^i, l_k^i}$. Selecting a subsequence and renaming it we get

$$\alpha = \alpha_k^1, \ldots \alpha_k^{N'}, \qquad A_\alpha \geqq \sum_1^{N'} D_{\alpha_k^i, l_k^i} - o(1)$$

and minimizers $f_k^i \in H_{\alpha_k^i, l_k^i}$ with $f_k^i(0) \notin \mathcal{O}_{\varepsilon_o}$, $N' = (N-1) + N_1 + N_2 \geqq N+1$. In particular, $A_\alpha \geqq (N+1)a_1$.

Now if none of $\{f_k^i\}$, $i = 1, \ldots, N'$ has singular points, then we are done. Otherwise we can repeat the above construction. Since $A_\alpha < \infty$, this process must stop after finitely many steps and thus the existence of a minimizing α-disk is proved. QED

Proof of the estimate (2.1). One can argue in terms of singular points as above, but we choose a somewhat different approach. For a given $C_1 > 0$ we put $F_{C_1} = \{f : f$ is a minimizing α-disk for some $\alpha \in K$ such that $A_\alpha \leqq C_1$, α satisfies the Douglas condition and $f(0) \notin \mathcal{O}_{\varepsilon_0}\}$. We claim that there exists $r_o \in (0,1)$ such that $f(B \setminus B_{r_o}(0)) \subset \mathcal{O}_{\varepsilon_0}$. Otherwise, there would be a sequence $\alpha_k \in K$ and for each k a minimizing α_k-disk $f_k \in F_{C_1}$ together with $w_1, w_2, \ldots \in B$ such that $\lim w_k = w \in \partial B$ and $\operatorname{dist}(f_k(w_k), \mathcal{M}) \geqq \varepsilon_o$. By the lemma of Courant–Lebesgue and the fact $D(f_k) \leqq C_1$ we can find $r_k \to 0$ so that $\int_{C_{r_k}(w)} \left\| \frac{\partial f_k}{\partial \theta} \right\|^2 d\theta \leqq 4D(f_k)|\log r_k|^{-1}$, $w_k \in B_{r_k}(w)$. Using Replacement Procedure, Lemma 6, Lemma 7 and arguing the same way as above, we then obtain a subsequence $\alpha_{k'}$, decompositions $\alpha_{k'} = \alpha_k^1 \ldots \alpha_{k'}^N$, positive numbers $l_{k'}^i$ and minimizers $f_{k'}^i$ of the Dirichlet intergral in $H_{\alpha_{k'}^i, l_{k'}^i}$, so that $A_{\alpha_{k'}} \geqq \sum_{i=1}^N D(f_{k'}^i) - o(1)$, $N > 1$, and each sequence $\{f_{k'}^i|_{\partial B}\}$, $i = 1, \ldots, N$ converges uniformly. By this uniform convergence we may assume that all $\alpha_{k'}$ are equal to some α and all $\alpha_{k'}^i$ are equal to some α^i, $i = 1, \ldots, N$. It follows $A_\alpha \geqq \sum_1^N A_{\alpha^i}$, contradicting the Douglas condition.

By the uniform relation $f(B \setminus B_{r_o}) \subset \mathcal{O}_{\varepsilon_o}$, the condition $f \in F_{C_1}$, the remark to Proposition 1 and the compactness of $\mathcal{M}$ we obtain a uniform Hölder norm estimate. The higher order estimates follow from the regularity proofs (see the references for Proposition 1). QED

Theorem 2 *Let $\mathcal{N}$ be homogeneously regular and $\mathcal{M}$ compact. If G is the complement of a normal subgroup of $\ker i_*$, then there exist minimizing G-disks and every minimizing G-disk f with $f(0) \notin \mathcal{O}_{\varepsilon_o}$ satisfies (2.4) for some C depending only on A_G.*

Proof. We first prove the existence. Pick $\alpha_1, \alpha_2, \ldots \in K$ such that $\alpha_k \subset G$ and $A_{\alpha_k} = A_G$ for large k. Fixing such a k we get, by Corollary 1 of Theorem 1,

$$\alpha_k = \beta_1 \ldots \beta_N, \qquad A_{\alpha_k} = A_{\beta_1} + \ldots + A_{\beta_N}$$

such that there exist minimizing β_i-disks. Since G is the complement of a normal subgroup, we have $\beta_i \subset G$ for some i, Clearly $A_{\beta_i} = A_G$.

To show the *a priori* estimate we proceed as in the corresponding part of the proof of Theorem 1. If no uniform relation $f(B \setminus B_{r_o}) \subset \mathcal{O}_{\varepsilon_o}$ for $f \in F_C := \{f : f$ is a minimizing G-disk for some G so that G is the complement of a normal subgroup

of $\ker i_*$, $A_G \leqq C$ for a given, fixed C, and $f(0) \notin \mathcal{O}_{\varepsilon_0}\}$ holds, then we can obtain a sequence G_k, for each k a minimizing G_k-disk f_k together with the associated big and small maps f_k^1, f_k^2 which satisfy $A_{G_k} \geqq D(f_k^1) + D(f_k^2) - \bar{C} A_{G_k} |\log r_k|^{-1}$ for some $\bar{C} > 0$ and a sequence $r_k \to 0$, as well as $D(f_k^1), D(f_k^2) \geqq a_1$. By the assumption on G_k, for each k either $f_k^1 \in H_{G_k}$ or $f_k^2 \in H_{G_k}$. Hence $A_{G_k} \geqq A_{G_k} + a_1 - \bar{C} A_{G_k} |\log r_k|^{-1}$, a contradiction. We see that in fact we can choose $r_0 = 1 - e^{-2C/a_1}$. QED

Corollary *There exist generators $e_1, \ldots, e_N$ of $\ker i_*$ with minimizing $[e_i]$-disks. Also, among all disk type surfaces whose boundaries lie on $\mathcal{M}$, are noncontractible in $\mathcal{M}$ and contractible in $\mathcal{N}$, there is one with least area and Dirichlet integral. Finally, if $\mathcal{F}$ is a subset of $\mathcal{N}$ disjoint from $\mathcal{M}$, then among all disk type surfaces whose boundaries lie on $\mathcal{M}$, are non-contractible in $\mathcal{N} \setminus \mathcal{F}$ and contractible in $\mathcal{N}$ (assuming there are such surfaces) there exists one with least area and Dirichlet integral (This is the setting in* [C], [D], [Ha], *and* [Hi] *where one demands* $\operatorname{dist}(\mathcal{M}, \mathcal{F}) > 0$ *but allows more general $\mathcal{M}$. In $\mathbb{R}^3$, by duality, one can choose Jordan loops $\mathcal{F}$ to prescribe 1-dimensional homology classes in $\mathcal{M}$).*

Now let us treat noncompact supporting submanifolds $\mathcal{M}$.

Definition 5. We call $\mathcal{M}$ "big at infinity", if $\mathcal{M}$ satisfies a chord-arc condition and there is a neighborhood $\mathcal{O}$ of $\mathcal{M}$ such that $\mathcal{M}$ is a retract of $\mathcal{O}$ and $\operatorname{dist}(p, \mathcal{M}) \to \infty$ as $p \in \partial\mathcal{O}$ goes to infinity. (Of course, we assume here that $\mathcal{N}$ is noncompact.)

Theorem 3 *Theorem* 1, *Theorem* 2 *and their corollaries are true for a noncompact supporting submanifold $\mathcal{M}$ if $\mathcal{M}$ is big at infinity and $\mathcal{N}$ has bounded metric.*

Proof. We only indicate how to carry over the proof of the existence part of Theorem 1. First we note that Replacement Procedure can be applied. On the other hand, for any $C > 0, l > 0$, all noncontractible loops in $\mathcal{M}$ whose boundary lengths $\leqq l$ and which bound surfaces in H of Dirichlet integral $\leqq C$ must lie in some compact region of $\mathcal{N}$. Otherwise, we would have such a loop τ so that $\operatorname{dist}(p_o, \tau)$ is arbitrarily large for some fixed point $p_o \in \mathcal{M}$. Arguing as in the proof of Lemma 3 we find a Jordan loop τ' near τ and a solution f of the Plateau problem for τ' such that $\tau' \subset \mathcal{O}_\varepsilon \subset \mathcal{O}$ for arbitrarily small $\varepsilon > 0$, τ' is noncontractible in $\mathcal{O}$ and $D(f) \leqq 2C$. Lemma 2 together with the "big at infinity" assumption then leads to a contradiction. It follows that Lemma 6 carries over. Lemma 5 also carries over, since for each given $g \in H_\alpha$ the set $g(\partial B)$ is compact and a C^1 mapping $\phi_\varepsilon : \mathcal{N} \to \mathcal{N}$ can be constructed such that $\|d\phi_\varepsilon\| \leqq 1 + C\varepsilon$ and ϕ_ε is a deformation retract from some neighborhood of $\mathcal{M}$ onto $\mathcal{M}$. It is seen that Lemma 7 carries over, too. QED.

References

[C] Courant, R.: Dirichlet's principle, conformal mapping and minimal surfaces. New York: Interscience Publication 1950

[D] Davids, N.: Minimal surfaces spanning closed manifolds and having prescribed topological position. Am. J. Math. **64**, 348–362 (1942)

[GHN] Grüter, M., S. Hildebrandt, Nitsche, J.C.C.: On the boundary behavior of minimal surfaces with a free boundary which are not minima of the area. Manuscr. Math. **35**, 387–410 (1981)

[Ha] Harth, F.P.: Minimalflächen in Riemannschen Mannigfaltigkeiten mit freiem Rand. Dissertation, Mainz 1970

[Hi] Hildebrandt, S.: Randwertprobleme für Flächen mit vorgeschriebener mittlerer Krümmung und Anwendungen auf die Kapillantätstheorie. II. Freie Ränder. Arch. Ration. Mech. Anal. **39**, 275–293 (1970)

[HN] Hildebrandt, S., Nitsche, J.C.C.: Minimal surfaces with free boundaries. Acta Math. **143**, 251–272 (1979)

[Jä] Jäger, W.: Behavior of minimal surfaces with free boundaries. Commun. Pure Appl. Math. **23**, 803–818 (1970)

[Jo] Jost, J.: Conformal mappings and the Plateau-Douglas problem. J. Reine Angew. Math. **359**, 37–54 (1985)

[JK] Jost, J., Karcher, H.: Geometrische Methoden zur Gewinnung von a-priori-Schranken für harmonische Abbildungen. Manuscr. Math. **40**, 27–77 (1982)

[L] Lewy, H.: On the boundary behavior of minimal surfaces. Proc. Natl. Acad. Sci. **37**, 103–110 (1951)

[M1] Morrey, C.B.: The problem of Plateau in a Riemannia manifold. Ann. Math. **49**, 807–851 (1948)

[M2] Morrey, C.B.: Multiple integrals in the calculus of variations. Berlin Heidelberg New York: Springer 1966

[MY] Meeks, B., Yau, S.T.: Topology of three-dimensional manifolds and the embedding problems in minimal surface theory. Ann. Math. **112**, 441–484 (1980)

[N] Nitsche, J.C.C.: Minimal surfaces with partially free boundary. Least area property and Hölder continuity for boundaries satisfying a chord-arc condition. Arch. Ration. Mech. Anal. **39**, 131–145 (1970)

[T] Tolksdorf, P.: On minimal surfaces with free boundaries in given homotopy classes. Ann. Inst. Henri Poincaré **2**, 157–165 (1985)

[Y] Ye, R.: Existence, regularity and finiteness of minimal surfaces with free boundary. SFB 256 no. 1, Preprint Bonn 1987

Math. Z. 206, 333–339 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

Complete space-like submanifolds in a de Sitter space with parallel mean curvature vector

Qing-ming Cheng

Department of Mathematics, Northeast University of Technology, Shenyang, Liaoning, China

Received January 3, 1990, in final form February 21, 1990

1 Introduction

Let $M_p^{n+p}(c)$ be an $n+p$-dimensional connected semi-Riemannian manifold of index p and of constant curvature c, which is called as indefinite space form of index p. If $c>0$, we call it as a de Sitter space of index p. It is seen that a complete space-like hypersurface of a Minkowski space R_1^{n+p} possesses a remarkable Bernstein property in the maximal case by Calabi [2] and Cheng and Yau [3]. As a generalization of the Bernstein type problem, a complete space-like maximal submanifold M^n of $M_p^{n+p}(c)$ was characterized by Ishihara [5] under a certain conditions. An entire space-like hypersurface with constant mean curvature of a Minkowski space is investigated by Goddard [4] and Treibergs [10]. Akutagawa [1] and Ramanathan [9] investigated the complete space-like hypersurfaces in a de Sitter space. They obtained independently that a complete space-like hypersurface in a de Sitter space with constant mean curvature is totally umbilical if the following are satisfied

$$H^2 \leqq c, \qquad \text{when } n=2\,; \tag{1.1}$$

$$n^2H^2 < 4(n-1)c, \qquad \text{when } n \geqq 3\,. \tag{1.2}$$

In this paper, we consider general submanifolds in an indefinite space form, we obtain that a complete space-like submanifold in a de Sitter space with parallel mean curvature vector is totally umbilical if the conditions (1.1) or (1.2) are satisfied. Conditions (1.1) and (1.2) are best possible.

2 Local formulas and lemmas

Let $M_p^{n+p}(c)$ be an $(n+p)$-dimensional semi-Riemannian manifold of constant curvature c whose index is p. Let M^n be an n-dimensional Riemannian manifold immersed in $M_p^{n+p}(c)$. As the semi-Riemannian metric of $M_p^{n+p}(c)$ induced the

Riemannian metric of M^n, the immersion is called space-like. We choose a local field of semi-Riemannian orthonormal frames $e_1, \ldots, e_{n+p}$ in $M_p^{n+p}(c)$ such that at each point of M^n, $e_1, \ldots, e_n$ span the tangent space of M^n and forms an orthonormal frame there. We make us of the following convention on the range of indices: $1 \leqq A, B, C, \ldots \leqq n+p$; $1 \leqq i, j, k, \ldots \leqq n$; $n+1 \leqq \alpha, \beta, \gamma, \ldots \leqq n+p$; and we shall agree that repeated indices under summation sign are summed over respective ranges. Let $\omega_1, \ldots, \omega_{n+p}$ be its dual frame field so that the semi-Riemannian metric of $M_p^{n+p}(c)$ is given by $ds^2_{M_p^{n+p}} = \sum \omega_i^2 - \sum \omega_\alpha^2 = \sum \varepsilon_A \omega_A^2$, where $\varepsilon_i = 1$ and $\varepsilon_\alpha = -1$. Then the structure equations of $M_p^{n+p}(c)$ are given by

$$d\omega_A = -\sum \varepsilon_B \omega_{AB} \wedge \omega_B, \qquad \omega_{AB} + \omega_{BA} = 0\,, \tag{2.1}$$

$$d\omega_{AB} = -\sum \varepsilon_C \omega_{AC} \wedge \omega_{CB} - \tfrac{1}{2}\sum K_{ABCD}\omega_C \wedge \omega_D\,, \tag{2.2}$$

$$K_{ABCD} = c\varepsilon_A\varepsilon_B(\delta_{AD}\delta_{BC} - \delta_{AC}\delta_{BD}) \tag{2.3}$$

We restricted these forms to M^n, then

$$\omega_\alpha = 0, \qquad \text{for } \alpha = n+1, \ldots, n+p, \tag{2.4}$$

and the Riemannian metric of M^n is written as $ds^2 = \sum \omega_i^2$. We put $\omega_{\alpha i} = \sum h_{ij}^\alpha \omega_j$. From Cartan's Lemma we have $h_{ij}^\alpha = h_{ji}^\alpha$, where h_{ij}^α are the components of the second fundamental form of M^n. Let

$$h = \frac{1}{n}\sum_i \sum (h_{ii}^\alpha) e_\alpha \tag{2.5}$$

$$H^2 = \frac{1}{n^2}\sum \left(\sum_i h_{ii}^\alpha\right)^2 \tag{2.6}$$

where H is called mean curvature. From (2.1), we obtain the structure equations of M^n

$$d\omega_i = -\sum \omega_{ij} \wedge \omega_j, \qquad \omega_{ij} + \omega_{ji} = 0\,, \tag{2.7}$$

$$d\omega_{ij} = -\sum \omega_{ik} \wedge \omega_{kj} - \tfrac{1}{2}\sum R_{ijkl}\omega_k \wedge \omega_l\,, \tag{2.8}$$

and the Gaussian formula

$$R_{ijkl} = c(\delta_{il}\delta_{jk} - \delta_{ik}\delta_{jl}) - \sum (h_{il}^\alpha h_{jk}^\alpha - h_{ik}^\alpha h_{jl}^\alpha) \tag{2.9}$$

The components of the Ricci curvature tensor Ric are given by

$$R_{jk} = c(n-1)\delta_{jk} - \sum h_{il}^\alpha h_{jk}^\alpha + \sum h_{ik}^\alpha h_{ji}^\alpha \tag{2.10}$$

We have also the structure equations of the normal bundle of M^n

$$d\omega_\alpha = -\sum \omega_{\alpha\beta} \wedge \omega_\beta, \qquad \omega_{\alpha\beta} + \omega_{\beta\alpha} = 0\,, \tag{2.11}$$

$$d\omega_{\alpha\beta} = -\sum \omega_{\alpha\gamma} \wedge \omega_{\gamma\beta} - \tfrac{1}{2}\sum R_{\alpha\beta ij}\omega_i \wedge \omega_j\,, \tag{2.12}$$

$$R_{\alpha\beta ij} = -\sum (h_{il}^\alpha h_{jl}^\beta - h_{jl}^\alpha h_{il}^\beta) \tag{2.13}$$

For indefinite Riemannian manifolds, refer to O'Neill [8].

Let h_{ijk}^α denote the covariant derivatives of h_{ij}^α so that

$$\sum h_{ijk}^\alpha \omega_k = dh_{ij}^\alpha - \sum h_{ki}^\alpha \omega_{kj} - \sum h_{jk}^\alpha \omega_{ki} - \sum h_{ij}^\beta \omega_{\beta\alpha}\,. \tag{2.14}$$

Similarly, let h^α_{ijkl} denote the covariant derivative of h^α_{ijk} so that

$$\sum h^\alpha_{ijkl}\omega_l = dh^\alpha_{ijk} - \sum h^\alpha_{ljk}\omega_{li} - \sum h^\alpha_{ilk}\omega_{lj} - \sum h^\alpha_{ijl}\omega_{lk} - \sum h^\beta_{ijk}\omega_{\beta\alpha} \tag{2.15}$$

Then we obtain

$$h^\alpha_{ijk} = h^\alpha_{ikj}, \qquad \text{for any } \alpha\,, \tag{2.16}$$

and the Ricci formula

$$h^\alpha_{ijkl} - h^\alpha_{ijlk} = -\sum h^\alpha_{im}R_{mjkl} - \sum h^\alpha_{jm}R_{mikl} - \sum h^\beta_{ij}R_{\alpha\beta kl}\,. \tag{2.17}$$

The Laplacian Δh^α_{ij} of the second fundamental form h is defined by $\Delta h^\alpha_{ij} = \sum h^\alpha_{ijkk}$. From (2.17) we have

$$\Delta h^\alpha_{ij} = \sum h^\alpha_{kkij} - \sum h^\alpha_{km}R_{mijk} - \sum h^\alpha_{mi}R_{mkjk} - \sum h^\beta_{ki}R_{\alpha\beta jk}\,. \tag{2.18}$$

Lemma 2.1 (see Omori [7] and Yau [11]). *Let M^n be an n-dimensional complete Riemannian manifold whose Ricci curvature is bounded from below. Let F be a C^2-function bounded from above on M^n, then for any $\varepsilon > 0$, there exists a point p in M^n such that*

$$\begin{aligned} \sup F - \varepsilon &< F(p)\,, \\ |\operatorname{grad} F| &< \varepsilon\,, \\ \Delta F &< \varepsilon\,. \end{aligned} \tag{2.19}$$

Lemma 2.2 (see Okumura [6]). *Let $a_1, \ldots, a_n$ be real numbers satisfies $\sum a_j = 0$ and $\sum a_j^2 = K^2 (K > 0)$, then we have*

$$|\sum a_j^3| \leqq (n-2)[n(n-1)]^{-1/2}K^3\,.$$

3 Theorem and its proof

Theorem. *Let M^n be an n-dimensional complete space-like submanifold in $M_p^{n+p}(c)$ with parallel mean curvature vector. If*

$$H^2 \leqq c, \qquad \text{when } n = 2\,, \tag{3.1}$$

$$n^2H^2 < 4(n-1)c, \qquad \text{when } n \geqq 3\,. \tag{3.2}$$

then, M^n is totally umbilical.

Proof. From (2.18) and Gaussian formula, we have

$$\begin{aligned} \Delta h^\alpha_{ij} = {} & \sum h^\alpha_{kkij} + nch^\alpha_{ij} - c\sum h^\alpha_{kk}\delta_{ij} + \sum h^\alpha_{km}h^\beta_{mk}h^\beta_{ij} - 2\sum h^\alpha_{mk}h^\beta_{mj}h^\beta_{ik} \\ & + \sum h^\alpha_{mi}h^\beta_{mk}h^\beta_{kj} - \sum h^\alpha_{mi}h^\beta_{mj}h^\beta_{kk} + \sum h^\alpha_{jm}h^\alpha_{mk}h^\beta_{ki}\,. \end{aligned} \tag{3.3}$$

Because the mean curvature vector is parallel, we have that mean curvature H is constant.

If $H = 0$, M^n is maximal. From the theorem 1.1 in [5], we know that M^n is totally geodesic.

If $H \neq 0$, we can choose e_{n+1} in such a way that its direction coincides with that of mean curvature vector h. Then

$$\omega_{\beta,n+1} = 0, \quad H = \text{constant}\,, \tag{3.4}$$

$$H^\alpha H^{n+1} = H^{n+1}H^\alpha\,, \tag{3.5}$$

$$\operatorname{tr}H^{n+1} = nH, \qquad \operatorname{tr}H^\alpha = 0, \quad \alpha \neq n+1\,, \tag{3.6}$$

where H^α denote the matrix (h_{ij}^α).

Putting

$$\mu_{ij} = h_{ij}^{n+1} - H\delta_{ij}, \qquad \tau_{ij}^\alpha = h_{ij}^\alpha, \quad \alpha \neq n+1, \tag{3.7}$$

we have

$$\|\mu\|^2 = \operatorname{tr}(\mu)^2 = \sum \mu_{ij}^2 = \operatorname{tr}(H^{n+1})^2 - nH^2, \tag{3.8}$$

$$\|\tau\|^2 = \sum_{\beta \neq n+1} \operatorname{tr}(\tau^\beta)^2 = \sum_{\beta \neq n+1} \|\tau^\beta\|^2 = \sum_{\beta \neq n+1} (\tau_{ij}^\beta)^2 = \sum_{\beta \neq n+1} (h_{ij}^\beta)^2, \tag{3.9}$$

$$\operatorname{tr}\mu = 0, \qquad \operatorname{tr}(\tau^\beta) = 0, \quad \beta \neq n+1, \tag{3.10}$$

$$S = \|\mu\|^2 + \|\tau\|^2 + nH^2, \tag{3.11}$$

where S is the square of length of the second fundamental form. Hence, it may be seen that $\|\tau\|^2$ as well $\|\mu\|^2$ are independent of the choice of the frame field and are functions globally defined on M^n.

A submanifold M^n is said to be pseudo-umbilical if it is umbilical with respect to the direction of the mean curvature vector h, i.e., $h_{ij}^{n+1} = H\delta_{ij}$. From (3.7) $\sim$ (3.11) one can easily see that M^n is pseudo-umbilical if and only if $\|\mu\|^2 = 0$, M^n is totally umbilical if and only if it is pseudo-umbilical and $\|\tau\|^2 = 0$.

$$\begin{aligned}\Delta h_{ij}^{n+1} &= nch_{ij}^{n+1} - ncH\delta_{ij} + \sum h_{km}^{n+1} h_{mk}^\beta h_{ij}^\beta - 2\sum h_{mk}^{n+1} h_{mj}^\beta h_{ik}^\beta \\ &\quad + \sum h_{mi}^{n+1} h_{mk}^\beta h_{kj}^\beta - nH\sum h_{mi}^{n+1} h_{mj}^{n+1} + \sum h_{jm}^{n+1} h_{mk}^\beta h_{ki}^\beta, \end{aligned} \tag{3.12}$$

$$\begin{aligned}\Delta h_{ij}^\alpha &= nch_{ij}^\alpha + \sum h_{km}^\alpha h_{mk}^\beta h_{ij}^\beta - 2\sum h_{mk}^\alpha h_{mj}^\beta h_{ik}^\beta + \sum h_{mi}^\alpha h_{mk}^\beta h_{kj}^\beta \\ &\quad - nH\sum h_{mi}^\alpha h_{mj}^{n+1} + \sum h_{jm}^\alpha h_{mk}^\beta h_{ki}^\beta, \quad a \neq n+1. \end{aligned} \tag{3.13}$$

$$\begin{aligned}\tfrac{1}{2}\Delta\|\mu\|^2 &= \sum (h_{ijk}^{n+1})^2 + nc\sum (h_{ij}^{n+1})^2 - n^2cH^2 + \sum h_{mk}^{n+1} h_{mk}^\beta h_{ij}^\beta h_{ij}^{n+1} \\ &\quad - 2\sum h_{mk}^{n+1} h_{mj}^\beta h_{ik}^\beta h_{ij}^{n+1} + \sum h_{mi}^{n+1} h_{mk}^\beta h_{ki}^\beta h_{ij}^{n+1} - nH\sum h_{mi}^{n+1} h_{mj}^{n+1} h_{ij}^{n+1} \\ &\quad + \sum h_{jm}^{n+1} h_{mk}^\beta h_{ki}^\beta h_{ij}^{n+1} \\ &= \sum (h_{ijk}^{n+1})^2 + nc\sum (h_{ij}^{n+1})^2 - n^2cH^2 - nH\operatorname{tr}(H^{n+1})^3 \\ &\quad + \sum_{\beta \neq n+1} \operatorname{tr}(H^{n+1}H^\beta)^2 + [\operatorname{tr}(H^{n+1})^2]^2. \end{aligned} \tag{3.14}$$

Here we use $H^{n+1}H^\beta = H^\beta H^{n+1}$.

$$\begin{aligned}\tfrac{1}{2}\Delta\|\tau\|^2 &= \sum_{\alpha \neq n+1} (\tau_{ijk}^\alpha)^2 + nc\|\tau\|^2 + \sum_{\alpha \neq n+1} h_{km}^\alpha h_{mk}^\beta h_{ij}^\beta h_{ij}^\alpha \\ &\quad - 2\sum_{\alpha \neq n+1} h_{mk}^\alpha h_{mj}^\beta h_{ik}^\beta h_{ij}^\alpha + \sum_{\alpha \neq n+1} h_{mi}^\alpha h_{mk}^\beta h_{kj}^\beta h_{ij}^\alpha \\ &\quad - nH\sum_{\alpha \neq n+1} h_{mi}^\alpha h_{ij}^\alpha h_{mj}^{n+1} + \sum_{\alpha \neq n+1} h_{jm}^\alpha h_{mk}^\beta h_{ki}^\beta h_{ij}^\alpha. \end{aligned} \tag{3.15}$$

On the other hand,

$$\operatorname{tr}(H^{n+1})^3 = \operatorname{tr}\mu^3 + 3H[\operatorname{tr}(H^{n+1})^2 - nH^2] + nH^3. \tag{3.16}$$

(3.14) and (3.16) imply

$$\begin{aligned}\tfrac{1}{2}\Delta\|\mu\|^2 &\geqq (\|\mu\|^2 + nH^2)^2 - nH[\operatorname{tr}(\mu)^3 + 3H\|\mu\|^2 + nH^3] + nc\|\mu\|^2 \\ &= \|\mu\|^2[\|\mu\|^2 + nc - nH^2] - nH\operatorname{tr}(\mu)^3. \end{aligned} \tag{3.17}$$

Since $\operatorname{tr}\mu = 0$, we can apply Lemma 2.2 to the eigenvalues of μ and hence

$$|\operatorname{tr}(\mu)^3| \leqq (n-2)[n(n-1)]^{-1/2} \|\mu\|^3 . \tag{3.18}$$

Hence,

$$\begin{aligned} \tfrac{1}{2}\Delta \|\mu\|^2 &\geqq \|\mu\|^2[\|\mu\|^2 + nc - nH^2] - n|H|(n-2)[n(n-1)]^{-1/2} \|\mu\|^3 \\ &= \|\mu\|^2\{\|\mu\|^2 - n|H|(n-2)[n(n-1)]^{-1/2} \|\mu\| + nc - nH^2\} . \end{aligned} \tag{3.19}$$

From (2.10), we know that the Ricci curvature of M^n is bounded from below. Putting $F = -(\|\mu\|^2 + a)^{-1/2}$, since M^n is space-like and F is bounded, we can apply Lemma 2.1 to the function F. For given any positive number $\varepsilon > 0$, there exists a point p at which F satisfies the properties (2.19) in Lemma 2.1. Consequently, the following relationship.

$$\tfrac{1}{2}F^4(p)\Delta \|\mu(p)\|^2 < 3\varepsilon^2 - F(p)\varepsilon . \tag{3.20}$$

can be derived by the simple and direct calculation. For a convergent $\{\varepsilon_m\}$ such that $\varepsilon_m \to 0$ $(m \to \infty)$, there exists a point sequence $\{p_m\}$ such that the sequence $\{F(p_m)\}$ converges to F_0 because $\{F(p_m)\}$ is a bounded sequence, by taking a subsequence, if necessary. From the difinition of the supremum and (2.19), we have $F_0 = \sup F$ and hence the definition of F gives rise to $\lim \|\mu(p_m)\|^2 = \sup \|\mu\|^2$.

On the other hand, it follows from (3.20) that we have

$$\tfrac{1}{2}F^4(p_m)\Delta \|\mu(p_m)\|^2 < 3\varepsilon_m^2 - F(p_m)\varepsilon_m . \tag{3.21}$$

The right hand side of (3.21) converges to 0 because F is bounded. Accordingly, for any positive number $\varepsilon > 0$ $(\varepsilon < 2)$ there is a sufficiently large integer m for which we have

$$F(p_m)^4 \Delta \|\mu(p_m)\|^2 < \varepsilon . \tag{3.22}$$

This relationship and (3.19) yield

$$\begin{aligned} &(2-\varepsilon)\|\mu(p_m)\|^4 - 2(n-2)[n(n-1)]^{-1/2} n|H| \|\mu(p_m)\|^3 \\ &\quad + 2(nc - nH^2 - \varepsilon a)\|\mu(p_m)\|^2 - \varepsilon a^2 < 0 . \end{aligned}$$

Hence, $\{\|\mu(p_m)\|\}$ is bounded. Thus the supremum of F satisfies $F_0 = \sup F > 0$. According to (3.22), we have

$$\limsup_{m \to \infty} \Delta \|\mu(p)\|^2 \leqq 0 . \tag{3.23}$$

(3.19) implies

$$\sup \|\mu\|^2 [\sup \|\mu\|^2 - (n-2)[n(n-1)]^{-1/2} n|H| \sup \|\mu\| + (nc - nH^2)] \leqq 0 . \tag{3.24}$$

1), when $n = 2$ and $H^2 \leqq c$,

$$\sup \|\mu\|^2 [\sup \|\mu\|^2 + nc - nH^2] \leqq 0 .$$

Hence $\sup \|\mu\|^2 = 0$, i.e., $\|\mu\|^2 \equiv 0$.

2), when $n \geqq 3$ and $n^2H^2 < 4(n-1)c$, we have also

$$\sup \|\mu\|^2 = 0, \text{ hence } \|\mu\|^2 \equiv 0 .$$

That is, M^n is pseudo-umbilical. Hence $h_{ij}^{n+1} = H\delta_{ij}$.

From (3.15), we get

$$\begin{aligned}\tfrac{1}{2}\Delta\|\tau\|^2 &= \sum_{\alpha=n+1}(\tau^\alpha_{ijk})^2 + nc\|\tau\|^2 + \sum_{\alpha=n+1} h^\alpha_{km}h^\beta_{mk}h^\beta_{ij}h^\alpha_{ij}\\ &\quad - 2\sum_{\alpha=n+1} h^\alpha_{mk}h^\beta_{mj}h^\beta_{ik}h^\alpha_{ij} + \sum_{\alpha=n+1} h^\alpha_{mi}h^\beta_{mk}h^\beta_{kj}h^\alpha_{ij}\\ &\quad - nH\sum_{\alpha=n+1} h^\alpha_{mi}h^{n+1}_{mj}h^\alpha_{ij} + \sum_{\alpha=n+1} h^\alpha_{jm}h^\beta_{mk}h^\beta_{ki}h^\alpha_{ij}\\ &= \sum_{\alpha=n+1}(\tau^\alpha_{ijk})^2 + nc\|\tau\|^2 + \sum_{\alpha,\beta=n+1}[\mathrm{tr}(H_\alpha H_\beta)]^2\\ &\quad - nH\|\tau\|^2 - 2\sum_{\beta,\alpha=n+1} h^\alpha_{mk}h^\beta_{mj}h^\beta_{ik}h^\alpha_{ij}\\ &\quad + \sum_{\alpha,\beta=n+1} h^\alpha_{mi}h^\beta_{mk}h^\beta_{kj}h^\alpha_{ij} + \sum_{\alpha,\beta=n+1} h^\alpha_{jm}h^\beta_{mk}h^\beta_{ki}h^\alpha_{ij}\,. \qquad (3.25)\end{aligned}$$

Here we make use of $h^{n+1}_{ij} = H\delta_{ij}$.

We put $S_{\alpha\beta} = \sum h^\alpha_{ij}h^\beta_{ij}$ for $\alpha,\ \beta = n+1$. Then $(S_{\alpha\beta})$ is a $(p-1)\times(p-1)$ symmetrix matrix.. It can be assumed to be diagonal for a suitable choice of $e_{n+2}, \ldots, e_{n+p}$. Set $S_\alpha = S_{\alpha\alpha}$ and we have $\|\tau\|^2 = \sum S_\alpha$. In general, for a matrix $A = (a_{ij})$, we put $N(A) = \mathrm{tr}(A^tA)$. Now we have from (3.25)

$$\begin{aligned}\tfrac{1}{2}\Delta\|\tau\|^2 &= \sum_{\alpha=n+1}(\tau^\alpha_{ijk})^2 + (nc - nH^2)\|\tau\|^2 + \sum_{\alpha=n+1} S^2_\alpha\\ &\quad + \sum_{\alpha,\beta=n+1} N(H^\alpha H^\beta - H^\beta H^\alpha)\,.\end{aligned}$$

Obviously, $N(H^\alpha H^\beta - H^\beta H^\alpha) \geqq 0$. Let

$$(p-1)\sigma_1 = \sum_{\alpha=n+1} S_\alpha = \|\tau\|^2,\ [(p-1)(p-2)\sigma_2]/2 = \sum_{\substack{\alpha<\beta\\ \alpha,\beta=n+1}} S_\alpha S_\beta$$

Thus we have

$$\sum_{\alpha=n+1} S^2_\alpha = (p-1)\sigma_1^2 + (p-1)(p-2)(\sigma_1^2 - \sigma_2)\,,$$

$$(p-1)^2(p-2)(\sigma_1^2 - \sigma_2) = \sum_{\substack{\alpha<\beta,\\ \alpha,\beta=n+1}} (S_\alpha - S_\beta)^2\,.$$

Hence, $\sum_{\alpha=n+1} S^2_\alpha = [1/(p-1)]\|\tau\|^4 + [1/(p-1)] \sum_{\substack{\alpha<\beta\\ \alpha,\beta=n+1}} (S_\alpha - S_\beta)^2$. Thus we obtain

$$\tfrac{1}{2}\Delta\|\tau\|^2 \geqq (nc - nH^2)\|\tau\|^2 + [1/(p-1)]\|\tau\|^4\,.$$

We make use of the similar methods of proof of $\|\mu\|^2$ for $\|\tau\|^2$. We have $\|\tau\|^2 \equiv 0$. Hence, M^n is totally umbilical.

Remark. When $n = 2$, condition (3.1) is best possible from [1] and [9]. When $n \geqq 3$, condition (3.2) is also best possible from following example.

Example. We consider Riemannian product $H^1(c_1) \times S^{n-1}(c_2)$, where $n \geqq 3$, $c_1 = (2-n)c$ and $c_2 = [(n-2)/(n-1)]c$. Because $H^1(c_1) \times S^{n-1}(c_2) \subset S^{n+1}_1(c)$, then we can get $n^2H^2 = 4(n-1)c$.

References

1. Akutagawa, K.: On space-like hypersurfaces with constant mean curvature in the de Sitter space. Math. Z. **196**, 13–19 (1987)
2. Calabi, E.: Examples of Bernstein problems for some nonlinear equations. Proc. Symp. Pure Appl. Math. **15**, 223–230 (1970)
3. Cheng, S.Y., Yau, S.T.: Maximal space-like hypersurfaces in the Lorentz-Minkowski. Ann. Math. **104**, 407–419 (1976)
4. Goddard, A.J.: Some remarks on the existence of spacelike hyperbolic of constant mean curvature. Math. Proc. Camb. Philos. Soc. **82**, 489–495 (1977)
5. Ishihara, T.: Maximal spacelike submanifolds of pseudo-hyperbolic space with second fundamental form of maximal length (Preprint)
6. Okumura, M.: Hypersurfaces and a pinching problem on the second fundamental tensor. Am. J. Math. **96**, 207–213 (1974)
7. Omori, H.: Isometric immersion of Riemannian manifolds. J. Math. Soc. Jpn **19**, 205–214 (1967)
8. O'Neill, B.: Semi-Riemannian Geometry. New York: Academic Press 1983
9. Ramanathan, J.: Complete space-like hypersurfaces of constant mean curvature in the de Sitter space. Indiana Univ. Math. J. **36**, 349–359 (1987)
10. Treibergs. T.E.: Entire hypersurfaces of constant mean curvature in Minkowski 3-space. Invent. Math. **66**, 39–56 (1982)
11. Yau, S.T.: Harmonic functions on complete Riemannian manifolds. Comm. Pure and Appl. Math. **28**, 201–228 (1975)

Math. Z. 206, 341–344 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Picard numbers of surfaces in 3-dimensional weighted projective spaces

A.J. de Jong* **and J.H.M. Steenbrink**

Department of Mathematics, Catholic University, Toernooiveld, NL-6525 ED Nijmegen, The Netherlands

Received January 21, 1990

Let $q_0, \ldots, q_3$ be positive integers. Consider the weighted complex projective 3-space $\mathbb{P} = \mathbb{P}(q_0, q_1, q_2, q_3) = \mathrm{Proj}(\mathbb{C}[x_0, x_1, x_2, x_3])$ where x_i has weight q_i. We will assume that $\gcd\{q_\alpha, q_\beta, q_\gamma\} = 1$ whenever $\#\{\alpha, \beta, \gamma\} = 3$ (this is not a restriction in fact, see [3] §1).

For $d \in \mathbb{N}$ we let $\mathscr{F}_d$ be the family of surfaces of degree d in $\mathbb{P}$, i.e. the schemes of zeros of sections in $\mathcal{O}_{\mathbb{P}}(d)$. Such a section is called *quasi-smooth* if its affine cone in $\mathbb{C}^4$ has an isolated singular point at the origin. If $X \in \mathscr{F}_d$ is quasi-smooth, then X is a V-manifold ([5], Lemma 1 p. 216). Hence $H^2(X, \mathbb{Q})$ has a pure Hodge structure and we can ask

Question. If the general element $x \in \mathscr{F}_d$ is quasi-smooth and has $p_g(X) > 0$, does it follow that the generic member of $\mathscr{F}_d$ has Picard number $\rho = 1$?

The classical Noether–Lefschetz theorem says that the answer is yes for surfaces in $\mathbb{P}^3(\mathbb{C})$. The condition $p_g(X) > 0$ is equivalent with $d \geqq 4$. In [6] the second author dealt with the case $\mathbb{P}(1, 1, a, b)$, $gcd(a, b) = 1$, using a generalization of the notion of Lefschetz pencils. In [2], this question is answered affirmatively for $d = km$, $m = lcm(q_0, q_1, q_2, q_3) > 1$ and $k \geqq 2$. We would like to mention here that his method also applies to the case $d = km + s$, $k \geqq 2$, where $s = q_0 + q_1 + q_2 + q_3$. However, it cannot work for all degrees d big enough because the use of Delorme's theorem ([3], Prop. 2.2) automatically gives the restriction $d = \mathrm{km}$ or $d = \mathrm{km} + s$.

In this article we will describe a method which can deal also with the case $d = m$ (but we do not solve this case completely).

The method of [2] is based on the study of the map

$$\mu\colon H^1(X, \Theta_X)_0 \otimes H^{1,1}_{\mathrm{prim}}(X) \to H^{0,2}(X)$$

for $X \in \mathscr{F}_d$ quasi-smooth. Here $H^1(X, \Theta_X)_0$ is the image of the Kodaira–Spencer

* Supported by the Netherlands Organization for Scientific Research (N.W.O.)

mapping for the family $\mathscr{F}_d$ and the subscript prim denotes primitive cohomology. One has the following

Criterion 1. *If for any quasi-smooth $X \in \mathscr{F}_d$ the map*

$$ad(\mu): H^{1,1}_{\text{prim}}(X) \to \text{Hom}(H^1(X, \Theta_X)_0, H^{0,2}(X))$$

$$\xi \mapsto (\theta \mapsto \mu(\theta, \xi))$$

is injective, then a generic $X \in \mathscr{F}_d$ has $\rho = 1$ ([1]).

Using the pairing Q of [1] it follows that this is equivalent to the surjectivity of the natural map

$$^t\mu: H^{2,0}(X) \otimes H^1(X, \Theta_X)_0 \to H^{1,1}_{\text{prim}}(X) .$$

However, as shown in [2], surjectivity of this map does not hold in general for $d = m$. To overcome this difficulty we put $H^{1,1}_{\text{prim}}(X, \mathbb{Q}) = H^{1,1}_{\text{prim}}(X) \cap H^2(X, \mathbb{Q})$ (i.e. the set of primitive Hodge classes) and use

Criterion 2. *If for any quasi-smooth $X \in \mathscr{F}_d$ the mapping*

$$\text{ad}(\mu): H^{1,1}_{\text{prim}}(X, \mathbb{Q}) \to \text{Hom}(H^1(X, \Theta_X)_0, H^{0,2}(X))$$

is injective, then a generic $X \in \mathscr{F}_d$ has $\rho = 1$.

Let us consider the Pham–Brieskorn variety of degree m in $\mathbb{P}$: $X = V(F)$ with

$$F(x_0, x_1, x_2, x_3) = x_0^{m/q_0} + x_1^{m/q_1} + x_2^{m/q_2} + x_3^{m/q_3} .$$

Then X is quasi-smooth.

Recall that $\mathbb{P} = (\mathbb{C}^4 \setminus \{0\})/\mathbb{C}^*$ where the action of $t \in \mathbb{C}^*$ on $x \in \mathbb{C}^4$ is given by $t.x = (t^{q_0}x_0, \ldots, t^{q_3}x_3)$. Hence $(\mathbb{C}^*)^4$ acts on $\mathbb{P}$ and the subgroup $\prod_i \mu_{m/q_i}$ acts on X. Let G be the image of this group in $\text{Aut}(X)$. We have the exact sequence

$$1 \to \mu_m \to \prod_i \mu_{m/q_i} \to G \to 1$$

$$\zeta \mapsto \prod_i \zeta^{q_i} .$$

The character group of G can be naturally identified with

$$\hat{G} = \{(a_0, a_1, a_2, a_3) \in (\mathbb{Z}/m\mathbb{Z})^4 \mid a_i \equiv 0 \pmod{q_i} \text{ and } \sum_i a_i = 0\} .$$

As in [4] we use the notation

$$V(\alpha) = \{\xi \in H^2_{\text{prim}}(X, \mathbb{C}) \mid g^*(\xi) = \alpha(g)\xi \text{ for all } g \in G\} \quad (\alpha \in \hat{G}) ;$$

$$\mathscr{A} = \{\alpha = (a_0, a_1, a_2, a_3) \in \hat{G} \mid a_i \neq 0 \text{ for all } i\} .$$

For $\alpha \in \mathscr{A}$ we set $l(\alpha) = \sum_i \langle \alpha_i \rangle$ where $\langle \alpha_i \rangle$ is the unique representative of α_i in the interval $[0, m-1]$. Exactly as in [4] we get

Theorem. (*i*) $\dim V(\alpha) = 1$ *for* $\alpha \in \mathscr{A}$ *and else* $\dim V(\alpha) = 0$;

(*ii*) $H^{2,0}(X) = \sum_{l(\alpha)=m} V(\alpha)$, $H^{1,1}_{\text{prim}}(X) = \sum_{l(\alpha)=2m} V(\alpha)$,

$$H^{0,2}(X) = \sum_{l(\alpha)=3m} V(\alpha);$$

(*iii*) *let* $\mathscr{B} = \{\alpha \in \mathscr{A} \mid l(t\alpha) = 2m \text{ for all } t \in (\mathbb{Z}/m\mathbb{Z})^*\}$.

Then $H^{1,1}_{\text{prim}}(X, \mathbb{Q}) \otimes \mathbb{C} = \sum_{\alpha \in \mathscr{B}} V(\alpha)$.

One way to prove this theorem is to use the fact that X is the quotient of a Fermat surface in ordinary projective 3-space and the description of the cohomology of Fermat varieties from [4]. Another way is to use the identifications

$$H^{2,0}(X) \cong R_{m-s}.\Omega/F, \quad H^{1,1}_{\text{prim}}(X) \cong R_{2m-s}.\Omega/F^2, \quad H^{0,2}(X) \cong R_{3m-s}.\Omega/F^3$$

where

$$\Omega = dx_0 \wedge dx_1 \wedge dx_2 \wedge dx_3;$$

$$R = \mathbb{C}[x_0, \ldots, x_3]/(\partial F/\partial x_0, \ldots, \partial F/\partial x_3)$$

(see [5]). In fact, for $\alpha \in \mathscr{A}$ a generator for $V(\alpha)$ is

$$\omega_\alpha := \operatorname{res}_X x_0^{\langle \alpha_0 \rangle / q_0 - 1} \ldots x_3^{\langle \alpha_3 \rangle / q_3 - 1} \Omega / F^{l(\alpha)/m}.$$

Moreover, using the identification $H^1(X, \Theta_X)_0 \cong R_m$ (see [1]) we find that the map μ is identified up to sign with the multiplication map

$$R_m \otimes R_{2m-s} \to R_{3m-s}.$$

As R_k has a natural basis of monomials for any $k \in \mathbb{N}$ and the product of monomials in R is again a monomial (or zero), it follows that the kernel of $ad(\mu)$ is spanned by a set of monomials in R_{2m-s}.

Next remark that if $\eta \in H^{1,1}_{\text{prim}}(X, \mathbb{Q})$, $\eta = \sum_{\alpha \in \mathscr{B}} \lambda_\alpha \omega_\alpha$, then $\lambda_\alpha \neq 0 \Leftrightarrow \forall t \in (\mathbb{Z}/m\mathbb{Z})^*$: $\lambda_{t\alpha} \neq 0$. Indeed, η lies in a $\mathbb{Q}$-rational representation space of G, so if it has a non-zero projection to some $V(\alpha)$, then it has a non-zero projection to each conjugate character space $V(t\alpha)$, $t \in (\mathbb{Z}/m\mathbb{Z})^*$ (this argument is used by Shioda to prove part (iii) of the theorem). Using the compatibility of the pairing Q from [1] with the action of the group G we can conclude:

Criterion 3. *If for all* $\beta \in \mathscr{B}$ *there exist* $t \in (\mathbb{Z}/m\mathbb{Z})^*$ *and* $\alpha \in \mathscr{A}$ *with* $l(\alpha) = 3m$ *such that* $\langle t\beta_i \rangle \leqq \langle \alpha_i \rangle$ *for* $i = 0, \ldots, 3$, *then a generic* $X \in \mathscr{F}_m$ *has* $\rho = 1$.

We will now prove that this criterion holds for certain choices of weights $q_0, \ldots, q_3$. In fact, a computer search has given no case where the criterion does not apply if all q_i are $\leqq 50$. According to [2] there are three cases to consider (up to permutation of the weights):

(i) $lcm(q_0, q_1) < m$ and $lcm(q_2, q_3) < m$;
(ii) $q_i = m$ for some i:
(iii) $q_0 = 1$ and $q_i \neq m$ for all i.

Case (ii) is easy, for then $p_g(X) = 0$. We can handle case (iii). Suppose that $m \geqq s$ (otherwise $p_g(X) = 0$) and $q_0 = 1$. Let $\beta \in \mathscr{B}$. By possibly changing β into $-\beta$ we may suppose that $\langle \beta_0 \rangle \geqq m/2$. Put $i_j = \langle \beta_j \rangle / q_j - 1, j = 0, \ldots, 3$. We must find a factor of degree m of $\prod_j x_j^{i_j}$. As $q_i \neq m$ for all i we have $q_i \leqq m/2$ and so all conditions of the sub lemma of [2] p. 186 are fulfilled and the claim follows.

Case (i) we cannot do in complete generality. Again we start with $\beta \in \mathscr{B}$. Write $m' = lcm(q_0, q_1)$, $m'' = lcm(q_2, q_3)$, $i_j = \langle \beta_j \rangle / q_j - 1$ and

$$\langle \beta_0 \rangle = m' s_0 + \rho_0, \quad 0 < \rho_0 \leqq m' ;$$

$$\langle \beta_1 \rangle = m' s_1 + \rho_1, \quad 0 < \rho_1 \leqq m' ;$$

$$\langle \beta_2 \rangle = m'' s_2 + \rho_2, \quad 0 < \rho_2 \leqq m'' ;$$

$$\langle \beta_3 \rangle = m'' s_3 + \rho_3, \quad 0 < \rho_3 \leqq m'' ;$$

As in [2] p. 186 we deduce that $\prod_j x_j^{i_j}$ has a factor of degree m as soon as

$$m'(s_0 + s_1) + m''(s_2 + s_3) > 2m - m' - m''$$

which is equivalent to

$$\sum_j \rho_j < m' + m'' \ (*)$$

Suppose that we have $\rho_0 \neq m'$, $\rho_1 \neq m'$, $\rho_2 \neq m''$, $\rho_3 \neq m''$ and $\sum_j \rho_j \neq m' + m''$. Then $(*)$ will be true either for β or for $-\beta$.

Of course this argument does not cover all possibilities. The main difficulty in case (i) remains to detect when one has $p_g(X) = 0$.

References

1. Carlson, J., Green, M., Griffiths, P., Harris, J.: Infinitesimal variation of Hodge structure I. Compos. Math. **50**, 109–205 (1983)
2. Cox, D.: Picard numbers of surfaces in 3-dimensional weighted projective spaces. Math. Z. **201**, 183–189 (1989)
3. Delorme, Ch.: Espaces projectifs anisotropes. Bull. Soc. Math. Fr. **103**, 203–223 (1975)
4. Shioda, T.: The Hodge conjecture for Fermat varieties. Math. Ann. **245**, 175–184 (1979)
5. Steenbrink, J.H.M.: Intersection form for quasi-homogeneous singularities. Compos. Math. **34**, 221–223 (1977)
6. Steenbrink, J.H.M.: On the Picard group of certain smooth surfaces in weighted projective spaces. In: Aroca, J.M., Buchweitz, R., Giusti, M., Merle, M. (ed.) Algebraic Geometry. Proceedings, La Rábida 1981. (Lecture Notes Math. vol. 961, pp. 302–313) Berlin Heidelberg New York Springer 1982

Note added in proof

D. Cox communicated to us that he has determined all cases where a generic $X \in \mathscr{F}_d$ has $p_g(X) = 0$. This occurs exactly when, after permutation, either $q_0 = m$ or $(q_0, q_1, q_2, q_3) = (1, 1, 2, 3), (1, 2, 2, 3), (1, 3, 4, 6), (1, 6, 10, 15), (1, 2, u, u)$ with u odd or (a, a, b, b) with $a > 1$, $b > 1$ and $gcd(a, b) = 1$.

Math. Z. 206, 345–350 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Knot and link projections in 3-manifolds*

Richard K. Skora**

Department of Mathematics, State University of New York, Stony Brook, NY 11794-3651, USA

Received September 26, 1989; in final form May 26, 1990

The study of knots and links in the 3-sphere is greatly enhanced by a theorem of K. Reidemeister [Re1], [Re2] and J.W. Alexander and G.B. Briggs [Al-Br]. This theorem says that two links in the 3-sphere are combinatorially equivalent if and only if the projection of one link can be deformed by certain operations to a projection of the other link. The operations and the theorem have come to be known as the Reidemeister Operations and the Reidemeister Theorem, respectively. Some classical link invariants [Re1], [Re2], [Al2], [Al-Br] and many of the new link polynomials [Jo], [F-Y-H-L-M-O], [P-T], [Ka] are easily defined and calculated from a link projection.

In this paper we extend the definition of a link projection to an arbitrary closed 3-manifold and we prove a generalization of the above theorem. This result is analogous to the generalization of the Markov Theorem [Mar] proved in [Sk2]. Both results were announced in [Sk1].

In §1 we define link projection. Our definition of projection depends on the choice of an *axis*, i.e. a fibered knot or link in the ambient manifold. Every closed 3-manifold has a fibered knot. The classical notion of link projection is recovered by taking the unknot in S^3 as an axis.

The generalized Reidemeister Operations are defined in §2. The first three operations are local and are identical to the three Reidemeister Operations. The fourth operation is new. It is the only global operation and it incorporates the choice of axis into its definition. We state and prove the main result, Theorem 3.1, in §3.

1 The definition of knot and link projections

In this paper all manifolds and maps between manifolds are understood to be piecewise-linear. A *link* in a 3-manifold M is a 1-dimensional, closed submanifold. A *knot* is a link with one component.

* Supported in part by a SUNY Research Development Grant and by an NSF Postdoctoral Research Fellowship.

** *Present address:* Department of Mathematics, Mathematics Building, Columbia University in the City of New York, New York, N.Y. 10027, USA

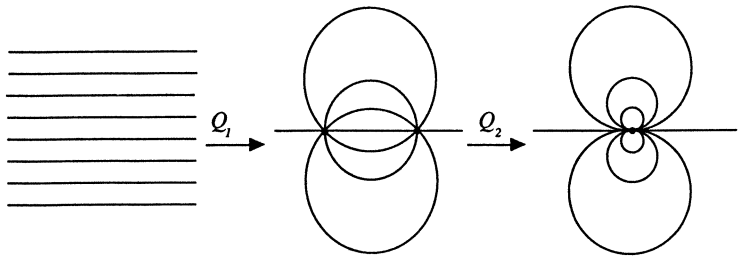


Fig. 1. Fibering $M-S$

Let D be a disk in M and let β, β' be arcs in ∂D with disjoint interior and union equal to ∂D. Let L, L' be links in M. If $\beta = L \cap D$, $\beta' = L' \cap D$ and $L' = (L-\beta) \cup \beta'$, then say L' is obtained from L by an *elementary deformation* through D.

Two links are *combinatorially equivalent* if there is a sequence of links

$$L = L_0, L_1, \ldots, L_n = L',$$

such that for each k, L_{k+1} is obtained from L_k by an elementary deformation.

It is well known that two links L, L' are combinatorially equivalent if and only if there is an isotopy $H: M \times [0, 1] \to M$, such that $H_0 = Id_M$, $H_1(L) = L'$. It is easy to see that combinatorial equivalence implies the latter. The proof of the converse is an easy adaptation of Theorem 6.2 (Isotopy by Moves) in [Hud].

Let $Q: \mathbb{R}_2 \times (0, 1) \to S^3 - \{+\infty\}$ be a homeomorphism. Notice this fibering looks quite strange near $+\infty$. The projection $\tilde{p}: \mathbb{R}^2 \times (0, 1) \to \mathbb{R}^2$ covers a projection $p: S^3 - \{+\infty\} \to \mathbb{R}^2$. This is the projection used to define the projection of a link in S^3.

We will construct an analogous fibering and projection for an arbitrary closed 3-manifold. The reward will be that the proof of our main theorem will be easy.

Alexander [Al1] showed that every orientable, closed 3-manifold has a fibered link. Myers [My] and Gonzales-Acuna [GA] showed that the link may be taken to be a knot. And Berstein and Edmonds [Be-Ed] showed that every closed 3-manifold contains a fibered knot. Harer [Ha] described the difference between all fibered links in a 3-manifold.

Let M be a closed 3-manifold and fix a fibered link A in M. Call A the *axis*. By definition there is a compact surface with boundary F and a homeomorphism $\psi: F \to F$ which is the identity in a neighborhood of ∂F, such that

$$\frac{(F, \partial F) \times [0, 1]}{\sim} = (M, A),$$

where $(\psi(x), 0) \sim (x, 1)$, for all $x \in F$; and $(x, s) \sim (x, t)$, for all $x \in \partial F$, $s, t \in [0, 1]$. Let

$$Q_1: F \times [0, 1] \to M$$

be the corresponding quotient map. See Figure 1.

We will now define a better set of coordinates for M. Let $F_0 = Q_1(F \times \{0\})$ and S be a 1-dimensional complex in $\mathring{F}_0$, such that $S \subset F_0$ is a homotopy equivalence. Then there is a quotient map $q\colon \partial F_0 \times ([0, 1], \{1\}) \to (F_0, S)$, where $q|\partial F_0 \times [0, 1)$ is a homeomorphism. This determines a retract $r\colon F_0 \to S$ by $q(w, t) \mapsto q(w, 1)$.

Extend r to a quotient map

$$Q_2\colon M \to M,$$

so that $Q_2|M - F_0$ is a homeomorphism (see Figure 1). Let $Q = Q_2 \circ Q_1 : (F, \partial F) \times [0, 1] \to (M, S)$. Notice $Q|\mathring{F} \times (0, 1)\colon \mathring{F} \times (0, 1) \to M - S$ is a homeomorphism. The projection $\tilde{p}\colon \mathring{F} \times (0, 1) \to \mathring{F}$ covers a projection $p\colon M - S \to \mathring{F}$.

This projection has the following nice property. Suppose some disk meets S transversely at the point a, and N is a neighborhood of a. Then there is a disk $D \subset N$ intersecting S transversely at a which satisfies

(*) (i) $\alpha_0 - D = \alpha_1 - D$;
(ii) $p|\alpha_0$, $p|\alpha_1$ are proper embeddings;
(iii) $p(\alpha_0) = \psi(p(\alpha_1))$; and
(iv) $p(\alpha_0) = p(D \cap Q(\mathring{F} \times (0, \frac{1}{2}]))$ and $p(\alpha_1) = p(D \cap Q(\mathring{F} \times [\frac{1}{2}, 1)))$, where $\alpha_0 = \partial(D \cap Q(\mathring{F} \times (0, \frac{1}{2}])), \alpha_1 = \partial(D \cap Q(\mathring{F} \times [\frac{1}{2}, 1)))$.

1.1 Definition. Let L be a link in $M - S$. Then $p|L\colon L \to \mathring{F}$ is a *projection* for L if it is a regular map.

Recall that $p|L$ is *regular* if no vertex of L is a critical point and every singular point is a transverse double point.

Every link is approximable arbitrary closely by a combinatorially equivalent link L, such that $p|L$ is a projection.

Notice that a projection implicitly includes both the link L and the surface $\mathring{F}$. Thus it includes sufficient information to reconstruct the link upto combinatorial equivalence.

It may be convenient to think of a link projection informally. Let $p|L$ be a link projection and β be an arc in L. Say $p(\beta)$ contains only *over crossings* if $s \leqq t$, whenever $Q(x, s) \in L$, $Q(x, t) \in \beta$. Similarly, say $p(\beta)$ contains only *under crossings* if $s \geqq t$, whenever $Q(x, s) \in L$, $Q(x, t) \in \beta$. One may think of a link projection simply as $p(L)$ together with over/under crossing information at the singular points.

2 The generalized Reidemeister operations

Two projections $p|L$, $p|L'$ are *equivalent* if there is a sequence of link projections $p|L = p|L_0, p|L_1, \ldots, p|L_n = p|L'$, such that for each k, L_{k+1} is obtained from L_k by an elementary deformation through D_k; and $p|D_k$ is an embedding whose interior either misses $p(L_k)$ or intersects it in one embedded arc running from $p(\partial D) \cap p(L_k)$ to $p(\partial D) \cap p(L_{k+1})$.

The definitions of the operations are motivated by the fact that though two links may be combinatorially equivalent, their projections may not be equivalent. The operations are defined to exactly capture the exceptions which occur in trying to deform one projection to the other.

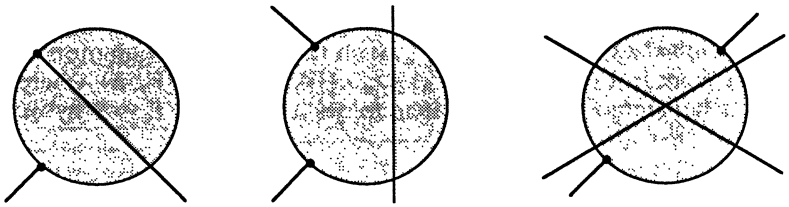


Fig. 2. Reidemister Operations 1, 2 and 3

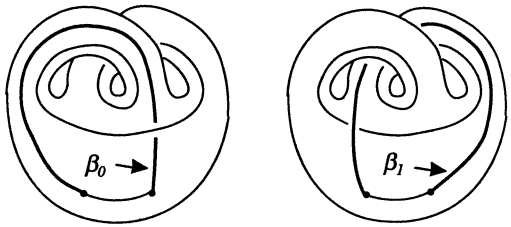


Fig. 3. Operation 4

2.1 Definition. Let $p|L$, $p|L'$ be two link projections. Say that $p|L'$ is obtained from $p|L$ by *Reidemeister Operation 1, 2,* or *3* if L' is obtained from L by an elementary deformation through D; and $p(L)$, $p(L')$, $p(D)$ looks like Figure 2: 1, 2, or 3, respectively.

If two links in S^3 which are disjoint with $+\infty$ are combinatorially equivalent, then there is a sequence of elementary deformations from one to the other through disks which miss $+\infty$. This is false for arbitrary M and S. Thus we need one more operation.

2.2 Definition. Let $p|L$, $p|L'$ be two link projections. Say that $p|L'$ is obtained from $p|L$ by *operation* 4 if L' is obtained from L by an elementary deformation through D; and D satisfies (*) of §1, where $D \cap L \cap L' = \partial D \cap Q(F \times \{\frac{1}{2}\})$.

We should point out that there are only finitely many possibilities for $p(\alpha_0)$ upto proper isotopy. The possible $p(\alpha_0)$'s is determined by the choice of spine S and the retract $r\colon F_0 \to S$. One gets one $p(\alpha_0)$ for each 1-simplex in S.

Let us explain informally how operation 4 changes the image of the link projection. Suppose $p|L$, $p|L'$ are link projections and $p|L'$ is obtained from $p|L$ by operation 4. Let D be as in Definition 2.2 and $\beta = D \cap L$, $\beta' = D \cap L'$. For specificity suppose $\beta = \alpha_0 \cap L$, $\beta' = \alpha_1 \cap L'$. Then the properly embedded open arcs $y = p(\alpha)$, $y' = p(\alpha_1)$ satisfy

(**) (i) $y - p(\beta) = y' - p(\beta')$
(ii) y, y' are embedded;
(iii) $y = \psi(y')$; and
(iv) $y \cap p(L) = p(\beta)$ contains only under crossing and $y' \cap p(L') = p(\beta')$ contains only over crossings.

There is a similar statement for $\beta = \alpha_1 \cap L$, $\beta' = \alpha_0 \cap L'$. See Figure 3.

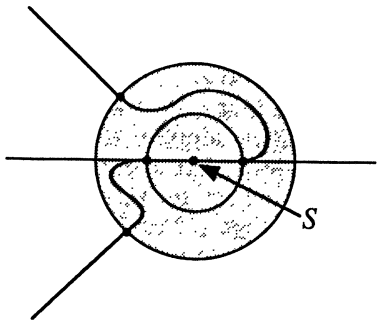


Fig. 4. Subdivision

Conversely, suppose $p|L$, $p|L'$ are link projections and $p(L')=p(L-\beta)\cup p(\beta')$, for arcs $\beta\subset L$, $\beta'\subset L'$. If there exists properly embedded open arcs y, y' in $\hat{F}$ satisfying (**), then L is obtained from L' through a sequence of operations 1, 2, 3 and 4. Operations 1, 2 and 3 may be necessary because operation 4 is very specialized.

3 The main theorem

3.1 Theorem. *Let $p|L$, $p|L'$ be link projections. Then L is combinatorially equivalent to L' if and only if there is a sequence of link projections*

$$p|L=p|L_0, p|L_1, \ldots, p|L_n=p|L',$$

such that for each k, either $p|L_{k+1}$ is equivalent to $p|L_k$; or $p|L_{k+1}$ is obtained from $p|L_k$ by operation 1, 2, 3 or 4.

Proof. Suppose there is a sequence of link projections

$$p|L=p|L_0, p|L_1, \ldots, p|L_n=p|L',$$

such that for each k; either $p|L_{k+1}$ is obtained from $p|L_k$ by either operation 1, 2, 3 or 4. By definition L is combinatorially equivalent to L'.

Conversely, suppose L is combinatorially equivalent to L'. By definition there is a sequence of links

$$L=L_0, L_1, \ldots, L_n=L',$$

such that for each k, L_{k+1} is obtained from L_k by an elementary deformation through D_k.

We will first focus on operation 4. Without loss of generality suppose each disk is transverse to S and $p|L_k$ is a projection. After possibly subdividing and general positioning the disks we may further suppose that each D_k either misses S or satisfies property (*) of §1, where $D_k\cap L_k\cap L_{k+1}=\partial D_k\cap Q(F\times\{\frac{1}{2}\})$. See Figure 4. Thus our sequence satisfies for each k either D_k misses S; or L_{k+1} is obtained from L_k by operation 4.

From above it sufficies to assume that each disk misses S. We will now focus on operations 1, 2 and 3. The rest of the proof is identical to Reidemenster's proof and Alexander-Briggs' proof. After possibly subdividing and general positioning the disks we may suppose that $p|D_k$ is an embedding. And after possibly subdividing one last time, we may suppose that either $p|L_{k+1}$ is equivalent to $p|L_k$; or $p(L_k)$, $p(L_{k+1})$, $p(D_k)$ looks like Figure 2. Thus our sequence satisfies for each k either $p|L_{k+1}$ is equivalent to $p|L_k$; or L_{k+1} is obtained from L_k by one of the Reidemeister operations 1, 2 or 3. □

When $M=S^3$ and $S=\{+\infty\}$ is a point, the D_k's miss S and only operations 1, 2 and 3 are used.

Acknowledgement. I would like to thank Charles Livingston for helpful conversations.

References

[Al1] Alexander, J.W.: A Lemma on Systems of Knotted Curves. Proc. Nat. Acad. Sci. USA **9**, 93–95 (1923)

[Al2] Alexander, J.W.: Topological Invariants of Knots and Links. Trans. Am. Math. Soc. **30**, 275–306 (1928)

[Al-Br] Alexander, J.W., Briggs, G.B.: On Types of Knotted Curves. Ann. Math., II. Ser. **28**, 562–586 (1927)

[Be-Ed] Berstein, I., Edmonds, A.L.: On the Construction of Branched Coverings of Low-Dimensional Manifolds. Trans. Am. Math. Soc. **247**, 87–124 (1979)

[F-Y-H-L-M-O] Freyd, P., Yetter, D., Hoste, J., Lickorish, W.B.R., Millett, K., Ocneanu, A.: A New Polynimial Invariant of Knots and Links. Bull. Am. Math. Soc. **12**, 239–246 (1985)

[GA] González-Acuña, F.: (unpublished)

[Ha] Harer, J.: How to Construct all Fibered Knots. Topology **21**, 263–280 (1982)

[Hud] Hudson, J.F.P.: Piecewise Linear Topology. New York: Benjamin 1969

[Jon] Jones, V.F.R.: A polynomial Invariant for Knots via von Neumann Algebras. Bull. Am. Math. Soc. **12**, 103–111 (1985)

[Ka] Kauffman, L.H.: State Models and the Jones Polynomial. Topology **26**, 395–407 (1987)

[Mar] Markov, A.A.: Über die freie Äquivalenz geschlossener Zöpfe. Recueil Mathematique Moscou **1**, 73–78 (1935)

[My] Myers, R.: Open Book Decomposition of 3-Manifolds. Proc. Am. Math. Soc. **72**, 397–402 (1978)

[P-T] Przytycki, J.H., Traczyk, P.: Invariants of Links of Conway Type. Kobe J. Math. **4**, 115–139 (1988)

[Re1] Reidemeister, K.: Elementare Begründung der Knotentheorie. Abh. Math. Semin. Univ. Hamb. **5**, 24–32 (1925)

[Re2] Reidemeister, K.: Knotentheorie. Ergeb. Math. Grenzgeb. **1**, Berlin: Julius Springer 1932

[Sk1] Skora, R.K.: Research Announcement: Knots and Links in 3-Manifolds. Preprint

[Sk2] Skora, R.K.: Closed Braids in 3-Manifolds. Preprint

Math. Z. 206, 351–361 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On certain properties of the relative entropy of states of operator algebras

Dénes Petz*,**

Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

Received October 12, 1988; in final form March 26, 1990

In 1977 Araki defined the relative entropy of two normal positive functionals of a von Neumann algebra by means of the relative modular operator. Let the von Neumann algebra M be in standard form and set Ω in the natural positive cone as the vector representative of the normal positive functional ω. If φ is another normal positive functional with relative modular operator $\Delta(\varphi, \omega)$ with respect to ω then

$$S(\varphi, \omega)=\begin{cases} -\langle \log \Delta(\varphi, \omega)\Omega, \Omega\rangle & \text{if } \operatorname{supp}\varphi \geqq \operatorname{supp}\omega \\ +\infty & \text{otherwise.} \end{cases} \tag{1}$$

The convexity and monotonicity of the relative entropy functional are easily read off from the following variational expression of Kosaki ([10]).

$$S(\varphi, \omega)=\sup\sup\left\{\omega(I)\log n-\int_{1/n}^{\infty} t^{-1}\,\omega(y(t)^* y(t))+t^{-2}\,\varphi(x(t)\,x(t)^*)\,dt\right\}. \tag{2}$$

Here $y(t)=I-x(t)$, the first sup is taken over the positive integers and the second one is over all step functions x: $[1/n, \infty)\to M$ such that the range of x is finite and $x(t)=I$ for t large enough.

There are several ways for the extension of the relative entropy functional to two positive functionals of a C^*-algebra, and all reasonable ones give the same notion. If Kosaki's formula serves as a definition then the monotonicity in the first variable and joint convexity follow straightforwardly.

The aim of the present paper is to establish certain properties of the relative entropy functional in the general context of C^*-algebras. Most of these properties have been observed under more restrictive conditions in earlier papers. (The celebrated strong subadditivity of the entropy gives such an example.) The gener-

* On leave of absence from the Mathematical Institute HAS, Budapest. This research was partially supported by the Hungarian National Foundation for Scientific Research grant No. 1171

** *Present address:* Mathematical Institute of the Hungarian Academy of Sciences, P.O.B. 127, H-1364 Budapest, Hungary

al approach is applied to prove the existence of the mean relative entropy for a noncommutative stationary Markov chain with conditional expectations.

It is assumed throughout the paper that C^*-algebras are unital and subalgebras do contain the unit of the superalgebra.

It follows from the monotonicity of $S(\varphi, \omega)$ under strongly positive (i.e., Schwarz positive) unital mappings that for a positive contraction $a \in M$

$$\omega(a)(\log \omega(a) - \log \varphi(a)) + \omega(b)(\log \omega(b) - \log \varphi(b)) \leqq S(\varphi, \omega) \tag{3}$$

where b is written for $I - a$. It is also clear by a density argument that if π is a faithful representation of a C^*-algebra $\mathscr{A}$ and the positive functionals φ and ω admit a normal extension $\tilde{\varphi}$ and $\tilde{\omega}$ to $\pi(\mathscr{A})''$ then $S(\varphi, \omega) = S(\tilde{\varphi}, \tilde{\omega})$.

If φ is a positive functional of the C^*-algebra $\mathscr{A}$ then we shall write $(\Phi, \pi_\varphi, H_\varphi)$ for the corresponding GNS triplet.

Lemma 1 *Let φ and ω be arbitrary positive functionals of the C^*-algebra $\mathscr{A}$. If there exists no normal positive functional $\tilde{\omega}$ on $\pi_\varphi(\mathscr{A})''$ such that $\tilde{\omega}(\pi_\varphi(a)) = \omega(a)$ for all $a \in \mathscr{A}$, then there is a net (x_i) of positive contractions in $\mathscr{A}$ such that*

$$\lim_i \varphi(x_i) = 0 \quad \textit{and} \quad \lim_i \omega(x_i) > 0.$$

Proof. Let π_0 be a faithful representation of the C^*-algebra $\mathscr{A}$ and set $\pi = \pi_0 \oplus \pi_\varphi \oplus \pi_\omega$. We denote by Z_Φ (Z_Ω) the smallest central projection in $M = \pi(\mathscr{A})''$ which leaves $0 \oplus \Phi \oplus 0$ ($0 \oplus 0 \oplus \Omega$) fixed. If the above described $\tilde{\omega}$ does not exist (that is, π_φ does not quasi-contain π_ω, cf. [3]) then Z_Ω is not smaller than Z_Φ. So $e = Z_\Omega(I - Z_\Phi)$ is a nonzero projection with the properties

$$\langle e(0 \oplus \Phi \oplus 0), (0 \oplus \Phi \oplus 0) \rangle = 0 \quad \text{and} \quad \langle e(0 \oplus 0 \oplus \Omega), (0 \oplus 0 \oplus \Omega) \rangle > 0.$$

The rest follows from the Kaplansky density theorem. There exists a net (x_i) of positive contractions in $\mathscr{A}$ such that $\pi(x_i)$ converges to e in the strong operator topology. □

As a consequence of Lemma 1 we may arrive at another definition of the relative entropy of arbitrary positive functionals ω and φ of a C^*-algebra $\mathscr{A}$. Let M be $\pi_\varphi(\mathscr{A})''$ and denote by $\tilde{\varphi}$ and $\tilde{\omega}$ the normal positive functionals on M such that $\tilde{\omega}(\pi_\varphi(a)) = \omega(a)$ and $\tilde{\varphi}(\pi_\varphi(a)) = \varphi(a)$ for all $a \in \mathscr{A}$ (provided that the later one exists). Then

$$S(\varphi, \omega) = \begin{cases} S(\tilde{\varphi}, \tilde{\omega}) & \text{if } \tilde{\omega} \text{ exists} \\ +\infty & \text{otherwise.} \end{cases} \tag{4}$$

To prove (4) it suffices to consider the case when $\tilde{\omega}$ does not exists. Then the net (x_i) of Lemma 1 is at our disposal. We set

$$s_i = \omega(x_i)(\log \omega(x_i) - \log \varphi(x_i)) + \omega(I - x_i)(\log \omega(I - x_i) - \log \varphi(I - x_i))$$

and due to (3) we have $s_i \leqq S(\varphi, \omega)$. One can easily check that the condition in Lemma 1 implies $s_i \to \infty$, therefore $S(\varphi, \omega)$ must be $+\infty$.

We recall that a conditional expectation is defined as a projection of norm one onto a subalgebra (see [18]; Chapter II).

Theorem 2 *Let M be a von Neumann algebra and M_1 its von Neumann subalgebra. Assume that there exists a faithful normal conditional expectation E of M onto M_1. If φ_1 and ω are states of M_1 and M, respectively then*

$$S(\varphi_1 \circ E, \omega) = S(\varphi_1, \omega | M_1) + S(\omega \circ E, \omega). \tag{5}$$

First we prove (5) in a special case.

Lemma 3 (5) *holds if φ_1 is faithful and there are positive constants λ and μ such that $\varphi_1 \leqq \lambda \omega_1 \leqq \mu \varphi_1$ (writing ω_1 for $\omega | M_1$).*

Proof. (5) is obvious if $S(\varphi_1 \circ E, \omega) = \infty$ and $S(\omega \circ E, \omega) = \infty$. Therefore we may assume that one of them is finite. Let us see that in this case all the three quantities in (5) are finite. Indeed,

$$S(\varphi_1 \circ E, \omega) \leqq S(\lambda \mu^{-1} \omega \circ E, \omega) = \log \mu - \log \lambda + S(\omega_1 \circ E, \omega)$$

$$S(\varphi_1, \omega_1) \leqq S(\lambda \mu^{-1} \omega_1, \omega_1) = \log \mu - \log \lambda$$

$$S(\omega \circ E, \omega) \leqq S(\lambda^{-1} \varphi_1 \circ E, \omega) = \log \lambda + S(\varphi_1 \circ E, \omega).$$

We use the differentiation formula ([8], Theor. 4) including the Radon-Nikodym cocycle of states:

$$S(\psi_1, \psi_2) = i \lim_{t \to +0} \frac{\psi_1([D\psi_1, D\psi_2]_t - I)}{t}$$

whenever ψ_1 and ψ_2 are normal states and $S(\psi_1, \psi_2)$ is finite. The chain rule of the Radon-Nikodyn cocycles ([18]; 3.5) says that

$$[D(\varphi_1 \circ E), D\omega]_t = [D(\varphi_1 \circ E), D(\omega \circ E)]_t [D(\omega \circ E), D\omega]_t .$$

Note that under the assumptions $\omega \circ E$ is faithful and

$$[D(\varphi_1 \circ E), D(\omega \circ E)]_t = [D\varphi_1, D\omega_1]_t$$

for every real t ([18], 10.5). It is convenient to use a short hand notation for the three cocycles, say, u_t, v_t, w_t, respectively.

We have

$$\frac{\omega(u_t) - 1}{t} = \frac{\omega(v_t) - 1}{t} + \frac{\omega(w_t) - 1}{t} + t^{-1} \langle (\omega_t - I)\Omega, (v_t^* - I)\Omega \rangle$$

and show that the third term of the right side tends to 0 as $t \to 0$. Here $(w_t - I)\Omega \to 0$ since the cocycle w_t is so-continuous. On the other hand, $v_t = [D\varphi_1, D\omega_1]_t$ and the function

$$it \to v_t \Omega$$

admits an analytical extension to the strip $\{z \in \mathbb{C}: -1/2 < \mathrm{Re}\, z < 1/2\}$ (see [7] or [18]; 3.13). Therefore, $t^{-1}(v_t - I)\Omega$ is bounded in a neighbourhood of 0. □

Lemma 4 (5) *holds if there are positive constants λ and μ such that $\varphi_1 \leqq \lambda \omega_1 \leqq \mu \varphi_1$.*

Proof. Let $p_1 = \operatorname{supp} \varphi_1$. We consider the von Neumann algebras $N = p_1 M p_1$ and $N_1 = p_1 M_1 p_1$. The mapping $F = E|N$ is a faithful normal conditional expectation of N onto N_1. Using the fact that

$$S(\psi_1, \psi_2) = S(\psi_1|p_1 M p_1, \psi_2|p_1 M p_1)$$

provided that $\operatorname{supp}\psi_1$, $\operatorname{supp}\psi_2 \leqq p_1$, we can reduce Lemma 4 to the faithful case contained in Lemma 3. □

Lemma 5 (5) *holds if* $\varphi_1 \leqq \lambda \omega_1$ *for certain positive constant* λ.

Proof. For $1 > \varepsilon > 0$, let $\nu_1^\varepsilon = (1-\varepsilon)\varphi_1 + \varepsilon\omega_1$. Then $\nu_1^\varepsilon \leqq ((1-\varepsilon)\lambda + \varepsilon)\omega_1$ and $\omega_1 \leqq \varepsilon^{-1}\nu_1^\varepsilon$. We can benefit from Lemma 4.

$$S(\nu_1^\varepsilon \circ E, \omega) = S(\nu_1^\varepsilon, \omega_1) + S(\omega \circ E, \omega).$$

Now we perform the limit $\varepsilon \to +0$. We may assume that $S(\omega_1 \circ E, \omega)$ is finite. By convexity we have

$$S(\nu_1^\varepsilon \circ E, \omega) \leqq (1-\varepsilon) S(\varphi_1 \circ E, \omega) + \varepsilon S(\omega_1 \circ E, \omega)$$

and infer

$$\limsup_{\varepsilon \to 0} S(\nu_1^\varepsilon \circ E, \omega) \leqq S(\varphi_1 \circ E, \omega).$$

A combination with lower semicontinuity yields

$$\lim_{\varepsilon \to 0} S(\nu_1^\varepsilon \circ E, \omega) = S(\varphi_1 \circ E, \omega).$$

By a similar argument

$$\lim_{\varepsilon \to 0} S(\nu_1^\varepsilon, \omega_1) = S(\varphi_1, \omega_1)$$

and the proof of the lemma is complete. □

Proof of Theorem 2 For $1 > \varepsilon > 0$ let $\psi^\varepsilon = (1-\varepsilon)\omega + \varepsilon\varphi_1 \circ E$. Then $\varphi_1 \leqq \varepsilon^{-1}(\psi^\varepsilon | M_1)$ and due to Lemma 5 we have

$$S(\varphi_1 \circ E, \psi^\varepsilon) = S(\varphi_1, \psi^\varepsilon | M_1) + S(\psi^\varepsilon \circ E, \psi^\varepsilon).$$

If we apply convexity and lower semicontinuity as in the proof of Lemma 5 then letting $\varepsilon \to +0$ we conclude (5). □

It must be noted that (5) was proved in [9, Theor. 3.2] under the restrictive condition that the subalgebra M_1 is in the centralizer of the faithful normal state $\varphi_1 \circ E$.

Let E be as in Theorem 2 and φ a normal state such that $\varphi \circ E = \varphi$. For a fixed normal state ω_1 of M_1 set $\mathcal{K}$ for the set of all normal extensions of ω_1 to M. $\mathcal{K}$ is a convex set and we can ask which $\omega \in \mathcal{K}$ minimizes $S(\varphi, \omega)$. If there exists a $\psi \in \mathcal{K}$ such that $S(\varphi, \psi)$ is finite then the minimizer is unique and it is $\omega_1 \circ E$. This is evident from (5).

Theorem 2 may be formulated also in the setting of C^*-algebras as follows. If φ is a separating state of the C^*-algebra $\mathcal{A}$ admitting a conditional expectation F onto its subalgebra $\mathcal{A}_1$ which preserves φ then

$$S(\varphi, \omega) = S(\varphi|\mathcal{A}_1, \omega|\mathcal{A}_1) + S(\omega \circ F, \omega) \tag{6}$$

holds for every state ω on $\mathscr{A}$. (Recall that a state is termed separating if the cyclic vector in the GNS construction is separating for the generated von Neumann algebra.) To see (6) it is sufficient to observe that F induces a faithful normal conditional expectation in the GNS representation.

Theorem 6 *Let φ and $\omega_1, \omega_2, \ldots, \omega_n$ be arbitrary normal positive functionals on the von Neumann algebra M. Then the equality*

$$S(\varphi, \omega)+\sum_{i=1}^{n} S(\omega, \omega_i)=\sum_{i=1}^{n} S(\varphi, \omega_i) \tag{7}$$

holds for $\omega=\sum_{i=1}^{n} \omega_i$.

Proof. We carry out the proof for $n=2$. Define $N=M \oplus M$ and set the functionals $\varphi_{12}, \omega_{12}, \psi_{12}$ by the following formulas:

$$\varphi_{12}(a \oplus b)=\tfrac{1}{2}(\varphi(a)+\varphi(b))$$

$$\omega_{12}(a \oplus b)=\tfrac{1}{2}(\omega_1(a)+\omega_2(b))$$

$$\psi_{12}(a \oplus b)=\tfrac{1}{4}(\omega_1(a)+\omega_1(b)+\omega_2(a)+\omega_2(b))$$

for $a, b \in M$. The mapping $E: a \oplus b \to \frac{1}{2}(a+b) \oplus \frac{1}{2}(a+b)$ is a φ_{12} preserving faithful normal conditional expectation onto the diagonal subalgebra $N_1=\{a \oplus a: a \in M\}$. Since $\psi_{12}=\omega_{12} \circ E$ Theorem 2 gives that

$$\frac{1}{2} S(\varphi, \omega_1)+\frac{1}{2} S(\varphi, \omega_2)=S\left(\varphi, \frac{\omega_1+\omega_2}{2}\right)+\frac{1}{2} S\left(\frac{\omega_1+\omega_2}{2}, \omega_1\right)+\frac{1}{2} S\left(\frac{\omega_1+\omega_2}{2}, \omega_2\right)$$

which is equivalent with (7). □

We note that (7) was obtained for injective von Neumann algebras in [8] and it was pointed out without proof also in [5].

Corollary 7 (7) *holds if φ and $\omega_1, \ldots, \omega_n$ are arbitrary positive functionals of a C*-algebra $\mathscr{A}$.*

Proof. We should investigate the existence of normal extensions of ω and ω_i to $\pi_\varphi(\mathscr{A})''$. If ω does not admit a normal extension then $S(\varphi, \omega)=+\infty$ and $\varphi(a_k) \to 0$, $\omega(a_k) \to \varepsilon>0$ for a net (a_k) of positive contractions (see Lemma 1). Then $\limsup \omega_i(a_k)>0$ at least for one index i and we have $S(\varphi, \omega_i)=+\infty$. Arguing in the same way we obtain that if one ω_i does not possess a normal extension to $\pi_\varphi(\mathscr{A})''$ then $S(\varphi, \omega)=+\infty$ and (7) holds with $+\infty$ on both sides.

Theorem 8 *Let $\mathscr{K}$ be a closed convex set in the state space of the C*-algebra $\mathscr{A}$ and φ is a state of $\mathscr{A}$. If*

$$d=\inf\{S(\varphi, \omega): \omega \in \mathscr{K}\}$$

is finite then there exists a unique $\omega \in \mathscr{K}$ such that $d=S(\varphi, \omega)$.

Proof. We choose a sequence $(\omega_n)\subset\mathscr{K}$ such that $\lim_{n\to\infty} S(\varphi,\omega_n)=d$. Corollary 7 says that

$$S(\varphi,\omega_n)+S(\varphi,\omega_m)=2S\left(\varphi,\frac{\omega_m+\omega_n}{2}\right)+S\left(\frac{\omega_m+\omega_n}{2},\omega_m\right)+S\left(\frac{\omega_m+\omega_n}{2},\omega_n\right).$$

Hence

$$S\left(\frac{\omega_m+\omega_n}{2},\omega_n\right)\leqq\varepsilon$$

if n and m are big enough. The inequality

$$\|\psi_1-\psi_2\|^2\leqq 2S(\psi_1,\psi_2)$$

(see [9], Theor. 3.1) supplies that (ω_n) must be a Cauchy sequence and it converges to a state $\omega\in\mathscr{K}$. The chain of inequalities

$$d\leqq S(\varphi,\omega)\leqq\operatorname{liminf} S(\varphi,\omega_n)\leqq d$$

makes the proof complete. □

We note that a less general version of Theorem 8 was obtained in [8] (for faithful states on injective von Neumann algebras) and it was connected to statistical inferences.

Theorem 9 *Let φ and ψ_i be normal states of a von Neumann algebra M $(i=1,2,\ldots,n)$. Then*

$$\sum_{i=1}^n\lambda_i S(\varphi,\psi_i)\geqq S\left(\varphi,\sum_{i=1}^n\lambda_i\psi_i\right)\geqq\sum_{i=1}^n\lambda_i S(\varphi,\psi_i)+\sum_{i=1}^n\lambda_i\log\lambda_i \tag{8}$$

provided that $\lambda_i>0$ and $\sum_{i=1}^n\lambda_i=1$.

Proof. The first inequality follows from the celebrated joint convexity of the relative entropy [4–6]. We concentrate on the second one. By Theorem 6 we have

$$S\left(\varphi,\sum_{i=1}^n\lambda_i\psi_i\right)=\sum_{i=1}^n\lambda_i S(\varphi,\psi_i)-\sum_{j=1}^n\lambda_j S\left(\sum_{j=1}^n\lambda_i\psi_i,\psi_j\right).$$

Since

$$S\left(\sum_{i=1}^n\lambda_i\psi_i,\psi_j\right)\leqq S(\lambda_j\psi_j,\psi_i)=-\log\lambda_j$$

we arrive at (8). □

Corollary 10 (8) *holds if φ and ψ_i's are arbitrary states of a C*-algebra.*

The proof is similar to that of Corollary 7.

It is worthwhile to note that (8) gives the strict convexity of the relative entropy functional in the second variable.

Given a separating state φ of the C^*-algebra $\mathscr{A}$ the cyclic and separating vector $\Phi \in H_\varphi$ (with respect to $\pi_\varphi(\mathscr{A})''$) determines the corresponding modular operator Δ. The formula

$$\Phi^h = \exp(\tfrac{1}{2}(\log \Delta + \pi_\varphi(h))\,\Phi$$

defines the perturbed vector for $h = h^* \in \mathscr{A}$ and

$$\varphi^h(a) = \langle \pi_\varphi(a)\,\Phi^h, \Phi^h \rangle \qquad (a \in \mathscr{A})$$

is the (unnormalized) perturbed functional (see [4]; p. 191 and 4.1 of [2]).

Theorem 11 *Let φ be a separating and ω be an arbitrary positive functional of a C^*-algebra $\mathscr{A}$. Then*

$$S(\varphi, \omega) = \sup\{\omega(h) - \log \varphi^h(I) : h = h^* \in \mathscr{A}\}. \tag{9}$$

Proof. First we note that $S(\varphi, \omega) \geqq \omega(h) - \log \varphi^h(I)$ always holds [16]. To prove (9) we go through several cases.

Case I: ω does not have a normal extension to $\pi_\varphi(\mathscr{A})''$.

We show that in this case there is a net $(h_i) \subset \mathscr{A}^{sa}$ such that $\omega(h_i) - \log \varphi^{h_i}(I) \to \infty$. Let (x_i) be the net guaranteed by Lemma 1. We set a mapping $\alpha_i : \mathbb{C}^2 \to \mathscr{A}$ as

$$\alpha_i(t, s) = t\,x_i + s(I - x_i).$$

We saw in the proof of (4) that

$$S(\varphi \circ \alpha_i, \omega \circ \alpha_i) \to \infty.$$

There exist $k_i \in \mathbb{C}^2$ such that

$$S(\varphi \circ \alpha_i, \omega \circ \alpha_i) = \omega \circ \alpha_i(k_i) - \log(\varphi \circ \alpha_i)^{k_i}(I).$$

According to Theorem 4 of [16] we have

$$\omega(\alpha_i(k_i)) - \log \varphi^{\alpha_i(k_i)}(I) \geqq S(\varphi \circ \alpha_i, \omega \circ \alpha_i)$$

which gives that $h_i = \alpha_i(k_i)$ has the required property.

Case II: ω has a normal extension to $\pi_\varphi(\mathscr{A})''$. (It is denoted also by ω.)

Case II/a: in addition, ω is faithful. We refer to [16] and get a sequence $k_n = k_n^* \in \pi_\varphi(\mathscr{A})''$ such that

$$S(\varphi, \omega) = \lim_{n \to \infty} \{\omega(k_n) - \log \varphi_{k_n}(I)\}.$$

Since $x \mapsto \omega(x) - \log \varphi^x(I)$ is a so-continuous function on the selfadjoint part of $\pi_\varphi(\mathscr{A})''$ (see [2]) we can choose k_n easily from $\pi_\varphi(\mathscr{A})$.

Case II/b: there is a constant λ such that $\omega \leqq \lambda \varphi$. Consider $\psi_\varepsilon = \varepsilon \cdot \varphi + \omega$ for $0 < \varepsilon < 1$. Then $\liminf_{\varepsilon \to +0} S(\varphi, \varphi_\varepsilon) \geqq S(\varphi, \omega)$. The functional ψ_ε is faithful and

$$\varphi \leqq \varepsilon^{-1} \psi_\varepsilon, \qquad \psi_\varepsilon \leqq (\varepsilon + \lambda)\,\varphi.$$

The relative Hamiltonian $h(\psi_\varepsilon, \varphi)$ exists and

$$\log \varepsilon \leqq h_\varepsilon = h(\psi_\varepsilon, \varphi) \leqq \log(\varepsilon + \lambda)$$

(see Prop. 4.1 of [2]). If $\varepsilon < \varepsilon_0$ then we have

$$S(\varphi, \omega) \leqq S(\varphi, \psi_\varepsilon) + \delta$$

and

$$\begin{aligned} S(\varphi, \psi_\varepsilon) &= \psi_\varepsilon(h_\varepsilon) - \log \varphi^{h_\varepsilon}(I) \\ &\leqq \omega(h_\varepsilon) - \log \varphi^{h_\varepsilon}(I) + \varepsilon \log(\varepsilon + \lambda) \\ &\leqq \omega(h_\varepsilon) - \log \varphi^{h_\varepsilon}(I) + \delta. \end{aligned}$$

We conclude that

$$S(\varphi, \omega) \leqq \omega(h_\varepsilon) - \log \varphi^{h_\varepsilon}(I) + 2\delta$$

and this completes the proof in the case $\omega \leqq \lambda \varphi$.

Now we turn to case II in general. By Lemma 7 of [16] we find a projection $p_n \in \pi_\varphi(\mathscr{A})''$ such that

$$\omega(p_n a p_n) \leqq 2^n \varphi(p_n a p_n) \quad \text{and} \quad \varphi(I - p_n) \leqq 2^{-n} \omega(I)$$

(here $a \in \pi_\varphi(\mathscr{A})''_+$). Let $\mathscr{M}_n$ be the von Neumann algebra $p_n \pi_\varphi(\mathscr{A})'' p_n + \mathbb{C}(I - p_n)$. It follows from Theorem 6 of [16] that

$$\lim_{n \to \infty} S(\varphi | \mathscr{M}_n, \omega | \mathscr{M}_n) = S(\varphi, \omega).$$

By invocation to the case IIb we find $h_{n,\varepsilon}$ in $\mathscr{A}^{sa}$ such that

$$S(\varphi | \mathscr{M}_n, \omega | \mathscr{M}_n) \leqq \omega(h_{n,\varepsilon}) - \log(\varphi | \mathscr{M}_n)^{h_{n,\varepsilon}}(I) + \varepsilon.$$

However,

$$\log(\varphi | \mathscr{M}_n)^{h_{n,\varepsilon}}(I) \geqq \log \varphi^{h_{n,\varepsilon}}(I)$$

and we can approximate $S(\varphi, \omega)$ by $\omega(h) - \log \varphi^h(I)$ arbitrarily well. □

Theorem 12 *Let $\mathscr{A}$ be a C*-algebra with a separating state φ and subalgebras $\mathscr{A}_{12}$ and $\mathscr{A}_{23}$. Assume that the φ-preserving conditional expectation E_{12} of $\mathscr{A}$ onto $\mathscr{A}_{12}$ exists. If $E_{12}(\mathscr{A}_{23}) = \mathscr{A}_2$ is a C*-subalgebra of $\mathscr{A}$ then for every state ω we have*

$$S(\varphi, \omega) + S(\varphi_2, \omega_2) \geqq S(\varphi_{12}, \omega_{12}) + S(\varphi_{23}, \omega_{23}) \tag{10}$$

where the subscripts denote the restriction of states to the corresponding subalgebra.

Proof. We may assume that $S(\varphi, \omega) < +\infty$. By the C*-algebraic version of Theorem 2 we have

$$S(\varphi, \omega) = S(\varphi_{12}, \omega_{12}) + S(\omega_{12} \circ E_{12}, \omega)$$

and

$$S(\varphi_2, \omega_2) + S(\omega_2 \circ F, \omega_{23}) = S(\varphi_{23}, \omega_{23})$$

where $F=E_{12}|\mathscr{A}_{23}$ is the conditional expectation of $\mathscr{A}_{23}$ onto $\mathscr{A}_2$ (which preserves φ_{23}). Since

$$S(\omega_{12}\circ E_{12},\omega)\geqq S(\omega_2\circ F,\omega_{23})$$

we obtain (10) by addition. □

Let us discuss some consequences of (10) and make some comments. (10) is a kind of strong superadditivity of the relative entropy and has been observed in the case $\mathscr{A}=\mathscr{A}_1\otimes\mathscr{A}_2\otimes\mathscr{A}_3$, $\mathscr{A}_{12}=\mathscr{A}_1\otimes\mathscr{A}_2\otimes\mathbb{C}$, $\mathscr{A}_{23}=\mathbb{C}\otimes\mathscr{A}_2\otimes\mathscr{A}_3$ and $\varphi=\varphi_1\otimes\varphi_2\otimes\varphi_3$ for von Neumann algebras on different levels of generality in [6, 12, 13] and [14].

If $\mathscr{A}_1$ and $\mathscr{A}_2$ are subalgebras of $\mathscr{A}$ and

$$\varphi(a_1\,a_2)=\varphi(a_1)\,\varphi(a_2)\qquad(a_i\in\mathscr{A}_i;\ i=1,2)$$

then we may say that $\mathscr{A}_1$ and $\mathscr{A}_2$ are independent for φ (see [10] for example). If $\mathscr{A}_1$ is also expected (by a φ-preserving conditional expectation E_1) then $E_1(a_2)=\varphi(a_2)\cdot I$ and we have

$$S(\varphi,\omega)\geqq S(\varphi_1,\omega_1)+S(\varphi_2,\omega_2)\tag{11}$$

for every state ω. While independency is related to the superadditivity (11), the strong superadditivity (10) is based on the Markov property.

Let us recall the notion of a stationary Markov chain with conditional expectations. See [1] and [11] for details of a general theory. Let $\mathscr{A}_0\subset\mathscr{A}$ be C^*-algebras, α an automorphism of $\mathscr{A}$ and φ a state on $\mathscr{A}$. If J is an interval in $\mathbb{Z}$ then we denote by $\mathscr{A}_J$ the C^*-subalgebra of $\mathscr{A}$ generated by $\cup\{\alpha^i(\mathscr{A}_0)\colon i\in J\}$. The quadruple $(\mathscr{A},\mathscr{A}_0,\alpha,\varphi)$ is called a stationary Markov chain with conditional expectations if

(i) $\varphi\circ\alpha=\varphi$.

(ii) The φ-preserving conditional expectation E_J of $\mathscr{A}$ onto $\mathscr{A}_J$ exists for every interval $J=[n,k]$.

(iii) $E_{[n,k]}(\mathscr{A}_{[k,\infty]})=\mathscr{A}_k$.

We define the mean relative entropy $S_M(\varphi,\omega)$ of a Markovian state φ and an arbitrary stationary state ω as

$$S_M(\varphi,\omega)=\limsup_{n\to\infty} n^{-1}\,S(\varphi|\mathscr{A}_{[0,n]},\omega|\mathscr{A}_{[0,n]}).$$

Theorem 13 *If $(\mathscr{A},\mathscr{A}_0,\alpha,\varphi)$ is a stationary Markov chain with conditional expectations, φ is a separating and ω is an arbitrary α-invariant state then*

$$S_M(\varphi,\omega)=\lim_{n\to\infty} n^{-1}\,S(\varphi|\mathscr{A}_{[0,n]},\omega|\mathscr{A}_{[0,n]}).$$

Moreover, $\omega\mapsto S_M(\varphi,\omega)$ is an affine weakly lower semicontinuous function on the α-invariant states.

Proof. We may assume that

$$S_n=S(\varphi|\mathscr{A}_{[0,n]},\omega|\mathscr{A}_{[0,n]})$$

is finite for every $n \in \mathbb{N}$. By Theorem 12 we have

$$S(\varphi|\mathcal{A}_{[0,n+k]}, \omega|\mathcal{A}_{[0,n+k]}) + S(\varphi|\mathcal{A}_n, \omega|\mathcal{A}_n)$$
$$\geqq S(\varphi|\mathcal{A}_{[0,n]}, \omega|\mathcal{A}_{[0,n]}) + S(\varphi|\mathcal{A}_{[n,n+k]}, \omega|\mathcal{A}_{[n,n+k]}).$$

Due to the α-invariance of φ and ω this reads as

$$S_{n+k} + S_0 \geqq S_n + S_k. \tag{12}$$

We need the following

Lemma 14 *If (S_n) is a real sequence with the property* (14) *then the limit of $n^{-1} S_n$ exists as $n \to \infty$ and equals to* $\sup\{n^{-1}(S_n - S_0): n \in \mathbb{N}\}$.

This lemma is standard and its proof is omitted. The affinity of $\omega \mapsto S_M(\varphi, \omega)$ is a consequence of Corollary 10. To show its lower semicontinuity we note that

$$S_n - S_0 = S(\omega \circ E|\mathcal{A}_{[0,n]}, \omega|\mathcal{A}_{[0,n]})$$

where E is the φ-preserving conditional expectation from $\mathcal{A}$ onto $\mathcal{A}_0$ (see Theorem 2), and we may refer to the weak lower semicontinuity of the relative entropy. □

The main result of [17] is based on some properties of the relative and mean relative entropy. Since all properties collected in the appendix of [17] and proved there for nuclear C^*-algebras are showed in the present paper without the assumption of nuclearity, this hypothesis may be omitted in [17].

References

1. Accardi, L., Frigerio, A., Lewis, J.T.: Quantum stochastic processes. Publ. RIMS, Kyoto Univ. **181**, 97–133 (1982)
2. Araki, H.: Relative Hamiltonian for faithful normal states of a von Neumann algebra. Publ. RIMS, Kyoto Univ. **9**, 165–209 (1973)
3. Araki, H.: On the uniqueness of KMS states of one-dimensional quantum lattice systems. Commun. Math. Phys. **44**, 1–7 (1975)
4. Araki, H.: Relative entropy of states of von Neumann algebras II. Publ. RIMS, Kyoto Univ. **13**, 173–192 (1977)
5. Araki, H.: Recent progress on entropy and relative entropy. In: Mebkhout, M., Senor, R. (eds.) Proceedings of the VIIIth Congress on Mathematical Physics (pp. 354–365). Singapore: World Sci. Publisher 1987
6. Bratteli, O., Robinson, D.W.: Operator algebras and quantum statistical mechanics II. Berlin Heidelberg New York: Springer 1981
7. Connes, A.: Sur la théorème de Radon-Nikodym pour les poids normaux fidèles sémifinis. Bull. Sci. Math., II. Ser. **97**, 253–258 (1973)
8. Donald, M.: Further results on the relative entropy. Math. Proc. Camb. Philos. Soc. **101**, 363–373 (1987)
9. Hiai, F., Ohya, M., Tsukada, M.: Sufficiency, KMS condition and relative entropy in von Neumann algebras. Pac. J. Math. **96**, 99–109 (1981)
10. Kosaki, H.: Relative entropy of states: a variational expression. J. Oper. Theory **16**, 335–348 (1986)
11. Kümmerer, B.: Survey on a theory of noncommutative stationary Markov processes. In: Accardi, L., von Waldenfels, W. (eds.) Quantum Probability and Applications III (Lect. Notes Math., vol. 1303, pp. 154–182). Berlin Heidelberg New York: Springer 1988

12. Lieb, E.H., Ruskai, M.B.: Proof of the strong subadditivity of quantum mechanical entropy. J. Math. Phys. **14**, 1938–1941 (1973)
13. Lieb, E.H.: Some convexity and subadditivity properties of entropy. Bull. Am. Math. Soc. **81**, 1–13 (1975)
14. Lindblad, G.: Expectations and entropy inequalities for finite quantum systems. Commun. Math. Phys. **39**, 111–119 (1974)
15. Petz, D.: Properties of the relative entropy of states of von Neumann algebras. Acta Math. Hung. **47**, 65–72 (1986)
16. Petz, D.: A variational expression for relative entropy. Commun. Math. Phys. **114**, 345–349 (1988)
17. Petz, D., Raggio, G.A., Verbeure, A.: Asymptotics of Varadhan-type and the Gibbs variational principle. Commun. Math. Phys. **121**, 271–282 (1989)
18. Strătilă, Ş.: Modular theory of operator algebras. Tunbridge Wells: Abacus Press 1981

Math. Z. 206, 363–407 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Action-angle coordinates at singularities for analytic integrable systems*

Hidekazu Ito

Mathematical Institute, Tôhoku University, Sendai 980, Japan

Received July 18, 1989; in final form April 3, 1990

1 Introduction

In this paper, we consider integrable Hamiltonian systems and study their singularities. Let (M, σ) be a smooth symplectic manifold of dimension $2n$, i.e., σ is a nondegenerate closed two form on M. A smooth function $H: M \to \mathbf{R}$ defines a Hamiltonian vector field X_H by the relation $\sigma(X_H, \cdot) = dH$. The Poisson bracket of two functions $G, H: M \to \mathbf{R}$ is defined by

$$\{G, H\} := \sigma(X_G, X_H) = X_H G$$

and G, H are said to be *Poisson commuting* or *in involution* if $\{G, H\} = 0$ identically. The identity $\{G, H\} \equiv 0$ implies that G is an integral of the vector field X_H.

An integrable system on M is defined by the mapping

$$F := (F_1, \ldots, F_n): M \to \mathbf{R}^n$$

with n functions $F_i: M \to \mathbf{R}$ satisfying the following conditions:

(i) $\{F_i, F_j\} \equiv 0$ for all $i, j = 1, \ldots, n$;
(ii) $dF_1, \ldots, dF_n$ are linearly independent on an open and dense subset of M.

We call the triple (M, σ, F) an *integrable system*. The functions $F_1, \ldots, F_n$ are said to be *functionally independent* if the above condition (ii) holds. Clearly each vector field X_F $(i = 1, \ldots, n)$ possesses n functionally independent and Poisson commuting integrals $F_1, \ldots, F_n$. In this sense, we also say that each vector field X_{F_i} is integrable.

Since $F_1, \ldots, F_n$ are integrals of the vector fields X_{F_i} $(i = 1, \ldots, n)$, the orbits of X_{F_i} passing through a point $p_0 \in M$ are confined to a connected component of the level set

$$F^{-1}(c) = \{p \in M \mid F(p) = c \in \mathbf{R}^n\},$$

where $c = F(p_0)$. Let $F^{(p_0)}$ denote the connected component containing p_0. Suppose that $F^{(p_0)}$ is compact and that $dF_1, \ldots, dF_n$ are linearly independent every-

* Dedicated to Professor Tosihusa Kimura on his 60th birthday

where on $F^{(p_0)}$. Then the manifold $F^{(p_0)}$ is an n-dimensional torus $T^n=\mathbf{R}^n/\mathbf{Z}^n$ and in its neighbourhood there exists a canonical coordinate system $(\theta, I), \theta\in T^n, I\in\mathbf{R}^n$ such that F_i are functions of n variables $I=(I_1, \ldots, I_n)$ only. This implies that orbits of X_{F_i} are periodic or quasi-periodic motions on n-tori defined by $I=\text{const}$. The coordinates $I=(I_1, \ldots, I_n)$ and $\theta=(\theta_1, \ldots, \theta_n)$ are called *action coordinates* and *angle coordinates* respectively. This is proved by V.I. Arnol'd [2] under some additional conditions, which was removed by Jost [8] (see also Duistermaat [4] or Markus and Meyer [9]).

In this theorem, it is a crucial assumption that the set $F^{(p_0)}$ does not contain any critical point of F. What happens at singularities of F? In this paper, we are concerned with the singularities of F and our purpose is to introduce a canonical coordinate system generalizing the above action-angle coordinates to singular cases. It will naturally lead to a normal form of F at singularities. The results will be stated in the next section and the main results are Theorems 2.6 and 2.8. The later sections are devoted to their proofs and their organization will be described in detail at the end of the next section.

2 Statement of the main results

Let (M, σ, F) be an integrable system. Throughout this paper from now on, we assume that both the symplectic manifold (M, σ) and the mapping F are real analytic. In this case, we call the triple (M, σ, F) a *real analytic integrable system*. To state the results, we need some preliminary considerations.

(a) Definition of an orbit of rank k

Let ϕ_i^t be the flow of the Hamiltonian vector field $X_{F_i}(i=1, \ldots, n)$. For a point $p_0\in M$, we set

$$\Phi(p_0):=\{\phi_1^{t_1}\circ\phi_2^{t_2}\circ\ldots\circ\phi_n^{t_n}(p_0)\in M\,|\,(t_1, \ldots, t_n)\in\mathbf{R}^n\}.$$

We call $\Phi(p_0)$ the *orbit for the integrable system* passing through p_0 if n vector fields $X_{F_i}(i=1, \ldots, n)$ are complete on $\Phi(p_0)$, that is, the point $\phi_1^{t_1}\circ\phi_2^{t_2}\circ\ldots\circ\phi_n^{t_n}(p_0)$ is well-defined for all $(t_1, \ldots, t_n)\in\mathbf{R}^n$. This is the case if, for instance, the set $F^{(p_0)}$ is compact. Also, the set $\Phi(p_0)$ is called the orbit of n vector fields $X_{F_1}, \ldots, X_{F_n}$ passing through p_0 if these vector fields are complete on $\Phi(p_0)$. In this sense, we can define an orbit of k vector fields with any positive integer k.

We are interested in the case where $F^{(p_0)}$ admits singularities of F. Suppose that $\operatorname{rank}(dF_1, \ldots, dF_n)=k$ at the point p_0. Then one can choose a set of k functions whose derivatives are linearly independent at p_0. Let us assume that $F_1, \ldots, F_k$ are such functions. Then we have

$$\text{(2.1)}\quad \operatorname{rank}(dF_1, \ldots, dF_k)=k \quad\text{and}\quad dF_{k+i}=\sum_{j=1}^{k} c_{ij}\,dF_j \qquad (i=1, \ldots, n-k)$$

at p_0 with some constants $c_{ij}\in\mathbf{R}$. Actually one can prove that condition (2.1) holds everywhere on the orbit $\Phi(p_0)$ with the fixed constants $c_{ij}\in\mathbf{R}$. We call $\Phi(p_0)$ an *orbit of rank k* if $\operatorname{rank}(dF_1, \ldots, dF_n)=k$ on it.

We formulate this observation more generally. Associated with the orbit $\Phi(p_0)$ for the integrable system (M, σ, F), we define a Lie algebra $\mathfrak{F}$ as follows:

$$\mathfrak{F} := \{G \in C^\omega(\Phi(p_0)) \mid \{F_i, G\} \equiv 0, i = 1, \ldots, n\}.$$

Here $C^\omega(\Phi(p_0))$ denotes the set of all functions which are real analytic in some neighbourhood of $\Phi(p_0)$. Then we can prove

Proposition 2.1 (i) *Any two functions $G, G' \in \mathfrak{F}$ are in involution, i.e., $\{G, G'\} \equiv 0$.*

(ii) *Let $\Phi(p_0)$ be the orbit of rank k for the integrable system passing through a point p_0. Let $H_1, \ldots, H_k \in \mathfrak{F}$ be k functions such that*

$$\text{(2.2)} \qquad \operatorname{rank}(dH_1, \ldots, dH_k) = k \quad \text{and} \quad dF_i = \sum_{j=1}^{k} c_{ij}\, dH_j \qquad (i = 1, \ldots, n)$$

at p_0 with some constants $c_{ij} \in \mathbf{R}$. Then condition (2.2) holds everywhere on the orbit $\Phi(p_0)$ with the fixed constants $c_{ij} \in \mathbf{R}$ and $\Phi(p_0)$ is an orbit of k vector fields $X_{H_1}, \ldots, X_{H_k}$.

The proof of this proposition will be given in the next section. In the following, let ϕ_i^t denote the flow of the vector field X_{H_i}. Since ϕ_i^t defines a diffeomorphism on $\Phi(p_0)$ for all $t \in \mathbf{R}$, we can define a transitive action of $\mathbf{R}^k$ on $\Phi(p_0)$ by

$$\text{(2.3)} \qquad (t, p) \mapsto \phi^t(p) := \phi_1^{t_1} \circ \phi_2^{t_2} \circ \ldots \circ \phi_k^{t_k}(p);$$
$$t = (t_1, \ldots, t_k) \in \mathbf{R}^k, \qquad p \in \Phi(p_0).$$

Let

$$P_H(p_0) = \{t \in \mathbf{R}^k \mid \phi^t(p_0) = p_0\}$$

be the isotropy group of p_0 under this action. An element $T \in P_H(p_0)$ is called a *period* of ϕ^t. One can easily see that the group $P_H(p_0)$ is a $d(\leqq k)$-dimensional lattice (discrete subgroup of $\mathbf{R}^k$) and that the action (2.3) gives rise to a diffeomorphism $\mathbf{R}^k/P_H(p_0) \cong \Phi(p_0)$. Therefore $\Phi(p_0) \cong T^d \times \mathbf{R}^{k-d}$, where $T^d = \mathbf{R}^d/\mathbf{Z}^d$ is a d-dimensional torus (see Arnol'd [2] for details). Note that in the regular case (i.e., $\operatorname{rank}(dF_1, \ldots, dF_n) = n$) the component $F^{(p_0)}$ of the level set is the orbit $\Phi(p_0)$ itself [2]. If $\Phi(p_0)$ is compact, $\Phi(p_0) \cong T^k$. Our purpose is to introduce a special coordinate system in which the vector fields X_{F_i} can be solved explicitly. We will achieve this by imposing appropriate conditions from the viewpoint of the theory of dynamical systems.

(b) Parametrized reduction

The conditions to be imposed on the orbit $\Phi(p_0)$ are introduced via the reduction of Hamiltonian systems which will be described below (see [1]).

For the orbit $\Phi(p_0)$ of rank k passing through $p_0 \in M$, let $H_1, \ldots, H_k \in \mathfrak{F}$ be k functions satisfying condition (2.2). Assume that $dH_1, \ldots, dH_k$ are linearly independent in a neighbourhood U of p_0. We set $H = (H_1, \ldots, H_k)$ and $H(p_0) = h_0 \in \mathbf{R}^k$. For any $h \in \mathbf{R}^k$ in a small neighbourhood of h_0, we consider a level

set $H^{-1}(h) := \{p \in M \mid H(p) = h\}$. The flows of $X_{H_1}, \ldots, X_{H_k}$ give rise to a local action of $\mathbf{R}^k$ on $H^{-1}(h) \cap U$ and the quotient space

$$\Sigma_h := H^{-1}(h) \cap U / \mathbf{R}^k$$

is obtained by identifying the orbit of this action. It is a symplectic manifold of dimension $2m$ $(m = n - k)$ with the symplectic structure reduced from that of $U \subset M$. The manifold Σ_h depends on the parameter h analytically. Since we consider the reduction locally, Σ_h is a submanifold of U which is a local transversal section to the k vector fields X_{H_i}. We denote the projection map by π_h, i.e.,

$$\pi_h : H^{-1}(h) \cap U \to \Sigma_h.$$

We say that the k functions $H_1, \ldots, H_k$ define the (parametrized) reduction $\pi_h : H^{-1}(h) \cap U \to \Sigma_h$.

Associated with this reduction, we can reduce functions and mappings on U to those on Σ_h as follows: Let $G \in \mathfrak{F}$ and assume that it is real analytic on U. It is invariant under the flows of $X_{H_1}, \ldots, X_{H_k}$ and therefore induces a function $G_h : \Sigma_h \to \mathbf{R}$ such that $G = G_h \circ \pi_h$ on $H^{-1}(h) \cap U$. If the function G satisfies $dG = \sum_{j=1}^{k} c_j dH_j$ with $c_j \in \mathbf{R}$ at $p \in H^{-1}(h) \cap U$, then $\pi_h(p)$ is a critical point of G_h. It is an equilibrium point of the reduced vector field X_{G_h} on Σ_h and is called a *relative equilibrium.*

On the other hand, assume that for a function $G \in \mathfrak{F}$ the mapping $\exp X_G$ is well defined on U and that there exists a neighbourhood U' of p_0 such that $(\exp X_G)(U') \subset U$. Here and in what follows, we use the notation $\exp(tX_G)$ to denote the flow of the vector field X_G and hence $\exp X_G$ is obtained by setting $t = 1$ (it is called the *time 1 map*). For example, if the functions $H_1, \ldots, H_k$ satisfy the relations (2.2) at p_0, then for any period $T = (t_1, \ldots, t_k) \in P_H(p_0)$ the map ϕ^T defines a local symplectic diffeomorphism with the fixed point p_0. Since the vector fields $X_{H_1}, \ldots, X_{H_k}$ commute with one another in a neighbourhood of $\Phi(p_0)$, the map ϕ^T is equal to $\exp X_G$ with $G = \sum_{j=1}^{k} t_j H_j$. Let

$$\pi'_h : H^{-1}(h) \cap U' \to \Sigma'_h := H^{-1}(h) \cap U' / \mathbf{R}^k.$$

Since the functions $H_1, \ldots, H_k$ are invariant under the mapping $\exp X_G$, we can define a mapping $f_h : \Sigma'_h \to \Sigma_h$ by

$$f_h := \pi_h \circ \exp X_G \circ (\pi'_h)^{-1} : \Sigma'_h \to \Sigma_h. \tag{2.4}$$

One can see that f_h is a symplectic mapping depending on h analytically (see Lemma 3.1).

In what follows, for a vector field X we denote by DX the linearized vector field of X (at an equilibrium point), and for a mapping f we denote by df the linearized mapping of f (at a fixed point). For these two objects, we review the following

Definition 2.2 (i) Let $\pi_h(p)$ be an equilibrium point of X_{G_h} and let $\pm\lambda_1, \ldots, \pm\lambda_m$ be the eigenvalues of the linearized vector field DX_{G_h} at $\pi_h(p)$. (The eigenvalues

always occur in pairs in this way [2].) The equilibrium point $\pi_h(p)$ is said to be

(1) *nondegenerate* if $\lambda_j \neq 0$ for all $j=1, \ldots, m$;
(2) *elliptic* if $\pm\lambda_1, \ldots, \pm\lambda_m$ are all distinct and purely imaginary;
(3) *non-resonant* if the following condition holds:

$$\sum_{k=1}^{m} j_k \lambda_k \neq 0 \quad \text{for any nonzero integer vector } (j_1, \ldots, j_m) \in \mathbf{Z}^m \setminus \{0\}. \tag{2.5}$$

(ii) Let $\pi_h(p)$ be a fixed point of $f_h = \pi_h \circ \exp X_G \circ (\pi_h')^{-1}$ and let $\lambda_j, \lambda_j^{-1}$ $(j=1, \ldots, m)$ be the eigenvalues of the linearized mapping df_h at $\pi_h(p)$. (The eigenvalues always occur in pairs in this way [2].) The fixed point $\pi_h(p)$ is said to be

(1) *nondegenerate* if $\lambda_j \neq 1$ for all $j=1, \ldots, m$;
(2) *elliptic* if $\lambda_j, \lambda_j^{-1}$ $(j=1, \ldots, m)$ are all distinct and of modulus 1, i.e., $|\lambda_j|=1$;
(3) *non-resonant* if the following condition holds:

$$\prod_{k=1}^{m} \lambda_k^{j_k} \neq 1 \quad \text{for any nonzero integer vector } (j_1, \ldots, j_m) \in \mathbf{Z}^m \setminus \{0\}. \tag{2.6}$$

In the above, we require the simplicity of eigenvalues for the definition of an elliptic equilibrium point or an elliptic fixed point. This is an additional requirement to the usual definition of ellipticity.

Let $\Phi(p)$ be the orbit for the integrable system passing through $p \in M$ and let p' be an arbitrary point of $\Phi(p)$. Then the reduction is well defined also in a neighbourhood of p' and $\pi_h(p')$ is an equilibrium point of X_{G_h} or a fixed point of f_h. One can easily see that the eigenvalues in Definition 2.2 (i) and (ii) do not change when $\pi_h(p)$ is replaced by $\pi_h(p')$.

(c) Statement of the results

We consider the reduction $\pi_h : H^{-1}(h) \cap U \to \Sigma_h$ defined in the previous subsection (b). In order to state our result, we introduce the following definition.

Definition 2.3 Let $\Phi(p)$ be the orbit of rank k for the integrable system (M, σ, F) passing through a point $p \in H^{-1}(h) \cap U$. It is said to be *nondegenerate* if there exists a function $G \in \mathfrak{F}$ such that either of the following two conditions is satisfied:

(i) $\pi_h(p)$ is a nondegenerate equilibrium point of the reduced vector field X_{G_h}.
(ii) $\pi_h(p)$ is a nondegenerate fixed point of the mapping $f_h = \pi_h \circ \exp X_G \circ (\pi_h')^{-1}$.

The following remarks will be proved in Section 3.

Remark 2.4 Definition 2.3 does not depend on the choice of the k functions $H_1, \ldots, H_k \in \mathfrak{F}$. Namely, let $\hat{H}_1, \ldots, \hat{H}_k \in \mathfrak{F}$ be another set of functions such that

$$\operatorname{span}(d\hat{H}_1, \ldots, d\hat{H}_k) = \operatorname{span}(dH_1, \ldots, dH_k) \quad \text{at } p$$

and let us consider the reduction defined by the k functions $\hat{H}_1, \ldots, \hat{H}_k$. We denote the reduction by $\hat{\pi}_c: \hat{H}^{-1}(c) \cap U \to \hat{\Sigma}_c := \hat{H}^{-1}(c) \cap U/\mathbf{R}^k$, where $\hat{H} = (\hat{H}_1, \ldots, \hat{H}_k)$ and $\hat{H}(p) = c$. If $\pi_h(p)$ is an equilibrium point of X_{G_h}, then $\hat{\pi}_c(p)$ is an equilibrium point of X_{G_c} and the eigenvalues of the linearized vector field DX_{G_c} at $\hat{\pi}_c(p)$ are the same as those of DX_{G_h} at $\pi_h(p)$. Similarly, if $\pi_h(p)$ is a fixed point of f_h, then $\hat{\pi}_c(p)$ is a fixed point of $\hat{f}_c := \hat{\pi}_c \circ \exp X_G \circ (\hat{\pi}_c')^{-1}$ and the eigenvalues of the linearized mapping $d\hat{f}_c$ at $\hat{\pi}_c(p)$ are the same as those of df_h at $\pi_h(p)$.

Remark 2.5 In Definition 2.3, condition (i) implies condition (ii) in the following sense: Let $G \in \mathfrak{F}$ be a function such that $\pi_h(p)$ is a nondegenerate (or elliptic, nonresonant, resp.) equilibrium point of X_{G_h}. Then one can construct a function $\tilde{G} \in \mathfrak{F}$ such that the time 1 map $\exp X_{\tilde{G}}$ is well-defined and $\pi_h(p)$ is a nondegenerate (or elliptic, non-resonant, resp.) fixed point of $f_h := \pi_h \circ \exp X_{\tilde{G}} \circ (\pi_h')^{-1}$.

In spite of Remark 2.5, we presented the two conditions (i) and (ii) in Definition 2.3 and we will present the similar two conditions also in Definition 2.7 below since condition (i) is natural and easy to be checked in application. Moreover, the function $\tilde{G} \in \mathfrak{F}$ to be constructed in the proof of Remark 2.5 has a special nature that p is an equilibrium point of $X_{\tilde{G}}$ as well as a fixed point of $\exp X_{\tilde{G}}$ (see Sect. 3). It would be more natural to consider, for instance, a linear combination of $H_1, \ldots, H_k$ as a candidate for the function G satisfying condition (ii).

Our first result is concerned with the existence of a family of k-dimensional invariant tori.

Theorem 2.6 *Let $\Phi(p_0)$ be the compact orbit of rank k passing through a point p_0 for a real analytic integrable system (M, σ, F). Then the following holds:*

(i) *$\Phi(p_0)$ is diffeomorphic to a k-dimensional torus $T^k = \mathbf{R}^k/\mathbf{Z}^k$.*

(ii) *Let $H_1, \ldots, H_k \in \mathfrak{F}$ be k functions satisfying condition (2.2) at p_0, and let $H = (H_1, \ldots, H_k)$ and $H(p_0) = h_0$. Assume that $\Phi(p_0)$ is nondegenerate. Then there exists a unique real analytic family $\{\Phi_h\}$, depending on the parameter $h \in \mathbf{R}^k$ close to h_0, of compact orbits Φ_h of rank k for the integrable system such that (a) $\Phi_h \subset H^{-1}(h)$ and $\Phi_{h_0} = \Phi(p_0)$, and (b) any function $G \in \mathfrak{F}$ satisfies the relation*

$$dG = \sum_{i=1}^{k} c_i(h)\, dH_i \tag{2.7}$$

along the orbit Φ_h, where $c_i(h)$ are real analytic functions of h which are constant along Φ_h. This family $\{\Phi_h\}$ forms a $2k$-dimensional symplectic submanifold of M.

In the above, the real analytic family $\{\Phi_h\}$ means a family of orbits given in the form

$$\Phi_h := \Phi(\mathfrak{z}_h) \qquad (h \in V),$$

where V is a neighbourhood of h_0 in $\mathbf{R}^k$ and $\mathfrak{z}_h$ are points on $H^{-1}(h) \cap U$ such that not only $\mathfrak{z}_h$ but also every generator of the period lattice $P_h(\mathfrak{z}_h)$ depends on h analytically and $\mathfrak{z}_h$ satisfies $\mathfrak{z}_{h_0} = p_0$. Here $H = (H_1, \ldots, H_k)$ and U is the neighbourhood of p_0 in which the parametrized reduction is defined. We denote

the family of orbits by $\{\Phi_h\}_V$ to indicate the neighbourhood V in which the parameter h runs over.

In Theorem 2.6, the first assertion (i) is already proved and it follows that each orbit Φ_h in (ii) is a k dimensional torus. We also call $\{\Phi_h\}_V$ a *real analytic family of k-dimensional invariant tori for the integrable system.* We note that the family $\{\Phi_h\}_V$ does not depend on the choice of the functions $H_1, \ldots, H_k$. In fact, the condition that $\Phi(p_0)$ be nondegenerate does not depend on it. If $\hat{H}_1, \ldots, \hat{H}_k$ are another set of functions satisfying (2.2), then the corresponding assertion (ii) of Theorem 2.6 gives a different parametrization of the same family of k-dimensional invariant tori (see Remark 4.2).

For this family of k-dimensional tori, we introduce the following definition.

Definition 2.7 The family $\{\Phi_h\}_V$ is said to be *non-resonant (and elliptic)* if there exists a function $G \in \mathfrak{F}$ such that either of the following two conditions is satisfied:

(i) $\pi_h(\Phi_h \cap U)$ are equilibrium points of the vector fields X_{G_h} such that they are non-resonant for some $h \in V$ (and are elliptic for all $h \in V$).

(ii) $\pi_h(\Phi_h \cap U)$ are fixed points of the mappings $f_h = \pi_h \circ \exp X_G \circ (\pi'_h)^{-1}$ such that they are non-resonant for some $h \in V$ (and are elliptic for all $h \in V$).

By Remarks 2.4 and 2.5, Definition 2.7 is also independent of the choice of the k functions $H_1, \ldots, H_k$, and moreover condition (i) implies condition (ii) in the sense of Remark 2.5. One can easily see that if the equilibrium points (or fixed points) $\pi_h(\Phi_h \cap U)$ are non-resonant for some $h \in V$, then so they are generically for $h \in V$, namely, they are non-resonant for any h belonging to a residual subset of V (see Lemma 5.2).

Our main result is stated as follows:

Theorem 2.8 *Let (M, σ, F) be a real analytic integrable system and $\{\Phi_h\}_V$ a real analytic family of k-dimensional invariant tori for the integrable system as given in Theorem 2.6. Assume that this family is non-resonant and elliptic. Then there exists an open and dense subset $\hat{V}$ of V with the following property: For any $h \in \hat{V}$, there exist a neighbourhood U_h of Φ_h and a real analytic symplectic diffeomorphism ψ_h satisfying the following conditions:*

$$\text{(i)}\quad \psi_h = \psi_h(\theta, I, \zeta)\colon\ T^k \times D_1 \times D_2 \to U_h; \qquad \psi_h(T^k \times \{0\} \times \{0\}) = \Phi_h,$$

where T^k is a k-dimensional torus with coordinate $\theta = (\theta_1, \ldots, \theta_k)$, D_1 is a domain of $\mathbf{R}^k$ containing the origin with coordinate $I = (I_1, \ldots, I_k)$ and D_2 is a domain of $\mathbf{R}^{2m}$ containing the origin with coordinate $\zeta = (\xi, \eta) = (\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_m)$;

$$\text{(ii)}\qquad \psi_h^* \sigma = \sum_{j=1}^{k} d\theta_j \wedge dI_j + \sum_{j=1}^{m} d\xi_j \wedge d\eta_j \qquad (k+m=n);$$

(iii) *for any function $G \in \mathfrak{F}$, $G \circ \psi_h$ is a real analytic function of n variables $I_1, \ldots, I_k, \tau_1, \ldots, \tau_m$ with $\tau_j = (\xi_j^2 + \eta_j^2)/2$ in a neighbourhood of the origin.*

In this coordinate system (θ, I, ξ, η), the vector field X_G with $G \in \mathfrak{F}$ is expressed as

$$\dot{\theta}_j = \frac{\partial G}{\partial I_j}, \qquad \dot{I}_j = 0 \qquad (j = 1, \ldots, k),$$

$$\dot{\xi}_j = \frac{\partial G}{\partial \tau_j}\eta_j, \qquad \dot{\eta}_j = -\frac{\partial G}{\partial \tau_j}\xi_j \qquad (j = 1, \ldots, m).$$

Hence I_j and τ_j yield n integrals in involution and this system is solved explicitly as periodic or quasi-periodic motions on the tori $I_j=\text{const.}$, $\tau_l=\text{const.}\geqq 0$ ($j=1, \ldots, k; l=1, \ldots, m$) of dimension $\geqq k$.

In Theorem 2.8, the assumption that the family $\{\Phi_h\}_V$ be non-resonant and elliptic is empty in the case $k=n$, but the assertions (i)–(iii) are valid also for this case. It corresponds to the theorem of Arnol'd and Jost. On the other hand, if $k=0$, it reduces to our previous result [7, Theor. 1.4] asserting the convergence of the Birkhoff normal form near a non-resonant equilibrium point.

In Eliasson [5], the normal form given in the above condition (iii) is obtained near the invariant k-tori for the integrable system (M, σ, F). Having been motivated by the previous work by Vey [15], he established the above normal form under the assumption that the quadratic terms of the Taylor expansions of $F_j|\Sigma_{h_0}$ ($j=1, \ldots, n$) at $\pi_{h_0}(p_0)$ satisfy a nondegeneracy condition (see (5.4) in Sect. 5). In this case, condition (i) of Definition 2.7 is satisfied as we shall see in Section 5. On the other hand, condition (ii) of Definition 2.7 is a new condition that is not included in Eliasson's theorem. In this sense, Theorem 2.8 is a generalization of Eliasson's theorem as far as the system is real analytic. It should be mentioned that Eliasson's theorem holds true also for smooth case.

The so-called KAM theory (see [3, 10, 13]) is concerned with persistence of invariant n-tori under perturbation. Recently it has been generalized for that of invariant k-tori ($k<n$) such as given in Theorem 2.8 (see [6, 12]).

(d) Strategy of the proof

As we stated in the definition of the family $\{\Phi_h\}_V$ to be non-resonant and elliptic, $\{\pi_h(\Phi_h \cap U)\}$ is a family of equilibrium points of the vector fields X_{G_h} or a family of fixed points of the mappings f_h. Here G_h and f_h are parametrized by $h \in V$ and we deal with the Birkhoff normalization of these functions and mappings in the proof of Theorem 2.8. It is the crucial part of the proof. Here the Birkhoff normalization consists of finding a local canonical coordinate system (ξ, η) in which G_h is a function of m variables $\tau_1, \ldots, \tau_m$ with the parameter h and in which f_h is the time 1 map of the vector field with the Hamiltonian of this form. The function and the mapping of this form are called the *Birkhoff normal forms* with the parameter h. Notice that the equilibrium points (or the fixed points) $\pi_h(\Phi_h \cap U)$ are non-resonant for h belonging to a dense subset of V. However the set of h for which $\pi_h(\Phi_h \cap U)$ are resonant (i.e., condition (2.5) or (2.6) is violated) is also dense in V. Nevertheless we can carry out the Birkhoff normalization for any h belonging to an *open* and dense subset $\hat{V}$ of V. We owe it to the existence of m additional integrals for the vector field X_{G_h} or for the mapping f_h. We shall observe that, roughly speaking, additional integrals of X_{G_h} (or f_h) become functionally dependent for the parameter h belonging to the complement of $\hat{V}$ in V. It is to be noted that, without the assumption of integrability, the Birkhoff normalization is not be possible in general even formally for the resonant case.

To prove the convergence of this Birkhoff normalization, we encounter the so called small divisor difficulty. However, we can avoid it by using the existence of the additional integrals and can prove the convergence by using the rapidly convergent iteration method [10] without any Diophantine condition. We owe

the idea of its proof to [7] and the proof will be done by generalizing the argument of [7].

This paper is organized as follows: In Section 3, we introduce a local coordinate system in which our basic consideration will be done. Also we prove some preliminary facts including Proposition 2.1 and Remarks 2.4 and 2.5. In Section 4, we shall prove Theorem 2.6. Its proof also plays the role of introducing a coordinate system for which we can define the Birkhoff normalization. The aim of Section 5 and later ones is to prove Theorem 2.8. Sections 5 through 7 will be concerned with the Birkhoff normalization. In Section 5, we shall state the results asserting the convergence of the Birkhoff transformation in two theorems with corollaries. They will be formulated independently of the integrable system (M, σ, F). The first theorem and its corollary (Theor. 5.4 and Coroll. 5.5) are concerned with normal forms near an equilibrium point. The second ones (Theor. 5.6 and Coroll. 5.7) are concerned with normal forms near a fixed point. We shall state the results on complex normal forms in the theorems and those on real normal forms in the corollaries. Furthermore, we shall discuss the connection between the theorem of Eliasson and ours at the end of Section 5. Section 6 contains the proof of Theorem 5.4 and Corollary 5.5. We shall describe there how to avoid the small divisor difficulty associated with the Birkhoff normalization procedure. It constitutes the main part of the proof of Theorem 2.8. Section 7 is devoted to the proof of Theorem 5.6 and Corollary 5.7. In the final section (Section 8), we shall complete the proof of Theorem 2.8 by globalizing the Birkhoff transformation to a neighbourhood of the orbit Φ_h.

3 Preliminaries

In this section, we introduce a local coordinate system whose choice will be basic for our discussion. We will use it to prove Proposition 2.1, Remarks 2.4 and 2.5, and other preliminary facts stated in the previous section.

Let $\Phi(p_0)$ be the orbit of rank k passing through $p_0 \in M$ for the real analytic integrable system (M, σ, F). Suppose that condition (2.2) holds at a point $p_0 \in M$ for a set of k functions $H_1, \ldots, H_k \in \mathfrak{F}$. Since $dH_1, \ldots, dH_k$ are linearly independent in a neighbourhood of p_0, it is well known [1, 3] that there exists a real analytic symplectic diffeomorphism $\psi = \psi(u, v, x, y)$, $u = (u_1, \ldots, u_k)$, $v = (v_1, \ldots, v_k) \in \mathbf{R}^k$ and $x = (x_1, \ldots, x_m)$, $y = (y_1, \ldots, y_m) \in \mathbf{R}^m$ such that

$$\psi : (\mathbf{R}^{2n}, w_0) \to (M, p_0); \qquad w_0 = (0, h_0, 0, 0), \qquad h_0 = H(p_0) \tag{3.1}$$

and

$$H_i \circ \psi = v_i \qquad (i = 1, \ldots, k), \tag{3.2}$$

$$\psi^* \sigma = \sum_{j=1}^{k} du_j \wedge dv_j + \sum_{j=1}^{m} dx_j \wedge dy_j, \qquad (k + m = n). \tag{3.3}$$

In this local coordinate system (u, v, x, y), the Poisson bracket of G and $G' \in \mathfrak{F}$ is expressed as

$$\{G, G'\} = \sum_{j=1}^{k} \left(\frac{\partial G}{\partial u_j} \frac{\partial G'}{\partial v_j} - \frac{\partial G}{\partial v_j} \frac{\partial G'}{\partial u_j} \right) + \sum_{j=1}^{m} \left(\frac{\partial G}{\partial x_j} \frac{\partial G'}{\partial y_j} - \frac{\partial G}{\partial y_j} \frac{\partial G'}{\partial x_j} \right).$$

For any function $G \in \mathfrak{F}$, $G \circ \psi$ is independent of $u=(u_1, \ldots, u_k)$ since we have $\{H_i, G\} \equiv 0$ for all $i=1, \ldots, k$. In what follows, we write $G(=G(v, x, y))$ in place of $G \circ \psi$, which is defined in a neighbourhood of $w_0 \in \mathbf{R}^{2n}$. The reduced phase space Σ_h can be identified with

$$\Sigma_h = \{(u, v, x, y) \in U \mid u=0, v=h\},$$

where U is a neighbourhood of $(0, h_0, 0, 0) \in \mathbf{R}^{2n}$. This is a symplectic submanifold with the symplectic structure $\sum_{j=1}^{m} dx_j \wedge dy_j$ and is transversal to the flow ϕ^t of $X_{H_1}, \ldots, X_{H_k}$ on the level set $H^{-1}(h)$ for any h close to h_0.

By using this coordinate system, we will prove Proposition 2.1.

Proof of Proposition 2.1 (i) Let U be a neighbourhood of $p_0 \in M$. Since the 1-forms $dF_1, \ldots, dF_n$ are linearly independent on an open and dense subset of M, there exists a local coordinate system $(u, v)=(u_1, \ldots, u_n, v_1, \ldots, v_n)$ in some open subset of U such that conditions (3.1)–(3.3) hold with $H_i=F_i$ and $k=n$. This implies that any function $G \in \mathfrak{F}$ depends on n variables $v_1, \ldots, v_n$ only. Therefore we have $\{G, G'\}=0$ in the open subset of U for any G, $G' \in \mathfrak{F}$. By the analyticity of functions, we have the identity $\{G, G'\} \equiv 0$.

(ii) The relations (3.2) and (3.3) imply that X_{H_i} is equal to $\partial/\partial u_i$ and therefore the coordinates v, x, y are invariant under the flows of $X_{H_1}, \ldots, X_{H_k}$. This implies that along the orbit of k vector fields $X_{H_1}, \ldots, X_{H_k}$ near the point p_0, condition (2.2) holds with the fixed constants c_{ij}. The same discussion is valid along the whole orbit of $X_{H_1}, \ldots, X_{H_k}$ by using the above local coordinate system successively. Hence condition (2.2) holds everywhere on the orbit of $X_{H_1}, \ldots, X_{H_k}$ with the fixed constants $c_{ij} \in \mathbf{R}$. Since $\Phi(p_0)$ is an orbit of rank k, we have $\operatorname{rank}(dF_1, \ldots, dF_n)=k$ on this orbit. Consequently there exist k functions among $F_1, \ldots, F_n$ whose derivatives are linearly independent on $\Phi(p_0)$. If we assume that, for instance, $F_1, \ldots, F_k$ are such functions, then the orbit of $X_{H_1}, \ldots, X_{H_k}$ is also the orbit of $X_{F_1}, \ldots, X_{F_k}$, which is $\Phi(p_0)$ itself. This completes the proof. □

Next we will prove that the mapping f_h introduced by (2.4) is symplectic. In the coordinate system (u, v, x, y), the mapping $\phi := \exp X_G$ has the form

$$(u', v', x', y') = \phi(u, v, x, y); \quad u, v, u', v' \in \mathbf{R}^k, \quad x, y, x', y' \in \mathbf{R}^m \tag{3.4}$$

satisfying $v'=v$ and

$$\sum_{j=1}^{k} du'_j \wedge dv'_j + \sum_{j=1}^{m} dx'_j \wedge dy'_j = \sum_{j=1}^{k} du_j \wedge dv_j + \sum_{j=1}^{m} dx_j \wedge dy_j. \tag{3.5}$$

Here the condition $v'=v$ follows from the fact that $H_1, \ldots, H_k$ be integrals of the flow $\exp(tX_G)$. The mapping $f_h = \pi_h \circ \exp X_G \circ (\pi'_h)^{-1}$ is defined by the correspondence $(x, y) \mapsto (x', y')$ under the restriction $v=h$. We will consider h as a

parameter running over a neighbourhood of $h_0 \in \mathbf{R}^k$. The following lemma shows that f_h is a symplectic mapping depending on the parameter h. The converse assertion in the lemma will be for the later use.

Lemma 3.1 *Let A_1 be an invertible constant matrix of degree k and let ϕ be a local symplectic diffeomorphism* (3.4) *satisfying $v' = A_1 v$ and* (3.5). *Then ϕ can be expressed in the form*

$$u' = {}^t A_1^{-1} u + \alpha(v, x, y), \qquad v' = A_1 v, \tag{3.6}$$

$$(x', y') = \varphi(v, x, y), \tag{3.7}$$

where α and φ are vector functions of $2n-k$ variables v, x, y. Moreover, if v is fixed, the transformation φ is symplectic, i.e.,

$$\sum_{j=1}^{m} dx'_j \wedge dy'_j = \sum_{j=1}^{m} dx_j \wedge dy_j. \tag{3.8}$$

Conversely if $\varphi: (x, y) \mapsto (x', y')$ is a local symplectic diffeomorphism in a neighbourhood of the origin which depends on a parameter $v \in \mathbf{R}^k$, then it can be extended to a local symplectic diffeomorphism ϕ of the form (3.6) *and* (3.7) *in (u, v, x, y)-space. Here φ and ϕ are symplectic in the sense that they satisfy conditions* (3.8) *and* (3.5) *respectively.*

Remark. For the later use, all diffeomorphisms are assumed to be real analytic. In the above, if ϕ is real analytic, then so φ is. Conversely if φ is real analytic (in v, x, y), then so its extension ϕ is.

Proof. The symplectic character of ϕ implies that the Jacobian matrix $d\phi$ satisfies

$${}^t(d\phi)\, J\, (d\phi) = J, \tag{3.9}$$

where ${}^t(d\phi)$ is the transposed matrix of $d\phi$,

$$J = \begin{pmatrix} J_k & O \\ O & J_m \end{pmatrix}; \qquad J_l = \begin{pmatrix} O & I_l \\ -I_l & O \end{pmatrix} \quad (l = k, m)$$

and I_l is the identity matrix of degree l. Let us denote

$$d\phi = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where $A = \partial(u', v')/\partial(u, v)$, $B = \partial(u', v')/\partial(x, y)$, $C = \partial(x', y')/\partial(u, v)$, $D = \partial(x', y')/\partial(x, y)$. Then the relation (3.9) is equivalent to

$${}^t A J_k A + {}^t C J_m C = J_k, \qquad {}^t B J_k B + {}^t D J_m D = J_m, \qquad {}^t A J_k B + {}^t C J_m D = O.$$

By these relations together with $v' = A_1 v$, one can easily see that

$$A = \begin{pmatrix} {}^t A_1^{-1} & A_2 \\ O & A_1 \end{pmatrix}, \quad B = \begin{pmatrix} B_1 & B_2 \\ O & O \end{pmatrix}, \quad C = \begin{pmatrix} O & C_1 \\ O & C_2 \end{pmatrix}, \quad {}^t D J_m D = J_m, \tag{3.10}$$

where $A_2=\partial u'/\partial v$, $B_1=\partial u'/\partial x$, $B_2=\partial u'/\partial y$, $C_1=\partial x'/\partial v$ and $C_2=\partial y'/\partial v$. We have thus proved (3.6)–(3.8).

Conversely suppose that we are given a parametrized symplectic mapping $(x', y')=\varphi(v, x, y)$ satisfying (3.8). Since $\varphi_z(h_0, 0)=(\partial\varphi/\partial z)(h_0, 0)$ is a symplectic matrix, we can introduce a parametrized symplectic mapping $\varphi': z\mapsto\varphi'(v, z)$ by

$$\varphi=\varphi_z(h_0, 0)\circ\varphi'.$$

Then for any v sufficiently close to h_0 φ' is close to the identity in a small neighbourhood of the origin $z=0$. If we denote φ' by $(x', y')=\varphi'(v, x, y)$, it can be expressed by means of a generating function as follows:

$$x=\frac{\partial S'}{\partial y}, \quad y'=\frac{\partial S'}{\partial x'}; \quad S'(v, x', y)=\langle x', y\rangle+Q(v, x', y),$$

where $\langle\cdot,\cdot\rangle$ denotes the usual Euclidean scalar product, and $Q(v, x', y)$ is a real analytic function of v, x', y whose power series expansion with respect to x', y begins with terms of order $\geqq 3$. We extend this parametrized mapping to a symplectic mapping $\phi': (u, v, x, y)\mapsto(u', v', x', y')$ defined by

$$u=\frac{\partial S}{\partial v}, \quad v'=\frac{\partial S}{\partial u'}, \quad x=\frac{\partial S}{\partial y}, \quad y'=\frac{\partial S}{\partial x'}$$

with a generating function

$$S(u', v, x', y)=\langle u', v\rangle+S'(v, x', y).$$

If we define a symplectic mapping ϕ by

$$\phi=\begin{pmatrix} A & O \\ O & \varphi_z(h_0, 0)\end{pmatrix}\circ\phi'; \quad A=\begin{pmatrix} {}^tA_1^{-1} & O \\ O & A_1\end{pmatrix},$$

then it has the form (3.6) with (3.7). This completes the proof. □

In the following sections, we call a mapping φ a *parametrized symplectic mapping* (or *parametrized canonical transformation*) if it has the form (3.7) and satisfies (3.8).

Finally we will prove Remarks 2.4 and 2.5.

Proof of Remark 2.4 For the set of functions H_j such that $dH_1, \ldots, dH_k$ are linearly independent at p, let $\psi_1=\psi_1(u, v, x, y), u, v\in\mathbf{R}^k$, $x, y\in\mathbf{R}^m$, be a local symplectic diffeomorphism such that

$$\psi_1: (\mathbf{R}^{2n}, 0)\to(M, p), \quad \text{i.e., } \psi_1(0)=p$$

and

$$H\circ\psi_1-h=v, \tag{3.11}$$

$$\psi_1^*\sigma=\sum_{j=1}^{k} du_j\wedge dv_j+\sum_{j=1}^{m} dx_j\wedge dy_j,$$

where $H=(H_1, \ldots, H_k)$ and $H(p)=h$. We defined here, for the simplicity of notation below, the diffeomorphism ψ_1 in a different way from ψ defined by (3.1)–(3.3).

Let $\hat{H}_1, \ldots, \hat{H}_k \in \mathfrak{F}$ be functions such that

$$\operatorname{span}(d\hat{H}_1, \ldots, d\hat{H}_k)=\operatorname{span}(dH_1, \ldots, dH_k) \quad \text{at } p.$$

Associated with this set of functions $\hat{H}_j$, we introduce a local symplectic diffeomorphism $\psi_2=\psi_2(\hat{u}, \hat{v}, \hat{x}, \hat{y}), \hat{u}, \hat{v} \in \mathbf{R}^k, \hat{x}, \hat{y} \in \mathbf{R}^m$, such that

$$\psi_2 : (\mathbf{R}^{2n}, 0) \to (M, p), \quad \text{i.e., } \psi_2(0)=p$$

and

$$\hat{H} \circ \psi_2 - c = \hat{v},$$

$$\psi_2^* \sigma = \sum_{j=1}^{k} d\hat{u}_j \wedge d\hat{v}_j + \sum_{j=1}^{m} d\hat{x}_j \wedge d\hat{y}_j,$$

where $\hat{H}=(\hat{H}_1, \ldots, \hat{H}_k)$ and $c=\hat{H}(p)$. Let $\psi := \psi_1^{-1} \circ \psi_2$ and represent it as

$$w=\psi(\hat{w}); \quad w=(u, v, x, y), \quad \hat{w}=(\hat{u}, \hat{v}, \hat{x}, \hat{y}).$$

We also write the linearized mapping of ψ at $\hat{w}=0$ as $w=d\psi_0(\hat{w})$, where we use the same coordinates w, $\hat{w}$ for tangent vectors. Since $d\hat{H}_i=\sum_{j=1}^{k} c_{ij} dH_j$ at p, the linear map $d\psi_0$ satisfies $v=A_1 \hat{v}$, A_1 being an invertible constant matrix of degree k. Therefore, by Lemma 3.1, $d\psi_0$ can be expressed as

$$d\psi_0=\begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{3.12}$$

with matrices A, B, C, D of the form (3.10).

From now on, we divide the arguments into two cases according to conditions (i) and (ii) of Definition 2.3. In what follows, we write G in place of $G \circ \psi_1$ and $\hat{G}$ in place of $G \circ \psi_2$.

(i) *nondegenerate equilibrium point case.* Since $\hat{G}=G \circ \psi$, we have $\hat{G}_{\hat{w}}={}^t(d\psi)\, G_w$. Then, since $d\psi_0$ is of the form (3.12) with (3.10) and $G(w)$ is independent of u, it turns out that

$$\hat{G}_{\hat{z}}(0)={}^t D G_z(0),$$

where $z=(x, y)$ and $\hat{z}=(\hat{x}, \hat{y})$. Hence $\hat{z}=0$ is an equilibrium point of $X_{\hat{G}_c}$ if $z=0$ is that of X_{G_h}. Let $T := G_{ww}(0)$, where $G_{ww}(0)$ is the Hessian matrix of G at $w=0$. Then it has the form

$$T=\begin{pmatrix} P & Q \\ R & S \end{pmatrix}$$

with

$$P=\begin{pmatrix} O & O \\ O & P_1 \end{pmatrix}, \quad Q=\begin{pmatrix} O & O \\ Q_1 & Q_2 \end{pmatrix}, \quad R={}^t Q, \quad S={}^t S,$$

where $P_1=G_{vv}(0)$, $Q_1=G_{vx}(0)$, $Q_2=G_{vy}(0)$ and $S=G_{zz}(0)$. Since $\hat{G}(\hat{w})$ is also independent of $\hat{u}$, $\hat{G}_{\hat{w}\hat{w}}(0)$ has the same form as T above. On the other hand,

we have $\hat{G}_{\hat{w}\hat{w}}(0)={}^t(d\psi_0)\,T(d\psi_0)$ and hence, by a direct computation, we can easily see that

$$\hat{G}_{\hat{z}\hat{z}}(0)={}^tDG_{zz}(0)\,D. \tag{3.13}$$

The linearized vector field DX_{G_h} at $\pi_h(p)$ is given by the matrix $J_m G_{zz}(0)$ and the relation (3.13) implies that it is transformed by $d\psi_0$ into $D^{-1}J_m G_{zz}(0)\,D$. Therefore the eigenvalues of $J_m \hat{G}_{\hat{z}\hat{z}}(0)$ are the same as those of $J_m G_{zz}(0)$. This includes the assertion of Remark 2.4 for case (i).

(ii) *nondegenerate fixed point case.* Let $G\in\mathfrak{F}$ be a function such that $\pi_h(p)$ is a fixed point of $f_h:=\pi_h\circ\exp X_G\circ(\pi_h')^{-1}$. We set $\hat{f}_c:=\hat{\pi}_c\circ\exp X_G\circ(\hat{\pi}_c')^{-1}$. Let $\phi:=\psi_1^{-1}\circ\exp X_G\circ\psi_1$ and $\hat{\phi}:=\psi_2^{-1}\circ\exp X_G\circ\psi_2$. Then we have $\hat{\phi}=\psi^{-1}\circ\phi\circ\psi$. Since the map $w'=\phi(w)$ satisfies $v'=v$, $d\phi_0$ can be written as

$$d\phi_0=\begin{pmatrix}P & Q\\ R & S\end{pmatrix}$$

with matrices P, Q, R, S of the same form as A, B, C, D in (3.10) respectively, where $A_1=I_k$. Notice that $\hat{\phi}$ also satisfies the assumption of Lemma 3.1 with $A_1=I_k$. Then it turns out that $d\hat{\phi}_0=d\psi_0^{-1}\circ d\psi_0$ has the same form as $d\phi_0$ given above, where a direct computation shows that S is replaced $D^{-1}SD$. This implies that the eigenvalues of the linearized mapping $d\hat{f}_c$ at $\hat{\pi}_c(p)$ are the same as those of df_h at $\pi_h(p)$. This completes the proof of Remark 2.4. □

Proof of Remark 2.5 Let $G\in\mathfrak{F}$ be a function such that $\pi_h(p)$ is an equilibrium point of the reduced vector field X_{G_h}. In the local coordinate system (u, v, x, y) introduced in the proof of Remark 2.4 above, G is a function of v and $z=(x, y)$ only and hence we write the function G as $G=G(v, z)$ as before. Then $G(v, z)$ satisfies

$$G_z(h, 0)=0.$$

We set $g(v):=G(v, 0)$, which can be considered as a real analytic function of v in a neighbourhood of $v=h$. For any constant $\alpha\in\mathbf{R}\setminus\{0\}$, let us introduce a function

$$\tilde{G}:=\alpha(G-g(H)),$$

where $H=(H_1, \ldots, H_k)$. Then, since the value of H is close to h in a neighbourhood of $\Phi(p)$, $\tilde{G}$ is a real analytic function in a neighbourhood of $\Phi(p)$ and clearly satisfies the identities $\{\tilde{G}, F_i\}\equiv 0$ for all $i=1, \ldots, n$. Therefore we have $\tilde{G}\in\mathfrak{F}$. We note that the vector field $X_{\tilde{G}}$ has an equilibrium point at $p\in M$. This equilibrium point is also a fixed point of the time 1 map $\exp X_{\tilde{G}}$. Let $\pm\lambda_1, \ldots, \pm\lambda_m$ be the eigenvalues of the linearized vector field DX_{G_h} at $\pi_h(p)$. Then the eigenvalues of the linearized mapping df_h at $\pi_h(p)$ are given by $\exp(\pm\alpha\lambda_k)$ $(k=1, \ldots, m)$, where $f_h=\pi_h\circ\exp X_{\tilde{G}}\circ(\pi_h')^{-1}$. Therefore, choosing α with $|\alpha|>0$ sufficiently small, the fixed point $\pi_h(p)$ of f_h is nondegenerate or elliptic provided that it is a nondegenerate or elliptic equilibrium point of X_{G_h} respectively. In the case where $\pi_h(p)$ is a non-resonant equilibrium point of X_{G_h}, we can choose

the constant α so that $\lambda_1, \ldots, \lambda_m$ and $2\pi i/\alpha$ are rationally independent. Then we have $\sum_{k=1}^{m} j_k \alpha \lambda_k \notin 2\pi i \mathbf{Z}$ for all vectors $(j_1, \ldots, j_m) \in \mathbf{Z}^m \setminus \{0\}$ and hence $\pi_h(p)$ is a non-resonant fixed point of f_h. This proves the assertion of Remark 2.5. □

4 Existence of a family of k-dimensional invariant tori

The aim of this section is to prove Theorem 2.6. To this end, we first work in a neighbourhood of p_0 with the local canonical coordinate system (u, v, x, y) $(u, v \in \mathbf{R}^k; x, y \in \mathbf{R}^m)$ introduced in the previous section. In the following, the symbol $O(|z|^d)$ denotes a holomorphic (vector) function in $2n-k$ variables v and $z=(x, y)$ such that its power series expansion with respect to z begins with terms of order $\geqq d$.

Proposition 4.1 *Let $\Phi(p_0)$ be the orbit of rank k passing through a point p_0 for a real analytic integrable system (M, σ, F) and let $H_1, \ldots, H_k \in \mathfrak{F}$ be k functions satisfying condition* (2.2) *at p_0. Assume that $\Phi(p_0)$ is nondegenerate. Then in a neighbourhood of p_0 there exists a real analytic symplectic diffeomorphism $\psi = \psi(u, v, z)$ $(u, v, \in \mathbf{R}^k; z=(x, y), x, y \in \mathbf{R}^m)$ such that in addition to* (3.1)–(3.3) *the following two conditions hold in this coordinate system:*

(a) *Any function $G \in \mathfrak{F}$ has the form*

$$G(v, z) = G(v, 0) + O(|z|^2). \tag{4.1}$$

(b) *For any function $G \in \mathfrak{F}$, the time 1 map $\exp X_G$ is well-defined in a neighbourhood of $\Phi(p_0)$. If p_0 is a fixed point of $\exp X_G$, then in this coordinate system the parametrized symplectic mapping $f(v, \cdot) = \pi_v \circ \exp X_G \circ (\pi'_v)^{-1}$ has the form*

$$f(v, z) = A(v)\, z + O(|z|^2). \tag{4.2}$$

In the above and in what follows, we use the notation $f(v, \cdot)$ in place of f_v to denote the mapping $\pi_v \circ \exp X_G \circ (\pi'_v)^{-1}$.

Proof. For the proof, we use the canonical coordinate system (u, v, z) with $z=(x, y)$ satisfying (3.1)–(3.3). By Remark 2.5, it suffices to prove the assertions under the nondegeneracy condition (ii) of Definition 2.3.

Suppose that condition (ii) of Definition 2.3 is satisfied. Then, in the coordinate system (u, v, z), $z=0$ is a nondegenerate fixed point of the mapping $f(h_0, \cdot) = \pi_{h_0} \circ \exp X_G \circ (\pi'_{h_0})^{-1}$ for some $G \in \mathfrak{F}$. Consider a system of $2m$ equations

$$f(v, z) = z. \tag{4.3}$$

Here the origin $z=0$ is a fixed point of $f(h_0, z)$ and $\det(f_z(h_0, 0) - I) \neq 0$ by the nondegeneracy condition. Hence, by the implicit function theorem, there exists a unique real analytic function $z=\beta(v)$ satisfying equation (4.3) with $\beta(h_0)=0$. Let us carry out a parallel translation $z \mapsto z+\beta(v)$ depending on the parameter v. By Lemma 3.1, this can be extended to a symplectic mapping in (u, v, x, y)-space in such a way that v is invariant under the mapping. In this new canonical coordinate system (u, v, z), $z=0$ is a fixed point of the mapping $f(v, z)$ for any v and therefore $f(v, z)$ has the form (4.2).

To prove that any function $G' \in \mathfrak{F}$ satisfies $G'_z(v, 0) = 0$ for any v, we note that G' is an integral of the mapping $f(v, \cdot)$, i.e.,

$$G'(v, f(v, z)) = G'(v, z).$$

By differentiating this identity with respect to z at $z = 0$, we have

$$f_z(v, 0)\, G'_z(v, 0) = G'_z(v, 0).$$

By the nondegeneracy condition $\det(f_z(h_0, 0) - I) \neq 0$, this implies that $G'_z(v, 0) \equiv 0$, which proves the assertion (a).

The assertion (a) particularly implies that the relation

$$dG' = \sum_{j=1}^{k} c_j\, dH_j$$

holds at p_0. Therefore by the same argument as in the proof of Proposition 2.1 (ii), it turns out that this relation holds along the orbit $\Phi(p_0)$ with the fixed constants $c_j \in \mathbf{R}$ and hence we have

$$\exp(t X_{G'})(p_0) \in \Phi(p_0) \quad \text{for all } t \in \mathbf{R}.$$

This implies that the time 1 map $\exp X_{G'}$ is well-defined in a small neighbourhood of $\Phi(p_0)$. It follows from the relation $\{G, G'\} \equiv 0$ that the mappings $\exp X_G$ and $\exp X_{G'}$ commute. By Lemma 3.1, this leads to the relation

$$f(v, f'(v, z)) = f'(v, f(v, z)),$$

where $f'(v, \cdot) = \pi_v \circ \exp X_{G'} \circ (\pi'_v)^{-1}$ and $f'(h_0, \cdot)$ is assumed to have a fixed point at the origin. Since $f(v, 0) = 0$ for any v, we have

$$f(v, f'(v, 0)) = f'(v, 0),$$

which implies that $f'(v, 0)$ is a fixed point of $f(v, z)$. Since $f'(h_0, 0) = 0$, it follows from the uniqueness of $\beta(v)$ that $f'(v, 0) = 0$ for any v. Thus we have proved the assertion (b). □

We now turn to the proof of Theorem 2.6. In the following, we denote by ϕ_i^t the flow of the vector field X_{H_i} and we set $\phi_1^{t_1} \circ \cdots \circ \phi_k^{t_k}$ for $t = (t_1, \ldots, t_k) \in \mathbf{R}^k$.

Proof of Theorem 2.6 Recall that the isotropy group $P_H(p_0)$ of p_0 under the action (2.3) is a k-dimensional lattice in $\mathbf{R}^k$. Let

$$T_i = (t_1^{(i)}, \ldots, t_k^{(i)}) \in P_H(p_0) \quad (i = 1, \ldots, k)$$

be generators of the lattice $P_H(p_0)$. Then p_0 is a fixed point of the mappings $\phi^{T_i} = \exp X_{H^{(i)}}$ with $H^{(i)} = \sum_{j=1}^{k} t_j^{(i)} H_j$ $(i = 1, \ldots, k)$. Let us introduce the canonical coordinate system (u, v, z) by means of the symplectic diffeomorphism ψ described in Proposition 4.1. We note that $X_{H_i} = \partial/\partial u_i$ in the neighbourhood of p_0 because $H_i = v_i$. Then by Proposition 4.1 (b), we have

$$\psi^{-1} \circ \phi^{T_i} \circ \psi(0, v, 0) = (u^{(i)}(v), v, 0).$$

Here $u^{(i)}(v)$ are real analytic functions of v sufficiently close to h_0 such that $u^{(i)}(h_0)=0$. Therefore $\mathfrak{z}_h:=\psi(0,h,0)$ is a fixed point of the mapping ϕ^t with periods $T^{(i)}(h)=T_i-u^{(i)}(h)\,(i=1,\ldots,k)$, which are linearly independent over $\mathbf{R}$ for any h close to h_0. Consequently the point $\phi^t(\mathfrak{z}_h)$ is well-defined for all $t\in\mathbf{R}^k$ and

$$\Phi_h:=\{\phi^t(\mathfrak{z}_h)\mid t\in\mathbf{R}^k\}$$

are k-dimensional tori which are orbits for the integrable system (M,σ,F). By Proposition 4.1 (a), any function $G\in\mathfrak{F}$ satisfies the relation (2.7) at $\mathfrak{z}_h$ and by the same argument as before it turns out that the relation (2.7) holds everywhere on the orbit $\Phi_h(=\Phi(\mathfrak{z}_h))$. Hence Φ_h are compact orbits of rank k for the integrable system. Moreover it is easily seen that $\{\Phi_h\}_V$ is a symplectic submanifold of dimension $2k$.

It remains to prove the uniqueness of the family $\{\Phi_h\}$. Suppose that there exists another real analytic family of invariant k-tori Φ'_h such that $\Phi'_h=\Phi(\mathfrak{z}'_h)\subset H^{-1}(h)$, $\mathfrak{z}'_{h_0}=p_0$ and the relation (2.7) holds along Φ'_h for any function $G\in\mathfrak{F}$. Here, in the local coordinate system (u,v,z), the points $\mathfrak{z}'_h$ can be expressed as

$$\mathfrak{z}'_h=(0,h,z(h))\in\Sigma_h,$$

where $z(h)$ is a real analytic function of h such that $z(h_0)=0\in\mathbf{R}^{2m}$. We will prove below that $z(h)\equiv 0$, i.e., $\mathfrak{z}'_h=\mathfrak{z}_h$.

First of all, by assumption and Remark 2.5, we note that condition (ii) of Definition 2.3 is satisfied for a function $G\in\mathfrak{F}$. Since the relation (2.7) holds along Φ'_h for the function G, we have

$$\exp X_G=\phi_1^{c_1(h)}\circ\ldots\circ\phi_k^{c_k(h)}\quad\text{on}\quad\Phi'_h, \tag{4.4}$$

where $c_i(h)$ are real analytic functions of h. Here we may assume that $\mathfrak{z}'_{h_0}=p_0$ is a fixed point of the mapping $\exp X_G$ and hence $(c_1(h_0),\ldots,c_k(h_0))\in P_H(p_0)$. In fact, for the parametrized mapping $f(h,\cdot):=\pi_h\circ\exp X_G\circ(\pi'_h)^{-1}$, $\pi_{h_0}(p_0)$ is a fixed point of $f(h_0,\cdot)$ by assumption. This implies that

$$\exp X_G(0,h_0,0)=(u_0,h_0,0),$$

where $u_0:=(u_0^1,\ldots,u_0^k)\in\mathbf{R}^k$ is close to the origin in $\mathbf{R}^k$. Therefore p_0 is a fixed point of the mapping $\exp X_{\tilde G}$ with $\tilde G:=G-\sum_{i=1}^{k}u_0^iH_i$ because $X_{H_i}=\partial/\partial u_i$ in a local neighbourhood of p_0. It is clear that $f(h,\cdot)$ is equal to $\pi_h\circ\exp X_{\tilde G}\circ(\pi'_h)^{-1}$ and hence condition (ii) of Definition 2.3 is satisfied also for $\tilde G\in\mathfrak{F}$. Replacing $\tilde G$ by G, we may assume that p_0 is a fixed point of $\exp X_G$.

By the definition of a real analytic family, generators of the period lattice $P_H(\mathfrak{z}'_h)$ depend on h analytically. Therefore there exists a function $T(h)\in\mathbf{R}^k$ which is real analytic in h and satisfies the condition

$$\phi^{T(h)}(\mathfrak{z}'_h)=\mathfrak{z}'_h\quad\text{and}\quad T(h_0)=(c_1(h_0),\ldots,c_k(h_0)).$$

If we replace $\phi^{T(h)}$ by $\phi^{T(h_0)}$, we have

$$\phi^{T(h_0)}(0,h,z(h))=(u(h),h,z(h))$$

with $u(h)$ being a real analytic function of h such that $u(h_0)=0$. From the relation (4.4) it follows that

$$f(h,\cdot)=\pi_h\circ\phi^{T(h_0)}\circ(\pi'_h)^{-1} \quad \text{on} \quad \Phi'_h$$

and consequently $f(h,z(h))=z(h)$, that is, $(v,z)=(h,z(h))$ is a solution of the system (4.3). By the nondegeneracy condition (ii), the system (4.3) is solved uniquely for $z=z(v)$ holomorphic in v and satisfying $z(h_0)=0$. Since we have $f(h,0)\equiv 0$ by Proposition 4.1, it follows that $z(h)\equiv 0$ and hence $\Phi_h=\Phi'_h$. This completes the proof of Theorem 2.6. □

Remark 4.2 As we stated in Section 2, the family $\{\Phi_h\}$ given in Theorem 2.6 is independent of the choice of the k functions $H_1,\ldots,H_k$. In order to prove this, let $\hat{H}_1,\ldots,\hat{H}_k$ be another set of functions satisfying (2.2) and we set $\hat{H}=(\hat{H}_1,\ldots,\hat{H}_k)$. Then it follows from (2.7) that

$$d\hat{H}_i=\sum_{j=1}^{k} c_{ij}(h)\,dH_j \quad \text{along} \quad \Phi_h, \tag{4.5}$$

where $c_{ij}(h)$ are real analytic functions of h such that $\det(c_{ij}(h))\neq 0$. This implies that Φ_h are also orbits of k vector fields $X_{\hat{H}_1},\ldots,X_{\hat{H}_k}$ and that for any function $G\in\mathfrak{F}$ the relation (2.7) holds along Φ_h with H_i replaced by $\hat{H}_i$. Moreover it turns out that generators of the period lattice $P_{\hat{H}}(\mathfrak{d}_h)$ are linear combinations of those of $P_H(\mathfrak{d}_h)$ and they depend on h analytically. Also, if we set $\hat{h}:=\hat{H}(\mathfrak{d}_h)$, one can see from (4.5) that $\hat{h}$ runs over a neighbourhood of $\hat{H}(p_0)$ in $\mathbf{R}^k$ when h runs over a neighbourhood of h_0 in $\mathbf{R}^k$. Therefore $\{\Phi_h\}$ is also a real analytic family of invariant k-tori for the integrable system such that conditions (a) and (b) of Theorem 2.6 hold with H replaced by $\hat{H}$. By the uniqueness of the real analytic family in Theorem 2.6, this proves that the family $\{\Phi_h\}$ is independent of the choice of k functions $H_1,\ldots,H_k$.

5 Birkhoff normal form theorems

Let $\{\Phi_h\}$ be the family of invariant k-tori for the integrable system (M,σ,F) described in Theorem 2.6. In the course of the proof of Theorem 2.6, we introduced the coordinate system (u,v,x,y) in a neighbourhood of p_0 so that conditions (3.1)–(3.3) and (4.1)–(4.2) are satisfied. In this and the next two sections, we will show that the functions G and the mappings f in Proposition 4.1 can be generally transformed into Birkhoff normal forms with the parameter v provided that the family is non-resonant and elliptic.

For this purpose, we will proceed more generally, independently of the integrable system (M,σ,F). In Definition 5.3 below, we will define precisely (complex) normal forms as well as real normal forms, where the Birkhoff normal forms stated above correspond to the real ones. In the former part of this section, we will formulate results on normalization of functions and mappings containing parameters. They are a series of theorems and corollaries 5.4–5.7 and their proofs will be given in Sections 6 and 7. In the latter part of this section, we will show that these theorems and corollaries imply the existence of a parame-

trized symplectic transformation which takes any function G and any mapping f in Proposition 4.1 (a) and (b) respectively into Birkhoff normal forms.

We begin by preparing some notations. Let v_0 be a point in $\mathbf{C}^k$ or $\mathbf{R}^k$ and let v and z be vectors of dimension k and $2m$ respectively.

Definition 5.1 (i) By $A_{\mathbf{C}}(v_0, 0)$ (or $A_{\mathbf{R}}(v_0, 0)$) we denote the set of all holomorphic (or real analytic resp.) functions $G=G(v, z)$ of v and z on some neighbourhood of $(v_0, 0)$ such that $G(v, 0)\equiv 0$ and $G_z(v, 0)\equiv 0$.

(ii) By $\mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ (or $\mathrm{Diff}_{\mathbf{R}}(v_0, 0)$) we denote the set of all parametrized symplectic mappings $f=f(v, \cdot): z\mapsto f(v, z)$, $f(v, z)$ being a holomorphic (or real analytic resp.) $2m$-dimensional vector function of v and z on some neighbourhood of $(v_0, 0)$ such that $f(v, 0)\equiv 0$.

For any function G or any mapping f in the definition above, the origin $z=0$ in an equilibrium point of the vector field $X_{G(v,\cdot)}$ for any v or is a fixed point of the mapping $f(v, \cdot)$ for any v. In the following, v is always considered as a parameter and we write simply X_G and f to reprersent the parametrized vector field $X_{G(v,\cdot)}$ and the parametrized symplectic mapping $f(v, \cdot)$ respectively. Furthermore, we sometimes omit the term 'parametrized'.

In order to state the theorems, we first prove the following fact.

Lemma 5.2 (i) *Let $G=G(v, z)\in A_{\mathbf{C}}(v_0, 0)$ (or $A_{\mathbf{R}}(v_0, 0)$). Assume that $z=0$ is a non-resonant equilibrium point of X_G for some v in a neighbourhood V of v_0 in $\mathbf{C}^k$ (or $\mathbf{R}^k$). Then it is non-resonant for any v belonging to a residual subset of V.*

(ii) *Let $f=f(v, z)\in\mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ (or $\mathrm{Diff}_{\mathbf{R}}(v_0, 0)$). Assume that $z=0$ is a non-resonant fixed point of f for some v in a neighbourhood V of v_0 in $\mathbf{C}^k$ (or $\mathbf{R}^k$). Then it is non-resonant for any v belonging to a residual subset of V.*

Remark. The redisual subset of V is defined as the intersection of countable numbers of open and dense subsets of V. It is dense in V by the Baire's property,

Proof. (i) Let $\pm\lambda_i(v)$ $(i=1, \ldots, m)$ be the eigenvalues of $JG_{zz}(v, 0)$. For any non-zero integer vector $j=(j_1, \ldots, j_m)\in\mathbf{Z}^m\setminus\{0\}$, we introduce a function

$$f_j(v) := \sum_{\nu=1}^{m} j_\nu \lambda_\nu(v).$$

The origin $z=0$ is a non-resonant equilibrium point of X_G if the parameter v belongs to a set V_{NR} defined by

$$V_{NR} := \bigcap_{j\in\mathbf{Z}^m\setminus\{0\}} V_j; \qquad V_j=\{v\in V \mid f_j(v)\neq 0\}.$$

By assumption, this set V_{NR} is not empty. For any integer vector j, we observe that $V_j^c=\{v\in V \mid f_j(v)=0\}$ is a closed subset of V without inner points. Indeed, if it has an inner point, the analyticity of f_j implies that $f_j(v)\equiv 0$, which contradicts the fact that V_{NR} be non-empty. Therefore V_j is an open and dense subset of V. This implies the desired assertion.

(ii) Let $\lambda_i(v), \lambda_i^{-1}(v)\,(i=1, \ldots, m)$ be the eigenvalues of $f_z(v, 0)$. For any non-zero integer vector $j=(j_1, \ldots, j_m)\in \mathbf{Z}^m\setminus\{0\}$, we introduce a function

$$f_j(v) := \prod_{\nu=1}^{m} (\lambda_\nu(v))^{j_\nu}.$$

The fixed point $z=0$ of $f(v, \cdot)$ is non-resonant for any v belonging to a set V_{NR} defined by

$$V_{NR} := \bigcap_{j\in \mathbf{Z}^m\setminus\{0\}} V_j; \qquad V_j=\{v\in V \mid f_j(v)\neq 1\}.$$

Similarly as above, the set V_j turns out to be open and dense in V. This completes the proof. □

In view of the lemma above, for a function $G\in A_{\mathbf{C}}(v_0, 0)$ we say that the equilibrium point $z=0$ of X_G is *generically non-resonant* if it is non-resonant for some v in a neighbourhood of v_0 and if for $v=v_0$ the eigenvalues of the linearized vector field of X_G at $z=0$ are all distinct. Similarly for a mapping $f\in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$, we say that the fixed point $z=0$ of f is *generically non-resonant* if is non-resonant for some v in a neighbourhood of v_0 and if for $v=v_0$ the eigenvalues of the linearized mapping of f at $z=0$ are all distinct.

We define *normal forms* as follows:

Definition 5.3 (i) A function $G=G(v, z)\in A_{\mathbf{C}}(v_0, 0)$ is said to be in *normal form* (or *real normal form*) if it is a holomorphic function of n variables $v=(v_1, \ldots, v_k)$ and $\tau=(\tau_1, \ldots, \tau_m)$ with $\tau_j=x_j y_j$ (or $\tau_j=(x_j^2+y_j^2)/2$ resp.).

(ii) A mapping $f=f(v, z)\in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ is said to be in *normal form* (or *real normal form*) if it is a mapping of the form $f(v, z)=\exp X_G$ with a function $G\in A_{\mathbf{C}}(v_0, 0)$ which is in normal form (or real normal form resp.) in the sense of (i) above.

Remark. In the definition of the real normal form, we do not require the function or the mapping to be real analytic.

In the following, we consider functions $G\in A_{\mathbf{C}}(v_0, 0)$ (or $A_{\mathbf{R}}(v_0, 0)$) as power series in z whose coefficients are holomorphic (or real analytic resp.) functions of v. We expand it in homogeneous polynomials in z as follows:

$$(5.1) \qquad G(v, z)=G^0(v, z)+G^1(v, z)+\ldots; \qquad G^0(v, z)\not\equiv 0, \deg G^0(v, z)=s\geqq 2,$$

where $G^j(v, z)$ are homogeneous polynomials of degree $s+j$ in z whose coefficients are holomorphic functions of v. We call $G^0(v, z)$ the *lowest order part* of $G(v, z)$. Here it is to be noted that $G^0(v, z)$ can vanish identically for some fixed v.

Now we state theorems establishing normal forms of functions and mappings. For this purpose, we need the existence of m integrals of the parametrized vector fields or mappings such that their lowest order parts are functionally independent functions of z for $v=v_0$ fixed. Here and in what follows, we say that a function $F\in A_{\mathbf{C}}(v_0, 0)$ (or $A_{\mathbf{R}}(v_0, 0)$) is an integral of the parametrized vector field X_G

if $F(v, z)$ is an integral of X_G for any v fixed. This implies that the Poisson bracket $\{F, G\}$ vanishes identically with respect to v and z, where

$$\{F, G\} = \sum_{j=1}^{m} \left(\frac{\partial F}{\partial x_j} \frac{\partial G}{\partial y_j} - \frac{\partial F}{\partial y_j} \frac{\partial G}{\partial x_j} \right).$$

On the other hand, we call a function $G \in A_{\mathbf{C}}(v_0, 0)$ (or $A_{\mathbf{R}}(v_0, 0)$) an integral of the mapping f if $G(v, z)$ is an integral of f for any v fixed, that is, $G(v, f(v, z)) = G(v, z)$ identically with respect to v and z. In what follows, we write $G \circ f$ in place of $G(v, f(v, \cdot))$ for the simplicity of notation.

First we consider the case of a parametrized vector field near an equilibrium point.

Theorem 5.4 *Let $G(v, z) \in A_{\mathbf{C}}(v_0, 0)$ and assume that the origin $z = 0$ is a generically non-resonant equilibrium point of X_G. Assume that the vector field X_G possesses m integrals $G_1, \ldots, G_m \in A_{\mathbf{C}}(v_0, 0)$ such that $G_i^0(v_0, z)$ $(i = 1, \ldots, m)$ are functionally independent functions of z. Then there exists a transformation $\varphi \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ such that $G \circ \varphi$ is in normal form. Moreover the following holds:*

(i) *If a function $F \in A_{\mathbf{C}}(v_0, 0)$ is an integral of X_G, then $F \circ \varphi$ is in normal form.*

(ii) *If G is an integral of a mapping $f \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$, then $\varphi^{-1} \circ f \circ \varphi$ is in normal form.*

(iii) *If $\varphi' \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ is another transformation which takes G into normal form, then $\varphi^{-1} \circ \varphi'$ is in normal form.*

Remark. The property (iii) implies that the normal forms in the above are uniquely determined independently of the choice of φ.

Corollary 5.5 *In addition to the assumptions of Theorem 5.4, assume that $G(v, z) \in A_{\mathbf{R}}(v_0, 0)$ and that the equilibrium point $z = 0$ of X_G is elliptic for $v = v_0$. Then there exists a transformation $\varphi \in \mathrm{Diff}_{\mathbf{R}}(v_0, 0)$ such that all assertions of Theorem 5.4 hold with 'normal form' replaced by 'real normal form'.*

In the case of a parametrized symplectic mapping near a fixed point, we have

Theorem 5.6 *Let $f(v, z) \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ and assume that the origin $z = 0$ is a generically non-resonant fixed point of f. Assume that there exist m integrals $G_1, \ldots, G_m \in A_{\mathbf{C}}(v_0, 0)$ of f such that $G_i^0(v_0, z)$ $(i = 1, \ldots, m)$ are functionally independent functions of z. Then there exists a transformation $\varphi \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ such that $\varphi^{-1} \circ f \circ \varphi$ is in normal form. Moreover the following holds:*

(i) *If a function $F \in A_{\mathbf{C}}(v_0, 0)$ is an integral of f, then $F \circ \varphi$ is in normal form.*

(ii) *If a mapping $f' \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ satisfies the relation $f \circ f' = f' \circ f$, then $\varphi^{-1} \circ f' \circ \varphi$ is in normal form.*

(iii) *If $\varphi' \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ is another transformation which takes f into normal form, then $\varphi^{-1} \circ \varphi'$ is in normal form.*

Corollary 5.7 *In addition to the assumptions of Theorem 5.6, assume that $f(v, z) \in \mathrm{Diff}_{\mathbf{R}}(v_0, 0)$ and that the fixed point $z = 0$ of f is elliptic for $v = v_0$. Then there exists a transformation $\varphi \in \mathrm{Diff}_{\mathbf{R}}(v_0, 0)$ such that all assertions of Theorem 5.6 hold with 'normal form' replaced by 'real normal form'.*

In the above series of theorems and corollaries 5.4–5.7, the transformations φ are called the (parametrized) Birkhoff transformations. We will prove these theorems and corollaries in Sections 6 and 7.

In the rest of this section, we will show that the above results are applicable to the situation of Theorem 2.8. Let $\{\Phi_h\}_V$ be the family of invariant k-tori for the integrable system (M, σ, F) satisfying the assumptions of Theorem 2.8. Since the algebra $\mathfrak{F}$ contains n functionally independent functions, we can find a set of m functions $G_1, \ldots, G_m \in \mathfrak{F}$ such that $H_1, \ldots, H_k$ and $G_1, \ldots, G_m$ are functionally independent. For these n functions we have the local expressions $H_i = v_i\,(i=1, \ldots, k)$ and $G_i = G_i(v, z)\,(i=1, \ldots, m)$. Suppose that the functions $G_i(v, z)$ are holomorphic on a domain $V \times \Xi$, where V is a domain of $\mathbf{C}^k$ containing $v = v_0$ and Ξ a domain of $\mathbf{C}^{2m}$ containing $z=0$. Then the functional independence of $H_1, \ldots, H_k, G_1, \ldots, G_m$ is equivalent to the condition

$$\text{(5.2)} \qquad \operatorname{rank}\left(\frac{\partial(G_1, \ldots, G_m)}{\partial(z_1, \ldots, z_{2m})}\right) = m \qquad \text{on an open and dense subset of } V \times \Xi.$$

In the following, we say that the functions $G_1, \ldots, G_m$ are functionally independent if condition (5.2) is satisfied. This functional independence does not necessarily imply that of their lowest order parts $G_1^0, \ldots, G_m^0$. However we can show the following

Lemma 5.8 *Let $G_1, \ldots, G_m \in A_{\mathbf{C}}(v_0, 0)$ be functionally independent functions. Then there exist m functions $\hat{G}_1, \ldots, \hat{G}_m \in A_{\mathbf{C}}(v_0, 0)$ such that* (i) *they are polynomials of $G_1, \ldots, G_m$ whose coefficients are holomorphic functions of v in a neighbourhood of v_0, and* (ii) *their lowest order parts $\hat{G}_j^0$ $(j=1, \ldots, m)$ are functionally independent.*

Proof. This lemma is a direct generalization of Lemma 2.1 of [7] and we give the outline of its proof (see [7, Sect. 9] for details). We owe the proof basically to Ziglin [16]. The following algebraic lemma is crucial for the construction of $\hat{G}_j$ $(j=1, \ldots, m)$.

Lemma. *Let k be a field of characteristic 0 and $k[z_1, \ldots, z_n]$ the polynomial ring over k. If $f_1, \ldots, f_r \in k[z_1, \ldots, z_n]$ are functionally dependent, then they are algebraically dependent.*

For the proof of this fact, we refer to [7, Lemma 9.1], where the field k is assumed to be $\mathbf{R}$ or $\mathbf{C}$ that can be generalized as above without any modification.

Now, by assumption $G_1, \ldots, G_r$ (for any $r \geqq 1$) are functionally independent and therefore one can find a non-vanishing minor $M = M(v, z)$ in the form

$$M := \det\left(\frac{\partial(G_1, \ldots, G_r)}{\partial(z_{i_1}, \ldots, z_{i_r})}\right) \not\equiv 0.$$

We denote by M^0 the lowest order part of M. Observe that if we set

$$\mu(G_1, \ldots, G_r) := \deg M^0 + r - \sum_{i=1}^{r} \deg G_i^0,$$

then $\mu(G_1, \ldots, G_r) \geqq 0$ and $G_1^0, \ldots, G_r^0$ are functionally independent provided that $\mu(G_1, \ldots, G_r) = 0$. Suppose that $G_1^0, \ldots, G_{r-1}^0$ $(2 \leqq r \leqq m)$ are functionally indepen-

dent and that $G_1^0, \ldots, G_r^0$ are functionally dependent. Let $A(v)$ be the ring of convergent power series in v in a neighbourhood of v_0 and let $K(v)$ be the quotient field of this ring (integral domain) $A(v)$. Applying the above lemma to $f_i = G_i^0 \in K(v)[z_1, \ldots, z_{2m}]$ $(i=1, \ldots, r)$, we can find a polynomial $P \in K(v)[X_1, \ldots, X_r]$ such that

$$P(G_1^0, \ldots, G_r^0) \equiv 0. \tag{5.3}$$

Multiplying P by the least common multiplier of its divisors of the coefficients, the polynomial P can be chosen so that $P \in A(v)[X_1, \ldots, X_r]$. If we choose the polynomial P in such a way that its degree is minimal among those satisfying the identity (5.3), the function $G := P(G_1, \ldots, G_r)$ satisfies a strict inequality

$$\mu(G_1, \ldots, G_{r-1}, G) < \mu(G_1, \ldots, G_{r-1}, G_r)$$

(see [7, Sect. 9]). Therefore, repeating this process finite times, one can reach a function $\hat{G}_r = P_r(G_1, \ldots, G_r)$ with $P_r \in A(v)[X_1, \ldots, X_r]$ such that $\mu(G_1, \ldots, G_{r-1}, \hat{G}_r) = 0$. This implies that $G_1^0, \ldots, G_{r-1}^0, \hat{G}_r^0$ are functionally independent. The lemma is proved by using this procedure inductively. □

Suppose that the functions $G_1, \ldots, G_m$ are integrals of the vector field X_G or those of the mapping f in Proposition 4.1. Then the functions $\hat{G}_1, \ldots, \hat{G}_m$ obtained above are also their integrals. Let us now rewrite $G_1, \ldots, G_m$ in place of $\hat{G}_1, \ldots, \hat{G}_m$. To apply the above theorems and corollaries, we have to require that $G_i^0(v_0, z)$ $(i=1, \ldots, m)$ are functionally independent functions of z. This is not the case in general even though the lowest order parts $G_1^0, \ldots, G_m^0$ are functionally independent (i.e., condition (5.2) holds with G_i replaced by G_i^0). However it turns out that there exists an open and dense subset $\hat{V}$ of V such that the following condition holds for any $v \in \hat{V}$ fixed:

$$\operatorname{rank}\left(\frac{\partial(G_1^0, \ldots, G_m^0)}{\partial(z_1, \ldots, z_{2m})}\right) = m \quad \text{on an open and dense subset of } \Xi.$$

Let us now take $v \in \hat{V}$ arbitrarily and call it v_0. Here the letter v_0 denotes in general a different point from v_0 in the symbol $A_{\mathbf{C}}(v_0, 0)$, but we use it for the two meanings to avoid the complexity of notation. Suppose that the family $\{\Phi_h\}_V$ is non-resonant (and elliptic). Then, from theorems and corollaries 5.4–5.7, we conclude that any function $G \in \mathfrak{F}$ and any mapping f in Proposition 4.1 (a) and (b) respectively can be taken into (real) normal forms by a suitable transformation $\varphi \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ (or $\mathrm{Diff}_{\mathbf{R}}(v_0, 0)$). This assertion will be formulated precisely in Theorem 8.1 for the extension of φ to (u, v, x, y)-space.

Remark. In view of Remark 2.5, we can also conclude the above assertion from Theorem 5.6 and Corollary 5.7 only. Incidentally we will not need Theorem 5.4 nor Corollary 5.5 to prove the main theorem 2.8. We stated them and will prove them from the viewpoint independent of the proof of Theorem 2.8.

Finally let us discuss the relation between the result by Eliasson [5] and ours. In the proof of our above results, the assumption that $G_i^0(v_0, z)$ be functionally independent function of z plays an important role. This type of assumption is related to that of Eliasson's theorem. In his theorem, he assumed that the

lowest order parts of functions $G_i|\Sigma_{v_0}=G_i(v_0,z)$ at the relative equilibrium point $z=0$ have the following form:

$$G_i^0(v_0,z)=\sum_{j=1}^{m}\beta_{ij}x_jy_j \quad (i=1,\ldots,m) \quad \text{with } \det(\beta_{ij})\neq 0, \tag{5.4}$$

where G_i are Poisson commuting functions which come from the integrable system (M,σ,F) in the same way as above. Clearly G_i^0 are functionally independent functions of z and hence it is a special case satisfying the assumptions of Theorem 5.4 and Corollary 5.5. More precisely, in his case there is no assumption that $z=0$ be a generically non-resonant equilibrium point. However one can easily see that, if $G_i^0(v_0,z)$ satisfies condition (5.4), then $z=0$ is a generically non-resonant equilibrium point of a vector field X_G with some function $G\in A_{\mathbf{C}}(v_0,0)$ (see [7, Sect. 1]). Therefore our result includes Eliasson's one for analytic case. It is to be noted that the assumption of Theorem 5.6 (and Coroll. 5.7), which comes from condition (ii) of Definition 2.7, is new.

6 Proof of Theorem 5.4 and Corollary 5.5

In this section, we prove Theorem 5.4 and Corollary 5.5 in several steps.

(a) Linear transformation

First of all, we normalize the quadratic terms of $G(v,z)$. It corresponds to diagonalization of the matrix $JG_{zz}(v,0)$.

Proposition 6.1 *For a function $G(v,z)\in A_{\mathbf{C}}(v_0,0)$, let $\pm\lambda_j(v)\,(j=1,\ldots,m)$ be the eigenvalues of the matrix $JG_{zz}(v,0)$ and assume that they are distinct to each other for any $v\in V$, where V is a neighbourhood of v_0 in $\mathbf{C}^k$. Then there exists a symplectic matrix $C(v)$ whose components are holomorphic functions of v such that the transformation $z\mapsto C(v)\,z$ takes $G(v,z)$ into the following form*

$$G(v,z)=\sum_{j=1}^{m}\lambda_j(v)\,x_jy_j+O(|z|^3). \tag{6.1}$$

Proof. By the simplicity of the eigenvalues, $\lambda_j(v)$ are holomorphic functions of $v\in V$. Also the eigenvectors corresponding to these eigenvalues can be taken as holomorphic functions of $v\in V$ since each eigenspace is 1-dimensional. Let $c^{(j)}(v)$ and $c^{(j+m)}(v)$ be the eigenvectors corresponding to $\lambda_j(v)$ and $-\lambda_j(v)$ respectively. Then the matrix $C(v)=(c^{(1)},\ldots,c^{(2m)})$ diagonalizes the matrix $JG_{zz}(v,0)$ for any $v\in V$. This matrix $C(v)$ is not symplectic in general. However, following the discussion by Siegel [14, pp. 100–101], we can construct a symplectic matrix $\hat{C}(v)$ which diagonalizes $JS(v)$, where $S(v)=G_{zz}(v,0)$. In this step, the analyticity of the components of the matrices is preserved and the components of $\hat{C}(v)$ are also holomorphic functions of $v\in V$. If we set

$$\Lambda=\operatorname{diag}(\lambda_1(v),\ldots,\lambda_m(v),-\lambda_1(v),\ldots,-\lambda_m(v)), \tag{6.2}$$

we have $\hat{C}^{-1}JS\hat{C}=\Lambda$, and it follows from this that ${}^t\hat{C}S\hat{C}=-J\Lambda$ by using the symplectic character ${}^t\hat{C}J\hat{C}=J$. Consequently the function $G(v,\hat{C}(v)z)$ is in the desired form (6.1). □

(b) Structure of normalization of Poisson commuting functions

In this subsection we discuss the structure of normalization of Poisson commuting functions and prove the assertions (i) and (ii) of Theorem 5.4. To this end, we first define normal forms up to terms of finite order for functions and mappings.

We say that a function $G(v,z)\in A_{\mathbf{C}}(v_0,0)$ is in *normal form up to terms of order* $s+d-1$ ($s=\deg G^0, d\geqq 1$) if for the expansion (5.1) the polynomial $G^0+G^1+\ldots+G^d$ is in normal form, i.e., G can be written in the form

$$G(v,z)=g(v,z)+\hat{G}(v,z);\quad g=P_N g,\quad \hat{G}=O(|z|^{s+d}). \tag{6.3}$$

Here and in what follows, P_N is the operator defined by

$$P_N g(v,z)=\sum_{2|\alpha|\geqq s} c_{\alpha\alpha}(v)\,x^\alpha y^\alpha \quad \text{for } g(v,z)=\sum_{|\alpha|+|\beta|\geqq s} c_{\alpha\beta}(v)\,x^\alpha y^\beta,$$

where we used the multi-index notation $x^\alpha=x_1^{\alpha_1}\ldots x_m^{\alpha_m}$, $y^\beta=y_1^{\beta_1}\ldots y_m^{\beta_m}$, $|\alpha|=\sum_{i=1}^m \alpha_i$, $|\beta|=\sum_{i=1}^m \beta_i$. The above condition $g=P_N g$ implies that g is in normal form.

In the case of mappings, we say that a mapping $f\in \mathrm{Diff}_{\mathbf{C}}(v_0,0)$ is in *normal form up to terms of order d* if it is expressed as

$$f=f_d\circ\chi$$

with a normal form $f_d\in\mathrm{Diff}_{\mathbf{C}}(v_0,0)$ and $\chi\in\mathfrak{S}_d$, where

$$\mathfrak{S}_d=\{\chi\in\mathrm{Diff}_{\mathbf{C}}(v_0,0)\,|\,\chi(v,\zeta)=\zeta+O(|\zeta|^{d+1})\}.$$

For the later purpose, it is useful to notice that a transformation $\chi\in\mathfrak{S}_d$ can be written in the form

$$\chi(v,\zeta)=\zeta+JVW^{d+2}(v,\zeta)+O(|\zeta|^{d+2}),$$

where W^{d+2} is a homogeneous polynomial of degree $d+2$ in ζ with coefficients holomorphic in v. This can be easily seen from the expression of χ by using a generating function (see e.g., [7, Lemma 2.7]).

Our purpose is to prove the following proposition which includes assertions (i) and (ii) of Theorem 5.4.

Proposition 6.2 *Let $G=G(v,z)\in A_{\mathbf{C}}(v_0,0)$ and assume that $z=0$ is a generically non-resonant equilibrium point of the vector field X_G. Suppose that G is in normal form up to terms of order $s+d-1$ with $s=\deg G^0=2$. Then the following holds:*

(i) *If* $F \in A_{\mathbf{C}}(v_0, 0)$ *is a function such that* $\{F, G\} \equiv 0$, *then it is in normal form up to terms of order* $s' + d - 1$ *with* $s' = \deg F^0$.

(ii) *If* G *is an integral of a mapping* $f \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$, *then* f *is in normal form up to terms of order* d.

Remark. The above assertions hold also when the parameter v is fixed at any v for which $z=0$ is a non-resonant equilibrium point of $X_{G(v,\cdot)}$. More precisely, suppose that, for example, v_0 is such a parameter value. Then, if we replace $G(v, z)$ by $G(v_0, z)$ in the assumption, we have the same conclusions as above with $F(v, z)$ and $f(v, z)$ replaced by $F(v_0, z)$ and $f(v_0, z)$ respectively. This will be easily seen from the proof of this proposition below.

Proof. (i) We write G and F in the form (5.1). Then comparison of terms of order $s' + d - 1$ in the identity $\{G, F\} = 0$ gives

$$\{G^0, F^{d-1}\} + \{G^1, F^{d-2}\} + \ldots + \{G^{d-1}, F^0\} = 0. \tag{6.4}$$

If $d=1$, this implies that $\{G^0, F^0\} = 0$. By assumption, G^0 has the form

$$G^0 = \sum_{j=1}^{m} \lambda_j(v)\, x_j y_j.$$

Therefore the identity $\{G^0, F^0\} = 0$ leads to

$$\sum_{j=1}^{m} \lambda_j(v)(y_j F^0_{y_j} - x_j F^0_{x_j}) = 0.$$

Setting $F^0 = \sum_{|\alpha|+|\beta|=s'} c_{\alpha\beta}(v)\, x^\alpha y^\beta$, we obtain the identity

$$\sum_{|\alpha|+|\beta|=s'} c_{\alpha\beta}(v)\, \gamma(v)\, x^\alpha y^\beta = 0; \qquad \gamma(v) = \sum_{j=1}^{m} \lambda_j(v)(\beta_j - \alpha_j).$$

By Lemma 5.2, $\lambda_j(v)\,(j=1, \ldots, m)$ are rationally independent generically with respect to v and therefore we have $\gamma(v) \neq 0$ for $\alpha \neq \beta$ densely with respect to v. Consequently $c_{\alpha\beta}(v) \equiv 0$ for $\alpha \neq \beta$ by the continuity of $c_{\alpha\beta}(v)$ and hence the lowest order part F^0 is in normal form. For $d \geqq 2$, assume that G and F are in normal form up to terms of order $s+d-1$ and $s'+d-2$ respectively. Then since G^j and $F^{d-j-1}\,(j=1, \ldots, d-1)$ are in normal form, we have $\{G^j, F^{d-j-1}\} = 0$. Therefore the identity (6.4) leads to

$$\{G^0, F^{d-1}\} = 0.$$

In the same way as above, this implies that F^{d-1} is in normal form. This completes the proof of (i).

(ii) By assumption, we have the identity $G(v, f(v, z)) = G(v, z)$. Comparing the lowest order parts in this identity, we have

$$G^0(v, A(v)\, z) = G^0(v, z), \tag{6.5}$$

where $A(v)=f_z(v,0)$. This implies that the linear vector field X_{G^0} is invariant under the linear transformation $z\mapsto A(v)z$, i.e., $A^{-1}JSA=JS$, where $S=G^0_{zz}(v,0)$. Since G^0 is in normal form, JS is a diagonal matrix of the form (6.2) and their components are distinct to each other by the non-resonance condition. Therefore one can easily see from the relation $A^{-1}JSA=JS$ that $A(v)$ has the form

$$A(v)=\mathrm{diag}(\mu_1(v),\ldots,\mu_m(v),\mu_1^{-1}(v),\ldots,\mu_m^{-1}(v)) \tag{6.6}$$

for some $\mu_i(v)\in\mathbf{C}(i=1,\ldots,m)$. This implies that

$$A(v)z=\exp X_G(z);\qquad G=\sum_{j=1}^{m}\lambda_j(v)x_jy_j\quad(\lambda_j(v)=\log\mu_j(v))$$

and hence f is in normal form up to terms of order 1.

To prove the assertion by induction, suppose that f is in normal form up to terms of order $d-1$. Then f is written in the form $f=f_{d-1}\circ\chi$ with a normal form f_{d-1} and $\chi\in\mathfrak{S}_{d-1}$ which can be written as

$$\chi(v,\zeta)=\zeta+J\nabla W^{d+1}(v,\zeta)+O(|\zeta|^{d+1}),$$

where W^{d+1} is a homogeneous polynomial of degree $d+1$ in ζ with coefficients holomorphic in v. Let us define a function G' by

$$G'(v,z)=G(v,f_{d-1}(v,z)).$$

If we write G in the form (6.3), then G' has the form

$$G'(v,z)=g(v,z)+\hat{G}'(v,z);\qquad P_N g=g,\qquad \hat{G}'(v,z)=O(|z|^{s+d})$$

because f_{d-1} preserves the products x_iy_i. Therefore we have

$$\begin{aligned}G(v,f(v,\zeta))&=G'(v,\chi(v,\zeta))\\&=g(v,\zeta)+\{g(v,\zeta),W^{d+1}(v,\zeta)\}+O(|\zeta|^{s+d}).\end{aligned}\tag{6.7}$$

Since $G(v,f(v,\zeta))=G(v,\zeta)$, which is in normal form up to terms of order $s+d-1$, it follows from (6.7) that $\{G^0,W^{d+1}\}$ is in normal form. By the non-resonance condition, it turns out that W^{d+1} is in normal form by the same arguments as above. This implies that $f=f_{d-1}\circ\chi$ is in normal form up to terms of order d. This completes the proof. □

(c) Construction of the Birkhoff transformation

The desired transformation φ will be obtained as a composition of infinite number of transformations $\varphi_\nu\in\mathrm{Diff}_{\mathbf{C}}(v_0,0)\,(\nu=0,1,\ldots)$, where φ_ν provides a better approximation to the normal form of G. In this subsection, we construct the transformations φ_ν.

To describe one iteration step, let us assume that the function G is in normal form up to terms of order $s+d-1\,(s=\deg G^0)$. Then by Proposition 6.2, the integrals G_i of $X_G(i=1,\ldots,m)$ are in normal form up to terms of order s_i+d

-1 ($s_i = \deg G_i^0$). In the following, we put $G = G_0$ and $s = s_0 (= 2)$ to treat G and G_i in a unified manner. Then G_i ($i = 0, 1, \ldots, m$) are written in the form

$$G_i(v, z) = g_i(v, z) + \hat{G}_i(v, z); \quad g_i = P_N g_i, \quad \hat{G}_i = O(|z|^{s_i + d}). \tag{6.8}$$

The iteration step consists of finding a transformation $\varphi \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ which normalizes G_i up to terms of order $s_i + 2d - 1$ ($i = 0, 1, \ldots, m$). For this purpose, we define the transformation φ by the time 1 map $\varphi = \exp X_W$ with the Hamiltonian

$$W = W(v, z) = W^{d+2} + W^{d+3} + \ldots + W^{2d+1}, \tag{6.9}$$

where W^j are homogeneous polynomials of degree j in z with coefficients holomorphic in v. Since $\varphi^t(v, \zeta) := \exp(tX_W)$ is a solution of the system $\dot{z} = J\nabla W$, the transformation φ is written as

$$\varphi\colon\ \zeta \mapsto \varphi^1(v, \zeta) = \zeta + \int_0^1 J\nabla W(v, \varphi^t(v, \zeta))\, dt.$$

Here ∇W is the gradient vector with components W_{z_i} ($i = 1, \ldots, 2m$). One can easily see that this transformation φ has the form

$$\varphi(v, \zeta) = \zeta + J\nabla W(v, \zeta) + O(|\zeta|^{2d+1})$$

(see [7, Sect. 3]). Hence it follows from (6.8) that

$$G_i(v, \varphi(v, \zeta)) = g_i(v, \zeta) + \{g_i(v, \zeta), W(v, \zeta)\} + \hat{G}_i(v, \zeta) + O(|\zeta|^{s_i + 2d}).$$

Since g_i are in normal form, we have the identities $P_N\{g_i, W\} = 0$ and therefore the functions $G_i(v, \varphi(v, \zeta))$ ($i = 0, 1, \ldots, m$) are in normal form up to terms of order $s_i + 2d - 1$ if and only if W satisfies equations

$$\{g_i(v, \zeta), W(v, \zeta)\} + \hat{G}_i(v, \zeta) = P_N \hat{G}_i(v, \zeta) + O(|\zeta|^{s_i + 2d}). \tag{6.10}_i$$
$$(i = 0, 1, \ldots, m)$$

By comparison of the homogeneous parts of degree $s_i + l$ ($d \leqq l \leqq 2d - 1$) in $(6.10)_i$, we obtain

$$\{g_i^0, W^{l+2}\} = -(I - P_N)\hat{G}_i^l - \sum_{\nu=1}^{l-d} \{g_i^\nu, W^{l+2-\nu}\} \quad (i = 0, 1, \ldots, m),$$

where I is the identity operator. Notice that these equations can be written as

$$\sum_{j=1}^{m} \frac{\partial g_i^0}{\partial \omega_j} D_j W^{l+2} = F_i^l; \quad F_i^l := -(I - P_N)\hat{G}_i^l - \sum_{\nu=1}^{l-d} \{g_i^\nu, W^{l+2-\nu}\}, \tag{6.11}_i$$
$$(i = 0, 1, \ldots, m; l = d, \ldots, 2d - 1)$$

where D_j is the operator defined by

$$D_j = \{\omega_j, \cdot\}, \quad \omega_j = \xi_j \eta_j \quad (j = 1, \ldots, m).$$

Since $g_i^0 (= G_i^0)$ are polynomials in m variables $\omega_1, \ldots, \omega_m$ with coefficients holomorphic in v, the assumption of the functional independence of $G_i^0(v_0, z)$ $(i = 1, \ldots, m)$ implies that the condition

$$\det\left(\frac{\partial g_i^0}{\partial \omega_j}\right)_{i,j=1,\ldots,m} \not\equiv 0 \tag{6.12}$$

holds at $v = v_0$.

The existence of a formal Birkhoff transformation φ is reduced to the following proposition establishing one iteration step of normalization.

Proposition 6.3 *Under the assumptions of Theorem 5.4, suppose that $G(v, z)$ is in normal form up to terms of order $s+d-1$ with $s = \deg G^0 = 2$. Then there exists a holomorphic function $W(v, \zeta)$ of the form (6.9) such that, for $\varphi = \exp X_W$, the function $G \circ \varphi$ is in normal form up to terms of order $s+2d-1$. It is uniquely determined by imposing the condition $P_N W \equiv 0$. Moreover, for each homogeneous polynomial $W^{l+2}(v, \zeta)$ in ζ $(l = d, \ldots, 2d-1)$, the functions $D_j W^{l+2}$ $(j = 1, \ldots, m)$ are expressed as*

$$D_j W^{l+2}(v, \zeta) = \frac{q_j^l(v, \zeta)}{p(v, \zeta)} \qquad (j = 1, \ldots, m;\ l = d, \ldots, 2d-1) \tag{6.13}$$

with

$$q_j^l(v, \zeta) = \det \begin{pmatrix} g^0_{1\omega_1} & \cdots & \overset{(j)}{F_1^l} & \cdots & g^0_{1\omega_m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ g^0_{m\omega_1} & \cdots & F_m^l & \cdots & g^0_{m\omega_m} \end{pmatrix}, \tag{6.14}$$

$$p(v, \zeta) = \det\left(\frac{\partial g_i^0}{\partial \omega_j}\right)_{i,j=1,\ldots,m},$$

where F_i^l are given in $(6.11)_i$. In the formula (6.13), the numerator $q_j^l(v, \zeta)$ is divisible by $p(v, \zeta)$ and it defines a homogeneous polynomial in ζ with coefficients holomorphic in v.

Since the assumption of this proposition is satisfied for $d = 1$ by Proposition 6.1, we can take $d = 2^\nu$ $(\nu = 0, 1, \ldots)$ successively. Let φ_ν be the transformation described in this proposition with $d = 2^\nu$. Then it turns out that there exists a formal transformation $\varphi = \lim_{\nu \to \infty} (\varphi_0 \circ \varphi_1 \circ \ldots \circ \varphi_\nu)$ such that $G \circ \varphi$ is in normal form. By Proposition 6.2, the integrals $G_i \circ \varphi$ $(i = 1, \ldots, m)$ are also in normal form.

Proof of Proposition 6.3 Assume that $G(v, z)$ is holomorphic in a neighbourhood $V \times \Xi$ of $(v_0, 0)$, where V is a neighbourhood of v_0 in $\mathbf{C}^k$ and Ξ a neighbourhood of $z = 0$ in $\mathbf{C}^{2m}$. Let V_{NR} be the set of v for which $z = 0$ is a non-resonant equilibrium point of X_G. By Proposition 6.1, we may assume that the lowest order part $g^0 (= G^0)$ has the form $g^0 = \sum_{j=1}^{m} \lambda_j(v)\, x_j y_j$, where $\lambda_1(v), \ldots, \lambda_m(v)$ are rationally

independent generically with respect to v. In the following, we put $G=G_0$ and $s=s_0$ as before.

By assumption and by Proposition 6.2, the m integrals $G_i (i=1, \ldots, m)$ of X_G are also in normal form up to terms of order s_i+d-1. Our purpose is to find a holomorphic function W of the form (6.9) satisfying $m+1$ equations $(6.10)_0$–$(6.10)_m$, which implies that the $m+1$ functions $G_i (i=0, 1, \ldots, m)$ are in normal form up to terms of order s_i+2d-1. To this end, let us consider the system of equations $(6.11)_0$–$(6.11)_m$ for $W^{l+2}(v, \zeta)$, where condition (6.12) holds under the assumption. First we fix $v \in V_{NR}$ arbitrarily. Then, as solutions of equation $(6.11)_0$ satisfying $P_N W^{l+2} \equiv 0$, we can determine uniquely the homogeneous polynomials $W^{l+2}(v, \zeta)$ in ζ successively for $l=d, \ldots, 2d-1$. This is easily done by comparison of coefficients and the non-resonance condition is crucial for the argument (e.g. see [7, Prop. 3.3]). This implies that $G_0(v, \varphi(v, \zeta))$ $(\varphi = \exp X_W)$ is in normal form up to terms of order s_0+2d-1. Then by the remark to Proposition 6.2, the functions $G_i(v, \varphi(v, \zeta))$ $(i=1, \ldots, m)$ are also in normal form up to terms of order s_i+2d-1. Consequently it follows that for each $l=d, \ldots, 2d-1$ $W^{l+2}(v, \zeta)$ satisfies the system of $m+1$ equations $(6.11)_0$–$(6.11)_m$.

Consider a system of m equations $(6.11)_1$–$(6.11)_m$ as a linear (algebraic) equations for $D_j W^{l+2}$ $(j=1, \ldots, m)$. Then at any point ζ on which $\det(\partial g_i^0/\partial \omega_j) \neq 0$, one can solve this system uniquely as (6.13) with (6.14). Under condition (6.12), the inequality $p(v, \zeta) \neq 0$ is satisfied on an open and dense subset Ξ' of the neighbourhood Ξ of the origin $\zeta=0$. Since $D_j W^{l+2}(v, \zeta)$, $q_j^l(v, \zeta)$ and $p(v, \zeta)$ are holomorphic on the whole neighbourhood Ξ, the relation $D_j W^{l+2}(v, \zeta) \cdot p(v, \zeta) = q_j^l(v, \zeta)$ holds on Ξ and hence $D_j W^{l+2}$, q_j^l and p are actually polynomials in ζ. Therefore the polynomial q_j^l is divisible by p and the expression (6.13) is valid on the whole neighbourhood Ξ with any $v \in V_{NR}$ fixed.

Now we prove that $q_j^l(v, \zeta)$ is divisible by $p(v, \zeta)$ for any $v \in V$ and that the quotient $q_j^l(v, \zeta)/p(v, \zeta)$ defines a homogeneous polynomial in ζ with coefficients holomorphic in v. For this purpose, since $p(v_0, \zeta) \not\equiv 0$, we may assume that

$$p(v_0, 0, \ldots, 0, \zeta_{2m}) \not\equiv 0$$

after a suitable linear change of variables ζ. Let s be the degree of the polynomial $p(v_0, \zeta)$ in ζ and we set $\hat{\zeta}=(\zeta_1, \ldots, \zeta_{2m-1})$. Then, in a neighbourhood of $(v_0, 0)$ it follows from the division theorem that

$$q_j^l(v, \zeta) = p(v, \zeta)\, Q_j^l(v, \zeta) + \sum_{\nu=1}^{s} h_{j\nu}^l(v, \hat{\zeta})\, \zeta_{2m}^{s-\nu},$$

where $Q_j^l(v, \zeta)$ and $h_{j\nu}^l(v, \hat{\zeta})$ are holomorphic functions of v and ζ (or $\hat{\zeta}$). For any $v \in V_{NR}$ fixed, $q_j^l(v, \zeta)$ is divisible by $p(v, \zeta)$ and therefore the holomorphic functions $h_{j\nu}^l(v, \hat{\zeta})$ in the above vanish identically with respect to $\hat{\zeta}$ for any $v \in V_{NR}$. Since V_{NR} is dense in V, this implies that $h_{j\nu}^l(v, \hat{\zeta})$ $(\nu=1, \ldots, s)$ are identically zero. Thus we have proved that $q_j^l(v, \zeta)$ is divisible by $p(v, \zeta)$ and therefore the quotient $q_j^l(v, \zeta)/p(v, \zeta)$ is a holomorphic function of v and ζ in a neighbourhood of $(v_0, 0)$. For $v \in V_{NR}$, it is equal to $D_j W^{l+2}$ obtained by the operation of D_j on the homogeneous polynomial W^{l+2} determined above. Consequently it is a homogeneous polynomial in ζ with coefficients holomorphic in $v \in V$. Therefore W^{l+2} determined above for $v \in V_{NR}$ can be extended to a homogeneous polynomi-

al in ζ with coefficients holomorphic in $v \in V$ so that it satisfies the system of equations $(6.11)_1$–$(6.11)_m$. The polynomial W^{l+2} also satisfies equation $(6.11)_0$ for any $v \in V$. This completes the proof. □

(d) Convergence proof

The aim of this subsection is to prove that the sequence of transformations $\varphi^{(\nu)} := \varphi_0 \circ \ldots \circ \varphi_\nu$ converges uniformly in a neighbourhood of $(v, \zeta) = (v_0, 0)$. The nonresonance condition imposed on the equilibrium point of X_{G_0} is not important for the convergence proof although it was crucial for the construction of the transformation φ defining each iteration step. Actually we will prove the following

Theorem 6.4 *Let $G_i(v, z) \in A_{\mathbf{C}}(v_0, 0)$ $(i = 1, \ldots, m)$ be m functions whose lowest order parts $G_i^0(v, z)$ are in normal form and are functionally independent functions of z for $v = v_0$ fixed. Suppose that there exists a sequence of transformations $\varphi_\nu \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ $(\nu = 0, 1, \ldots)$ satisfying the following two conditions:*

(i) *For any $\nu = 0, 1, \ldots$, $G_i^{(\nu+1)} = G_i^{(\nu)} \circ \varphi_\nu$ $(i = 1, \ldots, m)$ are in normal form up to terms of order $s_i + 2^{\nu+1} - 1$, where $G_i^{(0)} = G_i$ and s_i is the degree of the lowest order part of G_i.*

(ii) *Each transformation φ_ν is defined by $\varphi_\nu = \exp X_W$ with Hamiltonian W of the form (6.9) with $d = 2^\nu$ which satisfies the condition $P_N W \equiv 0$.*

Let $\varphi^{(\nu)} := \varphi_0 \circ \ldots \circ \varphi_\nu$. Then the sequence $\{\varphi^{(\nu)}(v, \zeta)\}$ is convergent uniformly in a neighbourhood of the point $(v_0, 0)$ and therefore $\varphi := \lim_{\nu \to \infty} \varphi^{(\nu)}$ is a holomorphic parametrized canonical transformation such that $G_i \circ \varphi$ $(i = 1, \ldots, m)$ are in normal form.

In the previous paper [7, Theor. 3.6], this theorem is already proved when the parameter v is fixed. Since the uniform convergence with respect to v does not follow from that theorem, we have to prove the above theorem directly. Fortunately the proof of Theorem 3.6 in [7] is valid for the above theorem with slight modifications. In the following, we will show how to modify it for our purpose.

Proof. The proof is based on the so-called rapidly convergent iteration procedure which was developed by the KAM theory (see [3, 10, 13]). Our purpose is to construct a decreasing sequence of domains $\Omega^{(\nu)}$ $(\nu = 0, 1, \ldots)$ containing the point $(v, \zeta) = (v_0, 0)$ such that φ_ν maps $\Omega^{(\nu+1)}$ into $\Omega^{(\nu)}$ and $\Omega^{(\nu)}$ converges to a domain Ω on which the sequence of transformations $\varphi^{(\nu)}$ is uniformly convergent. In the following, let us assume that $v_0 = 0$ by a parallel translation and specify the domains $\Omega^{(\nu)}$ as polydisks around the origin $(v, \zeta) = 0$.

Since φ_ν is defined by the time 1 map $\exp X_W$, the estimates of W and its derivatives are crucial for our purpose. Our estimates of them are based on the expression (6.13). First we will estimate $p(v, \zeta)$ and it will also play a role of specifying the polydisks $\Omega^{(\nu)}$.

Suppose that the functions $G_i(v, \zeta)$ $(i = 1, \ldots, m)$ are holomorphic in a polydisk

$$\{(v, \zeta) \in \mathbf{C}^{k+2m} \mid |v_i| < r_0, |\zeta_j| < r_0 \ (i = 1, \ldots, k; j = 1, \ldots, 2m)\}$$

with some $r_0>0$ and suppose that they are expressed as convergent power series of the form (5.1). Since $p(v,\zeta)$ is a homogeneous polynomial in $\omega_1, \ldots, \omega_m$, it can be written as

$$p(v,\zeta)=\sum_{|\alpha|=s/2} c_\alpha(v)\,\xi^\alpha\eta^\alpha. \tag{6.15}$$

Here s is the degree of $p(v,\zeta)$ as a polynomial in ζ and $c_\alpha(v)$ are holomorphic functions in $v\in V_0$, where

$$V_0=\{v=(v_1,\ldots,v_k)\,|\,|v_i|<r_0, i=1,\ldots,k\}.$$

Let C be a set of multi indices α defined by $C=\{\alpha\,|\,c_\alpha(v)\not\equiv 0\}$. Then we can carry out a linear transformation of variables v so that

$$c_\alpha(0,\ldots,0,v_k)\not\equiv 0 \quad \text{for any } \alpha\in C. \tag{6.16}$$

In the following, we assume that the functions $c_\alpha(v)$ satisfy this condition in the polydisk V_0.

Lemma 6.5 *Under the above assumptions, there exist $k+m(=n)$ constants $0<\gamma_i<r_0 (i=1,\ldots,k)$ and $0<\delta_j<1 (j=1,\ldots,m)$ such that*

$$|p(v,\zeta)|\geqq c_1 r^s \quad \text{on} \quad \Delta_r \quad \text{for any } r>0, \tag{6.17}$$

where

$$\Delta_r=\{(v,\zeta)\in\mathbf{C}^{k+2m}\,|\,|v_i|=\gamma_i, |\xi_j|=|\eta_j|=\delta_j r\ (i=1,\ldots,k; j=1,\ldots,m)\}$$

and c_1 is a constant which is independent of r.

Proof. Since condition (6.16) is satisfied, it follows from the Weierstrass preparation theorem that

$$c_\alpha(v)=a_\alpha(v)\,b_\alpha(v),$$

where $a_\alpha(v)$ is a distinguished pseudo polynomial and $b_\alpha(v)$ is a holomorphic function such that $b_\alpha(0)\neq 0$. The function $a_\alpha(v)$ takes the form

$$a_\alpha(v)=v_k^N+f_1(\hat{v})\,v_k^{N-1}+\ldots+f_{N-1}(\hat{v})\,v_k+f_N(\hat{v}) \tag{6.18}$$

with an integer $N\geqq 0$ and holomorphic functions $f_i(\hat{v})$ in $\hat{v}$ satisfying $f_i(0)=0$. For a set of k constants γ_i such that $0<\gamma_i<r_0 (i=1,\ldots,k)$, let us define a set V_γ by

$$V_\gamma=\{v\in\mathbf{C}^k\,|\,|v_i|=\gamma_i\ (i=1,\ldots,k)\}.$$

Since $f_j(0)=0$ in (6.18), we can estimate $|a_\alpha(v)|$ on V_γ from below by choosing $\gamma_1,\ldots,\gamma_{k-1}$ sufficiently small in comparison with γ_k. This leads to the estimate of $|c_\alpha(v)|$ because $b_\alpha(0)\neq 0$. Namely there exists a constant $c_0>0$ which depends on $\gamma_1,\ldots,\gamma_k$ such that

$$|c_\alpha(v)|\geqq c_0>0 \quad \text{on} \quad V_\gamma \quad \text{for any } \alpha\in C. \tag{6.19}$$

Next we set

$$\xi_j=t^{\rho_j}\hat{\xi}_j, \quad \eta_j=t^{\rho_j}\hat{\eta}_j \quad (j=1,\ldots,m),$$

where $t>0$ and $\rho_j>0$ are constants to be chosen appropriately below. Let us denote the vector with the above components by $\zeta=(t^{\rho}\hat{\xi}, t^{\rho}\hat{\eta})$. Then it follows from (6.15) that

$$p(v,\zeta)=\sum_{|\alpha|=s/2} c_\alpha(v)\, t^{2\langle\alpha,\rho\rangle}\hat{\xi}^\alpha\hat{\eta}^\alpha; \qquad \langle\alpha,\rho\rangle=\sum_{j=1}^{m}\alpha_j\rho_j.$$

We choose $\rho_1, \ldots, \rho_m$ to be rationally independent. Then there exists one and only one $\alpha_*\in C$ such that $\langle\alpha_*,\rho\rangle=\min_{\alpha\in C}\langle\alpha,\rho\rangle$. Setting $d=2\langle\alpha_*,\rho\rangle$, we have

$$p(v,\zeta)=c_{\alpha_*}(v)\, t^d\hat{\xi}^{\alpha_*}\hat{\eta}^{\alpha_*}+t^d\hat{p}(v,\hat{\xi},\hat{\eta},t), \qquad |\hat{p}(v,\hat{\xi},\hat{\eta},t)|\to 0 \quad (t\to 0).$$

Here $\hat{p}(v,\hat{\xi},\hat{\eta},t)$ is a homogeneous polynomial of degree $s/2$ in $\hat{\xi}_1\hat{\eta}_1, \ldots, \hat{\xi}_m\hat{\eta}_m$ whose coefficients tend to 0 as $t\to 0$. Therefore, noting that $c_\alpha(v)$ does not vanish on V_γ by (6.19), there exists a sufficiently small positive number t_* such that

$$|\hat{p}(v,\hat{\xi},\hat{\eta},t_*)|\leqq\tfrac{1}{2}\min_{v\in V_\gamma}|c_{\alpha_*}(v)|\, r^s \quad \text{on} \quad V_\gamma\times\{|\hat{\xi}_j|=|\hat{\eta}_j|=r\,(j=1,\ldots,m)\}.$$

Consequently we have

$$|p(v,\zeta)|\geqq\tfrac{1}{2}\min_{v\in V_\gamma}|c_{\alpha_*}(v)|\, t_*^d\, r^s\geqq\tfrac{1}{2}c_0\, t_*^d\, r^s$$

on $V_\gamma\times\{(\xi,\eta)=(t_*^{\rho}\hat{\xi}, t_*^{\rho}\hat{\eta})\,|\,|\hat{\xi}_j|=|\hat{\eta}_j|=r(j=1,\ldots,m)\}$. Setting $\delta_j=t_*^{\rho_j}$ $(j=1,\ldots,m)$ and $c_1=(1/2)\,c_0\,t_*^d$, we obtain the estimate (6.17). □

For any positive number $r<r_0$, let us define a polydisk Ω_r by

$$\Omega_r=V\times\Xi_r \tag{6.20}$$

with

$$V=\{v\in\mathbf{C}^k\,|\,|v_i|<\gamma_i\ (i=1,\ldots,k)\},$$
$$\Xi_r=\{\zeta\in\mathbf{C}^{2m}\,|\,|\xi_j|<\delta_j r, |\eta_j|<\delta_j r\ (j=1,\ldots,m)\},$$

where γ_i and δ_j are the constants given in Lemma 6.5. Our purpose is to show that there exists a decreasing sequence of positive numbers r_ν $(\nu=0,1,\ldots,)$ such that $r_\nu\to r_0/2$ $(\nu\to\infty)$ and Ω_{r_ν} $(\nu=0,1,\ldots)$ is the desired polydisk $\Omega^{(\nu)}$. Its proof is the same as the corresponding part of the proof of Theorem 3.6 in [7]. The parameter v does not enter into the arguments of [7] and hence the polydisk Ω_r in [7] corresponds to Ξ_r in the above. However all estimates in [7] are valid for the above case if we replace the polydisk Ω_r in [7, p. 429] by (6.20). We give a sketch of the proof below and refer to Sections 4–6 of [7] for details.

Given a polydisk Ω_r, let $A(\Omega_r)$ be the set of all holomorphic functions in v and ζ whose power series expansions are absolutely convergent in a neighbourhood of $\bar{\Omega}_r$. Let us represent $f\in A(\Omega_r)$ in the form

$$f=f_0+f_1+f_2+\ldots,$$

where $f_j = f_j(v, \zeta)$ are homogeneous polynomials of degree j in ζ with coefficients holomorphic in $v \in V$. We define a norm $\| f \|_r$ of $f \in A(\Omega_r)$ by

$$\| f \|_r := \sum_{j=0}^{\infty} |f_j|_r; \quad |f_j|_r = \max_{(v,\zeta)\in\bar{\Omega}_r} |f_j(v, \zeta)|.$$

We note that for any function $f \in A(\Omega_r)$ the maximum $|f|_r$ on $\bar{\Omega}_r$ is attained already on Δ_r which was defined in Lemma 6.5, i.e., $|f|_r = \max_{(v,\zeta)\in\Delta_r} |f|$.

Recall that for the function $W = W^{d+2} + \ldots + W^{2d+1}$ defining φ_ν, the functions $D_j W^{l+2}$ are expressed as (6.13) with (6.14). Then it follows from (6.17) that

$$|D_j W^{l+2}|_r \leqq \max_{(v,\zeta)\in\Delta_r} \left| \frac{q_j^l(v, \zeta)}{p(v, \zeta)} \right| \leqq \frac{|q_j^l(v, \zeta)|_r}{c_1 r^s}.$$

This is the key formula to avoid the small divisor difficulty associated with the iteration procedure. By using this formula, one can estimate $|D_j W^{l+2}|_r$ ($l = d, \ldots, 2d-1$) successively and reach the estimate of $\| D_j W \|_r$ in terms of the norm $\| \hat{G}_i^{(\nu)} \|_r$, where $\hat{G}_i^{(\nu)}$ are defined by (6.8) with G_i replaced by $G_i^{(\nu)}$. One can derive from this the estimate of W and its derivatives. It provides us the estimates of the new error terms $\| \hat{G}_i^{(\nu+1)} \|_{r_{\nu+1}}$ as well as the range of the transformation φ_ν. The final estimate associated with one iteration step is given by Proposition 5.3 in [7]. Moreover the choice of the desired sequence $\{r_\nu\}$ is given in Section 6 of [7]. For the details, we refer to [7, Sect. 4–6]. This completes the proof of Theorem 6.4. □

(e) Uniqueness of normal forms

By the proceeding discussions, we have proved the existence of a transformation $\varphi \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ such that it takes G into normal form and possesses properties (i) and (ii) of Theorem 5.4. To complete the proof of Theorem 5.4, it remains to prove its final assertion (iii).

Let $\varphi' \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ be another transformation which takes G into normal form. Let $\psi = \varphi^{-1} \circ \varphi'$. We note that the matrix $C'(v) := d\varphi'(v, 0)$ also diagonalizes $JG_{zz}(v, 0)$. Here $d\varphi'(v, 0) := \varphi'_z(v, 0)$, i.e., the linearized mapping of $\varphi'(v, \cdot)$ at $z = 0$. Since every eigenspace of $JG_{zz}(v, 0)$ is 1-dimensional, $C^{-1}(v) C'(v)$ is a diagonal matrix, where $C(v) = d\varphi(v, 0)$. Hence it follows from its symplectic character that $d\psi(v, 0)$ is a diagonal matrix of the form (6.6), i.e., in normal form. We will prove more generally that ψ is in normal form. To this end, we note that any transformation $\chi \in \mathfrak{S}_d$ can be written as

$$\chi = (\exp X_W) \circ \chi'; \quad W = W^{d+2} + \ldots + W^{2d+1}, \quad \chi' \in \mathfrak{S}_{2d},$$

where W^j are homogeneous polynomials of degree j in z with coefficients holomorphic in v. For this fact, see [7, Lemma 2.7]. This implies that the transformation ψ can be written as

$$\psi = \psi^{(\nu)} \circ \chi; \quad \psi^{(\nu)} = \psi_0 \circ \psi_1 \circ \ldots \circ \psi_\nu, \quad \chi \in \mathfrak{S}_d \quad (d = 2^\nu; \nu = 0, 1, 2, \ldots)$$

with

$$\begin{cases}\psi_0 = d\psi(v, 0), \\ \psi_\nu = \exp X_W; \quad W = W^{d+2} + \ldots + W^{2d+1} \quad (d = 2^{\nu-1}; \nu = 1, 2, \ldots).\end{cases}$$

Let us denote $G \circ \varphi$ by the same symbol G. Then not only G but also $G \circ \psi$ is in normal form. We prove inductively that each transformation ψ_ν is in normal form. Suppose that $\psi^{(\nu-1)}$ is in normal form. Then $G \circ \psi^{(\nu-1)}$ is in normal form and the transformation ψ_ν takes it into normal form (at least) up to terms of order $2d$ with $d = 2^{\nu-1}$. By the discussion of subsection (c), this implies that for each $l = d, \ldots, 2d-1$ the polynomial W^{l+2} satisfies equation $(6.11)_0$. Here the right-hand side of $(6.11)_0$ is equal to 0 for $l = d$ since $G = G_0$ is already in normal form. Recalling that $\partial g_0^0 / \partial \omega_j = \lambda_j(v)$ are rationally independent generically with respect to v, one can easily see that $D_j W^{d+2} \equiv 0$ $(j = 1, \ldots, m)$, which proves that W^{d+2} is in normal form. Similarly we can prove inductively that W is in normal form and hence ψ_ν is in normal form. This implies that the transformation $\psi = \varphi^{-1} \circ \varphi'$ is in normal form. Thus we have completed the proof of Theorem 5.4. □

(f) Real analytic case (proof of Corollary 5.5)

In the following we will prove Corollary 5.5. First let us consider the linear transformation $z \mapsto C(v)z$ described in Proposition 6.1. We may assume that each component of the symplectic matrix $C(v)$ is a convergent power series of v in a neighbourhood of v_0. Let us represent this matrix in the form

$$C(v) = (c^{(1)}(v), \ldots, c^{(2m)}(v)),$$

where $c^{(j)}(v)$ are $2m$-dimensional column vectors. Recall that $c^{(j)}(v)$ are the eigenvectors of $JG_{zz}(v, 0)$ corresponding to the eigenvalues $\lambda_j(v)$ $(j = 1, \ldots, 2m)$. Here we assume that

$$\lambda_{j+m}(v) = -\lambda_j(v) \quad (j = 1, \ldots, m).$$

Let $\overline{c^{(j)}}(v)$ denote the column vectors obtained from $c^{(j)}(v)$ by taking the complex conjugate of the coefficients of the power series corresponding to each component of $c^{(j)}(v)$. Then we can prove the following

Lemma 6.6 *Assume that $G(v, z) \in A_{\mathbb{R}}(v_0, 0)$ and that the eigenvalues of $JG_{zz}(v, 0)$ are all distinct and purely imaginary. Then the symplectic matrix $C(v)$ described in Proposition* 6.1 *can be chosen so that*

$$c^{(j+m)}(v) = -i\overline{c^{(j)}}(v) \quad (j = 1, \ldots, m). \tag{6.21}$$

Proof. Let $A(v) = JG_{zz}(v, 0)$. By the reality of $A(v)$, one can easily see that $\overline{c^{(j)}}$ are the eigenvectors of $A(v)$ corresponding to the eigenvalues $\lambda_{j+m} = -\lambda_j (= \overline{\lambda_j})$. Since each eigenspace is 1-dimensional, we have relations

$$c^{(j+m)}(v) = \overline{c^{(j)}}(v)\, \rho_j(v) \quad (j = 1, \ldots, m), \tag{6.22}$$

where $\rho_j(v)$ are holomorphic functions of v in a neighbourhood of v_0. On the other hand, the symplectic character of $C(v)$ means the relation ${}^tCJC=J$, which leads to

$$ {}^tc^{(j)} J c^{(j+m)} = 1. \tag{6.23}$$

From (6.22) and (6.23) it follows that

$$ \overline{\rho_j}(v) = -\rho_j(v). $$

This implies that $\rho_j(v)$ are power series of v with purely imaginary coefficients. Here we may assume that $i\rho_j(v)>0$ in a real neighbourhood of v_0. If it is not the case, we change $\lambda_j(v)$ and $\lambda_{j+m}(v)$ with each other, and change $c^{(j)}(v)$ and $c^{(j+m)}(v)$ for $c^{(j+m)}(v)$ and $-c^{(j)}(v)$ respectively. Then we achieve the relation $i\rho_j(v)>0$. Let $P(v)$ be a diagonal matrix of the form

$$ P(v) = \begin{pmatrix} P_1(v) & O \\ O & P_1(v)^{-1} \end{pmatrix} $$

with a diagonal matrix $P_1(v)$. Note that the transformation $z \mapsto P(v)\,z$ preserves the form (6.1) of G. If we set $P_1(v) = \operatorname{diag}(p_1(v), \ldots, p_m(v))$ with $p_j(v) = \sqrt{i\rho_j(v)}$ and replace $C(v)$ by $C(v)\,P(v)$, then we have $\rho_j(v) = -i$ in (6.22) for the the column vectors of $C(v)\,P(v)$. Hence we have the relation (6.21). □

Now let us consider a linear transformation $z^* = C(v)z$, where we use z^* as the original variables. Then z^* is real, namely $z^* \in \mathbf{R}^{2m}$ if and only if z satisfies the condition

$$ \bar{z} = T(v)\,z; \qquad T(v) = \bar{C}^{-1} C, $$

where $\bar{C} := \bar{C}(v) = (\overline{c^{(1)}}(v), \ldots, \overline{c^{(2m)}}(v))$. From Lemma 6.6, one can see that

$$ T(v) = \begin{pmatrix} O & -iI_m \\ -iI_m & O \end{pmatrix} $$

and the reality condition $\bar{z} = T(v)\,z$ is expressed as

$$ y_j = i\overline{x_j} \quad (j=1, \ldots, m). \tag{6.24}$$

Therefore, for the expression (6.1) of $G \in A_{\mathbf{R}}(v_0, 0)$, the coordinate vector $z = (x_1, \ldots, x_m, y_1, \ldots, y_m)$ satisfies condition (6.24) when $v \in \mathbf{R}^k$. By the nonlinear transformation $z = \varphi(v, \zeta)$ described in Proposition 6.3, condition (6.24) is transformed into the same form

$$ \eta_j = i\overline{\xi_j} \quad (j=1, \ldots, m). \tag{6.25}$$

For this fact, we refer to [7, Sect. 7]. Finally, if we carry out a linear canonical transformation

$$ \zeta = M \begin{pmatrix} q \\ p \end{pmatrix}; \qquad M = \frac{1}{\sqrt{2}} \begin{pmatrix} I_m & iI_m \\ iI_m & I_m \end{pmatrix}, \tag{6.26}$$

then condition (6.25) holds if $(q, p) \in \mathbf{R}^{2m}$ and it follows that $\xi_j \eta_j = (q_j^2 + p_j^2)/2$. Consequently, using z and ζ in place of z^* and (q, p) respectively, $z = \varphi(v, M\zeta)$ is a real analytic canonical transformation (with the parameter v) satisfying the assertions of Corollary 5.5. This completes the proof. □

7 Proof of Theorem 5.6 and Corollary 5.7

The proof of Theorem 5.6 is similar to that of Theorem 5.4. We follow the same argument as in the previous section.

Let $f(v, z) \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ be a transformation satisfying the assumptions of Theorem 5.6. First we carry out a parametrized holomorphic linear canonical transformation so that the linear part $A(v)$ of f takes the form (6.6). This is possible because the eigenvalues of $A(v)$ are simple. Therefore we may assume that f is in normal form up to terms of order 1.

Next we describe the structure of the normalization of f and its integrals $G_1, \ldots, G_m$.

Proposition 7.1 *Let $f = f(v, z) \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ and assume that $z = 0$ is a generically non-resonant fixed point of the mapping f. Suppose that f is in normal form up to terms of order d. Then the following holds:*

(i) *If a function $G \in A_{\mathbf{C}}(v_0, 0)$ is an integral of f, then it is in normal form up to terms of order $s + d - 1$ with $s = \deg G^0$.*

(ii) *If a mapping $f' \in \mathrm{Diff}_{\mathbf{C}}(v_0, 0)$ satisfies the relation $f \circ f' = f' \circ f$, then it is in normal form up to terms of order d.*

Remark. The remark to Proposition 6.2 is valid also to this proposition, that is, the above assertions hold also when the parameter v is fixed at any v for which $z = 0$ is a non-resonant fixed point of $f(v, \cdot)$.

Proof. (i) If $d = 1$, the comparison of the lowest order parts in the identity $G(v, f(v, z)) = G(v, z)$ yields (6.5) with (6.6), where $\mu_i(v)$ $(i = 1, \ldots, m)$ satisfy generically the non-resonance condition (2.6) with λ_j replaced by $\mu_j(v)$. Setting $G^0 = \sum_{|\alpha| + |\beta| = s} c_{\alpha\beta}(v) x^\alpha y^\beta$, this leads to

$$\sum_{|\alpha| + |\beta| = s} \gamma(v) c_{\alpha\beta}(v) x^\alpha y^\beta = 0; \qquad \gamma(v) = \prod_{i=1}^{m} \mu_i^{\alpha_i - \beta_i} - 1.$$

By the non-resonance condition, we have $c_{\alpha\beta}(v) \equiv 0$ for $\alpha \neq \beta$ and therefore G^0 is in normal form. In the case when f is in normal form up to terms of order $d \geqq 2$, notice that the products $x_i y_i$ are transformed by f into the form $x_i y_i + O(|z|^{d+2})$. If the integral G is in normal form up to terms of order $s + d - 2$, then comparison of terms of order $s + d - 1$ in the identity $G(v, f(v, z)) = G(v, z)$ gives the relation (6.5) with G^0 replaced by G^{d-1}. In the same way as above, one can easily see that G^{d-1} is in normal form. This completes the proof by induction.

(ii) Let $A'(v) = f_z'(v, 0)$. Then the relation $f \circ f' = f' \circ f$ implies that $A(v) A'(v) = A'(v) A(v)$. Since $A(v)$ is a diagonal matrix of the form (6.6) with distinct eigenvalues, one can easily see that $A'(v)$ is also in the same form as (6.6). Consequently

the assertion holds for the case $d=1$. Suppose that f' is in normal form up to terms of order $d-1$ under the assumption. Then the mappings f and f' are written in the form

$$f=f_d\circ\chi,\quad \chi\in\mathfrak{S}_d \quad\text{and}\quad f'=f'_{d-1}\circ\chi',\quad \chi'\in\mathfrak{S}_{d-1} \tag{7.1}$$

where f_d and f'_{d-1} are in normal form. Here we recall that $\chi'\in\mathfrak{S}_{d-1}$ can be written in the form

$$\chi'(v,z)=z+J\nabla W^{d+1}(v,z)+O(|z|^{d+1}), \tag{7.2}$$

where W^{d+1} is a homogeneous polynomial of degree $d+1$ in z with coefficients holomorphic in v. Let us rewrite $f\circ f'$ and $f'\circ f$ by using these expressions. First we rewrite $f\circ f'$. Notice that $\chi\circ f'_{d-1}=f'_{d-1}\circ\varphi$ for some $\varphi\in\mathfrak{S}_d$ and that $f'_{d-1}\circ f_d=f_d\circ f'_{d-1}$. Then we have

$$\begin{aligned} f\circ f' &= f_d\circ f'_{d-1}\circ\varphi\circ\chi'=f'_{d-1}\circ f_d\circ(z+J\nabla W^{d+1}(v,z)+O(|z|^{d+1})) \\ &= f'_{d-1}\circ(f_d(v,z)+AJ\nabla W^{d+1}(v,z)+O(|z|^{d+1})). \end{aligned} \tag{7.3}$$

On the other hand, $f'\circ f$ can be rewritten as follows:

$$\begin{aligned} f'\circ f &= f'_{d-1}\circ\chi'\circ f_d\circ\chi \\ &= f'_{d-1}\circ(f_d(v,z)+J\nabla W^{d+1}(v,Az)+O(|z|^{d+1})). \end{aligned} \tag{7.4}$$

Therefore the commuting relation $f\circ f'=f'\circ f$ implies that

$$AJ\nabla W^{d+1}(v,z)=JA^{-1}\nabla\hat{W}^{d+1}(v,z),$$

where $\hat{W}^{d+1}(v,z)=W^{d+1}(v,Az)$. Since A is a diagonal matrix of the form (6.6), we have $AJ=JA^{-1}$ and therefore this identity leads to

$$W^{d+1}(v,z)=W^{d+1}(v,Az).$$

This implies that W^{d+1} is in normal form because of the non-resonance condition. Consequently the mapping χ' is in normal form up to terms of order d and we conclude that f' is in normal form up to terms of order d. This completes the proof. □

To prove the existence of a transformation $z=\varphi(v,\zeta)$ which takes f into normal form, we use the same iteration procedure as in the proof of Theorem 5.4. One iteration step is described as follows:

Proposition 7.2 *Under the assumptions of Theorem* 5.6, *suppose that* $f(v,z)$ *is in normal form up to terms of order d. Then there exists a holomorphic function* $W(v,\zeta)$ *of the form* (6.9) *such that, for* $\varphi=\exp X_W$, *the mapping* $\varphi^{-1}\circ f\circ\varphi$ *is in normal form up to terms of order* $2d$. *It is uniquely determined by imposing the condition* $P_N W\equiv 0$. *Each homogeneous polynomial* $W^{l+2}(v,\zeta)$ *satisfies the identity* (6.13) *with* (6.14).

Proof. The proof is based on an inductive argument. Let us define a transformation $\rho_l\in\mathrm{Diff}_{\mathbf{C}}(v_0,0)$ $(l=d,d+1,\ldots,2d)$ by $\rho_d=\mathrm{id}$ and

$$\rho_l=\exp X_{W^{d+2}}\circ\exp X_{W^{d+3}}\circ\ldots\circ\exp X_{W^{l+1}}\qquad (l=d+1,\ldots,2d).$$

Since the mapping $\exp X_{W^{l+1}}$ has the form

$$\exp X_{W^{l+1}} = \zeta + J\nabla W^{l+1}(v,\zeta) + O(|\zeta|^{2d+1}) \quad (l=d+1, \ldots, 2d), \tag{7.5}$$

it follows that

$$\exp X_W = \rho_l \circ \chi \quad \text{with } \chi \in \mathfrak{S}_{2d}$$

for

$$W = W^{d+2} + \ldots + W^{l+1}.$$

Suppose that $W^{d+2}, \ldots, W^{l+1}$ are determined so that the mapping $\rho_l^{-1} \circ f \circ \rho_l$ is in normal form up to terms of order l. In the following, let us rewrite $\rho_l^{-1} \circ f \circ \rho_l$ as f. Our purpose is to determine W^{l+2} so that, for $\varphi = \exp X_{W^{l+2}}$, the mapping $\varphi^{-1} \circ f \circ \varphi$ is in normal form up to terms of order $l+1$. By assumption, we can write f in the form

$$f = f_l \circ \chi, \quad \chi \in \mathfrak{S}_l$$

with a normal form f_l. Recall that $\varphi = \exp X_{W^{l+2}}$ has the form (7.5) with $l+1$ replaced by $l+2$ and note that $\chi \in \mathfrak{S}_l$ can be written as

$$\chi(v,\zeta) = \zeta + J\nabla U^{l+2} + O(|\zeta|^{l+2})$$

with some homogeneous polynomial $U^{l+2}(v,\zeta)$ of degree $l+2$ in ζ. Writing W^{l+2} and U^{l+2} as W and U respectively, we have

$$\begin{aligned}
\varphi^{-1} \circ f \circ \varphi &= \varphi^{-1} \circ f_l \circ \chi \circ \varphi \\
&= (f_l - J\nabla W(v, A\zeta) + O(|\zeta|^{l+2})) \circ (\zeta + J\nabla(U+W)(v,\zeta) + O(|\zeta|^{l+2})) \\
&= f_l \circ (\zeta - A^{-1} J A^{-1} \nabla \hat{W}(v,\zeta) + O(|\zeta|^{l+2})) \circ (\zeta + J\nabla(U+W)(v,\zeta) + O(|\zeta|^{l+2})) \\
&= f_l \circ (\zeta + J\nabla(U+W-\hat{W})(v,\zeta) + O(|\zeta|^{l+2})),
\end{aligned}$$

where $\hat{W}(v,\zeta) = W(v, A\zeta)$ and we used the relation $A^{-1}J = JA$. Hence we have

$$\varphi^{-1} \circ f \circ \varphi = f_{l+1} \circ \chi'; \quad \chi' \in \mathfrak{S}_{l+1}$$

with

$$f_{l+1} = f_l \circ \exp X_K; \quad K = U + W - \hat{W}.$$

If $W (= W^{l+2})$ satisfies the equation

$$P_N U = U + W - \hat{W}, \tag{7.6}$$

then the above relation implies that f_{l+1} is in normal form and hence $\varphi^{-1} \circ f \circ \varphi$ is in normal form up to terms of order $l+1$.

Let V_{NR} be the set of v for which $z=0$ is a non-resonant fixed point of $f(v,z)$. For any $v \in V_{NR}$ fixed, one can determine the coefficients of the homogeneous polynomial $W^{l+2}(v,\zeta)$ in ζ so that the identity (7.6) holds. They are determined uniquely under the condition $P_N W^{l+2} \equiv 0$. This is easily seen from the non-resonance condition (see e.g., [7, Prop. 8.3]). Then it follows from the remark to Proposition 7.1 that the integrals $G_i (i=1, \ldots, m)$ are in normal form up to terms of order $s_i + l$. By the same arguments as in the proof of Proposition 6.3, the polynomial $W^{l+2}(v,\zeta)$ satisfies the system of m equations $(6.11)_1$–$(6.11)_m$. By using the expression (6.13) with (6.14), the $W^{l+2}(v,\zeta)$ can be extended to

a homogeneous polynomial in ζ with coefficients holomorphic in $v \in V$ so that it satisfies the system of m equations $(6.11)_1$–$(6.11)_m$. One can easily see that this solution W^{l+2} satisfies also equation (7.6). This completes the proof. □

Proposition 7.2 establishes the existence of a formal transformation $\varphi = \lim_{\nu \to \infty} \varphi^{(\nu)}$ with $\varphi^{(\nu)} = \varphi_0 \circ \varphi_1 \circ \ldots \circ \varphi_\nu$ such that $\varphi^{-1} \circ f \circ \varphi$ is in normal form, where φ_ν is the transformation φ described in Proposition 7.2 with $d = 2^\nu$. The transformation φ_ν satisfies the assumption of Theorem 6.4 and therefore the convergence of $\{\varphi^{(\nu)}\}$ follows immediately. Moreover we can prove the uniqueness of the normal forms (the final assertion of Theor. 5.6) in the same way as in the part (e) of Section 6. This completes the proof of Theorem 5.6. □

Finally we will prove Corollary 5.7. Let $C(v)$ be the matrix which diagonalizes the linear part $A(v)$ of $f \in \mathrm{Diff}_{\mathbf{R}}(v_0, 0)$. The column vectors $c^{(j)}(v)$ of $C(v)$ are the eigenvectors corresponding to the eigenvalues $\mu_j(v)$, where $\mu_{j+m} = \mu_j^{-1}$. Here we have $|\mu_j| = 1$ by assumption and hence $\mu_j^{-1} = \overline{\mu_j}$. In the same way as the proof of Lemma 6.6, the matrix $C(v)$ can be chosen so that condition (6.21) holds. Hence by the argument of Section 6 (f), we can see that $z = \varphi(v, M\zeta)$ with a matrix M given by (6.26) is a real analytic transformation satisfying the assertions of Corollary 5.7. This completes the proof. □

8 Globalization of the Birkhoff transformation

We are now in the final stage of the proof of Theorem 2.8. We will show that there exists the desired coordinate system (θ, I, ξ, η) defined in a neighbourhood of the invariant k-tori Φ_h. Such global coordinates will be introduced by using the flows of $X_{H_1}, \ldots, X_{H_k}$. Our argument is similar to the proof by Moser [11] of the theorem of Arnol'd and Jost. In what follows, we denote by ϕ_i^t the flow of X_{H_i} and let $\phi^t := \phi_1^{t_1} \circ \ldots \circ \phi_k^{t_k}$ for $t = (t_1, \ldots, t_k) \in \mathbf{R}^k$ as in the previous sections.

First of all, we extend the Birkhoff transformation φ described in Corollaries 5.5 and 5.7 to a local symplectic diffeomorphism in (u, v, x, y)-space.

Theorem 8.1 *Let (M, σ, F) be a real analytic integrable system and $\{\Phi_h\}_V$ the real analytic family of invariant k-tori for the integrable system as given in Theorem 2.6. Assume that this family is non-resonant and elliptic. Then there exists an open and dense subset $\hat{V}$ of V with the following property: For any $h \in \hat{V}$ and any point $p \in \Phi_h$, there exists a local real analytic diffeomorphism*

$$\psi_0 = \psi_0(u, v, x, y): (\mathbf{R}^{2n}, w) \to (M, p);$$
$$w = (0, h, 0, 0), \quad u, v \in \mathbf{R}^k, \quad x, y \in \mathbf{R}^m$$

such that

(i) $$\psi_0^* \sigma = \sum_{j=1}^{k} du_j \wedge dv_j + \sum_{j=1}^{m} dx_j \wedge dy_j,$$

(ii) $$H_i \circ \psi_0 = v_i \quad (i = 1, \ldots, k),$$

(iii) $$G \circ \psi_0 = \tilde{G}(v, \tau) \quad \textit{for any function } G \in \mathfrak{F},$$

(iv) $$\psi_0^{-1} \circ \phi^T \circ \psi_0 = \exp X_{\tilde{H}(v,\tau)} \quad \textit{for any period } T \in P_H(p),$$

where $\tilde{G}$, $\tilde{H}$ are functions in real normal form, i.e., real analytic functions of n variables $v_1, \ldots, v_k, \tau_1, \ldots, \tau_m$ with $\tau_j = (x_j^2 + y_j^2)/2$.

Remark. If the family $\{\Phi_h\}_V$ is not assumed to be elliptic, there exists a holomorphic diffeomorphism $\psi_0:(\mathbf{C}^{2n}, w)\to(M_{\mathbf{C}}, p)$ satisfying the above conditions (i)–(iv) with 'real normal form' replaced by 'normal form', i.e., holomorphic functions of n variables $v_1, \ldots, v_k, \tau_1, \ldots, \tau_m$ with $\tau_j=\xi_j\eta_j$.

Proof. Let $h\in V$ and p an arbitrary point on the orbit Φ_h. Let $\psi=\psi(u, v, x, y)$ be a real analytic symplectic diffeomorphism described in Proposition 4.1 with $p_0=p$ and $h_0=h$ in this case. Recall that there exists an open and dense subset $\hat{V}$ of V such that for any $v_0\in\hat{V}$ the assumptions of the Birkhoff normalization theorem (Corollary 5.7) are satisfied. Here we note that, in view of Remark 2.5, it suffices to consider the case where condition (ii) of Definition 2.7 is satisfied. Let φ be the Birkhoff transformation described in Corollary 5.7. By Lemma 3.1, it can be extended to a canonical transformation of the form (3.6) and (3.7) in (u, v, x, y)-space. Notice that the variables v, x and y are preserved under this extension. Let ψ' denote this transformation. We will prove that the composition $\psi_0=\psi\circ\psi'$ satisfies the above conditions (i)–(iv). It is trivial that conditions (i) and (ii) are satisfied. Since $G\circ\psi$ does not depend on u, condition (iii) is also satisfied.

It remains to prove (iv). For any period $T\in P_H(p)$, the mapping $\psi_0^{-1}\circ\phi^T\circ\psi_0$ can be written as

$$(u', v', x', y')=\psi_0^{-1}\circ\phi^T\circ\psi_0(u, v, x, y) \quad \text{with } v'=v, \qquad (x', y')=\phi_v^T(x, y),$$

where ϕ_v^T is the mapping $\phi_v^T=\pi_v\circ\phi^T\circ(\pi_v')^{-1}:\Sigma_v'\to\Sigma_v$, which is considered in the coordinate system (u, v, x, y) introduced by ψ_0. We note that $X_{H_i}=\partial/\partial u_i$ in a neighbourhood of p. By Corollary 5.7 and the argument in the later part of Section 5, ϕ_v^T for any $T\in P_H(p)$ is in normal form, i.e.,

$$\phi_v^T=\exp X_{H(v,\tau)}$$

with some function $H(v, \tau)$ which is in real normal form. Let us consider the right-hand side as a mapping in (u, v, x, y)-space, namely the time 1 map of the vector field

$$\dot{u}=\frac{\partial H}{\partial v}, \qquad \dot{v}=0, \qquad \dot{x}=\frac{\partial H}{\partial y}, \qquad \dot{y}=-\frac{\partial H}{\partial x}.$$

Let $(\tilde{u}, \tilde{v}, \tilde{x}, \tilde{y})=\psi_0^{-1}\circ\phi^T\circ\psi_0\circ(\exp X_H)^{-1}(u, v, x, y)$. Then it follows that $\tilde{x}=x$, $\tilde{y}=y$ as well as $\tilde{v}=v$. Therefore by means of a generating function, we have

$$\tilde{u}=u+\frac{\partial Q}{\partial v}; \qquad Q=Q(v).$$

Hence the mapping $\psi_0^{-1}\circ\phi^T\circ\psi_0$ can be written as

$$\psi_0^{-1}\circ\phi^T\circ\psi_0=\exp X_{\tilde{H}(v,\tau)}; \qquad \tilde{H}(v, \tau)=H(v, \tau)+Q(v).$$

This completes the proof. □

Let $T_i(i=1, \ldots, k)$ be generators of the period lattice $P_H(p)$. We note that $T_1, \ldots, T_k$ are linearly independent over $\mathbf{R}$. Let $V'\subset\hat{V}$ and Ξ be small neighbourhoods of $v=h$ and $z=0\in\mathbf{R}^{2m}$ respectively. We set $\Sigma:=\psi_0(0, V', \Xi)$. The vector

fields $X_{H_1}, \ldots, X_{H_k}$ are transversal to this cross section Σ. We take a smaller neighbourhood $\Xi' \subset \Xi$ and set $\Sigma' := \psi_0(0, V', \Xi')$. Let $p' := \psi_0(0, v, z) \in \Sigma'$ for $v \in V'$ and $z \in \Xi'$. Then there exists a real analytic function $T^{(i)} = T^{(i)}(v, z) \in \mathbf{R}^k$ such that

$$\phi^{T^{(i)}}(p') \in \Sigma; \qquad T^{(i)}(h, 0) = T_i.$$

By Theorem 8.1, we have

$$\psi_0^{-1} \circ \phi^{T_i} \circ \psi_0 = \exp X_{\tilde{H}_i} \qquad (i = 1, \ldots, k)$$

for some function $\tilde{H}_i = \tilde{H}_i(v, \tau)$ in real normal form. Therefore the u-coordinate vector of $\psi_0^{-1} \circ \phi^{T_i} \circ \psi_0(0, v, z)$ is given by $\partial \tilde{H}_i / \partial v$. Here we note that $\partial \tilde{H}_i / \partial v = 0$ at $(u, v, z) = (0, h, 0)$ because p is a fixed point of ϕ^{T_i}. Since $X_{H_i} = \partial / \partial u_i$ in a neighbourhood of p, $T^{(i)}(v, z)$ is given by

$$T^{(i)}(v, z) = \frac{\partial K_i}{\partial v}; \qquad K_i = \langle T_i, v \rangle - \tilde{H}_i(v, \tau).$$

Let us consider the vector field X_{K_i} as defined on $\mathbf{R}^k \times \Omega$, where $\Omega = V' \times \Xi'$. In the following, we assume that the neighbourhood Ξ' is of the form

$$\Xi' = \{(\xi, \eta) \in \mathbf{R}^{2m} \mid \tau_j = \tfrac{1}{2}(\xi_j^2 + \eta_j^2) < \varepsilon \qquad (j = 1, \ldots, m)\} \tag{8.1}$$

with some small constant $\varepsilon > 0$. The vector field X_{K_i} is written as

$$\dot{u} = \frac{\partial K_i}{\partial v} = T^{(i)}(v, \tau), \qquad \dot{v} = 0 \qquad \dot{x} = -\frac{\partial \tilde{H}_i}{\partial y}, \qquad \dot{y} = \frac{\partial \tilde{H}_i}{\partial x}$$

and is solved explicitly because v_j and τ_j are integrals of the system. We introduce a coordinate vector $I = (I_1, \ldots, I_k)$ by

$$K_j(v, \tau) - K_j(h, 0) = I_j \qquad (j = 1, \ldots, k). \tag{8.2}$$

Notice that $\partial K_i / \partial v = T_i$ at $(u, v, z) = (0, h, 0)$ and hence $\det(\partial K_i / \partial v_j) \neq 0$ there because $T_1, \ldots, T_k$ are linearly independent. Then we can solve equation (8.2) uniquely for $v = v(I, \tau)$ satisfying $v(0, 0) = h$. Let

$$\hat{\phi}^t = \exp\left(\sum_{j=1}^{k} t_j X_{K_j}\right) \qquad \text{for } t = (t_1, \ldots, t_k) \in \mathbf{R}^k.$$

Then we have the following

Lemma 8.2 *Let $\psi_1 : \mathbf{R}^k \times \Gamma \times \Xi' \to \mathbf{R}^k \times V' \times \Xi'$ be a transformation defined by*

$$\psi_1 : (\theta, I, \xi, \eta) \mapsto (u, v, x, y) := \hat{\phi}^{\theta}(0, v(I, \tau), \xi, \eta),$$

where $v(I,\tau)$ is the solution of equation (8.2) *as above and $\Gamma\subset\mathbf{R}^k$ is a suitable neighbourhood of $I=0$. Then ψ_1 is symplectic, i.e.,*

$$\psi_1^*\left(\sum_{j=1}^{k} du_j\wedge dv_j+\sum_{j=1}^{m} dx_j\wedge dy_j\right)=\sum_{j=1}^{k} d\theta_j\wedge dI_j+\sum_{j=1}^{m} d\xi_j\wedge d\eta_j \tag{8.3}$$

and it satisfies $\psi_1(0,0,0,0)=(0,h,0,0)$ and

$$(K_j-K_j(h,0))\circ\psi_1(\theta,I,\xi,\eta)=I_j \qquad (j=1,\ldots,k), \tag{8.4}$$

$$\psi_1(\theta+\mathbf{e}_i,I,\xi,\eta)=(u+T^{(i)},v,x',y'); \qquad (x',y')=\phi_v^{-T_i}(x,y), \tag{8.5}$$

where $\mathbf{e}_i=(0,\ldots,0,\overset{(i)}{1},0,\ldots,0)\in\mathbf{R}^k$ and $(u,v,x,y)=\psi_1(\theta,I,\xi,\eta)$.

Proof. Let us consider a mapping

$$\Psi\colon (t,w)\mapsto(t,\hat{\phi}^t(w)), \qquad w=(u,v,x,y)$$

which is defined on $\mathbf{R}^k\times\Omega$, where $\Omega=\mathbf{R}^k\times V'\times\Xi'$. Let σ and $\tilde{\sigma}$ be 2-forms on $\mathbf{R}^k\times\Omega$ defined by

$$\sigma=\sum_{j=1}^{k} du_j\wedge dv_j+\sum_{j=1}^{m} dx_j\wedge dy_j, \qquad \tilde{\sigma}=\sigma+\sum_{j=1}^{k} dK_j\wedge dt_j.$$

Then the representation matrix of $\tilde{\sigma}$ is

$$A=\begin{pmatrix} O & -K_w \\ {}^tK_w & J \end{pmatrix}$$

and the linearized mapping $d\Psi$ is expressed as

$$d\Psi=\begin{pmatrix} I_k & O \\ J\,{}^tK_w & d\hat{\phi}^t \end{pmatrix}.$$

Hence, since $\{K_i,K_j\}\equiv 0$ for all $i,j=1,\ldots,k$, we can easily see that

$${}^t(d\Psi)\,A(d\Psi)=\begin{pmatrix} O & O \\ O & J \end{pmatrix}.$$

This implies the relation $\Psi^*\tilde{\sigma}=\sigma$. If we set $(u,v,x,y)=\hat{\phi}^t(q,v,\xi,\eta)$, it can be written as

$$\Psi^*\left(\sum_{j=1}^{k} du_j\wedge dv_j+\sum_{j=1}^{m} dx_j\wedge dy_j+\sum_{j=1}^{k} dK_j\wedge dt_j\right)=\sum_{j=1}^{k} dq_j\wedge dv_j+\sum_{j=1}^{m} d\xi_j\wedge d\eta_j.$$

Restricting the mapping Ψ to $\{(q, v, \xi, \eta) | q = 0\}$ and introducing new coordinates θ_j and I_j by $t_j = \theta_j$ and (8.2), it turns out that the transformation ψ_1 satisfies (8.3) and (8.4). Moreover we have $\psi_1(0, 0, 0, 0) = (0, h, 0, 0)$ and

$$\begin{aligned}\psi_1(\theta + \mathbf{e}_i, I, \xi, \eta) &= \hat{\phi}^{\theta + \mathbf{e}_i}(0, v, \xi, \eta) = \hat{\phi}^{\mathbf{e}_i}(u, v, x, y) \\ &= \left(u + \frac{\partial K_i}{\partial v}, v, x', y'\right); \quad (x', y') = \phi_v^{-T_i}(x, y),\end{aligned}$$

which proves (8.5). Since $x_j^2 + y_j^2 = \xi_j^2 + \eta_j^2$ under ψ_1 and Ξ' is of the form (8.1), ψ_1 maps $\mathbf{R}^k \times \Gamma \times \Xi'$ into $\mathbf{R}^k \times V' \times \Xi'$, where $\Gamma \subset \mathbf{R}^k$ is a suitable neighbourhood of $I = 0$. This completes the proof. □

Since $\phi^t \circ \psi_0(0, h, 0, 0)$ is defined for all $t \in \mathbf{R}^k$, it follows that for any compact ball D about 0 in $\mathbf{R}^k$ there exists a neighbourhood $U(D)$ of $(v, z) = (h, 0)$ in $\mathbf{R}^k \times \mathbf{R}^{2m}$ such that $\phi^t(0, v, x, y)$ is defined for all $t \in D$ and for all $(v, x, y) \in U(D)$.

Lemma 8.3 *Let $\psi_2 : D \times U(D) \to M$ be a mapping defined by*

$$\psi_2 : (u, v, x, y) \mapsto \phi^u \circ \psi_0(0, v, x, y).$$

Then it is symplectic, i.e.,

$$\psi_2^* \sigma = \sum_{j=1}^{k} du_j \wedge dv_j + \sum_{j=1}^{m} dx_j \wedge dy_j. \tag{8.6}$$

Proof. Let us take $s \in D$ arbitrarily and let $u \in D$ be a point sufficiently close to s. We note that

$$\psi_2(u, v, x, y) = \phi^s \circ \phi^{u-s} \circ \psi_0(0, v, x, y).$$

Here $\phi^{u-s} \circ \psi_0(0, v, x, y) = \psi_0(u - s, v, x, y)$ since $u - s$ is sufficiently close to the origin. Since ϕ^s is symplectic and s is fixed, the relation (8.6) follows immediately. □

Now observe that

$$\begin{aligned}\psi_2(u + T^{(i)}, v, x, y) &= \phi^{u + T^{(i)}} \circ \psi_0(0, v, x, y) \\ &= \phi^u \circ \psi_0(0, v, \phi_v^{T_i}(x, y)) \\ &= \psi_2(u, v, \phi_v^{T_i}(x, y)).\end{aligned} \tag{8.7}$$

Here we recall that Ξ' is of the form (8.1) and that $\phi_v^{T_i}$ preserves $\tau_1, \ldots, \tau_m$. Since $T^{(1)} \ldots, T^{(k)}$ are linearly independent over $\mathbf{R}$, the relation (8.7) implies that the mapping ψ_2 is well defined on $\mathbf{R}^k \times V' \times \Xi'$ if V' and Ξ' are taken appropriately small. Let us consider the composition $\psi := \psi_2 \circ \psi_1 : \mathbf{R}^k \times \Gamma \times \Xi' \to M$, where we take the neighbourhoods Γ and Ξ' sufficiently small. Then it follows from (8.5) and (8.7) that

$$\begin{aligned}\psi(\theta + \mathbf{e}_i, I, \xi, \eta) &= \psi_2(u + T^{(i)}, v, \phi_v^{-T_i}(x, y)) \\ &= \psi_2(u, v, x, y) \\ &= \psi(\theta, I, \xi, \eta).\end{aligned}$$

Setting $D_1 = \Gamma$ and $D_2 = \Xi'$, one can easily see that this gives rise to a diffeomorphism

$$\psi = \psi_h \colon (\mathbf{R}^k/\mathbf{Z}^k) \times D_1 \times D_2 \to M$$

which induces the coordinate system (θ, I, ξ, η) in a neighbourhood of the orbit Φ_h. It is easily seen that conditions (i) and (ii) of Theorem 2.8 are satisfied for this diffeomorphism ψ_h. Moreover, for any function $G \in \mathfrak{F}$ we have

$$G \circ \psi(\theta, I, \xi, \eta) = G \circ \psi_0(0, v, x, y),$$

where $(u, v, x, y) = \psi_1(\theta, I, \xi, \eta)$. Here τ_j are invariant under ψ_1, i.e., $x_j^2 + y_j^2 = \xi_j^2 + \eta_j^2$ and v can be expressed as a function of I_j and τ_j only. Therefore the assertion (iii) of Theorem 2.8 follows from Theorem 8.1. This completes the proof of Theorem 2.8. □

Acknowledgement. I would like to thank the referees for their valuable comments and suggestions.

References

1. Abraham, R., Marsden, J.E.: Foundations of mechanics (2nd ed.). Reading, Massachusetts: Benjamin/Cummings 1978
2. Arnol'd, V.I.: Mathematical methods of classical mechanics. Grad. Texts Math. 60. Berlin Heidelberg New York: Springer 1978
3. Arnol'd, V.I., Kozlov, V.V., Neishtadt, A.I.: Dynamical systems III, Mathematical aspects of classical and celestial mechanics. Encycl. Math. Sci. 3. Berlin Heidelberg New York: Springer 1988
4. Duistermaat, J.J.: On global action-angle coordinates. Comm. Pure Appl. Math. **38**, 687–706 (1980)
5. Eliasson, H.: Hamiltonian systems with Poisson commuting integrals. Thesis, Univ. of Stockholm (1984)
6. Eliasson, H.: Perturbations of stable invariant tori for Hamiltonian systems. Ann. Sc. Norm. Super. Pisa, Cl. Sci. **15**, 115–148 (1988)
7. Ito, H.: Convergence of Birkhoff normal forms for integrable systems. Comment. Math. Helv. **64**, 412–461 (1989)
8. Jost, R.: Winkel- und Wirkungsvariable für allgemeine mechanische Systeme. Helv. Phys. Acta **41**, 965–968 (1968)
9. Markus, L., Meyer, K.R.: Generic Hamiltonian dynamical systems are neither integrable nor ergodic. Mem. Am. Math. Soc. **144** (1974)
10. Moser, J.: Stable and random motions in dynamical systems, with special emphasis on celestial mechanics. Ann. Math. Stud. **77** (1973)
11. Moser, J.: Drei Kapitel über Integrable Hamiltonsche Systeme. Lect. Note ETH-Zürich (1984)
12. Pöschel, J.: On elliptic lower dimensional tori in Hamiltonian systems. Math. Z. **202**, 559–608 (1989)
13. Salamon, D., Zehnder, E.: KAM theory in configuration space. Comment. Math. Helv. **64**, 84–132 (1989)
14. Siegel, C.L., Moser, J.K.: Lectures on Celestial Mechanics. Berlin Heidelberg New York: Springer 1971
15. Vey, J.: Sur certain systèmes dynamiques séparables. Am. J. Math. **100**, 591–614 (1978)
16. Ziglin, S.L.: Branching of solutions and nonexistence of first integrals in Hamiltonian mechanics I. Funct. Anal. Appl. **16**, 181–189 (1983)

Math. Z. 206, 409–422 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

An upper bound for the first eigenvalue of the Dirac operator on compact spin manifolds

Helga Baum
Mathematisches Institut, Humboldt Universität, O-1086 Berlin, FRG

Received Februar 20, 1990; in final form March 16, 1990

1 Introduction

Let (M^n, g) be a compact Riemannian spin manifold with the Dirac operator D. If (M^n, g) has positive scalar curvature, then D has no kernel and there is a well known *lower* bound for the smallest positive and negative eigenvalue λ_+ and λ_- of D

$$|\lambda_\pm| \geqq \frac{1}{2}\sqrt{\frac{R_0 n}{n-1}},$$

where R_0 denotes the minimum of the scalar curvature (see [5]). C. Vafa and E. Witten proved in [9] the existence of *universal* upper bounds for the eigenvalues of the twisted Dirac operator on compact manifolds (see also [1]). Using the method of Vafa and Witten in the present paper we derive an *upper* bound for the smallest eigenvalue of the Dirac operator D in geometrical terms of (M^n, g).

Let $(M^{2m}; g)$ be an even-dimensional compact manifold. Then the spectrum of the Dirac operator D is symmetric to zero. Let $0 = \lambda_0^2 < \lambda_1^2 < \lambda_2^2 < \ldots$ denote the eigenvalues of D^2 and let m_j be the multiplicity of λ_j^2. In the second part of the paper we prove the basic Theorem

Theorem. *Let $f\colon M^{2m} \to S^{2m}$ be a smooth map of M into the standard sphere with degree*

$$\deg(f) \geqq 2^{m-1}(m_0 + \ldots + m_{k-1}) + 1.$$

Then the k^{th} eigenvalue $\pm\lambda_k$ of D is bounded by

$$|\lambda_k| \leqq 2^{m-1}\sqrt{m} \cdot \|df\|_\infty$$

where

$$\|df\|_\infty := \max_{x \in M} \|df_x\|.$$

In the third part of the paper we consider compact manifolds (M^n, g) of positive scalar curvature. As an application of the basic Theorem we obtain the following estimates for the first eigenvalue λ (λ = eigenvalue of smallest absolut value) of D.

Proposition 1 *Let (M^n, g), $n=2m, 2m-1$, be a compact spin manifold with positive scalar curvature, let K^M be the sectional curvature and r_{inj}^M be the injectivity radius of (M^n, g). Then the first eigenvalue λ of the Dirac operator is bounded by*

$$|\lambda| \leqq 2^{m-1}\sqrt{m}\cdot \min_{p\in M}\sqrt{K(p)} \leqq 2^{m-1}\sqrt{m}\cdot \max\left\{\sqrt{K^M}, \frac{\pi}{r_{\text{inj}}^M}\right\},$$

where

$$K(p) := \max\left\{K^M|_{B(p, r_{\text{inj}}(p))}, \frac{\pi^2}{r_{\text{inj}(p)}^2}\right\}.$$

Corollary 1 *Let (M^{2m}, g) be a compact even-dimensional spin manifold of positive sectional curvature $0<K^M$ and let $K_{\max}$ be the maximum of K^M. Then the first eigenvalue $\pm\lambda_1$ of D is bounded by*

$$(*) \qquad |\lambda_1| \leqq 2^{m-1}\sqrt{m}\cdot\sqrt{K_{\max}}.$$

Corollary 2 *Let (M^{2m-1}, g) be a simply-connected, compact, odd-dimensional spin manifold with sectional curvature $0<\frac{k}{4}<K^M\leqq k$. Then the first eigenvalue λ of D is bounded by*

$$|\lambda| \leqq 2^{m-1}\sqrt{m}\cdot\sqrt{k}.$$

In the fourth part of the paper we consider compact oriented surfaces in $\mathbb{R}^3$ and prove

Proposition 2 *Let $M^2 \subset \mathbb{R}^3$ be a compact oriented surface of genus $g \neq 1$ and let $\mu(x)$ be the principal curvature in $x\in M$ of maximal absolute value. Then the first eigenvalue $\pm\lambda_1$ of the Dirac-operator satisfies*

$$(**) \qquad |\lambda_1| \leqq c(g)\cdot \max_{x\in M}|\mu(x)|,$$

where

$$c(g) = \begin{cases} 1 & \text{if } g=0 \\ 3 & \text{if } g=2, 3 \\ 2 & \text{if } g\geqq 4. \end{cases}$$

For the standard sphere S^2 we have the equality in (*) and (**).

2 The basic theorem

Let (M^{2m}, g) be an even-dimensional compact Riemannian spin manifold with the spinor bundle $\mathcal{S}$ and the Dirac operator $D: \Gamma(\mathcal{S}) \to \Gamma(\mathcal{S})$. Since D exchanges the positive and negative part φ^+ and φ^- of the spinor bundle $\mathcal{S} = \mathcal{S}^+ \oplus \mathcal{S}^-$, the spectrum of D is symmetric to zero. We denote by $0=\lambda_0^2<\lambda_1^2<\lambda_2^2<\ldots$ the eigenvalues of D^2 and by m_j the multiplicity of λ_j^2 ($m_0 \geqq 0$). In this part we will prove the

Theorem. *Let f: $M^{2m} \to S^{2m}$ be a smooth map of M into the standard sphere of degree*

$$\deg(f) \geqq 2^{m-1}(m_0 + \ldots + m_{k-1}) + 1.$$

Then the k^{th} eigenvalue $\pm\lambda_k$ of D is bounded by

$$|\lambda_k| \leqq 2^{m-1}\sqrt{m} \cdot \|df\|_\infty,$$

where

$$\|df\|_\infty := \max_{x \in M} \|df_x\|.$$

The proof of the Theorem based on the comparison of the spectrum of two twisted Dirac operators on M with values in the bundle $f^*\mathcal{S}_0$, where $\mathcal{S}_0$ denotes the spinor bundle of the standard sphere S^{2m}.

First we prove some topological properties of the spinor bundle $\mathcal{S}_0$.

Lemma 1 1) *$\mathcal{S}_0$ is a trivial complex vector bundle of dimension 2^m. The spinor fields $\Gamma(\mathcal{S}_0)$ on S^{2m} can be identified with the space of smooth functions $C^\infty(S^{2m}; \mathbb{C}^{2^m})$ such that the spinor derivative of $u \in \Gamma(\mathcal{S}_0)$ is given by*

$$\nabla_X^{\mathcal{S}_0} u = X(u) + \tfrac{1}{2} i(-1)^m X \cdot (u^+ - u^-)$$

for each vector field X on S^{2m}, where $u = u^+ \oplus u^-$ is the decomposition of u with respect to $\mathcal{S}_0 = \mathcal{S}_0^+ \oplus \mathcal{S}_0^-$ and $\cdot$ the Clifford product in $\mathcal{S}_0$.

2) *The Chern character of $\mathcal{S}_0^+$ and $\mathcal{S}_0^-$ is given by*

$$\mathrm{ch}(\mathcal{S}_0^\pm) = 2^{m-1}\alpha_0 \pm \alpha_{2m},$$

where α_i are the positive generators of $H^i(S^{2m}; \mathbb{Z})$.

Proof. Let $F^{2m} \subset N^{2m+1}$ be an oriented submanifold of codimension one of (N, g) and let ξ be the normal unit vector field to F with respect to the choosen orientation of F and N. Then $(F, g|_F)$ has a spinor structure such that the spinor bundle $\mathcal{S}_F$ of $(F, g|_F)$ is isomorphic to the restriction $\mathcal{S}_N|_F$ of the spinor bundle $\mathcal{S}_N$ of (N, g) to F. By this identification, the spinor derivative of $\varphi \in \Gamma(\mathcal{S}_N)$ for vectors X tangent to F is given by

$$\nabla_X^{\mathcal{S}_N} = \nabla_X^{\mathcal{S}_F}(\varphi|_F) - \tfrac{1}{2} i(-1)^m \nabla_X^N \xi \cdot (\varphi^+ - \varphi^-),$$

where $\varphi|_F = \varphi^+ \oplus \varphi^- \in \mathcal{S}_F^+ \oplus \mathcal{S}_F^-$ and $\cdot$ denotes the Clifford product in $\mathcal{S}_F$ (see [3]).

Now, let us apply this to the standard sphere $S^{2m} \subset \mathbb{R}^{2m+1}$. The spinor bundle of the Euklidean space $\mathbb{R}^{2m+1}$ is the trivial bundle $\mathcal{S} = \mathbb{R}^{2m+1} \times \Delta_{2m+1}$. Hence the spinor bundle $\mathcal{S}_0 = \mathcal{S}|_{S^{2m}}$ of S^{2m} is trivial. The spinor derivative of a spinor field $\varphi \in \Gamma(\mathcal{S}) = C^\infty(\mathbb{R}^{2m+1}, \mathbb{C}^{2^m})$ on $\mathbb{R}^{2m+1}$ is $\nabla_X^{\mathcal{S}} \varphi = X(\varphi)$. For the normal vector field $\xi(x) = x$ of $S^{2m} \subset \mathbb{R}^{2m+1}$ we have

$$\nabla_X^{\mathbb{R}^{2m+1}} \xi = X(\xi) = X \quad \text{if } X \in T_x S^{2m}.$$

Hence for $\varphi \in C^\infty(\mathbb{R}^{2m+1}; \mathbb{C}^{2^m})$ the spinor derivative of $\varphi|_{S^{2m}} \in \Gamma(\mathcal{S}_0)$ is given by

$$\nabla_X^{\mathcal{S}_0}(\varphi|_{S^{2m}}) = X(\varphi) + \tfrac{1}{2} i(-1)^m X \cdot (\varphi^+ - \varphi^-),$$

where $\varphi|_{S^{2m}} = \varphi^+ \oplus \varphi^- \in \mathcal{S}_0^+ \oplus \mathcal{S}_0^-$. This proves the first statement of Lemma 1.

Now, we calculate the Chern character of $\mathcal{S}_0^+$ and $\mathcal{S}_0^-$. Let E^r be a r-dimensional complex vector bundle on M with hermitian covariant derivative ∇^E and curvature form R^E: $\Lambda^2 M \to \mathrm{End}(E)$. Then the Chern character of E is the cohomology class in $H^{**}(M; \mathbb{Z})$ defined by the closed even form

$$(*) \qquad \widehat{\mathrm{ch}}(E) = \mathrm{Tr}_E(e^{-\frac{1}{2\pi i}R^E}) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}\left(\frac{1}{2\pi i}\right)^k \mathrm{Tr}_E((R^E)^k).$$

For the curvature form $R^{\mathcal{S}}$ in the spinor bundle $\mathcal{S}$ of a spin manifold (M^n, g) we have

$$R^{\mathcal{S}} = \sum_{i<j} R^{\mathcal{S}}_{ij}\, \sigma^i \wedge \sigma^j = \tfrac{1}{2} \sum_{\substack{i<j \\ k<l}} R^M_{ijkl}\, s_k \cdot s_l\, \sigma^i \wedge \sigma^j,$$

where $(s_1, \ldots, s_n)$ is a local ON-basis of (M, g), $(\sigma^1, \ldots, \sigma^n)$ its dual and R^M_{ijkl} are the components of the curvature tensor of M with respect to $(s_1, \ldots, s_n)$ (see [2], Chap. 3.1). Since the components of the curvature tensor of the sphere S^{2m} are

$$R_{ijkl} = \delta_{jk}\delta_{il} - \delta_{jl}\delta_{ik},$$

we obtain for the curvature form in the spinor bundle $\mathcal{S}_0^{\pm}$ of S^{2m}

$$R^{\mathcal{S}_0^{\pm}} = -\tfrac{1}{2} \sum_{i<j} s_i \cdot s_j\, \sigma^i \wedge \sigma^j.$$

Hence the trace of the endomorphism $(R^{\mathcal{S}_0^{\pm}})^k$ for $0 < k \leqq m$ is

$$\mathrm{Tr}((R^{\mathcal{S}_0^{\pm}})^k) = \frac{(-1)^k}{2^k} \sum_{\substack{i_1<j_1 \\ i_k<j_k}} \mathrm{Tr}_{\mathcal{S}_0^{\pm}}(s_{i_1} \cdot s_{j_1} \cdot \ldots \cdot s_{i_k} \cdot s_{j_k})\, \sigma^{i_1} \wedge \sigma^{j_1} \wedge \ldots \wedge \sigma^{i_k} \wedge \sigma^{j_k}$$

$$= \frac{(-1)^k}{2^k} \frac{(2k)!}{2^k} \binom{2m}{2k} \sum_{1 \leqq r_1 < \ldots < r_{2k} \leqq 2m} \mathrm{Tr}_{\mathcal{S}_0^{\pm}}(s_{r_1} \cdot s_{r_2} \cdot \ldots \cdot s_{r_{2k}})\, \sigma^{r_1} \wedge \ldots \wedge \sigma^{r_{2k}}.$$

Using the definition of the Clifford multiplication (see [5]) it is easy to verify that for a positively oriented ON-basis $(s_1, \ldots, s_{2m})$ and $1 \leqq r_1 < \ldots < r_{2k} \leqq 2m$

$$\mathrm{Tr}_{\mathcal{S}_0^{\pm}}(s_{r_1} \cdot \ldots \cdot s_{r_{2k}}) = \begin{cases} 0 & 0 < k < m \\ \pm i^m \cdot 2^{m-1} & k = m. \end{cases}$$

Therefore, we obtain

$$\mathrm{Tr}_{\mathcal{S}_0^{\pm}}((R^{\mathcal{S}_0^{\pm}})^k) = \begin{cases} \dim \mathcal{S}_0^{\pm} = 2^{m-1} & k = 0 \\ 0 & 0 < k < m \\ \pm(-1)^m i^m \dfrac{(2m)!}{2^{m+1}}\, dS^{2m} & k = m. \end{cases}$$

For the Chern character form of $\mathcal{S}_0^+$ and $\mathcal{S}_0^-$ follows

$$\widehat{\mathrm{ch}}(\mathcal{S}_0^{\pm}) = 2^{m-1} \pm \frac{(2m)!}{2 \cdot 4^m \cdot \pi^m \cdot m!}\, dS^{2m}.$$

Since

$$\mathrm{Vol}(S^{2m}) = \int_{S^{2m}} dS^{2m} = \frac{2\pi^m 4^m m!}{(2m)!}$$

and

$$\alpha_{2m} = \left[\frac{dS^{2m}}{\mathrm{Vol}(S^{2m})}\right]$$

we obtain

$$\mathrm{ch}(\mathcal{S}_0^{\pm}) = 2^{m-1}\alpha_0 \pm \alpha_{2m}. \quad \square$$

Let $f\colon M^{2m} \to S^{2m}$ be a smooth map. The spinor derivative $\nabla^{\mathcal{S}_0}$ of $\mathcal{S}_0$ defines a metric covariant derivative ∇^f in the induced bundle $f^*\mathcal{S}_0$ on M by

$$\nabla_X^f (f^* u) := \nabla^{\mathcal{S}_0}_{df(X)} u.$$

Furthermore, let $(a_1, \ldots, a_{2m})$ be a fixed trivialisation of $\mathcal{S}_0$ by constant functions $a_i \in C^\infty(S^{2m}; \mathbb{C}^{2^m})$ and $f^* a_i$ the induced trivialisation in $f^*\mathcal{S}_0$. We define a second metric covariant derivative ∇^0 in $f^*\mathcal{S}_0$ by

$$\nabla^0 (f^* a_i) \equiv 0 \qquad i = 1, \ldots, 2^m.$$

We denote by D_f, $D_0\colon \Gamma(\mathcal{S} \otimes f^*\mathcal{S}_0) \to \Gamma(\mathcal{S} \otimes f^*\mathcal{S}_0)$ the Dirac operators with coefficients in $f^*\mathcal{S}_0$ defined by the covariant derivatives ∇^f and ∇^0, respectively:

$$D_f = \sum_{j=1}^{n} s_j \cdot (\nabla^{\mathcal{S}}_{s_j} \otimes 1 + 1 \otimes \nabla^f_{s_j})$$

$$D_0 = \sum_{j=1}^{n} s_j \cdot (\nabla^{\mathcal{S}}_{s_j} \otimes 1 + 1 \otimes \nabla^0_{s_j}),$$

where $(s_1, \ldots, s_n)$ is a local ON-basis. The twisted Dirac operators D_0 and D_f are essentially selfadjoint. Let us denote by $L_f\colon \Gamma(\mathcal{S} \otimes f^*\mathcal{S}_0) \to \Gamma(\mathcal{S} \otimes f^*\mathcal{S}_0)$ the difference

$$L_f := D_f - D_0.$$

Lemma 2 *The operator L_f is a selfadjoint bundle morphism in the bundle $\mathcal{S} \otimes f^*\mathcal{S}_0$ which satisfies*

$$L_f(\varphi \otimes f^* u) = \tfrac{1}{2} i (-1)^m \sum_{j=1}^{n} s_j \cdot \varphi \otimes df(s_j) \cdot (u^+ - u^-).$$

Proof. Let $u=\sum_{k=1}^{2m}\alpha_k a_k$. Using Lemma 1(1) we obtain

$$\begin{aligned}L_f(\varphi\otimes f^*u)&=(D_f-D_0)(\varphi\otimes f^*u)=\sum_k(D_f-D_0)(\alpha_k\cdot\varphi\otimes f^*a_k)\\&=D(\alpha_k\varphi)\otimes f^*a_k+\alpha_k(s_j\cdot\varphi\otimes\nabla^f_{s_j}(f^*a_k))\\&\quad-D(\alpha_k\varphi)\otimes f^*a_k-\alpha_k(s_j\cdot\varphi\otimes\nabla^0_{s_j}(f^*a_k))\\&=\sum_{k,j}\alpha_k(s_j\cdot\varphi\otimes\nabla^{\mathscr{S}_0}_{df(s_j)}(a_k))\\&=\tfrac{1}{2}i(-1)^m\sum_{j=1}^{n}s_j\cdot\varphi\otimes df(s_j)\cdot(u^+-u^-).\quad\square\end{aligned}$$

If we identify the bundle $\mathscr{S}\otimes f^*\mathscr{S}_0$ with the sum $\mathscr{S}\oplus\ldots\oplus\mathscr{S}$ using the trivialisation $(f^*a_1,\ldots,f^*a_{2m})$ of $f^*\mathscr{S}_0$ we have $D_0=(D,\ldots,D)$. Hence the spectrum of D_0 is the same as the spectrum of D, where each eigenvalue λ_k^2 of D_0^2 occurs with the multiplicity $2^m\cdot m_k$. Now, we estimate the dimension of the kernel of the operator D_f.

Lemma 3 *The dimension of the kernel of the operator D_f satisfies*

$$\dim\operatorname{Ker}D_f\geqq|2^{m-1}\,\hat{\mathscr{A}}(M)+\deg(f)|+|2^{m-1}\,\hat{\mathscr{A}}(M)-\deg(f)|,$$

where $\hat{\mathscr{A}}(M)=\hat{\mathscr{A}}(TM)[M]$ denotes the $\hat{\mathscr{A}}$-genus of M.

Proof. We denote by $D_f^{\pm}$ the parts

$$D_f^{\pm}\colon\ \Gamma(\mathscr{S}\otimes f^*\mathscr{S}_0^{\pm})\to\Gamma(\mathscr{S}\otimes f^*\mathscr{S}_0^{\pm})$$

of D_f. Using the index theorem for $D_f^{\pm}$ and Lemma 1(2) we obtain

$$\begin{aligned}\operatorname{index}(D_f^{\pm})&=\dim(\operatorname{Ker}(D_f^{\pm})^+)-\dim(\operatorname{Ker}(D_f^{\pm})^-)\\&=\int_M\hat{\mathscr{A}}(TM)\,\widehat{\operatorname{ch}}(f^*\mathscr{S}_0^{\pm})\\&=\int_M\hat{\mathscr{A}}(TM)\,f^*(\widehat{\operatorname{ch}}(\mathscr{S}_0^{\pm}))\\&=\int_M\hat{\mathscr{A}}(TM)\{2^{m-1}+f^*\,\widehat{\operatorname{ch}}_{2m}(\mathscr{S}_0^{\pm})\}\\&=2^{m-1}\,\hat{\mathscr{A}}(M)+\deg(f)\cdot\int_{S^{2m}}\widehat{\operatorname{ch}}_{2m}(\mathscr{S}_0^{\pm})\\&=2^{m-1}\,\hat{\mathscr{A}}(M)\pm\deg(f).\end{aligned}$$

Hence

$$\dim\operatorname{Ker}D_f^{\pm}=\dim(\operatorname{Ker}(D_f^{\pm})^+)+\dim(\operatorname{Ker}(D_f^{\pm})^-)\geqq|2^{m-1}\,\hat{\mathscr{A}}(M)\pm\deg(f)|$$

and

$$\begin{aligned}\dim\operatorname{Ker}D_f&=\dim\operatorname{Ker}D_f^+\,+\dim\operatorname{Ker}D_f^-\\&\geqq|2^{m-1}\,\hat{\mathscr{A}}(M)+\deg(f)|+|2^{m-1}\,\hat{\mathscr{A}}(M)-\deg(f)|.\quad\square\end{aligned}$$

Lemma 4 *Let f: $M^{2m} \to S^{2m}$ be a smooth map with degree* $\deg(f) \geqq 2^{m-1}(m_0 + \ldots + m_{k-1}) + 1$. *Then the k^{th} eigenvalue $\pm\lambda_k$ of the Dirac operator D of M is bounded by*

$$|\lambda_k| \leqq \|L_f\|.$$

Proof. Since $D_0 = D_f - L_f$, from perturbation theory follows, that in the interval $[-\|L_f\|, \|L_f\|]$ lie at least dim Ker D_f eigenvalues of D_0.

Because of

$$|\hat{\mathscr{A}}(M)| = |\hat{\mathscr{A}}(TM)[M]| = |\text{index } D| = |m_0^+ - m_0^-| \leqq m_0$$

we have by assumption

$$2^{m-1}|\hat{\mathscr{A}}(M)| \leqq 2^{m-1} m_0 < \deg(f).$$

Hence from Lemma 3 follows

$$\begin{aligned} \dim \text{Ker } D_f &\geqq |2^{m-1}\hat{\mathscr{A}}(M) + \deg(f)| + |2^{m-1}\hat{\mathscr{A}}(M) - \deg(f)| \\ &= 2 \cdot \deg(f) \\ &\geqq 2^m(m_0 + \ldots + m_{k-1}) + 2. \end{aligned}$$

Therefore, in the interval $[-\|L_f\|, \|L_f\|]$ lie at least $2^m(m_0 + \ldots + m_{k-1}) + 2$ eigenvalues of D_0. But, as mentioned above, D_0 has the same eigenvalues as D and the number of the eigenvalues $\pm\lambda_j$, $0 \leqq j \leqq k-1$, of D_0 with their multiplicities is just $2^m(m_0 + \ldots + m_{k-1})$.

Hence, the eigenvalues $\pm\lambda_k$ of D_0 lie in the intervall $[-\|L_f\|, \|L_f\|]$, too. This proves the lemma. □

Finally, we calculate the norm of L_f.

Lemma 5 *Let f: $M^{2m} \to S^{2m}$ be a smooth map. Then the operator norm of L_f satisfies*

$$\|L_f\| \leqq 2^{m-1}\sqrt{m}\,\|df\|_\infty.$$

Proof. L_f is a selfadjoint endomorphism in the bundle $\mathscr{S} \otimes f^* \mathscr{S}_0$. For the operator norm of L_f we have

$$\begin{aligned} \|L_f\|^2 &= \sup_{\|\varphi\|=1} \|L_f \varphi\|^2 = \sup_{\|\varphi\|=1} \int_M (L_f(x)\,\varphi(x), L_f(x)\,\varphi(x))\, dM \\ &= \sup_{\|\varphi\|=1} \int_M (L_f^2(x)\,\varphi(x), \varphi(x))\, dM \\ &\leqq \max_{x \in M} \|L_f^2(x)\|. \end{aligned}$$

Since $L_f(x)$ commutes the components of $\mathscr{S} \otimes f^* \mathscr{S}_0 = \mathscr{S}^+ \otimes f^* \mathscr{S}_0 \oplus \mathscr{S}^- \otimes f^* \mathscr{S}_0$, each non-zero eigenvalue of $L_f^2(x)$ occurs twice. Hence

$$\begin{aligned} \|L_f^2(x)\| &= \max\{\mu_i^2 \mid \mu_i \text{ eigenvalue of } L_f(x)\} \\ &\leqq \tfrac{1}{2} \text{Trace } L_f^2(x). \end{aligned}$$

From Lemma 2 follows

$$L_f^2(x)=\tfrac{1}{4}\sum_{j,k} s_j\cdot s_k\otimes df(s_j)\cdot df(s_k).$$

Using

$$\mathrm{Tr}_{\mathscr{S}}(s_j\cdot s_k)=-\delta_{kj}\dim\mathscr{S}$$

we obtain

$$\begin{aligned}\mathrm{Tr}\,L_f^2(x)&=\tfrac{1}{4}\sum_{j,k=1}^{n}\mathrm{Tr}_{\mathscr{S}}(s_k\cdot s_j)\,\mathrm{Tr}_{\mathscr{S}_0}(df(s_k)\cdot df(s_j))\\&=\tfrac{1}{4}\dim\mathscr{S}\cdot\dim\mathscr{S}_0\cdot\sum_{j=1}^{n}\|df(s_j)\|^2\\&\leqq 4^{m-1}\cdot 2m\cdot\|df_x\|^2.\end{aligned}$$

This provides

$$\|L_f\|\leqq 2^{m-1}\sqrt{m}\cdot\|df\|_\infty.\quad\square$$

From Lemma 4 and 5 the Theorem follows.

3 An upper bound for the first eigenvalue of the Dirac operator on manifolds of positive scalar curvature

Let (M^n, g) be a compact spin manifold with positive scalar curvature. It is well-known that the Dirac operator of (M^n, g) has no kernel (see [8]) and that the first eigenvalue (=eigenvalue of smallest absolute value) satisfies

$$|\lambda|\geqq\frac{1}{2}\sqrt{\frac{R_0 n}{n-1}},$$

where R_0 is the minimum of the scalar curvature of (M^n, g) (see [5]). We want to prove an *upper* bound for $|\lambda|$. Let $p\in M$ be a fixed point of M. We denote by K^M the sectional curvature of (M^n, g), by $r_{\mathrm{inj}}(p)$ the injectivity radius of $p\in M$ and by $B(p, r)$ the geodesic ball around p with radius r. We will prove

Proposition 1 *Let (M^n, g), $n=2m$, $2m-1$, be a compact spin manifold of positive scalar curvature. Then the first eigenvalue λ of the Dirac operator is bounded by*

$$|\lambda|\leqq 2^{m-1}\sqrt{m}\cdot\min_{p\in M}\sqrt{K(p)},$$

where

$$K(p):=\max\left\{K^M|_{B(p,r_{\mathrm{inj}}(p))},\frac{\pi^2}{r_{\mathrm{inj}}(p)^2}\right\}.$$

Let us denote by $r(p)$ the number $r(p):=\dfrac{\pi}{\sqrt{K(p)}}\leqq r_{\mathrm{inj}}(p)$. We consider the standard sphere $S^n(K(p))$ of constant sectional curvature $K(p)$. We denote the north pole of $S^n(K(p))$ by x_0 and the exponential map in x_0 by $\widetilde{\exp}$. Let I_p: $T_pM\to T_{x_0}S^n(K(p))$ be a fixed isometry.

For a sufficient small real number $\varepsilon>0$ we define a map $f_\varepsilon^p\colon M^n\to S^n(K(p))$ by

$$f_\varepsilon^p(x):=\begin{cases}\widetilde{\exp}\left(I_p\left(\dfrac{\exp_p^{-1}(x)}{\beta_\varepsilon^p(\|\exp_p^{-1}(x)\|)}\right)\right) & x\in B(p,r(p))\subset M\\ -x_0 & \text{otherwise.}\end{cases}$$

Here $\beta_\varepsilon^p\in C^\infty(\mathbb{R})$ denotes an appropriate smoothing function with the following properties

$$\beta_\varepsilon^p(t)=\begin{cases}1 & t\leqq\varepsilon\\ \dfrac{t}{r(p)} & t\geqq r(p)-\varepsilon\\ \in\left[\dfrac{r(p)-\varepsilon}{r(p)}-\varepsilon,1\right] & \text{otherwise}\end{cases}$$

and

$$\left|\frac{d}{dt}\left(\frac{t}{\beta_\varepsilon^p(t)}\right)\right|\leqq 1+\alpha(\varepsilon),$$

$$\left|\frac{\sin\left(\sqrt{K(p)}\,\dfrac{t}{\beta_\varepsilon^p(t)}\right)}{\sin(\sqrt{K(p)}\cdot t)}\right|\leqq 1+\alpha(\varepsilon)$$

for all $t\in(0,r(p))$, where $\alpha(\varepsilon)>0$ is a constant which tends to zero if ε goes to zero.

Lemma 6 *$f_\varepsilon^p\colon M^n\to S^n(K(p))$ is a smooth map of degree one and $\|df_\varepsilon^p\|_\infty\leqq 1+\alpha(\varepsilon)$.*

Proof. The injectivity radius of $S^n(K(p))$ is $\dfrac{\pi}{\sqrt{K(p)}}=r(p)$. Hence, according to the properties of the smoothing function β_ε^p, f_ε^p is smooth and of degree one. It remains to calculate the norm of df_ε^p. df_ε^p maps the radial vectors of $B(p,r(p))$ into radial vectors of the geodesic ball $\tilde{B}(x_0,r(p))$ in $S^n(K(p))$ and vectors, perpendicular to the radial directions in $B(p,r(p))$ into the same kind of vectors in $\tilde{B}(x_0,r(p))$. Hence, to calculate $\|df_\varepsilon^p\|_\infty$ it is enough to estimate $\|df_\varepsilon^p(X)\|$ for radial vectors X and vectors X, perpendicular to the radial directions.

1. Case. Let X be a radial unit vector in T_xM, $x\in B(p,r(p))$. Then there is a geodesic curve $c(t)=\exp_p(t\,v)$, $v\in T_pM$, $\|v\|=1$, such that $c(t_0)=x$ and $c'(t_0)=X$. The curve $f_\varepsilon^p(c(t))$ is given by means of the geodesic $\gamma_{I_p(v)}(s):=\widetilde{\exp}(s\,I_p(v))$ in $S^n(K(p))$ by the formula

$$f_\varepsilon^p(c(t))=\gamma_{I_p(v)}\left(\frac{t}{\beta_\varepsilon^p(t)}\right).$$

Therefore, we obtain

$$\|df_\varepsilon^p(X)\|=\left\|\frac{d}{dt}\,(f_\varepsilon^p(c(t))|_{t=t_0}\right\|=\left|\left(\frac{t}{\beta_\varepsilon^p(t)}\right)'\Big|_{t=t_0}\right|\leqq 1+\alpha(\varepsilon).$$

2. Case. Let $X \in T_x B(p, r(p))$ be a unit vector, perpendicular to the radial directions of $B(p, r(p))$. We choose a curve $c(t)$ in the geodesic sphere of radius $d := \operatorname{dist}(p, x)$ around p such that $c(0) = x$ and $c'(0) = X$. Then

$$f_\varepsilon^p(c(t)) = \widetilde{\exp}\left(\frac{1}{\beta_\varepsilon(d)} \cdot I_p(\exp_p^{-1}(c(t)))\right).$$

Let $v(t) := I_p(\exp_p^{-1}(c(t)) \subset T_{x_0} S^n(K(p))$.

Using the formula for the exponential map of the sphere (see [6], p. 117) we obtain

$$f_\varepsilon^p(c(t)) = x_0 \cdot \cos\left(\sqrt{K(p)}\, \frac{d}{\beta_\varepsilon^p(d)}\right) + \frac{v(t)}{d \cdot \sqrt{K(p)}} \cdot \sin\left(\sqrt{K(p)} \cdot \frac{d}{\beta_\varepsilon^p(d)}\right).$$

Hence

$$df_\varepsilon^p(X) = \frac{d}{dt} f_\varepsilon^p(c(t))|_{t=0} = \frac{v'(0)}{d \cdot \sqrt{K(p)}} \cdot \sin\left(\sqrt{K(p)} \cdot \frac{d}{\beta_\varepsilon^p(d)}\right).$$

On the other hand, we consider the curve

$$\begin{aligned} c_0(t) &:= \widetilde{\exp}(I_p \exp_p^{-1}(c(t)) = \widetilde{\exp}(v(t)) \\ &= x_0 \cdot \cos(\sqrt{K(p)} \cdot d) + \frac{v(t)}{d \cdot \sqrt{K(p)}} \cdot \sin(\sqrt{K(p)} \cdot d). \end{aligned}$$

Applying the Comparison Theorem of Rauch (see [4]) we obtain

$$\|c_0'(0)\| \leqq \|c'(0)\| = \|X\| = 1.$$

Therefore,

$$\begin{aligned} 1 \geqq \|c_0'(0)\| &= \frac{1}{d \cdot \sqrt{K(p)}} \cdot \|v'(0)\| \, |\sin(\sqrt{K(p)} \cdot d)| \\ &= \|df_\varepsilon^p(X)\| \left| \frac{\sin(\sqrt{K(p)} \cdot d)}{\sin\left(\sqrt{K(p)} \cdot \frac{d}{\beta_\varepsilon^p(d)}\right)} \right|. \end{aligned}$$

It follows

$$\|df_\varepsilon^p(X)\| \leqq \left| \frac{\sin\left(\sqrt{K(p)} \cdot \frac{d}{\beta_\varepsilon^p(d)}\right)}{\sin(\sqrt{K(p)} \cdot d)} \right| \leqq 1 + \alpha(\varepsilon).$$

This proves $\|df_\varepsilon^p\|_\infty \leqq 1 + \alpha(\varepsilon)$. □

Now, we consider the map

$$\hat{f}_\varepsilon^p := \sqrt{K(p)} \cdot f_\varepsilon^p \colon M^n \to S^n.$$

$\hat{f}_\varepsilon^p$ is a smooth map of degree one into the standard sphere and satisfies

$$\|d\hat{f}_\varepsilon^p\|_\infty = \sqrt{K(p)} \cdot \|df_\varepsilon^p\|_\infty \leqq \sqrt{K(p)}(1 + \alpha(\varepsilon)).$$

Using these maps $\hat{f}_\varepsilon^p$ in the Theorem and going with ε to zero, we obtain the statement of Proposition 1 for even-dimensional manifolds (M^{2m}, g), $n = 2m$.

Now, let (M^n, g) be of odd dimension $n=2m-1$. We denote by $\tilde{M}$ the Riemannian product $\tilde{M}=M^n\times S^1$ of M and the standard sphere S^1. Let us denote by $\pi\colon \tilde{M}\to M$ the projection onto M and by $\mathcal{S}$ and $\hat{\mathcal{S}}$, respectively, the spinor bundles of M with respect to the two coppies of the spinor representation in odd dimension. $\tilde{M}$ has a spinor structure such that the spinor bundle $\tilde{\mathcal{S}}$ of $\tilde{M}$ can be identified with the sum

$$\tilde{\mathcal{S}}=\pi^*\mathcal{S}\oplus\pi^*\hat{\mathcal{S}}$$

(see [3]). Hence, a spinor field on $\tilde{M}$ is the sum of two S^1-parametric spinor fields on M.

Lemma 7 *The Dirac operator $\tilde{D}$ on $\tilde{M}$ satisfies*

$$\tilde{D}^2(\varphi_1\oplus\hat{\varphi}_2)=(D^2\varphi_1-\varphi_1'')\oplus(\widehat{D^2\varphi_2}-\hat{\varphi}_2'').$$

Proof. If we identify $\tilde{\mathcal{S}}$ with $\pi^*\mathcal{S}\oplus\pi^*\hat{\mathcal{S}}$ we obtain for the Clifford product and the spinor derivative in $\tilde{\mathcal{S}}$

$$\begin{aligned}
&\nabla^{\tilde{\mathcal{S}}}_X(\varphi_1\oplus\hat{\varphi}_2)=\nabla^{\mathcal{S}}_X\mathcal{S}_1\oplus\widehat{\nabla^{\mathcal{S}}_X\mathcal{S}_2}\\
&X\cdot(\varphi_1\oplus\hat{\varphi}_2)=X\cdot\varphi_1\oplus-\cdot\widehat{X\cdot\varphi_2}\qquad\text{for } X\in TM\\
&\nabla^{\tilde{\mathcal{S}}}_{\frac{\partial}{\partial t}}(\varphi_1\oplus\hat{\varphi}_2)=\varphi_1'+\hat{\varphi}_2'\\
&\frac{\partial}{\partial t}\cdot(\varphi_1\oplus\hat{\varphi}_2)=(-1)^{m-1}i(\varphi_2\oplus\hat{\varphi}_1)
\end{aligned}$$

(see [3]). Therefore,

$$\tilde{D}(\varphi_1\oplus\hat{\varphi}_2)=(D\varphi_1+(-1)^{m-1}i\varphi_2')\oplus(-\widehat{D\varphi_2}+(-1)^{m-1}i\varphi_1')$$

and

$$\tilde{D}^2(\varphi_1\oplus\hat{\varphi}_2)=(D^2\varphi_1-\varphi_1'')\oplus(\widehat{D^2\varphi_2}-\hat{\varphi}_2''). \quad\square$$

According to Lemma 7, the spectrum of $\tilde{D}^2$ is the set

$$\{\mu^2+k^2\,|\,\mu\in\mathrm{spec}(D),\ k\in\mathbb{N}_0\},$$

where the multiplicity of μ^2+k^2 is twice the multiplicity of μ in $\mathrm{spec}(D)$. (If $\varphi_\mu\in\Gamma(\mathcal{S})$ is an eigenspinor to $\mu\in\mathrm{spec}(D)$, $\varphi=\varphi_\mu e^{\mathrm{kit}}\in\Gamma(\pi^*\mathcal{S})$ and $\hat{\varphi}=\hat{\varphi}_\mu e^{\mathrm{kit}}\in\Gamma(\pi^*\hat{\mathcal{S}})$ are eigenspinors of $\tilde{D}^2$ to the eigenvalue μ^2+k^2.) In particular, the first eigenvalue of D and $\tilde{D}$ coincide. Now, let be $q\in S^{2m-1}\times S^1$. The injectivity radius of q in $S^{2m-1}\times S^1$ is π and the sectional curvature of $S^{2m-1}\times S^1$ is smaller or equal one. We consider the map

$$h^q_\varepsilon\colon S^{2m-1}\times S^1\to S^{2m},$$

defined in analogy to the map f^p_ε above:

$$h^q_\varepsilon(x)=\begin{cases}\widetilde{\exp}\left(I_q\left(\dfrac{\exp_q^{-1}(x)}{\beta^q_\varepsilon(\exp_q^{-1}(x))}\right)\right) & x\in B(q,\pi)\\ \text{south pole} & \text{otherwise.}\end{cases}$$

According to Lemma 6, h_ε^q is smooth, of degree one and satisfies

$$\|dh_\varepsilon^q\|_\infty \leqq 1+\hat{\alpha}(\varepsilon),$$

where $\hat{\alpha}(\varepsilon)$ tends to zero if ε goes to zero.

Then the map

$$F_\varepsilon := h_\varepsilon^q \circ (\hat{f}_\varepsilon^p, id): \tilde{M} = M^{2m-1} \times S^1 \to S^{2m-1} \times S^1 \to S^{2m}$$

is a smooth map of degree one and

$$\|dF_\varepsilon\|_\infty \leqq \sqrt{K(p)}(1+\alpha(\varepsilon))(1+\hat{\alpha}(\varepsilon)).$$

Using these maps F_ε in the Theorem and going with ε to zero, we obtain the statement of Proposition 1 for odd-dimensional manifolds (M^{2m-1}, g). □

Corollary 1 *Let (M^{2m}, g) be a compact, even-dimensional spin manifold of positive sectional curvature $0<K^M$ and let $K_{\max}$ be the maximum of the sectional curvature. Then the first eigenvalue λ of the Dirac operator satisfies*

$$(*) \qquad |\lambda| \leqq 2^{m-1}\sqrt{m}\cdot\sqrt{K_{\max}}.$$

Proof. If $K^M>0$, then the scalar curvature of (M, g) is positive. For an even-dimensional, oriented manifold of positive sectional curvature holds by the Theorem of Klingenberg ([6], p. 227)

$$r_{\text{inj}}(p) \geqq \frac{\pi}{\sqrt{K_{\max}}}.$$

Hence $K(p) \leqq K_{\max}$ for all $p \in M$ and the statement follows from Proposition 1. □

Remark. The first eigenvalue of the 2-dimensional standard sphere $S^2(K)$ of constant sectional curvature K is $\lambda_1 = \sqrt{K}$ (see [5]). Hence, in this case we have the equality in formula (*).

Corollary 2 *Let (M^{2m-1}, g) be a simply-connected, compact, odd-dimensional spin manifold with sectional curvature $0<\frac{k}{4}<K^M \leqq k$. Then the first eigenvalue λ of D satisfies*

$$|\lambda| \leqq 2^{m-1}\sqrt{m}\cdot\sqrt{k}.$$

Proof. Under the assumption, by the Theorem of Klingenberg ([6], p. 254) follows $r_{\text{inj}}(p) \geqq \frac{\pi}{\sqrt{k}}$.

Hence $K(p) \leqq k$ for all $p \in M$. □

4 An upper bound for the first eigenvalue of the Dirac operator on compact hypersurfaces in $\mathbb{R}^3$

Proposition 2 *Let $M^2 \subset \mathbb{R}^3$ be an oriented compact surface of genus $g \neq 1$ with induced metric and let $\mu(x)$ be the principal curvature in $x \in M$ with the maximal absolute value. Then the first eigenvalue λ_1 of the Dirac operator is bounded by*

$$\|\lambda_1\| \leqq c(g) \cdot \max_{x \in M} |\mu(x)|$$

where

$$c(g) = \begin{cases} 1 & \text{if } g = 0 \\ 3 & \text{if } g = 2, 3 \\ 2 & \text{if } g \geqq 4. \end{cases}$$

Proof. For an oriented compact surface of genus g N. Hitchin proved in [7], that the dimension m_0 of the kernel of the Dirac operator is bounded by

$$m_0 \leqq 2\left[\frac{g+1}{2}\right].$$

Let us denote by $\mathrm{e}: M^2 \to S^2$ the Gauß map of M^2. Then from the Gauß-Bonet-Theorem

$$4\pi(1-g) = \int_M K\, dM = \int_M \mathrm{e}^*\, dS^2 = \deg(\mathrm{e})\, \mathrm{Vol}(S^2)$$

it follows $\deg(\mathrm{e}) = 1 - g$.

Let us first consider the case $g = 0$. Then $m_0 = 0$, $\deg(\mathrm{e}) = 1$ and $\|d\mathrm{e}_x\| = |\mu(x)|$. Using the Gauß map e in the Theorem, we obtain the statement of Proposition 2 for $g = 0$.

We fix a number $k \in \mathbb{N}$ and consider the points

$$p_j = e_3 \cos\left(\frac{2\pi j}{k}\right) + e_1 \cdot \sin\left(\frac{2\pi j}{k}\right) \in S^2 \subset \mathbb{R}^3 \qquad 0 \leqq j < k.$$

For a small real number $\varepsilon > 0$ we consider (as in Chap. 3) a smooth map $f_{k,\varepsilon}$: $S^2 \to S^2(k^2)$ of S^2 into the sphere of constant sectional curvature k^2

$$f_{k,\varepsilon}(x) := \begin{cases} \widetilde{\exp}\left(I_{p_j}\left(\dfrac{\exp_{p_j}^{-1}(x)}{\beta_\varepsilon(\|\exp_{p_j}^{-1}(x)\|)}\right)\right) & x \in B\left(p_j, \dfrac{\pi}{k}\right) \\ & 0 \leqq j < k \\ \text{south pole} & \text{otherwise.} \end{cases}$$

Then we have $\deg(f_{k,\varepsilon}) = k$ and $\|df_{k,\varepsilon}\|_\infty \leqq 1 + \alpha(\varepsilon)$ (see Lemma 6). Now, let $\hat{f}_{k,\varepsilon}$: $M^2 \to S^2$ be the map

$$\hat{f}_{k,\varepsilon} := -k f_{k,\varepsilon} \circ \mathrm{e}.$$

We obtain

$$\deg(\hat{f}_{k,\varepsilon}) = \deg(f_{k,\varepsilon})(-\deg(\mathrm{e})) = k \cdot (g-1)$$

and

$$\begin{aligned} \|d\hat{f}_{k,\varepsilon}\|_\infty &\leqq k(1+\alpha(\varepsilon)) \cdot \|d\mathrm{e}\|_\infty \\ &\leqq k(1+\alpha(\varepsilon)) \cdot \max_{x \in M} |\mu(x)|. \end{aligned}$$

Now, let the genus g of M^2 be $g=2, 3$. Then

$$\deg(\hat{f}_{3,\varepsilon})=3(g-1)\geqq 2\left[\frac{g+1}{2}\right]+1\geqq m_0+1$$

and by the Theorem follows the statement of Proposition 2. In case of genus $g>3$ it is

$$\deg(\hat{f}_{2,\varepsilon})=2(g-1)\geqq 2\left[\frac{g+1}{2}\right]+1\geqq m_0+1$$

and we apply the Theorem again. □

References

1. Atiyah, M.F.: Eigenvalues of the Dirac operator. AT Bonn, Preprint 1984
2. Baum, H.: Spin-Strukturen und Dirac-Operatoren über pseudo-Riemannschen Mannigfaltigkeiten. Teubner-Texte Math. 41, Leipzig 1981
3. Baum, H.: Complete Riemannian manifolds with imaginary Killing spinors. Ann. Global Anal. Geom. **7** (3), 205–226 (1989)
4. Cheeger, J., Ebin, D.G.: Comparison Theorems in Riemannian Geometry. North-Holland Math. Libr. 9, 1975
5. Friedrich, Th.: Der erste Eigenwert des Dirac-Operators einer Kompakten, Riemannschen Mannigfaltigkeit nichtnegativer Skalarkrümmung. Math. Nachr. **97**, 117–146 (1980)
6. Gromoll, D., Klingenberg, W., Meyer, W.: Riemannsche Geometrie im Großen. (Lect. Notes Math., vol. 55) Berlin Heidelberg New York: Springer 1968
7. Hitchin, N.: Harmonic spinors. Adv. Math. **14**, 1–55 (1974)
8. Lichnerowicz, A.: Spineurs harmoniques. C.R. Acad. Sci., Paris, Ser. A–B **257**, 7–9 (1963)
9. Vafa, C., Witten, E.: Eigenvalue Inequalities for Fermions in Gauge Theories. Commun. Math. Phys. **95**, 257–276 (1984)

Math. Z. 206, 423–442 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

The Artin algebras associated with differential operators on singular affine curves

Kenneth A. Brown

Department of Mathematics, University Gardens, University of Glasgow, Glasgow G12 8QW, UK

Received September 25, 1989; in final form April 4, 1990

1 Introduction

1.1 It was shown in [13] that the ring $\mathscr{D}(X)$ of differential operators of the irreducible affine curve X over $\mathbb{C}$ is a Noetherian affine domain of Krull dimension 1 with a unique minimal non-zero ideal $J(X)$, so $H(X) := \mathscr{D}(X)/J(X)$ is a finite dimensional $\mathbb{C}$-algebra. Moreover, $H(X) = \sum^{\oplus}(H(X, x))$, the indecomposable algebra $H(X, x)$ being associated to the singularity x of X [13, (4.13)]. The purposes of this paper are to throw light (a) on the relationship between the structure of $H(X, x)$ and the nature of the singularity x, and also (b) to indicate how the structure of $H(X)$ influences that of $\mathscr{D}(X)$.

1.2 For (a), the link is afforded by the maximal unramified extension $\bar{\mathscr{O}}$ of the local ring $\mathscr{O}$ at x in its complete integral closure $\tilde{\mathscr{O}}$, and by the conductor E of $\bar{\mathscr{O}}$ in $\mathscr{O}$ – see (3.1). We shall show that $\mathscr{O}/E$ is a faithful $H(X, x)$-module. More precisely, let $E = E_1' \cap E_2' \cap \ldots \cap E_t'$ be the primary decomposition of E in $\bar{\mathscr{O}}$, and set $E_i = E_i' \cap \mathscr{O}$. *Then*

$$H(X, x) = \{\psi \in \mathrm{End}_{\mathbb{C}}(\mathscr{O}/E) : \psi(E_i/E) \subseteq E_i/E \ \forall i\},$$

(Theor. 4.6).

Notice from this that, writing $D := \sum E_i$, then D/E is the unique maximal proper $H(X, x)$-submodule of $\mathscr{O}/E$, of codimension s, say. This leads in § 5 to *a block upper triangular decomposition of* $H(X, x)$,

$$H(X, x) = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

where the maximal ideal $M = \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix}$ *is the annihilator of* $\mathscr{O}/D$, *so* $C \cong M_s(\mathbb{C})$, *and* $S = \begin{pmatrix} 0 & B \\ 0 & C \end{pmatrix}$, *the right socle of* $H(X, x)$, *is isomorphic as a left ideal to the direct sum of* s *copies of* $\mathscr{O}/E$.

1.3 Let S_0 and M_0 be the inverse images of S and M in $\mathscr{D}(X, x)$. Then S_0 and M_0 can be described in terms of the ring $\mathscr{D}(\mathscr{O}+D\bar{\mathscr{O}})$ of differential operators on $(\mathscr{O}+D\bar{\mathscr{O}})$. We prove in § 6 *that* $\mathscr{D}(\mathscr{O})\subseteq\mathscr{D}(\mathscr{O}+D\bar{\mathscr{O}})$, *that* S_0 *consists of those operators mapping* $(\mathscr{O}+D\bar{\mathscr{O}})$ *into* $\mathscr{O}$, *and that* $\mathscr{D}(\mathscr{O})$ *is the idealizer of* S_0 *in* $\mathscr{D}(\mathscr{O}+D\bar{\mathscr{O}})$.

1.4 To prove the above results we need an improvement of a result of W.C. Brown [2]: we show in § 3 that $\mathscr{D}(\mathscr{O})\subseteq\mathscr{D}(\bar{\mathscr{O}})$, in the notation of (1.2); *a similar result holds globally.*

1.5 The importance of $H(X)$ for the structure of $\mathscr{D}(X)$ is demonstrated in § 7, where we show that $\mathscr{D}(\bar{X})$ *is a perfect left localization of* $\mathscr{D}(X)$, the $H(X)$-modules being the torsion modules under this localization. A consequence of this is *a description of the Grothendieck groups of* $\mathscr{D}(X)$ *in terms of those of* $H(X)$ *and of the ring* $\mathscr{O}(\tilde{X})$ *of regular functions of the normalisation of* X. In particular,

$$G_0(\mathscr{D}(X)) \cong G_0(H(X))\oplus K_0(\mathscr{O}(\tilde{X})).$$

1.6 The results of (1.3) and (1.5) could – and perhaps should – have been expressed in terms of stratification and "recollement" of the derived categories of bounded complexes of modules over $\mathscr{D}(\mathscr{O})$, $\mathscr{D}(\bar{\mathscr{O}})$, $\mathscr{D}(\mathscr{O}+D\bar{\mathscr{O}})$ and other rings of differential operators. Consulting [3], [10] and references therein, the interested reader will have no difficulty in effecting this translation. It seems possible that this is the most convenient formalism through which to approach the problem of understanding the "top left corner" of $H(X, x)$, denoted by A in (1.2). And using this language leads naturally to the question of whether the rings $H(X)$ are quasi-hereditary [10, p. 280]; in fact all examples of $H(X)$ calculated to date are even hereditary.

1.7 In § 8 various families of examples are described illustrating the above results. The conditions under which $H(X, x)$ is simple are also determined. This section can be read independently of the others; it may be helpful to glance through it before reading the rest of the paper. Some open questions and speculations are gathered in § 9.

2 Notation and background

2.1 Let X be an affine curve over $\mathbb{C}$, and let $\tilde{X}$ denote its normalisation. Let $\mathscr{O}(X)$ denote the ring of regular functions on X and $\mathscr{O}(\tilde{X})$ the integral closure of $\mathscr{O}(X)$ in its quotient field K, so $\mathscr{O}(\tilde{X})$ is the ring of regular functions on $\tilde{X}$. We shall reserve the symbols $\mathscr{O}$, $\tilde{\mathscr{O}}$ for the localizations of $\mathscr{O}(X)$ and $\mathscr{O}(\tilde{X})$ at a point x of X. By [13, Prop. 4.2], the ring $\mathscr{D}(X)$ of differential operators on X has a unique minimal non-zero ideal, which we denote by $J(X)$. Set $H(X)=\mathscr{D}(X)/J(X)$, a finite dimensional $\mathbb{C}$-algebra, [13, Prop. 4.2]. The results proved about $H(X)$ in [13] (and in some cases [8]) are as follows:

(a) [13, Theor. 4.13] Let $x\in X$, and let $\mathscr{D}(X, x)$ be the ring of differential operators on the local ring $\mathscr{O}(X, x)$ at x. Then $\mathscr{D}(X, x)$ has a unique minimal non-zero

ideal $J(X, x)$, $\mathscr{D}(X, x)=\mathscr{D}(X)\otimes_{\mathscr{O}(X)}\mathscr{O}(X, x)$, and $J(X, x)=J(X)\otimes_{\mathscr{O}(X)}\mathscr{O}(X, x)$. Set $H(X, x)=\mathscr{D}(X, x)/J(X, x)$. Then

$$H(X)\cong\sum{}^{\oplus}\{H(X, x): x\in\mathrm{Sing}(X)\},$$

where $\mathrm{Sing}(X)$ is the singular locus of X.
(b) [13, Theor. 3.4 and 3.7] Let $\pi: \tilde{X}\to X$ be the projection map. Let $x\in\mathrm{Sing}(X)$; then $H(X, x)=0$ if and only if $\pi^{-1}(x)$ is a single point. Hence $H(X)=0$ if and only if π is injective. In this case, $\mathscr{D}(X)$ is Morita equivalent to $\mathscr{D}(\tilde{X})$.
(c) [13, Theor. 4.2 and 3.6] Let

$$P=\mathscr{D}(\tilde{X}, X)=\{\partial\in\mathscr{D}(\tilde{X}): \partial*\mathscr{O}(\tilde{X})\subseteq\mathscr{O}(X)\},$$

a right ideal of $\mathscr{D}(\tilde{X})$ and a left ideal of $\mathscr{D}(X)$. Then $J(X)=P\mathscr{D}(X)$. If X is unramified at all points of $\mathrm{Sing}(X)$, then $J(X)=P$.
(d) [12, Theor. 5] View $\mathscr{O}(X)$ as a (left) $\mathscr{D}(X)$-module. It has finite length, and has a unique minimal submodule of finite codimension, namely $J(X)*\mathscr{O}(\tilde{X})=J(X)*\mathscr{O}(X)$. This ideal is the socle of $\mathscr{O}(X)$ as a $\mathscr{D}(X)$-module.
(e) [12, Theor. 5] Write W for the finite dimensional $\mathscr{D}(X)$-module $\mathscr{O}(X)/J(X)*(\mathscr{O}(X))$. Then W is a faithful $H(X)$-module.

Let $\mathscr{O}$ denote either a one-dimensional affine domain over $\mathbb{C}$, or the localization of such a ring at a finite collection of maximal ideals. If m is a semimaximal ideal of $\mathscr{O}$, we say that a subring A of the integral closure $\tilde{\mathscr{O}}$ of $\mathscr{O}$ in its field of quotients K is *unramified over* $\mathscr{O}$ *at* m if $\mathscr{O}\subseteq A$ and mA is a semiprime ideal of A; and A is *an unramified extension* of $\mathscr{O}$ if $\mathscr{O}\subseteq A\subseteq\tilde{\mathscr{O}}$ and A is unramified at all the maximal ideals of $\mathscr{O}$.
(f) [13, Lemmas 5.2, 5.3 and 5.4] Let I be the conductor of $\mathscr{O}(\tilde{X})$ in $\mathscr{O}(X)$. Then $I\mathscr{D}(\tilde{X})\subseteq J(X)$. If X is unramified at all points of $\mathrm{Sing}(X)$, then $I\mathscr{D}(\tilde{X})=J(X)$, and so $W=\mathscr{O}/I$.

3 Extension of operators over unramified extensions

Our purpose in this section is to generalise a result of W.C. Brown [2, Theor. 1]. First, though, we need some elementary facts about the "unramified closure".

3.1 Unramified extensions were defined before (2.1, f). Let $\mathscr{O}$, $\tilde{\mathscr{O}}$ and A be as stated there. The following can readily be proved:

(i) A is unramified over $\mathscr{O}$ at $m\Leftrightarrow A_m$ is unramified at $m\mathscr{O}_m\Leftrightarrow A_n$ is unramified at $n\mathscr{O}_n$ for all maximal ideals n of $\mathscr{O}$ containing m.
(ii) If A and B are unramified over $\mathscr{O}$ at m, then $\langle A, B\rangle$ is unramified over $\mathscr{O}$ at m. (For one sees from Nakayama's lemma that $A_m=\sum_{i=1}^{t} e_i\mathscr{O}_m$ where $e_i-e_i^2\in mA_m$. A similar statement applies to B_m. Hence one deduces that $\langle A, B\rangle_m/m\langle A, B\rangle_m$ is spanned by idempotents.)

In view of (ii), and the fact that $\tilde{\mathscr{O}}$ is a finitely generated $\mathscr{O}$-module, it makes sense to define the *maximal unramified extension* $\tilde{\mathscr{O}}^{(m)}$ of $\mathscr{O}$ at m.

Similarly, *the maximal unramified extension* of $\mathscr{O}$ is the join of all unramified extensions of $\mathscr{O}$ – in this paper it will be denoted by $\bar{\mathscr{O}}$. Since the composition of unramified extensions is unramified (use (i)), one sees:

(iii) There are no proper unramified extensions of $\bar{\mathcal{O}}^{(m)}$ at $m\bar{\mathcal{O}}^{(m)}$. There are no proper unramified extensions of $\bar{\mathcal{O}}$.
(iv) $\bar{\mathcal{O}} = \cap\{\bar{\mathcal{O}}^{(m)}: m \text{ a maximal ideal of } \mathcal{O}\}$
$= \cap\{\overline{(\mathcal{O})_m}^{(m)}: m \text{ a maximal ideal of } \mathcal{O}\}$
$= \cap\{(\bar{\mathcal{O}}^{(m)})_m: m \text{ a maximal ideal of } \mathcal{O}\}$,
where the last equality is a consequence of:
(v) $(\bar{\mathcal{O}}^{(m)})_m = \overline{(\mathcal{O}_m)}^{(m)}$ for all maximal ideals of $\mathcal{O}$.
It follows easily from (iii) that:
(vi) The projection $\mathrm{Spec}(\tilde{\mathcal{O}}) \to \mathrm{Spec}(\bar{\mathcal{O}}^{(m)})$ is $1-1$ on the fibre lying over m; the projection $\mathrm{Spec}(\tilde{\mathcal{O}}) \to \mathrm{Spec}(\bar{\mathcal{O}})$ is $1-1$.

All of the above (and the appropriate generalisation to higher dimensions) is presumably well-known; but we have not been able to locate a convenient reference.

3.2 Theorem. *Let X be a reduced affine curve and let x be a point of X. Let B denote the local ring of X at x, and let $\bar{B}$ be the maximal unramified extension of B. Then $\mathcal{D}(B) \subseteq \mathcal{D}(\bar{B})$.*

Proof. Let m be the maximal ideal of B. As in [2], we may replace B and $\bar{B}$ by their m-adic completions. In this case, since idempotents lift over mB, $\bar{B} = V_1 \oplus \ldots \oplus V_t$, where V_i is a local domain with maximal ideal m_i, say; and $\bar{B}$ is reduced, with minimal primes $p_1, \ldots, p_t$, where p_i is the kernel of the natural homomorphism ψ_i from B to V_i, $1 \leqq i \leqq t$.

We claim that each ψ_i is an epimorphism. For this, it is enough, by Nakayama's Lemma, to show that $m_i = \psi_i(m) + \psi_i(m)\, m_i$. But this follows from the fact that $(mB)_{m_i} = (m_i)_{m_i}$, which is a consequence of the unramified nature of the extension $B \subseteq \bar{B}$.

Each differential operator δ on B fixes $p_1, \ldots, p_t$ [2, Lemma]. In view of the previous paragraph, δ induces a differential operator, call it δ_i, on V_i, for $i = 1, \ldots, t$. Setting $\bar{\delta} = \delta_1 + \ldots + \delta_t \in \mathcal{D}(\bar{B})$ and noting that $\bar{\delta}|_B = \delta$, the result follows.

3.3 Applying (3.2) to each point of $\mathcal{O}$, and using the good properties of localization mentioned in (2.1, a), we deduce the

Theorem. $\mathcal{D}(\mathcal{O}) \subseteq \mathcal{D}(\bar{\mathcal{O}})$.

4 $H(X)$ as a ring of vector space endomorphisms

4.1 Continue with the notation introduced in (2.1), and assume further (for convenience) that X has a unique singular point, namely x, with corresponding maximal ideal m of $\mathcal{O}(X)$. Note that $\tilde{\mathcal{O}} = \mathcal{O}(\tilde{X})_m$ is the integral closure of $\mathcal{O} = \mathcal{O}(X)_m$. By (2.1, a), $H(X) \cong H(\mathcal{O})$. In general we shall leave to the reader the routine task of deriving the global versions of our results, for a curve with an arbitrary number of singular points, using the fact that $H(X) = \sum^{\oplus}\{H(X, x): x \in \mathrm{Sing}(X)\}$, (2.1, a).

Factorise the normalisation map as $\tilde{X} \xrightarrow{\theta} \bar{X} \xrightarrow{\mu} X$, where $\pi = \mu \circ \theta$, μ is unramified, and θ is injective. As before, write $\bar{\mathcal{O}}$ for $\mathcal{O}(\bar{X})_m$.

4.2 Retain the notation just introduced, but temporarily weaken the hypothesis that $\mathcal{O}$ is local: Suppose rather that $\mathcal{O}$ is semilocal, with Jacobson radical m. (We do this to permit a proof by induction of the lemma below.)

Lemma. *Suppose that $\mathcal{O} \subsetneq \bar{\mathcal{O}}$, and let E be the conductor of $\bar{\mathcal{O}}$ in $\mathcal{O}$. Then there exists a ring A with $\mathcal{O} \subsetneq A \subseteq \bar{\mathcal{O}}$, such that F/E is a cyclic $\bar{\mathcal{O}}/m\bar{\mathcal{O}}$-module, where F is the conductor of $\bar{\mathcal{O}}$ in A.*

Proof. If $m\bar{\mathcal{O}} \subseteq E$ the lemma is true with $A = \bar{\mathcal{O}}$, so we may assume that

$$m\bar{\mathcal{O}} \nsubseteq E. \tag{1}$$

It is convenient to replace $\mathcal{O}$ and $\bar{\mathcal{O}}$ by their m-adic completions. Since some power of $m\bar{\mathcal{O}}$ is contained in E, say $m^e\bar{\mathcal{O}}$, and the factors of $\mathcal{O}$ and $\bar{\mathcal{O}}$ by this ideal are isomorphic to the corresponding factors of the completed rings, this does not affect the conclusion of the lemma. (Or the reader may prefer simply to convert the argument below to one concerning $\mathcal{O}/m^e\bar{\mathcal{O}} \subseteq \bar{\mathcal{O}}/m^e\bar{\mathcal{O}}$.)

Having completed at m, $\mathcal{O}$ is now a finite subdirect product of local domains, $\mathcal{O}_1, \ldots, \mathcal{O}_s$, say. Clearly

$$\bar{\mathcal{O}} = \sum_{i=1}^{s}{}^{\oplus} \bar{\mathcal{O}}_i.$$

So if rings A_i can be found between $\mathcal{O}_i$ and $\bar{\mathcal{O}}_i$ satisfying the desired property, for each i, then $A := \sum_{i=1}^{s}{}^{\oplus} A_i$ will suffice. Thus we may assume that $\mathcal{O}$ is a complete local ring.

As in (3.2), $\bar{\mathcal{O}} = \sum_{i=1}^{t}{}^{\oplus} \bar{\mathcal{O}} e_i$, where $e_1, \ldots, e_t$ are orthogonal indempotents and $\mathcal{O}$ maps onto $\bar{\mathcal{O}} e_i$ via the maps $\sigma \to \sigma e_i$ ($\sigma \in \mathcal{O}$, $1 \leqq i \leqq t$). Hence $E = \sum^{t\,\oplus} Ee_i$, and clearly, for each j, $Ee_j \subseteq m \cap \bar{\mathcal{O}} e_j$. In fact, this is an equality: for if $x = xe_j \in m \cap \bar{\mathcal{O}} e_j$ and $r = \sum_i r_i e_i \in \bar{\mathcal{O}}$, then the elements r_i can be chosen from $\mathcal{O}$, as noted above, and so

$$xr = xr_j e_j = xr_j \in m \cap \bar{\mathcal{O}} e_j.$$

Thus $m \cap \bar{\mathcal{O}} e_j$ is an ideal of $\bar{\mathcal{O}}$ contained in $\mathcal{O}$, and so contained in $E \cap \bar{\mathcal{O}} e_j = Ee_j$.

We have shown that

$$E = \sum_{i=1}^{t}{}^{\oplus} (m \cap \bar{\mathcal{O}} e_i) = \sum_{i=1}^{t}{}^{\oplus} (\mathcal{O} \cap \bar{\mathcal{O}} e_i); \tag{2}$$

to see the second equality, observe that $\mathcal{O} \cap \bar{\mathcal{O}} e_i$ is a proper ideal of $\mathcal{O}$, and so lies in m. By (1), $E \subsetneq m$, so there exists at least one j for which $Ee_j \subsetneq me_j$. Reorder so that $j = 1$; in view of (2) we have

$$\mathcal{O} \cap \bar{\mathcal{O}} e_1 \subsetneq me_1.$$

Choose $x \in me_1 \setminus (\mathcal{O} \cap \bar{\mathcal{O}} e_1)$ such that $xm \subseteq \mathcal{O} \cap \bar{\mathcal{O}} e_1$. Set $A = \mathcal{O} + \mathbb{C}x \supsetneq \mathcal{O}$, and let F be the conductor of $\bar{\mathcal{O}}$ in A. As in (2),

$$F = \sum_{i=1}^{t}{}^{\oplus} (A \cap \bar{\mathcal{O}} e_i). \tag{3}$$

Since $\dim_{\mathbb{C}}(A/\mathcal{O})=1$,

$$\dim_{\mathbb{C}}(A\cap\bar{\mathcal{O}}e_j/\mathcal{O}\cap\bar{\mathcal{O}}e_j)\leqq 1, \tag{4}$$

for each $j=1, \ldots, t$. In view of (2) and (3), $A\cap\mathcal{O}e_j/\mathcal{O}\cap\bar{\mathcal{O}}e_j$ is the homogeneous component of F/E with annihilator $(\sum_{i\neq j}^{\oplus}\bar{\mathcal{O}}e_i+me_j)$; so the lemma follows from (4).

4.3 By (3.3), $\mathcal{D}(X)\subseteq\mathcal{D}(\bar{X})$. Set $\bar{P}=\mathcal{D}(\bar{X}, X)=\{\partial\in\mathcal{D}(K)\colon \partial(\mathcal{O}(\bar{X}))\subseteq\mathcal{O}(X)\}$, so $\bar{P}$ is a right ideal of $\mathcal{D}(\bar{X})$ and an ideal of $\mathcal{D}(X)$. The next result generalises [13, Lemma 5.3].

Lemma. *Let E be the conductor of $\mathcal{O}(\bar{X})$ in $\mathcal{O}(X)$. Then $\bar{P}=E\mathcal{D}(\bar{X})$.*

Proof. Clearly, $E\mathcal{D}(\bar{X})\subseteq\bar{P}$. Let $m_1, \ldots, m_t$ be the distinct maximal ideals of $\mathcal{O}(\bar{X})$ lying over m, so that $m\mathcal{O}(\bar{X})=\prod_{i=1}^{t}m_i$. For some $j_1, \ldots, j_t$, $\prod_{i=1}^{t}m_i^{j_i}\subseteq E$. The argument is by induction on $\dim_{\mathbb{C}}(\mathcal{O}(\bar{X})/E)$. Suppose first that this is t; then $E=m\mathcal{O}(\bar{X})\subseteq\mathcal{O}(X)$. Since $\mathcal{D}(\bar{X})$ is Morita equivalent to $\mathcal{D}(\tilde{X})$, and $E=\bigcap_{i=1}^{t}m_i$, $\mathcal{D}(\bar{X})/E\mathcal{D}(\bar{X})$ is the direct sum of the t non-isomorphic irreducible right $\mathcal{D}(\bar{X})$-modules $\mathcal{D}(\bar{X})/m_i\mathcal{D}(\bar{X})$, $i=1, \ldots, t$, [13, Prop. 4.15] Hence, if $\bar{P}\supsetneq E\mathcal{D}(\bar{X})$, then $\bar{P}\cap\bar{\mathcal{O}}\supsetneq E$, which is impossible. So the induction begins.

Suppose now that $E\subsetneq m\mathcal{D}(\bar{X})$. Let A be the (global form of the) ring described in (4.2), with conductor F of $\mathcal{O}(\bar{X})$ in A, so that F/E is a cyclic $\mathcal{O}(\bar{X})/m\mathcal{O}(\bar{X})$-module. Now

$$E\mathcal{D}(\bar{X})\subseteq\bar{P}\subseteq\mathcal{D}(\mathcal{O}(\bar{X}), A)=F\mathcal{D}(\bar{X}), \tag{5}$$

where the equality is got by the induction hypothesis. By construction, each homogeneous component of the completely reducible $\mathcal{O}(\bar{X})$-module F/E has dimension 1. Therefore, as in the first paragraph of the proof, $F\mathcal{D}(\bar{X})/E\mathcal{D}(\bar{X})$ is the direct sum of some of the non-isomorphic irreducible modules $\mathcal{D}(\bar{X})/m_i\mathcal{D}(\bar{X})$, $(i=1, \ldots, t)$. It follows from (5) that $\bar{P}$ has the form $G\mathcal{D}(\bar{X})$ for some ideal G of $\mathcal{O}(\bar{X})$ with $E\subseteq G\subseteq F$. Since $G\mathcal{D}(\bar{X})*(\bar{\mathcal{O}})=G\subseteq\mathcal{O}$, G must equal E. This proves the induction step.

4.4 Lemma. (i) $J(X)=\bar{P}=E\mathcal{D}(\bar{X})$.
(ii) $J(X)*(\mathcal{O}(X))=E$, *and* $W=\mathcal{O}(X)/E$.

Proof. Since $J(X)*(\mathcal{O}(\tilde{X}))=J(X)*(\mathcal{O}(X))$ and $W=\mathcal{O}(X)/J(X)*(\mathcal{O}(X))$, by (2.1, d), (ii) will follow from (i). For (i), in view of (4.3), we need only prove that $J(X)=\bar{P}$. Since $\bar{P}$ is an ideal of $\mathcal{D}(X)$, $J(X)\subseteq\bar{P}$ by minimality of $J(X)$. On the other hand,

$$\bar{P}*(\mathcal{O}(\bar{X}))=E\mathcal{D}(\bar{X})*(\mathcal{O}(\bar{X}))=E.$$

Since $\mathcal{D}(\tilde{X}, \bar{X})*(\mathcal{O}(\tilde{X}))=\mathcal{O}(\bar{X})$ by (2.1, d), since $\mathcal{D}(\mathcal{O}(\bar{X}))$ is simple,

$$E=\bar{P}*(\mathcal{D}(\tilde{X}, \bar{X})*(\mathcal{O}(\tilde{X})))\subseteq\mathcal{D}(\tilde{X}, X)*(\mathcal{O}(\tilde{X})).$$

Therefore $\bar{P}$ annihilates $W=\mathcal{O}(X)/\mathcal{D}(\tilde{X}, X)*(\mathcal{O}(\tilde{X}))$. Since W is a faithful $H(X)$-module, by (2.1, e), $\bar{P}\subseteq J(X)$.

4.5 Let the primary decomposition of $\mathcal{O}(\bar{X})/E$ be

$$\mathcal{O}(\bar{X})/E \cong \sum_{i=1}^{t}{}^{\oplus} R_i,$$

so that $R_i \cong \bar{\mathcal{O}}/\bigcap_{e\geq 1}(E+m_i^e)$. If U is a right ideal of a ring R, the *idealizer* $\mathrm{Id}_R(U)$ of U in R is

$$\mathrm{Id}_R(U)=\{r\in R: ru\in U \text{ for all } u\in U\}.$$

Recall that $\mathrm{Id}_R(U)/U \cong \mathrm{End}_R(R/U)$. We can now state the

Theorem. *Retain the notation of* (4.1)–(4.4). *Then*

$$H(X)=\left\{\psi\in \sum_{i=1}^{t}{}^{\oplus} \mathrm{End}_{\mathbb{C}}(R_i): \psi(\mathcal{O}(X)/E)\subseteq \mathcal{O}(X)/E\right\}.$$

Proof. By (3.1), $\mathcal{D}(X)\subseteq\mathcal{D}(\bar{X})$, and so

$$\mathcal{D}(X)=\{\delta\in\mathcal{D}(\bar{X}): \delta(\mathcal{O}(X))\subseteq\mathcal{O}(X)\}. \tag{6}$$

By (6) and (4.3), $\mathcal{D}(X)\subseteq \mathrm{Id}_{\mathcal{D}(\bar{X})}(E\mathcal{D}(\bar{X}))$. Hence, by (4.4),

$$\begin{aligned} H(X)=\mathcal{D}(X)/E\mathcal{D}(\bar{X}) &\subseteq \mathrm{Id}_{\mathcal{D}(\bar{X})}(E\mathcal{D}(\bar{X}))/E\mathcal{D}(\bar{X}) \\ &\cong \mathrm{End}_{\mathcal{D}(\bar{X})}(\mathcal{D}(\bar{X})/E\mathcal{D}(\bar{X})). \end{aligned} \tag{7}$$

Consider the right $\mathcal{D}(\bar{X})$-module $\mathcal{D}(\bar{X})/E\mathcal{D}(\bar{X})$: As in the proof of Lemma (4.3), $\mathcal{D}(\bar{X})/E\mathcal{D}(\bar{X})$ is completely reducible, with distinct irreducible components $\mathcal{D}(\bar{X})/m_i\,\mathcal{D}(\bar{X})$ of multiplicities $\dim_{\mathbb{C}}(R_i)$, for $i=1,\ldots,t$. In particular,

$$\mathrm{End}_{\mathcal{D}(\bar{X})}(\mathcal{D}(\bar{X})/E\mathcal{D}(\bar{X}))\cong \sum_{i=1}^{t}{}^{\oplus} \mathrm{End}_{\mathbb{C}}(R_i). \tag{8}$$

The theorem now follows from (8), (7) and (6).

4.6 To transcribe Theorem (4.5) into a form intrinsic to $\mathcal{O}(X)$, define

$$E_i=\bigcap_{e\geq 1}(E+m_i^e)\cap\mathcal{O}(X)=\bigcap_{e\geq 1}E+(m_i^e\cap\mathcal{O}(X)),$$

for $i=1,\ldots,t$. Thus E_i is an ideal of $\mathcal{O}(X)$, and $\cap E_i=E$. Since $\mathcal{O}(\bar{X})$ is unramified over $\mathcal{O}(X)$, the natural embedding of $\mathcal{O}(X)/E_i$ into R_i is also surjective, (by the same argument as was used in the proof of (3.2) to show that the maps ψ_i are onto). That is,

$$\mathcal{O}(X)/E_i\cong R_i \tag{9}$$

for $i=1,\ldots,t$.

Theorem. *With the above notation,*

$$H(X)\cong\{\psi\in\mathrm{End}_{\mathbb{C}}(\mathcal{O}(X)/E): \psi(E_i)/E\subseteq E_i/E, \forall i\}.$$

Proof. Let S denote the right hand ring in the statement. If $\psi \in S$ then ψ defines an element $\hat{\psi}$ of $\sum_{i=1}^{t}{}^{\oplus} \mathrm{End}_{\mathbb{C}}(\mathcal{O}(X)/E_i)$, and thus $\hat{\psi} \in \sum_{i=1}^{t}{}^{\oplus} \mathrm{End}(R_i)$, by (9). By construction, $\hat{\psi}(\mathcal{O}(X)/E) \subseteq \mathcal{O}(X)/E$. The map $\psi \to \hat{\psi}$ therefore defines an embedding of S in $H(X)$, in view of Theorem (4.5).

Now take $v \in H(X)$, viewing $H(X)$ in terms of the isomorphism (4.5). Thus the restriction v' of v to $\mathcal{O}(X)/E$ is in $\mathrm{End}_{\mathbb{C}}(\mathcal{O}(X)/E)$, and clearly $v'(E_i/E) \subseteq E_i/E$ for all i. That is, $v' \in S$. Since the image of v' under $\hat{}$ is v, $S \cong H(X)$ as claimed.

5 Representing $H(X)$ on its right socle

5.1 Continue with the notation introduced in previous sections. Thus W, the finite dimensional faithful $H(X)$-module $\mathcal{O}/J(X)*(\mathcal{O})$, is $\mathcal{O}/E$ by (4.4, ii). Since the local ring $\mathcal{O}$ is contained in $\mathcal{D}(\mathcal{O})$, and W has a unique maximal proper submodule as an $\mathcal{O}$-module, W has a unique maximal submodule as a $\mathcal{D}(\mathcal{O})$-module. We shall denote this module by V. Observe that, by Theorem (4.6),

$$V = \left(\sum_{i=1}^{t} E_i \right) / E.$$

We denote the ideal $\sum_{i=1}^{t} E_i$ of $\mathcal{O}$ by D, and set

$$M_0 = \mathrm{Ann}_{\mathcal{D}(\mathcal{O})}(\mathcal{O}/D),$$

a maximal ideal of $\mathcal{D}(\mathcal{O})$.

5.2 Let $\mu^{-1}(m) = \{m_1, \ldots, m_t\}$, the maximal ideals of $\bar{\mathcal{O}}$ lying over m. Let $\mathcal{O}_Y$ be any local ring such that $\mathcal{O} \subseteq \mathcal{O}_Y \subseteq \bar{\mathcal{O}}$ and the following conditions are satisfied: (a) The conductor C_Y of $\bar{\mathcal{O}}$ in $\mathcal{O}_Y$ contains D; and (b) $\mathcal{O}_Y \subseteq \mathbb{C} + \prod_{i=1}^{t} m_i$. (Note that $\hat{\mathcal{O}} := \mathbb{C} + \prod_{i=1}^{t} m_i$ is one possible choice for $\mathcal{O}_Y$.) Set $P_Y = \mathcal{D}(Y, X) = \{\partial \in \mathcal{D}(\mathcal{O}_Y) : \partial(\mathcal{O}_Y) \subseteq \mathcal{O}\}$, a right ideal of $\mathcal{D}(Y) := \mathcal{D}(\mathcal{O}_Y)$, and a left ideal of $\mathcal{D}(\mathcal{O})$. Write $\hat{P}$ for $\mathcal{D}(\hat{\mathcal{O}}, \mathcal{O})$ Observe that $\bar{P} \subseteq \hat{P} \subseteq P_Y$ for every choice of $\mathcal{O}_Y$ satisfying the above conditions, since $\mathcal{D}(Y, X) = \mathcal{D}_Y(Y, X)$ as in [13, 2.7].

Lemma. $P_Y D \subseteq P_Y C_Y \subseteq \bar{P}$.

Proof. By (a), $D \subseteq C_Y \subseteq \mathcal{D}(\bar{\mathcal{O}}, \mathcal{O}_Y)$. Thus

$$P_Y D \subseteq P_Y C_Y \subseteq P_Y \mathcal{D}(\bar{\mathcal{O}}, \mathcal{O}_Y) \subseteq \mathcal{D}(\bar{\mathcal{O}}, \mathcal{O}) = \bar{P}.$$

5.3 Lemma. $P_Y M_0 \subseteq P\mathcal{D}(\mathcal{O}) = \bar{P}$.

Proof. By the definition of M_0 and Lemma 5.2,

$$P_Y M_0 * (\mathcal{O}) \subseteq P_Y * (D) \subseteq P_Y D * (\mathcal{O}) \subseteq P * (\mathcal{O}).$$

Since $P*(\mathcal{O})=J(X)*(\mathcal{O})$ by (2.1, c), and $W=\mathcal{O}/J(X)*(\mathcal{O})$ is a faithful module over $H(X)=\mathcal{D}(\mathcal{O})/P\mathcal{D}(\mathcal{O})$ by (2.1, e), this proves the lemma.

5.4 Lemma. $P_Y*(\mathbb{C})=\mathcal{O}$.

Proof. By the definition of P_Y, $P_Y*(\mathbb{C})\subseteq\mathcal{O}$. Recall that $\hat{\mathcal{O}}=\mathbb{C}+\prod_i m_i$. The proofs of [13, 4.7–4.10] (which are claimed to prove that $\mathcal{D}(\hat{\mathcal{O}}, \mathcal{O})*(\hat{\mathcal{O}})=\mathcal{O}$), in fact prove more, namely that $\mathcal{D}(\hat{\mathcal{O}}, \mathcal{O})*(\mathbb{C})=\mathcal{O}$. Since $\mathcal{O}_Y\subseteq\hat{\mathcal{O}}$, P_Y contains $\mathcal{D}(\hat{\mathcal{O}}, \mathcal{O})$ and it follows that $\mathcal{O}\subseteq P_Y*(\mathbb{C})$.

5.5 Set $S=\hat{P}\mathcal{D}(\mathcal{O})/\bar{P}$ and $M=M_0/\bar{P}$, ideals of $H(X)$. In view of (5.2), since $\bar{\mathcal{O}}$ has conductor $m\bar{\mathcal{O}}$ in $\hat{\mathcal{O}}$, $\hat{P}m\subseteq\bar{P}$, so that

$$\hat{P}*(m/E)=\hat{P}m*(\mathcal{O}/E)\subseteq\bar{P}*(\mathcal{O}/E)=0.$$

Thus, by (5.4), there exists $g_0\in\hat{P}$ with $g_0*(m/E)=0$ and $g_0*(1+E)=1+E$. Therefore, setting $g=g_0+\bar{P}$,

$$\hat{P}/\bar{P}=H(X)g\cong W,$$

via the map $hg\to hg*(1)$.

Let G/E be a vector space complement to $V=D/E$ in $W=\mathcal{O}/E$, with basis $\{1=g_1, \ldots, g_s\}$, say. Since $H(X)/M\cong M_s(\mathbb{C})$, for each $i=2, \ldots, s$ there exists $\delta_i\in H(X)$ with $\delta_i(g_i)=1+\alpha_i$, where $\alpha_i\in V$, and $\delta_i(g_j)\in V$ for $j\neq i$. Set $\delta_1=g$, where g is defined above, and define

$$f=1-\sum_{i=1}^{s} g_i\, g\, \delta_i=1-g-\sum_{i=2}^{s} g_i\, g\, \delta_i, \tag{1}$$

yielding an element f of $H(X)$ projecting W onto V, with kernel G/E. Thus, $f=f^2\in M$, and, for $u\in M$ and $w\in W$, $fu*(w)=u*(w)$, so that $fu=u$. That is, $M=fH(X)$. On the other hand

$$1-f=\sum_{i=1}^{s} g_i\, g\, \delta_i\in\hat{P}\mathcal{D}(\mathcal{O})/\bar{P}=S, \tag{2}$$

so that

$$l-\mathrm{Ann}_{H(X)}(M)=H(X)(1-f)\subseteq S.$$

But (5.3) asserts that $SM=0$, so

$$S=l-\mathrm{Ann}_{H(X)}(M)=H(X)(1-f).$$

By construction, $(1-f)H(X)f=0$, so

$$H(X)=\begin{pmatrix} fH(X)f & fH(X)(1-f) \\ 0 & (1-f)H(X)(1-f) \end{pmatrix}.$$

Here, $(1-f)H(X)(1-f)\cong M_s(\mathbb{C})=H(X)/M$. It is easy to deduce from (2) that

$$H(X)(1-f)=H(X)g\oplus\sum_{i=2}^{s\,\oplus} H(X)\, g\, \delta_i,$$

and that each of the s summands on the right is isomorphic to $H(X)g$ and so to W. Moreover it is clear from the upper triangular decomposition of $H(X)$ that $S=H(X)(1-f)$ is the right socle of $H(X)$.

Let Y be any curve satisfying (5.2, a and b). As noted in (5.2), $P_Y\mathscr{D}(\mathscr{O})\supseteq\hat{P}\mathscr{D}(\mathscr{O})$, so $P_Y\mathscr{D}(\mathscr{O})/\bar{P}$ contains the right socle S of $H(X)$. However, $P_Y\mathscr{D}(\mathscr{O})M_0=P_YM_0\subseteq\bar{P}$, by (5.3), and so $S=P_Y\mathscr{D}(\mathscr{O})/\bar{P}$.

We summarise these observations in the

Theorem. *Retain the notation of this section, (so $H(X)$ is the Artinian factor of $\mathscr{D}(X)$ associated to the singularity x of the irreducible curve $\mathscr{O}(X)$).*

(i) *There exists $g=g^2\in\hat{P}/\bar{P}$ with $g*(m/E)=0$ and $g*(1+E)=1+E$. Thus $\hat{P}/\bar{P}=H(X)g\cong W$.*
(ii) *There exists $f=f^2\in H(X)$ such that $M=fH(X)$ and $S=H(X)(1-f)$.*
(iii) $H(X)=\begin{pmatrix} fH(X)f & fH(X)(1-f) \\ 0 & (1-f)H(X)(1-f)\end{pmatrix}$. *Let $s=\dim_{\mathbb{C}}(W/V)=\dim_{\mathbb{C}}(\mathscr{O}/D)$. Then $(1-f)H(X)(1-f)\cong M_s(\mathbb{C})$ and $S=H(X)(1-f)$.*
(iv) *S is the right socle of $H(X)$. It is right homogeneous, with $SM=0$. As a left ideal it is isomorphic to the direct sum of s copies of W. In particular, it is left faithful.*
(v) $S=\hat{P}\mathscr{D}(\mathscr{O})/\bar{P}=P_Y\mathscr{D}(\mathscr{O})/\bar{P}$ *for all Y satisfying* (5.2, a *and* b).

6 Describing $\mathscr{D}(X)$ and its ideals using over-rings

6.1 Let $\bar{\mathscr{O}}$ be the maximal unramified extension of $\mathscr{O}$, as introduced in (3.1), and define

$$R=\langle\bar{\mathscr{O}},\mathscr{D}(\mathscr{O})\rangle\subseteq\mathscr{D}(\bar{\mathscr{O}}),$$

the inclusion being a consequence of (3.2). More precisely, we have, (in the notation of § 3), the

Proposition. *Let $e_1,\ldots,e_t\in\bar{\mathscr{O}}$ be representatives for the distinct primitive idempotents of $\bar{\mathscr{O}}/E$, and let $e_i'=e_i+E$, $i=1,\ldots,t$.*

(i) $R\subseteq\mathrm{Id}_{\mathscr{D}(\bar{\mathscr{O}})}(E\mathscr{D}(\bar{\mathscr{O}}))$.
(ii) *$\bar{\mathscr{O}}/E$ is a subring of $R/E\mathscr{D}(\bar{\mathscr{O}})$, the elements e_i' are central idempotents of $R/E\mathscr{D}(\bar{\mathscr{O}})$, and*

$$R/E\mathscr{D}(\bar{\mathscr{O}})=\sum_i{}^{\oplus}e_i'[\mathscr{D}(\mathscr{O})/E\mathscr{D}(\bar{\mathscr{O}})],$$

a direct sum of rings.
(iii) $R=\sum_i e_i\mathscr{D}(\mathscr{O})=\sum_i\mathscr{D}(\mathscr{O})e_i$.

Proof. By (4.4, i), $\mathscr{D}(\mathscr{O})\subseteq\mathrm{Id}_{\mathscr{D}(\bar{\mathscr{O}})}(E\mathscr{D}(\bar{\mathscr{O}}))$, so (i) is clear. Primary decomposition and (9) of § 4 yield, as $\mathscr{O}/E$-modules,

$$\bar{\mathscr{O}}/E\cong\sum_i{}^{\oplus}(\mathscr{O}/E)e_i'.$$

Since $E\mathscr{D}(\bar{\mathcal{O}}) = \mathrm{Ann}_{\mathscr{D}(\mathcal{O})}(\mathcal{O}/E)$, $\bar{\mathcal{O}}/E$ embeds in $R/E\mathscr{D}(\bar{\mathcal{O}})$ and Theorem (4.5) shows that e_i' is a central indempotent of $R/E\mathscr{D}(\bar{\mathcal{O}})$ for $i=1, \ldots, t$. As $\bar{\mathcal{O}} = \sum_i e_i\mathcal{O} = \sum_i \mathcal{O}e_i$, (ii) and (iii) now follow.

6.2 The ring R will enable us to prove inclusions of rings of differential operators additional to (3.2). The key here is the

Lemma. *Let A be a $\mathscr{D}(\mathcal{O})$-submodule of $\mathcal{O}$. Then the R-submodule of $\bar{\mathcal{O}}$ generated by A is $A\bar{\mathcal{O}}$. In particular, $\mathscr{D}(\mathcal{O})*(A\bar{\mathcal{O}}) \subseteq A\bar{\mathcal{O}}$.*

Proof. Clearly, $A\bar{\mathcal{O}} \subseteq R*(A)$; and by (iii) of (6.1)

$$R*(A\bar{\mathcal{O}}) = R*(A) = \sum_i e_i\,\mathscr{D}(\mathcal{O})*(A) \subseteq \sum_i e_i A = A\bar{\mathcal{O}}.$$

6.3 One might hazard a guess that (3.2) is the shadow cast by a stronger result, namely that $\mathscr{D}(\mathcal{O}) \subseteq \mathscr{D}(B)$ whenever B is a ring lying between $\mathcal{O}$ and $\bar{\mathcal{O}}$. But in general this is false. For example, if $D \subsetneq m$, take $B = \mathcal{O} + m\bar{\mathcal{O}}$; so $m\bar{\mathcal{O}}$, being the conductor of $\bar{\mathcal{O}}$ into B, is a $\mathscr{D}(B)$-submodule of B. However, since D is the unique maximal $\mathscr{D}(\mathcal{O})$-submodule of $\mathcal{O}$, by (5.1), $\mathscr{D}(\mathcal{O})*(m\bar{\mathcal{O}})$ contains $\mathbb{C}$. Nevertheless, some inclusions of this type are valid, as we can now see.

Theorem. (i) *If A is a $\mathscr{D}(\mathcal{O})$-submodule of $\mathcal{O}$, then $\mathscr{D}(\mathcal{O}) \subseteq \mathscr{D}(\mathcal{O} + A\bar{\mathcal{O}})$.*
(ii) *The conductor of $\bar{\mathcal{O}}$ into $\mathcal{O} + D\bar{\mathcal{O}}$ is $D\bar{\mathcal{O}}$.*

Proof. (i) follows from (6.2).
(ii) Clearly the conductor C of $\bar{\mathcal{O}}$ into $\mathcal{O} + D\bar{\mathcal{O}}$ contains $D\bar{\mathcal{O}}$. Suppose the inclusion is strict, so that $D \subsetneq C \cap \mathcal{O}$. By (4.4, ii), C is a $\mathscr{D}(\mathcal{O} + D\bar{\mathcal{O}})$-submodule of $\mathcal{O} + D\bar{\mathcal{O}}$. Thus (i) shows that $C \cap \mathcal{O}$ is a $\mathscr{D}(X)$-submodule of $\mathcal{O}$, which, in view of our supposition and the maximality of D amongst proper submodules, must equal $\mathcal{O}$. That is, $C \supseteq \mathcal{O}$, so $\bar{\mathcal{O}} = \mathcal{O} + D\bar{\mathcal{O}}$. But $\mathcal{O} + D\bar{\mathcal{O}}$ is local, whereas $\bar{\mathcal{O}}$ is unramified over $\mathcal{O}$. Hence the two rings coincide only when $\mathcal{O} = \bar{\mathcal{O}}$, and then $\mathcal{O} = E \subseteq D \subseteq \mathcal{O}$, so $D = C = D\bar{\mathcal{O}}$ in this case, contradicting our initial supposition.

6.4 Recall from (5.1) that M_0 denotes the maximal ideal of $\mathscr{D}(\mathcal{O})$ such that $M_0/\bar{P} = M$. Analogously, we denote by S_0 the ideal of $\mathscr{D}(\mathcal{O})$ for which $S_0/\bar{P} = S$, the right socle of $H(X)$. (We have already seen in (5.5) that $S_0 = \hat{P}\mathscr{D}(\mathcal{O})$.)

Theorem. (i) $M_0 = D\mathscr{D}(\bar{\mathcal{O}}) \cap \mathscr{D}(\mathcal{O})$.
(ii) $S_0 = \mathscr{D}(\mathcal{O} + D\bar{\mathcal{O}}, \mathcal{O})$.

Proof. (i) The maximal unramified extension of $\mathcal{O} + D\bar{\mathcal{O}}$ is $\bar{\mathcal{O}}$ and the conductor of $\bar{\mathcal{O}}$ into $\mathcal{O} + D\bar{\mathcal{O}}$ is $D\bar{\mathcal{O}}$, by (6.3, ii), so

$$D\mathscr{D}(\bar{\mathcal{O}}) = \mathrm{Ann}_{\mathscr{D}(\mathcal{O} + D\bar{\mathcal{O}})}(\mathcal{O} + D\bar{\mathcal{O}}/D\bar{\mathcal{O}}) \tag{1}$$

by Lemma (4.4, ii). Now

$$D\bar{\mathcal{O}} \cap \mathcal{O} = D; \tag{2}$$

for the left side certainly contains the right, while the left side is a proper $\mathscr{D}(\mathcal{O})$-submodule of $\mathcal{O}$, by (6.2), and thus must be contained in D by the maximality of the latter, (5.1). Therefore

$$\begin{aligned} M_0 &= \mathrm{Ann}_{\mathscr{D}(\mathcal{O})}(\mathcal{O}/D) \\ &= \mathrm{Ann}_{\mathscr{D}(\mathcal{O})}(\mathcal{O}/D\bar{\mathcal{O}} \cap \mathcal{O}), && \text{by (2),} \\ &= \mathrm{Ann}_{\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})}(\mathcal{O}+D\bar{\mathcal{O}}/D\bar{\mathcal{O}}) \cap \mathscr{D}(\mathcal{O}), && \text{by (6.3, i),} \\ &= D\mathscr{D}(\bar{\mathcal{O}}) \cap \mathscr{D}(\mathcal{O}), && \text{by (1).} \end{aligned}$$

(ii) The ring $\mathcal{O}+D\bar{\mathcal{O}}$ satisfies hypotheses (a) and (b) of (5.2), so $S_0 = \mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}}, \mathcal{O})\mathscr{D}(X)$ by (5.5, v). But $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}}, \mathcal{O})$ is a right ideal of $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$, so (6.3, i) shows that

$$\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}}, \mathcal{O})\mathscr{D}(\mathcal{O}) = \mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}}, \mathcal{O}).$$

6.5 Further light on the connection between $\mathscr{D}(\mathcal{O})$ and $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$ suggested by (6.4) is shed by the final results of this section:

Theorem. (i) $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}}) = \mathscr{D}(\mathcal{O}) + D\mathscr{D}(\bar{\mathcal{O}})$.
(ii) $\mathscr{D}(\mathcal{O}) = \mathrm{Id}_{\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})}(S_0) = \mathrm{End}_{\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})}(S_0)$.
(iii) *If* $\mathcal{O} \subsetneq \bar{\mathcal{O}}$, *then* $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$ *has global dimension* 2. *(If* $\mathcal{O} = \bar{\mathcal{O}}$, *then* $\mathscr{D}(\mathcal{O})$ *is hereditary, Morita equivalent to* $\mathscr{D}(\bar{\mathcal{O}})$.*)*
(iv) S_0 *is a generative right ideal of* $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$, *and a projective left ideal of* $\mathscr{D}(X)$.
(v) *The following are equivalent:*
(a) S_0 *is a projective right ideal of* $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$;
(b) $\mathscr{D}(\mathcal{O})$ *is Morita equivalent to* $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$;
(c) $S_0 = \mathscr{D}(\mathcal{O}) = \mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$;
(d) $E = D$;
(e) $H(X)$ *is simple Artinian.*

Proof. (i) By (6.3) and (4.4, i), $\mathscr{D}(\mathcal{O}) + D\mathscr{D}(\bar{\mathcal{O}})$ is contained in $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$, and $D\mathscr{D}(\bar{\mathcal{O}})$ is an ideal of $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$. Moreover, $D\bar{\mathcal{O}}$ is the unique maximal proper $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$-submodule of $\mathcal{O}+D\bar{\mathcal{O}}$, as can be seen either "directly" by showing that, in the notation of (5.1), $V=0$ here; or "indirectly" using the fact that $\mathscr{D}(\mathcal{O}) \subseteq \mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$. Hence, by (4.6),

$$\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})/D\mathscr{D}(\bar{\mathcal{O}}) = H(\mathcal{O}+D\bar{\mathcal{O}}) \cong M_s(\mathbb{C}), \tag{3}$$

where $s = \dim_{\mathbb{C}}(\mathcal{O}+D\bar{\mathcal{O}}/D\bar{\mathcal{O}})$. However, (2) shows that $s = \dim_{\mathbb{C}}(\mathcal{O}/D)$, and so, by (6.4, i), and (5.5, iii),

$$\mathscr{D}(\mathcal{O})/M_0 = \mathscr{D}(\mathcal{O})/D\mathscr{D}(\bar{\mathcal{O}}) \cap \mathscr{D}(\mathcal{O}) \cong M_s(\mathbb{C}). \tag{4}$$

Therefore, comparing the codimensions of $D\mathscr{D}(\bar{\mathcal{O}})$ in $\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})$ and $\mathscr{D}(\mathcal{O}) + D\mathscr{D}(\bar{\mathcal{O}})$ given by (3) and (4), we deduce equality of these rings.
(ii) Since $\mathcal{O}+D\bar{\mathcal{O}}$ satisfies (a) and (b) of (5.2), (5.4) applies to show that $S_0 * (\mathbb{C}) = \mathcal{O}$, so $\mathscr{D}(\mathcal{O}) = \mathrm{End}_{\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})}(S_0)$ by [13, (3.2)]; and this immediately yields equality in the inclusion – given by (6.3, i) – of $\mathscr{D}(\mathcal{O})$ in $\mathrm{Id}_{\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})}(S_0)$.
(iii) By [5, Theor. 2.2] and (3) we have $\mathrm{gl.dim.}(\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})) \leqq 2$. If $\mathcal{O} \subsetneq \bar{\mathcal{O}}$, then $\mathcal{O}+D\bar{\mathcal{O}} \subsetneq \bar{\mathcal{O}}$, and $\mathrm{gl.dim.}(\mathscr{D}(\mathcal{O}+D\bar{\mathcal{O}})) \geqq 2$ by [13, Corollary 4.12]. The statement in brackets is [13, Theor. 3.4].

(iv) As observed in proving (ii), $S_0*(\mathbb{C})=\mathcal{O}$. Hence S_0 cannot be contained in the unique proper ideal $D\mathcal{D}(\bar{\mathcal{O}})$ of $\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})$, since $D\mathcal{D}(\bar{\mathcal{O}})*(\mathcal{O}+D\bar{\mathcal{O}})=D\bar{\mathcal{O}}$. Thus S_0 is a generative right ideal of $\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})$.

The projectivity of S_0 as a left ideal of $\mathcal{D}(\mathcal{O})$ now follows by general properties of idealizers of generative right ideals, as in [7, (5.5.7, i)].

(v) *(a)⇒(b)* In view of (ii) and (iv) this is just the definition of Morita equivalence.

(b)⇒(c) Suppose $\mathcal{D}(\mathcal{O})$ is Morita equivalent to $\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})$. By (3) and the correspondence of ideals across an equivalence [7, Theor. (3.5.9)], $H(X)$ is simple Artinian. By definition, $S_0/\bar{P}$ is the right socle of $H(X)$, so $S_0=\mathcal{D}(\mathcal{O})$. Thus $1\in\{\mathcal{D}(\mathcal{O})\subseteq\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}},\ \mathcal{O})$ and so $\mathcal{O}=\mathcal{O}+D\bar{\mathcal{O}}$.

(c)⇒(d) If $\mathcal{O}=\mathcal{O}+D\bar{\mathcal{O}}$, then $D\bar{\mathcal{O}}$ is an ideal of $\bar{\mathcal{O}}$ contained in $\mathcal{O}$, so $D\bar{\mathcal{O}}\subseteq E$.

(d)⇒(e) If $D=E$, then by (5.1) $\mathcal{O}/E$ is a faithful irreducible $H(X)$-module, so $H(X)$ is simple Artinian.

(e)⇒(c) If $H(X)$ is simple then $S_0/\bar{P}$, being the right socle of $\mathcal{D}(\mathcal{O})/\bar{P}$, equals $\mathcal{D}(\mathcal{O})/\bar{P}$, and so $S_0=\mathcal{D}(\mathcal{O})$. As in the proof of (b)⇒(c), this proves $\mathcal{D}(\mathcal{O})=\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})$.

(c)⇒(a) Trivial.

6.6 Corollary. *The subcategory $\mathcal{S}$ of $\mathcal{D}(\mathcal{O})$-*Mod *generated by S_0 is equivalent to $\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})$-*Mod.

Proof. Let F be the division ring of quotients of $\mathcal{D}(\mathcal{O})$. Define $Q=\{q\in F: S_0q\subseteq S_0\}$, so $Q=\mathrm{End}_{\mathcal{D}(\mathcal{O})}(_|S_0)$ and $\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})\subseteq Q$. Since $J(X)$ and $S_0/J(X)$ are idempotent, by (2.1) and (5.5), $S_0^2=S_0$. By (6.5, iv), S_0 is a generative right ideal of $\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})$, so

$$\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})=\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})\,S_0^2=\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})\,S_0\,QS_0=QS_0,$$

while

$$Q=Q\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})=Q\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})\,S_0=QS_0.$$

Thus $Q=\mathcal{D}(\mathcal{O}+D\bar{\mathcal{O}})$ and the result follows from Morita theory.

7 Localizations, subidealizers and Grothendieck groups

7.1 It is convenient in this section to work globally, so throughout (7.1) to (7.6) $\mathcal{O}(X)$ will denote the ring of regular functions of the affine irreducible curve X, and $\mathcal{O}(\bar{X})$ will be the maximal unramified extension of $\mathcal{O}(X)$ in its integral closure $\mathcal{O}(\tilde{X})$, E the conductor of $\mathcal{O}(\bar{X})$ in $\mathcal{O}(X)$. Recall that $\bar{P}=E\mathcal{D}(\bar{X})=\mathcal{D}(\bar{X},\ X)$ is the unique minimal non-zero ideal of $\mathcal{D}(X)$ (4.4), and a right ideal of $\mathcal{D}(\bar{X})$, while $\mathcal{D}(\bar{X})$ contains $\mathcal{D}(X)$ and is Morita equivalent to $\mathcal{D}(\tilde{X})$ by (3.2) and (2.1, b), all these rings being contained in $\mathcal{D}(K)$, (where K is the quotient field of $\mathcal{O}(X)$), which in turn lies in the quotient division ring F of $\mathcal{D}(X)$.

7.2 Our main reference for localization is [14], especially Ch.XI. *A perfect left localization* B of a ring A is a ring homomorphism $\varphi: A\to B$ which is an epimorphism in the category of rings, making B into a flat right A-module; a perfect *right* localization is defined analogously. When A has a ring of fractions F, and B is a subring of F containing A, then B is a perfect left localization

if and only if it is a flat right A-module [14, Prop. XI.2.4]. If B is actually a finitely generated projective right A-module, then B is called *a finite left localization* of A [14, § XI.7].

7.3 We shall apply these ideas by means of a slight generalisation of a result of Hodges [6, Theor. 2.1]. Denote the category of finitely generated right [resp. left] modules over a ring R by Mod-R [resp. R-Mod]. For K-theoretic notation we follow [9], but writing $G_i^l(R)$ and $G_i^r(R)$ for the Grothendieck groups $K_i(R\text{-Mod})$ and $K_i(\text{Mod-}R)$ respectively.

Theorem. *Let $\phi\colon R \to S$ be a perfect left localization, where R and S are left Noetherian rings. Suppose that there exists an $R-S$-bimodule B, such that*

(a) *B_S is flat;*
(b) *${}_RB$ is a finitely generated submodule of ${}_RS$; and*
(c) *$S \otimes_R B \cong S$ via the evaluation map.*

Let $\mathscr{T} = \{M \in R\text{-Mod}\colon S \otimes_R M = 0\}$. Then for each non-negative integer i there is an exact sequence

$$0 \to K_i(\mathscr{T}) \to G_i^l(R) \to G_i^l(S) \to 0.$$

Proof. As in [6], the long exact sequence of localization applied to the short exact sequence of Noetherian categories $0 \to \mathscr{T} \to R\text{-Mod} \to S\text{-Mod} \to 0$ yields

$$\ldots G_{i+1}^l(S) \xrightarrow{\partial} K_i(\mathscr{T}) \longrightarrow G_i^l(R) \xrightarrow{K_i(\phi_*)} G_i^l(S) \xrightarrow{\partial} K_{i-1}(\mathscr{T}) \longrightarrow \ldots$$

where $\phi_*\colon R\text{-Mod} \to S\text{-Mod}\colon M \to S \otimes_R M$. Set

$$\phi^*\colon S\text{-Mod} \to R\text{-Mod}\colon X \to B \otimes_S X.$$

By (c), $\phi_* \phi^*$ is equivalent to the identity functor, whilst (a) and (b) ensure that $K_i(\phi^*)$ is a well-defined homomorphism of K-groups. Thus $K_i(\phi_*) K_i(\phi^*)$ is the identity on $G_i^l(S)$. Therefore the connecting maps ∂ are zero and the sequence splits.

7.4 Corollary. *Let B be a generative right ideal of a left Noetherian ring S, and let R be a left Noetherian subring of* $\mathrm{Id}(B)$ *which contains B. Suppose that B_S is flat. Then the hypotheses of* (7.3) *are satisfied and $\mathscr{T} = (R/B)$-Mod.*

Proof. By [7, Lemma 5.5.7(i)], S is a finite left localization of R. Now (a) and (b) of the theorem follow from the hypotheses of the corollary, and (c) follows by applying the functor $S \otimes_R -$, (which is exact since S_R is flat), to the sequence $0 \to B \to R \to R/B \to 0$, since $S \otimes_R (R/B) = SB \otimes_R (R/B) = 0$.

Similarly, if $M \in (R/B)$-Mod, then $S \otimes_R M = SB \otimes_R M = S \otimes_R BM = 0$; and if $N \in R$-Mod with $S \otimes_R N = 0$, then $B \otimes_R N = B \otimes_S S \otimes_R N = 0$, so BN, an image of $B \otimes_R N$, is zero. That is, $\mathscr{T} = (R/B)$-Mod.

7.5 The hypothesis of (7.4) are satisfied when $R = \mathscr{D}(X)$, $S = \mathscr{D}(\bar{X})$ and $B = \bar{P}$, since $\mathscr{D}(\bar{X})$, being Morita equivalent to $\mathscr{D}(\tilde{X})$, is hereditary. The outcome is the

Theorem. *Let X be an irreducible affine curve, with integral closure $\tilde{X}$. Let r be the number of isomorphism classes of irreducible $H(X)$-modules. Let i be a non-negative integer. Then*

$$G_i^l(\mathscr{D}(X)) \cong K_i(\mathbb{C})^{(r)} \oplus K_i(\mathscr{O}(\tilde{X})).$$

Proof. By (7.3) and (7.4), for each $i \geqq 0$,

$$G_i^l(\mathscr{D}(X)) \cong G_i^l(H(X)) \oplus G_i^l(\mathscr{D}(\bar{X})).$$

Now $G_i^l(\mathscr{D}(\bar{X}) \cong G_i^l(\mathscr{D}(\tilde{X}))$, since $\mathscr{D}(\bar{X})$ is Morita equivalent to $\mathscr{D}(\tilde{X})$, (2.1, b). The group on the right of this isomorphism is in turn isomorphic to $K_i(\mathscr{D}(\tilde{X}))$, by [9, § 3, Coroll. 2], since $\mathscr{D}(\tilde{X})$ is hereditary Noetherian; and this group in turn is isomorphic to $K_i(\mathscr{O}(\tilde{X}))$, by Quillen's Theorem on filtered rings [9, § 6, Theor. 7]. For, $\mathscr{D}(\tilde{X})$ has a filtration given by the order of differential operators, with associated graded ring $\mathscr{O}(T^*\tilde{X})$, the regular functions on the cotangent bundle over $\tilde{X}$ [1, p. 95]. Since $T^*\tilde{X}$ is non-singular, $\mathscr{O}(T^*\tilde{X})$ has finite global dimension (equal to 2, in fact), and so also does $\mathscr{D}(\tilde{X})_0 = \mathscr{O}(\tilde{X})$. Thus the hypotheses of Quillen's theorem are satisfied.

Finally, $G_i^l(H(X))$ is the direct sum of r copies of $K_i(\mathbb{C})$ by dévissage, [9].

7.6 In all known cases, $H(X)$ (and hence $\mathscr{D}(X)$ [4]) have finite global dimension, so that $G_i^l(\mathscr{D}(X))$ and $G_i^r(\mathscr{D}(X))$ equal the corresponding K-groups on the categories of finitely generated *projective* left and right $\mathscr{D}(X)$-modules, respectively; and these are isomorphic by [9]. While regularity of $\mathscr{D}(X)$ remains for the moment conjectural, we can nevertheless confirm the identity of its left and right Grothendieck groups, as follows.

Set $T = \mathrm{End}(\bar{P}_{\mathscr{D}(\bar{X})})$, so $\mathscr{D}(X) \subseteq T$ and T is Morita equivalent to $\mathscr{D}(\bar{X})$ and so to $\mathscr{D}(\tilde{X})$ by (2.1, b). In particular, T is simple, so $\bar{P}T = T$; and T is a hereditary Noetherian ring. Thus the right-hand version of (7.4) applies, with $B = \bar{P}$, $R = \mathscr{D}(X)$ and $S = T$. We conclude as in (7.5), but using the right-hand version of (7.3), that, for all $i \geqq 0$,

$$G_i^r(\mathscr{D}(X)) \cong K_i(\mathbb{C})^{(r)} \oplus K_i(\mathscr{O}(\tilde{X})).$$

8 Examples

The notation is that of (2.1).

8.1 $H(X)$ simple Artinian

As was observed in (6.5, v), $H(X)$ is simple if and only if $E = D$. As in (4.6), let $R_1, \ldots, R_t$ be the distinct indecomposable summands of $\bar{\mathscr{O}}/E$, and let r_i be the index of nilpotency of $J(R_i)$, for $i = 1, \ldots, t$. Let $E = \cap_i E_i$, indexed correspondingly, so that $R_i \cong \mathscr{O}/E_i$ and $E_i = E + (m_i^{r_i} \cap \mathscr{O})$, for $i = 1, \ldots, t$. Thus $H(X)$ is simple if and only if each R_i is isomorphic to $\mathscr{O}/E$. In this event, $r_1 = r_2 = \ldots = r_t = r$, say. Now $m^r = (m_i \cap \mathscr{O})^r \subseteq m_i^r \cap \mathscr{O}$, for all i, so $(m\bar{\mathscr{O}})^r \subseteq \cap E_i = E$, and r is the least integer for which this inclusion holds.

Suppose now that X is unramified, so that $\bar{\mathcal{O}}=\tilde{\mathcal{O}}$. Then each R_i, (and so also $\mathcal{O}/E$), is a principal ideal ring. Let z be a generator of $m\tilde{\mathcal{O}}/E$, so that

$$\mathcal{O}=\mathbb{C}+\mathbb{C}z+\mathbb{C}z^2+\ldots+\mathbb{C}z^{r-1}+E.$$

Moreover, since $\tilde{\mathcal{O}}$ is Dedekind, $E=\prod_i m_i^r=(\prod_i m_i)^r=z^r\tilde{\mathcal{O}}$. That is,

$$\mathcal{O}=\mathbb{C}+\mathbb{C}z+\mathbb{C}z^2+\ldots+\mathbb{C}z^{r-1}+\tilde{\mathcal{O}}z^r. \tag{1}$$

Conversely, if $\mathcal{O}$ has the form (1), then one calculates easily from (4.6) that

$$H(X)=M_r(\mathbb{C}).$$

8.2 Examples satisfying (1)

(These are given in non-localised form. If desired, localise at the maximal ideal of the singularity for consistency with (8.1).) Let $\tilde{\mathcal{O}}=\mathbb{C}[x]$, and $z=x(x-1)(x-2)\ldots(x-n)$, (where $n\geqq 2$). Let $r\geqq 1$, and set $\mathcal{O}=\mathbb{C}+\mathbb{C}z+\ldots+\mathbb{C}z^{r+1}+\tilde{\mathcal{O}}z^r$. (These are the examples given in [12, Prop. 5].) Generalising (8.1) for the unramified case, we obtain:

8.3 Block upper triangular examples

Assume that X is *unramified* at its unique singular point x, and (as usual) that we have localised at the maximal ideal m of x. Let $m_1, \ldots, m_t$ be the maximal ideals of $\tilde{\mathcal{O}}$ lying over m, so I, the conductor of $\tilde{\mathcal{O}}$ in $\mathcal{O}$, equals $\prod m_i^{r_i}$, where $r_1, \ldots, r_t$ are positive integers. Number the branches so that $r_1\geqq r_2\geqq\ldots\geqq r_t$. Recall from (9) of (4.6) that, for each $i=1, \ldots, t$,

$$\mathcal{O}/m_i^{r_i}\cap\mathcal{O}\cong\tilde{\mathcal{O}}/m_i^{r_i},$$

a local principal ideal ring. Hence, for each i,

$$\dim_{\mathbb{C}}(\mathcal{O}/I)\geqq\dim_{\mathbb{C}}(\mathcal{O}/m_i^{r_i}\cap\mathcal{O})=r_i. \tag{2}$$

In particular,

$$\dim_{\mathbb{C}}(\mathcal{O}/I)\geqq r_1. \tag{3}$$

Suppose that equality holds in (3). Then $\mathcal{O}/I$ is a local principal ideal ring, (and so serial); thus

$$I=m_1^{r_1}\cap\mathcal{O}\subseteq\ldots\subseteq m_i^{r_i}\cap\mathcal{O}\subseteq\ldots\subseteq\mathcal{O}. \tag{4}$$

From Theorem (4.6) with (2) and (4) we deduce: Under the above hypotheses, $H(X)$ is block upper triangular,

$$H(X)=\begin{pmatrix} M_{s_1}(\mathbb{C}) & & & * \\ & M_{s_2}(\mathbb{C}) & & * \\ & & \ddots & \\ 0 & & & M_{s_l}(\mathbb{C}) \end{pmatrix},$$

where every entry above the diagonal is $\mathbb{C}$, the number of blocks, l, is the number of strictly positive numbers $(r_i - r_{i+1})$, for $i=1, \ldots, t$, where we set $r_{t+1}=0$, and where $s_1, \ldots, s_l$ are these positive differences, in order of occurrence.

8.4 Remarks on 8.3

(a) The unramified case of (8.1) falls into this pattern, with $r_t - r_{t+1} = r_t = r$ the *only* positive difference.

(b) Suppose that the Artinian factor H associated to a singular point x of X is a block upper triangular matrix ring. As usual, let $\mathcal{O}$ denote the localization at x. Since $\mathcal{O}/I$ is a faithful indecomposable projective H-module, by (5.5, v), it is a uniserial H-module. From this and (4.6) it follows at once that equality holds in (3), so that $\mathcal{O}/I$ is a principal ideal ring. Thus the hypotheses of (8.3) are necessary; to be precise: *The Artinian factor H associated to an unramified singularity x is block upper triangular if and only if $\mathcal{O}/I$ is a principal ideal ring, where I is the conductor of $\tilde{\mathcal{O}}$ in $\mathcal{O}$.*

(c) An analysis along the lines of (8.3) and (8.4, b) in the case of a *ramified* singularity yields a similar conclusion: *$H(X)$ is block upper triangular if and only if the ideals $E_1, \ldots, E_t$ are totally ordered by inclusion.* But there does not seem to be a neat way of characterising this occurrence in terms of $\mathcal{O}/E$.

8.5 Specific examples for (8.3)

As in (8.2) we give global examples.

Let $\tilde{\mathcal{O}}=\mathbb{C}[x]$, $z=\prod_{i=0}^{n-1}(x-i)$, (where $n\geqq 3$). Choose positive integers $r_0, \ldots, r_{n-1}$, and set

$$I=\prod_{i=0}^{n-1}(x-i)^{r_i}\mathbb{C}[x] \quad \text{and} \quad \mathcal{O}=I+\sum_{l\geqq 0}\mathbb{C}z^l.$$

It is routine to prove the

Lemma. (i) *$\mathcal{O}/I$ has a basis $\{z^l: 0\leqq l\leqq r_0-1\}$.*
(ii) *I is the conductor of $\tilde{\mathcal{O}}$ in $\mathcal{O}$ unless $r_0-1=r_1=r_{n-1}$. (When these equalities hold, z^{r_0-1} is in the conductor.)*

Thus, since $\mathcal{O}$ is visibly unramified, it satisfies the conditions for (8.3). For appropriate choices of n and r_i $(0 \leqq i \leqq n-1)$ one gets any desired block upper triangular matrix ring as $H(X)$. The examples with $H(X)=\begin{pmatrix} \mathbb{C} & \mathbb{C} \\ 0 & \mathbb{C} \end{pmatrix}$ of [12, Prop. 6] have this form.

8.6 [13, Prop. 5.6]. *Examples with* $H(X)=\begin{pmatrix} \mathbb{C} & \mathbb{C}\oplus\mathbb{C} \\ 0 & \mathbb{C} \end{pmatrix}$.

Let $\tilde{\mathcal{O}}=\mathbb{C}[x]$, $z=\prod_{i=0}^{n-1}(x-i)$, (where $n \geqq 3$). Set $I=z^2\tilde{\mathcal{O}}$ and $\mathcal{O}=I+\mathbb{C}+\mathbb{C}z+\mathbb{C}xz$. One checks easily that I is the conductor. Note that $(x-i)^2\tilde{\mathcal{O}}\cap\mathcal{O}=I+\mathbb{C}(x-i)z$, so that $D=m=I+\mathbb{C}z+\mathbb{C}xz$. Theorem (4.6) or (5.5) shows that $H(X)=\begin{pmatrix} \mathbb{C} & \mathbb{C}\oplus\mathbb{C} \\ 0 & \mathbb{C} \end{pmatrix}$.

8.7 Example with $H(X)=\begin{pmatrix} \mathbb{C}\oplus\mathbb{C} & \mathbb{C}\oplus\mathbb{C} \\ 0 & \mathbb{C} \end{pmatrix}$.

Let $\tilde{\mathcal{O}}=\mathbb{C}[x]$ and $z=\prod_{i=0}^{3}(x-i)$.

Set $a=zx(x-1)$ and $b=z(x-2)(x-3)$, and put

$$\mathcal{O}=z^2\tilde{\mathcal{O}}+\mathbb{C}a+\mathbb{C}b+\mathbb{C}.$$

Here, $\mathcal{O}$ is unramified with conductor $z^2\tilde{\mathcal{O}}$. Further,

$$x^2\tilde{\mathcal{O}}\cap\mathcal{O}=(x-1)^2\tilde{\mathcal{O}}\cap\mathcal{O}=I+\mathbb{C}a,$$

and

$$(x-2)^2\tilde{\mathcal{O}}\cap\mathcal{O}=(x-3)^2\tilde{\mathcal{O}}\cap\mathcal{O}=I+\mathbb{C}b.$$

Hence $D=m=I+\mathbb{C}a+\mathbb{C}b$. Recalling from (5.5) that $S=\mathrm{Ann}_{H(X)}(D/I)$, Theorem (4.6) shows

$$\begin{aligned} H(X)/S &= \{\psi\in\mathrm{End}_{\mathbb{C}}(\mathbb{C}a+\mathbb{C}b): \psi(a)\in\mathbb{C}a, \psi(b)\in\mathbb{C}b\} \\ &= \mathbb{C}\oplus\mathbb{C}. \end{aligned}$$

Hence, by Theorem 5.5, $H(X)=\begin{pmatrix} \mathbb{C}\oplus\mathbb{C} & \mathbb{C}\oplus\mathbb{C} \\ 0 & \mathbb{C} \end{pmatrix}$.

9 Questions and conjectures

For each $i=1, \ldots, t$, set $\mathcal{O}_i=\mathcal{O}+E_i\bar{\mathcal{O}}$, and $S_i=\mathcal{D}(\mathcal{O}_i, \mathcal{O})$. Thus, for all i, $\mathcal{O}\subseteq\mathcal{O}_i\subseteq\mathcal{O}+D\bar{\mathcal{O}}\subseteq\bar{\mathcal{O}}$, and by (6.3, i)

(1) $$\mathcal{D}(\mathcal{O})\subseteq\mathcal{D}(\mathcal{O}_i),$$

so each S_i is an ideal of $\mathscr{D}(\mathscr{O})$. Moreover,

$$\cap S_i = \mathscr{D}(\sum_i \mathscr{O}_i, \mathscr{O}) = \mathscr{D}(\mathscr{O} + D\bar{\mathscr{O}}, \mathscr{O}) = S.$$

We are thus led by (6.5, iv) to ask

9.1 Question. Is S_i a generative right ideal of $\mathscr{D}(\mathscr{O}_i)$?

Whatever the answer to (9.1), I believe that there is a close connection between the rings $H(\mathscr{O}_i)$, for $i=1, \ldots, t$, and $H(X)$. Let J_i be the minimal non-zero ideal of $\mathscr{D}(\mathscr{O}_i)$. It is not hard to see that

$$\cap \mathscr{O}_i = \mathscr{O}. \tag{2}$$

Now $J(X) = E\mathscr{D}(\bar{\mathscr{O}}) \subseteq J_i$ for all i, whereas, by (2),

$$\cap J_i = \cap \mathscr{D}(\bar{\mathscr{O}}, \mathscr{O}_i) \subseteq \mathscr{D}(\bar{\mathscr{O}}, \cap \mathscr{O}_i) = J(X).$$

It follows that

$$J(X) = \cap J_i. \tag{3}$$

From (1) and (3) we deduce

9.2 $H(X)$ embeds in $\sum_i^{\oplus} H(\mathscr{O}_i)$.

Consideration of examples suggests that it may be worth asking

9.3 Question. Does the embedding (9.2) *yield a surjection of $H(X)$ onto $H(\mathscr{O}_i)$, for all i? That is, is $\mathscr{D}(\mathscr{O}_i) = J_i + \mathscr{D}(\mathscr{O})$?*
By (1) and (2),

$$\mathscr{D}(\mathscr{O}) = \cap \mathscr{D}(\mathscr{O}_i).$$

A positive answer to (9.1) would therefore open the way to an analysis of $\mathscr{D}(\mathscr{O})$ in terms of its finite set of finite left localizations $\mathscr{D}(\mathscr{O}_i)$, $i=1, \ldots, t$, [7, Lemma 5.5.7(i)]. For example, we ask

9.4 Is there an embedding of the derived category of bounded complexes of $\mathscr{D}(\mathscr{O}_i)$-modules in that of $\mathscr{D}(\mathscr{O})$-modules, for each i, such that together their images "generate" the whole derived category of bounded complexes of $\mathscr{D}(\mathscr{O})$-modules?

Acknowledgements. I have been helped in this research by conservations with I.M. Musson, L. O'Carroll and S.P. Smith. In particular, I am grateful to Paul Smith – this work began as a joint project with him, and a version of Theorem (4.6) in the special case of an unramified singularity was proved by him in the course of that collaboration. A similar result was obtained independently by Ian Musson.

References

1. Bjork, J.-E.: Rings of differential operators. Mathematical Library, Amsterdam: North-Holland 1979
2. Brown, W.C.: A note on higher derivatives and ordinary points of curves. Rocky Mt. J. Math. **14**, 397–402 (1984)

3. Cline, E., Parshall, B., Scott, L.: Algebraic stratification in representation categories. J. Algebra **117**, 504–521 (1988)
4. Fields, K.L.: On the global dimension of residue rings. Pac. J. Math. **32**, 345–349 (1970)
5. Goodearl, K.R.: Subrings of idealizer rings. J. Algebra **33**, 405–429 (1975)
6. Hodges, T.J.: *K*-theory and right ideal class groups for *HNP* rings. Trans. Am. Math. Soc. **302**, 751–767 (1987)
7. McConnell, J.C., Robson, J.C.: Noncommutative Noetherian rings. New York: Wiley-Interscience 1987
8. Muhasky, J.L.: The differential operator ring of an affine curve. Trans. Am. Math. Soc. **307**, 705–723 (1988)
9. Quillen, D.: Higher algebraic *K*-theory I. In: Bass, H. (ed.) Algebraic *K*-theory. Vol. 1. Lect. Notes Math., vol. 341. Berlin Heidelberg New York: Springer 1973
10. Scott, L.L.: Simulating algebraic geometry with algebra I: The algebraic theory of derived categories. Proc. Symp. Pure Math. **47**, 271–281 (1987)
11. Smith, S.P.: Differential operators on commutative algebras. (Lect. Notes Math. vol. 1197) Berlin Heidelberg New York: Springer 1986
12. Smith, S.P.: Curves, differential operators and finite dimensional algebras. Seminaire d'algèbre P. Dubreil et M.-P. Malliavin (1985–86). Lect. Notes Math., vol. 1296. Berlin Heidelberg New York: Springer 1988
13. Smith, S.P., Stafford, J.T.: Differential operators on an affine curve. Proc. Lond. Math. Soc. **56**, 229–259 (1988)
14. Stenstrom, B.: Rings of Quotients. Berlin Heidelberg New York: Springer 1975

Math. Z. 206, 443–455 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

Applications of the Ochanine Genus

Martin Bendersky
CUNY, Hunter College, 695 Park Avenue, New York, NY 10021, USA

Received July 20, 1989; in final form April 2, 1990

1 Introduction

Results known to be true for the signature and $\hat{A}$ genus of a $4k$ manifold have been generalized to the elliptic genus of Ochanine, Landweber and Stong. For example the Bott-Taubes rigidity theorem, [B-T], may be viewed as an extension of the vanishing theorem of [A-H] to elliptic genera. More recently Hirzebruch has generalized a theorem about the G-signature of a manifold with orientation preserving involution, [H], to the elliptic genus of manifolds with orientation-preserving involutions.

In [O3], S. Ochanine extended the elliptic genus to $8k+1$ and $8k+2$ Spin manifolds. Ochanine's genus specializes in dimension $8k+2$ both to $\hat{A}$, a generalization of the invariant introduced by Atiyah in [A] to study Riemann surfaces, and in many cases (e.g. for SU-manifolds) to the Brown-Kervaire invariant. Hirzebruch raised the question of generalizing his result to Ochanine's genus. Such a generalization would imply interesting vanishing results for the Atiyah and Brown-Kervaire invariants. In light of [A-H] it is surprising that the $\hat{A}$ genus in fact need not be zero on an $8k+2$ manifold with a non-trivial circle action. The existence of a non-trivial Spin structure on S^1 precludes a general vanishing theorem. In this paper we study the question of the behavior of Ochanine's genus on Spin manifolds with non-trivial S^1 action.

The most interesting applications are specializations of the results in §6 and §7. We recall the definition of even type and odd type S^1 actions on a Spin manifold. An action has even type if it respects the Spin structure on each connected component, odd type if on each component it does not preserve the Spin structure.

Theorem A *If M is an $8k+i$, $(i=1,2)$ Spin manifold with a non-trivial odd type, semi-free S^1 action, then the $\hat{A}$ and Ochanine-Kervaire invariants of M vanish.*

Theorem B *If M is an $16k+10$ SU manifold with non-trivial, free, S^1 action which lifts to the SU structure, then the $\hat{A}$ and Ochanine-Kervaire invariants of M vanish.*

The Ochanine-Kervaire invariant is defined in (3.5), and is closely related to the Brown-Kervaire invariant.

We conjecture that Theorem A is true without the semi-free condition on the action. We also conjecture, following Ochanine, that (up to Spin manifolds

with all KO-characteristic numbers zero), the kernel of the Ochanine genus is precisely the elements in $\Omega_*^{\text{Spin}} = \pi_* M\,\text{Spin} = M\,\text{Spin}_*$ which contain a manifold with non-trivial, odd type, semi-free S^1 action.

2 Elliptic genera

Let G be SO or Spin, MG_* the G bordism group and Λ a ring. A *Λ-valued G genus*, φ, is a ring homomorphism

$$\varphi\colon MG_* \to \Lambda.$$

The "classical" case, $G = SO$ and Λ and $\mathbb{Q}$-algebra was studied in [L-S]. Since $MSO_* \otimes \mathbb{Q}$ is generated by $\{\mathbb{Q}P^{2n}\}$, $n = 1, 2, \ldots$, φ is determined by the *log series*, $g(x)$

$$g(x) = \Sigma \frac{\varphi(\mathbb{C}P^{2n})}{2n+1} x^{2n+1} = \int_0^x \Sigma \varphi(\mathbb{C}P^{2n})\, t^{2n} dt.$$

Following Ochanine [O2] we say φ is *elliptic* if $g(x)$ has the form

$$g(x) = \int_0^x \frac{dt}{\sqrt{1 - 2\delta t^2 + \varepsilon t^4}}$$

where $\delta, \varepsilon \in \Lambda$.

Call φ *non-degenerate* if $R(t) = 1 - 2\delta t^2 + \varepsilon t^4$ has distinct roots. The important *degenerate* cases are $\varepsilon = \delta = 1$, and $\varepsilon = 0$, $\delta = -\frac{1}{8}$. In the first case φ is the signature, in the second φ is the $\hat{A}$ genus [A-H].

The genus is recovered from the *characteristic power series*, $u/s(u)$ where $s(g(u)) = u$. If we write the total Pontrjagin class of a manifold M as a formal product

$$P(M) = \prod_{i=1}^{2k} (1 + x_i^2)$$

then

$$\varphi(M) = \{\prod x_i / s(x_i)\}\,[M].$$

The characteristic power series of the signature is $x/\tanh(x)$, and the characteristic power series of the $\hat{A}$ genus is $x/2 \sinh \frac{1}{2x}$.

An easy computation shows that $\delta = \varphi(\mathbb{C}P^2)$ and $\varepsilon = \varphi(\mathbb{H}P^2)$. We assign degrees of 2 and 4 to δ and ε respectively corresponding to the *complex* dimensions of the related manifolds.

We will have occasion to use two constructions of universal elliptic genera due to Witten [W]. One is related to the index of the signature operator, and the other is associated to the Dirac operator.

For E a real vector bundle over X, we define the total exterior and symmetric power operations

$$\Lambda_t(E) = \sum \Lambda^i(E)\, t^i$$
$$S_t(E) = \sum S^i(E)\, t^i.$$

Define

$$\mathscr{B}_q(E)=\bigotimes_{n\geq 1}(\Lambda_{q^{2n}}(E)\otimes S_{q^{2n}}(E))$$

and

$$\Theta_q(E)=\bigotimes_{n\geq 1}(\Lambda_{-q^{2n-1}}(E)\otimes S_{q^{2n}}(E))$$

$\Theta_q(E)$ and $\mathscr{B}_q(E)$ are formal power series in q with coefficients in $KO(X)$. Moreover

$$\Theta_q(E\oplus F)=\Theta_q(E)\,\Theta_q(F)$$

and

$$\mathscr{B}_q(E\oplus F)=\mathscr{B}_q(E)\,\mathscr{B}_q(F),$$

hence Θ_q and $\mathscr{B}_q$ may be extended to virtual bundles. Let $\tilde{E}=E-\dim E$. Define genera

$$\Phi_i\colon\ MSO_{4k}\to\mathbb{Q}[\![q]\!]$$

by

$$\Phi_1(M)=\langle 2^{2k}\hat{L}(M)\,ch(\mathscr{B}_q(\tilde{T}_M)\otimes\mathbb{C}),[M]\rangle \tag{2.1}$$
$$\Phi_2(M)=\langle\hat{A}(M)\,ch(\Theta_q(\tilde{T}_M)\otimes\mathbb{C}),[M]\rangle$$

where T_M is the tangent bundle bundle of M and $\hat{L}$ has characteristic power series $x/2\tanh(x/2)$. Zagier proves that Φ_1 and Φ_2 are universal elliptic genera with ε and δ generators of $M(\Gamma_0)$, the ring of modular forms over $\mathbb{C}$ for the subgroup of $SL_2(\mathbb{Z})$ given by matrices $\begin{bmatrix}a & b\\ c & d\end{bmatrix}$ with $c\equiv 0 \bmod 2$.

Theorem 2.2 [Z] (i) *Φ_1 is elliptic with*

$$\delta=1/2\prod_{n\geq 1}(1-q^{2n})^4\{\prod_{n\geq 1}((1+q^{2n-1})^8+(1-q^{2n-1})^8\}$$
$$\varepsilon=\prod_{n\geq 1}(1-q^{2n})^8(1+q^{2n-1})^8(1-q^{2n-1})^8.$$

(ii) *Φ_2 is elliptic with*

$$\delta=-\tfrac{1}{8}-3\sum_{n\geq 1}\Big(\sum_{\substack{d\mid n\\ d\text{ odd}}}d\Big)q^n$$
$$\varepsilon=\sum_{n\geq 1}\Big(\sum_{\substack{d\mid n\\ n/d\text{ odd}}}d^3\Big)q^n.\quad\square$$

A function, $f(\tau)$ is a *modular form for Γ_0 of weight $2k$* if f is a holomorphic function defined on the upper half-plane satisfying

$$f\left(\frac{a\tau+b}{c\tau+d}\right)=(c\tau+d)^{2k}f(\tau)$$

for all $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma_0$. Let $q = e^{2\pi i \tau}$, then f is assumed to be holomorphic at ∞ and at the other cusp $\tau = 0$. Hence f has Fourier expansion $f(\tau) = \sum_{n \geq 0} a_n q^n$, with a similar holomorphicity condition at $\tau = 0$. From [L] or [Z], $M_*(\Gamma_0) = \mathbb{C}[\delta, \varepsilon]$ as graded rings, where the grading on M_* is given by the weight of f, and the grading on $\mathbb{C}[\delta, \varepsilon]$ is induced by the degrees of δ and ε. Therefore, from (2.2), Φ_i are $M_*(\Gamma_0)$ valued SO genera.

Let $M_*^{\mathbb{Z}}(\Gamma_0)$ denote the ring of elements of $M_*(\Gamma_0)$ with integral Fourier coefficients.

Proposition 2.3 [C] *Let δ and ε be as in* 2.2 (ii). *Then*

$$M_*^{\mathbb{Z}}(\Gamma_0) = \mathbb{Z}[\delta_0, \varepsilon] \quad \textit{with } \delta_0 = -8\delta. \quad \square$$

Let Φ be an elliptic genus defined on MSO_*. Then φ, the restriction of Φ to $M\,\mathrm{Spin}_*$, is also an elliptic genus. By (2.1), φ may be described in terms of KO-characteristic numbers. The integrability theorems for KO-characteristic numbers [A-H] imply the image of φ on $M\,\mathrm{Spin}_{2*}$ lies in:

$$M_*^{\mathbb{Z}}(\Gamma_0) * \equiv 0 \mod 4$$
$$2M_*^{\mathbb{Z}}(\Gamma_0) * \equiv 2 \mod 4.$$

(In [L] it is shown that φ is surjective.)

It is interesting to note how the interplay between the two representations of the universal elliptic genus in (2.1) imply non-trivial integrability theorems. We used Φ_2 to compute the image of *any elliptic Spin genus*. In particular if M is a Spin manifold $\Phi_1(M^{8k+4})$ must be a polynomial in δ and ε with coefficients divisible by 16. This is essentially the proof given in [L, p. 63] of the following theorem of Ochanine.

Theorem 2.4 [O1] *The signature of an* $8k+4$ Spin *manifold is divisible by* 16.

Proof. Set q equal to zero in Φ_1. $\square$

We will use this dual representation of the universal elliptic genus to apply a theorem of Hirzebruch to the Ochanine genus, described in §3.

3 Ochanine's genus

Let M^n be a closed n-dimensional Spin manifold, and $[M] \in KO_n(M)$ the fundamental class.

Definition 3.1 [O3] $\beta_q[M] = \Theta_q(T_M - n)\,[M] \in KO_n(\mathrm{point})[\![q]\!]$.

The multiplicative property of Θ_q implies β_q is a $KO_*[\![q]\!]$ $(= KO_*(\mathrm{point})\,[\![q]\!])$ Spin genus. We call β_q the Ochanine genus. It is a natural extension of the elliptic genus to manifolds of dimension $\neq 4k$. We briefly summarize the results in [O3].

Recall that the $\mathbb{Z}/8\mathbb{Z}$-graded ring, KO_*, is generated by an element η in degree 1 and an element ω in degree 4 with relations

$$2\eta=\eta^3=\eta\omega=0, \quad \omega^2=4.$$

The *Pontrjagin character*, $ph\colon KO^*(X)\to H^{**}(X;\mathbb{Q})$, defined as complexification followed by the Chern character, satisfies

$$ph(\eta)=0, \quad ph(\omega)=2.$$

(The 2 appears because the complexification map from KO to KU is multiplication by two in dimensions congruent to 4 mod 8. This is the reason KO characteristic numbers are even on an $8k+4$ Spin manifold.)

The formula for evaluation of KO-characteristic numbers [S, p. 294] shows that

$$ph(\beta_q)=\Phi_2. \tag{3.2}$$

So we can identify β_q and Φ_2 in dimensions congruent to 0 mod 4. In particular the constant term in β_q is the extension of the $\hat{A}$ genus to dimensions not divisible by 4. In dimensions not congruent to 0 mod 4 this is called the Atiyah invariant.

Let $\bar{\varepsilon}$ denote the mod 2 reduction of ε defined in (2.2) (ii). Notice that $\delta_0\equiv 1 \bmod 2$.

Theorem 3.3 [O 3] *Let M^n be a* Spin *manifold. Then in* $(KO_n\otimes\mathbb{F}_2)[\![q]\!]$

$$\beta_q[M^n]\equiv a_0[M]+a_1[M]\,\bar{\varepsilon}+\ldots+a_m[M]\,\bar{\varepsilon}^m, \bmod 2, \quad m=[n/8]$$

for elements $a_i[M]\in KO_n\otimes\mathbb{F}_2$. □

The KO-characteristic number, $a_0[M]=\beta_0[M]$, is the $\hat{A}$-genus of M. Ochanine shows that the KO-characteristic number, $a_m[M]$ can be identified with more familiar invariants.

First suppose that M^{8k+3} is a spin manifold which represents zero in $M\,\mathrm{Spin}_{8k+3}$. In this dimension this is equivalent to all Stiefel-Whitney numbers vanishing. Let W^{8k+4} be a spin manifold with $\partial W=M$. Then (2.4) and the additivity of the signature, σ, imply $\sigma(W)$ mod 16 is well defined.

Definition 3.4. The generalized Rohlin invariant $R(M)$ is given by the formula:

$$R(M)=\sigma(W) \bmod 16.$$

Now suppose V^{8k+2} is a Spin manifold. Let $\bar{S}$ denote the circle with a non-trivial Spin structure. $\bar{S}$ represents the generator of $M\,\mathrm{Spin}_1\approx\mathbb{Z}/2\mathbb{Z}$. $V\times\bar{S}$ is zero in the Spin bordism group and, if W is a manifold with $\partial W=V\times\bar{S}$, $\sigma(W)$ is divisible by 8. (This follows since there is an orientation reversing, Spin preserving diffeomorphism of $\bar{S}$.)

Definition 3.5 The Ochanine-Kervaire invariant $k(V)$ is give by the following formula:

$$k(V)=\sigma(W)/8 \bmod 2$$

$k(V)$ is closely related to Brown's generalized Kervaire invariant [Br]. In [O1] Ochanine shows that the Brown-Kervaire invariant and $k(V)$ are the same if all tangential Stiefel-Whitney numbers divisible by an odd Stiefel-Whitney class vanish.

Definition 3.6 For a Spin manifold, M, of dimension n, $\mathscr{K}(M) \in KO_n \otimes \mathbb{F}_2$ is defined by

$$\mathscr{K}(M) = \begin{cases} \sigma(M), & n \equiv 0 \mod 8 \\ k(\bar{S} \times M)\,\eta, & n \equiv 1 \mod 8 \\ k(M)\,\eta^2, & n \equiv 2 \mod 8 \\ (\sigma(M)/16)\,\omega, & n \equiv 4 \mod 8. \end{cases}$$

Theorem 3.7 [O3] (i) $\mathscr{K}$ *is a* $KO_* \otimes \mathbb{F}_2$ *valued* Spin *genus.*
(ii) $a_m[M^n] = \mathscr{K}(M^n)$, *where* $m = [n/8]$.

4 Hirzebruch's theorem

Let M^{4k} be a closed, oriented manifold with an involution, I, which is homotopic to the identity. In [H], Hirzebruch derives the following relation between the signature of M and the signature of the fixed points.

$$\sigma(M) = \sigma(M^I \cdot M^I), \tag{4.1}$$

where M^I is the set of points in M fixed by the involution, and $M^I \cdot M^I$ is the self-intersection of M^I. Hirzebruch generalized this result to an elliptic genus, Φ, defined by

$$\Phi(M^{4k}) = \langle 2^{2k} \hat{L}(M)\, ch(\mathscr{B}_q(T_M) \otimes \mathbb{C}), [M] \rangle.$$

Let $\varphi = \prod_{n \geqq 1} \left(\frac{1+q^{2n}}{1-q^{2n}} \right)$. Then Φ is related to Φ_1 of (2.1) by the formula

$$\Phi(M^{4k}) = \Phi_1(M^{4k})\, \varphi^{4k}. \tag{4.2}$$

Using the rigidity theorem [B-T] Hirzebruch proves the following generalization of (4.1).

Theorem 4.3 [H1], [H-S] *Let* M^{4k} *be a* Spin *manifold with a non-trivial* S^1 *action. Let* M^I *be the set of points fixed by the involution, I, induced by the circle action. Then*

$$\Phi(M) = \Phi(M^I \cdot M^I). \quad \square$$

For our purposes it is convenient to rewrite (4.3) in terms of Φ_1. Recall that a circle may act on a connected, Spin manifold, M^{4k} in one of two ways. We may assume the action preserves a Riemannian metric and so lifts to an action on the principal tangential $SO(n)$ bundle. If the S^1 action further lifts to the Spin(n) bundle which defines the Spin structure on M we say the action has *even type.* If the action does not lift to Spin(n), then we say the action has *odd type.* The type of an action is characterized by the codimensions of the fixed point sets of the involution, [A-B]. If the action has even type then

each component of M^I has codimension congruent to 0 mod 4 in M. If the action has odd type then each component of M^I has codimension congruent to 2 mod 4 in M. In particular if the action has even type the codimension of $M^I \cdot M^I$ is congruent to 0 mod 8.

Proposition 4.4 (i) *If the action has even type then*

$$\Phi_1(M^{4k}) = \sum_{\text{codim. } M^I \cdot M^I = 8i} \Phi_1(M^I \cdot M^I) \cdot \varepsilon^i.$$

(ii) *If the action has odd type then* $\Phi_1(M^{4k}) = 0$.

Proof. From (2.2) (i) and the formula for φ we have

$$\varphi = (1/\varepsilon)^{1/8}.$$

Substituting (4.2) into (4.3) we obtain

$$\Phi_1(M)/(\varepsilon^{k/2}) = \sum_{\text{codim. } M^I \cdot M^I = 8i} \Phi_1(M^I \cdot M^I)/(\varepsilon^{k/2-i})$$

and (i) follows. Part (ii) is implicit in [O2]. □

We may replace Φ_1 and ε in (4.4) with Φ_2 and ε as in (2.2) (ii). In particular, since the $\hat{A}$ genus is obtained by setting ε to 0, (4.4) implies $\hat{A}(M^{4k}) = 0$ if M admits a non-trivial circle action (see [A-H]).

Hirzebruch asked if (4.4) could be true for the genus β_q. Such a theorem would have interesting consequences for the Ochanine-Kervaire invariant and the Atiyah invariant. Unfortunately there is a counter-example. Let $\bar{S}_0$ denote the circle with the non-trivial Spin structure with S^1 acting by rotation. Let V be any $8k+1$ Spin manifold with $\beta_q(V) \neq 0$. Then $V \times \bar{S}_0$ with S^1 acting trivially on V is a Spin manifold with free S^1 action. If (4.4) were true for β_q, we would have to have $\beta_q(V \times \bar{S}_0) = 0$. But since $\beta_q(\bar{S}) = \eta$ we have $\beta_q(V \times \bar{S}_0) = \beta_q(V)\,\eta \neq 0$. There are similar examples to show that $\hat{A}$ and $\mathscr{K}$ do not necessarily vanish on an $8k+2$ Spin manifold with free circle action. We will show that, in some sense, this is all that can go wrong if the action is semi-free.

5 The key lemma

Suppose M^n be a closed Spin manifold of dimension $n = 8k+2$ with a circle action, λ. Let W be an $8k+4$ dimensional Spin manifold with $\partial W = M \times \bar{S}$.

Lemma 5.1 *Suppose W has a circle action, α, such that α restricted to ∂W is the action $\lambda \times 1$. Then*

(i) *If λ is an odd type action $\beta_q(M) = 0$.*

(ii) *If λ is an even type action*

$$\beta_q(M) = \sum_{\text{codim. } M^I \cdot M^I = 8i} \beta_q(M^I \cdot M^I) \cdot \bar{\varepsilon}^i.$$

We first recall the geometric construction of the Toda bracket $\langle 2, \eta, - \rangle$ in [S, p. 341]. Suppose V is an $8k+2$ Spin manifold and W is an $8k+4$ Spin

manifold with $\partial W = V \times \bar{S}$. Since $\bar{S}$ is of order 2 in $M\,\mathrm{Spin}_1$, we have $2\bar{S} = \partial N^2$, for some 2 dimensional Spin manifold, N. Define an $8k+4$ dimensional Spin manifold $T(V)$ by

$$T(V) = (2W) \bigcup_{2V \times \bar{S}} (-V \times N^2).$$

Theorem 5.2 [O3] $\beta_q(T(V)) \otimes 1 = t(\beta_q(V))$ *where* t: $KO_2 \to KO_4 \otimes \mathbb{F}_2$ *is the isomorphism sending* η^2 *to* $\omega \otimes 1$. □

Proof of 5.1 The S^1 action, $\lambda \times 1$, on $M \times \bar{S}$ extends to an action, μ, on $T(M)$. We have

$$T(M^I) = T(M)^I$$

and

$$T(M^I \cdot M^I) = T(M)^I \cdot T(M)^I.$$

If λ is an odd action, then μ has odd type and (4.4) implies $\Phi_2(T(M)) = 0$; hence by (3.2) and (5.2), $\beta_q(T(M)) = \beta_q(M) = 0$.

If λ is an even type action, then μ has even type and (4.4) implies

$$\Phi_2(T(M)) = \sum \Phi_2(T(M^I \cdot M^I))\, \varepsilon^i.$$

Using (3.2) again, we obtain a similar formula with β_q replacing Φ_2. (5.1) now follows from (5.2). □

Corollary 5.3 *Suppose* M *is a* Spin *manifold of dimension* $8k+1$ *satisfying the conditions stated in the first paragraph of* §5. *Then Lemma* 5.1 *is true for* M.

Proof. $M \times \bar{S}$, with trivial action on $\bar{S}$, is an $8k+2$ manifold satisfying the conditions of (5.1). The corollary follows since $\beta_q(M \times \bar{S}) = \beta_q(M)\,\eta$. □

6 Free actions

Let $M\,\mathrm{Spin}_n^{F,\mathrm{odd}}$ denote the bordism group of free, odd type circle actions on closed n-dimensional Spin manifolds. Similarly $M\,\mathrm{Spin}_n^{F,\mathrm{even}}$ is the group of bordism classes of free, even type circle actions on closed n-dimensional Spin manifolds.

Proposition 6.1 *Let* $M \in M\,\mathrm{Spin}_n^{F,\mathrm{odd}}$. *Then* $M \times \bar{S} = 0$ *in* $M\,\mathrm{Spin}_{n+1}^{F,\mathrm{odd}}$.

Proof. Let μ: $M\,\mathrm{Spin}_n^{F,\mathrm{odd}} \to M\,\mathrm{Spin}_{n-1}^c$ be the map which sends the bordism class $[M, \alpha]$ (α a free, odd type circle action on M) to $[M/S^1]$. μ is a map of $M\,\mathrm{Spin}_*$-modules, and is an isomorphism, [B, p. 482]. Since $\bar{S}$ bounds as a Spin^c manifold, multiplication by $\bar{S}$ kills $M\,\mathrm{Spin}_n^{F,\mathrm{odd}}$. □

Corollary 6.2 *Let* M *be an* $8k+i$ Spin *manifold,* $i = 1, 2$ *with a free, odd type circle action. Then*

$$\beta_q(M) = 0.$$

Proof. (6.2) is a consequence of (6.1), (5.1) and (5.3). □

Remark. Corollary (6.2) is also a consequence of (7.3) (ii). The only difference is that the manifold with boundary $M \times \bar{S}$ has a free action in (6.1), but only a semi-free action if we use (7.3) (ii).

The action of S^1 on $\bar{S}_0$ has *even type*, so the examples at the end of §4 all have even type actions. Let M be any Spin $8k+2$ manifold with free, even type circle action.

Theorem 6.3 *Suppose M is as above. Then*

$$\beta_q(M) = \beta_q(M/S^1 \times \bar{S}).$$

Theorem (6.3) shows that all manifolds with free S^1 actions which are counter-examples to Hirzebruch's conjecture are of the type described at the end of §4.

Proof. Let σ: $M\,\mathrm{Spin}_n^{F,\mathrm{even}} \to M\,\mathrm{Spin}_{n-1}(\mathbb{C}P^\infty)$ be the map which sends $[M, \alpha]$ to $[M/S^1, f]$, where f: $M/S^1 \to \mathbb{C}P^\infty$ classifies the complex line bundle associated with $M \to M/S^1$. σ is a map of $M\,\mathrm{Spin}_*$ modules and is an isomorphism [B]. Consider the following composite of $M\,\mathrm{Spin}_*$ module isomorphisms

$$\begin{aligned} \mathscr{A}\colon\ M\,\mathrm{Spin}_n^{F,\mathrm{even}} &\xrightarrow{\ \sigma\ } M\,\mathrm{Spin}_{n-1}(\mathbb{C}P^\infty) \\ &\xrightarrow{\ \rho\ } M\,\mathrm{Spin}_{n-1} \oplus \tilde{M}\,\mathrm{Spin}_{n-1}(\mathbb{C}P^\infty) \\ &\xrightarrow{\ 1 \oplus \psi\ } M\,\mathrm{Spin}_{n-1} \oplus M\,\mathrm{Spin}^c_{n-3} \end{aligned}$$

where $\rho[V, f] = [V] \oplus ([V, f] - [V, *])$ ($*$ is the constant map), and ψ is the map in [S, p. 354].

Since $\bar{S}$ kills $M\,\mathrm{Spin}^c_*$, $\mathscr{A}(([M, \alpha] - [M/S^1 \times \bar{S}_0]) \times \bar{S}) = 0$. So $[M - M/S^1 \times \bar{S}_0]$ satisfies the hypothesis of (5.1). Since the action is free, $\beta_q(M - (M/S^1 \times \bar{S})) = 0$. □

Corollary 6.4 *Suppose M is a $16k+10$ Spin manifold with a free circle action. If M/S^1 is* Spin *cobordant to an SU manifold then $\beta_q(M) = 0$.*

Proof. The corollary follows directly from (6.2) if the action has odd type. If the action has even type then we have $\beta_q(M) = \beta_q(M/S^1)\,\eta$. Since M/S^1 is a $16k+9$ SU manifold, it is SU cobordant to $N \times \theta$, where θ is S^1 with the non-trivial SU structure [S, p. 354]. The forgetful functor from SU bordism to Spin bordism maps θ to $\bar{S}$. So $\beta_q(M) = \beta_q(N)\,\eta^2$ where N is a $16k+8$ SU manifold. Since the Pontrjagin character sends the generator to 1 in dimensions congruent to 0 mod 8, it suffices to show that $\Phi_2(N)$ is even. The argument in [C-F, p. 71] applies to $\Phi_2(N)$. To use the argument in [C-F] we must first show that Φ_2 is an integer on W (= bordism of weakly, almost complex manifolds such that all Chern numbers divisible by c_1^2 are zero. This group is denoted $\mathscr{W}(\mathbb{C}, 2)$ in [S]). The KU-characteristic number, $\langle e^{-c_1/2}\,\hat{A}(M)\,ch(s_\omega(\pi)), [M]\rangle$ is an integer if $M \in MU_*$ and $s_\omega(\pi)$ is the s_ω symmetric function for the KU theory Pontrjagin class. If dimension $M = 4k$ and $c_1^2 = 0$, then

$$\langle e^{-c_1/2}\,\hat{A}(M)\,ch(s_\omega(\pi)), [M]\rangle = \langle \hat{A}(M)\,ch(s_\omega(\pi)), [M]\rangle$$

since $\hat{A}(M)\cdot ch(s_\omega(\pi))$ is concentrated in degrees congruent to zero mod 4. In particular, since the complexification of the i-th KO-Pontrjagin class is the i-th KU-Pontrjagin class [S, p. 255], Φ_2 is an integer on W. Now we can define τ' in terms of Φ_2 rather than the signature. It is still the case that one may choose generators z'_{4k} ($k \neq$ a power of 2) for which τ' vanishes. (The notation is that of [C-F].) This follows since Φ_2 is multiplicative with respect to fibrations [O2] and z'_{4k} are defined as the difference of two manifolds, both of which are total spaces of fibrations for which Φ_2 vanishes on the base. □

Corollary 6.5 *Suppose M^{16k+10} is an SU manifold with a free circle action which preserves the SU structure. Then $\hat{A}(M)=0$ and $k(M)=0$.* □

7 Semi-free actions

Let $M\,\mathrm{Spin}_n^{SF,\mathrm{odd}}$ denote the bordism group of odd type, semi-free circle actions on closed Spin manifolds of dimension n. In this section we prove the following theorem.

Theorem 7.1 *Let $M^{8k+2}\in M\,\mathrm{Spin}_{8k+2}^{SF,\mathrm{odd}}$, then $M\times\bar{S}=0$ in $M\,\mathrm{Spin}_{8k+3}^{SF,\mathrm{odd}}$.*

Corollary 7.2 *Let M^{8k+i}, $i=1,2$ admit an odd type, semi-free circle action. Then*

$$\beta_q(M)=0. \quad \square$$

We study the bordism sequence of the pair (SF, F).

$$\ldots \to M\,\mathrm{Spin}_n^{F,\mathrm{odd}} \xrightarrow{\ j\ } M\,\mathrm{Spin}_n^{SF,\mathrm{odd}} \to M\,\mathrm{Spin}_n^{SF,F,\mathrm{odd}} \xrightarrow{\ \partial\ } \ldots$$

The following is proven in [B].

Theorem 7.3 (i) $M\,\mathrm{Spin}_n^{SF,F,\mathrm{odd}} \approx \bigoplus_k \tilde{M}\,\mathrm{Spin}_n(MU(2k+1))$

(ii) *j is trivial.* □

Proof of 7.1 By (7.3) is suffices to show that $\bar{S}$ kills $M\,\mathrm{Spin}_{8k+2}^{SF,F}$. Since the M Spin spectrum, M Spin, decomposes into a product of spectra, $\mathrm{bo}\langle 8k+2\rangle$, $\mathrm{bo}\langle 8k\rangle$ and $K(\mathbb{Z}/2\mathbb{Z})$ we have to understand the Sq^2 cohomology of $H^*(MU(2k+1);\mathbb{Z}/2\mathbb{Z})$. It is a bit easier to assemble the spaces $MU(s)$ together and use Snaith's stable splitting of BU, [Sn].

$$\bigvee_s MU(s)\sim BU.$$

There are stable equivalences

$$BU\sim \mathbb{C}P^\infty \times BSU \sim \mathbb{C}P^\infty \vee (\mathbb{C}P^\infty \wedge BSU)\vee BSU.$$

As in §6, $\bar{S}$ kills $\tilde{M}\,\mathrm{Spin}(\mathbb{C}P^\infty)$ and $\tilde{M}\,\mathrm{Spin}(\mathbb{C}P^\infty \wedge BSU)$, since these two groups are $M\,\mathrm{Spin}^c$-modules. It remains to show that $\bar{S}$ kills $\tilde{M}\,\mathrm{Spin}_{8k+2}(BSU)$.

Let G denote $H^*(BSU;\mathbb{Z}/2\mathbb{Z})$. Then $Sq^1(G)=0$ and Sq^2 is given by the formula

$$Sq^2 w_{2k}=(k-1)\,w_{2k+2}$$

Sq^2 is a derivation on G and a standard argument shows that

$$H^*(G; Sq^2) = \mathbb{Z}/2\mathbb{Z}[w_{4k}^2 | k = 1, 2, \ldots]$$

In particular $H^*(G; Sq^2)$ is concentrated in degrees congruent to 0 mod 8. As a consequence the spectrum M Spin $\wedge$ BSU is, cohomologically, a product of bo$\langle 8k \rangle$'s, bo$\langle 8k+2 \rangle$'s, bu$\langle k \rangle$'s and $K(\mathbb{Z}/2\mathbb{Z})$'s. Hence there is no homotopy in M Spin $\wedge$ BSU in dimension $8k+3$ with Adams filtration greater than zero. Since multiplication with $\bar{S}$ raises Adams filtration by at least 1, 7.1 follows. □

If the action has even type then the relative group in the bordism sequence of (semi-free actions, free actions), is the sum of $\tilde{M}$ Spin$(MU(2k))$'s, and j is not zero. In this case we have proven that $M \times \bar{S}$ is Spin, S^1 equivariantly, cobordant to a manifold, V^{8k+3} with a *free*, even type S^1 action. Proposition (7.3) below is a rather unsatisfactory generalization of (6.3). We first need some notation. Suppose V is an oriented manifold with a smooth, free circle action. Let

$$V' = (V \times D^2)/S^1$$

where the action on $V \times D^2$ is the product of the action on V and the rotation on D^2. V' has a semi-free S^1 action induced by the action on V, and $\partial(V') = V$ as S^1 manifolds. The fixed point set of this action is $V/S^1 \times \{0\} \subset V'$. Let $V/S^1 \cdot V/S^1$ denote the self-intersection of V/S^1 in V'. Define an integer, σ_S, which depends on the action, by

$$\sigma_S(V) = \sigma(V') - \sigma(V/S^1 \cdot V/S^1).$$

Proposition 7.4 *Let M^{8k+2} be a* Spin *manifold with a semi-free, even type circle action. Then there is an* $8k+3$ Spin *manifold, V, with a free, even type circle action such that*

$$8k(M) - 8k(M^I \cdot M^I) = \sigma_S(V) - R(V) \mod 16$$

where $R(V)$ is the generalized Rochlin invariant, (3.3).

Proof. Let X^{8k+4} denote the cobordism from $M \times \bar{S}$ to V. We construct a closed, oriented, manifold W^{8k+4} as follows

$$W = (-V') \bigcup_V (X) \bigcup_{M \times S^1} (-M \times D^2).$$

Then $\sigma(W) = \sigma(X) - \sigma(V')$. One the other hand, $\sigma(W) = \sigma(X^I \cdot X^I) - \sigma(V/S^1 \cdot V/S^1)$ by (4.1). Now V bounds a Spin manifold, Y (not necessarily with S^1 action). Let $Z = X \bigcup_V (-Y)$.

By definition, $8k(M) = \sigma(Z) = \sigma(X) - \sigma(Y) = \sigma(X) - R(V)$ mod 16. We also have $\sigma(X^I \cdot X^I) = 8k(M^I \cdot M^I)$. Eliminating $\sigma(X)$ proves the proposition. □

Notice that 7.4 is indeed a generalization of (6.3) since we can choose V to be $M/S^1 \times \bar{S}_0 \times \bar{S}$.

8 Applications

Let $\mathbf{X}=\mathbf{X}(d_1, \ldots, d_s)\subset \mathbb{C}P^{n+s}$ denote the non-singular hypersurface of multidegree $(d_1, \ldots, d_s)$. Its complex dimension is n. Let $d=d_1 d_2 \ldots d_s$, and $|d|=d_1+\ldots+d_s$. Using the Lefschetz hypersurface section theorem as stated in [A-F] we have:

Proposition 8.1 *Suppose* $n=4m+1$, $n>0$. *Then*

(i) $\mathbf{X}$ *has a* Spin *structure if and only if* $s\equiv|d| \bmod 2$.

(ii) $\mathbf{X}$ *has an SU structure if and only if* $n+1+s=|d|$.

Proof. Part (i) is proven in [O 1, p. 56]. Part (ii) is similar using the integral version of the Lefschetz theorem proven in [A-F]. □

Proposition 8.2 [O 1, p. 56] *Suppose* $\mathbf{X}$ *is* Spin *and* $n=4m+1$. *Then*

(i) *If* $d\equiv 1 \bmod 2$ $k(\mathbf{X})=\dfrac{d^2-1}{8} \bmod 2$.

(ii) *If* $d\equiv 0 \bmod 4$ *then* $k(\mathbf{X})=\dfrac{d(m+1)}{4} \bmod 2$. □

Corollary 8.3 *With* $\mathbf{X}$ *and* n *as in* (8.2).

(i) *If* d *is* odd *and* $\dfrac{d^2-1}{8}$ *is* odd, $\mathbf{X}$ *does not admit a non-trivial* odd *type, semi-free circle action. If in addition* $\mathbf{X}$ *is an SU manifold of dimension* $16k+10$, $\mathbf{X}$ *does not admit a free action preserving the SU structure.*

(ii) *If* $d\equiv 0 \bmod 4$ *and* $\dfrac{d(m+1)}{4}$ *is* odd, $\mathbf{X}$ *does not admit a non-trivial* odd *type, semi-free circle action. If in addition* $\mathbf{X}$ *is an SU manifold of dimension* $16k+10$, $\mathbf{X}$ *does not admit a free action preserving the SU structure.* □

The results in §8 are obtained by looking at the top term in (3.3). If our conjecture on the vanishing of β_q were true, then the KO numbers $a_i[M]$, $i\geqq 0$, would be a complete list of obstructions for M to be Spin bordant, up to Spin manifolds with all KO-characteristic numbers zero, to a manifold with semi-free, odd type circle action. The first few obstructions are:

$$1[M],\quad \pi_1[M],\quad (\pi_2+\pi_1)[M] \quad \text{and} \quad (\pi_3+\pi_1^2)[M].$$

Acknowledgements. I wish to thank A. Bahri and P. Landweber for many useful conversations. I am especially indebted to S. Ochanine for explaining this genus to me and for many helpful suggestions.

References

[A-F] Andreotti, A., Frankel, T.: The Lefschetz theorem on hyperplane sections. Ann. Math. **69**, 713–717 (1959)

[A] Atiyah, M.F.: Riemann surfaces and spin structures. Ann. Sci. Éc. Norm. Supér., IV Sér. **4**, 47–62 (1971)

[A-H] Atiyah, M.F., Hirzebruch, F.: Riemann-Roch theorems for differentiable manifolds. Bull. Am. Math. Soc. **65**, 276–281 (1959)

[A-S] Atiyah, M.F., Singer, I.M.: The index of elliptic operators: III. Ann. Math. **87**, 546–604 (1968)
[B] Borsari, L.D.: Bordism of semifree circle actions on Spin manifolds. Trans. Am. Math. Soc. **301**, 479–487 (1987)
[B-T] Bott, R., Taubes, C.: On the rigidity theorems of Witten. J.A.M.S. **2**, 137–186 (1989)
[Br] Brown, E.H.: Generalizations of the Kervaire invariant. Ann. Math. **95**, 368–383 (1972)
[C] Chudnovsky, D.V., Chudnovsky, G.V., Landweber, P.S., Ochanine, S., Stong, R.E.: Integrability and divisibility of elliptic genera (to appear)
[C-F] Conner, P.E., Floyd, E.E.: Torsion in SU-bordism. Mem. Am. Math. Soc. **60** (1966)
[H] Hirzebruch, F.: Involutionen auf Mannigfaltigkeiten. Proceedings, Conf. on Trans. Groups, Tulane 1967. Berlin Heidelberg New York: Springer 1968
[H1] Hirzebruch, F.: Involutions on manifolds, the signature and elliptic genera. Rutgers University (Lecture) May 18, 1989
[H-S] Hirzebruch, F., Slodowy: (to appear)
[L] Landweber, P.S.: Elliptic cohomology and modular forms. In: Landweber, P.S. (ed.) Elliptic Curves and Modular Forms in Algebraic Topology. Proceedings, Princeton 1986. (Lect. Notes Math. vol. 1326, pp. 55–68) Berlin Heidelberg New York: Springer 1988
[L-S] Landweber, P.S., Stong, R.E.: Circle actions on Spin manifolds and characteristic numbers. Topology **27**, 145–162 (1988)
[O1] Ochanine, S.: Signature modulo 16, invariants de Kervaire généralisés, et nombres charactéristiques dans la K-théorie réelle. Bull. Soc. Math. Fr. **109**, Memoire 5, Suppl. (1981)
[O2] Ochanine, S.: Sur les genres multiplicatifs définis par des intégrales elliptiques. Topology **26**, 143–151 (1987)
[O3] Ochanine, S.: Elliptic genera, modular forms over KO_*, and the Brown-Kervaire invariant. Math. Z. **206**, 277–291 (1991)
[Sn] Snaith, V.: Towards algebraic cobordism. Bull. Am. Math. Soc. **83**, 384–385 (1977)
[S] Stong, R.E.: Notes on Cobordism Theory. Princeton: Princeton University Press 1968
[W] Witten, E.: Elliptic genera and quantum field theory. Commun. Math. Phys. **109**, 525–536 (1987)
[Z] Zagier, D.: Note on the Landweber-Stong elliptic genus. In: Landweber, P.S. (ed.) Elliptic Curves and Modular Forms in Algebraic Topology. Princeton 1986. (Lect. Notes Math. vol. 1326, pp. 216–224) Berlin Heidelberg New York: Springer 1988

Math. Z. 206, 457–471 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Einstein Metrics on solvable groups

T.H. Wolter
Mathematisches Institut, Universität Zürich, CH-8001 Zürich, Switzerland

Received April 24, 1990; in final form May 22, 1990

Introduction

All symmetric spaces of noncompact type with metric induced by the Killing form have constant Ricci curvature. By the Iwasawa decomposition theorem, these spaces can be identified with solvable groups $\mathcal{S}$ having the following characteristic properties:

Definition. A simply connected solvable group $\mathcal{S}$ is of *Iwasawa type*, if its Lie algebra $\mathfrak{s}$ satisfies the conditions

(1) $\mathfrak{s}=\mathfrak{a}+\mathfrak{n}$, where $\mathfrak{n}=[\mathfrak{s},\mathfrak{s}]$ and $\mathfrak{a}$ is abelian.
(2) All operators ad_A, $A\in\mathfrak{a}$ are semisimple (i.e. diagonalizable over $\mathbb{C}$).
(3) There is an $A^0\in\mathfrak{a}$ such that $\mathrm{ad}_{A^0}|_{\mathfrak{n}}$ has eigenvalues with strictly positive real parts.

This class contains plenty of examples apart from those that arise as Iwasawa factors of symmetric spaces. Take for example any nilpotent algebra $\mathfrak{n}$ that admits a semisimple derivation D with eigenvalues of positive real part; define $\mathfrak{s}=\mathbb{R}A+\mathfrak{n}$ such that $\mathrm{ad}_A|_{\mathfrak{n}}=D$, clearly $\mathfrak{s}$ is of Iwasawa type.

In this paper we investigate when groups of Iwasawa type admit left invariant Einstein metrics. It turns out, that this depends chiefly on the eigenvalues of the Ricci form of $\mathfrak{n}$ relative to the induced metric. We point out a general method to construct new examples and apply it to several concrete situations.

Results of R. Azencott and E. Wilson imply that groups of Iwasawa type always admit left invariant metrics of nonpositive sectional curvature [4]. Thus one may ask for metrics having both properties. Employing a systematic approach, we rediscover all known homogeneous Einstein manifolds of nonpositive sectional curvature and in addition two new infinite series of (pairwise non isometric) examples.

Finally we investigate a certain tensorial equation for the Ricci curvature. If $\mathcal{S}$ is an Iwasawa group with a left invariant metric such that all geodesic symmetries are volume preserving, we can deduce that the metric is Einstein on every de Rham factor.

1 Einstein metrics

From now on $\mathfrak{s}$ will in general denote the Lie algebra of a group of Iwasawa type. A scalar product on $\mathfrak{s}$ induces a left invariant metric on the group $\mathcal{S}$. Two metric Lie algebras are called isometric, if the corresponding simply connected Lie groups are isometric as Riemannian manifolds.

1.1 Proposition. *Let* $\mathfrak{s}$ *carry a scalar product where* $\mathfrak{a}$ *and* $\mathfrak{n}$ *are orthogonal and all operators* ad_A, $A \in \mathfrak{a}$ *are normal, then it is isometric to a Lie algebra* $\mathfrak{s}'$, *where all maps* $\mathrm{ad}_{A'}$, $A' \in \mathfrak{a}'$ *are symmetric.*

Proof. Since ad_A is normal, its symmetric and skew-symmetric parts are also derivations of $\mathfrak{n}$ [15, Lemma 1]. Define a new Lie algebra $\mathfrak{s}'$, where the operation of $\mathfrak{a}$ on $\mathfrak{n}$ is given by the symmetric parts. From [5, Lemma 4.2] one concludes, that $\mathfrak{s}'$ is isometric to $\mathfrak{s}$. □

Note that in general, $\mathfrak{s}'$ will not be isomorphic to $\mathfrak{s}$. Prop. 1.1 motivates the next definition:

1.2 Definition. A solvable Lie algebra $\mathfrak{s}$ with scalar product is called a *metric Lie algebra of Iwasawa type*, if

(1) $\mathfrak{s}$ is the orthogonal sum of $\mathfrak{a}$ and $\mathfrak{n}$, $\mathfrak{a}$ abelian.
(2) All operators ad_A, $A \in \mathfrak{a}$ are symmetric.
(3) For some $A^0 \in \mathfrak{a}$, $\mathrm{ad}_{A^0}|_{\mathfrak{n}}$ has positive eigenvalues.

1.3 Proposition. *Two metric algebras of Iwasawa type are isometric iff they are isomorphic (i.e. there is a Lie algebra automorphism which preserves the scalar products).*

Proof. [1, Theor. 1].

For any (orthonormal) left invariant vector fields X, Y of a metric Lie algebra $\mathfrak{s}$, the Levi-Civita connection and sectional curvature can be computed as [2, p. 92]

$$(1.A) \qquad 2\langle \nabla_X Y, Z\rangle = \langle [X, Y], Z\rangle - \langle [Y, Z], X\rangle - \langle [X, Z], Y\rangle,$$

$$K(X \wedge Y) = |\nabla_X Y|^2 - \langle \nabla_X X, \nabla_Y Y\rangle - \langle \mathrm{ad}_Y^2 X, X\rangle - |[X, Y]|^2.$$

Suppose that $\mathfrak{s}$ is of Iwasawa type: $\mathfrak{a}$ operates on $\mathfrak{n}$ by commuting symmetric maps, so $\mathfrak{n}$ has an orthogonal decomposition into root spaces $\mathfrak{n}_\alpha$, where $\alpha \in \mathfrak{a}^*$ such that $\mathrm{ad}_A X_\alpha = \alpha(A) X_\alpha \forall X_\alpha \in \mathfrak{n}_\alpha$. Relative to the scalar product we define dual root vectors A_α in $\mathfrak{a}$. Let $A, A' \in \mathfrak{a}$, $X, Y \in \mathfrak{n}$; $\nabla^{\mathfrak{n}}$ be the induced connection on the subgroup $\mathcal{N}$, then

$$(1.B) \qquad \nabla_A A' = 0, \quad \nabla_A X = 0, \quad \nabla_X A = -\mathrm{ad}_A X$$

$$\nabla_X Y = \sum_\alpha \langle X_\alpha, Y_\alpha\rangle A_\alpha + \nabla^{\mathfrak{n}}_X Y.$$

For the sectional curvature we obtain (all vectors are assumed to be orthonormal)

$$(1.C) \qquad K(A, A') = 0, \quad K(A, X) = -|\mathrm{ad}_A X|^2,$$

$$K(X, Y) = K^{\mathfrak{n}}(X, Y) - \sum (|X_\alpha|^2 |Y_\beta|^2 - \langle X_\alpha, X_\beta\rangle^2)\langle A_\alpha, A_\beta\rangle.$$

Let H be the mean curvature vector of $\mathcal{N}$ in $\mathcal{S}$; from (1.B) we conclude that $H = \sum \dim \mathfrak{n}_\alpha \cdot A_\alpha$, hence $\langle H, A\rangle = \mathrm{tr}\, \mathrm{ad}_A$ for all $A \in \mathfrak{a}$. Choose an orthonormal basis $A_1 = \dfrac{H}{|H|}, \ldots, A_r$ in $\mathfrak{a}$.

1.4 Lemma. *Let* $\mathrm{ric}^{\mathfrak{n}}$ *be the Ricci curvature of* $\mathfrak{n}$, *then*

(1) $\mathrm{ric}(A, A') = -\operatorname{tr} \mathrm{ad}_A \mathrm{ad}_{A'}$
(2) $\mathrm{ric}(A, X) = 0$
(3) $\mathrm{ric}(X) = \mathrm{ric}^{\mathfrak{n}}(X) - \langle \mathrm{ad}_H X, X\rangle = \mathrm{ric}_{\mathfrak{n}}(X) - \operatorname{tr} \mathrm{ad}_{A_1} \cdot \langle \mathrm{ad}_{A_1} X, X\rangle$.

Proof. (1) $R(X, A)A' = -\mathrm{ad}_A \mathrm{ad}_{A'} X$.
(2) $\mathrm{ric}(A, X) = \operatorname{tr} V(\mathrm{ad}_A X) - \operatorname{tr} \mathrm{ad}_A V_X - \operatorname{tr} \mathrm{ad}_A \mathrm{ad}_X$. Define inductively $\mathfrak{n}^1 = \mathfrak{n}$, $\mathfrak{n}^k = [\mathfrak{n}, \mathfrak{n}^{k-1}]$. If $Y \in \mathfrak{n}^{k-1}$ is orthogonal to $\mathfrak{n}^k$, $\langle V_Y Z, Y\rangle = 0$ for all $Z \in \mathfrak{n}$. Choosing a suitable basis for $\mathfrak{n}$ we see that $\operatorname{tr} V \mathrm{ad}_A X = 0$. The second term vanishes because ad_A is symmetric and V_X skew-symmetric. By Lie's theorem there is a basis of $\mathfrak{n}$ such that all operators ad_Z, $Z \in \mathfrak{s}$ are represented by upper triangular matrices. As the matrix of ad_X is nilpotent, we conclude that $2 \operatorname{tr} \mathrm{ad}_A \mathrm{ad}_X = \operatorname{tr}(\mathrm{ad}_A + \mathrm{ad}_X)^2 - \operatorname{tr} \mathrm{ad}_A^2 - \operatorname{tr} \mathrm{ad}_X^2 = 0$. Since $V_X Y = V_X^{\mathfrak{n}} Y + \sum \langle \mathrm{ad}_{A_i} X, Y\rangle A_i$, we have

$$\mathrm{ric}(X) = -\sum_1^r |\mathrm{ad}_{A_i} X|^2 + \mathrm{ric}^{\mathfrak{n}}(X) - \sum_1^r \operatorname{tr} \mathrm{ad}_{A_i} \cdot \langle \mathrm{ad}_{A_i} X, X\rangle - |\mathrm{ad}_{A_i} X|^2. \tag{3}$$

For $i \neq 1$, $\operatorname{tr} \mathrm{ad}_{A_i} = \langle A_i, H\rangle = 0$. □

If the metric is Einstein, we deduce from 1.4 (3) that $\mathrm{ad}_{A_1}, \ldots, \mathrm{ad}_{A_r}$ and $\mathrm{ric}^{\mathfrak{n}}$ have common eigenvectors $X_1, \ldots, X_n$ with eigenvalues $\lambda_{1,i}, \ldots, \lambda_{r,i}$ and v_i respectively ($i = 1, \ldots, n$).

1.5 Proposition. *Under the above assumptions, the metric is Einstein with constant* $-c$ iff

(i) $\lambda_{1,i} = \dfrac{v_i + c}{d}$, $d = \sqrt{\sum \dfrac{(v_j + c)^2}{c}}$
(ii) $\sum v_i (v_i + c) = 0$
(iii) $\sum_i \lambda_{1,i}^2 = \ldots = \sum_i \lambda_{r,i}^2 = c$
(iv) $\sum_i \lambda_{s,i} \lambda_{t,i} = 0$, $1 \leqq s < t \leqq r$.

Proof. 1.4 (3) implies $v_i = (\sum \lambda_{1,j}) \lambda_{1,i} - c$. As

$$\sum (v_i + c)^2 = (\sum \lambda_{1,j})^2 \sum \lambda_{1,i}^2 = (\sum \lambda_{1,j})^2 \cdot c$$

we have $\sum \lambda_{1,j} = d$ and $v_i = d\lambda_{1,i} - c$. From $d = \sum \dfrac{v_i + c}{d}$ we derive (ii). The equation $\operatorname{tr} \mathrm{ad}_A \mathrm{ad}_{A'} = c \cdot \langle A, A'\rangle$ yields (iii) and (iv).

Now assume, that (i)–(iv) are fulfilled: $\operatorname{tr} \mathrm{ad}_{A_1} = \sum \dfrac{v_i + c}{d} = d$, so $\mathrm{ric}(X_i) = v_i - d\lambda_{1,i} = -c$. (iii) and (iv) give $\mathrm{ric}(A) = -c|A|^2$, by 4.7 (2) $\mathrm{ric}(A, X) = 0$. □

These equations can be used to construct new Einstein manifolds: Let $\mathfrak{n}$ be a metric nilpotent Lie algebra, the Ricci form shall be diagonalized for the vectors $X_1, \ldots, X_n$ with eigenvalues $v_1, \ldots, v_n$. For any $c > 0$, define a map D on $\mathfrak{n}$ such that $DX_i = \lambda_i X_i$, where the λ_i are given by 1.5 (i).

1.6 Theorem. *Suppose there is a constant c with*

$$v_k - v_i - v_j = c \quad \text{if } \langle [X_i, X_j], X_k\rangle \neq 0 \tag{1.D}$$
$$\text{and} \quad \sum v_i (v_i + c) = 0.$$

Define $\mathfrak{s} = \mathbb{R}A + \mathfrak{n}$ *such that* $\mathrm{ad}_A|_{\mathfrak{n}} = D$, *then* $\mathfrak{s}$ *is a solvable Lie algebra that carries a metric with* $\mathrm{ric}(Z) = -c|Z| \; \forall Z \in \mathfrak{s}$.

Proof. From the first equation in (1.D) we conclude that $\lambda_i+\lambda_j=\lambda_k$ if the structural constant $\alpha_{ijk}\neq 0$, so D is a derivation of $\mathfrak{n}$. Extend the scalar product of $\mathfrak{n}$ in such a way that A is a unit vector perpendicular to $\mathfrak{n}$. As $\sum\lambda_i^2=c$, Proposition 1.5 implies that the metric is Einstein. □

For certain $\mathfrak{n}$ one can also construct examples with $\dim\mathfrak{a}>1$: Choose $r-1$ vectors $(\lambda_{s,1}, \ldots, \lambda_{s,n})$, $1<s\leqq r$ which are pairwise orthogonal and perpendicular to $(\lambda_1, \ldots, \lambda_n)$.

1.7 Corollary. *If* $\lambda_{s,i}+\lambda_{s,j}=\lambda_{s,k}$ *whenever* $\alpha_{ijk}\neq 0$, $\sum\lambda_{s,i}=0$ *and* $\sum\lambda_{s,i}^2=c$ *for* $2\leqq s\leqq r$,

$$\mathfrak{s}=\mathfrak{a}+\mathfrak{n}, \quad \mathfrak{a}=\operatorname{span}\{A_1, \ldots, A_r\} \quad \text{with} \quad \operatorname{ad}_{A_s}X_i=\lambda_{s,i}X_i$$

carries an Einstein metric (with constant $-c$*).*

Proof. Define a scalar product on $\mathfrak{a}$ by

$$\langle A_s, A_t\rangle=\frac{1}{c}\operatorname{tr}\operatorname{ad}_{A_s}\operatorname{ad}_{A_t},$$

then $A_1, \ldots, A_r$ will be orthonormal, and for $H=d\cdot A_1$ we have

$$\langle H, A_s\rangle=0=\operatorname{tr}\operatorname{ad}_{A_s} \quad \text{for } s\neq 1, \quad \langle H, A_1\rangle=d=\operatorname{tr}\operatorname{ad}_{A_1}.$$

Hence we are in the situation of Proposition 1.5. □

1.8 Low dimensional examples. For nilpotent Lie algebras of small dimension, there is a general procedure to determine all metrics satisfying the equations of Theorem 1.6. Tables of all nilpotent algebras up to dimension 7 can be found in [3, 18].

Take e.g. $\mathfrak{n}=\operatorname{span}\{X_1, X_2, X_3, X_4, X_5\}$ with $[X_1, X_2]=X_4$, $[X_1, X_3]=X_5$. Choose some scalar product on $\mathfrak{n}$. We can assume that X_1, X_2, X_3 are orthogonal to the center. Now apply the orthonormalization process to X_2 and X_1 to obtain vectors X_1', X_2' with $[X_2', X_3]=0$ and $[X_1', X_2']$, $[X_1', X_3]$ linearly independent. By scaling the metric we achieve that $X_4'=[X_1', X_2']$ is a unit vector. Choose X_5' perpendicular to X_4', then $[X_1', X_3]=\varepsilon X_4'+\delta X_5'$ with $\delta\neq 0$. Changing the directions of X_3' and X_5' if necessary, we can assume that $\varepsilon\geqq 0$ and $\delta>0$.

Denote this orthonormal basis again by $X_1, \ldots, X_5$. For nilpotent groups, the Ricci curvature operator can be computed as [2, p. 106]

$$\tfrac{1}{4}\sum\operatorname{ad}_{X_i}\operatorname{ad}_{X_i}^t-\tfrac{1}{2}\sum\operatorname{ad}_{X_i}^t\operatorname{ad}_{X_i}.$$

We obtain for $2\cdot\operatorname{ric}^{\mathfrak{n}}$ the matrix

$$\begin{pmatrix} -(1+\varepsilon^2+\delta^2) & & & & \\ & -1 & -\varepsilon & & \\ & -\varepsilon & -(\varepsilon^2+\delta^2) & & \\ & & & 1+\varepsilon^2 & \varepsilon\delta \\ & & & \varepsilon\delta & \delta^2 \end{pmatrix}$$

If the metric on $2c$ is induced by an Einstein metric, there has to be a derivation D of $\mathfrak{n}$ with $\operatorname{ric}^{\mathfrak{n}}(X_i, X_j)=d\cdot\langle DX_i, X_j\rangle$ for $i\neq j$, $d=\operatorname{tr}D>0$, cf. 1.4(3). Represent

D by the matrix $((a_{ij}))$, then $a_{23}=-d\varepsilon$ and $a_{45}=d\varepsilon\delta$. The equation $DX_4=D[X_1, X_2]=[DX_1, X_2]+[X_1, DX_2]$ however implies that $a_{45}=\delta a_{23}$, hence $\varepsilon=0$ and $\mathrm{ric}^{\mathfrak{n}}$ must be diagonal with eigenvalues $-\frac{1+\delta^2}{2}, -\frac{1}{2}, -\frac{\delta^2}{2}, \frac{1}{2}, \frac{\delta^2}{2}$. It is easy to see that the equations (1.D) are fulfilled iff $\delta=1$, $c=2$. The canonical metric on $\mathfrak{s}=\mathbb{R}A+\mathfrak{n}$, where $\mathrm{ad}_A|_{\mathfrak{n}}$ has eigenvalues $\frac{1}{3}, \frac{1}{2}, \frac{1}{2}, \frac{5}{6}, \frac{5}{6}$, is of constant Ricci curvature -2, cf. 1.6.

The derivations with eigenvalues $0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}$ and $0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$ satisfy the conditions of Corollary 1.7, so we can define an Einstein metric on $\mathfrak{s}'=\mathfrak{a}'+\mathfrak{n}$ where $\dim \mathfrak{a}'=3$.

1.9 Infinite series. Following class of Lie algebras has been defined by E. Deloff [11, p. 85]:

Definition. A two step nilpotent algebra $\mathfrak{n}$ is *uniform* of type (m, n, r), if there are basis vectors $\{V_1, \ldots, V_n, Z_1, \ldots, Z_m\}$ and numbers r, s such that

(1) $[V_i, V_j]\in\{0, \pm Z_1, \ldots, \pm Z_m\}$, $[V_i, Z_l]=[Z_k, Z_l]=0$
(2) If $[V_i, V_j]=\pm[V_i, V_k]\neq 0$, then $V_j=V_k$
(3) For every Z_l there exist exactly r disjoint pairs $\{V_i, V_j\}$ with $[V_i, V_j]=Z_l$
(4) For every V_i there exist exactly s V_j with $[V_i, V_j]\neq 0$.

The number s is determined by m, n and r, since the number of non-vanishing brackets $[V_i, V_j]$ is just $sn=2rm$.

Example of an uniform algebra of type $\left(\frac{q(q-1)}{2}, pq, p\right)$:

$$\mathfrak{n}=\mathrm{span}\{V_{i\mu}: 1\leqq i\leqq p, 1\leqq\mu\leqq q\}+\mathrm{span}\{Z_{\mu\nu}: 1\leqq\mu<\nu\leqq q\}$$
$$\text{with}\quad [V_{i\mu}, V_{j\nu}]=\begin{cases} Z_{\mu\nu} & \text{for } i=j, \mu<\nu \\ 0 & \text{for } i\neq j. \end{cases}$$

If $q=2$, we obtain the Lie algebra of the Heisenberg group.

In general, it is difficult to decide for which numbers m, n, r with $\frac{2rm}{n}\in\mathbb{N}$ a uniform algebra of type (m, n, r) exists; the case $2r=n$ will be discussed below.

The basis given above defines a scalar product on $\mathfrak{n}$. It turns out that this basis diagonalizes the Ricci tensor:

1.10 Lemma [11, Lemma 4.4]. *The Ricci tensor of an uniform Lie algebra of type (m, n, r) has eigenvalues*

$$v_i=\mathrm{ric}(V_i)=-\frac{s}{2},\ i=1, \ldots, n; \qquad v_{n+k}=\mathrm{ric}(Z_k)=\frac{r}{2},\ k=1, \ldots, m.$$

For this metric, the first equation of (1.D) is fulfilled with $c=\frac{r}{2}+s$. As $v_i+c=\frac{1}{2}(r+s)$, $v_{n+k}+c=r+s$,

$$\sum_1^n v_i(v_i+c)+\sum_1^m v_{n+k}(v_{n+k}+c)=-\frac{ns}{4}(r+s)+\frac{mr}{2}(r+s)=0.$$

So we can apply the formulas of Proposition 1.5:

We have $d^2=\frac{2}{r+2s}\left(\frac{n}{4}+m\right)(r+s)^2=\left(\frac{r+s}{2}\right)^2\frac{2n}{r}$,

(1.E) $$\lambda_i=\sqrt{\frac{r}{2n}},\ i=1,\ldots,n;\qquad \lambda_{n+k}=\sqrt{\frac{2r}{n}},\ k=1,\ldots,m.$$

1.11 Modifications. (1) Assume that $\mathfrak{n}$ carries a metric for which (1.D) is valid. Let $\mathfrak{a}_k$ be an euclidean k-space (with bracket $=0$), then $\mathfrak{n}'=\mathfrak{n}+\mathfrak{a}_k$ also fulfills (1.D). The eigenvalues of the derivation D on $\mathfrak{n}'$ can be computed from 1.5(i), using $d'=\sqrt{d^2+kc}$.

(2) Let $\mathfrak{n}_1, \mathfrak{n}_2$ carry metrics with property (1.D). By scaling we can achieve that $c_1=c_2$. Their direct sum $\mathfrak{n}=\mathfrak{n}_1+\mathfrak{n}_2$ also satisfies (1.D), $d=\sqrt{d_1^2+d_2^2}$ in this case.

(3) Suppose that $\mathfrak{s}=\mathfrak{a}+\mathfrak{n}$ with $\dim\mathfrak{a}>1$ carries an Einstein metric. Let $\mathfrak{a}'$ be a subspace in $\mathfrak{a}$ that contains the mean curvature vector H of $\mathcal{N}$; then the induced metric on $\mathfrak{s}'=\mathfrak{a}'+\mathfrak{n}$ is again Einstein.

Take for example the symmetric space $M=SL(n,\mathbb{R})/SO(n,\mathbb{R})$ with metric induced by $\operatorname{tr}A\cdot B$ [14, p. 142]. M is Einstein with $c=\frac{n}{2}$; the Lie algebra of the nilpotent part in the Iwasawa decomposition is spanned by the canonical basis vectors E_{ij}, $i<j$. For the mean curvature vector we have $\operatorname{ad}_H E_{ij}=(j-i)E_{ij}$ and $|H|=(n-1)\sqrt{\frac{n}{3}}$. Define $\mathfrak{s}'=\mathbb{R}A+\mathfrak{n}$, where $\mathfrak{n}$ is the Lie algebra of upper triangular matrices with zero diagonal and $\operatorname{ad}_A=\frac{1}{|H|}\operatorname{ad}_H$. If $n=3$, we obtain complex hyperbolic space of dimension 4; for $n\geqq 4$, $\mathcal{S}'$ is not symmetric.

Remarks. (1) The Lie groups constructed in Theorem 1.6 and Corollary 1.7 are not necessarily of Iwasawa type, condition (3) may not be fulfilled. Nevertheless, the formulas of Lemma 1.4 are still valid since their proof only made use of properties (1) and (2). If we require in addition that all vectors $(\lambda_{1,i},\ldots,\lambda_{r,i})$, $i=1,\ldots,n$ are contained in an upper half space of $\mathfrak{a}$, $\mathcal{S}$ will be of Iwasawa type. For $\dim\mathfrak{a}=1$ this means $v_i+c>0\ \forall i$.

(2) Nilpotent or more general unimodular solvable groups do not admit left invariant Einstein metrics (not even metrics of negative Ricci curvature), cf. [13]. Furthermore, there are nilpotent Lie algebras that do not carry metrics with property (1.D): Dixmier-Lister defined an eight-dimensional nilpotent algebra, where all derivations are nilpotent endomorphisms [12], but a derivation D constructed in the fashion of Theorem 1.6 would have at least one eigenvalue $\neq 0$. Does every two step nilpotent algebra admit some scalar product with property (1.D)?

(3) Given a fixed nilpotent Lie algebra, one can use the procedure of 1.8 to determine all left invariant metrics satisfying the equations (1.D). In all examples we checked so far, the solution was unique (up to some constant factor). However, for certain 6- or 7-dimensional algebras there might be one-parameter families of non isometric solutions.

(4) Homogeneous bounded domains with their Bergmann metric can be identified with certain solvable groups which are of Iwasawa type [20], carrying

left invariant Kähler-Einstein metrics. On the other hand, all homogeneous Kähler manifolds with negative Ricci curvature are biholomorphically equivalent to bounded homogeneous domains [10, p. 17]; if the metric is Einstein, it must be the Bergmann metric [8]. So our methods do not yield new Kähler-Einstein manifolds.

2 Examples with nonpositive sectional curvature

All symmetric spaces of noncompact type with metric induced by the Killing form are Einstein manifolds with nonpositive sectional curvature. D. Alekseevskii classified in [2] all 5-dimensional homogeneous Einstein manifolds of nonpositive sectional curvature and discovered one non-symmetric example. E. Deloff constructed infinite series, which are generalizations of this 5-dimensional example, cf. [11] and 2.6 below. In 1985, J. Boggino (a student of I. Dotti Miatello) came accidentally upon a new class of examples: He modified complex hyperbolic space by taking groups of Heisenberg type (defined by A. Kaplan) as nilpotent parts and computed their Ricci and sectional curvature, cf. [7] and 2.3.

In the last paragraph we sketched a general method to find manifolds with constant negative Ricci curvature, yet most of the examples constructed this way have planes with positive sectional curvature. This can be explained as follows: A metric Lie algebra $\mathfrak{s}=\mathbb{R}A+\mathfrak{n}$ of Iwasawa type has negative curvature, if the eigenvalues of ad_A are large compared to the sectional curvature of $\mathcal{N}$ [15, p. 26]. Equation 1.5(i) however implies, that the smallest eigenvalue of ad_A is $\leqq\sqrt{\frac{c}{n}}$.

So far we have obtained only two new series of Einstein manifolds with nonpositive curvature cf. 2.6 and 2.8. None of these examples are isometric by Proposition 1.3.

To start with, we investigate the simplest case where the nilpotent part $\mathfrak{n}$ of $\mathfrak{s}$ is abelian. Following Proposition has been proved by D. Alekseevskii in [2, Proposition 4.2] for $\dim\mathfrak{a}=1$ and 2.

2.1 Proposition. *Let $\mathcal{S}$ carry an Einstein metric of nonpositive sectional curvature. If the nilpotent part $\mathfrak{n}$ of $\mathfrak{s}$ is abelian, $\mathcal{S}$ splits into the product of r real hyperbolic spaces, $r=\dim\mathfrak{a}$.*

Proof. By 1.4(3) $\langle \mathrm{ad}_H X, X\rangle = c|X|^2$ for all $X\in\mathfrak{n}$, hence $\mathrm{ad}_H|_{\mathfrak{n}} = c\cdot Id$. If A_α is a root vector, we have $\mathrm{tr}\,\mathrm{ad}_{A_\alpha}=\langle H, A_\alpha\rangle=\alpha(H)=c$. 1.4(1) implies $\mathrm{tr}\,\mathrm{ad}^2_{A_\alpha/|A_\alpha|}=c$, thus

$$\mathrm{tr}\,\mathrm{ad}_{A_\alpha}=\sum_\beta\langle A_\alpha, A_\beta\rangle\dim\mathfrak{n}_\beta=\mathrm{tr}\,\mathrm{ad}^2_{A_\alpha/|A_\alpha|}=\sum_\beta\left\langle\frac{A_\alpha}{|A_\alpha|}, A_\beta\right\rangle^2\dim\mathfrak{n}_\beta$$

$$\Leftrightarrow\sum_\beta\left(1-\left\langle\frac{A_\alpha}{|A_\alpha|},\frac{A_\beta}{|A_\alpha|}\right\rangle\right)\langle A_\alpha, A_\beta\rangle\dim\mathfrak{n}_\beta=0.$$

Take $X\in\mathfrak{n}_\alpha$, $Y\in\mathfrak{n}_\beta$. If $\mathfrak{n}$ is abelian, formula (1.C) simplifies to $K(X\wedge Y)=-\langle A_\alpha, A_\beta\rangle|X\wedge Y|^2$, which gives $\langle A_\alpha, A_\beta\rangle\geqq 0$. Assume that A_α is a root vector of maximal length. From Schwarz' inequality we conclude

$$\left\langle\frac{A_\alpha}{|A_\alpha|},\frac{A_\beta}{|A_\alpha|}\right\rangle<1\quad\text{for all }\beta\neq\alpha.$$

So the above sum can vanish only if $\langle A_\alpha, A_\beta\rangle = 0$ for all $\beta \neq \alpha$. $\mathfrak{s}$ decomposes into the orthogonal sum of ideals

$$\mathfrak{s} = \mathfrak{s}_1 + \mathfrak{s}_2 = (\mathbb{R}A_\alpha + \mathfrak{n}_\alpha) + (\operatorname{span}_{\beta\neq\alpha}\{A_\beta\} + \sum_{\beta\neq\alpha}\mathfrak{n}_\beta),$$

and S splits isometrically into $S_1 \times S_2$, where S_1 is a real hyperbolic space. The assertion follows by induction. □

Now we consider metric Lie algebras of Iwasawa type where $\mathfrak{n}$ is a uniform two step algebra. Let $\mathfrak{v} = \operatorname{span}\{V_1, \ldots, V_n\}$, $\mathfrak{z} = \operatorname{span}\{Z_1, \ldots, Z_m\}$. Define the map $j: \mathfrak{z} \to \operatorname{Alt}(\mathfrak{v})$ by

$$\langle j(Z)V, W\rangle = \langle [V, W], Z\rangle \qquad V, W \in \mathfrak{v},\ Z \in \mathfrak{z}.$$

2.2 Lemma. *Suppose $n = 2r$, then j has the properties*

(i) $j(Z_k)V_i \in \{\pm V_1, \ldots, \pm V_n\}$
(ii) $\langle j(Z_k)V_i, j(Z_l)V_i\rangle = \langle Z_k, Z_l\rangle$, $k \neq l$ *and* $j(Z_k)^2 V_i = -V_i$.

Proof. By Definition 1.9(2) and (3) there exists for V_i and Z_k exactly one V_j with $[V_i, V_j] = \pm Z_k$, thus $j(Z_k)V_i = \pm V_j$. 1.9(1) implies $[V_i, j(Z_k)V_i] \in \{\pm Z_1, \ldots, \pm Z_m\}$; since $\langle [V_i, j(Z_k)V_i], Z_k\rangle = |j(Z_k)V_i|^2 = 1$ we conclude $[V_i, j(Z_k)V_i] = Z_k$. Because of $[j(Z_k)V_i, j(Z_k)^2 V_i] = Z_k$, we have $j(Z_k)^2 V_i = -V_i$. □

If in addition $j(Z_k)j(Z_l)V_i = -j(Z_l)j(Z_k)V_i$ for all $Z_k \neq Z_l$, j can be extended to a representation of the Clifford algebra $Cl(\mathfrak{z})$ on $\mathfrak{v}$, whence $\mathfrak{n}$ is a so called Lie algebra of Heisenberg type, cf. [16]. Conversely, if $\mathfrak{n}$ is of Heisenberg type with rational structural constants, there are maximal lattices $L \subset \mathfrak{v}$, $L' \subset \mathfrak{z}$ such that $j(L')L \subset L$, so $\mathfrak{n}$ is uniform. Lie algebras of Heisenberg type exist for all numbers m, n with $m \leqq \rho(n)$, where ρ is the Hurwitz-Radon function [16, p. 128].

Using (1.B) we compute the covariant derivatives of $\mathfrak{s} = \mathbb{R}A + \mathfrak{n}$, where $\mathfrak{n}$ is of Heisenberg type and $\operatorname{ad}_A V = \lambda V$, $V \in \mathfrak{v}$, $\operatorname{ad}_A Z = 2\lambda Z$, $Z \in \mathfrak{z}$:

$$\text{(2.A)} \qquad \nabla_A = 0, \quad \nabla_Z A = -2\lambda Z, \quad \nabla_V A = -\lambda V, \quad \nabla_Z Z' = 2\lambda\langle Z, Z'\rangle A,$$

$$\nabla_V V' = \lambda\langle V, V'\rangle A + \tfrac{1}{2}[V, V'], \quad \nabla_Z V = \nabla_V Z = -\tfrac{1}{2}j(Z)V.$$

Let σ be spanned by the orthonormal vectors $X = a\cdot A + V + Z$, $Y = V' + Z'$ with $a \geqq 0$. Since $\operatorname{ad}_Y^2 X = a\cdot \operatorname{ad}_{V'}^2 A = 0$, formula (1.A) reduces to

$$K(\sigma) = |\nabla_X Y|^2 - \langle \nabla_Y Y, \nabla_X X\rangle - |[X, Y]|^2.$$

As

$$\nabla_X X = \lambda(|V|^2 + 2|Z|^2)A - a\lambda(V + 2Z) - j(Z)V$$

$$\nabla_Y Y = \lambda(|V'|^2 + 2|Z'|)A - j(Z')V'$$

$$\nabla_X Y = \lambda(\langle V, V'\rangle + 2\langle Z, Z'\rangle)A - \tfrac{1}{2}(j(Z')V + j(Z)V') + \tfrac{1}{2}[V, V']$$

$$[X, Y] = a\lambda(V' + 2Z') + [V, V']$$

and $\langle V, V'\rangle = -\langle Z, Z'\rangle$, $\langle j(Z')V, j(Z)V'\rangle + \langle j(Z)V, j(Z')V'\rangle = 2\langle Z, Z'\rangle\langle V, V'\rangle$, we deduce

$$\begin{aligned}(2.B)\quad K(\sigma) = &-\tfrac{3}{4}|[V, V']|^2 - \lambda^2|V|^2|V'|^2 - (2\lambda^2 - \tfrac{1}{4})(|V|^2|Z'|^2 + |V'|^2|Z|^2)\\ &-4\lambda^2|Z|^2|Z'|^2 - (1-\lambda^2)\langle Z, Z'\rangle^2 - a^2\lambda^2|V'|^2\\ &-4a^2\lambda^2|Z'|^2 - 3a\lambda\langle Z', [V, V']\rangle - \tfrac{3}{2}\langle j(Z)V, j(Z')V'\rangle.\end{aligned}$$

2.3 Theorem (J. Boggino [7]). *For $\lambda = \frac{1}{2}$, $\mathfrak{s}$ has nonpositive sectional curvature and constant Ricci curvature $-\left(\frac{n}{4}+m\right)$.*

Proof. For $r = \frac{n}{2}$, the eigenvalues in (1.E) are $\lambda_i = \frac{1}{2}$, $\lambda_{n+k} = 1$. Insert $\lambda = \frac{1}{2}$ into (2.B):

$$K(\sigma) = -\tfrac{3}{4}|[V, V'] + aZ'|^2 - \tfrac{1}{4} - \tfrac{3}{4}|Z|^2|Z'|^2 - \tfrac{3}{4}\langle Z\, Z'\rangle^2 - \tfrac{3}{2}\langle j(Z)V, j(Z')V'\rangle,$$

$K(\sigma) \leqq 0$ since $-\frac{3}{4}|Z|^2|Z'|^2 + \frac{3}{2}|Z||Z'||V||V'| \leqq \frac{1}{4}$ [7, p. 539]. □

These examples include all rank one symmetric spaces with curvature between -1 and $-\frac{1}{4}$ [15]. Furthermore, Proposition 3 of that paper implies that $\mathcal{S}$ is symmetric iff

$$(2.C)\qquad j(Z_k)j(Z_l)V_i \in \{j(Z_1)V_i, \ldots, j(Z_m)V_i\} \quad \text{for all } Z_k \neq Z_l.$$

2.4 Corollary. *Under the assumptions of Theorem 2.3, the curvature operator $\hat{R}$ is nonpositive iff $\mathcal{S}$ is symmetric.*

Proof. It is well known, that the curvature operator of a symmetric space of noncompact type has eigenvalues $\leqq 0$ [21]. If $\mathcal{S}$ is not symmetric, there exist orthonormal vectors V, V', Z, Z' such that $j(Z)V = -j(Z')V'$ and $[V, V'] = 0$, see (2.C). One computes that $\langle \hat{R}(V \wedge V' + \frac{1}{2}\cdot Z \wedge Z',\ V \wedge V' + \frac{1}{2}\cdot Z \wedge Z'\rangle = 0$. If $\hat{R} \leqq 0$, $\hat{R}(V \wedge V' + \frac{1}{2}\cdot Z \wedge Z')$ had to be 0, but $\langle \hat{R}(V \wedge V' + \frac{1}{2} Z \wedge Z'),\ j(Z')V \wedge j(Z)V\rangle = \frac{3}{4}$. □

These groups can be used as building blocks for new examples:

2.5 Lemma. *Let $\mathfrak{s}_1 = \mathbb{R}A_1 + \mathfrak{n}_1$ and $\mathfrak{s}_2 = \mathbb{R}A_2 + \mathfrak{n}_2$ be two metric Lie algebras of Iwasawa type with nonpositive (negative) sectional curvature. Define the Lie algebra $\mathfrak{s} = \mathbb{R}A + \mathfrak{n}_1 + \mathfrak{n}_2$ by $[\mathfrak{n}_1, \mathfrak{n}_2] = 0$, $\mathrm{ad}_A|_{\mathfrak{n}_1} = \mathrm{ad}_{A_1}$, $\mathrm{ad}_A|_{\mathfrak{n}_2} = \mathrm{ad}_{A_2}$; then $\mathfrak{s}$ also has nonpositive (negative) curvature.*

Proof. Let ρ denote the quadratic form $\rho(X \wedge Y) = \langle R(X, Y)Y, X\rangle$. Choose orthogonal vectors $a \cdot A + X_1 + X_2$ and $Y_1 + Y_2$. By (1.B),

$$\begin{aligned}\rho((A + X_1 + X_2) \wedge (Y_1 + Y_2)) = &\rho_1((A + X_1) \wedge Y_1) + \rho_2((A + X_2) \wedge Y_2)\\ &+ \rho(X_1 \wedge Y_2 + X_2 \wedge Y_1).\end{aligned}$$

The positive definite operator ad_A defines a new scalar product $\langle X, Y\rangle_A := \langle \mathrm{ad}_A X, Y\rangle$ on $\mathfrak{n}$, then

$$\rho(X_1 \wedge Y_2 + X_2 \wedge Y_1) = -(|X_1|_A^1 |Y_2|_A^2 + |X_2|_A^2 |Y_1|_A^2 - 2\langle X_1, Y_1\rangle_A \langle X_2, Y_2\rangle_A)$$
$$= -|X_1 \wedge Y_2 + X_2 \wedge Y_1|_A^2.$$

If the curvature of $\mathfrak{s}_1, \mathfrak{s}_2$ is negative, there exists a $\tau > 0$ with

$$\rho(A + X_1 + X_2, Y_1 + Y_2) \leqq -\tau |(A + X_1 + X_2) \wedge (Y_1 + Y_2)|^2. \quad \square$$

If $\mathfrak{s}_1, \mathfrak{s}_2$ satisfy the assumptions of Theorem 2.3, we can likewise define a metric of constant Ricci curvature on $\mathfrak{s} = \mathbb{R}A + \mathfrak{n}_1 + \mathfrak{n}_2$, but in this case, the eigenvalues of ad_A will be smaller than the eigenvalues of ad_{A_1} and ad_{A_2} cf. 1.11(2). To apply 2.5 one has to determine on each factor, for which values of λ the curvature term (2.B) is still nonpositive. Since $K^{\mathfrak{n}}(V_i, Z_{n+k}) = \frac{1}{4}$, we must certainly have $\frac{1}{4} - \lambda_i \lambda_{n+k} \leqq 0$, hence $\lambda \geqq \frac{1}{\sqrt{8}}$.

First assume, that $\mathfrak{n}$ is not a nilpotent Iwasawa factor of a rank one symmetric space. Choose vectors V, V', Z, Z' as in the proof of 2.4 and let σ be spanned by $V + \frac{Z}{\sqrt{2}}$, $V' + \frac{Z'}{\sqrt{2}}$, then $K(\sigma) = -\frac{16}{9}\lambda^2 + \frac{4}{9}$. Thus $K \leqq 0$ iff $\lambda \geqq \frac{1}{2}$, for $\lambda = \frac{1}{2}$ there are planes of curvature 0. This shows that non-symmetric examples cannot be composed in this fashion.

Next we consider the Lie algebra of the ordinary Heisenberg group, i.e. $\dim \mathfrak{z} = 1$. We evaluate (2.B) for $\lambda = \frac{1}{\sqrt{8}}$ and $Z = b \cdot Z'$, $Z' \neq 0$:

$$K(\sigma) = -\frac{3}{4}|[V, V']|^2 - \frac{1}{8}|V|^2 |V'|^2 - \frac{1}{2}b^2 |Z'|^4 - \frac{7}{8}b^2 |Z'|^4$$
$$-\frac{a^2}{8}|V'|^2 - \frac{a^2}{2}|Z'|^2 - \frac{3a}{\sqrt{8}}\langle Z', [V, V']\rangle - \frac{3}{2}b\langle j(Z')V, j(Z')V'\rangle$$
$$\leqq -\frac{3}{4}\left|[V, V'] - \frac{a}{\sqrt{2}}Z'\right|^2 - \frac{a^2}{8}|Z'|^2 - \frac{1}{8}|V|^2 |V'|^2 + \frac{b^2}{8}|Z'|^4,$$

which is $\leqq 0$ since $|V||V'| \geqq |\langle V, V'\rangle| = |\langle Z, Z'\rangle| = b|Z'|^2$. If $Z' = 0$, likewise

$$K(\sigma) = -\tfrac{3}{4}|[V, V']|^2 - \tfrac{1}{8}|V|^2 |V'|^2 - \tfrac{1}{8}a^2 |V'|^2 \leqq 0.$$

As $K(\sigma)$ is strictly decreasing in λ, the sectional curvature will be negative for $\lambda > \frac{1}{\sqrt{8}}$.

2.6 Proposition. *Assume that* $\mathfrak{n}$ *is the Lie algebra of the Heisenberg group of dimension* $n+1$ *(*n *even),* $\mathfrak{a}_k$ *abelian of dimension* k.

(i) *Define* $\mathfrak{s}=\mathbb{R}A+(\mathfrak{n}+\mathfrak{n})$, *where* ad_A *has eigenvalues* $\frac{1}{\sqrt{8}}$ *and* $\frac{1}{\sqrt{2}}$. *For a suitable metric,* $\mathfrak{s}$ *is Einstein with nonpositive sectional curvature* $\left(\text{Einstein constant } -\frac{n}{4}-1\right)$.

(ii) (E. Deloff [11, p. 95]) *Let* $k\leqq n$, *then* $\mathfrak{s}=\mathbb{R}A+(\mathfrak{n}+\mathfrak{a}_k)$ *where* ad_A *has eigenvalues* $\frac{1}{2d}\left(\frac{n}{2}+1\right)$, $\frac{1}{d}\left(\frac{n}{2}+1\right)$, $\frac{1}{d}\left(\frac{n}{4}+1\right)$, $d^2=\left(\frac{n}{2}+1\right)^2+k\left(\frac{n}{4}+1\right)$ *carries an Einstein metric of negative sectional curvature* $\left(c=\frac{n}{4}+1\right)$.

Proof. (i) As $\mathfrak{n}+\mathfrak{n}$ is uniform of type $\left(2, 2n, \frac{n}{2}\right)$, (1.E) implies $\lambda_i=\frac{1}{\sqrt{8}}$, $\lambda_k=\frac{2}{\sqrt{8}}$. For $\lambda=\frac{1}{\sqrt{8}}$, $\mathfrak{s}=\mathbb{R}A+\mathfrak{n}$ has nonpositive curvature, so we can apply 2.5.

(ii) The eigenvalues of ad_A can be computed using 1.11(1) and 1.5. For $\lambda^2=\frac{1}{4d^2}\left(\frac{n}{2}+1\right)^2>\frac{1}{8}$, $\mathbb{R}A+\mathfrak{n}$ has negative curvature, which implies $k<n+\frac{4}{n+4}$. The Lie algebra $\mathbb{R}A+\mathfrak{a}_k$ has constant negative curvature, now use 2.5. □

One may ask, if one can carry over these results to the Lie algebra $\mathfrak{n}$ of the quaternionic Heisenberg group, where $\mathfrak{z}=\mathrm{span}\{Z_1, Z_2, Z_3\}$ and $j(Z_1)j(Z_2)=j(Z_3)$. Unfortunately, $\mathbb{R}A+\mathfrak{n}$ has some positive sectional curvatures for $\lambda_i=\frac{1}{\sqrt{8}}$: Let σ be spanned by $V+Z$, $V'+Z'$, where $Z=\frac{1}{\sqrt{2}}Z_1$, $Z'=\frac{1}{\sqrt{2}}Z_2$, $V'=-j(Z_3)V$; (2.B) implies

$$K(\sigma)=-\tfrac{3}{4}|[V, V']|^2-\tfrac{1}{8}|V|^2|V'|^2-\tfrac{1}{2}|Z|^2|Z'|^2-\tfrac{3}{2}\langle j(Z)V, j(Z')V'\rangle.$$

As $|[V, V']|=|\langle[V, V'], Z_3\rangle|=|j(Z_3)V|^2=\frac{1}{2}$ and $j(Z')j(Z)V=-\frac{1}{2}j(Z_3)V$, $K(\sigma)=-\frac{3}{16}-\frac{1}{32}-\frac{1}{8}+\frac{3}{8}=\frac{1}{32}>0$.

We do not know the minimal value of λ for the Lie algebra $\mathfrak{n}$ of the quaternionic Heisenberg group such that $\mathbb{R}A+\mathfrak{n}$ still has nonpositive curvature, but we can specify some lower bound:

2.7 Lemma. *Let* $\mathfrak{n}$ *be the nilpotent part of a rank one symmetric space with curvature in* $[-1; -\frac{1}{4}]$. $\mathfrak{s}=\mathbb{R}A+\mathfrak{n}$ *has negative curvature for all* $\lambda\geqq\frac{19}{40}$.

Proof. Let K^λ be the curvature of $\mathfrak{s}$ when ad_A has eigenvalues λ, 2λ. If $|K^\lambda-K^{\frac{1}{2}}|\leqq\frac{1}{4}$, K^λ is negative. We will show below, that $\left|\frac{\partial}{\partial\lambda}K^\lambda\right|_{\frac{1}{2}}\Big|<10$, by the mean value theorem, $K^\lambda<0$ for $10\cdot(\frac{1}{2}-\lambda)\leqq\frac{1}{4}$, i.e. $\lambda\geqq\frac{19}{40}$. From (2.B) we calculate

$$\frac{\partial}{\partial\lambda}K^\lambda(\sigma)\Big|_{\frac{1}{2}}=-(|V|^2|V'|^2+2|V|^2|Z'|^2+2|V'|^2|Z|^2+4|Z|^2|Z'|^2-\langle Z, Z'\rangle^2)$$
$$-3a\langle Z', [V, V']\rangle-a^2|V'|^2-4a^2|Z'|^2.$$

As $|V|^2=1-a^2-|Z|^2$ and $|V'|^2=1-|Z'|^2$,

$$\left|\frac{\partial}{\partial\lambda}K^{\lambda}\Big|_{\frac{1}{2}}\right|\leqq(|V||V'|+\tfrac{3}{2}a|Z'|)^2+(2-\tfrac{5}{4}a^2)|Z'|^2+a^2+2|Z|^2,$$

which is <10, because $(|V'|+\frac{3}{2}|Z'|)^2\leqq 2(|V'|^2+\frac{9}{4}|Z'|^2)<5$. □

2.8 Proposition. *Assume that* $\mathfrak{n}$ *is the Lie algebra of the quaternionic Heisenberg group of dimension* $n+3$, $n\equiv 0(\mathrm{mod}\ 4)$. *For* $k\leqq\frac{n}{9}$, $\mathfrak{s}=\mathbb{R}A+(\mathfrak{n}+\mathfrak{a}_k)$ *carries an Einstein metric of negative sectional curvature* $\left(\text{Einstein constant } -\left(\frac{n}{2}+3\right)\right)$.

Proof. The eigenvalues of ad_A can be read off from 1.11(1) and 1.5:

$$\mathrm{ad}_A V=\frac{1}{2d}\left(\frac{n}{2}+3\right)V,\quad \mathrm{ad}_A Z=\frac{1}{d}\left(\frac{n}{2}+3\right)Z,\quad \mathrm{ad}_A|_{\mathfrak{a}_k}=\frac{1}{d}\left(\frac{n}{4}+3\right)\cdot Id$$

with $d^2=\left(\frac{n}{2}+3\right)^2+k\left(\frac{n}{4}+3\right)$. To apply 2.5, we must have $\frac{1}{4d^2}\left(\frac{n}{2}+3\right)^2>\frac{19^2}{40^2}$, i.e. $k<\left(\frac{400}{361}-1\right)\left(n-\frac{12}{n+12}\right)$, whence $k\leqq\frac{n}{9}$. □

Remarks. (1) It is easy to see that the curvature operators of all the examples in Proposition 2.6 and 2.8 have some positive eigenvalues. Along with Corollary 2.4 we are lead to conjecture that homogeneous Einstein manifolds with nonpositive curvature operators are symmetric.

(2) If one takes Heisenberg groups of different dimensions in 2.6(i), one can still define an Einstein metric on $\mathfrak{s}=\mathbb{R}A+\mathfrak{n}_1+\mathfrak{n}_2$ (cf. 1.11(2)) but the factor of lower dimension will have planes of positive sectional curvature.

(3) So far we have not found any non-symmetric example with $\dim\mathfrak{a}>1$. Following algebras seemed to be promising candidates at first sight: Suppose that $\mathfrak{n}$ is the Lie algebra of the Heisenberg group of dimension $2l+1$; set $\mathfrak{a}=\mathrm{span}\{A_1, A_2\}$, where ad_{A_1} and ad_{A_2} have eigenvalues $\frac{1}{2},\ldots,\frac{1}{2},1$ and $\frac{\sqrt{3}}{2},\ldots,\frac{\sqrt{3}}{2},-\frac{\sqrt{3}}{2},\ldots,-\frac{\sqrt{3}}{2},0$ on $\mathfrak{n}$. By 1.5, the canonical scalar product on $\mathfrak{s}=\mathfrak{a}+\mathfrak{n}$ is Einstein. For $l=1$, $\mathscr{S}$ is isometric to the symmetric space $SL(3,\mathbb{R})/SO(3,\mathbb{R})$ (which has of course nonpositive sectional curvature). However, if $l>1$, there are vectors $V_1, V_2\in\mathfrak{n}$ such that $[V_1, V_2]=0$ and $\mathrm{ad}_{A_2}V_1=\frac{\sqrt{3}}{2}V_1$, $\mathrm{ad}_{A_2}V_2=-\frac{\sqrt{3}}{2}V_2$. Set $X=V_1+V_2$, $Y=V_1-V_2$, using (1.C) one computes $K(X\wedge Y)=2|V_1\wedge V_2|^2>0$.

Perhaps modifications of symmetric spaces of higher rank as described in 1.11(3) may produce such examples. Observe that Lemma 2.5 does not hold for groups of Iwasawa type with $\dim\mathfrak{a}>1$, cf. following counterexample: Assume that $\mathfrak{s}$ is the Iwasawa factor of $M=SL(3,\mathbb{R})/SO(3,\mathbb{R})$, see above. Set $\mathfrak{s}=\mathfrak{a}+\mathfrak{n}'$ with $\mathfrak{n}'=\mathfrak{n}+\mathfrak{n}$, where the operation of $\mathfrak{a}$ is defined as in 2.5. Take X in the

first nilpotent factor with $\mathrm{ad}_{A_1} X = \frac{1}{2} X$, $\mathrm{ad}_{A_2} X = \frac{\sqrt{3}}{2} X$, Y in the second with $\mathrm{ad}_{A_1} Y = \frac{1}{2} Y$, $\mathrm{ad}_{A_2} Y = -\frac{\sqrt{3}}{2} Y$. For the second fundamental form of $\mathfrak{n}'$ we obtain

$$II_{n'}(X, X) = \frac{1}{2} A_1 + \frac{\sqrt{3}}{2} A_2, \quad II_{n'}(Y, Y) = \frac{1}{2} A_1 - \frac{\sqrt{3}}{2} A_2, \quad II_{n'}(X, Y) = 0,$$

whence $K(X, Y) = 0 - \langle II_{n'}(X, X), II_n, \langle Y, Y \rangle \rangle = +\frac{1}{2}$.
(4) For Kähler-Einstein metrics we refer to following result [10]: A bounded homogeneous domain has nonpositive sectional curvature relative to the Bergmann metric iff it is symmetric.

3 Weakening the Einstein condition

A simply connected Riemannian manifold M with parallel Ricci tensor has constant Ricci curvature on every de Rham factor [6, 1.100] Consider the less restrictive condition that

$$0 = (\nabla_Z \mathrm{ric})(Z, Z) = Z.\mathrm{ric}(Z, Z) + 2\,\mathrm{ric}(\nabla_Z Z, Z) \tag{3.A}$$

for all vector fields Z on M. This means that the Ricci curvature is invariant under the geodesic flow. If M is a compact manifold of nonpositive sectional curvature, (3.A) implies that the Ricci tensor is parallel [6, 16.54]. We prove the same thing for groups of Iwasawa type:

3.1 Theorem. *Suppose that $\mathfrak{s}$ is a metric Lie algebra of Iwasawa type. Condition* (3.A) *implies that $\mathscr{S}$ is Einstein on every de Rham factor.*

Proof. For left invariant vector fields $Z \in \mathfrak{s}$, (3.A) gives $\mathrm{ric}(\nabla_Z Z, Z) = 0$. Let $Z = A + X$, then by (1.B) $\nabla_Z Z = -\mathrm{ad}_A X + \sum |X_\alpha|^2 A_\alpha + \nabla^{\mathfrak{n}}_X X$, so

$$\mathrm{ric}(\mathrm{ad}_A X, X) = \sum |X_\alpha|^2 \,\mathrm{ric}(A_\alpha, A)$$

since $\mathrm{ric}(\nabla^{\mathfrak{n}}_X X, X) = \mathrm{ric}(\nabla_X X, X) = 0$. Both sides of this equation define for fixed A a symmetric form on $\mathfrak{n}$, thus we have

$$\tfrac{1}{2}(\mathrm{ric}(\mathrm{ad}_A X, Y) + \mathrm{ric}(X, \mathrm{ad}_A Y)) = \sum \mathrm{ric}(A_\alpha, A)/\langle X_\alpha, Y_\alpha \rangle. \tag{3.B}$$

Choose $X = X_\alpha$, $Y = Y_\beta$ with $\alpha \neq \beta$ and $A \in \mathfrak{a}$ with $(\alpha + \beta)(A) \neq 0$, then $\mathrm{ric}(X, Y) = 0$ by (3.B), hence different root spaces $\mathfrak{n}_\alpha$, $\mathfrak{n}_\beta$ are orthogonal with respect to the Ricci form. For $X, Y \in \mathfrak{n}_\alpha$ we have

$$\alpha(A)\,\mathrm{ric}(X, Y) = \mathrm{ric}(A, A_\alpha)\langle X, Y \rangle. \tag{3.C}$$

Define

$$c_\alpha = \frac{\mathrm{ric}(A_\alpha)}{\alpha(A_\alpha)} \quad \text{for all roots } \alpha,$$

then $\mathrm{ric}(X, Y)=c_\alpha\cdot\langle X, Y\rangle$ on $\mathfrak{n}_\alpha$ and $\mathrm{ric}(A, A_\alpha)=c_\alpha\cdot\langle A, A_\alpha\rangle$ for all $A\in\mathfrak{a}$ by (3.C). From the last equation we conclude, that root vectors A_α, A_β with $c_\alpha\neq c_\beta$ are orthogonal:

$$c_\alpha\langle A_\beta, A_\alpha\rangle=\mathrm{ric}(A_\alpha, A_\beta)=c_\beta\langle A_\alpha, A_\beta\rangle.$$

This again implies, that the root spaces $\mathfrak{n}_\alpha$, $\mathfrak{n}_\beta$ commute. Otherwise, $\alpha+\beta$ would be a root of $\mathfrak{a}$, since

$$c_{\alpha+\beta}=\frac{\mathrm{ric}(A_{\alpha+\beta})}{|A_{\alpha+\beta}|^2}=\frac{\mathrm{ric}(A_\alpha)+\mathrm{ric}(A_\beta)}{|A_\alpha|^2+|A_\beta|^2}\neq c_\alpha, c_\beta,$$

$A_{\alpha+\beta}=A_\alpha+A_\beta$ would have to be perpendicular to A_α and A_β.

Now sum up all subspaces $\mathfrak{n}_\alpha$ with $c_\alpha=c$ to some space $\mathfrak{n}_c$ and define $\mathfrak{a}_c$ to be the span of the corresponding root vectors. Then $\mathfrak{s}$ decomposes into the orthogonal sum of ideals $\mathfrak{s}_c=\mathfrak{a}_c+\mathfrak{n}_c$; the restriction of the metric to $\mathfrak{s}_c$ has constant Ricci curvature c. □

3.2. Corollary. *Let $\mathfrak{s}$ be an irreducible metric Lie algebra of Iwasawa type where the geodesic symmetries on $\mathcal{S}$ preserve the Riemannian volume, then $\mathcal{S}$ is Einstein.*

Proof. The invariance of the volume element in normal coordinates under reflections implies (3.A) [6, 16.55]. □

Remarks. For Iwasawa factors of symmetric spaces the reflections even preserve the metric. One may ask, if there are non-symmetric examples at all satisfying the assumptions of Corollary 3.2. For several reasons, the groups of Boggino (Theor. 2.3) seem to be good candidates. A positive answer to this question would be of interest in the following context: In [17] it is shown, that homogeneous commutative spaces (i.e. manifolds, where the algebra of isometry invariant differential operators is commutative) have volume preserving geodesic symmetries. Now a group of Iwasawa type with left invariant metric is commutative iff it is symmetric [22, Theor. 4.20]. So a non-symmetric example would show, that the class of spaces with volume preserving geodesic symmetries is strictly larger than the class of commutative spaces.

Groups of Iwasawa type have also been investigated in harmonic analysis, cf. [9]. It may be interesting to relate these results to their geometry.

References

1. Alekseevskiĭ, D.V.: Conjugacy of polar factorizations of Lie groups. Mat. Sb. **13**, 12–24 (1971); English translation: Math. USSR-Sb. **84**, 14–26 (1982)
2. Alekseevskii, D.V.: Homogeneous Riemannian spaces of negative curvature. Mat. Sb. **25**, 87–109 (1975); English translation: Math. USSR-Sb **96**, 93–117 (1975)
3. Ancochea Bermudez, J.M., Goze, M.: Classification des algèbres de Lie nilpotents complexes de dimension 7. Arch. Math. **52**, 175–185 (1989)
4. Azencott, H., Wilson, E.: Homogeneous manifolds with negative curvature I. Trans. Am. Math. Soc. **215**, 323–362 (1976)
5. Azencott, R., Wilson, E.: Homogeneous manifolds with negative curvature II. Mem. Am. Math. Soc. **8**, (1976)
6. Besse, A.: Einstein Manifolds. Berlin Heidelberg New York: Springer 1987
7. Boggino, J.: Generalized Heisenberg groups and solvmanifolds naturally associated. Rend. Semin. Mat., Torino **43**, 529–547 (1985)

8. Cartan, E.: Sur les domaines bornés homogènes de l'éspace de n variables complexes. Abh. Math. Semin. Univ. Hamburg **11**, 116–162 (1936)
9. Damek, E.: Left invariant degenerate elliptic operators on semidirect extensions of homogeneous groups. Stud. Math. **89**, 169–196 (1987)
10. D'Atri, J., Dotti Miatello, I.: A characterization of bounded symmetric domains by curvature, Trans. Am. Math. Soc. **276**, 531–540 (1983)
11. Deloff, E.: Naturally reductive metrics and metrics with volume preserving geodesic symmetries on NC-Algebras. New Brunswick: Thesis, Rutgers 1979
12. Dixmier, J., Lister, W.G.: Derivations of nilpotent Lie algebras. Proc. Am. Math. Soc. **8**, 155–158 (1957)
13. Dotti Miatello, I.: Ricci curvature of left invariant metrics on solvable unimodular Lie groups. Math. Z. **180**, 257–263 (1982)
14. Eberlein, P.: Structure of manifolds of nonpositive curvature. In: Proceedings of a Conf. on Diff. Geom. and Global Analysis. Berlin Heidelberg New York: Springer 1985
15. Heintze, E.: On homogeneous manifolds of negative curvature. Math. Ann. **211**, 23–34 (1974)
16. Kaplan, A.: Riemannian nilmanifolds attached to Clifford modules. Geom. Dedicata **11**, 127–136 (1981)
17. Kowalski, O., Vanhecke, L.: Opérateurs différentiels invariants et symétries géodesiques préservant le volume. C.R. Acad. Sci. Paris, Sér. I Math. **296**, 1001–1003 (1983)
18. Magnin, L.: Sur les algèbres de Lie nilpotentes de dimension $\leqq 7$. J. Geom. Phys. **3**, 119–144 (1986)
19. Nakajima, K.: On j-algebras and homogeneous Kähler manifolds. Hokkaido Math. J. **15**, 1–20 (1986)
20. Pyatetskii-Shapiro, I.: The geometry and classification of bounded homogeneous domains. Russ. Math. Surv. **20**, 1–48 (1965)
21. Simons, J.: On transitivity of holonomy systems. Ann. Math., II. Ser. **76**, 213–234 (1962)
22. Wolter, T.: Homogene Mannigfaltigkeiten nichtpositiver Krümmung. Zürich: Thesis 1989

Math. Z. 206, 473–499 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Some results on connecting orbits for a class of Hamiltonian systems

Paul H. Rabinowitz [1,*] **and Kazunaga Tanaka** [2]

[1] Department of Mathematics and Center for the Mathematical Sciences, University of Wisconsin, Madison, WI 53706, USA

[2] Department of Mathematics, Nagoya University, Nagoya 464, Japan

Received October 25, 1989; in final form April 20, 1990

§ 1 Introduction

This paper concerns the existence of various kinds of connecting orbits for second order Hamiltonian systems of the form

$$\ddot{q}+V'(q)=0. \tag{HS}$$

It will always be assumed that V has a global maximum, e.g. at $x=0$. Therefore $q\equiv 0$ is a solution of (HS). We are interested in nontrivial solutions of (HS) that terminate at $x=0$, i.e.

$$\lim_{t\to\infty} q(t)\equiv q(\infty)=0=\dot{q}(\infty). \tag{1.1}$$

If e.g. $\Omega\subset\mathbf{R}^n$ is bounded, $V\in C^1(\bar{\Omega},\mathbf{R})$ and

$$V(x)<V(0) \quad \text{for all } x\in\bar{\Omega}\setminus\{0\}, \tag{1.2}$$

Theorem (2.1) shows there is a solution of (HS) such that $q(0)\in\partial\Omega$, $q(t)\in\Omega$ for all $t>0$. and (1.1) holds. If (1.2) is replaced by

$$\begin{aligned} &V(x)<V(0) && \text{for all } x\in\Omega\setminus\{0\}\\ &V(x)=V(0) && \text{for } x\in\partial\Omega\\ &V'(x)\neq 0 && \text{for } x\in\partial\Omega, \end{aligned} \tag{1.3}$$

then Theorem (2.22) shows there exists a solution of (HS) as in Theorem (2.1) further satisfying $\dot{q}(0)=0$. Since (HS) is time reversible, extending this solution to $\mathbf{R}$ via $q(-t)=q(t)$ yields a homoclinic solution of (HS) emanating from 0.

* Sponsored in part by the National Science Foundation under grant #MCS-8110556, the US Army Research Office under contract #DAA L03-87-K-0043, and the Office of Naval Research under Grant No. N00014-88-K-0134. Any reproduction for the purposes of the United States Government is permitted.

These results are proved in § 2 by studying the functional

$$I(q)=\int_0^\infty [\tfrac{1}{2}|\dot{q}(t)|^2 - V(q)]\,dt \tag{1.4}$$

corresponding to (HS) and using elementary minimization arguments. Our arguments were motivated in part by [1] where related reasoning was employed to prove the existence of heteroclinic solutions of (HS) for functions $V(x)$ which are periodic in x.

In § 3, the setting of Theorem (2.22) is studied in more detail. If $\partial\Omega$ contains l components, we show in Theorem (3.1) that (HS) possesses at least l_h homoclinic orbits joining 0 to $\partial\Omega$ and at least l_p periodic solutions joining the components, where $l_h + l_p \geqq l$.

In § 4, we assume $\Omega = \mathbf{R}^n$. Then the methods of § 2 show that for each $\xi \in \mathbf{R}^n \setminus \{0\}$, there is a solution of (HS) satisfying $q(0)=\xi$ and (1.1) (Theor. (4.1)). Moreover the critical value of I in (1.4) has a variational characterization. The existence assertion of Theorem (4.1) without this variational characterization was already proved by Bolotin and Kozlov [2] in a more general setting by a less direct argument. It is also proved in § 4 that under the hypotheses of Theorem (4.1), there is a solution of (HS) emanating from infinity and terminating at 0.

Lastly in § 5, the existence of homoclinic orbits is studied for

$$\ddot{q} + V_q(t, q) = 0 \tag{1.5}$$

assuming that the potential energy V grows at a superquadratic rate as $|x| \to \infty$, i.e.

$$V(t, x)|x|^{-2} \to \infty \quad \text{as } |x| \to \infty. \tag{1.6}$$

Such questions have been studied recently by Coti-Zelati, Ekeland, and Sere [3], and Hofer and Wysocki [4] for general Hamiltonian systems:

$$\dot{z} = \mathcal{J} H_z(t, z) \tag{1.7}$$

and for (1.5) in [5]. It is assumed in [3–5] that H (or V) is T-periodic in t. Here (1.5) is treated without such an assumption for

$$V(t, x) = -\tfrac{1}{2} L(t)x \cdot x + W(t, x) \tag{1.8}$$

where $L(\cdot)$ is positive definite and W satisfies (1.6). Two different settings are studied. For the first, it is assumed that the smallest eigenvalue of $L(t)$ approaches ∞ as $|t| \to \infty$. For the second, we assume

$$L(t) \to L_\infty(t),$$

a T-periodic function, in an appropriate sense as $|t| \to \infty$. For both cases minimax arguments are employed to obtain the existence of the homoclinic solution.

This work was done while the second author was visiting the Center for the Mathematical Sciences, University of Wisconsin-Madison. He would like to thank the Center for its kind hospitality.

§ 2 Autonomous Hamiltonian systems

This section contains Theorems (2.1) and (2.22) as stated in the Introduction. In this and the sections that follow, in the proofs of the results, it will be assumed that $n \geqq 2$. The proofs for $n=1$ are much simpler.

(2.1) **Theorem.** *Let Ω be a bounded neighborhood of 0 in $\mathbf{R}^n$ and $V \in C^1(\bar{\Omega}, \mathbf{R})$ with $V(x) < V(0)$ for all $x \in \bar{\Omega} \setminus \{0\}$. Then there exists a solution, q, of* (HS) *such that $q(0) \in \partial\Omega$, $q(\infty) = 0 = \dot{q}(\infty)$, and $q(t) \in \Omega$ for all $t \in (0, \infty)$.*

Proof. Without loss of generality, it can be assumed that $V(0) = 0$. The proof consists of several steps and follows the lines of a related situation in [1] involving the existence of heteroclinic orbits of (HS).

Let $\mathbf{R}^+ = [0, \infty)$ and

$$E = \left\{ q \in W^{1,2}_{\mathrm{loc}}(\mathbf{R}^+, \mathbf{R}^n) \,\middle|\, \int_0^\infty |\dot{q}|^2 \, dt < \infty \right\}. \tag{2.2}$$

E is a Hilbert space under the norm

$$\|q\|^2 = \int_0^\infty |\dot{q}|^2 \, dt + |q(0)|^2$$

and $E \subset C(\mathbf{R}^+, \mathbf{R}^n)$. Let Γ be the subset of E defined by

$$\Gamma = \{ q \in E \mid q(0) \in \partial\Omega,\ q(\infty) = 0, \text{ and } q(t) \in \bar{\Omega} \text{ for all } t \in \mathbf{R}^+ \}. \tag{2.3}$$

For $q \in \Gamma$, consider the functional

$$I(q) = \int_0^\infty [\tfrac{1}{2} |\dot{q}|^2 - V(q)] \, dt. \tag{2.4}$$

Set

$$c = \inf_{q \in \Gamma} I(q). \tag{2.5}$$

Then Theorem (2.1) follows on establishing:

(2.6) **Proposition.** *c is a critical value of I with a corresponding critical point $q \in \hat{\Gamma}$ where*

$$\hat{\Gamma} = \{ w \in \Gamma \mid w(t) \in \Omega \text{ for all } t \in (0, \infty) \}. \tag{2.7}$$

Moreover

$$c = \inf_{w \in \hat{\Gamma}} I(w), \tag{2.8}$$

q satisfies (HS), *and $\dot{q}(\infty) = 0$.*

The first step in proving Proposition (2.6) is a variant of Lemma (3.6) of [1]. Let $B_\rho(\xi)$ denote the open ball of radius ρ about $\xi \in \mathbf{R}^n$. If $\xi = 0$, we simply write B_ρ.

(2.9) **Lemma.** *Let $\rho > 0$ such that $B_\rho \subset \bar{\Omega}$. Set*

$$\beta(\rho) = \min_{x \in \bar{\Omega} \setminus B_\rho} -V(x).$$

Suppose $w \in E$ and $w(t) \in \bar{\Omega} \setminus B_\rho$ for $t \in \bigcup_{j=1}^{k} [r_j, s_j]$. Then

$$I(w) \geqq \sqrt{2\beta(\rho)} \sum_{j=1}^{k} |w(r_j) - w(s_j)|. \tag{2.10}$$

Proof. The proof of Lemma (2.9) is same as its analogue in [1] so we refer to [1] for details.

An immediate consequence of Lemma (2.9) is

(2.11) **Lemma.** *If $w \in E$, $w(t) \in \bar{\Omega}$ for all $t \in \mathbf{R}^+$, and $I(w) < \infty$, then $w(\infty) = 0$. If in addition, $w(0) \in \partial\Omega$, then $w \in \Gamma$.*

Proof. Let $\omega(w)$ denote the set of limit points of the orbit $w(t)$ as $t \to \infty$. Since $\bar{\Omega}$ is compact, $\omega(w) \neq \phi$. Let $\xi \in \omega(w)$. If $\xi \neq \omega(\infty)$, there is a $\delta > 0$ and sequences $(t_m), (\tau_m) \subset \mathbf{R}^+$ such that $t_m \to \infty$, $w(t_m) \to \xi$, $\tau_m \to \infty$, and $w(\tau_m) \notin B_\delta(\xi)$ as $m \to \infty$. Applying Lemma (2.9) with $\rho = \delta/4$ shows $I(w) \geqq \frac{3k\delta}{4}\sqrt{2\beta(\delta/4)}$ for any $k \in \mathbf{N}$, i.e. $I(w) = \infty$, a contradiction. Therefore $\xi = w(\infty)$. Lastly observe that since $V(0) > V(x)$ for all $x \in \bar{\Omega} \setminus \{0\}$, the only possible value of ξ for which $I(w) < \infty$ is $\xi = 0$.

Now we can prove:

(2.12) **Proposition.** *There exists $q \in \Gamma$ such that $I(q) = c$.*

Proof. Let (q_m) be a minimizing sequence for (2.5). Since $V \leqq 0$ and $\bar{\Omega}$ is compact, the form of I shows (q_m) is bounded in E. Hence a subsequence of (q_m) converges weakly in E and strongly in $L^\infty_{\text{loc}}(\mathbf{R}^+, \mathbf{R}^n)$ to $q \in E$ satisfying $q(0) \in \partial\Omega$ and $q(t) \in \bar{\Omega}$ for all $t \in \mathbf{R}^+$. A simple lower semicontinuity argument – see Proposition (3.12) of [1] – shows $I(q) < \infty$ and

$$I(q) \leqq \inf_{w \in \Gamma} I(w). \tag{2.13}$$

Consequently $q \in \Gamma$ via Lemma (2.11) and equality holds in (2.13).

To complete the proof of Proposition (2.6) and Theorem (2.1), it remains to show that $q \in \hat{\Gamma}$, i.e. $q(t) \in \Omega$ for all $t \in (0, \infty)$, q satisfies (HS), and $\dot{q}(\infty) = 0$. Set

$$\mathcal{T} = \{t \in (0, \infty) \mid q(t) \in \partial\Omega\}.$$

Since $q(\infty) = 0$ and q is continuous, $\mathcal{T}$ is compact.

(2.14) **Proposition.** *$\mathcal{T} = \phi$ and $q \in \hat{\Gamma}$.*

Proof. If $\mathscr{T} \neq \phi$, define τ by

$$\tau \equiv \max\{t \in (0, \infty) \mid t \in \mathscr{T}\}. \tag{2.15}$$

Hence $\tau > 0$ and since $q(\infty) = 0$, $\tau < \infty$. Set $Q(t) = q(t+\tau)$. Then $Q \in \hat{\Gamma} \subset \Gamma$ and

$$I(Q) = \int_{\tau}^{\infty} [\tfrac{1}{2}|\dot{q}|^2 - V(q(t))]\, dt < I(q) \tag{2.16}$$

since $V(q(t)) \leqq 0$ and $\not\equiv 0$ for $t \in (0, \tau)$. But then (2.13) is violated. Hence $\mathscr{T} = \phi$ and $q \in \hat{\Gamma}$.

(2.17) **Proposition.** *q satisfies* (HS) *on* $(0, \infty)$.

Proof. The proof of this statement is identical to that of Proposition (3.18) of [1] and will be omitted.

The Proposition implies

(2.18) **Corollary.** $\dot{q}(\infty) = 0$.

Proof. Since q satisfies (HS), there is a constant A such that

$$\tfrac{1}{2}|\dot{q}(t)|^2 + V(q(t)) \equiv A \tag{2.19}$$

for all $t \in \mathbf{R}^+$. Consequently

$$I(q) = \int_{0}^{\infty} (|\dot{q}|^2 - A)\, dt. \tag{2.20}$$

Since $q \in E$ and $I(q) < \infty$, (2.20) shows $A = 0$. Finally $q(\infty) = 0$, $V(0) = 0$, and (2.19) imply that $\dot{q}(\infty) = 0$.

The proof of Theorem (2.1) is complete.

(2.21) *Remark.* If V merely belongs to $C^1(\bar{\Omega}, \mathbf{R})$, the initial value problem for (HS) need not have a unique solution. Therefore it is possible that the solution $q(t)$ just constructed satisfies $q(t) \equiv 0$ for large t. Of course this cannot happen if $V \in C^2(\bar{\Omega}, \mathbf{R})$.

As a consequence of Theorem (2.1), we have

(2.22) **Theorem.** *Let $V \in C^1(\mathbf{R}^n, \mathbf{R})$ with $V(0) = 0$ being a strict local maximum of V. Let*

$$V^0 = \{x \in \mathbf{R}^n \mid V(x) < 0\} \cup \{0\}$$

and let Ω be the component of V^0 containing 0. If Ω is bounded and $V'(x) \neq 0$ for all $x \in \partial\Omega$, then there is a solution, q, of (HS) *in $\hat{\Gamma}$ with $\dot{q}(0) = 0 = \dot{q}(\infty)$. Moreover $I(q) = c$ where c is given by* (2.5) *and* (2.8).

Proof. An approximation argument based on Theorem (2.1) will be employed. For $\varepsilon > 0$, let

$$\Omega_\varepsilon = \{x \in \Omega \mid \operatorname{dist}(x, \partial\Omega) > \varepsilon\}.$$

For small ε, Ω_ε is a neighborhood of 0 for which Theorem (2.1) is valid. Hence there exists a solution q_ε of (HS) such that if

$$\Gamma_\varepsilon = \{w \in E \mid w(0) \in \partial\Omega_\varepsilon, w(\infty) = 0, \text{ and } w(t) \in \bar{\Omega}_\varepsilon \text{ for all } t \in (0, \infty)\},$$

and

$$\hat{\Gamma}_\varepsilon = \{w \in E \mid w(0) \in \partial\Omega_\varepsilon, w(\infty) = 0, \text{ and } w(t) \in \Omega_\varepsilon \text{ for all } t \in (0, \infty)\},$$

then

(2.23) (i) $q_\varepsilon \in \hat{\Gamma}_\varepsilon$ and $\dot{q}_\varepsilon(\infty) = 0$

(ii) $I(q_\varepsilon) = \inf_{\hat{\Gamma}_\varepsilon} I = \inf_{\Gamma_\varepsilon} I \equiv c_\varepsilon$.

Since $\Omega_\varepsilon \subset \Omega$ which is bounded and q_ε is a solution of (HS), $\{q_\varepsilon\}$ is bounded in $C^2(\mathbf{R}^+, \mathbf{R}^n)$. Hence there is a subsequence $\varepsilon_k \to 0$ and $q \in C^2(\mathbf{R}^+, \mathbf{R}^n)$ such that $q_{\varepsilon_k} \to q$ in $C^2_{\text{loc}}(\mathbf{R}^+, \mathbf{R}^n)$. By (HS) and (2.23), q is a solution of (HS), $q(0) \in \partial\Omega$, $q(t) \in \bar{\Omega}$ for all $t \in (0, \infty)$, and

$$\tfrac{1}{2}|\dot{q}(t)|^2 + V(q(t)) = 0 \quad \text{for } t \in \mathbf{R}^+. \tag{2.24}$$

Since $q(0) \in \partial V^0$, (2.24) shows $\dot{q}(0) = 0$. It remains to show that $q \in \hat{\Gamma}$, $\dot{q}(\infty) = 0$, and (2.5) and (2.8) hold. These facts are consequences of the next 3 lemmas. For now let c be as given by (2.5) and $\hat{c}$ by (2.8) so

$$c \leqq \hat{c}. \tag{2.25}$$

(2.26) **Lemma.** *$q \in E$ and $I(q) \leqq c$.*

Proof. Note that for $0 \leqq \varepsilon < \delta$, if $p \in \Gamma_\varepsilon$, since $p(\infty) = 0$, there exists a $\tau > 0$ such that $p(\tau) \in \partial\Omega_\delta$ and $p(t) \in \Omega_\delta$ for all $t > \tau$. Hence $w(t) = p(t+\tau) \in \Gamma_\delta$ and $I(w) \leqq I(p)$. This implies that for such pairs p and w:

$$\inf_{p \in \Gamma_\varepsilon} I(p) \geqq \inf_{w \in \Gamma_\delta} I(w) \geqq \inf_{u \in \Gamma_\delta} I(u) \tag{2.27}$$

for $0 \leqq \varepsilon < \delta$ and therefore

$$\hat{c} \geqq c \geqq c_\varepsilon \geqq c_\delta. \tag{2.28}$$

Since $q_{\varepsilon_k} \to q$ in C^2_{loc}, for all $T > 0$,

$$\begin{aligned} \infty > \hat{c} \geqq c &\geqq \overline{\lim_{k \to \infty}} I(q_{\varepsilon_k}) \\ &\geqq \overline{\lim_{k \to \infty}} \int_0^T [\tfrac{1}{2}|\dot{q}_{\varepsilon_k}|^2 - V(q_{\varepsilon_k})]\, dt \\ &= \int_0^T [\tfrac{1}{2}|\dot{q}|^2 - V(q)]\, dt. \end{aligned} \tag{2.29}$$

T being arbitrary, (2.29) shows $q \in E$ and $I(q) \leqq c$.

(2.30) **Lemma.** $q(\infty) = 0 = \dot{q}(\infty)$.

Proof. If $q(\infty)=0$, by (2.24), $\dot{q}(\infty)=0$. To prove that $q(\infty)=0$, an indirect argument is employed. If $q(\infty)\neq 0$, as in the proof of Lemma (2.11), there is a sequence $t_m\to\infty$ and $\xi\neq 0$ such that $q(t_m)\to\xi$ as $m\to\infty$. A slight modification of the argument of Lemma (2.11) shows that $V(\xi)=0$, i.e. $\xi\in\partial\Omega$ and therefore $V'(\xi)\neq 0$. Consequently there is a neighborhood U of ξ in $\mathbf{R}^n$ such that

$$|V'(x)-V'(\xi)|\leqq\tfrac{1}{2}|V'(\xi)| \tag{2.31}$$

for all $x\in U$. Since by (2.24), $\dot{q}$ is bounded in L^∞, there is a $\delta>0$ and $m_0\in\mathbf{N}$ such that

$$q\Big(\bigcup_{m\geqq m_0}[t_m, t_m+\delta]\Big)\subset U. \tag{2.32}$$

Hence for $t\in(t_m, t_m+\delta)$, by (2.31),

$$\begin{aligned}|\dot{q}(t)-\dot{q}(t_m)| &= \left|\int_{t_m}^{t} V'(q(s))\,ds\right| \\ &= \left|(t-t_m)\,V'(\xi)-\int_{t_m}^{t}(V'(\xi)-V'(q(s)))\,ds\right| \\ &\geqq \tfrac{1}{2}(t-t_m)|V'(\xi)|\end{aligned} \tag{2.33}$$

and

$$\begin{aligned}\int_{t_m}^{t_m+\delta}|\dot{q}(s)|^2\,ds &\geqq \frac{1}{\delta}\left(\int_{t_m}^{t_m+\delta}|\dot{q}(s)|\,ds\right)^2 \\ &\geqq \frac{1}{\delta}\left(\frac{\delta^2}{4}|V'(\xi)|-\delta|\dot{q}(t_m)|\right)^2.\end{aligned} \tag{2.34}$$

By (2.24), $\dot{q}(t_m)\to 0$ as $m\to\infty$. Hence by (2.34),

$$I(q)\geqq\tfrac{1}{2}\sum_{m=m_0}^{\infty}\int_{t_m}^{t_m+\delta}|\dot{q}(t)|^2\,dt=\infty, \tag{2.35}$$

contrary to Lemma (2.26). Hence $\xi=0$ and $q(\infty)=0$.

Note that by Lemmas (2.26) and (2.30), $q\in\Gamma$.

(2.36) **Lemma.** *q satisfies* (2.5) *and* (2.8) *and $q(t)\in\Omega$ for all $t\in(0,\infty)$.*

Proof. Lemma (2.26) shows $I(q)\leqq c\leqq\hat{c}$. Since $q\in\Gamma$, (2.5) holds. To verify (2.8), it suffices to prove that $q(t)\in\Omega$ for all $t\in(0,\infty)$ and therefore $q\in\hat{\Gamma}$. Let

$$\mathscr{T}=\{t\in(0,\infty)\,|\,q(t)\in\partial\Omega\}.$$

By Lemma (2.30), $\mathscr{T}$ is a bounded set. If $\mathscr{T}\neq\phi$, set $\tau=\max\{t\,|\,t\in\mathscr{T}\}$ and $Q(t)=q(t+\tau)$. Then $Q\in\hat{\Gamma}$ and $I(Q)\leqq I(q)$ with equality holding if and only if

$$q(t)\equiv q(0)\in\partial\Omega \quad \text{for } t\in[0,\tau]. \tag{2.37}$$

But $q(t)$ is a solution of (HS) for all $t>0$ and by hypothesis $V'(x)\neq 0$ on $\partial\Omega$. Therefore (2.37) cannot hold. Hence

$$\hat{c}\leqq I(Q)<I(q)=c, \tag{2.38}$$

contrary to (2.25).

The proof of Theorem (2.22) is complete. As an immediate consequence of the theorem we have:

(2.39) **Corollary.** *Under the hypotheses of Theorem* (2.22). (HS) *possesses a homoclinic solution* q *with* $q(\pm\infty)=0=\dot{q}(\pm\infty)$ *and* $q(0)\in\partial\Omega$, $\dot{q}(0)=0$.

Proof. Observing that (HS) is time reversible, the solution q obtained in Theorem (2.22) can be extended to $\mathbf{R}$ via $q(-t)=q(t)$. Since $\dot{q}(0)=0$, this extension furnishes a solution of the desired type.

(2.40) *Remark.* By combining the ideas of Theorems (2.1) and (2.22), one can prove a stronger version of Theorem (2.1) where the assumption that $V(x)<V(0)$ for all $x\in\bar{\Omega}\setminus\{0\}$ is replaced by $V(x)<V(0)$ for all $x\in\Omega\setminus\{0\}$ and $V(x)=V(0)$ implies $V'(x)\neq 0$ for $x\in\partial\Omega$.

(2.41) *Remark.* In Theorem (2.22), the assumption that $V'(x)\neq 0$ for all $x\in\partial\Omega$ can be eliminated. However then the conclusion becomes: either there exists a homoclinic orbit joining 0 to $\partial\Omega$ (as in Coroll. (2.39)) or there is an orbit of heteroclinic type joining 0 and $\partial\Omega$, i.e. there exists a $T^*\geqq-\infty$ and a solution q of (HS) with $q\in C^2((T^*,\infty),\mathbf{R}^n)$, $q(\infty)=\dot{q}(\infty)=0$ and

$$\operatorname{dist}(q(t),\partial\Omega)\to 0 \quad \text{as } t\to T^*. \tag{2.42}$$

(If $V'=0$ at only finitely many points on $\partial\Omega$, (2.42) simplifies to $q(t)$ approaches one of these points as $t\to T^*$.) We will not carry out the details of the proof, which is based on the arguments of Theorem (2.22) and results from [1]. However we will give a quick sketch. Arguing as in the proof of Theorem (2.22), q_{ε_k} converges in C^2_{loc} to $q\in C^2$, a solution of (HS) with $q(t)\in\bar{\Omega}$ for all $t\in\mathbf{R}^+$, $q(0)\in\partial\Omega$, and (2.24) holds. If the functions $q_\varepsilon(t)$ avoid a neighborhood of the set

$$S=\{x\in\partial\Omega\,|\,V'(x)=0\},$$

the arguments of Theorem (2.22) and Corollary (2.39) carry over to show q is a homoclinic orbit of (HS). If however,

$$\operatorname{dist}(q_\varepsilon(t),S)\to 0,$$

by appropriately rescaling time and using (2.23)(ii) and arguments from [1], a subsequence of (q_{ε_k}) will converge to a heteroclinic or homoclinic orbit of (HS) emanating from 0. See also the proof of Theorem (3.1) for related rescaling arguments.

§ 3 A refined version of Theorem (2.22)

Returning to the setting of Theorem (2.22), note that $\partial\Omega$ can have at most finitely many components, for otherwise there would exist an accumulation point

z of these components. Consequently $z\in\partial\Omega$ and $V(z)=0$. Hence $V'(z)\neq 0$. But then the Implicit Function Theorem shows there is only one component of $\partial\Omega$ near z. Let $S_1, \ldots, S_l$ denote the components of $\partial\Omega$. Our main result in this section, Theorem (3.1), refines Theorem (2.22) and gives a lower bound for the number of "periodic" and "homoclinic" orbits in the sense of Corollary (2.39), i.e. the homoclinic begin at 0 at $t=-\infty$ bounce off of some S_j at $t=0$ and return to 0 at $t=\infty$. Likewise the periodic orbits bounce back and forth between a pair of components of $\partial\Omega$ in finite time.

A result due to Bolotin and Kozlov, related to Theorem (3.1), but when there are no homoclinic orbits present and involving a rather different proof can be found in [2, Theor. (3)].

(3.1) **Theorem.** *Under the hypotheses of Theorem* (2.22), *let* l_h *be the number of homoclinic orbits of* (HS) *emanating from* 0 *and let* l_p *be the number of periodic (after reflection) orbits of* (HS) *which have energy* 0 *and which join components of* $\partial\Omega$. *Then* $l_k+l_p\geqq l$.

Proof. Note that Γ is the union of l components:

$$\Gamma_i=\{q\in E\,|\,q(0)\in S_i,\ q(\infty)=0,\ \text{and}\ q(t)\in\bar{\Omega}\ \text{for all}\ t\in\mathbf{R}^+\},\qquad 1\leqq i\leqq l. \tag{3.2}$$

Similarly for ε small, Γ_ε consists of l components:

$$\Gamma_{i,\varepsilon}=\{q\in E\,|\,q(0)\in S_{i,\varepsilon},\ q(\infty)=0,\ \text{and}\ q(t)\in\bar{\Omega}_\varepsilon\ \text{for all}\ t\in\mathbf{R}^+\},\qquad 1\leqq i\leqq l, \tag{3.3}$$

where

$$S_{i,\varepsilon}=\{x\in\Omega\,|\,\operatorname{dist}(x,S_i)=\varepsilon\},\qquad 1\leqq i\leqq l. \tag{3.4}$$

Sets $\hat{\Gamma}_i$, $\hat{\Gamma}_{i,\varepsilon}$ can be defined similarly.

Let

$$c_i=\inf_{q\in\Gamma_i} I(q),\qquad 1\leqq i\leqq l \tag{3.5}$$

and

$$c_{i,\varepsilon}=\inf_{q\in\Gamma_{i,\varepsilon}} I(q),\qquad 1\leqq i\leqq l. \tag{3.6}$$

We fix i and consider $c_{i,\varepsilon}$. It follows from the proof of Theorem (2.1) that there is a $q_\varepsilon\in\Gamma_i$ such that $I(q_\varepsilon)=c_{i,\varepsilon}$. Moreover by the argument of Proposition (3.18) of [1], q_ε is a solution of (HS) for $t\in(0,\infty)\setminus\mathscr{T}_\varepsilon$ where

$$\mathscr{T}_\varepsilon=\left\{t\in(0,\infty)\,\middle|\,q_\varepsilon(t)\in\bigcup_{r=1}^{l} S_{r,\varepsilon}\right\}. \tag{3.7}$$

Two possibilities occur: (i) $\mathscr{T}_\varepsilon\neq\phi$ for all small $\varepsilon>0$ or (ii) $\mathscr{T}_\varepsilon=\phi$ for a sequence $\varepsilon_m\to 0$.

Case (i). $\mathscr{T}_\varepsilon\neq\phi$ for all small $\varepsilon>0$.

This is the more complicated situation. We will show that there is a "chain" which consists of pieces of periodic and homoclinic orbits of (HS) and which joins S_i and 0. More precisely we have:

(3.8) **Proposition.** *If $\mathscr{T}_\varepsilon \neq \phi$ for all small $\varepsilon > 0$, there is a $p = p(i) \leqq l$ and a map j: $\{1, \ldots, p\} \to \{1, \ldots, l\}$ where $j(1) = i$ and $j(s) \neq j(s')$ if $s \neq s'$, numbers $T_1, \ldots, T_p > 0$, and functions $Q_1, \ldots, Q_{p+1}$ such that:*

(a) *Q_s is a 2 T_s periodic solution of* (HS), $1 \leqq s \leqq p$,

(b) $Q_s(0) \in S_{j(s)}$, $Q_s(T_s) \in S_{j(s+1)}$, $\dot{Q}_s(0) = 0 = \dot{Q}_s(T_s)$, $1 \leqq s \leqq p$,

(c) $Q_s(t) \in \Omega$ *for all* $t \in (0, T_s)$,

(d) Q_{p+1} *is a "homoclinic" orbit of* (HS) *such that* $Q_{p+1} \in \hat{\Gamma}_{j(p)}$ *and* $\dot{Q}_{p+1}(0) = 0$.

(e) *Setting* $T_{p+1} = \infty$,

$$\sum_{s=1}^{p+1} \int_0^{T_s} [\tfrac{1}{2}|\dot{Q}_s|^2 - V(Q_s)]\, dt = \inf_{\Gamma_i} I = \inf_{\hat{\Gamma}_i} I,$$

(f) $\tfrac{1}{2}|\dot{Q}_s(t)|^2 + V(Q_s(t)) = 0$ *on* $[0, T_s]$, $1 \leqq s \leqq p+1$.

Proof. Since $\mathscr{T}_\varepsilon \neq \phi$ for all small $\varepsilon > 0$, numbers

$$\sigma_1^\varepsilon \geqq \tau_2^\varepsilon > \sigma_2^\varepsilon \geqq \ldots > \sigma_{p+1}^\varepsilon = 0 \tag{3.9}$$

can be defined as follows:

$$\sigma_1^\varepsilon = \sup\{t \in (0, \infty) \mid q_\varepsilon(t) \in \bigcup_r S_{r,\varepsilon}\}. \tag{3.10}$$

Suppose $q_\varepsilon(\sigma_1^\varepsilon) \in S_{r_1,\varepsilon}$, r_1 depending on ε. Clearly $r_1 \neq i$ via (3.6). Set

$$\tau_2^\varepsilon = \inf\{t \in [0, \sigma_1^\varepsilon] \mid q_\varepsilon(t) \in S_{r_1,\varepsilon}\} \tag{3.11}$$

so $\tau_2^\varepsilon \in (0, \sigma_1^\varepsilon]$. Set

$$\sigma_2^\varepsilon = \sup\{t \in [0, \tau_2^\varepsilon) \mid q_\varepsilon(t) \in \bigcup_r S_{r,\varepsilon}\} \tag{3.12}$$

If $\sigma_2^\varepsilon = 0$, $p = 1$ and we are through. If not, say $q_\varepsilon(\sigma_2^\varepsilon) \in S_{r_2,\varepsilon}$. Clearly $r_2 \neq i$ or r_1. Set

$$\tau_3^\varepsilon = \inf\{t \in [0, \sigma_2^\varepsilon] \mid q_\varepsilon(t) \in S_{r_2,\varepsilon}\} \tag{3.13}$$

Continuing in this fashion, in at most l steps, we find the σ's and τ's as in (3.9). Note that p and the indices r_k depend on ε as well as i and $r_{p+1} = i$. However since only at most $l-1$ sets are involved, a subsequence $\varepsilon_m \to 0$ can be chosen so that p and $r_1, \ldots$ are independent of ε. We set $j(s) = r_{p-s+1}$.

(3.14) **Proposition.** *For ε sufficiently small,*

(i) *There is a $\delta > 0$ (independent of ε) such that $q_\varepsilon(t) \notin B_\delta(0)$ for all $t \in \bigcup_{s=2}^{j+1} (\sigma_s^\varepsilon, \tau_s^\varepsilon)$,*

(ii) *$q_\varepsilon(t)$ is a solution of* (HS) *on $(\sigma_s^\varepsilon, \tau_s^\varepsilon)$ for each $s \in \{2, \ldots, p+1\}$,*

(iii) *There is an $M > 0$ (independent of ε) such that $|\sigma_s^\varepsilon - \tau_s^\varepsilon| \leqq M$ for $s \in \{2, \ldots, p+1\}$.*

Proof. If (i) is not true, for some $s\in\{2, \ldots, p+1\}$ there is a sequence $\varepsilon_m\to 0$ and $\mu_s^\varepsilon\in(\sigma_s^\varepsilon, \tau_s^\varepsilon)$ (with $\varepsilon=\varepsilon_m$) such that $q_\varepsilon(\mu_s^\varepsilon)\to 0$ as $m\to\infty$. Set

$$Q_\varepsilon(t)=\begin{cases} q_\varepsilon(t), & t\in(0, \mu_s^\varepsilon) \\ (\mu_s^\varepsilon+1-t)\, q_\varepsilon(\mu_s^\varepsilon), & t\in[\mu_s^\varepsilon, \mu_s^\varepsilon+1] \\ 0 & t\geqq\mu_s^\varepsilon+1. \end{cases} \tag{3.15}$$

Then $Q_\varepsilon(t)\in\Gamma_{i,\varepsilon}$ and

$$I(Q_\varepsilon)-I(q_\varepsilon)=\int_{\mu_s^\varepsilon}^{\mu_s^\varepsilon+1}[\tfrac{1}{2}|q_\varepsilon(\mu_s^\varepsilon)|^2-V(Q_\varepsilon)]\,dt - \int_{\mu_s^\varepsilon}^{\infty}[\tfrac{1}{2}|\dot q_\varepsilon(t)|^2-V(q_\varepsilon(t))]\,dt. \tag{3.16}$$

The first integral on the right hand side of (3.16) approaches 0 as $m\to\infty$ while Lemma (2.9) shows the second term is bounded away from 0. Hence $I(Q_{\varepsilon_m})<I(q_{\varepsilon_m})$ for m large, contrary to (3.6). To prove (ii), it suffices to show that $\mathscr{T}_\varepsilon\cap(\sigma_s^\varepsilon, \tau_s^\varepsilon)=\phi$. But this is immediate from the definition of the σ's and τ's and (3.6). Lastly to get (iii), suppose that $\tau_s^\varepsilon-\sigma_s^\varepsilon\to\infty$ along some sequence $\varepsilon_m\to 0$. Then by earlier arguments $q_{\varepsilon_m}(t-\sigma_s^{\varepsilon_m})$ converges in $C^2_{\mathrm{loc}}(\mathbf{R}^+, \mathbf{R}^n)$ to $\tilde q\in C^2(\mathbf{R}^+, \mathbf{R}^n)$ such that $\tilde q$ satisfies (HS), $\tilde q(t)\in\bar\Omega$ for all $t\in\mathbf{R}^+$, $\tilde q(0)\in S_{r_s}$, and

$$I(\tilde q)\leqq\varliminf_{m\to\infty} I(q_{\varepsilon_m})<\infty. \tag{3.17}$$

The argument of Lemma (2.30) shows $\tilde q(\infty)=0$. But $q_\varepsilon(t)\notin B_\delta(0)$ for all $t\in(\sigma_s^\varepsilon, \tau_s^\varepsilon)$ via (i) so this is impossible and there exists M as stated.

Now the functions Q_s can be constructed. Letting $m\to\infty$, by Proposition (3.14)(iii), a subsequence ε_m can be chosen so that

$$\varliminf_{m\to\infty}\tau_s^{\varepsilon_m}-\sigma_s^{\varepsilon_m}$$

exists. Denoting this limit by T_{p-s+2}, as above along a subsequence,

$$\begin{aligned} q_{\varepsilon_m}(t-\sigma_1^{\varepsilon_m}) &\to Q_{p+1}(t) && \text{in } C^2_{\mathrm{loc}}(\mathbf{R}^+, \mathbf{R}^n), \\ q_{\varepsilon_m}(t-\sigma_2^{\varepsilon_m}) &\to Q_p(t) && \text{in } C^2([0, T_p], \mathbf{R}^n), \\ &\vdots \\ q_{\varepsilon_m}(t-\sigma_{p+1}^{\varepsilon_m}) &\to Q_1(t) && \text{in } C^2([0, T_1], \mathbf{R}^n). \end{aligned}$$

By Proposition (3.14)(ii) and our construction, Q_s is a solution of (HS) on $(0, T_s)$ with $Q_s(0)\in S_{j(s)}$, $Q_s(T_s)\in S_{j(s+1)}$. Moreover weaker forms of (c) and (d) of Proposition (3.8) hold:

(3.18) ($\bar c$) $Q_s(t)\in\bar\Omega$ for all $t\in[0, T_s]$.
($\bar d$) Q_{p+1} is a "homoclinic" orbit of (HS) such that $Q_{p+1}\in\Gamma_{j(p)}$ and $Q_{p+1}(0)=0$.

Assume (e) and (f) of Proposition (3.8) for the moment. Then the remainder of (b) follows on extending Q_s as an even function about 0 and T_s.

Proof of (e). Clearly

$$(3.19)\quad \sum_{s=1}^{p+1}\int_0^{T_s}[\tfrac{1}{2}|\dot{Q}_s|^2-V(Q_s)]\,dt \leqq \varliminf_{m\to\infty}\sum_{s=1}^{p+1}\int_{\sigma_s^{\varepsilon_m}}^{\tau_s^{\varepsilon_m}}[\tfrac{1}{2}|\dot{q}_{\varepsilon_m}|^2-V(q_{\varepsilon_m})]\,dt$$
$$\leqq \varliminf_{m\to\infty} I(q_{\varepsilon_m})=\lim_{\varepsilon\to 0}\inf_{w\in\hat{\Gamma}_{i,\varepsilon}} I(w)\leqq \inf_{w\in\hat{\Gamma}_i} I(w),$$

the last inequality following as in (2.28). To show that equality holds in (3.19), note first that since $\partial\Omega$ is C^1, for any $w\in\Gamma_i$ with $I(w)<\infty$, there exists a sequence $w_k\in\hat{\Gamma}_i$ such that $w_k\to w$ in E and $I(w_k)\to I(w)$. Therefore it suffices to find a family of functions $(\eta_r(t))\subset\Gamma_i$ such that

$$(3.20)\quad I(\eta_r)\to\sum_{s=1}^{p+1}\int_0^{T_s}[\tfrac{1}{2}|\dot{Q}_s|^2-V(Q_s)]\,dt$$

as $r\to\infty$. Let $v_s\in C^1([0,1],S_{j(s+1)})$ such that $v_s(0)=Q_s(T_s)$ and $v_s(1)=Q_{s+1}(0)$. Let $v_{s,r}=v_s(t/r)\in C^1([0,r],S_{j(s+1)})$. Finally let $\eta_r(t)$ be the curve obtained in a natural way by joining the functions $Q_1(t)$, $v_{1,r}(t)$, $Q_2(t)$, $v_{2,r}(t)$, ..., $Q_{p+1}(t)$. Then $\eta_r\in\Gamma_i$. Since

$$(3.21)\quad \int_0^r[\tfrac{1}{2}|\dot{v}_{r,s}|^2-V(v_{r,s})]\,dt=\frac{1}{2r^2}\int_0^r|\dot{v}_s(t/r)|^2\,dt\to 0$$

as $r\to\infty$, (3.20) follows and (e) is proved.

Proof of (f). Since as in Lemma (2.30), $Q_{p+1}(\infty)=0=\dot{Q}_{p+1}(\infty)$, (f) holds for $s=p+1$. Thus suppose $1\leqq s\leqq p$. Since by the part of (b) already established, $V(Q_s(0))=0$, it suffices to prove $\dot{Q}_s(0)=0$. If not, a function $w\in\Gamma_i$ will be constructed such that

$$(3.22)\quad I(w)<\inf_{\Gamma_i} I.$$

But (3.22) is contrary to (e) so (f) must hold. To find w, let $a>0$ be small and $L>1$. Set

$$(3.23)\quad \xi(t)=Q_s\left(\frac{t}{L}\right),\qquad t\in(0,La)$$
$$=Q_s(t-La+a),\qquad t\in[La,T_s+La-a].$$

Then for small a,

$$(3.24)\quad \int_0^{T_s+La-a}(\tfrac{1}{2}|\dot{\xi}|^2-V(\xi))\,dt-\int_0^{T_s}[\tfrac{1}{2}|\dot{Q}_s|^2-V(Q_s)]\,dt$$
$$=\int_0^{La}\left[\frac{1}{2L^2}\left|\dot{Q}_s\left(\frac{t}{L}\right)\right|^2-V\left(Q_s\left(\frac{t}{L}\right)\right)\right]dt-\int_0^a[\tfrac{1}{2}|\dot{Q}_s(t)|^2-V(Q_s(t))]\,dt$$
$$=\left(\frac{1}{2L^2}|\dot{Q}_s(0)|^2+o(a)\right)La-(\tfrac{1}{2}|\dot{Q}_s(0)|^2+o(a))a<0.$$

Introducing a function η_r as in the proof of (e) of Proposition (3.8) with ξ replacing Q_s, (3.24) shows

$$\lim_{r\to\infty} I(\eta_r) < \inf_{\Gamma_i} I, \tag{3.25}$$

contrary to (e).

It remains to prove (c) and (d). Thus suppose

$$Q_s(0, T_s) \cap (\bigcup S_j) \neq \phi. \tag{3.26}$$

If $Q_s(\tau)$ lies in this intersection, $V(Q_s(\tau))=0$ while $V'(Q_s(\tau))\neq 0$. It follows that the set of such τ's must be isolated. Suppose there are $m(s)$ such points: $0<\tau_{1,s}<\ldots<\tau_{m(s),s}<T_s$. Set $T_{k,s}=t_{k,s}-t_{k-1,s}$ and

$$Q_{k,s}(t)=Q_s(t-\tau_{k-1,s})\colon [0, T_{k,s}]\to \mathbf{R}^n, \qquad 1\leqq k\leqq m(s).$$

Consider the new set of $\mu=\sum_{s=1}^{k+1} m(s)$ functions

$$P_1(t)=Q_{1,1}(t), \ldots, P_{m(1)}(t)=Q_{m(1),1}(t), P_{1+m(1)}(t)=Q_{1,2}(t), \text{ etc.}$$

on the associated t intervals. We claim $\mu\leqq l$. Otherwise for some $k<\hat{k}$, $P_k(\hat{T}_k)$ and $P_{\hat{k}}(\hat{T}_{\hat{k}})\in\partial S_\rho$ for associated $\hat{T}_k$, $\hat{T}_{\hat{k}}$ and some ρ. Then as in the proof of (e), the functions $P_1, \ldots, P_k, P_{\hat{k}+1}, \ldots, P_\mu$ can be made part of a curve η_r such that as $r\to\infty$,

$$I(\eta_r)\to \sum_{\substack{s=1\\ s\notin[k+1,\hat{k}]\cap\mathbf{N}}}^{\mu} \int_0^{\hat{T}_s} [\tfrac{1}{2}|\dot{P}_s|^2 - V(P_s)]\,dt < \sum_{s=1}^{\mu} \int_0^{\hat{T}_s} [\tfrac{1}{2}|\dot{P}_s|^2 - V(P_s)]\,dt, \tag{3.27}$$

contrary to (e). Thus (c) and (d) hold and the proof of Proposition (3.8) and analysis of case (i) are complete.

Now we turn to:

Case (ii). $\mathscr{T}_\varepsilon=\phi$ for a sequence $\varepsilon_m\to 0$.

Then by the arguments of Theorem (2.22), especially Lemmas (2.26) and (2.30), along a subsequence q_{ε_m} converges on C^2_{loc} to a solution q of (HS) with $q\in\Gamma_i$ and

$$I(q)=\inf_{w\in\Gamma_i} I(w). \tag{3.28}$$

Hence q reflected about $t=0$ provides the homoclinic orbit for this case.

(3.29) *Remark.* For case (ii), it is possible that for some $\gamma>0$, $q(\gamma)\in S_j$ for some $j\neq i$. However as in the construction of the σ's and τ's in (3.9), there can be at most $l-1$ such points, each corresponding to a different set S_j. If V is C^2, there cannot exist any such points γ since the solution to the initial value problem for (HS) with $w(\gamma)=x\in S_j$ and $\dot{w}(\gamma)=0$ would have as solutions both a homoclinic orbit joining x and 0 and a periodic orbit joining x and S_i.

(3.30) *Remark.* The case (ii) follows in particular for any i such that

$$c_i = \min_{1 \leqq j \leqq l} c_j.$$

Completion of the proof of Theorem (3.1) It remains only to show that

$$l_h + l_p \geqq l. \tag{3.31}$$

Let m be the number of chains of homoclinic and periodic solutions of energy 0 joining the set of S_j's to 0. We have shown for each S_i, there is at least one such chain. Hence $m \geqq l$. On the other hand, it is easy to see that $m \leqq l_h + l_p$ and (3.31) follows.

(3.32) *Remark.* The hypothesis that $V'(x) \neq 0$ on $\partial\Omega$ can be weakened. See Remark (2.41). One still gets an analogue of (3.31) where heteroclinic orbits are also included provided that one has enough regularity, e.g. C^1, for $\partial\Omega$.

§ 4 $\Omega = \mathbf{R}^n$

In this section, a variant of Theorem (2.1) will be proved for $\Omega = \mathbf{R}^n$. The existence of an orbit emanating from infinity and terminating at 0 will also be established.

(4.1) **Theorem.** *Suppose* $V \in C^1(\mathbf{R}^n, \mathbf{R})$, $V(0) > V(x)$ *for all* $x \in \mathbf{R}^n \setminus \{0\}$, *and*

$$\overline{\lim_{|x| \to \infty}} V(x) < V(0). \tag{4.2}$$

Then for each $\xi \in \mathbf{R}^n \setminus \{0\}$, *there is a solution* $q = q(t)$ *of* (HS) *such that* $q(0) = \xi$ *and* $q(\infty) = 0 = \dot{q}(\infty)$.

Proof. Again, without loss of generality, assume $V(0) = 0$. Set

$$\Gamma_\xi = \{w \in E \mid w(0) = \xi \text{ and } w(\infty) = 0\} \tag{4.3}$$

and

$$c_\xi = \inf_{w \in \Gamma_\xi} I(w). \tag{4.4}$$

We claim c_ξ is a critical value of I and any corresponding critical point q_ξ is the desired solution of (HS). The proof parallels that of Theorem (2.1) but is a bit simpler so we will omit it.

(4.5) *Remark.* A more general result than Theorem (4.1) has been obtained by Bolotin and Kozlov using arguments from Riemannian geometry. See e.g. [2, Theor. 10].

(4.6) *Remark.* Theorem (4.1) shows there is a solution of (HS) of energy 0 joining each point in $\mathbf{R}^n$ to 0. Thus it is natural to ask whether there is a 0 energy orbit joining 0 and ∞. The next result established the existence of such an

orbit emanating from ∞ and terminating at 0. Reversing time, we get an "escape orbit" of (HS). Since no assumptions have been made on the behavior of $V(x)$ as $|x|\to\infty$, this escape may occur in finite time. More precisely we have:

(4.7) **Theorem.** *Under the hypotheses of Theorem* (4.1), *there is a solution* q *of* (HS) *such that* $q(\infty)=0=\dot{q}(\infty)$ *and* $|q(t)|\to\infty$ *as* $t\to T^*$ *where* $T^*=0$ *or* $-\infty$.

Proof. We will find a solution q of (HS) such that $q(\infty)=0$, $q(0)\in\partial B_1$, and $|q(t)|\to\infty$ as $t\to T$ for some $T\in[-\infty, 0)$. Rescaling time if $T>-\infty$ yields the statement of the Theorem.

Two simple preliminary results are needed for the proof of Theorem (4.7).

(4.8) **Lemma.** *Let* Γ_ξ *and* c_ξ *be as defined in* (4.3)–(4.4). *Suppose* $q_\xi\in\Gamma_\xi$ *and* $I(q_\xi)=c_\xi$. *Let* $\tau\in(0,\infty)$ *and* $\eta=q_\xi(\tau)$. *Then* $w(t)=q_\xi(t+\tau)\in\Gamma_\eta$ *and* $I(w)=c_\eta$.

Proof. Clearly $w\in\Gamma_\eta$. If $I(w)>c_\eta=I(q_\eta)$, set

$$\begin{aligned} Q(t)&=q_\xi(t), && t\in[0,\tau]\\ &=q_\eta(t-\tau), && t\geqq\tau. \end{aligned}$$

Then $Q\in\Gamma_\xi$ but

$$\begin{aligned} I(Q)&=\int_0^\tau [\tfrac{1}{2}|\dot{q}_\xi|^2-V(q_\xi)]\,dt+c_\eta\\ &<\int_0^\tau [\tfrac{1}{2}|\dot{q}_\xi|^2-V(q_\xi)]\,dt+\int_0^\infty [\tfrac{1}{2}|\dot{w}|^2-V(w)]\,dt\\ &=I(q_\xi), \end{aligned} \tag{4.9}$$

contrary to (4.4).

(4.10) **Proposition.** *Let* $\xi_k\to\xi$ *in* $\mathbf{R}^n$. *Then there is a sequence* $q_{\xi_k}\in\Gamma_{\xi_k}$ *such that* $I(q_{\xi_k})=c_{\xi_k}$ *and* $q_{\xi_k}\to q_\xi$ *in* $C^2_{\mathrm{loc}}(\mathbf{R}^+,\mathbf{R}^n)$ *along a subsequence where* $q_\xi\in\Gamma_\xi$ *and* $I(q_\xi)=c_\xi$.

Proof. Since (ξ_k) is a bounded sequence, it is easily verified via (4.4) that the critical values c_{ξ_k} are bounded, e.g. by M, independently of k. It then follows from Lemma (2.9) that the functions q_{ξ_k} are bounded in $L^\infty(\mathbf{R}^+,\mathbf{R}^n)$. Hence (HS) implies (q_{ξ_k}) is bounded in $C^2(\mathbf{R}^+,\mathbf{R}^n)$. Consequently, as earlier, a subsequence of q_{ξ_k} converges in C^2_{loc} to a solution q of (HS) such that $q(0)=\xi$. Moreover $I(q_{\xi_k})\leqq M$ implies $I(q)\leqq M$. The form of I shows $q\in E$ and a slight variation of Proposition (2.11) gives $q(\infty)=0$. Thus $q\in\Gamma_\xi$.

It remains to prove that $I(q)=c_\xi$. If not, by (4.4), $I(q)>c_\xi$. Choose α so that

$$I(q)>\alpha>c_\xi. \tag{4.11}$$

Hence there is a $T>0$ such that

$$\int_0^T [\tfrac{1}{2}|\dot{q}|^2-V(q)]\,dt>\alpha>c_\xi. \tag{4.12}$$

Since $q_{\xi_k} \to q_\xi$ in C^2_{loc} along a subsequence, for large such k,

$$\int_0^T [\tfrac{1}{2}|\dot{q}_{\xi_k}|^2 - V(q_{\xi_k})]\,dt > \alpha > c_\xi. \tag{4.13}$$

Therefore

$$c_{\xi_k} = I(q_{\xi_k}) > \alpha > c_\xi \tag{4.14}$$

for such k. On the other hand, suppose $w \in \Gamma_\xi$ and $I(w) = c_\xi$. Let $\varepsilon > 0$ and set

$$\begin{aligned} Q_k(t) &= \left(1 - \frac{t}{\varepsilon}\right)\xi_k + \frac{t}{\varepsilon}\xi \qquad t \in [0, \varepsilon] \\ &= w(t-\varepsilon) \qquad\qquad t > \varepsilon. \end{aligned} \tag{4.15}$$

Then $Q_k \in \Gamma_{\xi_k}$ and

$$c_{\xi_k} \leqq I(Q_k) = \frac{1}{2\varepsilon}|\xi_k - \xi|^2 - \int_0^\varepsilon V(Q_k(t))\,dt + c_\xi. \tag{4.16}$$

Thus (4.15)–(4.16) show

$$\overline{\lim_{k\to\infty}}\, c_{\xi_k} \leqq -\varepsilon V(\xi) + c_\xi. \tag{4.17}$$

Since ε is arbitrary, (4.17) is contrary to (4.14). Hence $c_\xi = I(q)$.

Now we are ready for the:

Proof of Theorem (4.7) Let $(x_k) \subset \mathbf{R}^n$ satisfy $|x_k| \to \infty$ as $k \to \infty$. For each k, there is a solution $q_k \equiv q_{x_k}$ of (HS) given by Theorem (4.1). Since $q_k(\infty) = 0$, there is a $\tau_k > 0$ such that $\xi_k \equiv q_k(\tau_k) \in \partial B_1(0)$ and $q_k(t) \in B_1(0)$ for all $t > \tau_k$. Set $\tilde{q}_k(t) = q_k(t + \tau_k)$. Therefore $\tilde{q}_k(0) \in \partial B_1$ and $\tilde{q}_k$ is a solution of (HS). In fact, by Lemma (4.8), it can be assumed that $\tilde{q}_k = q_{\xi_k}$. Since $(\xi_k) \subset \partial B_1$, by choosing a subsequence if necessary, we have $\xi_k \to \xi \in \partial B_1$. Consequently by Proposition (4.10), a subsequence of $\tilde{q}_k$ converges in C^2_{loc} to $q \in \Gamma_\xi$ with $I(q) = c_\xi$. This function q is defined for $t \in \mathbf{R}^+$. We will show there is a $T < 0$ such that q extends to a solution of (HS) in such a way that it satisfies the conclusions of Theorem (4.7).

By the choice of (x_k), there is a $\tau_{k,1} < \tau_k$ such that $q_k(\tau_{k,1}) \in \partial B_2$ and $q_k(t) \in B_2$ for all $t > \tau_{k,1}$. Consider the sequence $(\tau_k - \tau_{k,1})$. We claim this sequence is bounded. If not, $\tau_k - \tau_{k,1} \to \infty$ along a subsequence. Clearly

$$\int_{\tau_{k,1}}^{\tau_k} [\tfrac{1}{2}|\dot{q}_k|^2 - V(q_k)]\,dt \leqq c_{q_k(\tau_k,\,1)} \leqq \sup\{c_\eta \mid \eta \in \partial B_2\} < \infty. \tag{4.18}$$

Since $-V$ is bounded away from 0 in any deleted neighborhood of 0, (4.18) shows for any $\varepsilon > 0$ and for all large k, the orbit q_k must enter B_ε. Therefore there is an $s_k \in (\tau_{k,1}, \tau_k)$ such that $q_k(s_k) \in B_\varepsilon$. Set

$$\begin{aligned} Q_k(t) &= q_k(t), && t \in [0, s_k] \\ &= (s_k + 1 - t)\,q_k(s_k), && t \in [s_k, s_k + 1] \\ &= 0, && t > s_k + 1 \end{aligned} \tag{4.19}$$

Then $Q_k \in \Gamma_{x_k}$ and

$$I(Q_k) - I(q_k) = \tfrac{1}{2}|q_k(s_k)|^2 - \int_{s_k}^{s_k+1} V(Q_k(t))\,dt - \int_{s_k}^{\infty} [\tfrac{1}{2}|\dot{q}_k|^2 - V(q_k)]\,dt. \tag{4.20}$$

The first two terms on the right hand side of (4.20) go to 0 as $\varepsilon \to 0$. The curve q_k joins B_ε to $B_2 \setminus B_1$ to 0 as t varies between s_k and ∞. Hence

$$\int_{s_k}^{\infty} [\tfrac{1}{2}|\dot{q}_k|^2 - V(q_k)]\,dt \tag{4.21}$$

has a positive lower bound independent of ε via Lemma (2.9). But then $I(Q_k) < I(q_k)$, contrary to the choice of q_k. It follows that $(\tau_k - \tau_{k,1})$ is a bounded sequence. Hence along a subsequence $\tau_k - \tau_{k,1} \to t_1 > 0$.

By appropriately rescaling time intervals again, a subsequence of $\tilde{q}_k$ (or q_k) converges to an extension of q (as a solution of (HS)) on $[-t_1, 0]$ such that $q(-t_1) \in \partial B_2$ and $q(t) \in B_2$ for $t \in (-t_1, 0)$. Continuing this construction yields an extension of q as a solution of (HS) to a maximal interval (T, ∞). Moreover there exists an increasing sequence t_k such that $q(-t_k) \in \partial B_{k+1}$ and $q(t) \in B_{k+1}$ for $t > -t_k$. This implies that

$$T = \inf_{k \in \mathbf{N}} \{-t_k\}$$

and $q(t)$ is not bounded as $t \to T$.

It remains to prove that $|q(t)| \to \infty$ as $t \to T$. Suppose that this is not the case. Then there is a $\mu > 0$ and sequences $s_k \to T$, $\sigma_k \to T$ such that

$$s_1 > \sigma_1 > s_2 > \sigma_2 > \ldots,$$

$|q(s_k)| = \mu$, $|q(\sigma_k)| = \mu + 1$, and $\mu < |q(t)| < \mu + 1$ for $t \in (\sigma_k, s_k)$. Now along a subsequence,

$$q(s_k) = \lim_{j \to \infty} \tilde{q}_j(s_k) \tag{4.22}$$

and by Proposition (4.10),

$$\int_{s_k}^{\infty} [\tfrac{1}{2}|\dot{q}|^2 - V(q)]\,dt = \inf_{w \in \Gamma_{q(s_k)}} I(w) \leqq \max\{c_\eta \mid \eta \in \partial B_\mu(0)\} \leqq M^*. \tag{4.23}$$

It is clear that M^* is bounded independently of k. Hence letting $k \to \infty$,

$$\int_{T}^{\infty} [\tfrac{1}{2}|\dot{q}|^2 - V(q)]\,dt \leqq M^*. \tag{4.24}$$

On the other hand, let

$$\beta_1 = \min_{\mu \leqq |x| \leqq \mu + 1} -V(x). \tag{4.25}$$

Then by Lemma (2.9),

$$\int_{\sigma_k}^{\tau_k} [\tfrac{1}{2}|\dot{q}|^2 - V(q)]\,dt \geqq \sqrt{2\beta_1}\,|q(\tau_k) - q(\sigma_k)| \geqq \sqrt{2\beta_1} \tag{4.26}$$

so

$$\int_{T}^{\infty} [\tfrac{1}{2}|\dot{q}|^2 - V(q)]\,dt = \infty \tag{4.27}$$

contrary to (4.24). Thus $|q(t)| \to \infty$ as $t \to T$ and the proof is complete.

For our final result in this section we will given an analogue of Theorem (4.7) for a potential V which is singular.

(4.28) **Theorem.** *Suppose there is an* $x_0 \in \mathbf{R}^n \backslash \{0\}$ *and* $V \in C_1(\mathbf{R}\backslash\{x_0\}, \mathbf{R})$ *such that* $V(0)=0$, $V(x)<0$ *for* $x \in \mathbf{R}^n \backslash \{0, x_0\}$

$$\overline{\lim_{|x| \to \infty}}\ V(x) < 0, \tag{4.29}$$

and $V(x) \to -\infty$ *as* $x \to x_0$. *Then there is a solution,* q, *of* (HS) *such that* $q(\infty)=0$ *and* $q(t) \to x_0$ *as* $t \to T^*$ *where* $T^*=0$ *or* $-\infty$.

Proof. Since the proof follows lines previously explored, we will only sketch the proof. Let $\varepsilon_k \to 0$. Let $\Omega_k = \mathbf{R}^n \backslash B_{\varepsilon_k}(x_0)$. Then by a combination of the proofs of Theorems (2.1) and (4.1), there is an $x_k \in \partial B_{\varepsilon_k}(x_0)$ and a solution q_k of (HS) such that $q_k(0)=x_k$, $q_k(\infty)=0$, and $q_k(t) \in \Omega_k$ for $t \in (0, \infty)$. Let $\rho \in (0, |x_0|)$. Hence there is a $\tau_k > 0$ such that $\xi_k \equiv q_k(\tau_k) \in \partial B_\rho$ and $q_k(t) \in B_\rho$ for all $t > \tau_k$. Now continuing as in the proof of Theorem (4.7), we get q as a limit of a subsequence of (q_k) after rescaling time.

§ 5 Some time dependent cases

In this section the existence of homoclinic orbits for some time-dependent Hamiltonian systems will be studied. Consider

$$\ddot{q} + V_q(t, q) = 0 \tag{5.1}$$

where V satisfies

(V_1) $V(t, x) = -\frac{1}{2} L(t)x \cdot x + W(t, x)$,

(V_2) $L(t) \in C(\mathbf{R}, \mathbf{R}^{n^2})$ is a positive definite symmetric matrix for all $t \in \mathbf{R}$,

(V_3) $W \in C^1(\mathbf{R} \times \mathbf{R}^n, \mathbf{R})$ and there is a constant $\mu > 2$ such that

$$0 < \mu W(t, x) \leqq x \cdot W_q(t, x)$$

for all $x \in \mathbf{R}^n \backslash \{0\}$ and $t \in \mathbf{R}$,

(V_4) $W_q(t, x) = o(|x|)$ as $x \to 0$ uniformly in $t \in \mathbf{R}$.

Note that (V_3)–(V_4) imply

$$W(t,x)=o(|x|^2) \tag{5.2}$$

as $x\to 0$, uniformly for $t\in\mathbf{R}$. Hence by (V_1)–(V_4), $x=0$ is a local maximum for all $t\in\mathbf{R}$. However it is not a global maximum since by (V_3) for each $t\in\mathbf{R}$ is an $a_1(t)$ such that

$$W(t,x)\geqq a_1(t)|x|^\mu \quad \text{for large } |x|. \tag{5.3}$$

We do not know if simple minimization arguments in the spirit of §2–4 can be used to treat (5.1). In [5] assuming (V_1)–(V_4), and that L and W are T-periodic in t, it was shown that (5.1) possesses a homoclinic orbit emanating from 0. Analogous results for general Hamiltonian systems were obtained by Coti-Zelati, Ekeland and Sere [3] and Hofer and Wysocki [4]. We study (5.1) without periodicity assumptions on L and W. Two results will be obtained. The first is:

(5.4) **Theorem.** *Suppose* V *satisfies* (V_1)–(V_4) *and*

(V_5) *The smallest eigenvalue of* $L(t)\to\infty$ *as* $|t|\to\infty$, *i.e.*

$$\inf_{|\xi|=1} L(t)\xi\cdot\xi\to\infty \quad \text{as } |t|\to\infty,$$

(V_6) *There is a* $\bar{W}\in C(\mathbf{R}^n,\mathbf{R})$ *such that*

$$|W(t,x)|+|W^q(t,x)|\leqq|\bar{W}(x)|$$

for all $x\in\mathbf{R}^n$, $t\in\mathbf{R}$.

Then there exists a (nontrivial) homoclinic orbit q *of* (5.1) *emanating from* 0 *and such that*

$$0<\int_{-\infty}^{\infty}[\tfrac{1}{2}|\dot{q}|^2-V(t,q)]\,dt<\infty.$$

Proof. Let

$$\tilde{E}=\left\{q\in W^{1,2}(\mathbf{R},\mathbf{R}^n)\,\middle|\,\int_{-\infty}^{\infty}[|\dot{q}|^2+L(t)q\cdot q]\,dt<\infty\right\}.$$

Then $\tilde{E}$ is a Hilbert space and as norm in $\tilde{E}$ we can take

$$\|q\|^2=\int_{-\infty}^{\infty}[|\dot{q}|^2+L(t)q\cdot q]\,dt. \tag{5.5}$$

Note that

$$\tilde{E}\subset W^{1,2}(\mathbf{R},\mathbf{R}^n)\subset L^p(\mathbf{R},\mathbf{R}^n) \tag{5.6}$$

for all $p\in[2,\infty]$ with the embedding being continuous. For $q\in\tilde{E}$ let

$$I(q)=\tfrac{1}{2}\|q\|^2-\int_{-\infty}^{\infty}W(t,q)\,dt. \tag{5.7}$$

Then $I \in C^1(\tilde{E}, \mathbf{R})$ and it is routine to verify that any critical point of I on $\tilde{E}$ is a classical solution of (5.1) with $q(\pm\infty)=0=q'(\pm\infty)$. See e.g. [1]. To establish the existence of a critical point of I, a variant of a standard "Mountain Pass" argument will be employed. The usual Mountain Pass Theorem does not apply here since the Palais-Smale condition does not hold due to the fact that we are working with functions on the unbounded set $\mathbf{R}$. However by (5.7) and (5.2), there are constants $\alpha, \rho > 0$ such that

$$I(q) \geqq \alpha \tag{5.8}$$

for $\|q\| = \rho$. Moreover by (V_3) – see e.g. [5] – there is a $q_0 \in \tilde{E}$ such that $\|q_0\| > \rho$ and

$$I(q_0) < 0 = I(0). \tag{5.9}$$

Set

$$\mathscr{K} = \{g \in C([0,1], \tilde{E}) \mid g(0)=0 \text{ and } g(1)=q_0\}$$

and

$$c = \inf_{g \in \mathscr{K}} \max_{s \in [0,1]} I(g(s)). \tag{5.10}$$

By (5.8)–(5.10)

$$c \geqq \alpha. \tag{5.11}$$

Now Ekeland's Variational Principle – see e.g. Theorems 4.10 and 4.3 in [6] – implies there is a sequence $(q_k) \subset \tilde{E}$ such that:

$$I(q_k) \to c \quad \text{and} \quad I'(q_k) \to 0 \tag{5.12}$$

as $k \to \infty$. We will show a subsequence of q_k converges to a critical point, q of I with $q \not\equiv 0$. Hence by earlier remarks q is a solution of (5.1) of the desired type.

As a first step, note that (q_k) is a bounded sequence. Indeed by (V_3) and (5.12), for large k,

$$\begin{aligned} c+1+\|q_k\| &\geqq I(q_k) - \frac{1}{\mu} I'(q_k) q_k \\ &= \left(\frac{1}{2} - \frac{1}{\mu}\right)\|q_k\|^2 - \int_{-\infty}^{\infty} \left[W(t, q_k) - \frac{1}{\mu} W_q(t, q_k) \cdot q_k\right] dt \\ &\geqq \left(\frac{1}{2} - \frac{1}{\mu}\right)\|q_k\|^2 \end{aligned} \tag{5.13}$$

from which the result follows.

Since (q_k) is bounded, it possesses a weakly convergent subsequence in $\tilde{E}$. Let q denote its weak limit. Note that q_k converges to q in $L^\infty_{\text{loc}}(\mathbf{R}, \mathbf{R}^n)$. Hence $I'(q_k) \to 0$ easily implies $I'(q)=0$, i.e. q is a critical point of I. It remains only to prove that $q \not\equiv 0$. Since $q_k \to q$ in L^∞_{loc}, $q_k \to q$ in $L^2([-A, A], \mathbf{R}^n)$ for all $A < \infty$. Hence it suffices to show there is an $A > 0$ such that $q_k \not\to 0$ in $L^2([-A, A], \mathbf{R}^n)$.

(5.14) **Proposition.** *If $q_k \to 0$ in $L^2([-A, A], \mathbf{R}^n)$, there exists an $a>0$ and independent of A such that*

$$\overline{\lim_{k\to\infty}} \|q_k\|^2_{L^2(\mathbf{R},\mathbf{R}^n)} \leqq \frac{a}{\beta(A)} \tag{5.15}$$

where

$$\beta(A) = \inf_{|\xi|=1,\,|t|\geqq A} L(t)\xi\cdot\xi.$$

Proof. Set $a = \sup_k \|q_k\|^2$. Consequently

$$\begin{aligned} \|q_k\|^2_{L^2(\mathbf{R},\mathbf{R}^n)} &= \int_{-A}^{A} |q_k|^2\,dt + \int_{\mathbf{R}\setminus[-A,A]} |q_k|^2\,dt \\ &\leqq \int_{-A}^{A} |q_k|^2\,dt + \frac{1}{\beta(A)} \int_{\mathbf{R}\setminus[-A,A]} L(t)q_k\cdot q_k\,dt \\ &\leqq \int_{-A}^{A} |q_k|^2\,dt + \frac{a}{\beta(A)}. \end{aligned} \tag{5.16}$$

Letting $k\to\infty$, (5.15) is immediate.

To complete the proof of Theorem (5.4), we show that for A sufficiently large, $q_k \not\to 0$ in $L^2([-A, A], \mathbf{R}^n)$. Indeed by (5.12) and the boundedness of (q_k),

$$I(q_k) - \tfrac{1}{2} I'(q_k)q_k = \int_{-\infty}^{\infty} [\tfrac{1}{2} W_q(t, q_k)\cdot q_k - W(t, q_k)]\,dt \to c > 0 \tag{5.17}$$

as $k\to\infty$. By (V_4) and (V_6) for each $M>0$, there exists a $K_M>0$ such that

$$|\tfrac{1}{2} W_q(t, x)\cdot x - W(t, x)| \leqq K_M |x|^2 \tag{5.18}$$

for all $|x|\leqq M$ and $t\in\mathbf{R}$. Taking

$$M = \sup_k \|q_k\|_{L^\infty},$$

by Proposition (5.14),

$$\begin{aligned} c &= \lim_{k\to\infty} \int_{-\infty}^{\infty} [\tfrac{1}{2} W_q(t, q_k)\cdot q_k - W(t, q_k)]\,dt \\ &\leqq K_M \overline{\lim_{k\to\infty}} \|q_k\|^2_{L^2(\mathbf{R},\mathbf{R}^n)} \leqq \frac{K_M a}{\beta(A)}. \end{aligned} \tag{5.19}$$

By (V_5), $\beta(A)\to\infty$ as $A\to\infty$. But then (5.19) contradicts (5.11). The proof is complete.

(5.20) *Remark.* By (5.11) and (5.19), the above argument works whenever there is an A such that

$$\beta(A) > \frac{K_M a}{\alpha}. \tag{5.21}$$

More careful estimates than those given above show that (V_5) can be weakened and (5.21) still obtains if $\beta(A)$ is large enough relative to

$$\underline{L} \equiv \inf_{t \in \mathbf{R}} \sup_{|\xi|=1} L(t)\xi \cdot \xi. \tag{5.22}$$

Indeed (5.13) provides an upper bound for $\|q_k\|$ depending only on μ and c. We assume $\underline{L} \leqq \sup_{|\xi|=1} L(t)\xi \cdot \xi \leqq 2\underline{L}$ on the interval $[-h, h]$. Note that we may assume supp $q_0 \subset [-h, h]$ without loss of generality. Let $g(s) = s q_0 \in \mathscr{K}$. Then

$$\bar{c} = \max_{s \in [0,1]} I(g(s))$$

is an upper bound for c depending only on $\underline{L}$. Since there is a constant $b > 0$ depending only on $\underline{L}$ such that

$$\|w\|_{L^\infty(\mathbf{R}, \mathbf{R}^n)} \leqq b \|w\|$$

for all $w \in \tilde{E}$, the upper bound for $\|q_k\|$ gives an upper bound for $\|q_k\|_{L^\infty}$ and therefore for M. We can also easily see that $\alpha > 0$ depends only on $\underline{L}$. Thus we see the right hand side of (5.21) depends only on $\underline{L}$. Therefore (5.21) holds and Theorem (5.4) remains valid if $\beta(A)$ is large enough relative to $\underline{L}$.

Our final result is a variant of Theorem (5.4):

(5.23) **Theorem.** *Suppose V satisfies (V_1)–(V_4),*

(V_7) *There is a T periodic function L_∞ satisfying (V_2) such that*

(i) $L_\infty(t) \geqq L(t)$ *for all* $t \in \mathbf{R}$ *i.e.* $L_\infty(t)x \cdot x \geqq L(t)x \cdot x$ *for all* $t \in \mathbf{R}$ *and* $x \in S^{n-1}$,

(ii) $L_\infty(\tau)x \cdot x > L(\tau)x \cdot x$ *for some* $\tau \in \mathbf{R}$ *and all* $x \in S^{n-1}$,

(iii) $|L(t) - L_\infty(t)| \to 0$ *as* $|t| \to \infty$.

(V_8) *$W(t, x)$ is T-periodic in t,*

(V_9) *the map $s \to s^{-1} W_q(t, sx) \cdot x$ is a strictly increasing function of $s \in (0, \infty)$ for all $t \in \mathbf{R}$ and $x \in \mathbf{R}^n \setminus \{0\}$.*

Then there exists a nontrivial homoclinic orbit of (5.1) *emanating from* 0.

(5.24) *Remark.* Condition (V_9) is related to hypotheses that have been used by Nehari [7, 8], Coffman [9], Hempel [10], and others. We suspect that Theorem (5.23) is true without this hypothesis.

Proof of Theorem (5.23) The proof follows the same lines as that of Theorem (5.4). Now we work in $E_1 \equiv W^{1,2}(\mathbf{R}, \mathbf{R}^n)$. Note that by (V_2) and (V_7),

$$\left(\int_{-\infty}^{\infty} (|\dot{q}|^2 + L(t)q \cdot q)\, dt \right)^{1/2}$$

and

$$\left(\int_{-\infty}^{\infty} (|\dot{q}|^2 + L_\infty(t)q \cdot q)\, dt \right)^{1/2}$$

are equivalent norms in E_1. Hence I as defined in (5.7) belongs to $C^1(E_1, \mathbf{R})$. Moreover (5.8)–(5.9) are still valid in E_1. Defining c by (5.10), with $\tilde{E}$ replaced by E_1 in $\mathscr{K}$, (5.11)–(5.13) still hold and $q_k \to q$ weakly in E_1 (along a subsequence) where $I'(q)=0$. It remains only to prove that $q \not\equiv 0$. This will take some work.

The idea of the proof is to consider the functional

$$I^\infty(q)=\int_{-\infty}^{\infty} [\tfrac{1}{2}|\dot{q}|^2+\tfrac{1}{2}L_\infty(t)q\cdot q-W(t,q)]\,dt. \tag{5.25}$$

By (V_7), (V_3), and (V_4), $I^\infty \in C^1(E_1, \mathbf{R})$ and by earlier remarks, critical points of I^∞ are homoclinic solutions of

$$\ddot{q}-L_\infty(t)q+W_q(t,q)=0 \tag{5.26}$$

which emanate from 0. We will show: (a) I^∞ also has a "mountain pass" critical value c^∞. Moreover by (V_9), c_∞ can also be characterized as: (b)

$$c^\infty=\inf\{I^\infty(y)\,|\,y\in E_1\setminus\{0\} \text{ and } (I^\infty)'(y)=0\}.$$

This enables us to prove: (c) $c<c^\infty$. Then: (d) an argument related to the one following Proposition (5.14) implies $q\not\equiv 0$.

The steps (a)–(d) will now be carried out in detail.

Step (a) There exists a nontrivial homoclinic solution of (5.26) emanating from 0.

Proof. The existence of such a solution was established in [5]. However we will give another proof here since we need a minimax characterization of its corresponding critical value. Note that the constants $\rho, \alpha \geqq 0$ and $q_0 \in E_1$ can be chosen so that the estimates (5.8)–(5.9) hold for both I and I^∞. Setting

$$c^\infty=\inf_{g\in\mathscr{K}}\max_{s\in[0,1]} I^\infty(g(s)). \tag{5.27}$$

as in the proof of Theorem (5.4), there is a bounded sequence $(w_k)\subset E_1$ such that

$$I^\infty(w_k)\to c^\infty, \quad (I^\infty)'(w_k)\to 0. \tag{5.28}$$

Moreover a subsequence of w_k converges weakly in E_1 and also in L^∞_{loc} to $w\in E_1$ such that $(I^\infty)'(w)=0$. Hence once we show $w\not\equiv 0$, as earlier it is a nontrivial homoclinic solution of (5.26) emanating from 0. Note that since L_∞ and W are T-periodic in t,

$$I^\infty(y(t))=I^\infty(y(t+lT)) \tag{5.29}$$

for all $y\in E_1$ and $l\in\mathbf{Z}$. The functions $w_k\in E_1$. Hence $w_k(t)\to 0$ as $|t|\to\infty$ – see e.g. (20)–(24) in [5]. Consequently by (5.29), without loss of generality it can be assume that w_k achieves its maximum in $[0, T]$. Thus if $w_k\to 0$ weakly in E_1 and in L^∞_{loc} as $k\to\infty$, in fact

$$w_k\to 0 \quad \text{in } L^\infty(\mathbf{R}, \mathbf{R}^n) \tag{5.30}$$

as $k \to \infty$. By (5.2),

$$|\tfrac{1}{2} W_q(t, x) \cdot x - W(t, x)| = o(1)|x|^2 \tag{5.31}$$

as $|x| \to 0$ uniformly for $t \in \mathbf{R}$. Hence by (5.30)–(5.31),

$$\int_{-\infty}^{\infty} [\tfrac{1}{2} W_q(t, w_k) \cdot w_k - W(t, w_k)]\, dt = o(1) \|w_k\|^2_{L^2(\mathbf{R}, \mathbf{R}^n)} \to 0 \tag{5.32}$$

as $k \to \infty$. On the other hand by (5.28),

$$\int_{-\infty}^{\infty} [\tfrac{1}{2} W_q(t, w_k) \cdot w_k - W(t, w_k)]\, dt = I^\infty(w_k) - \tfrac{1}{2}(I^\infty)'(w_k) w_k \to c^\infty \geqq \alpha \tag{5.33}$$

as $k \to \infty$, contrary to (5.32). Hence Step (a) is complete.

(5.34) *Remark.* (5.28), (5.33), Fatou's Lemma and $(I^\infty)'(w) = 0$ imply (along our subsequence)

$$c^\infty = \lim_{k \to \infty} \int_{-\infty}^{\infty} [\tfrac{1}{2} W_q(t, w_k) \cdot w_k - W(t, w_k)]\, dt$$

$$\geqq \int_{-\infty}^{\infty} [W_q(t, w) \cdot w - W(t, w)]\, dt = I^\infty(w).$$

Step (b) $c^\infty = \inf\{I^\infty(y) \mid y \in E_1 \setminus \{0\}, (I^\infty)' y) = 0\}$.

Proof. Let $b = \inf\{I^\infty(y) \mid y \in E_1 \setminus \{0\}, (I^\infty)'(y) = 0\}$. By Remark (5.34),

$$c^\infty \geqq b. \tag{5.35}$$

We claim for all $u \in E_1 \setminus \{0\}$ such that $(I^\infty)'(u) \neq 0$, there is a $g \in \mathscr{K}$ such that

$$\max_{s \in [0,1]} I^\infty(g(s)) = I^\infty(u). \tag{5.36}$$

Then by (5.27),

$$c^\infty \leqq b \tag{5.37}$$

so (5.35) and (5.37) yield Step (b).

To construct g, note that by (V_9), if $y \in E_1 \setminus \{0\}$, $I^\infty(sy)$ achieves its maximum along the ray $\{sy \mid s \geqq 0\}$ at a unique point $s_0 y$ characterized by $(I^\infty)'(s_0 y) y = 0$.

(See e.g. [9, 10].) Hence let g consist of a segment of the ray through 0 and u, a circular arc through $Ru/\|u\|$ and $Rq_0/\|q_0\|$, and q_0. Explicitly

$$g_R(s)=\begin{cases} 3sR\dfrac{u}{\|u\|} & s\in[0,\tfrac{1}{3}], \\ \dfrac{R}{\|u\|}\cos\dfrac{3\pi}{2}(s-\tfrac{1}{3})+\dfrac{Rq_0}{\|q_0\|}\sin\dfrac{3\pi}{2}(s-\tfrac{1}{3}) & s\in[\tfrac{1}{3},\tfrac{2}{3}], \\ 3(1-s)\dfrac{Rq_0}{\|q_0\|}+3(s-\tfrac{2}{3})q_0 & s\in[\tfrac{2}{3},1]. \end{cases} \tag{5.38}$$

Then $g_R\in\mathscr{K}$ for all $R>0$ and for R large, $I^\infty(g_R(s))<0$ for all $s\in[\frac{1}{3},1]$. Hence (5.36) holds and step (b) is verified.

Step (c) $c<c^\infty$.

Proof. Let g be the path constructed in Step (b) with $u=w$, w being as determined in Step (a). By (V_7)(i), for all $s\in[0,1]$,

$$I(g(s))\leqq I^\infty(g(s)). \tag{5.39}$$

Hence

$$c\leqq\max_{s\in[0,1]} I(g(s))\leqq I^\infty(w)=c^\infty. \tag{5.40}$$

If $c=c^\infty$, there is an $\bar{s}\in(0,1)$ such that

$$c=I(g(\bar{s}))=I^\infty(g(\bar{s}))=c^\infty. \tag{5.41}$$

But there is a unique $s\in(0,1)$ such that $I^\infty(g(s))=c^\infty$, namely $s=\|w\|/3R$ and $g(s)=w$. Hence to prove that $c<c^\infty$, it suffices to show that

$$I(w)<I^\infty(w). \tag{5.42}$$

If $I(w)=I^\infty(w)$, by (V_7)(ii), $w(t)\equiv 0$ in a neighborhood of $t=\tau$. Observe that w is a solution of the linear system of equations

$$\ddot{q}_i+\sum_{j=1}^{n} a_{ij}(t)q_j=0,\qquad 1\leqq i\leqq n \tag{5.43}$$

where $a_{ij}(t)=0$ if $w(t)=0$ and if $w(t)\neq 0$,

$$a_{ij}(t)=\frac{1}{|w(t)|}\frac{\partial V}{\partial q_i}(t,w(t))\,w_j(t).$$

The coefficients a_{ij} are continuous via (5.2). Since $w(t)\equiv 0$ near $t=\tau$ and satisfies a linear system with continuous coefficients, $w(t)\equiv 0$, a contradication. Thus (5.42) follows and Step (c) is proved.

Step (d) Completion of the proof of Theorem (5.23).

Recall that (q_k) is the "mountain pass" sequence for this problem satisfying (5.12) and q is the limit (along a subsequence) of q_k. We must show $q \not\equiv 0$. Due to the L^∞_{loc} convergence of q_k to q, it suffices to prove for some $A>0$,

$$\int_{-A}^{A} |q_k|^2\,dt \not\to 0 \tag{5.44}$$

along a subsequence. We will use a "concentration compactness" type argument. Suppose

$$\int_{-A}^{A} |q_k|^2\,dt \to 0 \tag{5.45}$$

as $k\to\infty$ for all $A>0$. Consider

$$\beta_k = \sup_{l\in\mathbf{Z}} \int_0^T |q_k(t+lT)|^2\,dt. \tag{5.46}$$

We claim

$$\lim_{k\to\infty} \beta_k \not= 0. \tag{5.47}$$

Indeed for any $u\in E_1$, s, $t\in[0, T]$, and $l\in\mathbf{Z}$, we have

$$\begin{aligned} |u(t+lT)|^2 &= |u(s+lT)|^2 + \int_s^t \frac{d}{dr}|u(r+lT)|^2\,dr \\ &\leqq |u(s+lT)|^2 + 2\int_0^T |u(r+lT)\cdot \dot{u}(r+lT)|\,dr. \end{aligned} \tag{5.48}$$

Integrating (5.48) for s over $[0, T]$ and setting $u=q_k$ yields:

$$\begin{aligned} |q_k(t+lT)|^2 &\leqq \frac{1}{T}\int_0^T |q_k(s+lT)|^2\,ds + \frac{2}{T}\left(\int_0^T |q_k(s+lT)|^2\,ds\right)^{1/2} \|\dot{q}_k\|_{L^2(\mathbf{R},\mathbf{R}^n)} \\ &\leqq \frac{\beta_k}{T} + \frac{2}{T}\beta_k^{1/2} M \end{aligned} \tag{5.49}$$

where M is an upper bound for $\|q_k\|$. Since l is arbitrary, if $\beta_k\to 0$, (5.49) shows

$$\lim_{k\to\infty} \|q_k\|_{L^\infty(\mathbf{R},\mathbf{R}^n)} = 0. \tag{5.50}$$

But then the argument of (5.30)–(5.33) implies $c=0$, a contradiction. Hence (5.47) holds.

Comparing (5.47) to (5.45), we see there must exist a sequence $l_k\in\mathbf{Z}$ such that

(5.51) (i) $|l_k|\to\infty$ as $k\to\infty$ and

(ii) $\int_0^T |q_k(t+l_kT)|^2\,dt \geqq \gamma > 0.$

Set $\tilde{q}_k(t)=q_k(t+l_k T)$. By (V_7)(iii) and a familiar argument, there is a subsequence of $\tilde{q}_k$ which converges weakly in E_1 and also in L^∞_{loc} to $\tilde{q}\in E_1$, a homoclinic solution of (5.26). Moreover by (5.51)(ii), $\tilde{q}$ is a nontrivial solution of (5.26). As in Remark (5.34), using (V_8),

$$c=\lim_{k\to\infty}\int_{-\infty}^{\infty}[\tfrac{1}{2}W_q(t,q_k)\cdot q_k-W(t,q_k)]\,dt \tag{5.52}$$

$$=\lim_{k\to\infty}\int_{-\infty}^{\infty}[\tfrac{1}{2}W_q(t\cdot\tilde{q}_k)\cdot\tilde{q}_k-W(t,\tilde{q}_k)]\,dt$$

$$\geqq\int_{-\infty}^{\infty}[\tfrac{1}{2}W_q(t,\tilde{q})\cdot\tilde{q}-W(t,\tilde{q})]\,dt=I^\infty(\tilde{q}).$$

By Step (b),

$$I^\infty(\tilde{q})\geqq c^\infty. \tag{5.53}$$

But (5.52)–(5.53) imply $c\geqq c^\infty$, contrary to Step (c). Hence (5.44) holds and $q\not\equiv 0$. The proof of Theorem (5.23) is complete.

References

1. Rabinowitz, P.H.: Periodic and heteroclinic orbits for a periodic Hamiltonian system. Analyse Nonlineaire **6**, 331–346 (1989)
2. Kozlov, V.V.: Calculus of variations in the large and classical mechanics. Russ. Math. Surv. **40**, 37–71 (1985)
3. Coti-Zelati, V., Ekeland, I., Sere, E.: A variational approach to homoclinic orbits in Hamiltonian systems. Math. Ann. **288**, 133–160 (1990)
4. Hofer, H., Wysocki, K.: First order elliptic systems and the existence of homoclinic orbits in Hamiltonian systems. (preprint)
5. Rabinowitz, P.H.: Homoclinic orbits for a class of Hamiltonian systems. Proc. Royal Soc. Edinburgh **114A**, 33–38 (1990)
6. Mawhin, J., Willem, M.: Critical Point Theory and Hamiltonian systems. Berlin Heidelberg New York: Springer 1989
7. Nehari, Z.: On a class of nonlinear second order differential equations. Trans. Am. Math. Soc. **95**, 101–123 (1960)
8. Nehari, Z.: Characteristic values associated with a class of nonlinear second-order differential equations. Acta Math. **105**, 141–175 (1961)
9. Coffman, C.: A minimum-maximum principle for a class of nonlinear integral equations. J. Anal. Math. **22**, 391–419 (1969)
10. Hempel, J.A.: Superlinear boundary value problems and nonuniqueness. University of New England, Armidale thesis 1970

Math. Z. 206, 501–504 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Erratum:

A precise result on the boundary regularity of biholomorphic mappings

László Lempert
Purdue University, West Lafayette, IN 47907, USA

Math. Z. **193**, 559–579 (1986)

1. Mr. Ralf Kampmann has called my attention to two errors in the paper in question, [3]. Although in the meantime another paper has appeared which proves essentially the same regularity result for biholomorphic mappings as does [3] (see [1]), I would like to correct the mistakes here. One of the reasons is that [3] discusses regularity properties of the Kobayashi metric, too, and precise regularity of the Kobayashi metric may be of interest in other types of questions as well (cf. [4, 5]).

The errors are in the proof of Theorem 4.1 on page 566, line-5 in [3]. The first is a simple misprint: Φ should read Θ. The second is that this mapping Θ is not Hölder continuous of order C^{k-s}. Therefore the implicit function theorem cannot be applied and does not yield the regularity of Φ claimed in Theorem 4.1. However, Θ is C^{k-s} when restricted to a smaller subset of its domain of definition in [3], and, as we are going to show here, this is sufficient to correct the proof of Theorem 4.1.

2. First recall Whitney's notion of Hölder continuity of functions defined on not necessarily open sets. Let A, E be Banach spaces, $\mathscr{A} \subset A$ a subset. A mapping $\phi: \mathscr{A} \to E$ is C^l $(0<l<1)$ if $\|\phi(a_1)-\phi(a_2)\|_E \leqq \text{const}\,\|a_1-a_2\|_A^l$ for $a_1, a_2 \in \mathscr{A}$. ϕ is called differentiable if for every $a \in \mathscr{A}$ there is a continuous linear mapping $L_a \in L(A, E)$ such that

$$\|\phi(a_1)-\phi(a)-L_a(a_1-a)\|_E = o(\|a_1-a\|_A) \quad \text{as} \quad \mathscr{A} \ni a_1 \to a.$$

Finally, if $l>1$ is non integral then ϕ is (inductively) defined to be C^l if it is differentiable and L_a can be chosen so that the mapping $\mathscr{A} \ni a \mapsto L_a \in L(A, E)$ is C^{l-1}. We shall denote L_a by $\partial\phi/\partial a$. Although $\partial\phi/\partial a$ is not uniquely determined, when ϕ is C^l, $\partial\phi/\partial a$ will always denote a C^{l-1} Hölder continuous choice of the partial derivative.

Let now $R>0$. With notation as in [3], let $\Xi_m(R)=\Xi_m\subset X_m$ denote the set of those mappings $f\colon \partial U\to\mathbb{C}^m$ whose components are in C^1 with C^1-norm less than R.

Proposition 1 *Suppose $K\colon\mathbb{C}^m\to\mathbb{C}$ is a function of class C^l, $l>0$, $\{l\}>s$. Then the mapping $\tilde{K}\colon\Xi_m\to X_1$ defined by $\tilde{K}(g)=K\circ g$ is of class C^{l-s}.*

Proof. Assume first $0<l<1$, and let $g, h\in\Xi_m$. For $x\in\partial U$

$$|K(g(x))-K(h(x))|\leqq\|K\|_{C^l}|g(x)-h(x)|^l \leqq\|K\|_{C^l}\|g-h\|^{l-s}_{X_m}; \tag{1}$$

and for

$$\Delta=(K(g(x))-K(h(x)))-(K(g(y)-K(h(y)))$$

with $x, y\in\partial U$ we have the estimates

$$\begin{aligned}|\Delta|&\leqq\|K\|_{C^l}(|g(x)-h(x)|^l+|g(y)-h(y)|^l)\\ &\leqq 2\|K\|_{C^l}\|g-h\|^l_{X_m},\\ |\Delta|&\leqq 2\|K\|_{C^l}R^l|x-y|^l.\end{aligned}$$

The two latter inequalities imply

$$|\Delta|\leqq\text{const}\,\|g-h\|^{l-s}_{X_m}|x-y|^s.$$

This, together with (1) prove that $\tilde{K}$ is C^l.

Next, if $1<l<2$, write

$$K(g(x)+h(x))-K(g(x))=K'(g(x))\,h(x)+R(x),$$

$$R(x)=\int_0^1(K'(g(x)+t\,h(x))-K'(g(x)))\,h(x)\,dt,$$

where, this time, $g, g+h\in\Xi_m$. From what has already been proved it follows that

$$\|R\|_{X_1}=o(\|h\|_{X_m})\quad\text{as } h\to 0.$$

Hence $\tilde{K}$ is differentiable and its derivative in g can be given by the map

$$X_m\ni\gamma\mapsto(K'\circ g)\,\gamma\in X_1.$$

Thus, by the first step of the proof, $\tilde{K}\in C^{l-s}$.

The proof for general l is done by an obvious induction.

Next we check a simple version of the implicit function theorem. Let A, B, and E be Banach spaces, $\mathscr{A}\in A$ a convex subset.

Proposition 2 *Let $\theta: \mathscr{A} \times B \to E$ be a C^l mapping ($l \geqq 1$). Assume that for a certain point $(a_0, b_0) \in \mathscr{A} \times B$ $\theta(a_0, b_0) = 0$ and that for an appropriate choice of the partial derivative $\partial\theta/\partial a$*

$$\frac{\partial \theta}{\partial a}(a_0, b_0): A \to E$$

is an isomorphism.

If now φ is a continuous mapping of some neighborhood G of $b_0 \in B$ into $\mathscr{A}$ such that $\varphi(b_0) = a_0$ and $\theta(\varphi(b), b) \equiv 0$, then φ is of class C^l near b_0.

Proof. For $b, b_1 \in G$ write

$$0 = \theta(\varphi(b_1), b_1) - \theta(\varphi(b), b) \tag{2}$$
$$= \frac{\partial \theta}{\partial a}(\varphi(b), b)(\varphi(b_1) - \varphi(b)) + \frac{\partial \theta}{\partial b}(\varphi(b), b)(b_1 - b) + R,$$

$$\|R\|_E = o(1)(\|\varphi(b_1) - \varphi(b)\|_A + \|b_1 - b\|_B) \quad (b_1 \to b). \tag{3}$$

When b is near b_0, $\partial\theta/\partial a(\varphi(b), b): A \to E$ is an isomorphism, and, indeed, the norm of the inverse mapping

$$\left(\frac{\partial \theta}{\partial a}(\varphi(b), b)\right)^{-1}: E \to A \tag{4}$$

is bounded. (2) and (3) now imply

$$\|\varphi(b_1) - \varphi(b)\|_A = O(\|b_1 - b\|_B) \quad (b_1 \to b),$$

and so by (3)

$$\|R\|_E = o(\|b_1 - b\|_B).$$

Applying the mapping (4) to (2), we see that φ is differentiable and

$$\frac{\partial \varphi(b)}{\partial b} = -\left(\frac{\partial \theta}{\partial a}(\varphi(b), b)\right)^{-1} \frac{\partial \theta}{\partial b}(\varphi(b), b).$$

Using this formula a simple induction proves that φ is l-Hölder continuous near b_0.

3. There is one more fact that we shall need. Let D be as in Lemma 3.1 of [3], and $f: U \to D$ an extremal map. By the above lemma, $f \in C^k(\bar{U})$.

Proposition 3 *If $f(0)$ belongs to a compact subset $\Sigma \subset D$ then $\|f\|_{C^k} \leqq c(\Sigma, D)$, where the constant $c(\Sigma, D)$ does not depend on f.*

Proof. That the $C^{1/2}$ norm of f is uniformly bounded follows from the a priori estimates of [2] when $k > 6$ and the approximation argument explained on bottom of page 562 and top of the next page in [3], when ∂D is just C^2. Having uniform $C^{1/2}$ estimates at hand, the proofs that $f \in C^k(\bar{U})$ also yield uniform C^k estimates, as is easily checked.

Corollary. *The mapping Φ in Theorem* 4.1 *in* [3] *is continuous.*

Proof. Suppose $w_j, w \in D$, $v_j, v \in \mathbb{C}^n \setminus \{0\}$, $w_j \to w$, $v_j \to v$, and the extremal mappings $f_j = \Phi(w_j, v_j)$ do not converge to $f = \Phi(w, v)$ in Y_n.

Since the C^k-norms of f_j are bounded, by passing to a subsequence, we can assume $f_j \to g \neq f$ in the space C^s. But the limit of extremal mappings being extremal, $g(0) = w = f(0)$, and $g'(0) = \lambda_1 v$, $f'(0) = \lambda_2 v$, we obtain $g = f$, a contradiction.

After these preparations, in the proof of Theorem 4.1, the following modifications have to be made. If we want to prove that Φ is C^{k-s} in a neighborhood of $(W, V) \in D \times \mathbb{C}^n \setminus \{0\}$, choose $R > 0$ so that for any extremal mapping $f: U \to D$ with $f(0)$ near W the C^1 norm of f is less than R. Next, the mapping Θ should be defined only on a neighborhood of (F, W, Λ, V) in

$$(Y_n \cap \Xi_m(R)) \oplus \mathbb{C}^n \oplus \mathbb{R} \oplus \mathbb{C}^n,$$

instead of $Y_n \oplus \mathbb{C}^n \oplus \mathbb{R} \oplus \mathbb{C}^n$. It follows from Proposition 1 that Θ is C^{k-s}.

We want to apply Proposition 2. Let $A = Y_n \oplus \mathbb{R}$, $B = \mathbb{C}^n \oplus \mathbb{C}^n$, $E = T \oplus Y'_{n-1} \oplus \mathbb{C}^n \oplus \mathbb{C}^n$, $\mathscr{A}$ a neighborhood of $a_0 = (F, \Lambda)$ in $Y_n \cap \Xi_n(R)$, $b_0 = (W, V)$, $\theta = \Theta$ and finally let φ defined by

$$\varphi(w, v) = (f, \lambda),$$

where $f = \Phi(w, v)$ and λ is such that $\lambda v = f'(0)$.

By the corollary above φ is continuous. The computations on pp 567–568 in [3] show that $\partial\theta/\partial a(a_0, b_0) = L$ is an isomorphism between A and E. Hence Proposition 2 implies that φ, hence also Φ, is C^{k-s}, as claimed.

References

1. Hasanov, S.V., Pinchuk, S.I.: Asymptotically holomorphic functions. Mat. Sb. **134(176)**, 546–555 (1987)
2. Lempert, L.: La métrique de Kobayashi et la représentation des domaines sur la boule. Bull. Soc. Math. Fr. **109**, 427–474 (1981)
3. Lempert, L.: A precise result on the boundary regularity of biholomorphic mappings. Math. Z. **193**, 559–579 (1986)
4. Lempert, L.: Complex geometry in convex domains. Proceedings of the International Congress of Mathematicians. Berkeley, California: 1986
5. Lempert, L.: Holomorphic invariants, normal forms, and the moduli space of convex domains. Ann. Math., II. Ser. **128**, 43–78 (1988)

Math. Z. 206, 505–526 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Finitistic dimension of artinian rings with vanishing radical cube

Edward L. Green [1] **and Birge Zimmermann-Huisgen** [2] ⋆

[1] Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA
[2] Department of Mathematics, University of California at Santa Barbara, Santa Barbara, CA 93106, USA

Received May 26, 1989; in final form December 28, 1989

1 Introduction and notation

The following conjecture, due to Rosenberg and Zelinsky (see [2]), dates back to the late 1950's: Given a finite dimensional algebra Λ over a field, the supremum of the projective dimensions of those finitely generated left Λ-modules which have finite projective dimension is finite. We denote this supremum by $\operatorname{fin\,dim}(\Lambda)$ and call it the (left) *finitistic dimension* of Λ. Although we will not explicitly stress the side, this invariant is not left-right symmetric.

While the finitistic dimension of commutative noetherian rings is well understood, (see [1, 3, 10]), work on the non-commutative artinian case is still in a pioneering phase. For more details on the history of the problem up to 1960 we refer the reader to Bass's paper [2], and for a summary of the later development we point to a recent article of Green, Kirkman and Kuzmanovich [6] where the conjecture is shown to be true for monomial algebras, i.e., for finite dimensional path algebras modulo ideals generated by paths.

In this paper we prove the conjecture under the restriction that the cube of the Jacobson radical J of Λ is zero (Theor. 16). In fact, our method of proof confirms the conjecture for a considerably larger class of rings including all those left artinian rings which have only simple composition factors of finite projective dimension in the left Λ-module J^3 (see Theor. 21); a trick due to Small, moreover, allows us to replace "$J^3=0$" by a finiteness condition on J^3 as a *right* module (Coroll. 17). A major portion of our argument is constructive and, in many cases, yields an explicit algorithm for computing the finitistic dimension. In all cases, it provides the following rough upper bound:

$$\operatorname{fin\,dim}(\Lambda)\leqq 2m+d+1,$$

where m is the number of isomorphism types of simple left Λ-modules of *infinite* projective dimension and d is the supremum of the projective dimensions of

⋆ Both authors were partially supported by grants from the N.S.F.

the simple left Λ-modules having *finite* projective dimension. For sharper bounds, see Theorems 2, 16, 21.

Our method appears to be of interest beyond the problem at hand. Based on composition series for the indecomposable projective left Λ-modules, we construct a certain group $G(\Lambda)$ of integral matrices in $Z^{m\times 2}$, together with a nilpotent endomorphism L on $G(\Lambda)$; in addition we assign a matrix $C=C(M)\in G(\Lambda)$ to any Λ-module M of finite projective dimension in such a fashion that the equality $L^c(C)=0$ places an upper bound of $c+d+1$ on the projective dimension of M.

The amount of homological information on Λ stored in the matrix group $G(\Lambda)$ and its operator L will further manifest itself in a forthcoming paper [13], where the present technique will be elaborated to yield accessible upper bounds on the number d introduced above and thus on fin dim(Λ). For $J^3=0$ this method provides explicit bounds in terms of the number of isomorphism types of simple left Λ-modules, valid also in case the left global dimension of Λ is finite.

As for the algorithmic side of our technique: once the simple modules of finite projective dimension have been identified and the number d has been computed or approximated, the procedure we develop for finding tight upper bounds on the finitistic dimension can be carried out by a computer; the method simultaneously singles out composition series of modules which lend themselves to testing sharpness of these bounds. The examples worked out in Sect. 4 illustrate how easy it is to apply this algorithm.

We will conclude by applying our results on dimensions to extend the class of rings for which the Nakayama Conjecture is known to hold.

Throughout the first four sections, Λ will be a basic left artinian ring with $J^3=0$, where J denotes the Jacobson radical of Λ. Note that the assumption that Λ is basic does not affect the generality of our treatment. The projective dimension of a Λ-module will be abbreviated by p dim M, or by p dim$_\Lambda M$ if we wish to emphasize the ring. Moreover, the first syzygy of M, the kernel of a projective cover of M, will be denoted by $\Omega_\Lambda(M)=\Omega^1_\Lambda(M)$; iterating Ω, we obtain the higher syzygies of M, $\Omega^k(M)=\Omega(\Omega^{k-1}(M))$. By Λ-mod, we will mean the category of all finitely generated left Λ-modules, and $L_2(\Lambda)$ will stand for the full subcategory of Λ-mod having as objects the modules of Loewy length at most two.

2 A matrix trick – simplest case

For the moment, suppose that all simple summands of the left Λ-module J^2 have infinite projective dimension. Letting $S_1, \ldots, S_n$ denote the representatives of the simple left Λ-modules with corresponding primitive idempotents $e_1, \ldots, e_n$, we define a map

$$L_2(\Lambda)\to \mathbf{Z}^{n\times 2} \quad \text{by} \quad M\mapsto [M]=\begin{pmatrix} a_1 & b_1 \\ \vdots & \vdots \\ a_n & b_n \end{pmatrix}$$

where a_i is the multiplicity of S_i in M/JM and b_i is the multiplicity of S_i in JM. Moreover, setting

$$[Je_i]=(\mathbf{u}_i,\mathbf{v}_i)=\begin{pmatrix}u_{i1} & v_{i1}\\ \vdots & \vdots\\ u_{in} & v_{in}\end{pmatrix},$$

we consider the $\mathbf{Z}$-linear map $L^*\colon \mathbf{Z}^{n\times 2}\to Z^{n\times 2}$ which sends

$$(\mathbf{a},\mathbf{b})=\begin{pmatrix}a_1 & b_1\\ \vdots & \vdots\\ a_n & b_n\end{pmatrix}$$

to the matrix $\left(\sum_{i=1}^{n} a_i\mathbf{u}_i-\mathbf{b},\ \sum_{i=1}^{n} a_i\mathbf{v}_i\right)$.

The following easy observation will turn out to be pivotal in the rest of the paper.

Observation 1 *Assume that all simple summands of J^2 have infinite projective dimension, and let $M\in L_2(\Lambda)$ be of finite projective dimension. Then $[\Omega^k_\Lambda(M)]=(L^*)^k[M]$. In particular,* p dim $M=a$ *if and only if $a+1$ is the least positive integer with $(L^*)^{a+1}[M]=0$.*

Proof. Let

$$[\mathbf{M}]=(\mathbf{a},\mathbf{b})=\begin{pmatrix}a_1 & b_1\\ \vdots & \vdots\\ a_n & b_n\end{pmatrix},$$

and let $0\to\Omega_\Lambda(M)=\ker(\alpha)\to\coprod_{i=1}^{n}(\Lambda e_i)^{a_i}\xrightarrow{\alpha}M\to 0$ be a projective cover of M. Since every simple summand of $\coprod_{i=1}^{n}(J^2e_i)^{a_i}$ has infinite projective dimension by our assumption and since p dim $\ker(\alpha)<\infty$, we clearly have $\coprod_{i=1}^{n}(J^2e_i)^{a_i}\subseteq J\ker(\alpha)$. Indeed, note that every simple submodule of $\ker(\alpha)$ not contained in $J\ker(\alpha)$ is a direct summand of $\ker(\alpha)$. On the other hand, $\ker(\alpha)\subseteq\coprod_{i=1}^{n}(Je_i)^{a_i}$, and therefore $J\ker(\alpha)\subseteq\coprod_{i=1}^{n}(J^2e_i)^{a_i}$. Thus, a straightforward count of simple composition factors yields $[\Omega_\Lambda(M)]=L^*[M]$. The observation follows by induction. □

Given $A\in\mathbf{Z}^{n\times 2}$, let $\tau(A)$ be the least non-negative integer a for which $(L^*)^a(A)=0$ if such an integer exists, and set $\tau(A)=\infty$ otherwise. Consider the subgroup $G_0^*(\Lambda)$ of $\mathbf{Z}^{n\times 2}$ consisting of all matrices A with $\tau(A)<\infty$. Moreover, suppose that $S_1,\ldots,S_p$ are the distinct isomorphism types of simple modules occurring as submodules of (J/J^2) viewed as a left Λ-module, and that $S_1,\ldots,S_q$

are the types of those simple Λ-modules occurring as submodules of *both* J/J^2 and J^2. By definition, we have that $q \leqq p$. We now define $G^*(\Lambda)$ as follows:

$$G^*(\Lambda) = G_0^*(\Lambda) \cap \begin{pmatrix} \mathbf{Z} & \mathbf{Z} \\ \vdots & \vdots \\ \cdot & \mathbf{Z} \\ \mathbf{Z} & 0 \\ \vdots & \vdots \\ \mathbf{Z} & \cdot \\ 0 & \cdot \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}$$

where the first column of the last matrix has p entries of $\mathbf{Z}$ and the second column has q entries of $\mathbf{Z}$. Next we note that, being subgroups of $\mathbf{Z}^{n \times 2}$, both of the groups $G_0^*(\Lambda)$ and $G^*(\Lambda)$ are free abelian. The main result of this section now follows via a straightforward argument from Observation 1.

Theorem 2 *Suppose that $J^3 = 0$, and keep the notation developed above. If all simple summands of J^2 have infinite projective dimension, then*

$$\text{fin dim}(\Lambda) \leqq \text{rank } G^*(\Lambda) \leqq p + q \leqq 2n.$$

Before we prove Theorem 2, we wish to point out the following:

(a) If Λ is a finite dimensional algebra over a field K, then $p + q \leqq \dim_K J$, and inequality holds if either J/J^2 or J^2 has repeated simple summands.

(b) The following upper bound for fin dim(Λ) is often better than rank $G^*(\Lambda)$:

$$c = \sup_{1 \leqq j \leqq k} \tau(A_j)$$

where $A_1, \ldots, A_k$ is any generating set for $G^*(\Lambda)$. The justification of this remark will be contained in the proof of Theorem 2 below.

Proof of Theorem 2 We start by noting that, given any $M \in \Lambda$-mod of finite projective dimension with $M/JM = S_1^{c_1} \coprod \ldots \coprod S_n^{c_n}$, the top $\Omega_\Lambda(M)/J\Omega_\Lambda(M)$ of the first syzygy is a homomorphic image of $(Je_1/J^2e_1)^{c_1} \coprod \ldots \coprod (Je_n/J^2e_n)^{c_n}$. Indeed, remember that none of the simple summands of $\Omega_\Lambda(M) \cap [(J^2e_1)^{c_1} \coprod \ldots \coprod (J^2e_n)^{c_n}]$ can belong to the top of $\Omega_\Lambda(M)$, since p dim$(\Omega_\Lambda(M)) < \infty$. Thus the matrix $[\Omega_\Lambda(M)]$ has zeros in positions $(p+1, 1), \ldots, (n, 1)$. Next, $J\Omega_\Lambda(M)$ is a submodule of a direct sum of copies of J^2 on one hand and a homomorphic image of a direct sum of copies of J/J^2 on the other; thus $[\Omega_\Lambda(M)]$ has zeros also in positions $(q+1, 2), \ldots, (n, 2)$. By Observation 1 we infer that $[\Omega_\Lambda(M)] \in G^*(\Lambda)$.

Next, consider the subgroup $G^{**}(\Lambda)$ of $G^*(\Lambda)$ generated by all the matrices $[\Omega_\Lambda(M)]$ where M runs through the objects of Λ-mod having finite projective dimension. Then $G^{**}(\Lambda)$ is in turn free abelian of some rank c which is smaller than or equal to the rank of $G^*(\Lambda)$. Since $L^*[\Omega_\Lambda(M)] = [\Omega_\Lambda^2(M)]$ by Observation 1, the map L^* restricts to an endomorphism L^{**} of $G^{**}(\Lambda)$. Let $\mathbf{M} \in \mathbf{Z}^{c \times c}$ be

the matrix of L^{**} with respect to any basis of $G^{**}(\Lambda)$. Clearly L^{**}, and hence $\mathbf{M}$, is nilpotent, because for each $A \in G^*(\Lambda)$ some power of L^* vanishes on A. Thus $\mathbf{M}^c = 0$ by elementary linear algebra, and, consulting Observation 1 once again, we infer that $\mathrm{p\,dim}(\Omega_\Lambda(M)) \leqq c-1$ for all finitely generated left Λ-modules M of finite projective dimension. Consequently, $\mathrm{fin\,dim}(\Lambda) \leqq c$. □

We conclude this section with some comments and examples, keeping the assumptions and notation developed above. To use the first remark towards the computation of the projective dimension of arbitrary finitely generated Λ-modules, keep in mind that their first syzygies belong to $L_2(\Lambda)$.

Remark 3 Let $M \in L_2(\Lambda)$. If it is known a priori that $\mathrm{p\,dim}\, M < \infty$, then the matrix $[M]$ completely determines the projective dimension of M (see Observation 1), and the pertinent calculation can easily be carried out on a computer. However, it may occur that the projective dimension of M is infinite and that nonetheless a power of L^* vanishes on $[M]$. We will give a "minimal" illustration of this phenomenon below. The following comments may help to determine the projective dimension of M without a priori information. If $(L^*)^{p+q}[M] \neq 0$, then $\mathrm{p\,dim}\, M = \infty$ by the argument above. If $(L^*)^{p+q}[M] = 0$, but $(L^*)^k[M]$ has either precisely one vanishing column or a negative entry for some $k < p+q$, we again know that the projective dimension of M is infinite. Only once all these straightforward tests do not yield infinite projective dimension do we have to compute syzygies and use the following equivalence: if $(L^*)^{p+q}[M] = 0$, then $\mathrm{p\,dim}\, M = \infty$ if and only if there exists a nonnegative integer $k < p+q$ such that $(L^*)^k[M] \neq [\Omega^k_\Lambda(M)]$.

Example 4 Let K be any field, $K\Gamma$ the path algebra of the quiver Γ

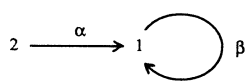


Fig. 1

and set $\Lambda = k\Gamma/\langle\beta^2\rangle$.

Then $[Je_2] = [\Lambda e_1] = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, $[Je_1] = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, and $G^*(\Lambda) = \mathbf{Z}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \subseteq \mathbf{Z}^{2\times 2}$. Hence $\mathrm{fin\,dim}(\Lambda) = 1 = \mathrm{rank}\, G^*(\Lambda)$. Moreover, if $M_1 = S_1 \coprod (\Lambda e_2 / J^2 e_2)$ and $M_2 = \Lambda e_1 \coprod S_2$, then $[M_1] = [M_2] = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, while $\mathrm{p\,dim}\, M_1 = \infty$ and $\mathrm{p\,dim}\, M_2 = 1$.

Example 5 The following is a finite dimensional algebra Λ such that $\mathrm{fin\,dim}(\Lambda) < \mathrm{rank}\, G^*(\Lambda) < p+q$.

Given a field K and the quiver Γ as shown below,

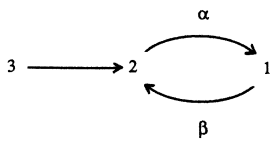


Fig. 2

let $\Lambda = K\Gamma/\langle \alpha\beta, \beta\alpha\rangle$. Clearly, $p=q=2$. Moreover,

$$G^*(\Lambda) = \mathbf{Z}\begin{pmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{pmatrix} \coprod \mathbf{Z}\begin{pmatrix}0 & 1\\ 1 & 0\\ 0 & 0\end{pmatrix};$$

indeed, a straightforward computation yields that

$$G^*(\Lambda) \subseteq \begin{pmatrix}\mathbf{Z} & \mathbf{Z}\\ \mathbf{Z} & \mathbf{Z}\\ 0 & 0\end{pmatrix} \quad \text{and} \quad L^*\begin{pmatrix}a_1 & 0\\ a_2 & 0\\ 0 & 0\end{pmatrix} = \begin{pmatrix}a_2 & 0\\ a_1 & 0\\ 0 & 0\end{pmatrix}.$$

Since $L^*A = 0$ for all $A \in G^*(\Lambda)$ and $\operatorname{p\,dim} S_1 = 1$, we obtain $\operatorname{fin\,dim}(\Lambda) = 1 < \operatorname{rank} G^*(\Lambda) = 2$. Note, however, that $\operatorname{fin\,dim}(\Lambda) = c = \sup\{\tau(A) \mid A \in G^*(\Lambda)\}$ as in the remark supplementing Theorem 2. □

To effectively compute the upper bound $\operatorname{rank} G^*(\Lambda)$ for $\operatorname{fin\,dim}(\Lambda)$ which, in most cases, is better than $p+q$, consider the matrix $\mathbf{A}$ of the transformation $1 \otimes L^*\colon \mathbf{Q} \otimes_{\mathbf{Z}} \mathbf{Z}^{n\times 2} \to \mathbf{Q} \otimes_{\mathbf{Z}} \mathbf{Z}^{n\times 2}$ with respect to the canonical basis. Let N be the nullspace of $\mathbf{A}^{2n}$, and

$$U = \begin{pmatrix}\mathbf{Q} & \mathbf{Q}\\ \vdots & \vdots\\ \cdot & \mathbf{Q}\\ \cdot & 0\\ \vdots & \vdots\\ \mathbf{Q} & \cdot\\ 0 & \cdot\\ \vdots & \vdots\\ 0 & 0\end{pmatrix}$$

where the first column has p entries of $\mathbf{Q}$ and the second column has q entries of $\mathbf{Q}$. Note that $\operatorname{rank} G^*(\Lambda) = \dim_{\mathbf{Q}}(N \cap U)$.

One should observe that the usually still smaller $1 \otimes L^*$-invariant subspace $\mathbf{Q} \otimes_{\mathbf{Z}} G^{**}(\Lambda)$ of U as in the proof of Theorem 2 is hard to find in general.

3 The general case

Keeping the hypothesis that $J^3 = 0$, we begin by conveniently grouping the isomorphism types of the simple left Λ-modules. Let $S_1, \ldots, S_m$ be the isomorphism types of the simple modules of infinite projective dimension, and let $\tilde{S}_{m+1}, \ldots, \tilde{S}_n$ be those of finite projective dimension. Moreover, let e_i, respectively $\tilde{e}_j$, be the primitive idempotents corresponding to the S_i, respectively the $\tilde{S}_j$, and set $e = e_1 + \ldots + e_m$, $\tilde{e} = \tilde{e}_{m+1} + \ldots + \tilde{e}_n$. Finally, define $d = \sup_{m+1 \leq j \leq n} \operatorname{p\,dim}_1 \tilde{S}_j$.

As might be expected, we will relate the dimensions $\mathrm{p\,dim}_A M$ and $\mathrm{p\,dim}_{eAe} eM$ for $M \in L_2(A)$. Our principal tool for doing this will be the modified matrix assignment

$$L_2(A) \to \mathbf{Z}^{m\times 2} \quad \text{with} \quad M \mapsto [M] = (\mathbf{a}, \mathbf{b}) = \begin{pmatrix} a_1 & b_1 \\ \vdots & \vdots \\ a_m & b_m \end{pmatrix},$$

where a_i, resp. b_i, denotes the multiplicity of S_i in M/JM, resp. JM. In other words, we are now recording only the infinite dimensional composition factors in the top and the radical of M. Note that even in the special case where all summands of J^2 have infinite projective dimension, the matrix $[M]$ introduced in Section 2 need not coincide with the one defined here. Next we modify the linear map L^* of Section 2 correspondingly. Letting $[Je_i] = (\mathbf{u}_i, \mathbf{v}_i) \in \mathbf{Z}^{m\times 2}$ for $1 \leqq i \leqq m$, we define

$$L\colon \mathbf{Z}^{m\times 2} \to \mathbf{Z}^{m\times 2} \quad \text{by} \quad A = (\mathbf{a}, \mathbf{b}) \mapsto \left(\sum_{i=1}^{m} a_i \mathbf{u}_i - \mathbf{b}, \sum_{i=1}^{m} a_i \mathbf{v}_i \right) = \sum_{i=1}^{m} a_i [Je_i] - (\mathbf{b}, \mathbf{0}),$$

where $\mathbf{a}$, $\mathbf{b}$ are the columns of the matrix A and the a_i denote the entries of $\mathbf{a}$.

Observation 6 and Definition. Given any $M \in L_2(A)$, let π be the canonical A-epimorphism $M/\tilde{e}JM \to M/JM$, and define

$$\varepsilon(M) = \pi^{-1}(e(M/JM)).$$

Then $\varepsilon(M)$ is the unique largest A-submodule of $M/\tilde{e}JM$ all of whose composition factors are copies of $S_1, \ldots, S_m$. It is easily checked that the inclusion map $eM \to M$ induces an eAe-module isomorphism $eM \cong \varepsilon(M)$, and hence eM inherits a natural A-module structure which extends the eAe-structure.

Moreover, observe that $\varepsilon\colon L_2(A) \to L_2(A)$ is a functor, and check that for any A-submodule A of M with $eA = eM$ the inclusion $A \subseteq M$ induces an isomorphism $\varepsilon(A) \cong \varepsilon(M)$. To see this, consult the following commutative diagram, where all maps are the canonical ones:

$$\begin{array}{ccccc} eA & \xrightarrow{\cong} & \varepsilon(A) & \longrightarrow & A/\tilde{e}JA \\ \| & & \downarrow & & \downarrow \\ eM & \xrightarrow{\cong} & \varepsilon(M) & \longrightarrow & M/\tilde{e}JM \end{array}$$

Lemma 7 *Let $M \in L_2(A)$. Then*

(a) $\mathrm{p\,dim}_A M < \infty \Leftrightarrow \mathrm{p\,dim}_A \varepsilon(M) < \infty$.

In fact, $\mathrm{p\,dim}_A \varepsilon(M) \leqq \max(\mathrm{p\,dim}_A M, d+1)$ *and* $\mathrm{p\,dim}_A M \leqq \max(\mathrm{p\,dim}_A \varepsilon(M), d)$. (b) *Let* $c \geqq 1$. *If* $(\varepsilon\Omega_A)^c\colon L_2(A) \to L_2(A)$ *is the c-fold iterate of the composite functor* $\varepsilon\Omega_A\colon L_2(A) \to L_2(A)$, *then* $(\varepsilon\Omega_A)^c(\varepsilon M) = 0$ *implies* $\mathrm{p\,dim}_A M \leqq d + c$.

Proof. Part (a) is straightforward. Part (b) we prove by induction on the exponent c. Clearly, the vanishing of $\varepsilon(\Omega_\Lambda(\varepsilon(M)))$ means that all composition factors of $\Omega_\Lambda(\varepsilon(M))$ belong to $\{\tilde{S}_{m+1}, \ldots, \tilde{S}_n\}$; hence $\operatorname{p\,dim}_\Lambda(\Omega_\Lambda(e(M))) \leqq d$, and therefore $\operatorname{p\,dim}_\Lambda M \leqq \max(\operatorname{p\,dim}_\Lambda \varepsilon(M), d) \leqq d+1$.

Letting $(\varepsilon\Omega_\Lambda)^c(\varepsilon M)=0$ for some $c \geqq 2$ and assuming our claim to be correct for $c-1$, we obtain $\operatorname{p\,dim}_\Lambda(\Omega_\Lambda(\varepsilon(M))) \leqq d+c-1$. It follows that $\operatorname{p\,dim}_\Lambda \varepsilon(M) \leqq d+c$, and hence $\operatorname{p\,dim}_\Lambda M \leqq \max(\operatorname{p\,dim}_\Lambda \varepsilon(M), d) \leqq d+c$. □

We point out the following immediate consequence of Lemma 7: For any two objects A, M in $L_2(\Lambda)$ such that $\varepsilon(A) \cong \varepsilon(M)$, the projective dimension of A is finite if and only if the projective dimension of M is finite. The latter equivalence applies in particular whenever A is a submodule of M such that $eA = eM$. This fact provides the background for Lemma 8.

Lemma 8 *Suppose that $M \in L_2(\Lambda)$ has finite projective dimension.*

(a) *If $A \subseteq M$ is a Λ-submodule with $eA = eM$, then*

$$[A] = [M] = [\varepsilon(M)].$$

(b) $[(\varepsilon\Omega_\Lambda)^k(\varepsilon(M))] = L^k[M] = L^k[\varepsilon(M)]$ *for all* $k \geqq 1$.

Proof. For part (a) it suffices to show that $[M] = [\varepsilon(M)]$, because $\varepsilon(A) \cong \varepsilon(M)$ by hypothesis. Setting $\bar{M} = M/\tilde{e}JM$, we clearly have $[M] = [\bar{M}]$ and $\varepsilon(M) \cong \varepsilon(\bar{M})$. Hence it is enough to show $[\bar{M}] = [\varepsilon(\bar{M})]$. Since $\operatorname{p\,dim}_\Lambda \varepsilon(\bar{M}) < \infty$, we know that $J\varepsilon(\bar{M}) = \operatorname{soc} \varepsilon(\bar{M}) \supseteq eJ\bar{M}$. The other inclusion, namely $J\varepsilon(\bar{M}) \subseteq eJ\bar{M}$, is obvious, and we obtain $J\varepsilon(\bar{M}) = eJ\bar{M}$. Thus the second columns of the matrices $[\varepsilon(\bar{M})]$ and $[\bar{M}]$ are the same. But, since the S_i occur in the same multiplicity as composition factors of $\varepsilon(\bar{M})$ and $\bar{M}$, the equality $[\varepsilon(\bar{M})] = [\bar{M}]$ follows.

Part (b) can now be established by a straightforward induction using the argument of Observation 1. □

Lemmas 7(b) and 8(b) indicate the direction which we will follow in the sequel. First we will show that, whenever $M \in L_2(\Lambda)$ has finite projective dimension k, we obtain $(\varepsilon\Omega_\Lambda)^k(\varepsilon(M)) = 0$ and thus $L^k[M] = 0$. In analogy to Section 2 we can then find a positive integer c such that $L^c[M] = 0$ for all $M \in L_2(\Lambda)$ of finite projective dimension. But in view of the preceding lemmas, the latter means $\operatorname{p\,dim}_\Lambda M \leqq c+d$ for all $M \in L_2(\Lambda)$ of finite projective dimension, which will place the upper bound $c+d+1$ on the finitistic dimension of Λ.

We will smooth the road with a few additional auxiliary remarks.

Remark 9 Let $K \in L_2(\Lambda)$ and $A \subseteq K$ be a submodule of K such that each indecomposable summand of A has Loewy length 2. Then, given any projective cover $f\colon P \to A$, there is a projective cover $g\colon P \oplus Q \to K$ with $g|_P = f$.

Proof. It suffices to observe that $\operatorname{soc} A = JA$ by hypothesis and hence $JA = A \cap JK$. □

Lemma 10 *Let $M \in L_2(\Lambda)$ with $\operatorname{p\,dim}_\Lambda M < \infty$, and suppose that $A \subseteq M$ is a submodule with $(A+JM)/JM = e(M/JM)$, and such that A has no simple direct summand isomorphic to some $\tilde{S}_j$. (Such a choice of A is always possible.) Then $[A] = [M]$, $eA = eM$, and $\operatorname{p\,dim}_\Lambda A < \infty$. Moreover, every nonzero direct summand of A has Loewy length 2.*

Proof. Let A be as described, and, for any subset X of M, let $\bar{X}$ be the canonical image of X in $\bar{M}=M/\tilde{e}JM$. In order to prove $[M]=[A]$, note first that $eJ\bar{A}\cong eJA$ as Λ-modules. We claim that it suffices to show the equality $J\varepsilon(M)=eJ\bar{A}$. Indeed, from $J\varepsilon(M)\cong eJA$ we deduce that the second columns of $[M]=[\varepsilon(M)]$ and $[A]$ are the same. On the other hand, $e(M/JM)=e((A+JM)/JM)$ is a homomorphic image of $e(A/JA)$, from which it follows that each coefficient in the first column of the matrix $[A]$ is larger than or equal to the corresponding coefficient of $[M]$. Since, on the other hand, the multiplicity of each simple S_i in A is smaller than or equal to its multiplicity in M, the matrices $[A]$ and $[M]$ must be equal.

To check that $J\varepsilon(M)=eJ\bar{A}$, observe that we have $\varepsilon(M)\subseteq\bar{A}+eJ\bar{M}$ by the choice of A, and thus $J\varepsilon(M)\subseteq J\bar{A}$. Since $J\varepsilon(M)=e(J\varepsilon(M))$, we deduce further tha $J\varepsilon(M)\subseteq eJ\bar{A}$. But, clearly, we also have $eJ\bar{A}\subseteq eJ\bar{M}\subseteq\varepsilon(M)$, because $\varepsilon(M)$ is the largest Λ-submodule of $\bar{M}$ with $\varepsilon(M)=e\varepsilon(M)$. Due to the finiteness of $\mathrm{p\,dim}_\Lambda\varepsilon(M)$, we even obtain that $eJ\bar{M}\subseteq\mathrm{soc}\,\varepsilon(M)=J\varepsilon(M)$. Thus $J(\varepsilon(M))=eJ\bar{A}$, and we conclude that $[A]=[M]$ by the first paragraph. The equality $eM=eA$ follows, and so does finiteness of the projective dimension of A (see Lemma 7 and the remark following it).

From $A=A'\oplus A''$ with Loewy length $(A'')\leqq 1$, finally, we deduce that A'' is a direct sum of copies of $S_1,\ldots,S_m$, because A has no summands isomorphic to $\tilde{S}_{m+1},\ldots,\tilde{S}_n$ by hypothesis. Therefore, the finiteness of $\mathrm{p\,dim}_\Lambda A''$ forces A'' to be zero. □

Proposition 11 *Suppose that $M\in L_2(\Lambda)$ has finite projective dimension, say $\mathrm{p\,dim}_\Lambda M=k$. Then there exists a minimal projective resolution*

$$0\longrightarrow P_k\oplus Q_k\xrightarrow{\alpha_k}\ldots\longrightarrow P_1\oplus Q_1\xrightarrow{\alpha_1}P_0\oplus Q_0\xrightarrow{\alpha_0}M\longrightarrow 0$$

of M with the following properties: Each P_i is a direct sum of copies of $\Lambda e_1,\ldots,\Lambda e_m$, and $\alpha_i(P_i)\subseteq\ker(\alpha_{i-1}|_{P_{i-1}})$; that is, the sequence

$$P_k\xrightarrow{\alpha_k|_{P_k}}P_{k-1}\longrightarrow\ldots\longrightarrow P_1\xrightarrow{\alpha_1-P_1}P_0\xrightarrow{\alpha_0-P_0}M$$

is a complex. Moreover, the induced sequence

$$0\to eP_k\to eP_{k-1}\to\ldots\to eP_1\to eP_0\to eM\to 0$$

of $e\Lambda e$-modules is a minimal projective resolution of eM (with extra zeros at the end, in general).

Proof. We construct α_i, P_i, and Q_i by induction as follows. Pick $A_0\subseteq M$ as in the hypothesis of Lemma 10. By Remark 9 and Lemma 10 there is a projective cover $\alpha_0\colon P_0\oplus Q_0\to M$ such that $\alpha_0|_{P_0}\colon P_0\to A_0$ is a projective cover of A_0. Since A_0/JA_0 is a direct sum of copies of $S_1,\ldots,S_m$, P_0 is a direct sum of copies of $\Lambda e_1,\ldots,\Lambda e_m$. Set $K_1=\ker(\alpha_0|_{P_0})$. Since $\mathrm{p\,dim}_\Lambda A_0<\infty$, by Lemma 10, we obtain $\mathrm{p\,dim}_\Lambda K_1<\infty$, and if we choose $A_1\subseteq K_1$ as in Lemma 10, then $\mathrm{p\,dim}_\Lambda A_1<\infty$ as well.

Now suppose that $0 \leqq i \leqq k$ and that we have already constructed a minimal projective resolution

$$P_i \oplus Q_i \xrightarrow{\alpha_i} P_{i-1} \oplus Q_{i-1} \xrightarrow{\alpha_{i-1}} \ldots \longrightarrow P_0 \oplus Q_0 \xrightarrow{\alpha_0} M \longrightarrow 0$$

and submodules $A_{j+1} \subseteq K_{j+1} = \ker(\alpha_j|_{P_j})$ for all $j \leqq i$ such that each P_j is a direct sum of copies of $\Lambda e_1, \ldots, \Lambda e_m$, $\mathrm{p\,dim}_\Lambda K_{j+1}$ and $\mathrm{p\,dim}_\Lambda A_{j+1}$ are both finite, and $\alpha_j|_{P_j}\colon P_j \to A_j$ is a projective cover. Moreover, suppose that each A_{j+1} is a submodule of K_{j+1} satisfying the hypothesis of Lemma 10.

If $i=k$, then $K_{i+1} = \ker(\alpha_k|_{P_k}) = 0$; we stop in that case. If $i<k$, Remark 9 and Lemma 10 permit us to choose a projective cover

$$\alpha_{i+1}\colon P_{i+1} \oplus Q_{i+1} \to \ker(\alpha_i)$$

such that $\alpha_{i+1}|_{P_{i+1}}\colon P_{i+1} \to A_{i+1}$ is a projective cover of A_{i+1}. Set $K_{i+2} = \ker(\alpha_{i+1}|_{P_{i+1}})$, choose $A_{i+2} \subseteq K_{i+2}$ according to the hypothesis of Lemma 10, and check that all features of our induction hypothesis are reestablished. Note in particular that $\mathrm{p\,dim}_\Lambda K_{i+2} < \infty$ by the induction hypothesis, which implies that $\mathrm{p\,dim}_\Lambda A_{i+2}$ is also finite.

Having constructed a minimal projective resolution of M as indicated, consider the complex

$$\begin{array}{ccccccccccc}
0 & \longrightarrow & P_k & \xrightarrow{\alpha_k - P_k} & P_{k-1} & \xrightarrow{\alpha_{k-1}|_{P_k-1}} & \ldots & \longrightarrow & P_1 & \xrightarrow{\alpha_1|_{P_1}} & P_0 \xrightarrow{\alpha_0|_{P_0}} A_0 \longrightarrow 0, \\
 & & \downarrow & & \uparrow & & & & \downarrow & & \uparrow \\
 & & A_k & \xrightarrow{\text{incl}} & K_k & & & & A_1 & \xrightarrow{\text{incl}} & K_1
\end{array}$$

where $K_i = \ker(\alpha_{i-1}|P_{i-1})$. Since $eA_i = eK_i$ for $1 \leqq i \leqq k$ by construction, this complex becomes exact upon multiplication from the left with e; that is, we obtain an exact sequence

$$0 \to eP_k \to eP_{k-1} \to \ldots \to eP_0 \to eA_0 = eM \to 0$$

of $e\Lambda e$-modules. Each P_i being a direct sum of copies of $\Lambda e_1, \ldots, \Lambda e_m$, this sequence is an $e\Lambda e$-projective resolution of eM. It is even a minimal resolution because the kernel of the i^{th} map is contained in eJP_i by construction. □

Corollary 12 *If $M \in L_2(\Lambda)$ has finite projective dimension, then* $\mathrm{p\,dim}_{e\Lambda e}\, eM \leqq \mathrm{p\,dim}_\Lambda M$. □

Note that the construction described in the proof of Proposition 11 can readily be adapted to yield a slight generalization of this inequality as follows.

Remark 13 Let $M \in L_2(\Lambda)$ and $N \subseteq M$ with $JN = N \cap JM$ such that both $\mathrm{p\,dim}_\Lambda M$ and $\mathrm{p\,dim}_\Lambda N$ are finite. Then $\mathrm{p\,dim}_{e\Lambda e}\, eN \leqq \mathrm{p\,dim}_\Lambda M$. For a proof, pick $A_0 \subseteq N$ according to the hypothesis of Lemma 10 and then follow the proof of Proposition 11 verbatim. □

Corollary 14 *If $M \in L_2(\Lambda)$ has finite projective dimension, then there exists a non-negative integer k such that $(\varepsilon\Omega_\Lambda)^k(\varepsilon(M))=0$.*

Proof. All we need to show in view of Corollary 12 is that $\Omega_{e\Lambda e}(eM) \cong (\varepsilon\Omega_\Lambda)(\varepsilon(M))$ as $e\Lambda e$-modules. Indeed, a Λ-projective cover

$$0 \to \Omega_\Lambda(\varepsilon(M)) \to P_0 \to \varepsilon(M) \to 0$$

induces an exact sequence of $e\Lambda e$-modules,

$$0 \to e\Omega_\Lambda(\varepsilon(M)) \to eP_0 \xrightarrow{\alpha} e\varepsilon(M) \cong eM \to 0$$

Since, as is easily checked, $\alpha: eP_0 \to eM \to 0$ is an $e\Lambda e$-projective cover, we obtain $\Omega_{e\Lambda e}(eM) \cong e\Omega_\Lambda(\varepsilon(M)) \cong (\varepsilon\Omega_\Lambda)(\varepsilon(M))$ as desired; all of the latter isomorphisms respect the $e\Lambda e$-structures. Inductively, we deduce that $\Omega^k_{e\Lambda e}(eM) \cong (\varepsilon\Omega_\Lambda)^k(\varepsilon(M))$, and the corollary follows. □

Proposition 15 *Let M be a Λ-module with $\mathrm{p\,dim}_\Lambda M < \infty$. If $M \in L_2(\Lambda)$, set $C(M)=[M]$, otherwise, set $C(M)=[\Omega_\Lambda(M)]$. Then there exists an integer c such that $L^c(C(M))=0$, and $\mathrm{p\,dim}_\Lambda M \leqq c+d$ in the first case, while $\mathrm{p\,dim}_\Lambda M \leqq c+d+1$ in the second.*

Proof. Combine Lemma 8 and Corollary 14 to show the existence of an integer c as desired. Then use Lemma 7 to deduce the bounds on the projective dimension of M. □

By analogy with our procedure in Sect. 2, we will now focus on the subgroup $G_0(\Lambda)$ of $\mathbf{Z}^{m\times 2}$, consisting of all matrices A such that $L^a(A)=0$ for some positive integer a. Moreover, we will assume that $S_1, \ldots, S_p$ are those simple Λ-modules of infinite projective dimension which occur in ${}_\Lambda(J/J^2)$, and $S_1, \ldots, S_q$ $(q \leqq p)$ the ones which occur in both ${}_\Lambda(J/J^2)$ and in ${}_\Lambda(J^2)$. Define

$$G(\Lambda) = G_0(\Lambda) \cap \begin{pmatrix} \mathbf{Z} & \mathbf{Z} \\ \vdots & \vdots \\ \vdots & \mathbf{Z} \\ \cdot & 0 \\ \vdots & \vdots \\ \mathbf{Z} & \vdots \\ 0 & \cdot \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix},$$

where the first column has p entries of $\mathbf{Z}$ and the second column has q. Let c be the smallest non-negative integer with the property that $L^c(G(\Lambda))=0$. We can then summarize our findings as follows:

Theorem 16 *If Λ is a left artinian ring with $J^3=0$, the finitistic dimension of Λ is finite. In fact,*

$$\text{fin dim}(\Lambda)\leqq c+d+1,$$

and $c\leqq \text{rank } G(\Lambda)\leqq p+q\leqq 2m$. In case all simple left Λ-modules have infinite projective dimension, $\text{fin dim}(\Lambda)\leqq \text{rank } G(\Lambda)$.

Proof. Consult the proof of Theorem 2 to see that $[\Omega_\Lambda(M)]\in G(\Lambda)$ whenever $M\in\Lambda$-mod has finite projective dimension, and that $c\leqq \text{rank } G(\Lambda)$. In combination with Proposition 15 this proves the first assertion. For the second, it suffices to observe that $G(\Lambda)=G^*(\Lambda)$ in case $m=n$. □

Theorem 16 invites attempts to replace the condition "$J^3=0$" by milder finiteness conditions on J^3. One such generalization, involving a finiteness condition for J^3 as a *right* Λ-module, can be derived without further effort from Theorem 16 and a result of Small; however, the resulting bound on $\text{fin dim}(\Lambda)$ described in Corollary 17 below is very rough and its computation is based on homological information about both Λ-mod and mod-Λ. In Section 5 we will, by contrast, consider a left finiteness condition for J^3 which will allow us to extend our matrix algorithm to facilitate computations of bounds on the finitistic dimension of Λ.

Let $I\subseteq J$ be a two-sided ideal of Λ. As is well-known, no helpful correlations between $\text{p dim}_\Lambda M$ and $\text{p dim}_{\Lambda/I} M/IM$ for left Λ-modules M can be expected in general. However, by Small's [11, Lemma 1] these dimensions are equal provided that $\text{Tor}_l^\Lambda(\Lambda/I, M)=0$ for all $l\geqq 1$. The authors are grateful to E. Kirkman for bringing this result to their attention. Combining it with Theorem 16 yields Corollary 17.

Corollary 17 *Suppose that Λ contains a two-sided ideal I with $J^3\subseteq I\subseteq J$ such that I has finite projective dimension k as a right Λ-module. Then* $\text{fin dim}(\Lambda)$ *is finite; more precisely,*

$$\text{fin dim}(\Lambda)\leqq \text{rank } G(\Lambda/I)+d+k+2,$$

where d is the supremum of the finite numbers among the projective dimensions of the simple left Λ/I-modules.

Proof. Let M be in Λ-mod. Clearly $\text{Tor}_l^\Lambda(\Lambda/I, \Omega^{k+1}(M))=0$ for all $l\geqq 1$ since the right Λ-module Λ/I has projective dimension $k+1$. Hence $\text{p dim}_\Lambda \Omega^{k+1}(M)=\text{p dim}_{\Lambda/I} \Omega^{k+1}(M)/I\Omega^{k+1}(M)$ by [11, Lemma 1] as quoted above. If $\text{p dim}_\Lambda M$ is finite, we thus obtain $\text{p dim}_\Lambda \Omega^{k+1}(M)\leqq \text{fin dim}(\Lambda/I)\leqq \text{rank } G(\Lambda/I)+d+1$ by Theorem 16, which implies $\text{p dim}_\Lambda M\leqq \text{rank } G(\Lambda/I)+d+k+2$ as desired. □

The special case where $I=J^2$ has been previously observed by Small.

4 Examples

While Theorem 16 covers the specialized situation of Section 2, the preliminary technique developed for Theorem 2 has an edge over the general version leading up to Theorem 16, whenever the former is applicable; in fact, this technique

does not require the computation of $d = \sup_{m+1 \leqq i \leqq n} p \dim \tilde{S}_i$, provided Λ does not have simple left modules of finite projective dimension occurring as summands of J^2.

An algorithm for the computation of rank $G(\Lambda)$ is available in complete analogy to that sketched at the end of Theorem 16; just replace the matrix $\mathbf{A}$ of $L^* \otimes \mathbf{Q}$ by the matrix of $L \otimes \mathbf{Q}$. The remarks following the statement of Theorem 2 extend to Theorem 16 as well: If you have enough computer time or stamina to compute $c = \sup_{1 \leqq i \leqq r} (\inf\{a \in \mathbf{N} \mid L^a(A_i) = 0\})$, where $A_1, \ldots, A_r$ is a generating set for $G(\Lambda)$, you will prefer to look at the (occasionally sharper) bound $\operatorname{fin\,dim}(\Lambda) \leqq c + d + 1$.

We illustrate the sharpness of our bound with a number of examples. Clearly, the bound in Theorem 16 exceeds $\operatorname{fin\,dim}(\Lambda)$ by at least 1 in case the left global dimension of Λ is finite. On the other hand, we will see that even when $d = 0$, the "corrective" term "$+1$" in Theorem 16, accounting for the coexistence of simple left modules of finite and of infinite projective dimension, cannot be dropped in general. All of our examples will be "path algebras modulo relations"; for general background on this type of construction we refer to the Appendix of [5].

Example 18 Given any non-negative integer n, we exhibit a finite dimensional factor $\Lambda = K\Gamma/\langle\rho\rangle$ of a path algebra such that all simple left Λ-modules have infinite projective dimension, and such that $\operatorname{fin\,dim}(\Lambda) = n = \operatorname{rank} G(\Lambda) = \operatorname{rank} G^*(\Lambda)$. Let K be any field, Γ the quiver

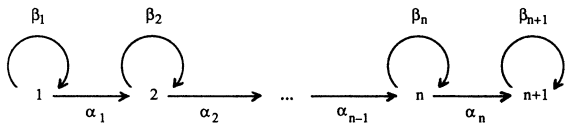


Fig. 3

and $\rho = \{\alpha_{i+1}\alpha_i, \alpha_i\beta_i \mid 1 \leqq i \leqq n\} \cup \{\beta_i^2 \mid 1 \leqq i \leqq n+1\}$. Clearly, $J^3 = 0$, and all simple left modules S_i have infinite projective dimension. It is readily checked that

$$[Je_i] = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \quad \text{for } 1 \leqq i \leqq n$$

where the entries 1 occur in the i^{th} and $(i+1)^{\text{st}}$ rows, and

$$[Je_{n+1}] = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Since S_1 does not occur in J^2, we know that

$$G(\Lambda)=G^*(\Lambda)\subseteq U\subseteq \mathbf{Z}^{(n+1)\times 2} \quad \text{where} \quad U=\begin{pmatrix}\mathbf{Z} & 0\\ \mathbf{Z} & \mathbf{Z}\\ \vdots & \vdots\\ \mathbf{Z} & \mathbf{Z}\end{pmatrix};$$

note that the role played by the simple modules $S_1, \ldots, S_q$ in our formal development is taken over by $S_2, \ldots, S_{n+1}$. Next, we set

$$V=\begin{pmatrix}\mathbf{Z} & 0\\ \vdots & \vdots\\ \mathbf{Z} & 0\end{pmatrix}$$

and verify that $L^a(\Lambda)\neq 0$ for all $a\geqq 1$ and all nonzero $A\in\mathbf{V}$. Therefore, $G(\Lambda)\cap V=0$, and $G(\Lambda)$ is isomorphic to a subgroup of

$$U/V\cong\begin{pmatrix}0\\ \mathbf{Z}\\ \vdots\\ \mathbf{Z}\end{pmatrix}.$$

We conclude that rank $G(\Lambda)\leqq n$. In fact, it is not difficult to check that $G(\Lambda)=\bigoplus_{2\leqq i<n+1}\mathbf{Z}(E_{i1}+E_{i2})$, where the E_{ij} denote the matrix units. By either Theorem 2 or Theorem 16 we obtain fin dim$(\Lambda)\leqq$ rank $G(\Lambda)=n$.

On the other hand, p dim$(\Lambda e_1/\Lambda\alpha_1)=n$, because $\Omega^{i-1}(\Lambda e_1/\Lambda\alpha_1)\cong\Lambda e_i/\Lambda\alpha_i$ for $2\leqq i\leqq n$, and $\Omega^n(\Lambda e_1/\Lambda\alpha_1)\cong\Lambda\alpha_n\cong\Lambda e_{n+1}$. Thus fin dim$(\Lambda)=$rank $G(\Lambda)=n$. □

We now apply our machinery to a non-monomial algebra.

Example 19 Let K be a field, Γ the following quiver

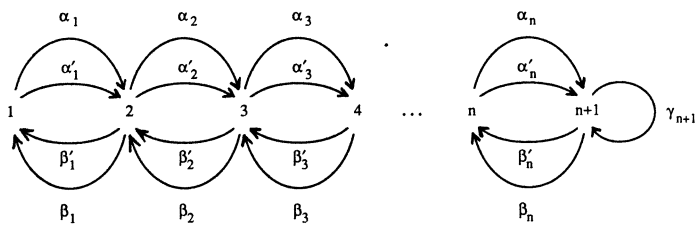


Fig. 4

and $\Lambda=K\Gamma/\langle\rho\rangle$, where ρ includes all paths of length 3, the relations $\beta_i\alpha_i-\beta_i'\alpha_i$, $\beta_i\alpha_i'$, $\alpha_i\beta_i$, $\alpha_i'\beta_i$, $\alpha_i\beta_i'$, $\alpha_i'\beta_i'$ for $1\leqq i\leqq n$, as well as γ_{n+1}^2 and all paths $\varepsilon\delta$, where ε and δ are arrows equipped with different indices.

We check that

$$[Je_1]=\begin{pmatrix}0 & 2\\ 2 & 0\\ 0 & 0\\ \vdots & \vdots\\ 0 & 0\end{pmatrix}, \quad [Je_i]=\begin{pmatrix}0 & 0\\ \vdots & \vdots\\ 0 & \cdot\\ 2 & 0\\ 0 & 2\\ 2 & 0\\ 0 & 0\\ \vdots & \vdots\\ 0 & 0\end{pmatrix}, \quad \text{for } 2\leqq i\leqq n$$

where the 2 in the second column occurs in the i^{th}-row, and

$$[Je_{n+1}]=\begin{pmatrix}0 & 0\\ \vdots & \vdots\\ 2 & \cdot\\ 1 & 0\end{pmatrix};$$

we readily derive that all simple modules have infinite projective dimension. Consequently,

$$G(\Lambda)=G^*(\Lambda)\subseteq U=\begin{pmatrix}\mathbf{Z} & \mathbf{Z}\\ \vdots & \vdots\\ \cdot & \mathbf{Z}\\ \mathbf{Z} & 0\end{pmatrix}\subseteq \mathbf{Z}^{(n+1)\times 2}.$$

On the basis of this information, it is easily checked that the restriction of L to U has trivial kernel. Thus $G(\Lambda)=0$, and therefore fin dim$(\Lambda)=0$. □

Next we demonstrate that for every pair (r, d) of non-negative integers there is a finite dimensional algebra Λ with fin dim$(\Lambda)=r+d+1$, where $r=\text{rank } G(\Lambda)$ and $d=\sup \text{p dim } \tilde{S}_i$, if the $\tilde{S}_i$ are the simple left modules of finite projective dimension as in Sect. 3.

Example 20 Start with any field K together with the following quiver Γ

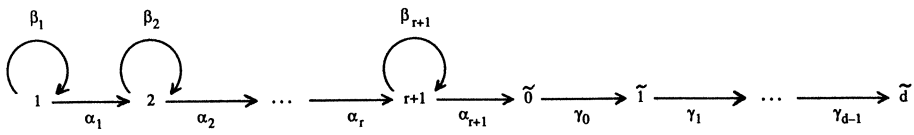


Fig. 5

and let ρ contain all those paths composed of the α_i, β_j, $1 \leqq i, j \leqq r+1$, which we listed in Example 18, as well as the paths $\gamma_0 \alpha_{r+1}$ and $\gamma_i \gamma_{i-1}$ for $1 \leqq i \leqq d-1$. If $\Lambda = K\Gamma/\langle\rho\rangle$, the computations carried out for Example 17 then show that rank $G(\Lambda) = r$. Since for $0 \leqq i \leqq d$ the projective dimension of the simple $\tilde{S}_i$ corresponding to $\tilde{i}$ is $d-i$, we obtain further that $d = \sup_{0 \leqq i \leqq d} \mathrm{p\,dim}\, \tilde{S}_i$, and hence $\mathrm{fin\,dim}(\Lambda) \leqq r+d+1$ by Theorem 16. Finally, check that $\mathrm{p\,dim}(\Lambda e_1/\Lambda\alpha_1) = r+d+1$ to conclude that $\mathrm{fin\,dim}(\Lambda) = r+d+1$ as desired. □

5 Extensions of the main result and applications

At this point we drop the hypothesis that $J^3 = 0$.

In [8] Mochizuki showed $\mathrm{fin\,dim}(\Lambda)$ to be finite provided that all simple composition factors of the *left* Λ-module J^2 have finite projective dimension; in this case, $d \leqq \mathrm{fin\,dim}(\Lambda) \leqq d+1$, where d is again the supremum of the projective dimensions of those simple left Λ-modules whose projective dimensions are finite. We include a brief argument establishing this fact for comparison with the much more complex situation where all composition factors of the left Λ-module J^3 have finite projective dimension.

Under the above hypothesis on J^2, let M be a left Λ-module of finite projective dimension. Then all composition factors of $J^2 M$ have finite projective dimension, hence $\mathrm{p\,dim}(J^2 M) \leqq d$ and $\mathrm{p\,dim}(M/J^2 M) < \infty$. Thus it clearly suffices to show that $\mathrm{p\,dim}(M/J^2 M) \leqq d+1$. Let $0 \to K \to P \to M/J^2 M \to 0$ be a projective cover. Due to the inclusion $JK \subseteq J^2 P$, our hypothesis forces all simple composition factors of JK to have finite projective dimension, and we conclude that $\mathrm{p\,dim}(JK) \leqq d$ and $\mathrm{p\,dim}(K/JK) < \infty$. But, since K/JK is semisimple, the latter implies that $\mathrm{p\,dim}(K/JK) \leqq d$, and we deduce $\mathrm{p\,dim}(K) \leqq d$. This proves that $\mathrm{p\,dim}(M/J^2 M) \leqq d+1$ as required.

The primary goal of this section is to push somewhat beyond the obvious simultaneous generalization of Theorem 16 and Mochizuki's result, that obvious generalization being finiteness of $\mathrm{fin\,dim}(\Lambda)$ in case all simple composition factors of the left Λ-module J^3 have finite projective dimension.

One of the reasons for our stepwise procedure lies in the fact that the bulk of definitions and notation required to set up our machinery for finding upper bounds on $\mathrm{fin\,dim}(\Lambda)$ will increase, while the bounds themselves will deteriorate. On the other hand, the bounds will still be attained, and the machinery will – aside from the computation of d – still be algorithmic in nature and no less efficient than that employed in the proof of Theorem 16. In fact, the technique leading up to Theorem 21 below may provide rougher bounds more quickly.

Recall that the left quiver of a basic left artinian ring Λ is the directed graph $\Gamma(\Lambda)$ which has a vertex for each primitive idempotent of Λ and contains l arrows $e_i \to e_j$ if and only if the simple module $\Lambda e_j/J e_j$ occurs (up to isomorphism) as a summand of multiplicity l in $J e_i/J^2 e_i$.

Given a finite directed graph Γ, we call a subgraph Δ of Γ *full* if all arrows in Γ between vertices in Δ belong to Δ. We will, moreover, say that Δ is a *sink* in Γ if Γ contains no arrows leaving Δ, i.e., if for any two vertices $v \in \Delta$ and $w \in \Gamma \setminus \Delta$ there is no arrow from v to w. The supremum of the lengths of paths in Γ will be denoted by $l(\Gamma)$. In particular, $l(\Gamma) < \infty$ precisely if Γ has no oriented cycles.

New machinery, including revised definitions of $S_1, \ldots, S_m, \tilde{S}_{m+1}, \ldots, \tilde{S}_n$, a group G of matrices and a linear operator L of G

Suppose in the following that Λ is a basic left artinian ring and that the quiver $\Gamma = \Gamma(\Lambda)$ contains a full subgraph Δ which is a sink in Γ and has the following additional property:

(*) For each vertex e_i in Δ the corresponding summand $J^3 e_i$ of J^3 has only simple composition factors of finite projective dimension.

Let $e_1, \ldots, e_m$ be those vertices in Δ which correspond to simple modules $S_i = \Lambda e_i / J e_i$ of infinitive projective dimension, and $\tilde{e}_{m+1}, \ldots, \tilde{e}_n$ the other vertices in Δ, i.e., those belonging to simple modules $\tilde{S}_j = \Lambda \tilde{e}_j / J \tilde{e}_j$ of finite projective dimension. Set $e = e_1 + \ldots + e_m$, $\tilde{e} = \tilde{e}_{m+1} + \ldots + \tilde{e}_n$, and $d = \sup_{m+1 \leqq j \leqq n} \operatorname{p\,dim}(\tilde{S}_j)$.

Observe that since Δ is a sink in Γ, if $M \in \Lambda$-mod has the property that M/JM is a direct sum of copies of $S_1, \ldots, S_m, \tilde{S}_{m+1}, \ldots, \tilde{S}_n$, then all composition factors of the module M and of its syzygies belong to this same list of simple modules. The category of all finitely generated left Λ-modules whose composition factors belong to $S_1, \ldots, S_m, \tilde{S}_{m+1}, \ldots, \tilde{S}_n$ will be denoted by Δ-mod. The role which was previously played by $L_2(\Lambda)$ will be taken over by the full subcategory $L_2'(\Lambda)$ of Δ-mod having as objects those modules M in which all simple composition factors of $J^2 M$ have finite projective dimension, i.e., all composition factors of $J^2 M$ belong to $\{\tilde{S}_{m+1}, \ldots, \tilde{S}_n\}$. Clearly, we have $\Omega^k_\Lambda(N) \in L_2'(\Lambda)$ for all $N \in \Delta$-mod and $k \geqq 1$.

Define a map $L_2'(\Lambda) \to \mathbf{Z}^{m \times 2}$ in analogy to that of Sect. 3 by assigning to a module $M \in L_2'(\Lambda)$ the matrix

$$[M] = \begin{pmatrix} a_1 & b_1 \\ \vdots & \vdots \\ a_m & b_m \end{pmatrix},$$

where a_i is the multiplicity of S_i in M/JM, and b_i is the multiplicity of S_i in JM/J^2M for $1 \leqq i \leqq m$. Again, we will consider an appropriate $\mathbf{Z}$-linear map, namely $L\colon \mathbf{Z}^{m \times 2} \to \mathbf{Z}^{m \times 2}$ given by $(\mathbf{a}, \mathbf{b}) \mapsto \sum_{i=1}^{m} a_i [J e_i] - (\mathbf{b}, \mathbf{0})$. In particular, our matrix set-up involves only those simple modules of infinite projective dimension whose vertices belong to Δ.

We start by extending the functor ε of Observation 6, which is defined on the objects of Loewy length $\leqq 2$ in $L_2'(\Delta)$, to a functor $\varepsilon\colon L_2'(\Delta) \to L_2(\Delta)$ as follows: Let $M \in L_2'(\Delta)$, and let $\mathscr{F}(M)$ be the unique largest Λ-submodule of JM all of whose simple composition factors belong to $\tilde{S}_{m+1}, \ldots, \tilde{S}_n$. Clearly, $J^2 M \subseteq \mathscr{F}(M)$ by our assumption on M. Moreover, we let $\pi\colon \bar{M} = M/\mathscr{F}(M) \to M/JM$ be the canonical epimorphism and define

$$\varepsilon(M) = \pi^{-1}(e(M/JM)).$$

The extended definition of ε clearly coincides with the old one on the overlap of $L_2'(\Delta)$ and $L_2(\Delta)$. One easily checks that all the statements under Observation 6 remain true in the wider context. Moreover, a straightforward verification shows that Lemmas 7 and 8 carry over verbatim if in each case M is chosen

in $L_2'(\Delta)$; in the proofs set $\bar{M}=M/\mathscr{F}(M)$. Call these upscaled versions Lemma 7′ and Lemma 8′, respectively, for clearer reference. By combining these results, we obtain the following implication for all M in $L_2'(\Delta)$:

$$L^c[M]=0 \text{ implies that either } \mathrm{p\,dim}_\Delta M \leqq c+d \text{ or } \mathrm{p\,dim}_\Delta M=\infty.$$

In showing the existence of an integer $c\geqq 1$ such that $L^c[M]=0$ for all modules M in $L_2'(\Delta)$ of finite projective dimension, we will try to relate the new situation back to that considered in Sect. 3, which requires upgrading of Remark 9 and Lemma 10.

For X in $L_2'(\Delta)$, let $\mathscr{G}(X)$ be the largest submodule of X all of whose simple composition factors belong to $\tilde{S}_{m+1}, \ldots, \tilde{S}_n$. Clearly, $\mathscr{F}(X)\subseteq\mathscr{G}(X)$, and equality holds if and only if $\mathscr{G}(X)\subseteq JX$.

Remark 9′ (1) Let $A\in L_2'(\Delta)$ with $\mathscr{G}(A)\subseteq JA$ and $\dim_\Delta A<\infty$. Then $A\cap JK=JA$ for any module $K\in L_2'(\Delta)$ containing A as a submodule. In particular, given any projective cover $f\colon P\to A$, there exists a projective cover $g\colon P\oplus Q\to K$ such that g restricted to P equals f.

(2) If $M\in L_2'(\Delta)$, there exists a submodule $A\subseteq M$ with $(A+JM)/JM=e(M/JM)$, having the additional properties that $A\cap JM=JA$ and $\mathscr{G}(A)\subseteq JA$.

Proof. (1) We only need to show $A\cap JK=JA$. To check this equation, consider $(A\cap JK+\mathscr{G}(A))/\mathscr{G}(A)\subseteq A/\mathscr{G}(A)$. Since $J(A\cap JK)\subseteq A\cap J^2K\subseteq\mathscr{G}(A)$, we see that $(A\cap JK+\mathscr{G}(A))/\mathscr{G}(A)$ is contained in the socle of $A/\mathscr{G}(A)$. By definition of $\mathscr{G}(A)$, the socle of $A/\mathscr{G}(A)$ has no simple component of finite projective dimension. On the other hand, $\mathrm{p\,dim}_\Delta A/\mathscr{G}(A)<\infty$ by hypothesis. We infer $\mathrm{soc}(A/\mathscr{G}(A))=J(A/\mathscr{G}(A))$, using the fact that every simple submodule of $A/\mathscr{G}(A)$ not contained in $J(A/\mathscr{G}(A))$ is a direct summand of $A/\mathscr{G}(A)$. Thus, $(A\cap JK+\mathscr{G}(A))/\mathscr{G}(A)\subseteq J(A/\mathscr{G}(A))$, and hence $A\cap JK\subseteq JA+\mathscr{G}(A)=JA$ by hypothesis. The other inclusion is obvious.

(2) Let $e(M/JM)=T_1\oplus\ldots\oplus T_l$, where each T_j is a simple module. If $T_j\cong S_{\sigma(j)}$, pick an element $t_j\in M$ such that t_j+JM generates T_j and $e_{\sigma(j)}t_j=t_j$, and define $A=\sum_{j=1}^{l}\Delta t_j$. Then clearly, $A+JM/JM=T_1\oplus\ldots\oplus T_l$. Moreover, for any $a\in A\cap JM$, say $a=\sum_{j=1}^{l}\lambda_j t_j=\sum_{j=1}^{l}\lambda_j e_{\sigma(j)}t_j$ with $\lambda_j\in\Delta$, we have $\lambda_j e_{\sigma(j)}\in J$ by the choice of the t_j. Thus, $a\in\sum_{j=1}^{l}Jt_j=JA$. We conclude that $A/JA\cong A+JM/JM$ is a direct sum of simple modules of infinite projective dimension and obtain $\mathscr{G}(A)\subseteq JA$ as desired. □

Lemma 10′ *Let $M\in L_2'(\Delta)$ with $\mathrm{p\,dim}_\Delta M<\infty$, and suppose that $A\subseteq M$ is chosen as in Remark 9′, part (2). Then $[A]=[M]$, $eA=eM$ and $\mathrm{p\,dim}_\Delta A<\infty$. In particular, A satisfies the hypothesis of Remark 9′, part (1).*

Proof. By the choice of A we have $A/JA\cong e(M/JM)$, and hence the first column of $[A]$ equals that of $[M]$. To show equality of the respective second columns, consider the canonical image $\bar{A}$ of A in $\bar{M}=M/\mathscr{F}(M)$. Start by noting that $eJ\bar{A}\cong eJA$, because $e(JA\cap\mathscr{F}(M))=0$. Thus, the entries of the second column

of $[A]$ are just the multiplicities of the S_i, $1 \leqq i \leqq m$, in $eJ\bar{A}$. Since moreover, $[M]=[\varepsilon(M)]$ by Lemma 8′, it suffices to show that $J\varepsilon(M)=eJ\bar{A}$.

For a proof of the latter equality we refer back to the second paragraph of the proof of Lemma 10 in Sect. 3. Keep in mind that $\varepsilon(M)$ and $\bar{A}$ are submodules of $\bar{M}$, and that the latter has Loewy length at most 2.

Thus $[A]=[M]$, which clearly entails $eA=eM$ and $\varepsilon(A)\cong\varepsilon(M)$. Lemma 7′ therefore yields $\mathrm{p\,dim}_A A<\infty$. □

Let $M\in L_2'(\Delta)$ with $\mathrm{p\,dim}_A M<\infty$. We call a submodule $A\subseteq M$ satisfying the conditions of Remark 9′, part (2), a *good* submodule of M. In view of Lemma 10′ and Remark 9′, part (1), we know that, given a good submodule A of M and any $K\in L_2'(\Delta)$ containing A as a submodule, there exists a projective cover $P\oplus Q\to K$ which restricts to a projective cover $P\to A$.

Bearing this in mind, we observe that an analogue of Proposition 11 of Section 3 – call it Proposition 11′ – remains true, where we allow for an arbitrary choice of M in $L_2'(\Delta)$ with $\mathrm{p\,dim}_A M<\infty$ in the hypothesis and keep all parts of the conclusion. Indeed, the old proof carries over verbatim if we take the following additional precautions: choose A_0 to be a good submodule of M and, for $j\geqq 1$, let A_j be a good submodule of $K_j=\ker(\alpha_j|_{P_j})$.

Corollary 12′ *If $M\in L_2'(\Delta)$ has finite projective dimension, then* $\mathrm{p\,dim}_{e\Lambda e} eM \leqq \mathrm{p\,dim}_A M$. □

Corollary 13′ *If $M\in L_2'(\Delta)$ has finite projective dimension, then there exists a non-negative integer a such that $(\varepsilon\Omega_\Delta)^a(\varepsilon(M))=0$ (the latter meaning that $L^a[M]=0$ in view of Lemma 8′).* □

We are now ready to proceed to the main result of this section. Let $G_0(\Delta)\subseteq \mathbf{Z}^{m\times 2}$ be the subgroup consisting of the matrices $A\in\mathbf{Z}^{m\times 2}$ which are annihilated by some power of L. To obtain the best possible bounds from our method, let $S_1,\ldots,S_p$ be those simple modules among $S_1,\ldots,S_m$ which occur in $Je/J^2e\oplus J\tilde{e}/J^2\tilde{e}$, and $S_1,\ldots,S_q$ those occurring in both $Je/J^2e\oplus J\tilde{e}/J^2\tilde{e}$ and in $J^2e/J^3e\oplus J^2\tilde{e}/J^3\tilde{e}$. Again, we set

$$G(\Delta)=G_0(\Delta)\cap\begin{pmatrix}\mathbf{Z} & \mathbf{Z}\\ \vdots & \vdots\\ \cdot & \mathbf{Z}\\ \cdot & 0\\ \vdots & \vdots\\ \mathbf{Z} & \cdot\\ 0 & \cdot\\ \vdots & \vdots\\ 0 & 0\end{pmatrix}$$

where the first column has p entries of $\mathbf{Z}$ and the second column has q.

Theorem 21 *Suppose that Λ is a basic left artinian ring and $\Gamma=\Gamma(\Lambda)$ contains a full subgraph Δ which is a sink in Γ and has the property that*

(*) *for each vertex e_i in Δ the corresponding summand J^3e_i of J^3 has only simple composition factors of finite projective dimension.*

If, moreover, the graph $\Gamma\backslash\mathrm{Arw}\,\Delta$, *obtained from* Γ *by deleting all arrows in* Δ, *has no oriented cycles, then*

$$\mathrm{fin\,dim}(\Delta)\leqq\mathrm{rank}\,G(\Delta)+d+l(\Gamma\backslash\mathrm{Arw}\,\Delta)+1,$$

and $\mathrm{rank}\,G(\Delta)\leqq p+q\leqq 2m$.

Proof. Suppose that $M\in\Delta$-mod has finite projective dimension. If $l=l(\Gamma\backslash\mathrm{Arw}\,\Delta)$, then, $\Gamma\backslash\mathrm{Arw}\,\Delta$ being cycle-free, the top of the l-th syzygy $\Omega^l_\Delta(M)$ is clearly a direct sum of copies of $S_1,\ldots,S_m,\widetilde{S}_{m+1},\ldots,\widetilde{S}_n$, plus, possibly, a direct sum of simple projective modules belonging to vertices in $\Gamma\backslash\Delta$. As we observed before, we then have $\Omega^{l+k}_\Delta(M)\in\Delta$-mod, and even $\Omega^{l+k}_\Delta(M)\in L'_2(\Delta)$ for all $k\geqq 1$; here we make use of both the fact that Δ is a sink and property (*).

By Corollary 13′ we infer that there exists an integer a with $L^a[\Omega^{l+1}_\Delta(M)]=0$, and hence the matrix $[\Omega^{l+1}_\Delta(M)]$ belongs to $G_0(\Delta)$. Using the argument in the proof of Theorem 2, we check that $[\Omega^1_\Delta(N)]\in G(\Delta)$ for any $N\in\Delta$-mod with $\mathrm{p\,dim}_\Delta N<\infty$. Since $\Omega^l_\Delta(M)$ is a direct sum of a module in Δ-mod and a projective module, we obtain, in particular, that $[\Omega^{l+1}_\Delta(M)]$ belongs to $G(\Delta)$. Clearly L induces a nilpotent endomorphism L^{**} of the subgroup $G^{**}(\Delta)$ of $G(\Delta)$ generated by all the matrices of the form $[\Omega^1_\Delta(N)]$ with $N\in\Delta$-mod and $\mathrm{p\,dim}_\Delta N<\infty$. Consequently there exists a number $c\leqq\mathrm{rank}\,G(\Delta)$ such that $(L^{**})^c=0$. Since $[\Omega^{l+1}_\Delta(M)]\in G^{**}(\Delta)$, we thus have $L^c[\Omega^{l+1}_\Delta(M)]=0$. By Lemmas 7′ and 8′ we deduce that $\mathrm{p\,dim}_\Delta(\Omega^{l+1}_\Delta(M))\leqq c+d$, and hence $\mathrm{p\,dim}_\Delta(M)\leqq c+d+l+1\leqq\mathrm{rank}\,G(\Delta)+d+l+1$ as claimed. □

Corollary 22 *If* Λ *is left artinian and all simple composition factors of the left* Λ*-module* J^3 *have finite projective dimension, then*

$$\mathrm{fin\,dim}(\Lambda)\leqq\mathrm{rank}\,G(\Gamma)+d+1,$$

where $\Gamma=\Gamma(\Lambda)$ *is the quiver of* Λ *and* d *is the supremum of the projective dimensions of those simple left* Λ*-modules which have finite projective dimension. Moreover,* $\mathrm{rank}\,G(\Gamma)\leqq p+q$, *where* p *is the number of isomorphisms types of simple left modules of infinite projective dimension which occur in* J/J^2, *and* q *is the number of those occurring simultaneously in* J/J^2 *and in* J^2/J^3. □

In [4, Coroll. to Theor. A] the Generalized Nakayama Conjecture, a strenghtening of the classical Nakayama Conjecture, was confirmed for Artin algebras with $J^3=0$. The latter conjecture asserts that any finite dimensional algebra of finite dominant dimension is quasi-Frobenius. Since, as was pointed out in [12, p. 105], the finitistic dimension conjecture is also stronger than the classical Nakayama Conjecture, the results of this article yield at least the latter in somewhat higher generality.

Recall that Λ is said to have *infinite left dominant dimension* if there is an exact sequence

$$0\to\Lambda\to Q_1\to Q_2\to Q_3\to\ldots$$

of left Λ-modules such that the Q_i are both injective and projective. By [7, Coroll. 1 and 9, Theor. 10], this condition is left-right symmetric provided that Λ is two-sided artinian, and we will in that case refer to it without mention of the side.

Theorem 23 *Suppose that* Λ *is two-sided artinian and has at least one of the following additional properties*:

(1) *J contains a two-sided ideal I with $J^3 \subseteq I$ such that I has finite projective dimension either as a left Λ-module or as a right Λ-module. (This is, of course, satisfied if J^3 has finite projective dimension on some side).*
(2) *Λ satisfies the hypothesis of Theorem 21.*
Then the Nakayama Conjecture holds for Λ, that is, infinite dominant dimension forces Λ to be quasi-Frobenius.

Proof. From either condition (1) or (2) we obtain that Λ has finite finitistic dimension on at least one side (Coroll. 17, resp., Theor. 21); let us assume that this is the case on the left. Assuming moreover that Λ has infinite dominant dimension, all we need to show is that Λ is left self-injective.

For the conveniene of the reader, we complete the argument along the line suggested in [12, p. 105]: Let $\operatorname{fin\,dim}(\Lambda) = a$. Since Λ has infinite dominant dimension, there clearly exists a minimal injective resolution

$$0 \longrightarrow \Lambda \longrightarrow Q_1^0 \xrightarrow{f_1} Q_2^0 \xrightarrow{f_2} Q_3^0 \xrightarrow{f_3} \dots$$

whose terms are injective and projective left Λ-modules. Now projectivity guarantees that the Q_i^0 are all finitely generated. Moreover, consulting the above exact sequence, we see that $\operatorname{Im}(f_a)$ has finite projective dimension, $0 \to \Lambda \to Q_1^0 \xrightarrow{f_1} \dots \xrightarrow{f_a} Q_{a+1}^0 \to \operatorname{Im}(f_a) \to 0$ being a projective resolution. From $\operatorname{p\,dim}_\Lambda \operatorname{Im}(f_a) \leqq a$ we deduce that $\operatorname{Im}(f_1)$ is a projective module, and hence that $\operatorname{Ker}(f_1) \cong \Lambda$ is a direct summand of Q_1^0. Thus Λ is injective as a left Λ-module as required. □

We conclude with an example of dual purpose: The first is to demonstrate the sharpness of the bound on $\operatorname{fin\,dim}(\Lambda)$ provided by Theorem 21; the second is to illustrate the techniques leading up to this Theorem with an example where both J^3 and J^4 have infinite projective dimension.

Example 24 The following path algebra modulo relations, $\Lambda = K\Gamma/\langle\rho\rangle$, has the property that $\operatorname{p\,dim}_\Lambda J^3 = \operatorname{p\,dim}_\Lambda J^4 = \infty$, where J^3 and J^4 are viewed as left Λ-modules.

Let K be any field and Γ the quiver shown below.

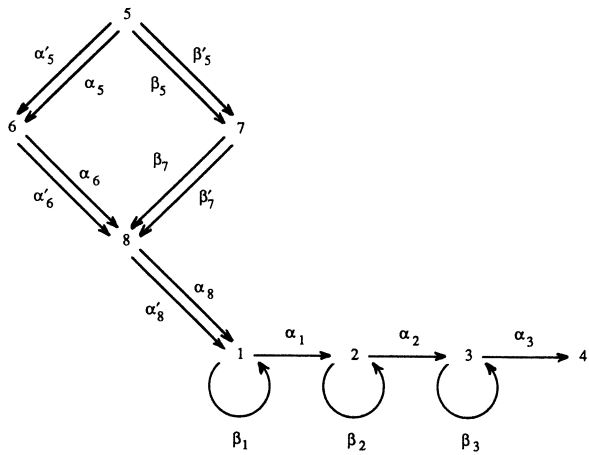


Fig. 6

Moreover, let ρ' be the set of all paths of the form $\gamma\delta$, where γ, δ are arrows, *at least* one of which is "primed", and set $\rho=(\rho'\setminus\{\beta_1\alpha_8'\})\cup\{\alpha_6\alpha_5-\beta_7\beta_5, \alpha_1\alpha_8, \alpha_2\alpha_1, \alpha_3\alpha_2, \beta_i^2, \alpha_i\beta_i, 1\leqq i\leqq 3\}$. Clearly, $J^3\cong S_1\oplus S_1\oplus S_1$ and $J^4\cong S_1$ have infinite projective dimension. Choosing Δ to be the full subgraph with vertex set $\{1, 2, 3, 4\}$, we obtain rank $G(\Delta)=2$ by Example 20, $d=0$ and $l(\Gamma\setminus\mathrm{Arw}\,\Delta)=3$; hence fin dim$(\Lambda)\leqq 6$ by Theorem 21. To ascertain that fin dim$(\Lambda)=6$, start by noting that p dim$_\Lambda Je_8=3$, which means that p dim$_\Lambda S_8=4$. Moreover, check that $\Omega_\Lambda^2(\Lambda e_5/\Lambda\alpha_5)\cong\Omega_\Lambda^1(\Lambda e_6/\Lambda\alpha_6')\cong\Lambda\alpha_6'\cong S_8$ to obtain that p dim$_\Lambda(\Lambda e_5/\Lambda\alpha_5)=2+4=6$. □

Acknowledgements. A major part of this work was done while both authors were visitors at the University of Washington; both wish to express their gratitude toward their host, R.B. Warfield, Jr. The authors also want to thank E. Kirkman for a helpful hint which led to Corollary 17 of this article.

References

1. Auslander, M., Buchsbaum, D.: Homological dimension in noetherian rings II. Trans. Am. Math. Soc. **88**, 194–206 (1958)
2. Bass, H.: Finitistic dimension and a homological generalization of semiprimary rings. Trans. Am. Math. Soc. **95**, 466–488 (1960)
3. Bass, H.: Injective dimension in noetherian rings. Trans. Am. Math. Soc. **102**, 18–29 (1962)
4. Fuller, K.R., Zimmermann-Huisgen, B.: On the Generalized Nakayama Conjecture and the Cartan determinant problem. Trans. Am. Math. Soc. **294**, 679–691 (1986)
5. Gordon, R., Green, E.L.: Modules with cores and amalgamations of indecomposable modules. Mem. Am. Math. Soc. **187**, 1–145 (1977)
6. Green, E.L., Kirkman, E., Kuzmanovich, J.: Finitistic dimension of finite dimensional monomial algebras. J. Algebra **136**, 37–50 (1991)
7. Harada, M.: QF-3 and semi-primary PP-rings I. Osaka J. Math. **2**, 357–368 (1965)
8. Mochizuki, H.: Finitistic global dimension for rings. Pac. J. Math. **15**, 249–258 (1965)
9. Mueller, B.J.: Dominant dimension of semi-primary rings. J. Reine Angew. Math. **232**, 173–179 (1968)
10. Raynaud, M., Gruson, L.: Critères de platitude et de projectivé. Invent. Math. **13**, 1–89 (1971)
11. Small, L.W.: A change of rings theorem. Proc. Am. Math. Soc. **19**, 662–666 (1968)
12. Tachikawa, H.: Quasi-Frobenius rings and their generalizations. (Lect. Notes Math., vol. 351), Berlin Heidelberg New York: Springer 1973
13. Zimmermann-Huisgen, B.: Bounds on finitistic and global dimension of Artinian rings with vanishing radical cube (preprint)

Math. Z. 206, 527–532 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On rings of semialgebraic functions*

J.M. Gamboa [1] **and J.M. Ruiz** [2]

[1] Depto. Algebra, F.C. Matemáticas, U. Complutense, E-28040 Madrid, Spain
[2] Depto. Geometría & Topología, F.C. Matemáticas, U. Complutense, E-28040 Madrid, Spain

Received June 7, 1989; in final form December 28, 1989

1 Introduction

Let X be a constructible subset of the real spectrum of a ring A, and let $\mathscr{C}(X)$ denote the ring of abstract semialgebraic functions on X. These are the suitable objects generalizing a semialgebraic subset M of R^n, R a real closed field, and its ring $\mathscr{S}(M)$ of real continuous functions with semialgebraic graph (cf. [B-C-R, Ch. 2] and [Br1]). In this note we compare the topological dimension of X and the Krull dimension of $\mathscr{C}(X)$ and study the finitely generated prime ideals of $\mathscr{C}(X)$. For the real spectrum we refer to [B-C-R, Ch. 7], and for the abstract semialgebraic functions, to [S, Ch. I] and [Br2].

For a locally closed semialgebraic subset M of R^n, the Krull dimension $\dim \mathscr{S}(M)$ of the ring $\mathscr{S}(M)$ coincides with the *topological* dimension $\dim(M)$ of M, i.e. with the maximal length of a specialization chain of M, [C-C]. This extends readily to the abstract setting: If X is locally closed, it is homeomorphic to the real spectrum Y of $\mathscr{C}(X)$, and since Y is homeomorphic to the Zariski spectrum of $\mathscr{C}(X)$, [S, Ch. I § 3], we have

$$\dim(X)=\dim(Y)=\dim \mathscr{C}(X)$$

This argument breaks down when X is not locally closed. Then X embeds into Y as a proper subspace and we only get

$$\dim(X)\leqq\dim \mathscr{C}(X).$$

This inequality can also be obtained with the prime ideals

$$\mathfrak{p}(x)=\{f\in\mathscr{C}(X): f(x)=0\},\quad x\in X,$$

because

(i) If $x_1\neq x_2$, then $\mathfrak{p}(x_1)\neq\mathfrak{p}(x_2)$ ($f(x_i)\leqq 0$ and $f(x_j)>0$ for some f, and $f+|f|\in\mathfrak{p}(x_i)\setminus\mathfrak{p}(x_j)$), and

* Partially supported by DCICYT, PB89-0379-C02-02

(ii) If x_1 specializes to x_2, denoted by $x_1 \to x_2$, then $\mathfrak{p}(x_1) \subset \mathfrak{p}(x_2)$ ($f(x_1)=0$ implies $f(x_2)=0$ by continuity).

In Section 2 we shall prove the converse inequality for an arbitrary X, when A is an *excellent* ring. Thus:

Theorem 1.1 *Let X be a constructible subset of the real spectrum of an excellent ring. Then* $\dim \mathscr{C}(X) = \dim(X)$. □

The proof of the theorem above uses the general theory of excellent rings, for which we refer the reader to [M, Ch. 13], and the good behaviour of closures in the real spectrum of an excellent ring (cf. [A-B-Rz], [Rz]).

By 1.1, chains of prime ideals of $\mathscr{C}(X)$ have bounded length (when $\dim X < +\infty$, otherwise 1.1 is trivial). However, the ring $\mathscr{C}(X)$ is noetherian very rarely. For, suppose it is: Every $h \in \mathscr{C}(X)$ gives the chain

$$(h) \subset (h^{1/3}) \subset \ldots \subset (h^{1/k}) \subset (h^{1/(k+2)}) \subset \ldots, \qquad k \text{ odd},$$

and, as $\mathscr{C}(X)$ is noetherian, $h^{1/(k+2)} = g h^{1/k}$ for some k, g. This can be rewritten $h^k(1 - g^{k(k+2)} h^2) = 0$, and so h is either a zero divisor or a unit. This means that $\mathscr{C}(X)$ has no other prime ideals than the associated primes of (0), and, $\mathscr{C}(X)$ being reduced, there are no inclusions among them. From this and (i), (ii) above, it follows that X consists of finitely many closed points. We have proved the non-trivial part of

Proposition 1.2 *The ring $\mathscr{C}(X)$ is noetherian if and only if X is discrete. In that case, $\mathscr{C}(X)$ is a finite product of fields.* □

In general, almost no prime ideal of $\mathscr{C}(X)$ is finitely generated. To make this more precise, consider a point $x \in X$, whose closure Z in X is open. Then Z is an open and closed constructible subset of X, and $\mathscr{C}(X) = \mathscr{C}(X \setminus Z) \times \mathscr{C}(Z)$. Clearly

$$\mathfrak{p} = \mathfrak{p}(x) = \text{ideal of } Z = \mathscr{C}(X \setminus Z) \times \{0\}$$

is finitely generated and prime. We shall see in Section 3 that $\mathscr{C}(X)$ has no other finitely generated prime ideals. If, furthermore, we deal with excellent rings, the conclusion is far more restrictive. Indeed, also in Section 3, we shall prove

Theorem 1.3 *Let X be a constructible subset of the real spectrum of an excellent ring. Then the finitely generated prime ideals of $\mathscr{C}(X)$ are the ideals of the open constructible points of X whose closure in X is open of dimension $\leqq 1$.* □

This implies that these points have at most one specialization in X. Actually, the result is stronger and they have at most one specialization even out of X. The essential reason behind is the richness of prime cones typical of excellent rings (cf. [Rz]).

Now, let us apply these results to the geometric case. This is possible because finitely generated algebras over fields are excellent rings. Then all assertions for constructible sets can be reformulated for semialgebraic sets using the tilda operation (cf. [B-C-R, Ch. 7], [S, Ch. III]). Furthermore, this produces some simplifications, since in this context "discrete" is equivalent to "finite", and constructible points are always closed. At the end, we obtain

Proposition 1.4 *Let M be a semialgebraic subset of R^n, R a real closed field, and let $\mathscr{S}(M)$ denote the ring of real continuous functions with semialgebraic graph. Then*

(a) $\dim \mathscr{S}(M) = \dim(M)$.

(b) *$\mathscr{S}(M)$ is noetherian if and only if M is finite.*

(c) *The finitely generated prime ideals of $\mathscr{S}(M)$ are the maximal ideals of the isolated points of M.* □

2 Proof of theorem 1.1

Let A be an excellent ring and X a constructible subset of the real spectrum $\mathrm{Spec}_r(A)$. Denote by $\mathscr{C}(X)$ the ring of abstract semialgebraic functions on X. We are going to prove that $\dim \mathscr{C}(X) \leqq \dim(X)$.

Consider a chain $\mathfrak{p}_0 \subsetneqq \mathfrak{p}_1 \subsetneqq \ldots \subsetneqq \mathfrak{p}_r$ of prime ideals of $\mathscr{C}(X)$. To show that $r \leqq \dim(X)$, pick functions $g_i \in \mathfrak{p}_i \setminus \mathfrak{p}_{i-1}$ for $i = 1, \ldots, r$. The following construction appears in [Br2]. The g_i's give a section $\varphi = (g_1, \ldots, g_r)$ over X of the canonical mapping $\pi\colon \mathrm{Spec}_r(A[t_1, \ldots, t_r]) \to \mathrm{Spec}_r(A)$, such that

(i) $\Gamma = \varphi(X)$ is a constructible subset of $\mathrm{Spec}_r(A[t_1, \ldots, t_r])$,

(ii) $\varphi\colon X \to \Gamma$ is a homeomorphism with inverse $\pi|\Gamma$, and

(iii) The condition $g(x) = h(\varphi(x))$, $x \in X$, defines a ring isomorphism $\varphi_*\colon \mathscr{C}(X) \to \mathscr{C}(\Gamma)\colon g \to h$, (cf. [Br2, 4.3, 4.4 and 4.5]) in other words, φ_* is induced by the canonical mapping $\pi|\Gamma\colon \Gamma \to X$.

Now, we apply the isomorphism φ_* of (iii) to the data $\{\mathfrak{p}_i, g_i\}$ given over X, to obtain similar ones $\{\mathfrak{q}_i, h_i\}$ over Γ. Furthermore, by the very definition of φ

$$t_i(\varphi(x)) = t_i(g_1(x), \ldots, g_r(x)) = g_i(x),$$

and so, $h_i = \varphi_*(g_i) = t_i|\Gamma$ and $h_i \in A[t_1, \ldots, t_r]$. Finally, by (i) and (ii), Γ is constructible with the same dimension as X. Whence, we are reduced to see $r \leqq \dim(\Gamma)$.

Consequently, we may assume $g_1, \ldots, g_r$ are the restrictions to X of elements of A.

We introduce a pro-constructible subset Y of $\mathrm{Spec}_r(A)$: Y is the intersection of all locally closed constructible subsets of $\mathrm{Spec}_r(A)$ which contain X. Although Y need not be constructible, the theory of abstract semialgebraic functions works the same (cf. [S, Ch. I]), and so we have the ring $\mathscr{C}(Y)$. Furthermore, $\dim \mathscr{C}(Y) = \dim(Y)$, repeating the proof given in the introduction for locally closed constructible sets, which is possible by [S, I.3.20]. Now, since $X \subset Y$, restriction gives a homomorphism $\mathscr{C}(Y) \to \mathscr{C}(X)$, and our initial chain $\mathfrak{p}_0 \subsetneqq \mathfrak{p}_1 \subsetneqq \ldots \subsetneqq \mathfrak{p}_r$ in $\mathscr{C}(X)$ lies over another one $\mathfrak{q}_0 \subset \mathfrak{q}_1 \subset \ldots \subset \mathfrak{q}_r$ in $\mathscr{C}(Y)$. Finally, since g_i is the restriction to X of an element of A, we can suppose $g_i \in \mathscr{C}(Y)$, and, of course, $g_i \in \mathfrak{q}_i \setminus \mathfrak{q}_{i-1}$. Whence $\mathfrak{q}_0 \subsetneqq \mathfrak{q}_1 \subsetneqq \ldots \subsetneqq \mathfrak{q}_r$, and $r \leqq \dim \mathscr{C}(Y) = \dim(Y)$.

So, we shall finish by showing $\dim(Y) = \dim(X)$.

Firstly, note that the closure $\mathrm{Cl}(X)$ of X is locally closed and constructible (because A is excellent, [A-B-Rz, 3.1]). Hence $X \subset Y \subset \mathrm{Cl}(X)$. In order to prove the above equality of dimensions we need the

Lemma 2.1 *Let $x_0 \to \ldots \to x_n \to x_{n+1}$ be a specialization chain in $\mathrm{Cl}(X)$. Then there is another one $x'_0 \to \ldots \to x'_n \to x_{n+1}$, with $x'_i \in X$.*

Proof. As $x_0 \in \mathrm{Cl}(X)$, there is $y \in X$ with $y \to x_0$, so $y \to x_{n+1}$ and $ht(\mathrm{supp}(x_{n+1})/\mathrm{supp}(y)) \geqq n+1$. Choose elements $f_1, \ldots, f_m, g \in A$ with

$$y \in \{x \in \mathrm{Spec}_r(A) : f_1(x) > 0, \ldots, f_m(x) > 0, g(x) = 0\} \subset X.$$

Then, by [Rz, 1.3], there is a chain $x'_0 \to \ldots \to x'_n \to x_{n+1}$ with

$$\mathrm{supp}(x'_0) = \mathrm{supp}(y), \quad f_1(x'_n) > 0, \ldots, f_m(x'_n) > 0;$$

these conditions imply $x'_0, \ldots, x'_n \in X$, and we are done. □

Now put $d = \dim(X)$. Immediately from 2.1 we deduce

$$d \leqq \dim(Y) \leqq \dim(\mathrm{Cl}(X)) \leqq d+1,$$

and so, if $\dim(Y) \neq \dim(X)$,

$$\dim(Y) = \dim(\mathrm{Cl}(X)) = d+1.$$

Let us see that this is impossible. Suppose Y contains a chain $x_0 \to \ldots \to x_d \to x_{d+1}$. If x_{d+1} were not closed, there would be another chain $x_{d+1} \to x_{d+2}$ and connecting both chains, $d+1 < \dim(\mathrm{Cl}(X))$. Hence, x_{d+1} is closed and we can write $\{x_{d+1}\} = \cap F$, where the intersection includes all closed constructible sets F with $x_{d+1} \in F$. On the other hand, $x_{d+1} \notin X$, because $x_{d+1} \in X$ would give $x'_0 \to \ldots \to x'_d \to x_{d+1}$, a chain too long in X (2.1 again). Consequently, $X \cap \cap F$ is empty, and by the compactness of the constructible topology, $X \cap F$ is empty for some F. But then X is contained in the complement U of F, which is locally closed and constructible. Therefore $Y \subset U$ and $x_{d+1} \notin U$, contradiction. □

Remark. The long argument above with the pro-constructible set Y is necessary because X may well have topological dimension strictly smaller than the one of its closure.

3 Proof of Theorem 1.3

Let X be a constructible subset of the real spectrum of a ring A, and denote by $\mathscr{C}(X)$ the ring of abstract semialgebraic functions on X. We start with the

Lemma 3.1 *If* $\mathfrak{p}$ *is a finitely generated prime ideal of* $\mathscr{C}(X)$, *then* $\mathfrak{p}$ *is the ideal of an open and closed specialization chain of* X.

Proof. It is clear that the zero set $Z \subset X$ of $\mathfrak{p}$ is closed, and by [S, I.3.14], there is $x \in X$ with $Z = \overline{\{x\}} \cap X$. Assume $h_1, \ldots, h_r \in \mathscr{C}(X)$ generate $\mathfrak{p}$ and set $h = h_1^2 + \ldots + h_r^2 \in \mathfrak{p}$. Then $g = h^{1/3} \in \mathscr{C}(X)$, and $\mathfrak{p}$ being prime, g belongs to $\mathfrak{p}$. We claim g generates $\mathfrak{p}$.

Indeed, we have

$$0 \leqq |h_i^3(u)/h(u)| = |h_i(u)[h_i^2(u)/h(u)]| \leqq |h_i(u)| \quad \text{if } h(u) \neq 0.$$

Using the remark after I.3.7 in [S], we can define an abstract semialgebraic function g_i on X by

$$g_i(u)=h_i^3(u)/h(u) \quad \text{if } h(u)\neq 0,$$
$$g_i(u)=0 \quad \text{otherwise.}$$

So, $f_i=g_i^{1/3}\in\mathscr{C}(X)$ and $h_i=f_i g$. We are done.

Next, Z is open in X. For, were it not, there would exist $z\in Z$, adherent to $X\setminus Z$. As $X\setminus Z=\{g\neq 0\}$ is constructible, some $y\in X\setminus Z$ specializes to z. On the other hand, also $g^{1/3}\in\mathfrak{p}$, i.e. $g^{1/3}=fg$ for some $f\in\mathscr{C}(X)$, and this gives $g(1-f^3g^2)=0$. As $y\notin Z$, $g(y)\neq 0$ and so $1-f^3(y)g^2(y)=0$, or in other words, $1-f^3g^2$ vanishes at y. But $y\to z$, and continuity implies $1=f^3(z)g^2(z)=0$. This contradiction shows Z must be open in X.

It remains to show that $\mathfrak{p}=\mathfrak{p}(x)$ or, equivalently, that any $f\in\mathscr{C}(X)$ vanishing on Z belongs to $\mathfrak{p}$. But Z being open, closed and constructible, the quotient f/g, which is defined on $X\setminus Z$, can be extended by zero to the whole X. Hence, g divides f, and f belongs to $\mathfrak{p}$. □

After this lemma, we can complete the proof of 1.3.

Assume A is excellent, and let $\mathfrak{p}$ be a finitely generated prime ideal of $\mathscr{C}(X)$. By Lemma 3.1, $\mathfrak{p}$ is the zero ideal of a specialization chain $Z=\overline{\{x\}}\cap X$, $x\in X$, which is open and closed in X. In particular, Z is constructible, and since A is excellent, the closure of Z in $\mathrm{Spec}_r(A)$ is constructible too ([A-B-Rz, 3.1]): This closure is $\overline{\{x\}}$. Now, let y denote the unique closed specialization of x in $\mathrm{Spec}_r(A)$, and consider the local excellent ring

$$B=A_{\mathfrak{q}}/\mathfrak{p}_A\cdot A_{\mathfrak{q}},\ \mathfrak{p}_A=\mathrm{supp}(x)\subset\mathrm{supp}(y)=\mathfrak{q}.$$

We claim that the Krull dimension of B is $\leqq 1$, which implies both x is open constructible and $\dim\overline{\{x\}}\leqq 1$.

By way of contradiction, suppose $\dim(B)\geqq 2$ and choose $\delta\in B$, $\delta\neq 0$, such that the localization B_δ is a regular ring. Now, the closure $\overline{\{x\}}$ is constructible, i.e. there are $f_1,\ldots,f_m, g\in A$ with

$$x\in\{z\in\mathrm{Spec}_r(A): f_1(z)>0,\ldots,f_m(z)>0, g(z)=0\}\subset\overline{\{x\}},$$

and we still denote by $f_1,\ldots,f_m$ the classes $\bmod\mathfrak{p}_A$ of $f_1,\ldots,f_m$. Summing up, x is a total ordering in B which makes $f_1,\ldots,f_m,\delta^2$ positive; x specializes to y, and y is a prime cone of B with support the maximal ideal. By [Rz, 1.3], there is a chain $x\to z\to y$ such that

$$f_1(z)>0,\ldots,f_m(z)>0,\ \delta^2(z)>0,$$

and, if $\mathfrak{n}$ is the support of z in B, $\dim(B_{\mathfrak{n}})=1$.

Now, as $\delta(z)\neq 0$, $\delta\notin\mathfrak{n}$ and $B_{\mathfrak{n}}$ is regular. Pick a uniformizer u of $B_{\mathfrak{n}}$. We know that any total ordering in $qf(B/\mathfrak{n})$ is the specialization of two different total orderings in $B_{\mathfrak{n}}$, determined by $u>0$ or $u<0$. Consequently, there are two different total orderings x' and x'' in B which specialize to z. Since $f_i(z)>0$, this means that $f_i(x')>0$, $f_i(x'')>0$. Furthermore, $g\in\mathrm{supp}(x)=\mathfrak{p}_A$, which is also the support of x' and x'' when considered as prime cones of A. In conclusion

x' and x'' belong to the set $\{f_1>0, \ldots, f_m>0, g=0\}$, contained in $\overline{\{x\}}$, which is impossible because $\overline{\{x\}}$ is totally ordered by inclusion. □

Acknowledgement. The authors wish to thank Professors C. Andradas and M. Coste for several helpful discussions during the preparation of this work.

References

[A-B-Rz] Andradas, C., Bröcker, L., Ruiz, J.M.: Minimal generation of basic open semianalytic sets. Invent. Math. **92**, 409–430 (1988)

[B-C-R] Bochnak, J., Coste, M., Roy, M.F.: Géométrie Algébrique Réelle. Ergeb. Math. Grenzgeb. **12**, Berlin Heidelberg New York: Springer 1987

[Br1] Brumfiel, G.W.: Witt rings and K-theory. Rocky Mt. J. Math. **14**, 733–765 (1984)

[Br2] Brumfiel, G.W.: The real spectrum of an ideal and KO-theory exact sequences. K-Theory **1**, 211–235 (1987)

[C-C] Carral, M., Coste, M.: Normal spectral spaces and their dimensions. J. Pure Appl. Algebra **30**, 227–235 (1983)

[M] Matsumura, H.: Commutative Algebra, (2. edition). Amsterdam: W.A. Benjamin Co. 1980

[Rz] Ruiz, J.M.: Cônes locaux et complétions. C.R. Acad. Sci., Paris, Sér. I **302**, 67–69 (1986)

[S] Schwartz, N.: The basic theory of real closed spaces. Mem. A.M.S. **397** (1989)

Math. Z. 206, 533–534 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

On the decomposition map of Grothendieck groups*

Xiaoming Zhou
School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA

Received August 2, 1989; in final form January 2, 1990

1 Introduction and statement

Our notation and terminology are standard and tend to follow [2]. All modules over an algebra are finitely generated, left and unital.

Throughout this paper, G denotes a finite group, H denotes a subgroup of G. Let A be a complete discrete valuation ring with quotient field K and residue field $k=A/(\pi)$, where (π) is the unique maximal ideal of A. We assume that $\operatorname{char} K=0$ and $\operatorname{char} k=p>0$.

Let $R_K(G)$ and $R_k(G)$ be the Grothendieck groups of the categories of finitely generated $K[G]$-modules and $k[G]$-modules, respectively. For each $K[G]$ (or $k[G]$) module V, $[V]$ denotes the representative of V in $R_K[G]$ (or $R_k[G]$).

Let $P_A(G|H)$ denote the subgroup of $R_K(G)$ generated by all $[K\otimes V]$, where V is an A-free $A[G]$-module and is H-projective and let $P_k(G|H)$ denote the subgroup of $R_k(G)$ generated by all $[V]$, where V is an H-projective $k[G]$-module.

Let $d: R_K(G)\to R_k(G)$

$$[V]\to[\bar{W}]$$

be the decomposition map, where W is a full AG-lattice of V, and $\bar{W}=V/\Pi V$. Clearly $d(P_A(G|H))\subseteq P_k(G|H)$. Let d_H: $P_A(G|H)\to P_k(G|H)$ be the restriction of d to $P_A(G|H)$.

The map d is surjective by [2, Theor. 33]. The map d_H is studied by M. Broué in [1]. He proved that d_H is surjective if either H is a normal p-subgroup of G or k is sufficiently large for G ([1, Theor. 2]).

This paper gives a complete answer to this natural question of the surjectivity of d_H. We will prove:

Theorem. *d_H is surjective.*

Our proof of our result uses Theorem 2 of [1].

2 Proof of the theorem

Let Q be a Sylow p-subgroup of H, then every H-projective $A[G]$-module (or $k[G]$-module) is Q-projective and conversely. Hence we have $P_A(G|H)=P_A(G|Q)$ and $P_k(G|H)=P_k(G|Q)$; thus we may assume that H is a p-subgroup of G.

* This article is a part of the doctoral dissertation of the author at the University of Minnesota under the guidance of Professor Morton E. Harris. The author wishes to express his gratitude to Professor Harris for his advice and suggestions

We prove the theorem by induction on $|H|$. The case that $H=1$ is a well-known result, since every $\{1\}$-projective $k[G]$-module is projective and can be lifted to a projective $A[G]$-module ([3, 18.1(iii)]).

Let V be an indecomposable H-projective $k[G]$-module. We shall show that there is an $x\in P_A(G|H)$ such that $d(x)=[V]$.

By the induction hypothesis, we may assume that H is the vertex of V since H contains a vertex of V.

Let $L=N_G(H)$. By [3, Theor. 20.6], there is an indecomposable $k[L]$-module U with vertex H such that

$$\mathrm{Ind}_L^G(U)=V\oplus X$$

for some $k[G]$-module X such that $X=\bigoplus_{i=1}^{s}X_i$ where the vertex of X_i $\leqq {}^G u_i H u_i^{-1}\cap H$, $u_i\notin L$, $1\leqq i\leqq s$.

Since $u_i H u_i^{-1}\cap H<H$ for $u_i\notin L$, by the induction hypothesis, there are $x_i\in P_A(G|H)$ such that $d(x_i)=[X_i]$, $1\leqq i\leqq s$.

Since U is H-projective and since $H\trianglelefteq L$, by [1, Theor. 2], there is $y\in P_A(L|H)$ such that $d(y)=[U]$.

We have $y=\sum_{j=1}^{t} a_j[K\otimes U_j]$, where U_j's are A-free $A[L]$-modules and are H-projective and $a_j\in\mathbb{Z}$, for $1\leqq j\leqq t$. Hence

$$\mathrm{Ind}_L^G(y)=\sum_{j=1}^{t} a_j[K\otimes \mathrm{Ind}_L^G U_j],$$

where $\mathrm{Ind}_L^G(U_j)$'s are A-free H-projective $A[G]$-modules, so that $\mathrm{Ind}_L^G(y)\in P_A(G|H)$.

By [3, 21 B], $\mathrm{Ind}\circ d=d\circ\mathrm{Ind}$, we have

$$d(\mathrm{Ind}_L^G(y))=\mathrm{Ind}_L^G(d(y))=\mathrm{Ind}_L^G([U])=[\mathrm{Ind}_L^G(U)],$$

therefore

$$\begin{aligned}[V]&=[\mathrm{Ind}_L^G(U)-[X]]\\&=d(\mathrm{Ind}_L^G(y))-\sum_{i=1}^{s} d(x_i)\\&=d\left(\mathrm{Ind}_L^G(y)-\sum_{i=1}^{s} x_i\right).\end{aligned}$$

Since $\mathrm{Ind}_L^G(y)-\sum_{i=1}^{s} x_i\in P_A(G|H)$, the proof is complete.

References

1. Broué, M.: Sur l'induction des modules indécomposables et la projectivité relative. Math. Z. **149**, 227–245 (1976)
2. Serre, J.P.: Linear Representations of Finite Groups. Berlin Heidelberg New York: Springer 1977
3. Curtis, C.W., Reiner, I.: I. Methods of Representation Theory with Applications to Finite Groups and Orders. New York: John Wiley 1981

Math. Z. 206, 535–550 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On a geometric property of the normal bundle of surfaces in $\mathbb{P}_4$

R. Braun *

Mathematisches Institut, Universität Bayreuth, Postfach 101251, Universitätsstrasse 30, W-8580 Bayreuth, Federal Republic of Germany

Received March 15, 1990; in final form June 11, 1990

0 Introduction

Let X be a smooth k-dimensional variety in complex projective space $\mathbb{P}_n$. From the restricted Euler sequence

$$0 \to \mathcal{O}_X \to \bigoplus_{i=1}^{n+1} \mathcal{O}_X(1) \to \mathcal{T}_{\mathbb{P}_n}|_X \to 0 \tag{0.1}$$

and the defining sequence of the normal bundle $\mathcal{N} := \mathcal{N}_{X|\mathbb{P}_n}$ of X in $\mathbb{P}_n$

$$0 \to \mathcal{T}_X \to \mathcal{T}_{\mathbb{P}_n}|_X \to \mathcal{N} \to 0 \tag{0.2}$$

we obtain a quotient

$$\bigoplus_{i=1}^{n+1} \mathcal{O}_X \to \mathcal{N}(-1) \to 0. \tag{0.3}$$

Hence the (-1)-twist of the normal bundle is generated by global sections and the normal bundle itself is ample. This leads to the following problem:

0.4 Problem. For which smooth varieties X in $\mathbb{P}_n$ is $\mathcal{N}(-1)$ ample?

In this paper we give a geometric answer to this question (see Prop. 1.2) and apply it to smooth surfaces in $\mathbb{P}_4$.

The main result concerns rational non-special surfaces in $\mathbb{P}_4$ whose classification was completed recently by Okonek and Alexander [O1, O2, A] (see Table 2.2). In Sect. 2 we shall prove the following theorem:

2.4 Theorem. *Let S be a smooth, nonspecial, non-degenerate and rational surface in $\mathbb{P}_4$.*

(1) If S is rigid then $\mathcal{N}(-1)$ is not ample.

(2) If S is not rigid then S can be deformed to a surface S' of the same type such that $\mathcal{N}_{S'|\mathbb{P}_4}(-1)$ is ample.

* *Present address:* Department of Mathematics, University of California, 405 Hilgard Avenue, Los Angeles, CA-90024-1555, USA

We call two non-special rational surfaces of the same type if the degree and the genus of a hyperplane section are the same. S is said to be rigid if all surfaces of the same type are projectively equivalent.

Moreover we show that abelian surfaces in $\mathbb{P}_4$ always have ample $\mathcal{N}(-1)$ (see Prop. 1.7) and that the non-ampleness of $\mathcal{N}(-1)$ is not bounded by the degree, i.e. there exist surfaces of arbitrary high degree where $\mathcal{N}(-1)$ is not ample (see Prop. 1.9).

This paper has grown out of a part of the author's dissertation [B] written at the University of Bayreuth. We refer to [B] for a more detailed exposition and further examples, e.g. rational ruled surfaces and K3-surfaces in higher-dimensional projective spaces.

1 The geometric criterion and first examples

In this paper we shall use Hartshorne's definition of ampleness, i.e. we call a vector bundle $\mathcal{E}$ ample if the tautological line bundle $\mathcal{O}_{\mathbb{P}(\mathcal{E})}(1)$ is ample.

Let X be a smooth k-dimensional variety in a complex projective space $\mathbb{P}_n$ $(1 \leqq k \leqq n)$ with normal bundle $\mathcal{N} := \mathcal{N}_{X|\mathbb{P}_n}$. We ask whether $\mathcal{N}(-1)$ is ample or not. There are the following two examples which are in some sense extremal:

1.1 *Examples.* (1) If X is degenerate, i.e contained in some hyperplane $H \simeq \mathbb{P}_{n-1}$ then $\mathcal{N}(-1)$ is not ample since

$$\mathcal{N} \simeq \mathcal{N}_{X|H} \oplus \mathcal{O}_X(1).$$

(2) If X is a non-degenerate complete intersection of hypersurfaces of degrees $d_1, \ldots, d_{n-k}$ then $\mathcal{N}(-1)$ is ample since $d_i \geqq 2$ for all i by assumption and

$$\mathcal{N} \simeq \bigoplus_{i=1}^{n-k} \mathcal{O}_X(d_i).$$

The next proposition gives a geometric answer to the above question.

1.2 Proposition. *The following conditions are equivalent:*

(1) $\mathcal{N}(-1)$ is not ample.

(2) There is a hyperplane H_0 in $\mathbb{P}_n$ which is tangent to X along a reduced and irreducible curve C. In this case $\mathcal{N}(-1)|_C$ is not ample.

Proof. Let $\mathbb{P}_n^\vee$ be the dual projective space of hyperplanes in $\mathbb{P}_n$ and let $C_X := \{(x, H) \in X \times \mathbb{P}_n^\vee : H \text{ is tangent to } X \text{ at } x\}$ be the conormal variety of X. $\mathbb{P}(\mathcal{N}(-1))$ can be identified via the exact sequences 0.1 and 0.2 with C_X and the two projections give the diagram

$$\begin{array}{ccc} \mathbb{P}(\mathcal{N}(-1)) = C_X & \xrightarrow{q} & X^\vee \subset \mathbb{P}_n^\vee \\ \downarrow p & & \\ X & & \end{array}$$

where $X^\vee$ is the dual variety of X [E2].

Since $q^* \mathcal{O}_{\mathbf{P}_n^\vee}(1) \simeq \mathcal{O}_{\mathbf{P}(\mathcal{N}(-1))}(1)$ and $\mathcal{O}_{\mathbf{P}(\mathcal{N}(-1))}(1)$ is globally generated by sections we see that $\mathcal{N}(-1)$ is not ample if and only if there exists a positive-dimensional fiber of q. Now obviously (2) implies (1).

Let conversely D be a reduced and irreducible curve in a fiber of q, say in $q^{-1}(H_0)$. Then $C := p(D)$ is again a reduced and irreducible curve since the restriction of q to any fiber of p is an isomorphism. By construction H_0 is tangent to X along C. □

1.3 *Remarks.* (1) If $k=1$, i.e. X is a smooth curve in $\mathbb{P}_n$, Proposition 1.2 yields: $\mathcal{N}(-1)$ is ample if and only if X is non-degenerate.

(2) Since the dimension of C_X is $N-1$ we see that $\mathcal{N}(-1)$ is not ample if the dimension of $X^\vee$ is less than $N-1$. Smooth varieties X with small dual varieties, i.e. $\dim(X^\vee) < N-1$, have been classified up to dimension 7 in [E1, E2, L-S].

(3) L. Ein proved that condition 1.2(2) is equivalent to any one of the following two conditions [E2]:

(i) $$\mathcal{N}(-1)|_C \simeq \mathcal{O}_C \oplus \mathcal{L}$$

where $\mathcal{L}$ is a vector bundle of rank $n-k-1$ on C.

(ii) $$H^0(C, \mathcal{N}^\vee(1)|_C) \neq 0.$$

Now we turn to the surface case.

1.4 Corollary. *For a smooth non-degenerate and linearly normal surface S in $\mathbb{P}_n$ the following statements are equivalent:*

(1) $\mathcal{N}(-1)$ is not ample.

(2) There exists a reduced and irreducible curve C on S such that

$$H^0(S, \mathcal{O}_S(H-2C)) \neq 0$$

where H is the class of a hyperplane section of S. In this case $\mathcal{N}(-1)|_C$ is not ample.

Proof. (1) implies (2) by Proposition 1.2 even without the assumption on linear normality.

Let conversely s be a non-trivial section of $\mathcal{O}_S(H-2C)$. Then s defines a non-trivial section of $\mathcal{O}_S(H)$ which lifts to a section $\tilde{s}$ of $\mathcal{O}_{\mathbf{P}_n}(1)$ by linear normality. The zero-set of $\tilde{s}$ is a hyperplane tangent to S along C. Hence Proposition 1.2 gives the assertion. □

1.5 *Remarks.* (1) Since all smooth surfaces in $\mathbb{P}_4$ except the Veronese surface of degree 4 are linearly normal, Corollary 1.4 applies in particular in this case.

(2) The same argument as in the proof of Corollary 1.4 allows us to construct many smooth projective varieties with non-ample $\mathcal{N}(-1)$: Let X be a smooth variety of dimension $k \geqq 2$ and $\mathcal{L}$ be an ample line bundle on X which has

a non-trivial section. Consider very ample tensor powers $\mathscr{L}^{\otimes r}$ $(r \geqq 2)$ of $\mathscr{L}$ and let X_r be the image of X under the embedding

$$X \hookrightarrow \mathbb{P}(H^0(X, \mathscr{L}^{\otimes r})^\vee) =: \mathbb{P}_n.$$

Then $\mathscr{N}_{X_r|\mathbb{P}_n}(-1)$ is not ample.

From now on we consider the normal bundle of smooth non-degenerate surfaces in $\mathbb{P}_4$. Note that for higher rank vector bundles there is no numerical criterion for ampleness in contrast to the Nakai-Moishezon criterion for line bundles. But for an ample rank 2 vector bundle $\mathscr{E}$ there are three necessary numerical conditions [F 1, Example 12.1.7]:

$$\det(\mathscr{E}) \text{ is ample}, \quad c_2(\mathscr{E}) > 0 \quad \text{and} \quad c_1^2(\mathscr{E}) - c_2(\mathscr{E}) > 0.$$

We first show that these three conditions are always fulfilled for $\mathscr{N}(-1)$ of a smooth non-degenerate surface in $\mathbb{P}_4$.

1.6 Proposition. *Let S be a smooth non-degenerate surface in $\mathbb{P}_4$ and $c_i := c_i(\mathscr{N}(-1))$. Then:*

(1) $\det(\mathscr{N}(-1))$ *is ample*

(2) $c_2 > 0$

(3) $c_1^2 - c_2 > 0$.

Proof. (1) From the two exact sequences 0.1 and 0.2 we calculate:

$$c_1(\mathscr{N}(-1)) = c_1(\det(\mathscr{N}(-1)) = 3H + K$$

where H is the class of a hyperplane section and K is the class of the canonical divisor of S.

By a more general result of Lanteri, Palleschi and Sommese [L-P-S] $3H + K$ is the class of an ample divisor and therefore (1) holds.

(2) Since $\mathscr{N}(-1)$ is globally generated we have $c_2 \geqq 0$. Assume $c_2 = 0$. Then we obtain an exact sequence [F 1, Example 12.1.8]:

$$0 \to \mathscr{O}_S \to \mathscr{N}(-1) \to \det(\mathscr{N}(-1)) \to 0.$$

By (1) and the vanishing theorem of Kodaira this sequence splits and consequently:

$$\mathscr{N}(-1) \simeq \mathscr{O}_S \oplus \det(\mathscr{N}(-1)).$$

But then S is necessarily contained in a hyperplane which contradicts the degeneracy of S.

(3) $c_1^2 - c_2$ is just the second Segre class $s_2(\mathscr{N}(-1))$ which is nothing but the degree of the dual variety $S^\vee$ of S in $\mathbb{P}_4^\vee$ since $S^\vee$ is a hypersurface [K, p. 361/362]. Therefore this number is positive. □

As a first example we consider abelian surfaces in $\mathbb{P}_4$ which can e.g. be realized as zero-sets of sections of the Horrocks-Mumford bundle $\mathscr{F}$.

1.7 Proposition. *Let A be an abelian surface in $\mathbb{P}_4$. Then $\mathscr{N}(-1)$ is ample.*

Proof. Let H be the class of a hyperplane section of A and assume that $\mathcal{N}(-1)$ is not ample. By Corollary 1.4 there exists a reduced and irreducible curve C on A such that

$$H^0(A, \mathcal{O}_A(H-2C)) \neq 0.$$

Therefore we obtain a linear equivalence

$$H \sim 2C + D$$

where D is some effective divisor on A.

Since the degree of an abelian surface in $\mathbb{P}_4$ is 10 (this follows e.g. from $c_2(\mathcal{F}) = 10$) we have:

$$(*) \qquad 10 = H^2 = 4C^2 + 4CD + D^2.$$

The intersection number of two effective divisors on an abelian surface is non-negative and by the theorem of Riemann-Roch selfintersection numbers are even. Hence we get two cases:

1. Case. $C^2 = 0$

From the adjunction formula and the fact that there are no rational curves on an abelian surface we obtain that C is a smooth elliptic curve in particular we have:

$$\text{degree}(C) = HC = DC \geqq 3.$$

Now (*) gives the contradiction

$$10 = 4CD + D^2 \geqq 12.$$

2. Case. $C^2 = 2$

Since an effective divisor with positive selfintersection on an abelian surface is ample we conclude from (*):

$$CD > 0$$

and get the contradiction:

$$10 = 8 + 4CD + D^2 \geqq 12. \quad \square$$

Our next aim is to show that there exist smooth non-degenerate surfaces in $\mathbb{P}_4$ of arbitrary high degree where $\mathcal{N}(-1)$ is not ample.

1.8 We look at symmetric Castelnuovo surfaces in $\mathbb{P}_4$ which are surfaces S_b of odd degree $2b+1$ ($b \geqq 1$), whose ideal sheaf $\mathcal{I}_b$ has a resolution of the following type

$$0 \longrightarrow \mathcal{O}_{\mathbb{P}_4} \oplus \mathcal{O}_{\mathbb{P}_4} \xrightarrow{\varphi} \mathcal{O}_{\mathbb{P}_4}(1) \oplus \mathcal{O}_{\mathbb{P}_4}(1) \oplus \mathcal{O}_{\mathbb{P}_4}(b) \longrightarrow \mathcal{I}_b(b+2) \longrightarrow 0$$

where the morphism φ is given by a 3×2-matrix

$$M = \begin{pmatrix} A \\ f_1 \quad f_2 \end{pmatrix}$$

and A is a symmetric 2×2-matrix of linear forms and f_1, f_2 are forms of degree b. The existence of such surfaces has been proved by Hulek, Okonek and Van de Ven in [H-O-V].

We shall need the following three properties of S_b which among many others can be found in [H-O-V]:

1.8.1 There exists a quotient

$$\mathcal{N}^{\vee}_{S_b|\mathbb{P}_4} \to \omega_b \otimes \mathcal{O}_{S_b}(2-2b) \to 0$$

where ω_b is the canonical bundle of S_b.

1.8.2 Let H_b resp. K_b be the class of a hyperplane section resp. the class of the canonical divisor of S_b. Then:

$$\begin{aligned} H_b^2 &= \deg(S_b) = 2b+1 \\ H_b K_b &= 2b^2-4b-3 \\ K_b^2 &= 2b^3-9b^2+6b+9. \end{aligned}$$

1.8.3 There exists a fibration $\psi_b\colon S_b \to \mathbb{P}_1$ where the fibers F_b are plane curves and their class is

$$F_b = K_b + (3-b)H_b.$$

1.9 Proposition. *Let S_b be a symmetric Castelnuovo surface in $\mathbb{P}_4$. Then $\mathcal{N}(-1)$ is not ample.*

Proof. Since $\mathcal{N}$ is of rank 2 we have

$$\mathcal{N}^{\vee} \simeq \mathcal{N} \otimes \det(\mathcal{N}^{\vee}).$$

From the exact sequences 0.1 and 0.2 follows:

$$\det(\mathcal{N}^{\vee}) \simeq \mathcal{O}_{S_b}(-5) \otimes \omega_b^{\vee}.$$

Now the quotient in 1.8.1 reads:

$$\mathcal{N}(-1) \to \omega_b \otimes \omega_b \otimes \mathcal{O}_{S_b}(6-2b) \to 0.$$

Let F be a smooth fiber of the fibration ψ_b. An easy calculation using the identities in 1.8.2 and 1.8.3 shows:

$$\deg((\omega_b \otimes \omega_b \otimes \mathcal{O}_{S_b}(6-2b))|_F) = 0.$$

Therefore $\mathcal{N}(-1)|_F$ has a quotient line bundle of degree 0. Hence $\mathcal{N}(-1)$ is not ample by a result of Gieseker [G, Prop. 2.1]. □

2 Non-special rational surfaces in $\mathbb{P}_4$

A smooth surface S in $\mathbb{P}_4$ is *non-special* if $h^1(S, \mathcal{O}_S(1)) = 0$. The classification of smooth, non-special rational surfaces in $\mathbb{P}_4$ has been completed by Okonek and Alexander [A, O1, O2] (see Table 2.2). They are all isomorphic to $\mathbb{P}_2$ blown-up in some points and embedded by a suitable very ample linear system.

2.1 *Notation.* For a smooth surface S in $\mathbb{P}_4$ we denote by d the degree of S, by π the genus of a smooth hyperplane section and by H the embedding linear system, i.e. $\mathcal{O}_S(H) \simeq \mathcal{O}_S(1)$. If S is isomorphic to $\mathbb{P}_2$ blown-up in the points $x_1, \ldots, x_r$ we define $\sigma\colon S \to \mathbb{P}_2$ to be the blow-up mapping, $E_i := \sigma^{-1}(x_i)$ the exceptional divisors and L a divisor on S such that $\sigma^*(\mathcal{O}_{\mathbb{P}_2}(1)) \simeq \mathcal{O}_S(L)$. Consequently $\mathrm{Pic}(S)$ is isomorphic to $\mathbb{Z}^{r+1}$ generated by $L, E_1, \ldots, E_r$.

2.2

d	π	S	H	
3	0	$\tilde{\mathbb{P}}_2(x_0)$	$2L-E_0$	cubic scroll
4	0	$\mathbb{P}_2$	$H \subset 2L$	Veronese surface
4	1	$\tilde{\mathbb{P}}_2(x_1 \ldots x_5)$	$3L-\sum_{i=1}^{5} E_i$	del Pezzo surface
5	2	$\tilde{\mathbb{P}}_2(x_0 \ldots x_7)$	$4L-2E_0-\sum_{i=1}^{7} E_i$	Castelnuovo surface
6	3	$\tilde{\mathbb{P}}_2(x_1 \ldots x_{10})$	$4L-\sum_{i=1}^{10} E_i$	Bordiga surface
7	4	$\tilde{\mathbb{P}}_2(x_1 \ldots x_6, y_1 \ldots y_5)$	$6L-2\sum_{i=1}^{6} E_i-\sum_{j=1}^{5} F_j$	
8	5	$\tilde{\mathbb{P}}_2(x_0 \ldots x_{10})$	$7L-E_0-2\sum_{i=1}^{10} E_i$	
9	6	$\tilde{\mathbb{P}}_2(x_1 \ldots x_{10})$	$13L-4\sum_{i=1}^{10} E_i$	

Table 1. Table of smooth non-special non-degenerate rational surfaces in $\mathbb{P}_4$

Notice that the property for a linear system H as in 2.2 to be very ample is an open condition on the set of points in $\mathbb{P}_2$.

The proof of Theorem 2.4 is based on the following elementary lemma.

2.3 Lemma. *Let D be a divisor on $S \simeq \tilde{\mathbb{P}}_2(x_1 \ldots x_r)$ of the class*

$$D = aL - \sum_{i=1}^{r} b_i E_i.$$

(1) If $h^0(S, \mathcal{O}_S(D)) \neq 0$ then $a \geqq 0$.

(2) If $b_j < 0$ then:

$$h^0(S, \mathcal{O}_S(D)) \neq 0 \quad \text{if and only if} \quad h^0(S, \mathcal{O}_S(D + b_j E_j)) \neq 0.$$

(3) If $h^0(S, \mathcal{O}_S(D)) \neq 0$ and $a = 0$ then $b_i \leqq 0$ for all i.

(4) If D is an irreducible curve then one of the following two conditions holds:

$$\begin{array}{llll} \text{either} & a>0 & \text{and} & b_i \geqq 0 \quad \text{for all } i \\ \text{or} & a=0 & \text{and} & D=-b_j E_j \quad \text{for some } b_j<0. \end{array}$$

In the latter case D is reduced if and only if $b_j = -1$.

(5) If D is an effective divisor with $a>0$ and $b_i \geqq 0$ for all i there is a 1 to 1 correspondence between sections of $\mathcal{O}_S(D)$ and curves of degree a in $\mathbb{P}_2$ having multiplicity at least b_i in x_i for all i.

Proof. The proof follows e.g. from [H, V, § 4]. □

We call two non-special rational surfaces S and S' in $\mathbb{P}_4$ *of the same type* if $d=d'$ and $\pi=\pi'$. We say that S is *rigid* if all surfaces of the same type are projectively equivalent to S.

2.4 Theorem. *Let S be a smooth, non-special, non-degenerate rational surface in $\mathbb{P}_4$.*

(1) If S is rigid then $\mathcal{N}(-1)$ is not ample.

(2) If S is not rigid then S can be deformed to a surface S' of the same type such that $\mathcal{N}_{S'|\mathbb{P}_4}(-1)$ is ample.

2.5 *Remarks.* (1) The rigid non-special rational surfaces in $\mathbb{P}_4$ are precisely the cubic scroll and the Veronese surface [O 3].

(2) Let $S \simeq \tilde{\mathbb{P}}_2(x_1 \dots x_r)$ be a non-rigid, non-special rational surface in $\mathbb{P}_4$. Then 2.4(2) is equivalent to: $\mathcal{N}(-1)$ is ample if the position of $x_1 \dots x_r$ in $\mathbb{P}_2$ is general enough. In fact we shall prove 2.4(2) by showing that the non-ampleness of $\mathcal{N}(-1)$ forces $x_1 \dots x_r$ to be in some special position in $\mathbb{P}_2$. Furthermore for the Castelnuovo surface of degree 5 and the Bordiga surface of degree 6 the exceptional case, namely that $\mathcal{N}(-1)$ is not ample does really occur (see Remarks 2.8 and 2.9).

Proof of Theorem 2.4 (1) Let $S \simeq \tilde{\mathbb{P}}_2(x_0)$ be the cubic scroll and F a line of the ruling. We look at the exact sequence of normal bundles

$$0 \to \mathcal{N}_{F|S} \to \mathcal{N}_{F|\mathbb{P}_4} \to \mathcal{N}_{S|\mathbb{P}_4}|_F \to 0.$$

Since F is a complete intersection of three hyperplanes and every bundle on $\mathbb{P}_1$ splits into a direct sum of line bundles this sequence reads

$$0 \to \mathcal{O}_{\mathbb{P}_1} \to \bigoplus_{i=1}^{3} \mathcal{O}_{\mathbb{P}_1}(1) \to \mathcal{O}_{\mathbb{P}_1}(a) \oplus \mathcal{O}_{\mathbb{P}_1}(b) \to 0$$

where $a+b=3$. W.l.o.g. $a \leqq b$.

The ampleness of $\mathcal{N}$ implies $a=1$ and $b=2$. Hence

$$\mathcal{N}(-1)|_F \simeq \mathcal{O}_{\mathbb{P}_1} \oplus \mathcal{O}_{\mathbb{P}_1}(1)$$

and $\mathcal{N}(-1)$ is not ample.

Consider now the Veronese embedding $\psi\colon \mathbb{P}_2 \xrightarrow{\sim} S \subset \mathbb{P}_4$ of $\mathbb{P}_2$ in $\mathbb{P}_4$ and let $\mathcal{E} := \psi^*(\mathcal{N}(-1))$.

We compute easily from the exact sequences 0.1 and 0.2:

$$c_1(\mathcal{E})=3 \quad \text{and} \quad c_2(\mathcal{E})=6.$$

Assume that $\mathcal{N}(-1)$ is ample. Then $\mathcal{E}$ is ample too and for every line ℓ in $\mathbb{P}_2$ we obtain

$$\mathcal{E}|_\ell \simeq \mathcal{O}_{\mathbb{P}_1}(1) \oplus \mathcal{O}_{\mathbb{P}_1}(2).$$

This means that $\mathscr{E}$ is a uniform rank 2 bundle on $\mathbb{P}_2$. By a theorem of Van de Ven [O-S-S, p. 211] and $c_1(\mathscr{E})=3$ we conclude:

$$\mathscr{E}\simeq\mathcal{O}_{\mathbb{P}_2}(1)\oplus\mathcal{O}_{\mathbb{P}_2}(2) \quad \text{or} \quad \mathscr{E}\simeq\mathscr{T}_{\mathbb{P}_2}.$$

Both cases contradict $c_2(\mathscr{E})=6$.

(2) The proof is case by case.

Note first that $\mathscr{N}(-1)$ of a del Pezzo surface is always ample since these are complete intersections of two hyperquadrics.

We restrict ourselves to the degree 9 case. The proofs in the other cases being similar. Let $S\simeq\tilde{\mathbb{P}}_2(x_1\dots x_{10})$ be embedded in $\mathbb{P}_4$ by the linear system $H=13L-4\sum_{i=1}^{10}E_i$.

Assume that $\mathscr{N}(-1)$ is not ample. We shall show that then the points $x_1\dots x_{10}$ have to be in special position in $\mathbb{P}_2$.

First of all from Corollary 1.4 we obtain the existence of a reduced and irreducible curve C on S such that $H-2C$ is an effective class. Consequently:

$$1\leqq\deg(C)=HC\leqq\tfrac{1}{2}H^2=4\tfrac{1}{2}.$$

Write

$$C=aL-\sum_{i=1}^{10}b_iE_i.$$

Then

$$H-2C=(13-a)L-\sum_{i=1}^{10}(4-2b_i)E_i.$$

Lemma 2.3 yields:

$$0\leqq a\leqq 6.$$

1. Case. $a=0$

Again by Lemma 2.3 C is one of the exceptional divisors, say $C=E_1$, i.e.

$$H-2C=13L-6E_1-4\sum_{i=2}^{10}E_i.$$

A section of $\mathcal{O}_S(H-2C)$ corresponds to a curve of degree 13 in $\mathbb{P}_2$ having at least a sixfold point in x_1 and at least fourfold points in $x_2\dots x_{10}$.

Claim. If $x_1\dots x_{10}$ are in general position there is no such curve.

Proof. Since the non-existence of a curve as above is an open condition on sets of 10 points in $\mathbb{P}_2$ it is enough to find one set of 10 points with this property. Choose 10 points as follows:

Let $L_1\dots L_4$ be four lines intersecting in one point, say x_1. Choose x_{10} outside of these lines and let $M_1\dots M_4$ be four lines through x_{10} not containing x_1. Let $x_2\dots x_5$ be the intersection points of L_1 and L_2 with M_1 and M_2 and let $x_6\dots x_9$ be the intersection points of L_3 and L_4 with M_3 and M_4.

Now assume that D_{13} is a curve with the prescribed multiplicities. Hence

$$D_{13} \cdot L_i \geqq 6+4+4=14 \quad \text{for } 1 \leqq i \leqq 4.$$

Bezout's theorem yields a decomposition

$$D_{13} = \bigcup_{i=1}^{4} L_i \cup D_9$$

where D_9 is a curve of degree 9 having at least a double point in x_1, at least triple points in $x_2 \ldots x_9$ and at least a fourfold point in x_{10}. Hence

$$D_9 \cdot M_i \geqq 4+3+3=10 \quad \text{for } 1 \leqq i \leqq 4.$$

Again by Bezout

$$D_9 = \bigcup_{i=1}^{4} M_i \cup D_5$$

where D_5 is a curve of degree 5 having at least double points in $x_1 \ldots x_9$. Hence

$$D_5 \cdot L_i \geqq 2+2+2=6 \quad \text{for } 1 \leqq i \leqq 4.$$

This means

$$D_5 = \bigcup_{i=1}^{4} L_i \cup L_0$$

where L_0 is a line through $x_2 \ldots x_9$. Contradiction.

2. Case. $a=1$

From Lemma 2.3 we have

$$C = L - \sum_{i=1}^{10} b_i E_i$$

with $0 \leqq b_i \leqq 1$ for all i. Furthermore

$$1 \leqq \deg(C) = 13 - 4 \sum_{i=1}^{10} b_i \leqq 4.$$

Consequently $\sum_{i=1}^{10} b_i = 3$, i.e. three points are collinear and therefore $x_1 \ldots x_{10}$ are in special position.

3. Case. $a=2$

Again by Lemma 2.3

$$C = 2L - \sum_{i=1}^{10} b_i E_i$$

with $0 \leqq b_i \leqq 1$ for all i. The inequalities

$$1 \leqq \deg(C) = 26 - 4 \sum_{i=1}^{10} b_i \leqq 4$$

yield $\sum_{i=1}^{10} b_i = 6$, i.e. six points lie on a conic which is also a special position.

4. Case. $a = 3$

In this case

$$C = 3L - \sum_{i=1}^{10} b_i E_i$$

with $0 \leqq b_i \leqq 2$ for all i and at most one b_i equals 2. From

$$1 \leqq \deg(C) = 39 - 4 \sum_{i=1}^{10} b_i \leqq 4$$

we obtain $\sum_{i=1}^{10} = 9$ which yields two subcases (after possible renumbering):

Subcase (4.1). $$C = 3L - 2E_i - \sum_{i=2}^{8} E_i.$$

Hence

$$H - 2C = 7L - 2 \sum_{i=2}^{8} E_i - 4E_9 - 4E_{10}.$$

A section of $\mathcal{O}_S(H-2C)$ corresponds to a curve of degree 7 in $\mathbb{P}_2$ having at least double points in $x_2 \ldots x_8$ and at least fourfold points in x_9 and x_{10}.

Claim. If $x_1 \ldots x_{10}$ are in general position there is no such curve.

Proof. We argue as in the first case.

Let $L_1 \ldots L_4$ be four lines intersecting in one point, say x_{10}. Choose $x_9 \neq x_{10}$ on L_4 and two lines L_5 and L_6 through x_9. Let $x_2 \ldots x_7$ be the six intersection points of L_5 and L_6 with L_1, L_2 and L_3. Let x_1 and x_8 be two points outside of these six lines.

Assume that D_7 is a curve with the prescribed multiplicities. Then

$$D_7 \cdot L_1 \geqq 8 \quad \text{for } 1 \leqq i \leqq 6.$$

By Bezout's theorem

$$D_7 = \bigcup_{i=1}^{6} L_i \cup L_0$$

where L_0 is a line having double points in x_1 and x_8. Contradiction.

Subcase (4.2). $$C = 3L - \sum_{i=1}^{9} E_i.$$

Hence

$$H-2C=7L-2\sum_{i=1}^{9}E_i-4E_{10}.$$

Claim. If $x_1 \dots x_{10}$ are in general position there is no curve of degree 7 in $\mathbb{P}_2$ having at least double points in $x_1 \dots x_9$ and at least a fourfold point in x_{10}.

Proof. Let $L_1 \dots L_4$ be four lines intersecting in one point, say x_{10}. Choose two lines L_5 and L_6 intersecting $L_1 \dots L_4$ in 8 points, say $x_2 \dots x_9$. Let x_1 be a point outside of these six lines.

Assume that D_7 is a curve with the above property. Then

$$D_7 \cdot L_i \geqq 8 \quad \text{for } 1 \leqq i \leqq 6.$$

Hence

$$D_7 = \bigcup_{i=1}^{6} L_i \cup L_0$$

where L_0 is a line having a double point in x_1. Again a contradiction.

5. Case. $a=4$

We have

$$C=4L-\sum_{i=1}^{10} b_i E_i$$

with $0 \leqq b_i \leqq 3$ for all i. Moreover

$$1 \leqq \deg(C)=52-4\sum_{i=1}^{10} b_i \leqq 4$$

which yields $\sum_{i=1}^{10} b_i = 12$. By the genus formula for plane curves [F2, p. 117] the singularities of a reduced and irreducible quartic can be either a triple point or at most three double points. Therefore we obtain three subcases (after possible renumbering):

Subcase (5.1). $$C=4L-3E_1-\sum_{i=2}^{10} E_i.$$

Then

$$H-2C=5L+2E_1-2\sum_{i=2}^{10} E_i.$$

Since

$$\left(5L-2\sum_{i=2}^{10} E_i\right)\cdot H=65-72=-7$$

we see by Lemma 2.3 that $H-2C$ is not effective. So this case does not occur.

Subcase (5.2). $$C=4L-2E_1-2E_2-\sum_{i=3}^{10} E_i.$$

Hence

$$H-2C=5L-2\sum_{i=3}^{10} E_i.$$

From $(H-2C)\cdot H=1$ we deduce that the corresponding divisor is a line in $\mathbb{P}_4$. On the other hand a plane quintic with at least 8 double points is by the genus formula necessarily reducible. Therefore this case also does not occur.

Subcase (5.3). $$C=4L-2\sum_{i=1}^{3} E_i-\sum_{i=4}^{9} E_i.$$

Hence

$$H-2C=5L-2\sum_{i=4}^{9} E_i-4E_{10}.$$

And $(H-2C)\cdot H=1$ plus the genus formula gives the same contradiction as in Subcase (5.2).

6. *Case.* $a=5$

Hence

$$C=5L-\sum_{i=1}^{10} b_i E_i$$

with $0\leqq b_i\leqq 4$ for all i. The inequalities

$$1\leqq \deg(C)=65-4\sum_{i=1}^{10} b_i\leqq 4$$

imply $\sum_{i=1}^{10} b_i=16$. Furthermore by the genus formula a reduced and irreducible plane quintic can have the following singularities:

either one fourfold point
or two triple points
or one triple point and at most three double points
or at most six double points.

From this we find after renumbering if necessary only one numerical possibility for the b_i's:

$$C=5L-2\sum_{i=1}^{6} E_i-\sum_{i=7}^{10} E_i.$$

Hence

$$H-2C=3L-2\sum_{i=7}^{10} E_i.$$

A section of $\mathcal{O}_S(H-2C)$ corresponds to a plane cubic with at least double points in $x_7 \dots x_{10}$. Consequently $x_7 \dots x_{10}$ have to lie on a line and therefore $x_1 \dots x_{10}$ are in special position in $\mathbb{P}_2$.

Last case. $a=6$

Then

$$C=6L-\sum_{i=1}^{10} b_i E_i$$

with $0 \leqq b_i \leqq 5$ for all i. From

$$1 \leqq \deg(C)=78-4\sum_{i=1}^{10} b_i \leqq 4$$

we deduce $\sum_{i=1}^{10} b_i=19$. Again by the genus formula a reduced and irreducible plane sextic allows the following singularities:

either one fivefold point
or one fourfold point
or one fourfold point and one triple point
or one fourfold point, one triple point and one double point
or one fourfold point and at most four double points
or three triple points and at most one double point
or two triple points and at most four double points
or one triple point and at most seven double points
or at most ten double points.

After possible renumbering we obtain two subcases:

Subcase (7.1). $$C=6L-3E_1-2\sum_{i=2}^{8} E_i-E_9-E_{10}.$$

Hence

$$H-2C=L+2E_1-2E_9-2E_{10}.$$

Subcase (7.2). $$C=6L-2\sum_{i=1}^{9} E_i-E_{10}.$$

Then

$$H-2C=L-2E_{10}.$$

Since a line can not have any double points both subcases do not occur. This finishes the proof of Theorem 2.4 in the degree 9 case. □

We conclude the section with several remarks.

2.6 *Remarks on scrolls.* (1) The non-ampleness of $\mathcal{N}(-1)$ of the cubic scroll follows also from Prop. 1.9 since every cubic scroll is a symmetric Castelnuovo

surface of degree 3. However the proof given in 2.4(1) works for the elliptic scroll of degree 5 too. Besides these two there are no further scrolls in $\mathbb{P}_4$ [A].

(2) Let $S \simeq \tilde{\mathbb{P}}_2(x_0)$ be embedded by the linear system $H = 2L - E_0$ as a cubic scroll in $\mathbb{P}_4$ and let $\sigma\colon S \to \mathbb{P}_2$ be the blow-up mapping. A suitable twist of the normal bundle of S in $\mathbb{P}_4$ descents to $\mathbb{P}_2$. It is not hard to see that [B, § 7]:

$$\mathcal{N}_{S|\mathbb{P}_4}(-2) \simeq \sigma^*(\mathcal{T}_{\mathbb{P}_2}(-2)).$$

2.7 *Remark on the Veronese surface.* Let $\psi\colon \mathbb{P}_2 \to S \subset \mathbb{P}_4$ be the Veronese embedding of $\mathbb{P}_2$ in $\mathbb{P}_4$ and let $\mathscr{E} := \psi^* \mathcal{N}(-1))$. We have seen that $c_1(\mathscr{E}) = 3$. Since the restriction of $\mathscr{E}$ to a general line ℓ in $\mathbb{P}_2$ has to be ample the generic splitting type of $\mathscr{E}$ is (1, 2). There is a one-dimensional family of jump lines in $\mathbb{P}_2$ parametrized by a smooth conic in $\mathbb{P}_2^\vee$, where the splitting type changes [Hu]. The splitting type on such lines is (0, 3) since $\mathscr{E}$ is globally generated. The images of these lines are precisely the curves on S where $\mathcal{N}(-1)$ fails to be ample.

2.8 *Remark on Castelnuovo surfaces of degree* 5. Let $S \simeq \tilde{\mathbb{P}}_2(x_0 \dots x_7)$ and $H = 4L - 2E_0 - \sum_{i=1}^{7} E_i$. The very-ampleness criterion of Biancofiore [Bi] gives:

2.8.1 Lemma. *H embeds S as a Castelnuovo surface of degree 5 in $\mathbb{P}_4$ if and only if the following three conditions are satisfied:*

(1) No four x_i's lie on a line.

(2) On any line through x_0 there is at most one further x_i.

(3) On any conic through x_0 there are at most 5 other x_i's.

According to the position of the points $x_0 \dots x_7$ in $\mathbb{P}_2$ we are able to decide whether $\mathcal{N}(-1)$ is ample or not. From similar calculations as in the proof of Theorem 2.4 we get the following result [B, § 7]:

2.8.2 Proposition. *Let S be a Castelnuovo surface of degree 5 in $\mathbb{P}_4$. Then $\mathcal{N}(-1)$ is not ample if and only if at least one of the following three conditions holds:*

(1) There is a quartic through $x_0 \dots x_7$ having at least a double point in x_0 and at least a triple point in some x_j $(j \neq 0)$. This condition is equivalent to: $\mathcal{N}(-1)|_{E_j}$ is not ample.

(2) There is a conic through $x_1 \dots x_{j-1}, x_{j+1} \dots x_7$ for some $1 \leqq j \leqq 7$. This condition is equivalent to: $\mathcal{N}(-1)|_{L-E_0-E_j}$ is not ample.

(3) There is a conic through $x_1 \dots x_7$. This condition is equivalent to: S is a symmetric Castelnuovo surface of degree 5.

Furthermore all these exceptional cases do in fact exist and there are surfaces which satisfy (1) *(resp.* (2)*) but not* (3).

2.9 *Remark on Bordiga surfaces of degree* 6. Let $S \simeq \tilde{\mathbb{P}}_2(x_1 \dots x_{10})$ and $H = 4L - \sum_{i=1}^{10} E_i$. From his criterion Biancofiore recovers the classically known

2.9.1 Lemma. *H embeds S as a Bordiga surface of degree 6 in $\mathbb{P}_4$ if and only if the following three conditions are satisfied:*

(1) No 4 x_i's lie on a line.

(2) No 8 x_i's lie on a conic.
(3) Not all x_i's lie on a cubic.

Again by the same methods we can prove [B, § 8]:

2.9.2 Proposition. *Let S be a Bordiga surface of degree 6 in $\mathbb{P}_4$. Then $\mathcal{N}(-1)$ is not ample if and only if there is a quartic through $x_1 \dots x_{10}$ having at least a triple point in x_j for some $1 \leqq j \leqq 10$. This condition is equivalent to: $\mathcal{N}(-1)|_{E_j}$ is not ample. Furthermore there exist Bordiga surfaces such that $\mathcal{N}(-1)|_{E_j}$ is not ample for all $1 \leqq j \leqq 10$.*

2.10 *Remark.* Consider the non-special rational surface of degree 9 in $\mathbb{P}_4$. A standard Chern class computation gives:

$$c_1^2 \mathcal{N}(-1)) - 2c_2(\mathcal{N}(-1)) = -2.$$

So we have found an example of an ample and globally generated rank 2 vector bundle with $c_1^2 < 2c_2$ (cf. also [F 3]).

Acknowledgement. I would like to thank K. Hulek for his encouragement during the preparation of this paper.

References

[A] Alexander, J.: Surfaces rationelles non-spèciales dans $\mathbb{P}_4$. Math. Z. **200**, 87–100 (1988)
[Au] Aure, A.: On surfaces of projective 4-space. Thesis, Oslo 1987
[Bi] Biancofiore, A.: On the hyperplane section of blow-ups of complex projective plane. Can. J. of Math. **41**, 1005–1020 (1989)
[B] Braun, R.: Über das Normalenbündel von projektiven Varietäten. Dissertation, Bayreuth 1989
[E 1, E 2] Ein, L.: Varieties with small dual varieties. I, II. Invent. Math. **86**, 63–74 (1986), Duke Math. J. **52**, 895–907 (1985)
[F 1] Fulton, W.: Intersection theory. Berlin Heidelberg New York: Springer 1984
[F 2] Fulton, W.: Algebraic curves. New York: Benjamin 1969
[F 3] Fulton, W.: Ample vector bundles, Chern classes and numerical criteria. Invent. Math. **32**, 171–178 (1976)
[G] Gieseker, D.: p-ample bundles and their Chern classes. Nagoya Math. J. **43**, 91–116 (1971)
[H 1] Hartshorne, R.: Algebraic geometry. Berlin Heidelberg New York: Springer 1977
[H 2] Hartshorne, R.: Ample vector bundles. In: Publ. Math. IHES **29**, pp. 63–94 (1966)
[Hu] Hulek, K.: Complete intersection curves, the splitting of the normal bundle and the Veronese surface. In: Casas-Alvero, E., Welters, G., Xambó-Descamps, S. (eds.) Algebraic geometry. Proceedings, Sitges 1983. (Lect. Notes Math., vol. 1124, pp. 132–145) Berlin Heidelberg New York: Springer 1985
[H-O-V] Hulek, K., Okonek, C., Van de Ven, A.: Multiplicity-2 structures on Castelnuovo surfaces. Ann. Sc. Norm. Sup. Pisa, Cl. Sci., IV. Ser. **13**, 427–448 (1986)
[K] Kleiman, S.L.: The enumerative theory of singularities. In: Holm, P. (ed.) Real and complex singularities. Proceedings of the Nordic Summer School, pp. 297–396, Oslo 1976. Alphen aan den Rijn: Sijthoff & Nordhoff International Publishers 1977
[L-P-S] Lanteri, A., Palleschi, M., Sommese, A.J.: Very ampleness of $\mathcal{K}_X \otimes \mathcal{L}^{\otimes \dim X}$ for ample and spanned line bundles $\mathcal{L}$. Osaka J. Math. **26**, 647–664 (1989)
[L-S] Lanteri, A., Struppa, D.: Projective 7-folds with positive defect. Comp. Math. **61**, 329–337 (1987)
[O 1] Okonek, C.: Über 2-codimensionale Untermannigfaltigkeiten vom Grad 7 in $\mathbb{P}_4$ und $\mathbb{P}_5$. Math. Z. **187**, 209–219 (1984)
[O 2] Okonek, C.: Flächen vom Grad 8 in $\mathbb{P}_4$. Math. Z. **191**, 207–223 (1986)
[O 3] Okonek, C.: On codimension-2 submanifolds in $\mathbb{P}_4$ and $\mathbb{P}_5$. Mathematica Gottingensis **50** (1986)
[O-S-S] Okonek, C., Schneider, M., Spindler, H.: Vector bundles on complex projective spaces. Progr. Math. **3**, Basel: Birkhäuser 1980

Math. Z. 206, 551–562 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On the number of positive solutions of some weakly nonlinear equations on annular regions

E.N. Dancer

Department of Mathematics, Statistics and Computing Science,
The University of New England, Armidale, N.S.W. 2351, Australia

Received March 22, 1990; in final form June 29, 1990

In this paper, we discuss the number of positive solutions of the equation

$$-\Delta u = u^{\alpha} \quad \text{in} \quad \Omega_{\varepsilon}$$
$$u = 0 \quad \text{on} \quad \partial\Omega_{\varepsilon}$$

where $1 < \alpha < (m+2)/(m-2)$, Ω_{ε} is a domain in R^m and Ω_{ε} is some form of annular domain that is, a domain with 'holes'. Our main interest is in the case where the holes are small and star shaped, there is a fixed number of holes and the holes are not close to each other or to the boundary. Our main result is that, if the holes are small enough and if a natural non-degeneracy assumption holds, then the number of positive solutions is the same with or without the holes. The idea in the proof is to obtain a bound in L^{∞} for solutions and then apply the techniques of [4]. We also present counterexamples showing that the result may be false (in particular, there may be a solution large in the sup norm) if we delete the star-shapedness assumption or the condition that the holes are not close to the boundary of if the holes are close together. This shows in conjunction with [5] how complicated the problem of large solutions is. Our techniques can also be used to show that there are large solutions in some cases even if there are holes which are not small but are close to the boundary.

The one way that we prove that there are large solutions is by a general result that, if $m=2$, if Ω_{ε} has a suitable Z_2 symmetry, if the strip $(-a, a) \times (-\varepsilon, \varepsilon) \subseteq \Omega_{\varepsilon}$ and $(-a, a) \times \{-\varepsilon, \varepsilon\} \subseteq \partial\Omega_{\varepsilon}$ and if Ω_{ε} has bounded measure uniformly in ε, then for small ε there is a large solution concentrated in the strip. This general result should be of use elsewhere. The proof is by a modification of the proof of Theorem 3 in [5].

As a simple application of our ideas, our main result implies that there is a unique non-trivial positive solution of (1) on the domain which consists of the unit ball with a small circular hole (not necessarily central but not close to the boundary). Even in the case of the true annulus (with small hole) this result is new. It was shown earlier by Ni and Nussbaum [16] that for any true annulus (1) has a unique non-trivial *radially symmetric* positive solution. Our result shows that this solution is the unique positive solution if the hole is small. On the other hand, if the hole is large there may also be non-radially

symmetric solutions. Some results similar to this last one were found earlier by Brezis and Nirenberg [1] and Coffman [3]. A discussion of why the equation is of interest can be found in [5] and [7].

We also briefly consider some more general nonlinearities.

In § 1, we obtain our main result, in § 2, we construct a number of counterexamples and prove our result on the existence of large solutions. In § 3 we consider the case of a true annulus and, in § 4, we briefly consider some more general nonlinearities.

Most of our results were announced in [7], though the results here improve those in [7].

§ 1 The main result

In this section, we prove our main result. *Assume that D is a bounded domain in R^m with C^2 boundary* (though this could be weakened a little). *Assume that $\{K_i\}_{i=1}^r$ are disjoint compact subsets of D with capacity zero and $\{U_i^j\}_{j=1,\ldots\infty}^{i=1,\ldots r}$ are open subsets of D such that U_i^j is star-shaped and the Hausdorff distance of U_i^j from K_i tends to zero as $j\to\infty$.* In addition, *assume that U_i^j has C^2 boundary.* (The easiest way that the Hausdorff distance condition is satisfied is that $U_j^{j+1}\subseteq U_i^j$ and $\bigcap_{j=1}^{\infty} U_i^j = K_i$. Let $D_n = D\setminus\bigcup_{i=1}^{r} U_i^n$.

We consider the equation

$$\begin{aligned} -\Delta u &= u^\alpha \quad \text{in} \quad \Omega \\ u &= 0 \quad \text{on} \quad \partial\Omega \end{aligned} \tag{1}$$

for various Ω's.

Theorem 1 *Assume that $m\geqq 2$, that the above assumptions on D hold, that $1<\alpha<(m+2)(m-2)^{-1}$ ($\alpha>1$ if $m=2$) and that every positive solution of* (1) *for $\Omega=D$ in $L^\infty(D)\cap \dot{W}^{1,2}(D)$ is non-degenerate (that is, has invertible linearization in $L^2(D)$). Then, for large n,* (1) *for $\Omega=D_n$ has the same number of positive solutions in $L^\infty(D_n)\cap \dot{W}^{1,2}(D_n)$ as for $\Omega=D$. Moreover, as $n\to\infty$, the solutions on D_n approach those on D in L^p for every p with $1<p<\infty$ (and are uniformly bounded in L^∞).*

Remark. When we talk of convergence in L^p, we extend a solution on D_n to be zero outside of D_n and we then mean convergence in $L^p(D)$.

Before proving Theorem 1, we need some comments. Firstly, it suffices to prove that there is a K independent of n such that $\|u\|_{\infty,D_n}\leqq K$ for every positive solution of (1) for Ω_n. Here $\|\ \|_{\infty,D_n}$ denotes the sup norm on D_n. For *fixed* n, this is a result of Gidas and Spruck [12]. The above bound suffices because under our above assumptions it follows easily that $D_n\to D$ in the sense of [6, p. 600]. (Note that $\bigcup_{i=1}^{r} K_i$ has capacity zero.) Hence the result will then follow from the proposition in [6] (which is a variant of the main result in [4]). Note that since solutions are uniformly bounded in L^∞, if they converge to solutions on D in some $L^p(D)$, Hölder's inequality ensures that this convergence holds in $L^p(D)$ for $1\leqq p<\infty$.

Hence it suffices to establish our apriori bound. Basic to this is the following variant of Lemma 1 in [5]. It is a slight strengthening of a result in [11].

Lemma 1 *Assume that u is a non-negative bounded solution of $-\Delta u = u^\alpha$ in $B_\mu(z)$ where $1<\alpha<(m+2)(m-2)^{-1}$ ($\alpha>1$ if $m=2$). Then $u(0)\leqq c(\mu, \alpha)$ where c is independent of the solution u. Here $B_\mu(z)$ denotes the ball of centre z and radius μ.*

Proof. By translation, we can assume $z=0$. By a simple rescaling, $v(x) = \tau^{2/(\alpha-1)}u(\tau x)$ is a solution of the same equation on $B_{R_0}(0)$ where R_0 is defined in [5], Lemma 1, and $\tau=\mu(R_0)^{-1}$. Hence the result follows from Lemma 1 in [5]. □

Proof of Theorem 1 We have to bound $\|u\|_\infty$ for a positive solution of (1) and $\Omega=D_n$. By Lemma 1, there is a bound $|u(x)|\leqq K_s$ whenever $x\in D_n$ and $d(x, \partial D_n)\geqq s$. Thus we only have to bound u near ∂D and close to $\bigcup_{i=1}^{r} K_i$.

We first bound u near ∂D. Suppose that the result is false. Then there exist positive solutions u_n of (1) for $\Omega=D_n$ and $x_n\in D_n$ near ∂D such that $u_n(x_n)\to\infty$ as $n\to\infty$. We can without loss of generality assume that x_n is the point of $S=\{x\in D: d(x, \partial D)\leqq\varepsilon\}$ where u achieves its maximum. Note that, by Lemma 1, u_n is bounded close to the 'inner' boundary of S. In particular, if n is large $u_n(x_n)$ is the maximum value of $u_n|\{x\in D: \|x-x_n\|\leqq\frac{1}{4}\varepsilon\}$. We can now use the same blowing up argument as in the proof of Theorem 1.1 in [12] to obtain a contradiction. Hence u_n is bounded close to ∂D. Thus u can only be large close to $\bigcup_{i=1}^{r} K_i$.

We now bound u close to $\bigcup_{i=1}^{r} K_i$. Fix i and choose a neighbourhood V_i of K_i such that V_i has smooth boundary and $\bar{V}_i$ does not intersect $\bigcup_{j\neq i} K_j$. For $n\geqq n_i$, $\bar{U}_i^n\subseteq\operatorname{int} V_i$.

Note that we have a uniform bound for ∇u on ∂V_i. (To see this, if $x\in V_i$, we choose a ball $\hat{B}$ with centre x and fixed small radius. By our comments above, solutions are uniformly bounded on $\hat{B}$ and hence by the equation, Δu is uniformly bounded on $\hat{B}$. Hence by using standard $W^{2,p}$ estimates on a concentric ball $\tilde{B}_1$ of smaller radius, we obtain bounds for ∇u on $\tilde{B}_1$ (cp part of the proof of Lemma 1 in [5]). Hence we have a bound for u in C^1 near ∂V_i.)

If $n\geqq n_i$, choose $y_i^n\in U_i^n$ so that U_i^n is star shaped about y_i^n. We now use a Pohozaev identity argument. We multiply the equation for u by $2(x-y_i^n)\cdot\nabla u$, integrate over $V_i\setminus U_i^n$ and integrate by parts. By a standard argument (cp part of the proof of Theorem 2 in [5]), we see that

$$2m(\alpha+1)^{-1}\int_{V_i\setminus U_i^n} u^{\alpha+1} - 2(\alpha+1)^{-1}\int_{\partial U_i^n} u^{\alpha+1}(x-y_i^n)\cdot\underline{n}\,dS$$

$$=(m-2)\int_{V_i-U_i^n}|\nabla u|^2+\int_{\partial U_i^n} 2(x-y_i^n)\cdot\nabla u\frac{\partial u}{\partial n}-|\nabla u|^2(x-y_i^n)\cdot\underline{n}\,dS+\tilde{C} \quad (2)$$

where $\tilde{C}$ represents similar terms to those on ∂U_i^n except on the surface ∂V_i and $\underline{n}$ denotes the outward normal. However, since u is bounded in C^1 near ∂V_i, we see that the terms in $\tilde{C}$ are uniformly bounded (in n). Note that $\underline{n}$

is the outward normal to $V_i \setminus U_i^n$ and hence is the *inward* normal to U_i^n. Thus, since U_i^n is star shaped, we see that the term

$$\begin{aligned}&\int_{\partial U_i^n} 2(x-y_i^n)\cdot \nabla u \frac{\partial u}{\partial n} - |\nabla u|^2 (x-y_i^n)\cdot \underline{n}\, dS \\ &= \int_{\partial U_i^n} (x-y_i^n)\cdot \underline{n} |\nabla u|^2 dS \\ &\leqq 0\end{aligned}$$

since $(x-y_i^n)\cdot \underline{n} \leqq 0$ by the star shapedness. (Note that for the first equality we have used that, since $u=0$ on ∂U_i^n, ∇u is parallel to $\underline{n}$ on ∂U_i^n. Note also that the second term in (2) is zero since $u=0$ on ∂U_i^n.) Hence we see from (2) that

$$2m(\alpha+1)^{-1} \int_{V_i \setminus U_i^n} u^{\alpha+1} - (m-2) \int_{V_i \setminus U_i^n} |\nabla u|^2 \leqq \tilde{C} \tag{3}$$

where $\tilde{C}$ is independent of n. On the other hand, by multiplying (1) by u and integrating over $V_i \setminus U_i^n$, we see by a similar but much easier argument that

$$-\int_{V_i \setminus U_i^n} u^{\alpha+1} + \int_{V_i \setminus U_i^n} |\nabla u|^2 \leqq \tilde{C}_1, \tag{4}$$

where $\tilde{C}_1$ is independent of n. By multiplying (4) by $(m-2)$ and adding this to (3), we see that

$$(2m(\alpha+1)^{-1} - (m-2)) \int_{V_i \setminus U_i^n} u^{\alpha+1} \leqq \tilde{C}_2,$$

where $\tilde{C}_2$ is independent of n. Since $\alpha < (m+2)(m-2)^{-1}$, it follows that $\int_{V_i \setminus U_i^n} u^{\alpha+1}$ is bounded independent of n. Since this is true for each i and since u is uniformly bounded on $D \setminus \bigcup_{i=1}^{r} V_i$ by our earlier comments, $\int_{D_n} u^{\alpha+1}$ is uniformly bounded. We can now use a bootstrapping argument. We have that u^α is bounded in $L_{1+\alpha^{-1}}(D_n)$ uniformly in n. By results in Gilbarg and Trudinger [13] (cp also Lemma 1 in [4]), u is bounded in $L_{m(1+\alpha^{-1})}(D_n)$ where $m(p) = mp/(m-2p)$ if $p < \frac{1}{2}m$ and $m(p) = \infty$ if $p > \frac{1}{2}m$ and where the bound is independent of n Since $m(1+\alpha^{-1}) < \alpha+1$ (because $\alpha < (m+2)/(m-2)$ if $m>2$), we have obtained an estimate in a better L_p norm on D_n. By repeating this process a finite number of times, we obtain a uniform bound for u in $L^\infty(D_n)$. Note that the number of steps is independent of n. By our earlier remarks, this completes the proof. □

Remarks. (1) The proof of the uniform bound does not use that the K_i have capacity zero. However, if some of the K_i have positive capacity, the limit problem would be the problem of the solution of (1) on $D \setminus \cup K_i^*$ vanishing on the boundary in the $W^{1,2}$ sense where the K_i^* are those of the K_i which have positive capacity.

(2) In Remark 1, the requirement on the star-shapedness of the U_i^n can sometimes be weakened a little, at least if $m=2$. We merely sketch this. We only

have to bound $\dfrac{\partial u}{\partial n}$ on the part of U_i^n where $(x-y_i^n)\cdot \underline{n}<0$. First note that a result of Gidas, Ni and Nirenberg [10, p. 223], can sometimes be used to show that if x is close to part T^n of ∂U_i^n and u is a solution of (1), then u is increasing in directions normal to T^n for a distance ε where ε is independent of n. For example, this holds if, at $x_n \in \partial U_i^n$, we can find a ball $\tilde{B}_n$ with radius $\tilde{\varepsilon}$ bounded from below (and independent of x_n) such that $\tilde{B}_n$ touches ∂U_i^n at x_n and is locally (that is near x_n) inside U_i^n and such that the component of $\partial U_i^n \cap \hat{B}_n$ containing x_n does not intersect the interior of $\tilde{B}_n$. (Here $\hat{B}_n$ is the concentric ball to $\tilde{B}_n$ with twice the radius.) This, combined with Lemma 1, gives a bound for u in the normal direction to U_i^n at x_n. If we can do this locally uniformly, we locally obtain L^∞ estimates and we can sometimes combine this with $W^{2,p}$ estimates to obtain local C^1 estimates. For example, by this method, we could obtain uniform bounds if K_i is a smooth curved line segment (and U_i^n are suitable neighbourhoods)

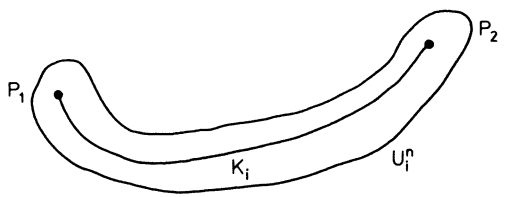


Fig. 1.

provided that there is an x_0 such that, for $i=1, 2$, the directed line segment from x_0 to p_i is the outward tangent to the curve at the endpoint p_i.

(3) The argument in the proof of Theorem 1 is related to an argument in [9] where they allow only one hole and do not try to make the estimates uniform in n.

Corollary. *Assume* $\|x_0\|<1$. *If* ε *is small and positive and* $1<\alpha<(m+2)/(m-2)$, *then* (1) *has a unique positive solution on* $\tilde{D}_\varepsilon=\{x\in R^m\colon \|x\|<1, \|x-x_0\|>\varepsilon\}$ *and vanishing on* $\partial\tilde{D}_\varepsilon$.

Proof. Since (1) has a unique positive solution on the unit ball which is non-degenerate (by [4], p. 140). Hence the result follows from Theorem 1. □

Remark. This result admits many variants. There can be several holes and the holes need not be circular (but should be star-shaped). If $m\geqq 3$, we could replace the circular hole by star shaped neighbourhoods of compact line segments in the open unit ball. Clearly we could use Theorem 1 to produce many other examples (for example when combined with [5]).

§ 2 Counterexamples

In this section, we give a number of counterexamples showing that, if we vary the assumptions slightly, then Theorem 1 fails because there can be solutions large in the sup norm. Note that the results in [4] or [6] imply that under weak assumptions, the results can only fail by the existence of solutions large

in the L^∞ norm. (More precisely if $\Omega_n \to \Omega_0$ as $n \to \infty$ in the sense of [6], then the only way we can have difficulties is if there exist positive solution u_n on Ω_n such that $\sup_{x \in \Omega_n} |u_n(x)| \to \infty$ as $n \to \infty$ (at least a subsequence).

We first show that some geometric (as distinct from topological) assumption on the U_i^n is necessary. We suppose that the assumptions of Theorem 1 hold and we assume $\bar{U}_i^{j+1} \subseteq U_i^j$ for every i, j. We now modify U_1^j by choosing a small ball B_j of radius $\varepsilon(j)$ so that $B_j \cap U_1^{j+1} = \phi$, B_j touches U_1^j from the inside at a unique point, and $\varepsilon(j) \to 0$ as $j \to \infty$. We can replace U_1^j by $U_1^j \backslash B_j$ except that we round off the corner in a smooth way (as in [4] or [5]). Note that the new hole $\tilde{U}_1^j$ which is $U_1^j \backslash B_j$ smoothed will be contractible if we do things carefully. All the assumptions of Theorem 1 hold except the star-shapedness condition. On the other hand, if we do the rounding off carefully, there exist positive solutions $\tilde{u}_n$ on $\tilde{D}_n = D \backslash \bigcup_{i=1}^{r} \tilde{U}_i^n$ such that $\|\tilde{u}_n\|_\infty \to \infty$ as $n \to \infty$. To prove this, we argue as follows. (This is similar to arguments in § 5 of [4].) On $D_n \cup B_n$ there is a positive solution V_n where

$$V_n(x) = \begin{cases} u_n(x) & \text{on } D_n \\ w_n(x) & \text{on } B_n. \end{cases}$$

Here u_n is a positive solution on D_n and w_n is the unique positive solution on B_n. By Theorem 1 and its proof, u_n is non-degenerate on D_n (at least for n large) while since B_n is a translated ball we see as in [4], p. 140 that w_n is non-degenerate. Thus by Theorem 1 in [4], there is a positive solution $\tilde{u}_n$ on $\tilde{D}_n$ close to V_n if the rounding off is small. Now, since B_n has small radius, a simple scaling argument shows that $\|w_n\|_\infty \to \infty$ as $n \to \infty$. Hence it follows that $\|\tilde{u}_n\|_\infty \to \infty$ as $n \to \infty$ provided that our rounding off is small. This gives the required example.

Before constructing more of our counterexamples, we obtain a general result on the existence of large solutions. We only prove the result if $m = 2$ though we *suspect* the same result is true if $m > 2$. We consider Lipschitz domains $\Omega_n \subseteq R^2$ such that (i) Ω_n is invariant under the map $(x, y) \to (-x, y)$, (ii) there is an $\tilde{a} > 0$ and $\varepsilon_n > 0$ such that $\varepsilon_n \to 0$ as $n \to \infty$, $(-\tilde{a}, \tilde{a}) \times (-\varepsilon_n, \varepsilon_n) \subseteq \Omega_n$, $(-\tilde{a}, \tilde{a}) \times \{-\varepsilon_n, \varepsilon_n\} \subseteq \partial\Omega_n$ and (iii) there exist $\mu_1, \mu_2 > 0$ such that the measure of $\Omega_n \backslash ((-\tilde{a}, \tilde{a}) \times (-\varepsilon, \varepsilon))$ is between μ_1 and μ_2 for all n.

Theorem 2 *Under the above assumptions there exists an n_0 such that, if $n \geqq n_0$ there is a positive solution v_n of (1) on Ω_n such that v_n is even in x and $\|v_n\|_\infty \to \infty$ as $n \to \infty$.*

Remark. v_n is concentrated in the strip $(-a, a) \times (-\varepsilon_n, \varepsilon_n)$. In fact, v_n is exponentially small off this strip when n is large.

Proof. This is really a simple variant of the proof of Theorem 3 in [5]. We merely point out the differences. We have an approximate solution u_n supported in the strip which is constructed exactly as in (5) and which is exponentially close to being a solution. Note that $\tilde{a}$ here corresponds to $b - a$ in [5]. The proof of Lemma 5 in [5] is unchanged. Before discussing Lemma 6 in [5], we

need some notation. Let $\tilde{\Omega}_n=\{\varepsilon_n^{-1}\underline{x}\colon \underline{x}\in\Omega_n\}$ where $\underline{x}=(x, y)$, let $\tilde{u}_{\varepsilon_n}(\underline{x})=\varepsilon_n^{2/(\alpha-1)}u_{\varepsilon_n}(\varepsilon_n^{-1}\underline{x})$ and let $L^2_\varepsilon(\tilde{\Omega}_n)$ denote the functions in $L^2(\tilde{\Omega}_n)$ which are even in x. Note that we have changed our notation here. We will use this notation for the rest of the section. The analogue of Lemma 6 in [5] is to prove that $H_{\varepsilon_n}=-\Delta-\alpha(\tilde{u}_{\varepsilon_n})^{\alpha-1}I$ with Dirichlet boundary conditions is invertible on $L^2_\varepsilon(\Omega_n)$ and satisfies an estimate $\|H_{\varepsilon_n}^{-1}\|_2\leqq C_5\varepsilon_n^{-2}$ where C_5 is independent of n. The proof of this is identical with Lemma 6 in [5] except at one point. We need to find a positive lower bound independent of n for the first eigenvalue of $\{(x, y)\colon (x, y)\in\Omega_n\colon x\geqq 0,\ x\geqq\frac{1}{2}(\gamma_1+\gamma_2)-\varepsilon_n$ or $|y|\geqq\varepsilon_n\}$. (γ_1 and γ_2 are defined in [5].) Since our assumptions ensure that the measure of this set is bounded above, the estimate follows from the well known result that, for given measure, the sphere has the smallest first eigenvalue or from work in Gilbarg and Trudinger [13] (cp Lemma 1 in [4]). The rest of the proof of Theorem 2 is essentially the same as the proof of Lemma 7 and Theorem 3 in [5]. □

Remarks. An example right at the end of the section shows that the above result fails if we delete the assumption of the invariance of Ω_n. Note however that the invariance could be varied slightly by replacing the reflection in x by the map $\underline{x}\to-\underline{x}$ as the symmetry. In fact, we do not really need the existence of μ_1. We also suspect that the requirement that the strip sides are exactly straight can be greatly weakened. Some weakening is certainly possible. As in [5], our methods can be easily modified to show that Theorem 2 is true if $\tilde{a}$ tends to zero as $n\to\infty$ provided that $\tilde{a}$ tends to zero sufficiently slowly (compared with ε_n). We expect but have not proved an analogue for $m>2$ where we replace the strip by $B_{\tilde{a}}\times(-\varepsilon_n, \varepsilon_n)$ and require that Ω_n is invariant under all rotations in $(x_1, \ldots, x_{m-1})$ (that is all rotations in $T=\{x\in R^m\colon x_m=0\}$). Here $B_{\tilde{a}}$ is the ball centre zero radius $\tilde{a}$ in T. The only difficulty in proving this is in proving the natural analogue of Lemma 5 in [5].

We now use Theorem 2 to construct a number of counterexamples. They are all counterexamples with $m=2$ though it seems very unlikely that the case where $m>2$ is any better behaved.

Our first counterexample shows that it is important that the holes do not get close to the boundary. We take a rectangle $[-a, a]\times[-b, b]$ except that we smooth the corners but so as to retain the symmetry. Denote this set by $\tilde{\Omega}$. We take U^n to be a similar smoothed rectangle except that it is a smoothing of $[-\frac{1}{2}a, \frac{1}{2}a]\times\left[-b+\frac{1}{n}, -b+\frac{2}{n}\right]$. Then by Theorem 2, there is a large solution of (1) on $\tilde{\Omega}\setminus U^n$ as well as the solution guaranteed by Theorem 1 in [6]. Thus the conclusion of Theorem 1 fails in this case. All the other assumptions of Theorem 1 are satisfied except that U^n approaches the boundary. (K is empty.) It might be objected that the U^n's do not have small diameter but we could avoid this by replacing $\frac{1}{2}a$ in the definition of U^n by something going to zero slowly with n. We could also use an approximation argument to replace the U^n's by strictly convex sets.

As our second example, we take D to be the unit ball in R^2 and take U_1^n and U_2^n to be 2 smoothed rectangles $[-a_n, a_n]\times\left[-\frac{2}{n}, -\frac{1}{n}\right]$ and $[-a_n, a_n]\times\left[\frac{1}{n}, \frac{2}{n}\right]$ where $a_n\to 0$ as $n\to\infty$ but rather slowly. Then all the assumptions of Theorem 1 hold except that U_1^n and U_2^n get close together. By the remarks after

Theorem 2, there is a large solution for n large concentrated in $[-a_n, a_n] \times \left[-\frac{1}{n}, \frac{1}{n}\right]$ and hence, as before, the conclusion of Theorem 1 fails.

We want to make one final remark in this section on the effect of geomery and how subtle it is. Suppose we put a hole near the boundary but now cut through so that there is an exit

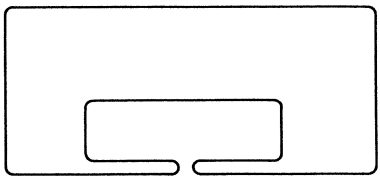


Fig. 2

from the hole to the outside nearby. See Fig. 2. In this case, if we do it carefully, then there are no large solutions. This contrasts with the results above and with the case of adding a small set on the outside. It is worth noting that the above construction does not affect the topology of Ω. We need to be a little more precise on how we do the construction. We take the smoothed rectangle $\tilde{\Omega}$ which is a smoothing of $[-a, a] \times [0, b]$ and then take $\Omega_n = \tilde{\Omega} \setminus \left(\frac{1}{4}\tilde{\Omega} + \left(0, \frac{1}{n}\right)\right)$ except that we cut through so that the symmetry of the set is preserved and so that the 'right hand tongue' T_n after cutting through is of the form $\{(x, y): f_1^n(x) \leqq y \leqq f_2^n(x): \mu_n \leqq x \leqq \frac{1}{4}b\}$ where f_1 is non-increasing and f_2 is non-decreasing. See Fig. 3.

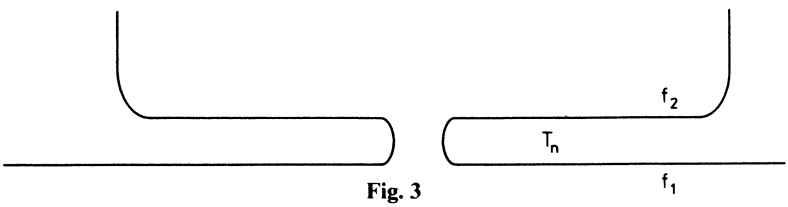


Fig. 3

We sketch very briefly the proof that there are no large solutions. By applying Gidas, Ni and Nirenberg arguments (as in [10]) purely to the tongue and its reflection in $x = \frac{1}{4}b$, we see that the maximum of a positive solution can only occur on T_n at $x = \frac{1}{4}b$. (This is a more local use of Gidas-Ni-Nirenberg ideas than usual.) Much as in the proof of Theorem 1, we see that the maximum of a large solution can only occur near the boundary and (cp [4], § 5), not near part of the boundary where the curvatures are all bounded (unless there are two parts of the boundary nearby). Thus, by our earlier remarks on the behaviour on the tongue, we see that the maximum of a large solution must occur near $x = \frac{1}{4}b$ (with an $x \geqq \frac{1}{4}b$) and near the end of the tongue. By some elementary geometry, one sees that there are no points with a $x \geqq \frac{1}{4}b$ which are close to two boundaries. Hence by our comments on curvatures of $\partial\Omega$ when a maximum occurs near the boundary, we see that there can be no large solutions. This example shows that Theorem 2 fails without the symmetry assumption. We have done this example in a special case purely for illustration. It is clear that much more general results can be proved. In particular, with care, one can allow the sides of the inside 'rectangle' to tend to zero if we

smooth its corners rather carefully. Here we have to be rather more careful and use a number of the ideas from the proof of Theorem 1 and 2 in [5] (including using a number of blowing up arguments).

§ 3 True annuli

In this section, we want to make a few remarks on the case of a true annulus $A_a=\{x\in R^m: a<\|x\|<1\}$.

It was noted by Brezis and Nirenberg [1] that for a fixed annulus and for α near (but less than) $(m+2)(m-2)^{-1}$ if $m>2$, there is a non-radially symmetric positive solution of (1). In fact, the maximizer of the natural variational principle is non-radially symmetric. Moreover, as we see below, if we fix α, then for a close to 1 the maximizer of the natural variational principle gives a non-radially symmetric positive solution of (1). Thus it is natural to ask if there is an $a(\alpha)$ in (0, 1) such that (1) has no non-symmetric positive solution on the annulus if $a<a(\alpha)$ while it has a non-symmetric positive solution if $a(\alpha)<a<1$. Moreover the Brezis and Nirenberg result suggests that $a(\alpha)\to 0$ as $\alpha\to(m+2)(m-2)^{-1}$ if $m>2$. In addition, with care, the estimates below can be done to show that on any compact subinterval C of $(1, (m+2)(m-2)^{-1})$ then there is a single $\tilde{a}$ such that we have a non-radially symmetric positive solution if $\alpha\in C$ and $a\geqq\tilde{a}$. In connection with this, it is of interest to note that, for any smooth bounded domain Ω, it can be proved that (1) has a unique non-trivial positive solution if $\alpha>1$ and α is near 1.

We sketch the proof that, if a is close to 1, there is a non-symmetric positive solution. We sketch it rather briefly because somewhat similar arguments appear in Coffman [3].

We maximize $\int_{A_a} u^{\alpha+1}$ subject to $u\in\dot{W}^{1,2}(A_a)$, $\|\nabla u\|_2=1$, $u\geqq 0$ on A_a. Standard arguments show that the solution of this after a rescaling gives a solution of (1). We let C^a denote the maximum value and C_r^a the maximum value on the subspace of symmetric functions. We will prove that $C_r^a<C^a$ if a is close to 1 and our claim of the previous paragraph then follows.

To estimate C_r^a is very easy. It is the maximum value of $c_m\int_a^1 u^{\alpha+1}r^{m-1}dr$ subject to $c_m\int_a^1 (u'(r))^2 r^{m-1}dr=1$. Here c_m is the surface area of the unit sphere. Since $a\leqq r\leqq 1$, we easily see that $C_r^a\leqq\beta\tilde{C}_r^a$ where $\tilde{C}_r^a$ is the maximum value of $\int_a^1 u^{\alpha+1}dr$ subject to $\int_a^1 (u'(r))^2 dr=1$ where in fact β tends to a constant as $a\to 1$ uniformly in α for bounded α. By the translation invariance of this new problem, we can replace $\int_a^1$ by $\int_0^{1-a}$ and then a change of variable $r=(1-a)t$ yields that $C_r^a\leqq\beta(1-a)^{\frac{1}{2}(\alpha+3)}\hat{C}$ where $\hat{C}$ is the maximum value of $\int_0^1 u^{\alpha+1}dt$ subject to $\int_0^1 u'(t)^2 dt=1$. Hence $C_r^a\leqq K(1-a)^{\frac{1}{2}(\alpha+3)}$. (In fact this estimate is asymptotically correct as $a\to 1$.)

To prove an estimate for C^a, we choose a fixed smooth positive function ϕ with support in the ball of radius $\frac{1}{2}$ and for a test function use $\tilde{\beta}\phi((1-a))^{-1}(x-\frac{1}{2}(1+a)e_1)$ where $\tilde{\beta}$ is chosen so this function satisfies the constraint. A simple but tedious computation shows that $\tilde{\beta}\sim(1-a)^{\frac{1}{2}(2-m)}$ and hence for this function $\int_{A_a} u^{\alpha+1}\sim(1-a)^{m+(\alpha+1)(1-\frac{1}{2}m)}$. By our earlier estimate for C_r^a, this is larger than C_r^a if a is close to 1 and hence our claim follows.

As Coffman noted when $m=2$, one can look for solutions with more symmetry. Indeed he used this to obtain many inequivalent solutions for a similar problem when $m=2$. This work was improved by Yan Yan Li [14].

In fact one can show that non-radially symmetric positive solutions bifurcate from radially symmetric ones as a increases to 1. The idea is to show (as suggested by our method of estimating C_r^a above) that the unique positive radial solution u_a on A_a when suitably rescaled approaches the positive solution of

$$-u''=u^\alpha, \qquad u(0)=u(1)=0 \tag{5}$$

as $a\to 1$. Moreover, the linearization of (1) at u_a on Ω_a when restricted to any irreducible representation (for the usual $O(n)$ group action) also approaches the linearization of (5) in a suitable sense as $a\to 1$. It follows easily that there is an infinite sequence $\{a_n\}_{n=1}^{\infty}$ increasing to 1 where non-symmetric positive solutions bifurcate from symmetric ones.

One final remark for this section. If we replaced the ball by a 2-dimensional rectangle Ω with the corners smoothed nicely (to preserve convexity and symmetry), we could ask the same question for $\Omega\backslash a\Omega$ for a close to 1. Once again, Theorem 2 implies that there are large solutions for a close to 1 (but less than 1). The advantage of this method is that the proof of Theorem 2 gives us a much better idea of what the solution which breaks the symmetries is like. Note that we obtain two distinct large solutions in this way by using different sides of the rectangle. In fact, we can obtain several other large solutions in this case. We could use the same ideas as in the proof of Theorem 2 to obtain a large solution which is close (after a scaling) to the Amick-Toland solution on *two* opposite sides of $\Omega\backslash a\Omega$ (and with a Z_2 symmetry). In addition, there is a large solution with a $Z_2\oplus Z_2$ symmetry (which is near the scaled Amick-Toland solution on 4 sides).

§ 4 More general nonlinearities

Finally we make some very brief remarks on how our assumptions on the type of nonlinearity can be weakened for some domains if there is only one hole. We replace u^α by $f(u)$ where f is C^1. The Pohozaev identity argument in the proof of Theorem 1 implies that we obtain a bound for $\int_{D_n} 2mF(u)-(m-2)uf(u)$ $\left(\text{where } F(t)=\int_0^t f(s)\,ds\right)$ provided that we have a bound for $\frac{\partial u}{\partial n}$ on ∂D. (We use the Pohozaev identity on $D\backslash U_1^n$.) Hence we easily see that, if $2mF(t)-(m-2)tf(t)\geqq\mu tf(t)$ for large positive t where $\mu>0$ and if $f(t)\leqq Kt^r$ for large positive t where $r<(m+2)(m-2)^{-1}$, we obtain uniform bounds in

L^∞ for positive solutions provided we have a uniform bound for $\frac{\partial u}{\partial n}$ on ∂D. (The Pohozaev identity gives a starting estimate and we bootstrap as in the proof of Theorem 1.) We can bound $\frac{\partial u}{\partial n}$ on ∂D provided that $t^{-1}f(t)\to\infty$ as $t\to\infty$ and provided that $m=2$ and f is non-negative or D has strong convexity properties (as in [9], step 2 of the proof of Theorem 1.1). To do this, we first multiply the equation by the first eigenfunction ϕ_1^n of $-\Delta$ on D_n to deduce a bound for $\int_{D_n} f(u)\,\phi_1^n$ and then use a variant of Gidas, Ni and Nirenberg ideas (as in [9]) to bound u near ∂D. (Here we use that ϕ_1^n converges in C^1 close to ∂D.) Finally, standard regularity theory ensures that $\frac{\partial u}{\partial n}$ is bounded close to ∂D. Thus, if we make the above assumptions on f and D, we have uniform bounds. In particular, the natural analogue of Theorem 1 holds for one hole in *this case* provided that $f(0)\geqq 0$. (We need this to ensure that positive solutions perturb to positive solutions.) Note that this argument is closely related to those in [9].

Alternatively, we could allow more general domain perturbations if we make stronger assumptions on f. The idea is to prove a variant of Lemma 1 by using again the ideas of Gidas and Spruck [11] (especially the proof of Theorem 6.1). The simplest result of this type is that the conclusion of Lemma 1 holds if f is non-negative and C^1 and if there exist A, $B>0$ and $q\in(1, m(m-2)^{-1})$ such that $A y^q\leqq f(y)\leqq B y^q$ for large positive y.

Note that our results do depend to some extent on the nonlinearity. In two dimensions, the problem where we replace y^p by λe^y behaves differently. In this case Lin [15] proved that on any annulus $\{x\in R^2\colon a<\|x\|<1\}$ there is a point $\tilde{\lambda}_a>0$ where non-radially symmetric positive solutions bifurcate from radially symmetric ones. This quite different from our case. The results in [15] could be improved by using the theory in § 5 of [8] to show that there is a non-radially symmetric positive solution if $0<\lambda<\tilde{\lambda}_a$. In fact there is an unbounded branch of non-symmetric positive solutions bifurcating from the symmetric ones at $\lambda=\lambda_a$. The proof of this uses that there is an apriori bound for positive solutions for fixed λ. Since radially symmetric solutions approach those on the ball (by the Propos. in [6] and by [15]), we easily see that $\tilde{\lambda}_a\to 0$ as $a\to 0$. Indeed, it is possible to show that sup $\{\lambda$: there exists a non-symmetric solution$\}$ tends to zero as $a\to 0$.

References

1. Brezis, H., Nirenberg, L.: Positive solutions of nonlinear elliptic equations involving critical Sobolev exponents. Comm. Pure App. Math. **36**, 437–444 (1983)
2. Cafarelli, L.A., Gidas, B., Spruck, J.: Asymptotic symmetry and local behaviour of semilinear elliptic equations with Sobolev exponent. Comm. Pure App. Math. **42**, 271–297 (1989)
3. Coffman, C.V.: A nonlinear boundary value problem with many positive solutions. J. Diff. Equations **54**, 429–437 (1984)
4. Dancer, E.N.: The effect of domain shape on the number of positive solutions of certain nonlinear equations. J. Diff. Equations **74**, 120–156 (1988)
5. Dancer, E.N.: On the influence of domain shape on the existence of large solutions of some superlinear problems. Math. Ann. **285**, 647–669 (1990)

6. Dancer, E.N.: A note on an equation with critical exponent. Bull. Lond. Math. Soc. **20**, 600–602 (1988)
7. Dancer, E.N.: Certain weakly nonlinear partial differential equations on annuli. Nonlinear Anal. **14**, 441–442 (1990)
8. Dancer, E.N.: A new degree for S^1-invariant gradient mappings and applications. Analyse Nonlineaire **2**, 329–370 (1985)
9. Figueredo, de, D.G., Lions, P.L., Nussbaum, R.: A priori estimates and existence of positive solutions of semilinear elliptic equations. J. Math. Pure et Appliquées **61**, 41–63 (1982)
10. Gidas, B., Ni, W., Nirenberg, L.: Symmetry and related properties by the maximum principle. Commun. Math. Phys. **68**, 209–243 (1979)
11. Gidas, B., Spruck, J.: Global and local behaviour of positive solutions of nonlinear elliptic equations. Commun. Pure Appl. Math. **34**, 525–598 (1981)
12. Gidas, B., Spruck, J.: A priori bounds for positive solutions of nonlinear elliptic equations. Commun. Partial Differ. Equations **6**, 883–901 (1981)
13. Gilbarg, D., Trudinger, N.: Elliptic partial differential equations of second order. Berlin Heidelberg New York: Springer 1977
14. Li, Yan Yan: Existence of many positive solutions of semilinear elliptic equations on annulus. J. Differ. Equations **83**, 348–367 (1990)
15. Lin, S.S.: On non radially symmetric bifurcation in the annulus. J. Differ. Equations **80**, 251–279 (1989)
16. Ni, W., Nussbaum, R.: Uniqueness and nonuniqueness for positive radial solutions of $\Delta u + f(u, r) = 0$. Commun. Pure Appl. Math. **38**, 67–108 (1985)

Math. Z. 206, 563–566 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Index and stability of harmonic Gauss maps

Bennett Palmer

Fachbereich 3 Mathematik, Technische Universität Berlin, Strasse des 17. Juni 135, W-1000 Berlin 12, Federal Republic of Germany

Received July 6, 1990

By a well-known result of Ruh [R] an immersed surface $M \in \mathbb{E}^3$ has constant mean curvature h if and only if its Gauss map

$$v: M \to S^2$$

is a harmonic map. Harmonic maps being critica of the energy functional, it is natural to attempt to relate the second variation of energy to the geometry of M when $h = \text{constant}$. We show

Theorem I *Let $M \subset \mathbb{E}^3$ be a compact surface of constant mean curvature and genus g. Then Index $(v) \geqq$ (largest integer $< 1 + g/2$).*

1 Preliminaries

The second variation of energy for harmonic maps was given by R.T. Smith [Sm], Here we recompute the second variation ina manner more suited to the present situation. The emphasis is on utilizing harmonic vector fields on M which may be considered as sections of the pull back bundle $v^* TS^2$ using the parallelism of $\mathbb{E}^3$. Let

$$X: M \to \mathbb{E}^3$$

be an immersion of a surface with constant mean curvature $h \neq 0$. Let

$$v: M \to S^2$$

be the Gauss map of M. The notation $\Omega \Subset M$ will indicate that Ω is a relatively compact subdomain of M with smooth boundary $\partial \Omega$. Let $\Gamma_c(v^* TS^2, \Omega)$ denote the space of smooth sections of the pull-back bundle $v^* TS^2$ with compact support in Ω.

Consider a variation of v of the form

$$v_\varepsilon = v + \varepsilon \xi + O(\varepsilon^2), \qquad \xi \in \Gamma_c(v^* TS^2, \Omega).$$

The energy of $v|_\Omega$ is given by

$$E = {}^1/_2 \int_\Omega dv \wedge *dv$$

where $*$ is the Hodge operator of M. Letting "$\cdot$" denote differentiation w.r.t. ε, we have the first variation

$$D_\xi E = \int d\dot{v} \wedge *dv$$

and the second variation

$$\begin{aligned} D_\xi^2 E &= \int (d\dot{v} \wedge *d\dot{v} + d\ddot{v} \wedge *dv \\ &= \int (d\dot{v} \wedge *d\dot{v} - \langle \ddot{v}, d*dv \rangle) \\ &= \int \|d\xi\|^2 - \langle \ddot{v}, \Delta v \rangle * 1. \end{aligned}$$

Using the fact that v is harmonic and hence satisfies

$$\Delta v + \|dv\|^2 v = 0$$

we obtain

$$D_\xi^2 E = \int \|d\xi\|^2 + \langle \ddot{v}, v \rangle \|dv\|^2 * 1$$

Since $\langle \ddot{v}, v \rangle = -\|\dot{v}\|^2 = -\|\xi\|^2$ one has

$$D_\xi^2 = \int_\Omega \|d\xi\|^2 - \|dv\|^2 \|\xi\|^2 * 1.$$

Def. $v|_\Omega$ is stable iff

$$D_\xi^2 E > 0$$

for all non-trivial $\xi \in \Gamma_c(v^* TS^2, \Omega)$. We say v is unstable if $v|_\Omega$ is not stable for some $\Omega \Subset M$.

Definition. The index of $v|_\Omega$, $\mathrm{Index}_\Omega(v)$, is the dimension of the largest subspace of $\Gamma_c(v^* TS^2, \Omega)$ on which

$$\xi \mapsto D_\xi^2 E$$

is negative definite. When M in compact we set $\mathrm{Index}(v) = \mathrm{Index}_M(v)$.

Observe that there is a natural inclusion of $\Gamma_c(TM, \Omega)$, the space of section of TM with support in Ω in $\Gamma_c(v^* TS^2, \Omega)$. For $\xi \in \Gamma_c(TM, \Omega)$, the Weingarten equation gives

$$d\xi = \nabla \xi - \langle \xi, dv \rangle v \tag{1.1}$$

where ∇ is the connection of TM. The two terms in the right hand side of (1.1) are orthogonal so that

$$\|d\xi\|^2 = \|\nabla \xi\|^2 + \|dv(\xi)\|^2.$$

By a well-known formula ([S], pg. 88, Propos. 6)

$$\|dv(\xi)\|^2 = -2h\langle dv(\xi), \xi \rangle - \kappa \|\xi\|^2$$

where κ is the curvature of M. Therefore

$$D_\xi^2 E = \int_\Omega \|\nabla \xi\|^2 - 2h\langle dv(\xi), \xi\rangle + \kappa\|\xi\|^2 - 4h^2\|\xi\|^2 * 1 \tag{1.2}$$

since $\|dv\|^2 = 4h^2 - 2\kappa$. Finally, using integration by parts we express this as

$$D_\xi^2 E = -\int_\Omega \langle \xi, J(\xi)\rangle * 1 \tag{1.3}$$

where J is the Jacobi operator

$$J(\xi) \equiv \text{tr. } \nabla\nabla\xi - \kappa\xi + 2h\,dv(\xi) + 4h^2\xi. \tag{1.4}$$

Here tr. denotes the trace taken with respect to the induced metric. Note that the operator

$$\Box\xi \equiv \text{tr. } \nabla\nabla\xi - \kappa\xi \tag{1.5}$$

is the Laplacian acting on vector fields to that

$$J = \Box + 2h\,dv + 4h^2. \tag{1.6}$$

(pf. of Theor. I). Let $H \in T(M)$ denote the space of harmonic vector fields on M. Then H is a $2g$ dimensional vector space equipped with an anti-involution

$$\xi \to \xi^\perp := \tau(\xi)$$

where τ denotes the almost complex structure of M. Note that $\|\xi\| = \|\xi^\perp\|$ holds.

Note that J is a strongly elliptic, self-adjoint operator acting on $\Gamma T(M)$. Denote its eigenvalues by $\lambda_1 \leqq \lambda_2 \leqq \ldots$.

Let $Y_1, Y_2, \ldots, Y_{m-1}$ be arbitrary elements in $\Gamma T(M)$. For $g > 2m-2$, there exists a non-trivial $\xi \in H$ satisfying $2m-2$ linear conditions

$$\int\langle \xi, Y_j\rangle = 0 = \int\langle \xi^\perp, Y_j\rangle; \quad j = 1, \ldots, m-1.$$

Since

$$J(\xi) = 2h\,dv(\xi) + 4h^2\xi$$
$$J(\xi^\perp) = 2h\,dv(\xi^\perp) + 4h^2\xi^\perp,$$

one has

$$D_\xi^2 E = -\int 2h\langle dv(\xi), \xi\rangle + 4h^2\|\xi\|^2 * 1$$
$$D_{\xi^\perp}^2 E = -\int 2h\langle dv(\xi^\perp), \xi^\perp\rangle + 4h^2\|\xi^\perp\|^2 * 1$$

and hence one finds by adding the two previous lines and using the definition of h,

$$\begin{aligned} D_\xi^2 E + D_{\xi^\perp}^2 E &= -\int(-4h^2\|\xi\| + 8h^2\|\xi\|^2) * 1 \\ &= -\int 4h^2\|\xi\|^2 * 1 < 0. \end{aligned}$$

It follows that at least one of $D_\xi^2 E$ or $D_{\xi^\perp}^2 E$ is negative.

Since λ_m is characterized by

$$\lambda_m = \sup_{\{Y_1,\ldots,Y_{m-1}\}\subset\Gamma(TM)} \inf_{\eta\perp\{Y_1,\ldots,Y_{m-1}\}} -\frac{\int\langle\eta, J(\eta)\rangle}{\int\|\eta\|^2}$$

one finds that λ_m is negative and the theorem follows.

Remark. Let $M\subset\mathbb{E}^3$ be a complete surface of constant non-zero mean curvature. By multiplying ξ by a suitable cut-off function and using the method of Theorem I, we can show that the index of v is bounded below by the largest integer less than $1+\gamma/2$ where 2γ denotes the dimension of the space of L^2 harmonic vector fields on M. Details along with results for surfaces in the three sphere will be given elsewhere.

Remark. Recall that surfaces of constant mean curvature $\hbar$ in $\mathbb{E}^3$ are derived from a variational principle

$$\delta J=0$$

where

$$J=\text{area}+2h \quad \text{(enclosed 3-volume)}.$$

For a variation of the form

$$X+\varepsilon\psi v+O(\varepsilon^2), \qquad \psi\in C_c^\infty(M)$$

the second variation of J is given by

$$\delta^2 J=-\int\psi L\psi$$

where

$$L\psi=\Delta\psi+(4h^2-2\mathscr{K})\psi.$$

Following Barbosa and DoCarmo [B, DC] we say a domain $\Omega\Subset M$ is stable iff $\delta^2 J\geqq 0$ for all nontrivial $\psi\in C_c^\infty(M)$ with

$$\int\psi=0.$$

A natural question is whether there is any connection between the stability of Ω and the stability of $v|_\Omega$. In this direction note that if $\xi\in\Gamma_c(v^*TS^2)$ then Kato's inequality gives

$$\|\nabla\|\xi\|\|^2\leqq\|d\xi\|^2$$

and therefore

$$\int(\|\nabla\|\xi\|\|^2-(4h^2-2\mathscr{K})\|\xi\|^2)*1\leqq\int\|d\xi\|^2-\|dv\|^2\|\xi\|^2*1.$$

It follows that if $\delta^2 J\geqq 0$ for all $\psi\in C_c^\infty$ then $v|_\Omega$ is stable.

References

[B, DC] Barbosa, J.L., DoCormo, M.: Stability of hyper surfaces with constant mean curvature. Math. Z. **185**, 339–353 (1984)

[R] Ruh, E.A.: Asymptotic behavior of minimal hypersurfaces. J. Differ. Geom. **4**, 509–513 (1970)

[Sm] Smith, R.T.: The second variation formula for harmonic maps. Proc. A.M.S. **47**, 229–236 (1975)

Math. Z. 206, 567–575 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

A note on extensions of algebraic and formal groups I

Tsutomu Sekiguchi [1,*] **and Noriyuki Suwa** [2,**]

[1] Department of Mathematics, Faculty of Science and Engineering, Chuo University, Kasuga 1-13-27, Bunkyo-ku, Tokyo 112, Japan
[2] Department of Mathematics, Tokyo Denki University, Kanda-nishiki-cho 2-2, Chiyoda-ku, Tokyo 101, Japan

Received April 4, 1990; in final form May 3, 1990

Introduction

Let A be a commutative ring. It is well known that $\mathrm{Ext}^1_A(\mathbb{G}_{m,A}, \mathbb{G}_{a,A})=0$ if A is a field and $\mathrm{Ext}^1_A(\mathbb{G}^\wedge_{a,A}, \mathbb{G}^\wedge_{m,A})=0$ if A is a perfect field. However, there may exist a non-trivial extension of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$ or of $\mathbb{G}^\wedge_{a,A}$ by $\mathbb{G}^\wedge_{m,A}$ if A has a nilpotent element. In this article, we give a very explicit description of $\mathrm{Ext}^1_A(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ and $\mathrm{Ext}^1_A(\mathbb{G}^\wedge_{a,A}, \mathbb{G}^\wedge_{m,A})$ when A is of characteristic $p>0$ as follows:

Let A be an $\mathbb{F}_p$-algebra. Then $\mathfrak{a}\mapsto F_p(\mathfrak{a}; X, Y)$ induces bijective homomorphisms $W(A)/F \xrightarrow{\sim} H^2_0(\mathbb{G}^\wedge_{a,A}, \mathbb{G}^\wedge_{m,A})$ and $W^\wedge(A)/F \xrightarrow{\sim} H^2_0(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ (Theorem 2.5.1. For the notation, see Sect. 2. It is more or less well known that $H^2_0(\mathbb{G}_{m,A}, \mathbb{G}_{a,A})=0$ and $H^2_0(\mathbb{G}^\wedge_{a,A}, \mathbb{G}^\wedge_{m,A})=0$ if A is of characteristic 0.) After a short review on Witt vectors and the Artin-Hasse exponential, we mention and prove the main theorem. We restrict ourselves to the case when A is an $\mathbb{F}_p$-algebra for simplicity though our method can be applied to the case when A is $\mathbb{Z}/p^n$-algebra.

Notation

Throughout the article, p denotes a prime number and A a $\mathbb{Z}_{(p)}$-algebra.

$\mathbb{G}_{a,A}$: the additive group scheme over A
$\mathbb{G}_{m,A}$: the multiplicative group scheme over A
W_A: the group scheme of Witt vectors over A
$\mathbb{G}^\wedge_{a,A}$: the additive formal group over A
$\mathbb{G}^\wedge_{m,A}$: the multiplicative formal group over A
$W^\wedge_A$: the formal group of Witt vectors over A

* Partially supported by Grant-in-Aid for Scientific Research #62540066
** Partially supported by Grant-in-Aid for Scientific Research #01540077

$\mathrm{Ext}^1_A(\mathbb{G}^{\wedge}_{a,A}, \mathbb{G}^{\wedge}_{m,A})$ or $\mathrm{Ext}^1_A(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ denotes the group of classes of *commutative* extensions of $\mathbb{G}^{\wedge}_{a,A}$ by $\mathbb{G}^{\wedge}_{m,A}$ or of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$, respectively.

For a commutative ring B, $B^{\times}$ denotes the multiplicative group $\mathbb{G}_m(B)$.

1 Recall: Witt vectors and the Artin-Hasse exponential

We recall here necessary facts on Witt vectors and the Artin-Hasse exponential. For details, see [DG, Chapt. V] or [HZ, Chapt. III].

1.1 Let $W(A)$ be the ring of Witt vectors over A. V denotes the additive endomorphism of $W(A)$, defined by $(a_0, a_1, \ldots) \mapsto (0, a_0, a_1, \ldots)$ and called the Verschiebung map. F denotes the map defined by $(a_0, a_1, \ldots) \mapsto (a_0^p, a_1^p, \ldots)$ and called the Frobenius map. F is a ring endomorphism of $W(A)$ and $FV = VF = p$, if A is of characteristic p. We denote often $F\mathfrak{a}$ by $\mathfrak{a}^{(p)}$.

We denote by $W^{\wedge}(A)$ the subset of $W(A)$ formed by the Witt vectors $\mathfrak{a} = (a_r)_{r \geqq 0}$ such that a_r is nilpotent for all r and $a_r = 0$ for almost all r (the A-valued points of the formal group of Witt vectors $W_A^{\wedge}$). It is easily seen that $W^{\wedge}(A)$ is an ideal of $W(A)$, stable under V and F.

1.2 Recall now the equality:

$$\exp(T) = \prod_{n \geqq 1} (1 - T^n)^{-\mu(n)/n} \in \mathbb{Q}[[T]]$$

and the definition of the Artin-Hasse exponential:

$$E_p(T) = \prod_{(n,p)=1} (1 - T^n)^{-\mu(n)/n} \in \mathbb{Z}_{(p)}[[T]].$$

Let $F(T) \in A[[T]]$ with $F(T) \equiv 1 \bmod \deg 1$. Then $F(T)$ is uniquely factorized in the form:

$$F(T) = \prod_{k \geqq 1} E_p(c_k T^k)$$

in $A[[T]]$.

Put $E_p(\mathfrak{a}; T) = \prod_{r \geqq 0} E_p(a_r T^{p^r})$ for $\mathfrak{a} \in W(A)$. Then $\mathfrak{a} \mapsto E_p(\mathfrak{a}; T)$ gives rise to injective homomorphisms $W(A) \to A[[T]]^{\times}$ and $W^{\wedge}(A) \to A[T]^{\times}$. The commutative square

$$\begin{array}{ccc} W^{\wedge}(A) & \xrightarrow{E_p(\ ;T)} & A[T]^{\times} \\ \cap & & \cap \\ W(A) & \xrightarrow{E_p(\ ;T)} & A[[T]]^{\times} \end{array}$$

is cartesian by the following lemma.

Lemma 1.3 *Let A be a $\mathbb{Z}_{(p)}$-algebra and $F(T) \in A[T]^{\times}$ with $F(0) = 1$. Put $F(T) = \prod_{k \geqq 1} E_p(c_k T^k)$. Then c_k is nilpotent for all k, and $c_k = 0$ for almost all k.*

Proof. Let $\mathfrak{a}$ be the ideal of A generated by the coefficients of $T^k (k \geqq 1)$ of $F(T)$ and $d = \deg F(T)$. Then $\mathfrak{a}$ is nilpotent, and $c_k \in \mathfrak{a}^s$ if $(s-1)d < k \leqq sd$.

2 Statement of the theorem

First we recall Hochschild cohomology groups. For details, see [DG, Chapt. II.5 and Chapt. III.6].

2.1 Let A be a $\mathbb{Z}_{(p)}$-algebra and $G(X, Y) \in A[[X, Y]]^{\times}$ (resp. $A[X, Y]^{\times}$). $G(X, Y)$ is called a symmetric 2-cocycle of $\mathbb{G}_{a,A}^{\wedge}$ (resp. $\mathbb{G}_{a,A}$) with coefficients in $\mathbb{G}_{m,A}^{\wedge}$ (resp. $\mathbb{G}_{m,A}$) if $G(X, Y)$ satisfies the following functional equations:

1) $G(X+Y, Z)\, G(X, Y) = G(X, Y+Z)\, G(Y, Z)$
2) $G(X, Y) = G(Y, X)$.

We denote by $Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ (resp. $Z^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$) the subgroup of $A[[X, Y]]^{\times}$ (resp. $A[X, Y]^{\times}$) formed by the symmetric 2-cocycles of $\mathbb{G}_{a,A}^{\wedge}$ (resp. $\mathbb{G}_{a,A}$) with coefficients in $\mathbb{G}_{m,A}^{\wedge}$ (resp. $\mathbb{G}_{m,A}$).

Let $F(T) \in A[[T]]^{\times}$ (resp. $A[T]^{\times}$). It is clear that

$$F(X)\, F(Y)\, F(X+Y)^{-1} \in Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge}) \text{ (resp. } Z^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})).$$

We denote by $B^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ (resp. $B^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$) the subgroup of $Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ (resp. $Z^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$) of the symmetric 2-cocycles of the form $F(X)\, F(Y)\, F(X+Y)^{-1}$. Put

$$H_0^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge}) = Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge}) / B^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$$

and

$$H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) = Z^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) / B^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}).$$

It is well known that $H_0^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ (resp. $H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$) is isomorphic to the group of classes of commutative extensions of $\mathbb{G}_{a,A}^{\wedge}$ by $\mathbb{G}_{m,A}^{\wedge}$ (resp. of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$), which split as extensions of formal A-schemes (resp. A-schemes).

Now we define symmetric 2-cocycles of $\mathbb{G}_{a,A}^{\wedge}$ with coefficients in $\mathbb{G}_{m,A}^{\wedge}$, using the Artin-Hasse exponential.

2.2 Noting that

$$E_p(T) = \exp(T) \cdot E_p(T^p)^{1/p} = \exp\Big(\sum_{i \geq 0} T^{p^i}/p^i\Big)$$

in $\mathbb{Q}[[T]]$, we obtain

$$\begin{aligned} &E_p(X)\, E_p(Y)\, E_p(X+Y)^{-1} \\ &\quad = [E_p(X^p)\, E_p(Y^p)\, E_p((X+Y)^p)^{-1}]^{1/p} \\ &\quad = \exp\left(\sum_{i \geq 1} \frac{X^{p^i} + Y^{p^i} - (X+Y)^{p^i}}{p^i}\right) \end{aligned}$$

in $\mathbb{Z}_{(p)}[[X, Y]]$.

Define now a formal power series

$$F_p(U; X, Y) = \exp\left(\sum_{i \geq 1} U^{p^i - 1} \frac{X^{p^i} + Y^{p^i} - (X+Y)^{p^i}}{p^i}\right) \in \mathbb{Z}_{(p)}[U][[X, Y]].$$

By the definition, we have immediately the following assertions.

Lemma 2.3

(1) $F_p(U; X, Y) \in Z^2(\mathbb{G}_{a,\mathbb{Z}_{(p)}[U]}^{\wedge}, \mathbb{G}_{m,\mathbb{Z}_{(p)}[U]}^{\wedge})$.
(2) $F_p(U^p; X, Y) \in B^2(\mathbb{G}_{a,\mathbb{Z}_{(p)}[U]}^{\wedge}, \mathbb{G}_{m,\mathbb{Z}_{(p)}[U]}^{\wedge})$.

Corollary 2.4 *Let A be a $\mathbb{Z}_{(p)}$-algebra and $a \in A$.*
(1) $F_p(a; X, Y) \in Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ *and* $F_p(a^p; X, Y) \in B^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$.
(2) *If a is nilpotent,* $F_p(a; X, Y) \in Z^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ *and* $F_p(a^p; X, Y) \in B^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$.

2.5 Assume now that A is an $\mathbb{F}_p$-algebra. Let $\mathfrak{a} = (a_r)_{r \geq 0} \in W(A)$. Define a formal power series

$$F_p(\mathfrak{a}; X, Y) = \prod_{r \geq 0} F_p(a_r; X^{p^r}, Y^{p^r}) \in A[[X, Y]].$$

Then:
(1) $F_p(\mathfrak{a}; X, Y) \in Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ and $F_p(\mathfrak{a}^{(p)}; X, Y) \in B^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$;
(2) if $\mathfrak{a} \in W^{\wedge}(A)$, $F_p(\mathfrak{a}; X, Y) \in Z^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ and $F_p(\mathfrak{a}^{(p)}; X, Y) \in B^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$.
Hence $\mathfrak{a} \mapsto F_p(\mathfrak{a}; X, Y)$ induces homomorphisms $W(A)/F \to H_0^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ and $W^{\wedge}(A)/F \to H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$.

Under these notations we can state our main theorem:

Theorem 2.5.1 *Let A be an $\mathbb{F}_p$-algebra. Then $\mathfrak{a} \mapsto F_p(\mathfrak{a}; X, Y)$ induces bijective homomorphisms $W(A)/F \xrightarrow{\sim} H_0^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ and $W^{\wedge}(A)/F \xrightarrow{\sim} H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$.*

3 Proof of the theorem

Lemma 3.1 *Let A be an $\mathbb{F}_p$-algebra and $G(X, Y) \in Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge}) \subset A[[X, Y]]^{\times}$. Then there exist a flat A-algebra B and $F(T) \in B[[T]]$ such that $F(X)\,F(Y)\,F(X+Y)^{-1} = G(X, Y)$.*

Proof. Dividing $G(X, Y)$ by its constant term, we may assume that $G(X, Y) \equiv 1$ mod. deg. 1.

Assume that there exist a flat A-algebra B and $F(T) \in B[[T]]^{\times}$ such that $F(X)\,F(Y)\,F(X+Y)^{-1} \equiv G(X, Y)$ mod. deg. r. Let $H(X, Y)$ denote the homogeneous component of degree r of $[F(X)\,F(Y)\,F(X+Y)^{-1}]^{-1}\,G(X, Y)$. Noting that $[F(X)\,F(Y)\,F(X+Y)^{-1}]^{-1}\,G(X, Y) \in Z^2(\mathbb{G}_{a,B}^{\wedge}, \mathbb{G}_{m,B}^{\wedge})$, we see that $H(X, Y)$ satisfies the functional equation:
1) $H(X+Y, Z) + H(X, Y) = H(X, Y+Z) + H(Y, Z)$;
2) $H(X, Y) = H(Y, X)$.
By Lazard's comparison lemma ([1], Lemme 3), there exists $a \in B$ such that

$$H(X, Y) = \begin{cases} a\{X^r + Y^r - (X+Y)^r\} & \text{if } r \text{ is not a power of } p \\ a\dfrac{X^r + Y^r - (X+Y)^r}{p} & \text{if } r \text{ is a power of } p. \end{cases}$$

(1) When r is not a power of p, put $E(T) = 1 + aT^r$. Then

$$E(X)\,E(Y)\,E(X+Y)^{-1} \equiv 1 + a\{X^r + Y^r - (X+Y)^r\} \text{ mod. deg. } r+1.$$

(2) When r is a power of p, put $b = \sqrt[p]{a} \in B[\sqrt[p]{a}] = B[T]/(T^p - a)$ and $E(T) = \sum_{i=0}^{p-1} (bT^{r/p})^i / i!$. Then

$$E(X)\,E(Y)\,E(X+Y)^{-1} \equiv 1 + a\frac{X^r + Y^r - (X+Y)^r}{p} \text{ mod. deg. } r+1.$$

In the both cases, we have

$$[F(X)\,E(X)][E(Y)\,F(Y)][F(X+Y)\,E(X+Y)]^{-1}\equiv G(X,Y)\bmod.\deg.r+1.$$

Replacing $F(T)$ by $F(T)\,E(T)$ and B by $B[\sqrt[p]{a}]$ (when r is a power of p), we may assume that there exist a flat A-algebra B and $F(T)\in B[[T]]^{\times}$ such that

$$[F(X)\,F(Y)\,F(X+Y)^{-1}]^{-1}\,G(X,Y)\equiv 1\bmod.\deg.r+1.$$

Continuing this process, we find a flat A-algebra B and $F(T)\in B[[T]]^{\times}$ such that $F(X)\,F(Y)\,F(X+Y)^{-1}=G(X,Y)$.

Lemma 3.2 *Let $B\supset A$ be $\mathbb{F}_p$-algebras and $F(T)\in B[[T]]^{\times}$. If $F(X)\,F(Y)\cdot F(X+Y)^{-1}\in A[[X,Y]]$, there exist $\mathfrak{b}\in W(B)$ and $G(T)\in A[[T]]$ such that $F(T)=E_p(\mathfrak{b},T)\,G(T)$ and $\mathfrak{b}^{(p)}\in W(A)$.*

Proof. Dividing $F(T)$ by its constant term, we may assume that $F(T)\equiv 1\bmod.\deg.1$. Put

$$F(T)=\prod_{k=1}^{\infty}E_p(c_k\,T^k)$$

and

$$F_r(T)=\prod_{\mathrm{ord}_p k\geqq r}E_p(c_k\,T^{k/p^r}),$$

$$G_r(T)=\prod_{\mathrm{ord}_p k=r}E_p(c_k\,T^{k/p^r})\quad\text{for } r\geqq 0,\text{ and } G_{-1}(T)=1.$$

Then $F_r(T)=\prod_{j\geqq r}G_j(T^{p^{j-r}})$. In particular, $F_0(T)=F(T)$.

We prove that

(S_r): (1) $c_{p^r}^p\in A$,
(2) $c_k\in A$ if $\mathrm{ord}_p k=r$ and $k\neq p^r$;

and

(T_r): (1) $G_r(X)\,G_r(Y)\,G_r(X+Y)^{-1}\in A[[X,Y]]$,
(2) $F_{r+1}(X)\,F_{r+1}(Y)\,F_{r+1}(X+Y)^{-1}\in A[[X,Y]]$;

by the induction on $r\geqq 0$.

If (S_r) holds true for each $r\geqq 0$, then $\mathfrak{b}=(c_{p^r})_{r\geqq 0}\in W(B)$ and $G(T)=\prod_{k:\text{not a power of }p}E_p(c_k\,T)$ satisfy the required condition.

First we verify that (S_r) and (T_{r-1}) imply (T_r). Indeed, by (S_r),

$$E_p(c_{p^r}X)\,E_p(c_{p^r}Y)\,E_p(c_{p^r}(X+Y))^{-1}=F_p(c_{p^r}^p;X,Y)\in A[[X,Y]]$$

and

$$G_r(T)/E_p(c_{p^r}T)=\prod_{\substack{\mathrm{ord}_p k=r\\ k\neq p^r}}E_p(c_k\,T^{k/p^r})\in A[[T]].$$

Hence we obtain

$$G_r(X)\,G_r(Y)\,G_r(X+Y)^{-1}\in A[[X,Y]].$$

Moreover,

$$F_{r+1}(T^p)=F_r(T)/G_r(T),$$

and by the assumption,

$$F_r(X)\,F_r(Y)\,F_r(X+Y)^{-1}\in A[[X,Y]].$$

Hence

$$F_{r+1}(X^p)\,F_{r+1}(Y^p)\,F_{r+1}(X^p+Y^p)^{-1}\in A[[X,Y]],$$

and therefore,

$$F_{r+1}(X)\,F_{r+1}(Y)\,F_{r+1}(X+Y)^{-1}\in A[[X,Y]].$$

Now we verify that (T_{r-1}) (2) implies (S_r). Note that

$$\frac{\partial}{\partial X}\left[\frac{F_r(X)\,F_r(Y)}{F_r(X+Y)}\right]=\left[\frac{F_r'(X)}{F_r(X)}-\frac{F_r'(X+Y)}{F_r(X+Y)}\right]\frac{F_r(X)\,F_r(Y)}{F_r(X+Y)}.$$

Hence, by the assumption,

$$\frac{F_r'(X)}{F_r(X)}-\frac{F_r'(X+Y)}{F_r(X+Y)}\in A[[X,Y]].$$

Note that

$$\begin{aligned}\frac{F_r'(T)}{F_r(T)}&=\frac{G_r'(T)}{G_r(T)}\\&=\sum_{\mathrm{ord}_p k=r}\frac{\{E_p(c_k\,T^{k/p^r})\}'}{E_p(c_k\,T^{k/p^r})}\\&=\sum_{\mathrm{ord}_p k=r}\sum_{i=0}^{\infty}\frac{k}{p^r}\,c_k^{p^i}\,T^{p^{i-r}k-1}.\end{aligned}$$

Looking at the coefficient of Y^{k/p^r-1} in

$$\frac{F_r'(X)}{F_r(X)}-\frac{F_r'(X+Y)}{F_r(X+Y)}=\frac{G_r'(X)}{G_r(X)}-\frac{G_r'(X+Y)}{G_r(X+Y)},$$

we see that $\dfrac{k}{p^r}\,c_k\in A$, and therefore $c_k\in A$, if $\mathrm{ord}_p k=r$ and $k\neq p^r$. Moreover,

$$\begin{aligned}&F_r(X)\,F_r(Y)\,F_r(X+Y)^{-1}\\&\quad=F_p(c_{p^r}^p;X,Y)\times\prod_{\substack{\mathrm{ord}_p k\geqq r\\ k\neq p^r}}[E_p(c_k\,X^{k/p^r})\,E_p(c_k\,Y^{k/p^r})\,E_p(c_k(X+Y)^{k/p^r})^{-1}].\end{aligned}$$

Looking at the coefficient of $X^{p^r-1}Y$, we see that $c_{p^r}^p\in A$.

Since (T_{-1}) holds true, (S_0) and (T_0) also hold true. Thus our induction works well.

Lemma 3.3 *Let $B\supset A$ be $\mathbb{F}_p$-algebras and $F(T)\in B[[T]]^\times$. If $F(X)\,F(Y)\cdot F(X+Y)^{-1}\in A[X,Y]^\times$, there exist $\mathbb{b}\in W\hat{}(B)$, $G_0(T)\in B[[T]]$ and $G_1(T)\in A[T]$ such that $F(T)=E_p(\mathbb{b},T)\,G_0(T)\,G_1(T)$, $\mathbb{b}^{(p)}\in W\hat{}(A)$ and $G_0(X)\,G_0(Y)\,G_0(X+Y)^{-1}=1$.*

Proof. Put $F(T)=\prod_{k=1}^{\infty} E_p(c_k T^k)$. Then

$$F(X)\,F(Y)\,F(X+Y)^{-1}$$
$$=\prod_{k\in\mathbb{P}} F_p(c_k^p; X, Y)\times \prod_{k\in\mathbb{N}-\mathbb{P}} [E_p(c_k X^k)\,E_p(c_k Y^k)\,E_p(c_k(X+Y)^k)^{-1}].$$

Here $\mathbb{P}=\{p^r; r\geqq 0\}$. Let $d=\deg. F(X)\,F(Y)\,F(X+Y)^{-1}$ and let $\mathfrak{a}$ denote the ideal of A generated by the coefficients of terms of degree $\geqq 1$ in $F(X)\,F(Y)\,F(X+Y)^{-1}$. Then $\mathfrak{a}$ is nilpotent.

We now observe that:

1) When $k=p^r$,

$$F_p(c_k^p; X^k, Y^k)=1+\sum_{j\geqq 1} H_j(X, Y),$$

where $H_j(X, Y)$ is a homogeneous polynomial of degree $j\,p^{r+1}$. The ideal generated by the coefficients of $H_1(X, Y)$ coincides with (c_k^p), and $(c_k^{p^j})$ contains the ideal generated by the coefficients of $H_j(X, Y)$ for $j>1$;

2) When k is not a power of p,

$$E_p(c_k X^k)\,E_p(c_k Y^k)\,E_p(c_k(X+Y)^k)^{-1}=1+\sum_{j\geqq 1} H_j(X, Y),$$

where $H_j(X, Y)$ is a homogeneous polynomial of degree $j\,k$. The ideal generated by the coefficients of $H_1(X, Y)$ coincides with (c_k), and (c_k^j) contains the ideal generated by the coefficients of $H_j(X, Y)$ for $j>1$.

It follows that

1) $c_k^p\in\mathfrak{a}^s$ if $k=p^r$ and $(s-1)\,d<p^{r+1}\leqq s\,d$;
2) $c_k\in\mathfrak{a}^s$ if k is not a power of p and $(s-1)\,d<k\leqq s\,d$.

Hence, c_k is nilpotent for all k, $c_k^p=0$ for almost all $k\in\mathbb{P}$ and $c_k=0$ for almost all $k\in\mathbb{N}-\mathbb{P}$. Therefore, it is enough to define $\mathfrak{b}=(b_r)_{r\geqq 0}$ by $b_r=c_{p^r}$ if $c_{p^r}^p\neq 0$ and $b_r=0$ if $c_{p^r}^p=0$ and to put $G_0(T)=\prod_{r:c_{p^r}^p=0} E_p(c_{p^r}; T^{p^r})$ and $G_1(T)=F(T)/[E_p(\mathfrak{b}; T)\,G_0(T)]$.

3.4 Now we give a proof of the main theorem.

Let $G(X, Y)\in Z^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$. By 3.1, there exist a flat A-algebra B and $F(T)\in B[[T]]^{\times}$ such that $F(X)\,F(Y)\,F(X+Y)^{-1}=G(X, Y)$. By 3.2, we may assume that $F(T)=E_p(\mathfrak{b}; T)=\prod_{r\geqq 0} E_p(b_r T^{p^r})$ with $\mathfrak{b}^{(p)}\in W(A)$. Then $G(X, Y)=F_p(\mathfrak{b}^{(p)}; X, Y)$. This proves that $W(A)/F\to H_0^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ is surjective.

Now assume that $F_p(\mathfrak{a}; X, Y)=\prod_{r\geqq 0} F_p(a_r; X^{p^r}, Y^{p^r})\in B^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$. Then there exists $F(T)\in A[[T]]^{\times}$ such that $F(X)\,F(Y)\,F(X+Y)^{-1}=F_p(\mathfrak{a}; X, Y)$. Put $F(T)=\prod_{k\geqq 1} E_p(c_k T^k)$. Then

$$F_p(\mathfrak{a}; X, Y)\,F_p(\mathfrak{b}^{(p)}; X, Y)^{-1}=\prod_{k\notin\mathbb{P}} [E_p(c_k X^k)\,E_p(c_k Y^k)\,E_p(c_k(X+Y)^k)^{-1}],$$

where $\mathbb{P}=\{p^r; r\geqq 0\}$ and $\mathfrak{b}=(c_{p^r})_{r\geqq 0}$. By comparing the coefficients, we see that $c_k=0$ if k is not a power of p, and therefore $F_p(\mathfrak{a}; X, Y)=F_p(\mathfrak{b}^{(p)}; X, Y)$. It follows that $W(A)/F\to H_0^2(\mathbb{G}_{a,A}^{\wedge}, \mathbb{G}_{m,A}^{\wedge})$ is injective.

Now let $G(X, Y) \in Z^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$. Take $F(T) \in B[[T]]^\times$ so that $F(X)F(Y)F(X+Y)^{-1} = G(X, Y)$. By 3.3, we may assume that $F(T) = E_p(\mathfrak{b}; T) = \prod_{r \geqq 0} E_p(b_r T^{p^r})$ with $\mathfrak{b} \in W^\wedge(B)$ and $\mathfrak{b}^{(p)} \in W(A)$. This proves that $W^\wedge(A)/F \to H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ is surjective. It is readily seen that $W^\wedge(A)/F \to H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ is injective, which completes the proof.

Corollary 3.5 *Let A be a local $\mathbb{F}_p$-algebra and E an extension of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$. Suppose that A_{red} is a unique factorization domain. Then there exists a finite flat A-algebra B such that $E \otimes_A B$ splits.*

Proof. Since $H^1(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) = \mathrm{Pic}(\mathbb{G}_{a,A}) = 0$, $H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) \simeq \mathrm{Ext}_A^1(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$. Let $G(X, Y) = \prod_{r \geqq 0} F_p(a_r; X^{p^r}, Y^{p^r})$ be a 2-cocycle which defines the extension E of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$. It is enough to take $B = A[\sqrt[p]{a_r}; r \geqq 0]$.

Example 3.6 Assume that A is an *artinian* local $\mathbb{F}_p$-algebra. Let ξ be a generator of the maximal ideal of A. If the residue field of A is perfect, *all* the extensions of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$ split over $B = A[\sqrt[p]{\xi}] = A[T]/(T^p - \xi)$.

Remark 3.7 $\mathrm{End}_{A-gr}(\mathbb{G}_{a,A}^\wedge)$ (resp. $\mathrm{End}_{A-gr}(\mathbb{G}_{a,A})$) acts on $H_0^2(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge)$ (resp. $H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$) by the pull-back. We can describe the action under the identifications $H_0^2(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge) = W(A)/F$ and $H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) = W^\wedge(A)/F$ as follows:

Let $[a]$ denote the endomorphism of $\mathbb{G}_{a,A}^\wedge$ or $\mathbb{G}_{a,A}$, defined by $a \in A$. Then $\mathfrak{a}[a] = a^p \mathfrak{a} = (a^{p^{r+1}} a_r)_{r \geqq 0}$.

Let F denote the Frobenius endomorphism of $\mathbb{G}_{a,A}^\wedge$ or $\mathbb{G}_{a,A}$. Then $\mathfrak{a} F = V \mathfrak{a}$.

Remark 3.8 We define bi-additive maps

$$H_0^2(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{a,A}^\wedge) \times \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge) \to H_0^2(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge)$$

and

$$H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{a,A}) \times \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) \to H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$$

by $(C(X, Y), \psi(T)) \mapsto \psi(C(X, Y))$.

In particular, $\psi(T) \mapsto \psi\left(\dfrac{X^p + Y^p - (X+Y)^p}{p}\right)$ induces homomorphisms

$$\gamma\colon \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge) \to H_0^2(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge)$$

and

$$\gamma\colon \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) \to H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}).$$

Recall that $\mathfrak{a} \mapsto E_p(\mathfrak{a}; T)$ gives rise to bijective homomorphisms

$${}_F W(A) \xrightarrow{\sim} \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge) \quad \text{and} \quad {}_F W^\wedge(A) \xrightarrow{\sim} \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}).$$

Under the identifications

$$\mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge) = {}_F W(A), \qquad H_0^2(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge) = W(A)/F,$$

and

$$\mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) = {}_F W^\wedge(A), \qquad H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) = W^\wedge(A)/F,$$

$\gamma\colon \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge) \to H_0^2(\mathbb{G}_{a,A}^\wedge, \mathbb{G}_{m,A}^\wedge)$ and $\gamma\colon \mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) \to H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ are given by $\mathfrak{a} \mapsto \mathfrak{a}$.

Example 3.9 Let A be an artinian local $\mathbb{F}_p$-algebra and $\mathfrak{m}$ the maximal ideal of A. Assume that $\mathfrak{m}^p=0$. Then $W^\wedge(A)$ is isomorphic to the additive group $\mathfrak{m}^{(\mathbb{N})}$ and the Frobenius map F is zero on $W^\wedge(A)$. Therefore $\mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ and $H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ are isomorphic to the additive group $\mathfrak{m}^{(\mathbb{N})}$, and the homomorphism γ: $\mathrm{Hom}_{A-gr}(\mathbb{G}_{a,A}, \mathbb{G}_{m,A}) \to H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})$ is bijective.

Remark 3.10 There may exist a non-commutative central extension of $\mathbb{G}_{a,A}^\wedge$ by $\mathbb{G}_{m,A}^\wedge$ or of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$. Indeed, assume that A is of characteristic $p>0$ but not reduced, and take $a\in A$ so that $a\neq 0$ and $a^p=0$. Put $G(X, Y)=\sum_{i=0}^{p-1}(aXY^p)^i/i!$. Then $G(X, Y)\neq 0$ and $G(X, Y)$ satisfies the functional equation:

$$G(X+Y, Z)\,G(X, Y)=G(X, Y+Z)\,G(Y, Z).$$

It follows that $G(X, Y)$ defines a non-commutative central extension of $\mathbb{G}_{a,A}^\wedge$ by $\mathbb{G}_{m,A}^\wedge$ or of $\mathbb{G}_{a,A}$ by $\mathbb{G}_{m,A}$.

Remark 3.11 It is more easily seen that $H_0^2(\mathbb{G}_{a,A}, \mathbb{G}_{m,A})=0$ if A is a $\mathbb{Q}$-algebra.

References

[DG] Demazure, M., Gabriel, P.: Groupes algébriques, Tome 1. Paris: Masson – Amsterdam: North-Holland 1970

[HZ] Hazewinkel, M.: Formal groups and applications. New York: Academic Press 1978

[1] Lazard, M.: Sur les groupes de Lie formels à un paramètre. Bull. Soc. Math. France **83**, 251–274 (1955)

Math. Z. 206, 577–587 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Finiteness conditions in Jordan pairs

Ottmar Loos

Institut für Mathematik, Universität Innsbruck, Technikerstrasse 25/7, A-6020 Innsbruck, Austria

Received August 21, 1989; in final form December 3, 1990

To Nathan Jacobson on his 80th birthday

Introduction

N. Jacobson defined the capacity of a Jordan algebra with unit element 1 as the minimum number of orthogonal division idempotents whose sum is 1, provided such idempotents exist. We extend this concept to Jordan pairs as follows. The *capacity* of a Jordan pair is the infimum of the cardinalities of all finite sets of orthogonal division idempotents (in the Jordan pair sense) whose common Peirce-0-space vanishes (if there are no such idempotents the capacity is $+\infty$, being the infimum of the empty set). A second invariant is the *principal length*, defined as the supremum of the lengths of all finite chains of principal inner ideals. Our main result is:

The capacity of a nondegenerate Jordan pair equals its principal length.

Nondegenerate Jordan pairs of finite principal length satisfy both chain conditions on principal inner ideals and thus have finite capacity [5, § 10]. This holds for linear Jordan algebras as well [2, p. 158, Theor. 3] and for quadratic algebras after passing to an isotope [4, 6.5.8]. As to the converse, one can show by a case-by-case verification, using N. Jacobson's classification of semiprimitive Jordan algebras with dcc and with a capacity, respectively [3, 4], and K. McCrimmon's work on inner ideals [9] that semiprimitive Jordan algebras of finite capacity have finite principal length equal to the capacity.

In contrast, the proof of this fact given here is classification-free and works for Jordan pairs as well. It rests on a diagonalization procedure by elementary automorphisms, similar to the reduction of a matrix by elementary row and column operations. As a consequence, a nondegenerate Jordan pair of finite capacity is von Neumann regular. A second ingredient is a conjugacy theorem for frames which can be considered the Jordan analogue of the Krull-Schmidt theorem. Here a *frame* means a finite set of orthogonal local idempotents having their common Peirce-0-space in the radical. Just as in ring theory, it can be shown that a Jordan pair contains a frame if and only if it is semiperfect [7]. The conjugacy theorem was first proved by H.P. Petersson [12] assuming dcc and acc on principal inner ideals but in fact holds without this hypothesis (which is important for the intended application). The main result then follows

by showing that a strictly decreasing chain of principal inner ideals gives rise, by completing elements to pair idempotents, to a chain of Peirce-2-spaces whose capacities decrease strictly as well, and are bounded by the capacity.

It seems appropriate to point out that it is Jordan *pairs* (as opposed to algebras or triple systems) that provide the natural framework for this proof. This is due to the fact that everything hinges on the concept of pair idempotent and the associated elementary automorphisms which have no natural counterpart for algebras or triple systems.

Our standard reference for terminology and notation is [5]. All Jordan pairs are modules over an arbitrary commutative ring of scalars which, however, plays no role whatsoever and may as well be taken as the integers.

§ 1 Elementary automorphisms

Let e be an idempotent of a Jordan pair V with Peirce spaces $V_i(e)$, $i=0, 1, 2$. For every $v\in V_1^-(e)$, the pair (e_+, v) is quasi-invertible, and the map $v\mapsto\beta(e_+, v)$ is a homomorphism of the additive group $V_1^-(e)$ into the inner automorphism group $\mathrm{Inn}(V)$. This follows from the formula $\beta(x, y)\,\beta(x^y, z)=\beta(x, y+z)$ and the fact that $e_+^v=e_+$ by the Peirce relations. Similarly, we have a homomorphism $u\mapsto\beta(u, e_-)$ from $V_1^+(e)$ into $\mathrm{Inn}(V)$. The subgroup of $\mathrm{Inn}(V)$ generated by $\beta(e_+, V_1^-(e))$ and $\beta(V_1^+(e), e_-)$ is denoted by $E(e, V)$ and called the group of *e-elementary automorphisms*. More generally, if S is a set of idempotents, then $E(S, V)$ is the group generated by all $E(e, V)$, $e\in S$. In case S is the set of all idempotents of V, we write $E(V)$ instead of $E(S, V)$ and call it the *elementary automorphism group* of V.

Recall that two idempotents e and e' are called *associated* ($e\approx e'$) if they have the same Peirce spaces. This is equivalent to e'_+ invertible in $V_2(e)$ with $e'_-=(e'_+)^{-1}$ (see [12, Lemma 1], [11, Theor. I.2.3]). Two sets S, S' of idempotents are associated ($S\approx S'$) if they are equal modulo the equivalence relation $\approx$. Two orthogonal idempotents $e_1\perp e_2$ are *connected* if $V_{12}=V_1(e_1)\cap V_1(e_2)$ contains invertible elements of $V_2(e_1+e_2)$.

Proposition 1 (a) $E(\varphi(e), V)=\varphi E(e, V)\,\varphi^{-1}$ *for an automorphism* φ *of* V.
(b) *If* $S\approx S'$ *then* $E(S, V)=E(S', V)$.
(c) *If* $e_1\perp e_2$ *then* $E(e_1+e_2, V)\subset E(\{e_1, e_2\}, V)$.
(d) *If* $e_1\perp e_2$ *are connected then* $e_2\approx\varphi(e_1)$ *for some* $\varphi\in E(e_1, V)$ *and* $E(e_1, V)=E(e_2, V)$.
(e) *If* c *is an idempotent and* $e\in V_j(c)$ *for* $j=2$ *or* $j=0$ *then every* $\varphi'\in E(e, V_j(c))$ *extends to a* $\varphi\in E(e, V)$ *which acts like the identity on* $V_{2-j}(c)$.

Proof. (a) is obvious from the definitions, and (b) follows from the formulas $\beta(a, v)=\beta(e_+, \{e_-av\})$ and $\beta(u, b)=\beta(\{e_+bu\}, e_-)$ for (a, b) in $V_2(e)$ and (u, v) in $V_1(e)$, see [1]. For (c), let V_{ij} $(0\leqq i\leqq j\leqq 2)$ be the Peirce spaces of V with respect to e_1, e_2. Then $V_1(e_1+e_2)=V_{10}+V_{20}$, and $\beta(e_1^+ + e_2^+, v_{10})\cdot\beta(e_1^+ + e_2^+, v_{20})=\beta(e_1^+ + e_2^+, v_{10}+v_{20})$. Furthermore, by the formula $\beta(x+z, y)=\beta(z, y^x)\beta(x, y)$ we have $\beta(e_1^+ + e_2^+, v_{20})=\beta(e_2^+, v_{20}^{e_1^+})\,\beta(e_1^+, v_{20})$. Now v_{20} is in the Peirce-0-space of e_1, hence $v_{20}^{e_1^+}=v_{20}$ and $\beta(e_1^+, v_{20})=\mathrm{Id}$. Similarly, $\beta(e_1^+ + e_2^+, v_{10})=\beta(e_1^+, v_{10})$, and $\beta(u_{10}+u_{20}, e_1^- + e_2^-)=\beta(u_{10}, e_1^-)\,\beta(u_{20}, e_2^-)$ whence (c) follows. To prove (d), let $u\in V_{12}^+$ be invertible in $V_2(e_1+e_2)$ with

inverse $v \in V_{12}^-$. Then a computation shows that $\varphi(e_1) \approx e_2$ where $\varphi = \beta(e_1^+, v)\,\beta(-u, e_1^-)\,\beta(e_1^+, v)$. Hence $E(e_2, V) = E(\varphi(e_1), V) = E(e_1, V)$ by (a) and (b). Finally, let $V' = V_j(c)$. Then $V_1'(e) = V' \cap V_1(e)$, showing that the generators $\beta'(e_+, v)$, $\beta'(u, e_-)$ of $E(e, V')$ (where (u, v) are in $V_1'(e)$) extend to $\beta(e_+, v)$, $\beta(u, e_-) \in E(e, V)$. The last assertion follows by orthogonality of $V_j(c)$ and $V_{2-j}(c)$.

Remark. Conjugacy of e_1 and e_2 up to association by an inner automorphism was proved in [12, Theor. 1 (i)]. However, the inner automorphism constructed there is in general not elementary.

§ 2 Weak diagonalization

We recall that an idempotent e in a Jordan pair V is *local* if $V_2(e)$ is a local Jordan pair in the sense that the non-invertible elements form a proper ideal. Clearly, a division idempotent is local, and the converse holds if the Jacobson radical Rad $V = 0$. The following remark is useful: *If e is contained in a diagonal Peirce space $V_j(c)$ $(j = 2, 0)$ of some other idempotent c then the Peirce-2-space of e computed in V or in $V_j(c)$ is the same.* This follows immediately from the Peirce relations. As a consequence, e is a local (division) idempotent of V if and only if e is such an idempotent of $V_j(c)$. (For $j = 1$ this fails, as simple examples show).

An idempotent e is called *complete* if $V_0(e) \subset \mathrm{Rad}\, V$. More generally, let $F = \{e_1, \ldots, e_n\}$ be a finite set of orthogonal idempotents, and let $\Sigma F = e_1 + \ldots + e_n$. We say F is complete if ΣF is complete; equivalently, if $V_0(F) = V_0(e_1) \cap \ldots \cap V_0(e_n) \subset \mathrm{Rad}\, V$. By a *frame* we mean a complete finite set of orthogonal local idempotents. In [5, 10.12], frames were defined for Jordan pairs with dcc on principal inner ideals in a somewhat different manner. Using [5, 10.10] or [6, Coroll. 2 of Theor. 2] it is easy to see that the two definitions are consistent, except for the fact that the frames considered here are *unordered* sets of idempotents. Note that the empty set is a frame if and only if V is radical. Also, if S is a subset of a frame F and $c = \Sigma S$ then S is a frame of $V_2(c)$ and $F - S$ is a frame of $V_0(c)$. This follows easily from the remark made at the beginning and the Peirce relations.

A frame F is called *strong* if it consists of division idempotents and $V_0(F) = 0$; equivalently, if the diagonal $\sum_{i=0}^{n} V_{ii}$ of F has radical zero. *Thus V has finite capacity if and only if it contains a strong frame*, and then the capacity, denoted by $\kappa(V)$, is the minimum cardinality of a strong frame. In particular, $\kappa(V) = 0$ if and only if $V = 0$.

Theorem 1 *Let F be a strong frame of a nondegenerate Jordan pair V and $x \in V^+$. Then there exists $\varphi \in E(F, V)$ and an ordering $F = \{e_1, \ldots, e_n\}$ such that*

$$\varphi_+(x) = \sum_{i=1}^{s} x_{2i-1,2i} + \sum_{j=2s+1}^{2s+t} x_{jj}$$

(Peirce components with respect to F) where $x_{2i-1,2i}$ is invertible in $V_{2i-1,2i-1} \oplus V_{2i-1,2i} \oplus V_{2i,2i} = V_2(e_{2i-1} + e_{2i})$ for $i = 1, \ldots, s$, and x_{jj} is invertible in V_{jj}, for $j = 2s+1, \ldots, 2s+t$.

Remark. The numbers s and t are not uniquely determined by x but $2s+t$ is; see [8].

We first prove a number of Lemmas.

Lemma 1 *Let e be an idempotent of a Jordan pair V, and let $x=x_2+x_1+x_0$ be the Peirce decomposition of $x\in V^+$ with respect to e. Suppose that x_2 is invertible in $V_2(e)$. Then an e-elementary transform of x is e-diagonal; more precisely,*

$$\varphi_+(x)=x_2+(x_0-Q_{x_1}x_2^{-1})\in V_2^+(e)\oplus V_0^+(e)$$

where $\varphi=\beta(x_1, x_2^{-1})\in E(e, V)$.

Proof. By Proposition 1(b) we may assume $x_2=e_+$, replacing e by $(x_2, x_2^{-1})\approx e$. Then

$$\begin{aligned}B(x_1, e_-)\cdot x&=e_++x_1+x_0-\{x_1, e_-, e_++x_1+x_0\}\\&\quad+Q_{x_1}Q_{e_-}(e_++x_1+x_0)\\&=e_++x_1+x_0-x_1-\{x_1, e_-, x_1\}+Q_{x_1}e_-\\&=e_++(x_0-Q_{x_1}e_-)\end{aligned}$$

by the Peirce relations.

Remark. In case of matrices subdivided into 4 blocks $\begin{pmatrix}a & b\\ c & d\end{pmatrix}$ with $a\sim x_2$ a square matrix, $\begin{pmatrix}0 & b\\ c & 0\end{pmatrix}\sim x_1$, $d\sim x_0$, this is equivalent to the matrix calculation

$$\begin{pmatrix}1 & 0\\ -ca^{-1} & 1\end{pmatrix}\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}1 & -a^{-1}b\\ 0 & 1\end{pmatrix}=\begin{pmatrix}a & 0\\ 0 & d-ca^{-1}b\end{pmatrix}.$$

Lemma 2 *Let V be nondegenerate with orthogonal division idempotents e_1 and e_2 such that $V=V_{11}\oplus V_{12}\oplus V_{22}$ with respect to e_1, e_2. If $0\neq x\in V_{12}^+$ is not invertible in V then $Q_x e_i^-=0$ and $\{x, y, e_1^++e_2^+\}\neq 0$ for some $y\in V_{12}^-$.*

Proof. Let $e=e_1+e_2$ and consider the unital Jordan algebra $J=V^+$ with unit element $1_J=e_+$ and U-operator $U_x=Q_xQ_{e_-}$. Then $1_J=e_1^++e_2^+$ is the sum of two orthogonal division idempotents of J. Hence the first assertion follows from [4, Theor. 6.3.1, part 1]. By part 2 of that same theorem, there exists $u\in V_{12}^+$ such that $x\circ u\neq 0$. Now $x\circ u=U_{x,1}\cdot u=Q_{x,e_+}Q_{e_-}u$ so $\{xye_+\}\neq 0$ for $y=Q_{e_-}u$.

Lemma 3 *Let V be nondegenerate and e an idempotent such that $V=V_2(e)\oplus V_1(e)$. If $0\neq x\in V_1^+(e)$ then $\{xye_+\}\neq 0$ for some $y\in V_1^-(e)$.*

Proof. By the Peirce relations, $Q_xV_2^-(e)\subset V_0^+(e)=0$, and by [5, 8.2] $Q_xy=\{\{xye_+\}e_-x\}$ for $y\in V_1^-(e)$. Since x is not an absolute zero divisor, the Lemma follows.

Lemma 4 *Let F be a strong frame of a nondegenerate Jordan pair V, and let $0\neq x\in V^+$. Then one of the following two alternatives holds.*
(a) *There exist $\varphi\in E(F, V)$ and $e\in F$ such that $\varphi_+(x)=x_2+x_0\in V_2^+(e)\oplus V_0^+(e)$ with x_2 invertible in $V_2(e)$.*
(b) *There exist $\varphi\in E(F, V)$ and $e_1, e_2\in F$ such that $\varphi_+(x)=x_{12}+x_{00}\in V_{12}^+\oplus V_{00}^+$ (Peirce spaces relative to e_1, e_2), with x_{12} invertible in $V_{11}\oplus V_{12}\oplus V_{22}=V_2(e_1+e_2)$.*

Proof. Let $F=\{e_1, \ldots, e_n\}$ and let V_{ij} be the Peirce spaces of V with respect to F. Then $V_{00}=V_0(F)=0$, so x decomposes as

$$x=\sum_{1\leqq i\leqq j\leqq n} x_{ij}+\sum_{i=1}^{n} x_{i0}.$$

We distinguish the following cases.

Case 1 $x_{ii}\neq 0$ for some i. Let $e=e_i$ and $x=x_2+x_1+x_0$ relative to e. Then $x_2=x_{ii}$ is invertible in $V_2(e)=V_{ii}$ since e is a division idempotent, and alternative (a) holds by Lemma 1 for $\varphi=\beta(x_1, x_2^{-1})$.

Case 2 All $x_{ii}=0$ but $x_{ij}\neq 0$ for some $1\leqq i<j\leqq n$. Renumbering F, we may assume $x_{12}\neq 0$. Set $e=e_1+e_2$.

Case 2.1 x_{12} is invertible in $V_2(e)=V_{11}\oplus V_{12}\oplus V_{22}$. Then $x=x_2+x_1+x_0$ relative to e where $x_2=x_{12}$. As $e\approx(x_{12}, x_{12}^{-1})$ we have $\varphi=\beta(x_1, x_2^{-1})\in E(F, V)$ by Proposition 1, and we are in alternative (b) by Lemma 1.

Case 2.2 x_{12} is not invertible in $V_2(e)$. From the Peirce relations, it is immediate that diagonal Peirce spaces inherit non-degeneracy. Hence by Lemma 2, applied to $V_2(e)$, we have $\{x_{12}, y_{12}, e_+\}\neq 0$ for some y_{12} in V_{12}^-. Now $\{x_{12}, y_{12}, e_+\}$ $=\{x_{12}, y_{12}, e_1^+\}+\{x_{12}, y_{12}, e_2^+\}\in V_{11}^+\oplus V_{22}^+$, so (say) $x'_{11}=\{x_{12}, y_{12}, e_1^+\}\neq 0$. Let $S=\{e_1, e_2\}$ and let $x=x_{12}+x_{10}^S+x_{20}^S+x_{00}^S$ be the Peirce decomposition of x with respect to S (observe $x_{12}^S=x_{12}$). By the Peirce relations, we then have

$$\begin{aligned} x'=B(e_1^+, -y_{12})\cdot x&=x+\{e_1^+, y_{12}, x_{12}+x_{10}^S+x_{20}^S+x_{00}^S\}\\ &\quad+Q(e_1^+)\,Q(y_{12})(x_{12}+x_{10}^S+x_{20}^S+x_{00}^S)\\ &=x+x'_{11}+\{e_1^+, y_{12}, x_{20}^S\}\end{aligned}$$

where the last term is in $V_{10}^+(S)$. Hence the $E(F, V)$-transform x' of x falls under case 1.

Case 3 $x_{ij}=0$ for all $1\leqq i\leqq j\leqq n$. Then $x_{i0}\neq 0$ for some i, say, $x_{10}\neq 0$. Let e'_2 $=e_2+\ldots+e_n$, and $T=\{e_1, e'_2\}$. Then $x=x_{10}+x_{20}^T$ is the Peirce decomposition of x relative to T and $V_{11}\oplus V_{10}=V_0(e'_2)$ is nondegenerate. By Lemma 3 (applied to $e=e_1$ in $V_0(e'_2)$) there exists $y_{10}\in V_{10}^-$ such that $\{x_{10}, y_{10}, e_1^+\}=x'_{11}\neq 0$. Therefore,

$$\begin{aligned} x'=B(e_1^+, -y_{10})\cdot x&=x+\{e_1^+, y_{10}, x_{10}+x_{20}^T\}\\ &\quad+Q(e_1^+)\,Q(y_{10})(x_{10}+x_{20}^T)\\ &=x+x'_{11}+\{e_1^+, y_{10}, x_{20}^T\}\end{aligned}$$

where the last summand belongs to $V_{12}^+(T)$. As before, the $E(F, V)$-transform x' of x falls under case 1, and the proof is complete.

Now Theorem 1 follows easily by induction on Card F from Lemma 4 and Proposition 1 (e) since $V_0(S)$ is nondegenerate with strong frame $F-S$ for any subset S of F.

Corollary. *A nondegenerate Jordan pair of finite capacity is von Neumann regular. In particular, it has radical zero, and hence every frame is strong.*

This is immediate from Theorem 1 and the fact that V^{op} satisfies the hypotheses along with V so Theorem 1 applies to elements of V^- as well.

Remarks. (1) It is of course well known that a nondegenerate Jordan pair with dcc on principal inner ideals is von Neumann regular (see [10] and [5, 10.17]). The point here is that the chain condition can be replaced by the existence of a strong frame; in fact, acc and dcc on principal inner ideals then follow (Theor. 3 below).
(2) Strict diagonal form ($s=0$ in the notation of Theor. 1) cannot always be achieved; see [8].

§ 3 Conjugacy of frames

We will show that frames are unique up to conjugacy by elementary automorphisms and association. The main tool is an analogue of Lemma 1 for idempotents instead of single elements. Let d and e be idempotents of V, and decompose d with respect to e into Peirce components:

$$(*) \qquad d_+ = x_2 + x_1 + x_0, \quad d_- = y_2 + y_1 + y_0.$$

Then a straightforward computation using the Peirce relations shows that the idempotence of d is equivalent to the following system of equations.

$$(0^+) \qquad x_0 = Q_{x_0} y_0 + Q_{x_1} y_2 + \{x_1\, y_1\, x_0\},$$
$$(0^-) \qquad y_0 = Q_{y_0} x_0 + Q_{y_1} x_2 + \{y_1\, x_1\, y_0\},$$
$$(1^+) \qquad x_1 = Q_{x_1} y_1 + \{x_2\, y_2\, x_1\} + \{x_2\, y_1\, x_0\} + \{x_0\, y_0\, x_1\},$$
$$(1^-) \qquad y_1 = Q_{y_1} x_1 + \{y_2\, x_2\, y_1\} + \{y_2\, x_1\, y_0\} + \{y_0\, x_0\, y_1\},$$
$$(2^+) \qquad x_2 = Q_{x_2} y_2 + Q_{x_1} y_0 + \{x_1\, y_1\, x_2\},$$
$$(2^-) \qquad y_2 = Q_{y_2} x_2 + Q_{y_1} x_0 + \{y_1\, x_1\, y_2\}.$$

Proposition 2 *With the above notations, suppose x_2 invertible in $V_2(e)$. Then $\varphi(d) = (x_2, x_2^{-1}) + f \approx e + f$ where*

$$\varphi = \beta(x_2, -y_1 - \{y_0\, x_1\, x_2^{-1}\})\, \beta(x_1, x_2^{-1})$$

is e-elementary, and f is the idempotent orthogonal to e given by

$$f_+ = x_0 - Q_{x_1} x_2^{-1}, \quad f_- = Q_{y_0} f_+.$$

Remark. In particular, then, $e \in V_2(\varphi(d))$. This fact is due to Petersson [12, Theor. 1 (ii)].

Proof. Replacing e by $(x_2, x_2^{-1}) \approx e$ we may assume $x_2 = e_+$ hence $x_2^{-1} = e_-$. Let $d' = \beta(y_1, e_-) \cdot d$. Then $d'_+ = e_+ + x'_0$ where $x'_0 = x_0 - Q_{x_1} e_- = f_+$ by Lemma 1. A computation using the Peirce relations shows that $d'_- = B(e_-, -x_1) \cdot d_- = y'_2 + y'_1 + y_0$ where $y'_1 = y_1 + \{y_0\, x_1\, e_-\}$. From (2^+) (applied to d') we get then $e_+ = Q_{e_+} y'_2$ whence $y'_2 = e_-$. Similarly, (2^-) and (1^+) show $Q_{y'_1} x'_0 = 0$ and $\{e_+\, y'_1\, x'_0\} = 0$ which implies $B(e_+, -y'_1) \cdot d'_+ = d'_+$ by an easy computation. From Lemma 1, applied to V^{op} and $d'_- = e_- + y'_1 + y_0$, we see that $B(y'_1, e_+) \cdot d'_- = e_- + (y_0 - Q_{y'_1} e_+)$. It remains to show that $y_0 - Q_{y'_1} e_+ = Q_{y_0} f_+$. Simplify $Q_{y'_1} e_+$

$= Q(y_1 + \{y_0\, x_1\, e_-\})\, e_+ = Q_{y_1} e_+ + \{y_1\, e_+\, \{y_0\, x_1\, e_-\}\} + Q(\{y_0\, x_1\, e_-\})\, e_+$ by using the identity JP14 [5, p. 14] on the second term and JP20 (loc. cit., p. 19) on the third, as well as the Peirce relations. The result is $Q_{y_1'} e_+ = Q_{y_1} e_+ + \{y_1\, x_1\, y_0\} + Q_{y_0} Q_{x_1} e_-$. Hence $y_0 - Q_{y_1'} e_+ = Q_{y_0} x_0 - Q_{y_0} Q_{x_1} e_-$ (by (0^-)) $= Q_{y_0} f_+$, and we have $\varphi(d) = e + f \in V_2(e) \oplus V_0(e)$ as asserted. Since φ is an automorphism, $e+f$ is an idempotent, and therefore f as well, by orthogonality of the Peirce spaces.

Corollary. *Let d and e be idempotents having the same V^+-component $d_+ = e_+$. Then $\varphi(d) = e$ for $\varphi = \beta(e_+, -y_1) \in E(e, V)$.*

Proof. Specialize $x_2 = e_+$, $x_1 = x_0 = 0$.

Theorem 2 *Let F be a frame of the Jordan pair V, and let $c_1, \ldots, c_n$ be orthogonal idempotents of V. Then there exists a partition $F = F_0 \cup F_1 \cup \ldots \cup F_n$ of F and $\varphi \in E(F, V)$ such that $\varphi(c_i) \approx \Sigma F_i$, for $i = 1, \ldots, n$.*

Proof. We argue by induction on n. Let $\bar{V} = V/\mathrm{Rad}\, V$ and denote the canonical map $V \to \bar{V}$ by $x \mapsto \bar{x}$. Observe that $F \to \bar{F}$ is bijective, and $E(F, V) \to E(\bar{F}, \bar{V})$ is surjective. Applying Theorem 1 to the element $\overline{c_1^+}$ of the Jordan pair $\bar{V}$ with strong frame $\bar{F}$, we find a $\psi \in E(F, V)$ such that $\overline{\psi_+(c_1^+)}$ has the weak diagonal form indicated in Theorem 1; in particular, it is invertible in $\bar{V}_2(\bar{e})$ where (with the notation of Theor. 1) $e = e_1 + \ldots + e_{2s+t}$. Decompose $d = \psi(c_1)$ with respect to e as in $(*)$. Then $\bar{d}_+ = \bar{x}_2 + \bar{x}_1 + \bar{x}_0 = \overline{\psi_+(c_1^+)} \in \bar{V}_2^+(\bar{e})$ implies $\bar{x}_1 = \bar{x}_0 = 0$ and x_2 invertible in $V_2(e)$, since the kernel of $V_2(e) \to \bar{V}_2(\bar{e})$ is the quasi-invertible ideal $V_2(e) \cap \mathrm{Rad}\, V$ of $V_2(e)$. Thus Proposition 2 furnishes a $\psi_1 \in E(e, V) \subset E(F, V)$ (by Propos. 1) and an idempotent $f \perp e$ such that $\psi_1(d) \approx e + f$. From the explicit form of f and $x_1, x_0 \in \mathrm{Rad}\, V^+$ it follows that $f \in \mathrm{Rad}\, V$. Since the radical contains no non-zero idempotents we have $f = 0$ and hence $\varphi_1(c_1) \approx e = \Sigma F_1$ where $\varphi_1 = \psi_1 \psi$ and $F_1 = \{e_1, \ldots, e_{2s+t}\}$.

For the induction step, let $V' = V_0(e)$ and $c_i' = \varphi_1(c_i)$ for $i = 2, \ldots, n$. Then $c_i' \in V_0(\varphi_1(c_1)) = V'$, and $F' = F - F_1$ is a frame of V'. By induction, there exists $\eta' \in E(F', V')$ and a partition $F' = F_0 \cup F_2 \cup \ldots \cup F_n$ with $\eta'(c_i') \approx \Sigma F_i$ for $i = 2, \ldots, n$. By Proposition 1(e), η' extends to an $\eta \in E(F', V) \subset E(F, V)$ acting like the identity on $V_2(e)$. Hence $\varphi = \eta \varphi_1$ has the required properties.

Corollary 1 *Let F be a frame of V. Then $E(V) = E(F, V)$; in other words, the group of all elementary automorphisms is generated by the e-elementary automorphisms where e runs through a single frame.*

Proof. By definition, $E(V)$ is generated by all $E(c, V)$ where c is an arbitrary idempotent. By Theorem 2, $\varphi(c) \approx e = \Sigma F_1$ for $\varphi \in E(F, V)$ and $F_1 \subset F$. Hence $E(c, V) = \varphi^{-1} E(e, V)\, \varphi \subset E(F, V)$, by Proposition 1.

Corollary 2 *Up to association, any two frames are conjugate by an elementary automorphism and have the same cardinality.*

Proof. Let F and $F' = \{c_1, \ldots, c_n\}$ be frames of V. By Theorem 2, $\varphi(c_i) \approx \Sigma F_i$, $i = 1, \ldots, n$, for some $\varphi \in E(F, V) = E(V)$. Since $\varphi(c_i)$ is local, Card $F_i = 1$. Also, $F_0 = \emptyset$ or else $V_0(F')$ would contain the non-zero idempotent $\varphi^{-1}(\Sigma F_0)$, contradicting completeness of F'. Hence $\varphi(F') \approx F$, in particular, Card $F =$ Card F'.

Remark. If V satisfies both chain conditions on principal inner ideals, this is due to Petersson (with inner instead of elementary automorphisms), see [12, Coroll. 3 of Main Theorem].

A Jordan pair V containing a frame is said to be *connected* if any two local orthogonal idempotents of V are connected. By Corollary 2, it suffices to have any two elements of one frame F connected. It can be shown [7] that this definition is equivalent to Petersson's [12]. For connected Jordan pairs, Corollaries 1 and 2 can be sharpened as follows (cf. [12, Main Theorem and Theor. 3]).

Corollary 3 *Let V be a connected Jordan pair containing a frame. Then any two ordered sets of orthogonal local idempotents of the same cardinality are conjugate by $E(V)$ up to association, and $E(V)=E(e, V)$ where e is any local idempotent.*

The proof is immediate in view of Proposition 1. Another application of Theorem 2 is the following result on the structure of idempotents in a Jordan pair containing a frame.

Proposition 3 *Suppose V has a frame, and let c be an idempotent of V.*
(a) *c can be imbedded in a frame: There exists a frame $F=\{e_1, \ldots, e_n\}$ such that $c \approx e_1+\ldots+e_r$.*
(b) *$V_2(c)$ and $V_0(c)$ contain frames.*
(c) *c is complete if and only if $c \approx \Sigma F$ for some frame F. Any two complete idempotents are conjugate (up to association) by an elementary automorphism.*
(d) *If V has invertible elements then c is complete if and only if c_+ is invertible in V. Hence in this case, any two complete idempotents are associated.*

Proof. (a) is immediate from Theorem 2.
(b) If F is as in (a) then $\{e_1, \ldots, e_r\}$ and $\{e_{r+1}, \ldots, e_n\}$ are frames of $V_2(c)$ and $V_0(c)$, respectively.
(c) The first part follows at once from (a), and the second from Corollary 2 of Theorem 2.
(d) If V has an invertible idempotent e then $V_1(e)=V_0(e)=0$ so e is complete. Conversely, let c be complete in $V=V_2(e)$. Then $c \approx \varphi(e)$ for some elementary automorphism, by (c). Thus $V_2(c)=V_2(\varphi(e))=V_2(e)=V$ and c is invertible.

Remarks. (1) It is not true in general that every idempotent can be written as a sum of orthogonal local idempotents, thus "$\approx$" cannot be replaced by "$=$" in (a); cf. [5, 5.12].
(2) Part (c) is due to Petersson [12, Coroll. 4 of Main Theorem], in case V has dcc and acc on principal inner ideals.
(3) In infinite-dimensional situations, (d) and the second part of (c) fail. For example, let M be a vector space with a countable basis $v_1, v_2, \ldots$, and let $V^+=V^-=\operatorname{End} M$ with $Q_x y=xyx$. Then the forward shift $c_+: v_i \mapsto v_{i+1}$ and the backward shift $c_-: v_{i+1} \mapsto v_i$, $v_1 \mapsto 0$ define an idempotent $c=(c_+, c_-)$ of V which is complete ($V_0(c)=0$) but not invertible ($V_1(c) \neq 0$).

§ 4 Capacity and principal length

Let V be a Jordan pair containing a frame F. By Corollary 2 of Theorem 2, the cardinality of F is an invariant of V which we denote by $\gamma(V)$. If V contains one strong frame (i.e., has finite capacity) then every frame is strong and $\kappa(V)=\gamma(V)$. In general, $\gamma(V)=\kappa(V/\operatorname{Rad} V)$ since a frame F of V gives rise to

a frame $\bar{F}$ of $\bar{V}=V/\mathrm{Rad}\,V$ of the same cardinality. Thus $\gamma(V)=0$ while $\kappa(V)=\infty$ for $0\neq V=\mathrm{Rad}\,V$.

By Proposition 3, we can define $\gamma(c)$ of an idempotent c by $\gamma(c)=\gamma(V_2(c))$. If $c\approx e_1+\ldots+e_r$ as in Proposition 3(a) then $\gamma(c)=r$.

Lemma 5 *Let V be a Jordan pair containing a frame, and let c, d, e be idempotents of V.*
(a) *$\gamma(c)=0$ if and only if $c=0$, and $\gamma(c)=1$ if and only if c is local.*
(b) *$\gamma(c+d)=\gamma(c)+\gamma(d)$ for $c\perp d$.*
(c) *$\gamma(c)+\gamma(V_0(c))=\gamma(V)$.*
(d) *If $c\in V_2(e)$ then $\gamma(c)\leqq\gamma(e)$, with equality if and only if $c\approx e$.*

The proof is an easy consequence of Theorem 2 and Proposition 3.

Lemma 6 *Let V be a Jordan pair containing invertible elements, say $V=V_2(c)$ where c_+ is invertible and $c_-=c_+^{-1}$. Let e be a local idempotent of V. Then there exists an idempotent $f\perp e$ such that $c\approx e+f$, and hence $V_0(e)=V_2(f)$. In particular, $V_0(e)$ contains invertible elements as well.*

Proof. If $c\approx e+f$ then it follows from the Peirce decomposition of V relative to e and f that $V_0(e)=V_2(f)$. Now assume first $\mathrm{Rad}\,V=0$ so e is a division idempotent. Decomposing $c_+=x_2+x_1+x_0$ and $c_-=y_2+y_1+y_0$ relative to e, we distinguish the following cases.

Case 1 $x_2\neq 0$ or $y_2\neq 0$. If $x_2\neq 0$ then x_2 is invertible in $V_2(e)$ since e is a division idempotent. By Proposition 2, $\varphi(c)\approx e+f$ where $f\perp e$, and $c\approx\varphi(c)$ by invertibility of c. Passing to V^{op}, one argues in the same way if $y_2\neq 0$.

Case 2 $x_2=y_2=0$. Let $e'=(Q(c_+)e_-,\ Q(c_-)e_+)$. Since $(Q(c_+),\ Q(c_-))\colon V^{\mathrm{op}}\to V$ is an isomorphism, e' is a division idempotent, and $e'\perp e$ by $Q(V_1^\sigma(e)\oplus V_0^\sigma(e))\cdot V_2^{-\sigma}(e)\subset V_0^\sigma(e)$. Let V_{ij} be the Peirce spaces of V with respect to $e_1=e$ and $e_2=e'$, and decompose $c_+=x_{12}+x_{22}+x_{10}+x_{20}+x_{00}$ accordingly (note $x_{11}=x_2=0$). Then x_{22} is the Peirce-2-component of c_+ relative to e'. Thus if $x_{22}\neq 0$ then $c\approx e'+f'$ by case 1, and hence $c\approx e+f$ where $f=(Q(c_+)f'_-,\ Q(c_-)f'_+)$ since $Q(c_\sigma)e'_{-\sigma}=e_\sigma$ by $Q(c_-)=Q(c_+)^{-1}$. This leaves the case $x_{22}=0$. But then x_{12} is invertible in $V_2(e'')=V_{11}\oplus V_{12}\oplus V_{22}$ where $e''=e_1+e_2$. Indeed, from $Q(c_+)e_1^-=e_2^+$ and $c_+=x_{12}+x_{10}+x_{20}+x_{00}$ it follows by expanding and comparing Peirce-components that $Q(x_{12})e_1^-=e_2^+$. Also, $Q(c_+)e_2^-=Q(c_+)Q(c_-)e_1^+=e_1^+$ implies $Q(x_{12})e_2^-=e_1^+$ in the same way. Hence $Q(x_{12})e''_-=e''_+$ and therefore $Q(x_{12})Q(e''_-)Q(x_{12})=Q(e''_+)$. Since $Q(e''_+)\colon V_2^-(e'')\to V_2^+(e'')$ is invertible so is $Q(x_{12})$. Now the Peirce decomposition of c_+ with respect to e'' is $c_+=x''_2+x''_1+x''_0$ where $x''_2=x_{12}$ is invertible in $V_2(e'')$. By Proposition 2 we have $c\approx\varphi(c)\approx e''+f''$ where $f''\perp e''$. This shows $c\approx e+f$ where $f=e'+f''$.

Finally, let V be arbitrary and $\bar{V}$ be arbitrary and $\bar{V}=V/\mathrm{Rad}\,V$. Then there exists an idempotent $g\perp\bar{e}$ such that $\bar{c}\approx\bar{e}+g$. Choose any $u\in V^+$ such that $\bar{u}=\bar{e}_++g_+$. Then $\bar{u}$ is invertible in $\bar{V}$ and $\mathrm{Rad}\,V$ is a quasi-invertible ideal which implies u is invertible in V. Hence $c'=(u,u^{-1})$ is an idempotent associated to c. Decompose $u=u_2+u_1+u_0$ relative to e. Then $\bar{u}_2=\bar{e}_+$ is invertible in $\bar{V}_2(\bar{e})$ so u_2 is invertible in $V_2(e)$. By Proposition 2, $c\approx c'\approx\varphi(c')\approx e+f$ for some $f\perp e$. This completes the proof.

Remark. This Lemma, which holds without any finiteness assumption on V, is really a cancellation theorem as the following special case shows. Let $M^\pm$ be right modules over some ring R and $V^\sigma=\mathrm{Hom}_R(M^{-\sigma},\ M^\sigma)$ with $Q_uv=uvu$.

Then V contains invertible elements if and only if $M^+ \cong M^-$. Let $M^\sigma = M_1^\sigma \oplus M_0^\sigma$ and $e_+ : M_1^- \to M_1^+$ an isomorphism with inverse e_-. Extending $e_\pm$ by 0 to $M^\pm$ we get an idempotent $e=(e_+, e_-)$ of V with Peirce spaces $V_2^\sigma(e) \cong \mathrm{Hom}_R(M_1^{-\sigma}, M_1^\sigma)$ and $V_0^\sigma(e) \cong \mathrm{Hom}_R(M_0^{-\sigma}, M_0^\sigma)$. The idempotent e is local if and only if $\mathrm{End}(M_1^+)$ ($\cong \mathrm{End}(M_1^-)$) is a local ring. By Lemma 6, $V_0(e)$ contains invertible elements. Thus $M^+ \cong M^-$ and $M_1^+ \cong M_1^-$ implies $M_0^+ \cong M_0^-$, a well known result from ring theory.

Let V be an arbitrary Jordan pair. For $x \in V^\sigma$ we denote $[x] = Q_x V^{-\sigma}$ the principal inner ideal determined by x. The length of a finite chain $[x_0] \subset [x_1] \subset \ldots \subset [x_n]$ of principal inner ideals is defined as the number of proper inclusions. The *principal length* of V, denoted $\pi(V)$, is the supremum of the lengths of all such chains. Clearly, $\pi(V)$ is a non-negative integer or ∞, and $\pi(V)=0$ if and only if V has trivial multiplication: $Q_x = 0$ for all x.

Theorem 3 *The following conditions on a nondegenerate Jordan pair V are equivalent.*

(i) *V has finite capacity (contains a strong frame);*

(i) *V has finite principal length;*

(iii) *V has acc and dcc on principal inner ideals;*

(iv) *V has dcc on principal inner ideals and is idempotent-finite (contains no infinite sets of orthogonal idempotents);*

(v) *V has dcc on principal inner ideals and contains a complete idempotent.*

If these conditions hold then $\kappa(V) = \pi(V)$, any frame of V has cardinality $\kappa(V)$, and any maximal chain of principal inner ideals has length $\pi(V)$.

Proof. (i) $\Rightarrow$ (ii): By Corollary of Theorem 1, V is regular. Hence every $x \in V^+$ is part of an idempotent: If $x = Q_x y$ then $e = (x, Q_y x)$ is an idempotent and $e_+ = x$ [5, 5.2]. Also, $[x] = V_2^+(e)$ and $V_2(e)$ inherits regularity, by the Peirce relations. Suppose $[u] \subset [x]$. Then $u \in [u] \subset V_2^+(e)$ by regularity, and we can extend $u = c_+$ to an idempotent $c \in V_2(e)$. By Lemma 5, $\gamma(c) \leqq \gamma(e) \leqq \gamma(V) = \kappa(V)$, and $\gamma(c) = \gamma(e)$ implies $c \approx e$ hence $[u] = V_2^+(c) = V_2^+(e) = [x]$. It follows that any chain of principal inner ideals has length at most $\kappa(V)$. On the other hand, if $\{e_1, \ldots, e_n\}$ is a frame of V then $0 \subset [e_1^+] \subset [e_1^+ + e_2^+] \subset \ldots \subset [e_1^+ + \ldots + e_n^+]$ is a chain of length $n = \kappa(V)$. This proves $\kappa(V) = \pi(V)$.

(ii) $\Rightarrow$ (iii): Obvious.

(iii) $\Rightarrow$ (iv): See [5, 10.7].

(iv) $\Rightarrow$ (v): By [5, 10.8], Rad $V = 0$. Let S be a maximal set of orthogonal idempotents and $c = \Sigma S$. Then $V_0(c)$ satisfies the same hypotheses as V. If $V_0(c) \neq 0$ it would contain a division idempotent by Fitting's Lemma ([6], Coroll. 2 of Theor. 2), contradicting maximality of S.

(v) $\Rightarrow$ (i): As before, Rad $V = 0$ thus a complete idempotent c has $V_0(c) = 0$. Also, $W = V_2(c)$ has dcc on principal inner ideals and Rad $W = 0$. By Fitting's Lemma, it contains a division idempotent e_1, and by Lemma 6, $W_0(e_1) = W_2(c_1)$ for an idempotent $c_1 \perp e_1$ such that $c \approx e_1 + c_1$. Also, $W_2(c_1) = V_2(c_1)$ by the Peirce relations. Continuing in this way, we get a descending chain $V_2(c) \supset V_2(c_1) \supset V_2(c_2) \supset \ldots$ of Peirce-2-spaces where $c \approx e_1 + c_1 \approx e_1 + e_2 + c_2 \approx \ldots$ and $e_1, e_2, \ldots$ are orthogonal division idempotents. By the dcc on principal inner ideals, $V_2(c_n) = 0$ hence $c_n = 0$ for some n, and then $\{e_1, \ldots, e_n\}$ is a strong frame.

It remains to prove the last statement. Suppose $[x_0] \subset [x_1] \subset \ldots \subset [x_m]$ is a maximal chain (one that cannot be imbedded into a longer chain). Extend

$x_m = f_m^+$ to an idempotent f_m, $x_{m-1} = f_{m-1}^+$ to an idempotent $f_{m-1} \in V_2(f_m)$ and so on. By applying Proposition 3 repeatedly we find a frame F and subsets $F_0 \subset F_1 \subset \ldots \subset F_m \subset F$ such that $f_i \approx \Sigma F_i$. Now it follows from maximality that Card $F_i = i$ and $F_m = F$, proving $m = \text{Card } F = \kappa(V)$.

Corollary. *Let J be a nondegenerate unital Jordan algebra. Then an isotope of J has finite capacity* (i.e., *J is classical in the terminology of* [4]) *if and only if J has dcc on principal inner ideals.*

Proof. Let J be classical. The Jordan pair $V = (J, J)$ associated to J depends up to isomorphy only upon the isotopy class of J. Thus we may replace J by an isotope and assume it has a supplementary set $\{a_1, \ldots, a_n\}$ of orthogonal (algebra) division idempotents. Then $\{(a_1, a_1), \ldots, (a_n, a_n)\}$ is a strong frame of V. By Theorem 3, V has dcc on principal inner ideals and so does J since principal inner ideals in J and V^+ are the same. The converse follows also from Theorem 3 or from [4, 6.5.8].

References

1. Faulkner, J.: Stable range and linear groups for alternative rings. Geom. Dedicata **14**, 177–188 (1983)
2. Jacobson, N.: Structure and Representations of Jordan Algebras. Am. Math. Soc. Coll. Publ., vol. 39. Rhode Island: Providence 1968
3. Jacobson, N.: Lectures on quadratic Jordan algebras. Lect. Notes. Bombay: Tata Institute 1969
4. Jacobson, N.: Structure Theory of Jordan algebras. Lect. Notes Math., vol. 5. Fayetteville: The University of Arkansas 1981
5. Loos, O.: Jordan Pairs. Lect Notes Math., vol. 460. Berlin Heidelberg New York: Springer 1975
6. Loos, O.: Fitting decomposition in Jordan systems. J. Algebra **136**, 92–102 (1991)
7. Loos, O.: On semiperfect Jordan pairs. Arch. Math (to appear)
8. Loos, O.: Diagonalization in Jordan pairs. J. Algebra (to appear)
9. McCrimmon, K.: Inner ideals in quadratic Jordan algebras. Trans. Am. Math. Soc. **159**, 445–468 (1971)
10. Meyberg, K.: A characterization of von Neumann regular Jordan triple systems. Proc. Am. Math. Soc. **49**, 25–27 (1975)
11. Neher, E.: Jordan Triple Systems by the Grid approach. Lect. Notes Math., vol. 1280. Berlin Heidelberg New York: 1987
12. Petersson, H.P.: Conjugacy of idempotents in Jordan pairs. Commun Algebra **6**, (7) 673–715 (1978)

Math. Z. 206, 589–604 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Pfaffians, central simple algebras and similitudes

M.-A. Knus [1], R. Parimala [2] and R. Sridharan [2]

[1] Mathematik, ETH-Zentrum, Rämistrasse 101, CH-8092 Zürich, Switzerland
[2] Tata Institute of Fundamental Research, School of Mathematics, Homi Bhabha Road, Bombay 400005, India

Received June 9, 1989; in final form January 22, 1990

In [KPS], pfaffians were constructed for certain modules over Azumaya algebras and this construction was used to classify rank 6 quadratic spaces over arbitrary commutative rings. The aim of this paper is to indicate some applications of this idea.

We prove here in a general context of Azumaya algebras of rank 16 over commutative rings, classical results of Albert [A] and Jacobson [J] on central simple algebras over fields.

Further, as a consequence of our computations, we get a necessary and sufficient condition for an involution of a rank 16 Azumaya algebra to admit an invariant quaternion subalgebra. This result seems to be of independent interest, even for algebras over fields, since there are examples of Amitsur-Rowen-Tignol of involutions on central simple algebras of dimension 16 which do not have any invariant quaternion algebras (see [ART]).

Finally, using the same techniques, we compute the group of special similitudes of rank 6 quadratic spaces over commutative rings. This puts in a general set up the classical computations of Dieudonné [D] for forms of dimension 6 over fields of characteristic not equal to 2.

We thank the referee for useful comments, in particular for his doubt which has led us to the present formulation of Theorem 5.2 and its Corollaries.

1 Quadratic spaces and Clifford algebras

In this section, we recall some properties of Clifford algebras, which, at least over fields of characteristic not equal to 2, are well-known. Let R be a commutative ring. We denote the group of units of R by R^*. Tensor products over R are unadorned. A *quadratic space* is by definition a pair (M, q), where M is a finitely generated projective R-module and $q\colon M \to R$ is a quadratic form such that the associated polar form $b_q(x, y) = q(x+y) - q(x) - q(y)$ is *nonsingular*, i.e. induces an isomorphism $M \tilde{\to} M^* = \mathrm{Hom}_R(M, R)$. Let $C(q) = C_0(q) \oplus C_1(q)$

be the Clifford algebra of (M, q). Any isometry $\varphi\colon (M, q) \xrightarrow{\sim} (M', q')$ induces an isomorphism $C(\varphi)\colon C(q) \xrightarrow{\sim} C(q')$. The involution of $C(q)$ which is $(-1)\cdot$Identity) on M is called the *standard involution* of $C(q)$. A *similitude* $\varphi\colon (M, q) \xrightarrow{\sim} (M', q')$ is an isomorphism $\varphi\colon M \xrightarrow{\sim} M'$ such that $q'(\varphi(x)) = \mu q(x)$ for some unit $\mu \in R^*$. The unit μ is called the *multiplier* of the similitude. Any similitude φ induces in a unique way an isomorphism of algebras with involution $C_0(\varphi)$: $C_0(q) \xrightarrow{\sim} C_0(q')$ such that $C_0(\varphi)(xy) = \mu^{-1}\varphi(x)\varphi(y)$, $x, y \in M$ (see [K] or [KPS]). Let now (I, h) be a *discriminant* module, i.e. I is an invertible R-module and $h\colon I \otimes I \xrightarrow{\sim} R$ is an isomorphism. We write (R, d) for the discriminant module $h(x, y) = xyd$, x, y in R, $d \in R^*$. For any discriminant module (I, h) and any quadratic space (M, q) we define a quadratic space $(I, h) \otimes (M, q)$ by putting

$$(h \otimes q)(x \otimes y) = h(x, x)\, q(y), \qquad x \in I, \qquad y \in M.$$

If $(I, h) = (R, d)$, any isometry $(I, h) \otimes (M, q) \xrightarrow{\sim} (M', q')$ can be viewed as a similarity with multiplier d. More generally we call an isometry $(I, h) \otimes (M, q) \xrightarrow{\sim} (M', q')$ a *similitude with multiplier* (I, h). Any such similitude φ induces an isomorphism of algebras with involution $C_0(\varphi)\colon C_0(q) \xrightarrow{\sim} C_0(q')$ (localize!).

If (M, q) is a quadratic space of even rank $n = 2m$, then $C(q)$ is an Azumaya algebra of rank 2^n. The centre Z of $C_0(q)$ is a quadratic etale R-algebra and $C_0(q)$ is Azumaya over Z. The isomorphism class of Z is called the *Arf invariant* of (M, q). We say that (M, q) has *trivial Arf invariant* if $Z \xrightarrow{\sim} R \times R$. As a quadratic algebra, Z has a nontrivial involution σ. Then $I(q) = \{x \in Z \mid x + \sigma(x) = 0\}$ is a discriminant module. The map $I(q) \otimes I(q) \to R$ is induced by the multiplication in Z. We call its isomorphism class the *signed discriminant* $\delta(q)$ of (M, q). Locally $\delta(q)$ is the class of $(R, (-1)^{\frac{n(n-1)}{2}}) \otimes \mathrm{disc}(q)$, where $\mathrm{disc}(q) = (\wedge^n M, \tilde{b}_q)$ and $\tilde{b}_q$ is the determinant of b_q with respect to a basis of M. We have $\delta(q_1 \perp q_2) = \delta(q_1) \otimes \delta(q_2)$. If (M_1, q_1) and (M_2, q_2) are quadratic spaces of even rank, then $C(q_1 \perp q_2) \xrightarrow{\sim} C(q_1) \otimes C(\delta(q_1) q_2)$.

2 Pfaffians revisited

In this section, we recall some properties of the pfaffian as defined in [KPS]. Let A be an Azumaya algebra of rank $4n^2$ over a commutative ring R such that its class in the Brauer group of R is of order 2. Let $\varphi\colon A \otimes_R A \to \mathrm{End}_R(P)$ be an isomorphism of R-algebras, P denoting a finitely generated faithful projective R-module. The switch map ω of $A \otimes_R A$, given by $\omega(a \otimes b) = b \otimes a$, is an inner automorphism $\mathrm{Int}\, u$, $u \in (A \otimes A)^*$, such that in any splitting $\alpha\colon S \otimes_R A \xrightarrow{\sim} \mathrm{End}_S(V)$, $(\alpha \otimes \alpha)(1 \otimes u)$ is the switch map of $V \otimes_S V$. If $\psi = \varphi(u) \in \mathrm{End}_R(P)$, then $\psi^2 = 1$. We call ψ the *module involution* of P (*induced by* φ). We call

$$\mathrm{Alt}(P) = \{x - \psi(x), x \in P\}$$

the set of *alternating elements* of P with respect to ψ. If $A = \mathrm{End}_R(V)$, $P = V \otimes V$ and φ is the canonical isomorphism $Can\colon \mathrm{End}_R(V) \otimes \mathrm{End}_R(V) \xrightarrow{\sim} \mathrm{End}_R(V \otimes V)$, then $\mathrm{Alt}(V \otimes V) = \{\Sigma(x_i \otimes y_i - y_i \otimes x_i), x_i, y_i \in V\}$ can be identified with $\wedge^2 V$ through the map $x \otimes y - y \otimes x \mapsto x \wedge y$. To the triple (A, P, φ) (which we call a *2-torsion datum* in [KPS]) we can associate in a functorial way, an invertible R-module $Pf(P)$, called the *pfaffian module* of P and a map $Pf\colon \mathrm{Alt}(P) \to Pf(P)$,

such that $Pf(V\otimes V)=\wedge^{2n}V$ and $Pf\colon \wedge^2 V\to\wedge^{2n}V$ is the standard pfaffian, if $(A, P, \varphi)=(\mathrm{End}_R(V), V\otimes V, Can)$.

The map Pf satisfies

$$Pf(\varphi(a\otimes a)(x))=n_A(a)\,Pf(x), \qquad a\in A, \qquad x\in \mathrm{Alt}(P),$$

where n_A is the reduced norm on A, and is homogeneous of degree n. Thus, if $n=2$, the map $Pf\colon \mathrm{Alt}(P)\to Pf(P)$ is quadratic. If $Pf(P)$ is free, a choice of a generator of $Pf(P)$ yields a quadratic space of rank 6 over R. It was shown in [KPS] that any quadratic space of rank 6 and trivial Arf invariant arises in this way from a 2-torsion datum.

Suppose that A is an Azumaya algebra with an involution σ, i.e. an R-linear automorphism of A such that $\sigma(xy)=\sigma(y)\,\sigma(x)$ and $\sigma^2=1$. There is a faithfully flat splitting $\varphi\colon S\otimes A\simeq \mathrm{End}_S(V)$ such that $\varphi\sigma\varphi^{-1}(x)=ux^tu^{-1}$, where x^t is the transpose of x and u is an isomorphism $V^*\xrightarrow{\sim} V$ such that $u^t=\varepsilon u$ for some $\varepsilon\in\mu_2(R)=\{\eta\in R\mid \eta^2=1\}$. The element ε is called the *type* of σ. We say that σ is of *orthogonal type* if $\varepsilon=1$, of *symplectic type* if $\varepsilon=-1$ and of *even symplectic type* if u is *alternating*, i.e. of the form $u=v-v^t$ for some $v\colon V^*\to V$.

If A has an involution, its class in the Brauer group $\mathrm{Br}(R)$ is of order 2. More precisely, the map $\varphi_\sigma\colon A\otimes_R A\to \mathrm{End}_R(A)$ defined by $\varphi_\sigma(x\otimes y)(a)=xa\sigma(y)$ is an isomorphism of R-algebras. Then the corresponding module involution $\psi_\sigma\in\mathrm{End}_R(A)$ is in fact $\varepsilon\sigma$, where $\varepsilon\in\mu_2(R)$ is the type of σ. We denote by $\mathrm{Alt}^\sigma(A)$ the set of alternating elements of A with respect to $\psi_\sigma=\varepsilon\sigma$ and by $Pf(A)$ the pfaffian of the datum (A, A, φ_σ). There is non-singular pairing $\tilde{\sigma}\colon Pf(A)\times Pf(A)\to R$ induced by σ and satisfying

$$\tilde{\sigma}(Pf(a), Pf(a))=n_A(a), \qquad a\in\mathrm{Alt}^\sigma(A).$$

In particular $(Pf(A), \tilde{\sigma})$ is a discriminant module. We call its isometry class the *pfaffian discriminant* of σ. If $A=\mathrm{End}_R(V)$, σ the involution given by $\sigma(f)=uf^tu^{-1}$, for $f\in\mathrm{End}_R(V)$ and $u\colon V^*\to V$ an isomorphism with $u^t=\varepsilon u$, $\varepsilon\in\mu_2(R)$, then $Pf(A)=\wedge^{2n}V$, where rank $V=2n$ and $Pf\colon \mathrm{Alt}(\mathrm{End}_R(V))\to Pf(\mathrm{End}_R(V))$ is the composite

$$\mathrm{Alt}(V\otimes V^*)\xrightarrow{1\otimes u}\mathrm{Alt}(V\otimes V)\simeq \bigwedge^2 V\xrightarrow{Pf}\bigwedge^{2n} V$$

(Note that under the identification $\mathrm{End}_R(V)=V\otimes V^*$, σ is simply the map $\omega(u^{-1}\otimes u)$). Further, the paring $\tilde{\sigma}\colon \wedge^{2n}V\to\wedge^{2n}V^*$ is $\wedge^{2n}u^{-1}$. The following property of the pfaffian discriminant is an immediate consequence of its definition:

Lemma 2.1 *Let u be a unit of A such that $\sigma(u)=\varepsilon u$ for some $\varepsilon\in\mu_2(R)$. Then $\tau=\mathrm{Int}\,u\circ\sigma$ is an involution of A such that $(Pf(A), \tilde{\tau})\simeq(Pf(A), n_A(u)\,\tilde{\sigma})$.*

We now describe two cases where the discriminant module $(Pf(A), \tilde{\sigma})$ is trivial.

Proposition 2.2 *The discriminant module $(Pf(A), \tilde{\sigma})$ is isometric to the trivial discriminant module $(R, 1)$ in the following two cases:*
1) *$A=A_1\otimes A_2$, $\sigma=\sigma_1\otimes\sigma_2$, A_1, A_2 Azumaya algebras of rank 4 with involutions σ_1, σ_2.*
2) *A is of rank $4n^2$ and σ is of even symplectic type.*

Proof. 1) We first consider the split case. Let $A_i = \mathrm{End}_R(V_i)$, V_i free of rank 4. There exist linear isomorphisms v_i: $V_i^* \xrightarrow{\sim} V_i$ with $v_i^t = \varepsilon_i v_i$, ($\varepsilon_i$ denoting the type of σ_i), such that $\sigma_i(x) = v_i x^t v_i^{-1}$, $i = 1, 2$. Let $\{e_1, e_2\}$ be a basis for V_1, $\{f_1, f_2\}$ a basis for V_2 and u_1, u_2 the matrices of v_1, v_2 with respect to these bases and the dual bases in V_1^*, V_2^* respectively. Let $\{g_i\}_{1 \leq i \leq 4}$ be the basis $\{e_i \otimes f_j\}$ of $V = V_1 \otimes V_2$, lexicographically ordered. The element

$$e_\sigma = \det u_1 \det u_2 \, g_1 \wedge g_2 \wedge g_3 \wedge g_4 \in \bigwedge^4 V$$

is independent of the choices of the bases $\{e_1, e_2\}$ and $\{f_1, f_2\}$; hence e_σ is a basis for $\wedge^4 V$. One checks that the map $\tilde{\sigma}$: $\wedge^4 V \to \wedge^4 V^*$ induced by σ sends e_σ to the dual basis element e_σ^*. Thus $(Pf(A), \tilde{\sigma})$ is trivial if $A = \mathrm{End}_R(V_1) \otimes \mathrm{End}_R(V_2)$. The general case follows by descent.

2) The proof for involutions of even symplectic type is similar to the proof of 1). In the split case, if the involution is given by $\sigma(x) = u x^t u^{-1}$, u an alternating matrix, one takes

$$e_\sigma = pf(u) \, e_1 \wedge \ldots \wedge e_{2n},$$

where $pf(u)$ is the classical pfaffian of u. □

Remark. For any Azumaya algebra A, let $n_A(a)$ denote the reduced norm and $tr_A(a)$ the reduced trace of $a \in A$. If A is of rank 4, then $\sigma_A(a) = tr_A(a) - a$ is an involution of A. If $A = M_2(R)$,

$$\sigma_A \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^t \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Thus σ_A is of even symplectic type. We call σ_A the *standard involution* of A. We have $x \sigma_A(x) = n_A(x)$ for all $x \in A$. Observe that, for any isomorphism α: $A \xrightarrow{\sim} B$ of Azumaya algebras of rank 4, we have $\alpha \sigma_A = \sigma_B \alpha$.

3 Algebras of rank 16

Let A be an Azumaya algebra of rank 16 over R with an involution σ. We assume that the invertible module $Pf(A)$ is free and fix an isometry λ: $(Pf(A), \tilde{\sigma}) \xrightarrow{\sim} (R, d)$, i.e. $\tilde{\sigma}(x, y) = \lambda(x) \, \lambda(y) \, d$. Let Pf_λ: $\mathrm{Alt}^\sigma(A) \to R$ be the corresponding quadratic form $Pf_\lambda = \lambda \circ Pf$. The form Pf_λ has the following remarkable property:

Proposition 3.1 *There exists an R-linear automorphism p of $\mathrm{Alt}^\sigma(A)$ such that $x p(x) = p(x) x = Pf_\lambda(x)$ for every $x \in \mathrm{Alt}^\sigma(A)$ and $p^2 = d$. If p' is an R-automorphism of $\mathrm{Alt}^\sigma(A)$ such that $x p'(x) \in R$ for every $x \in \mathrm{Alt}^\sigma(A)$ (or such that $p'(x) \, x \in R$ for every $x \in \mathrm{Alt}^\sigma(A)$), then $p' = \rho p$ for some unit $\rho \in R^\bullet$.*

Proof. We begin with the case $A = M_4(R)$, $\sigma(x) = x^t$ and we choose λ such that Pf_λ is the classical pfaffian $pf(x)$, i.e. $pf(x) = a_{12} a_{34} - a_{13} a_{24} + a_{14} a_{23}$ for

$$x = \begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{pmatrix}$$

We put $\mathrm{Alt}^\sigma(M_4(R))=\mathrm{Alt}_4(R)$. Then $p_1\colon \mathrm{Alt}_4(R)\to \mathrm{Alt}_4(R)$ given by

$$p_1(x)=\begin{pmatrix} 0 & -a_{34} & a_{24} & -a_{23} \\ a_{34} & 0 & -a_{14} & a_{13} \\ -a_{24} & a_{14} & 0 & -a_{12} \\ a_{23} & -a_{13} & a_{12} & 0 \end{pmatrix}$$

has the required property. In particular we get $p_1^2=1$. Let now $p_1'\colon \mathrm{Alt}_4(R)\xrightarrow{\sim}\mathrm{Alt}_4(R)$ be such that $p_1'(x)\,x\in R$ for every $x\in\mathrm{Alt}_4(R)$. We denote the standard basis of $M_4(R)$ by $\{E_{ij}\}$, $1\leqq i,j\leqq 4$. Let $\{g_i\}$, $1\leqq i\leqq 6$, be the basis $\{H_{ij}=E_{ij}-E_{ji}\}$ of $\mathrm{Alt}_4(R)$, ordered lexicographically. The condition $p_i'(g_1)\,g_1\in R$ implies that $p_1'(g_1)=\rho g_6$, $\rho\in R$ and $p_1'(g_1)\,g_6\in R$ implies that $p_1'(g_6)=\rho' g_1$ for $\rho'\in R$. The condition $p_1'(g_1+g_6)(g_1+g_6)\in R$ then implies $\rho=\rho'$ and $p_1'=\rho p_1$ follows by further similar computations. If A is not split, let $\alpha\colon S\otimes A\xrightarrow{\sim}\mathrm{End}_S(V)$ be a faithfully flat splitting of A with V free over S such that $\sigma_1=\alpha(1\otimes\sigma)\alpha^{-1}$ is induced by $h\colon V^*\xrightarrow{\sim}V$. Then $Pf\colon \mathrm{Alt}^\sigma(\mathrm{End}_S(V))\to Pf(\mathrm{End}_S(V))$ is the composite map $\mathrm{Alt}^\sigma(V\otimes_S V^*)\xrightarrow{1\otimes h}\mathrm{Alt}(V\otimes_S V)\xrightarrow{\sim}\wedge^2 V\xrightarrow{Pf}\wedge^4 V$. Let $\{e_i\}$, $1\leqq i\leqq 4$, be an ordered basis of V such that $(1\otimes\lambda)(e_1\wedge e_2\wedge e_3\wedge e_4)=1$. We identify $\mathrm{End}_S(V)$ and $V\otimes_S V$ with $M_4(S)$, $\mathrm{Alt}(V\otimes_S V)$ with $\mathrm{Alt}_4(S)$, and $Pf(\mathrm{End}_S(V))$ with S through the above choice of a basis of V. Let u be the matrix of $h\colon V^*\xrightarrow{\sim}V$ with respect to the basis $\{e_i\}$ and the dual basis $\{e_i^*\}$. We have $u^t=\varepsilon u$, where ε is the type of σ, and $\sigma_1(x)=uxu^{-1}$. Since

$$x-\varepsilon uxu^{-1}=(xu-(xu)^t)\,u^{-1}=u(u^{-1}x-(u^{-1}x)^t),$$

the maps $y\mapsto\alpha(1\otimes y)\,u$ and $y\mapsto u^{-1}\alpha(1\otimes y)$ induce isomorphisms $S\otimes\mathrm{Alt}^\sigma(A)\xrightarrow{\sim}\mathrm{Alt}_4(S)$. The map $1\otimes h\colon V\otimes V^*\to V\otimes V$ is given by $z\mapsto zu$, $z\in M_4(S)$. Thus we have $Pf_\lambda(y)=pf(\alpha(1\otimes y)\,u)$. We define $p(y)=\alpha^{-1}(up_1(\alpha(1\otimes y)\,u))$. By construction, we get

$$\alpha(yp(y))=\alpha(1\otimes y)\,up_1(\alpha(1\otimes y)\,u)=Pf_\lambda(y)$$

and

$$\begin{aligned}\alpha(p(y)\,y)&=up_1(\alpha(1\otimes y)\,u)\,\alpha(1\otimes y)\,uu^{-1}\\&=upf(\alpha(1\otimes y)\,u)\,u^{-1}=Pf_\lambda(y)\end{aligned}$$

The fact that $p(y)\in\mathrm{Alt}^\sigma(A)$ follows easily by faithfully flat descent. We now verify that $p^2=d$. It follows from the relation

$$pf(uxu^t)=\det(u)\,pf(x)$$

for the classical pfaffian, that

$$p_1(uxu^t)=\det(u)\,u^{-1}p_1(x)\,(u^t)^{-1}.$$

Thus we get

$$\begin{aligned}\alpha(p^2(y))&=up_1(up_1(\alpha(1\otimes y)\,u)\,u)\\&=\varepsilon\det(u)\,uu^{-1}p_1^2(\alpha(1\otimes y)\,u)\,(u^t)^{-1}\\&=d\alpha(1\otimes y)\end{aligned}$$

since $\det(u)=d$. The last claim, about p', follows easily by descent from the corresponding result for p'_1, proved above. □

We call p the *pfaffian adjoint* of A (for the involution σ and the choice $\lambda: Pf(A) \xrightarrow{\sim} R$).

Corollary 3.2 *Let σ_1, σ_2 be two involutions on A such that the corresponding pfaffians $Pf_1(A)$, $Pf_2(A)$ are free and let $\lambda_i: Pf_i(A) \xrightarrow{\sim} R$ be isomorphisms. Then the quadratic spaces $(\mathrm{Alt}^{\sigma_1}(A), Pf_{\lambda_1})$ and $(\mathrm{Alt}^{\sigma_2}(A), Pf_{\lambda_2})$ are similar.*

Proof. We first assume that R is local. The automorphism $\sigma_1\sigma_2^{-1}$ of A then is inner, say $\sigma_1(x)=v\sigma_2(x)v^{-1}$. We have $\sigma_i(v)=\mu v$ where $\mu=\varepsilon_1\varepsilon_2$, ε_i being the type of σ_i. Further $x\mapsto xv$ and $x\mapsto v^{-1}x$ are isomorphisms $\mathrm{Alt}^{\sigma_1}(A) \xrightarrow{\sim} \mathrm{Alt}^{\sigma_2}(A)$. Let $p'_2(x)=vp_1(xv)$. By 3.1, there is $\rho\in R^*$ such that $p'_2=\rho p_2$ and $x\mapsto xv$ is a similitude with multiplier ρ. In general, we put

$$I=\{x\in A \mid \sigma_1\sigma_2^{-1}(a)\,x=xa, \forall a\in A\}$$

so that I is locally generated by v as above. Since ρ (as above) is uniquely determined by v, we can define $b: I\otimes I\to R$ locally by $b(v,v)=\rho$ and the multiplication in A induces an isometry

$$(\mathrm{Alt}^{\sigma_1}(A), Pf_{\lambda_1})\otimes(I,b) \xrightarrow{\sim} (\mathrm{Alt}^{\sigma_2}(A), Pf_{\lambda_2}). \quad \square$$

We now compute the Clifford algebra of $(\mathrm{Alt}^{\sigma}(A), Pf_{\lambda})$. We take on $M_2(A)$ the checker-board gradation.

Proposition 3.3 *The map $\mathrm{Alt}^{\sigma}(A)\to M_2(A)$ given by $x\mapsto \begin{pmatrix} 0 & x \\ p(x) & 0 \end{pmatrix}$ yields an isomorphism of graded algebras $C(\mathrm{Alt}^{\sigma}(A), Pf_{\lambda}) \xrightarrow{\sim} M_2(A)$. In particular $C_0(\mathrm{Alt}^{\sigma}(A), Pf_{\lambda})$ is isomorphic to $A\times A$ and the centre of $C_0(\mathrm{Alt}^{\sigma}(A), Pf_{\lambda})$ is isomorphic to $R\times R$. Furthermore, the standard involution of $C(Alt^{\sigma}(A), Pf_{\lambda})$ corresponds to*

$$\tau\begin{pmatrix} a & b \\ c & d \end{pmatrix}=\begin{pmatrix} \sigma(d) & \varepsilon\sigma(b) \\ \varepsilon\sigma(c) & \sigma(a) \end{pmatrix}, \quad a,b,c,d\in M_2(A),$$

where ε is the type of σ.

Proof. It is immediate from 3.1, by the universal property of the Clifford algebra, that the given map induces a homomorphism of algebras $C(\mathrm{Alt}^{\sigma}(A), Pf_{\lambda}) \to M_2(A)$. Since both are Azumaya algebras of the same rank this map is an isomorphism. The last claim follows from the fact that τ is an involution and its restriction to the image of $\mathrm{Alt}^{\sigma}(A)$ in $M_2(A)$ is minus the identity. □

Let $Q(A)$ be the similitude class of $(\mathrm{Alt}^{\sigma}(A), Pf_{\lambda})$. Obviously $Q(A)$ does not depend on λ and, in view of 3.2, $Q(A)$ is independent of σ. It follows that $Q(A)=Q(B)$ if A and B are isomorphic Azumaya algebras of rank 16. We now prove the converse, so that:

Theorem 3.4 *Two Azumaya algebras of rank 16 with involutions are isomorphic (as algebras) if and only if $Q(A)=Q(B)$.*

Proof. As already remarked, $A \simeq B$ implies $Q(A)=Q(B)$. Conversely, if we have $(\mathrm{Alt}^\sigma(A), Pf_{\lambda_1}) \simeq (\mathrm{Alt}^\sigma(B), Pf_{\lambda_2}) \otimes (I, d)$, then (see Sect. 1)

$$C_0(\mathrm{Alt}^\sigma(A), Pf_{\lambda_1}) \simeq C_0(\mathrm{Alt}^\sigma(B), Pf_{\lambda_2})$$

and, by 3.3, $A \times A \simeq B \times B$. This implies that $A \simeq B$. □

Remark. The explicit description of the Clifford algebra of the pfaffian associated to a rank 16 algebra with involution, given in 3.3, will be used in Sect. 6 to compute the group of special similitudes of a rank 6 quadratic space.

4 The Albert form, a theorem of Jacobson

In this section, we assume that $2 \in R^*$. Let $A = A_1 \otimes A_2$, A_i being rank 4 Azumaya algebras over R. Let $\sigma = \sigma_1 \otimes \sigma_2$ denote the tensor product of the standard involutions on A_i (i.e. such that $\sigma_i(x)\, x \in R$ is the reduced norm). By 2.2 the discriminant module $(Pf(A), \tilde{\sigma})$ is trivial. Let $\lambda: (Pf(A), \tilde{\sigma}) \simeq (R, 1)$ be the isometry given in 2.2. We shall describe the space $(\mathrm{Alt}(A), Pf_\lambda)$ as the Albert form of the pair (A_1, A_2) defined as follows. Let A_i' denote the orthogonal complement of 1 in A_i with respect to the reduced norm, i.e. $A_i' = \{x \in A_i \mid tr_A(x) = 0\}$, tr_A denoting the reduced trace. Let $n_i: A_i' \to R$ be the restriction of the reduced norm to A_i'. The *Albert form* of the pair (A_1, A_2) is the quadratic space

$$Q(A_1, A_2) = (A_1', n_1) \perp (A_2', -n_2).$$

Proposition 4.1 $(\mathrm{Alt}^\sigma(A), Pf_\lambda) \simeq Q(A_1, A_2)$.

Proof. We obviously have $A_1' \otimes A_2' \subset \mathrm{Alt}^\sigma(A)$. Thus it clearly suffices to prove the proposition in a faithfully flat splitting. Let $\alpha: S \otimes A_i \xrightarrow{\sim} M_2(S)$, $i=1, 2$, be a faithfully flat splitting of A_i such that $\alpha_i(1 \otimes \sigma_i)\alpha_i^{-1}$ is the standard involution σ' of $M_2(S)$, i.e. $\sigma'(x) = u x^t u^{-1}$, $x \in M_2(S)$, $u = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. We put $\sigma'(x) = \bar{x}$. Then $\alpha_1 \otimes \alpha_2$: $S \otimes (A_1 \otimes A_2) \simeq M_2(S) \otimes M_2(S) = M_4(S)$ transports $\sigma_1 \otimes \sigma_2$ into the involution

$$\sigma' \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \bar{d} & -\bar{b} \\ -\bar{c} & \bar{a} \end{pmatrix}, \quad a, b, c, d \in M_2(S).$$

The type of σ' is 1 since the type of σ_i is -1. Thus,

$$\begin{aligned} \mathrm{Alt}^\sigma(M_4(S)) &= \{x - \sigma'(x) \mid x \in M_4(S)\} \\ &= \left\{ \begin{pmatrix} a & x \cdot 1 \\ y \cdot 1 & -\bar{a} \end{pmatrix}, x, y \in S, a \in M_2(S) \right\}. \end{aligned}$$

The basis e_σ of $Pf\, A$ constructed in 2.2 is

$$e_\sigma = (\det u)^2 g_1 \wedge g_2 \wedge g_3 \wedge g_4 = g_1 \wedge g_2 \wedge g_3 \wedge g_4$$

where $\{g_i\}$, $1 \leqq i \leqq 4$, is the standard basis of S^4. In view of the computation of Pf_λ given in Sect. 2,

$$\begin{aligned} Pf_\lambda \begin{pmatrix} a & x\cdot 1 \\ y\cdot 1 & -\bar{a} \end{pmatrix} &= pf\left(\begin{pmatrix} a & x\cdot 1 \\ y\cdot 1 & -\bar{a} \end{pmatrix} \begin{pmatrix} 0 & u \\ -u & 0 \end{pmatrix} \right) \\ &= Pf \begin{pmatrix} -xu & au \\ \bar{a}u & yu \end{pmatrix} \\ &= -xy - \det(a). \end{aligned}$$

Let $a = z\cdot 1 + a'$, $z \in S$, $a' \in M_2(S)$ with $tr(a') = 0$. Then,

$$\begin{pmatrix} a & x\cdot 1 \\ y\cdot 1 & -\bar{a} \end{pmatrix} = \begin{pmatrix} z & x \\ y & -z \end{pmatrix} \otimes 1 + 1 \otimes a' \in M_2(S) \otimes M_2(S)$$

with $\begin{pmatrix} z & x \\ y & -z \end{pmatrix}$, $a' \in M_2(S)' = \{a \in M_2(S) \mid tr(a) = 0\}$ and

$$\begin{aligned} Pf_\lambda \begin{pmatrix} a & x\cdot 1 \\ y\cdot 1 & -\bar{a} \end{pmatrix} &= \det \begin{pmatrix} z & x \\ y & -z \end{pmatrix} - \det(a') \\ &= n' \begin{pmatrix} z & x \\ y & -z \end{pmatrix} - n'(a') \end{aligned}$$

n' denoting the restriction of the determinant form to $M_2(S)'$. Thus,

$$\begin{aligned} (\mathrm{Alt}^\sigma(M_4(S)), Pf_\lambda) &\simeq (M_2(S)', n') \perp (M_2(S)', -n') \\ &\simeq Q(M_2(S), M_2(S)). \end{aligned}$$

This proves the proposition. □

The following theorem includes a result of Jacobson for fields (see [J]).

Theorem 4.2 *Let A_1, A_2, B_1, B_2 be Azumaya algebras of rank 4 over a ring R in which 2 is invertible. Then $A_1 \otimes A_2 \simeq B_1 \otimes B_2$ if and only if the Albert forms $Q(A_1, A_2)$ and $Q(B_1, B_2)$ are similar.*

Proof. The claim is an immediate consequence of 4.1 and 3.4. □

5 Splitting of rank 16 algebras into tensor products of rank 4 algebras, a theorem of Albert and an application

In this section, we give some conditions under which rank 16 Azumaya algebras, whose classes in the Brauer group $\mathrm{Br}(R)$ are of order 2, split into a tensor product of rank 4 algebras. All these results are generalizations of a result of Albert for fields (see [A]). We begin with the following

Theorem 5.1 *Let A be an Azumaya algebra of rank 16 over a semilocal ring R. If the class of A in the Brauer group of R is of order 2, then A is a tensor product of rank 4 Azumaya algebras.*

Proof. Since the class of A in the Brauer group of R is of order 2 and R is semilocal, there exists an involution σ on A (see [S]). Let λ: $Pf(A) \to R$ be chosen such that $(\mathrm{Alt}^\sigma(A), Pf_\lambda)$ represents 1. Any quadratic space over a semilocal ring is the orthogonal sum of binary spaces. So we can decompose $Pf_\lambda \simeq q_1 \perp q_2 \perp q_3$, q_i binary forms, and we can assume that q_1 represents 1. If $q = q_2 \perp q_3$, and if δ denotes the signed discriminant, we have $\delta(q) = \delta(q_1)$, $\delta(Pf_\lambda)$ being 1, further $C(Pf_\lambda) \simeq C(q_1) \otimes C(\delta(q_1)\, q)$ as graded algebras and $C(q_1) \simeq M_2(R)$. It follows that $C_0(Pf_\lambda)\, e \simeq C(\delta(q_1)\, q)$, where e is an idempotent generating the centre $R \times R$ of $C_0(Pf_\lambda)$. In view of 3.3 we get $A \simeq C(\delta(q_1)\, q) \simeq C(\delta(q_1)\, q_2) \otimes C(\delta(q_1)\, \delta(q_2)\, q_3)$. □

Observe that we do not assume 2 invertible in 5.1. For the next result we need that 2 is a unit. We have seen in Sect. 2 that if $A = A_1 \otimes A_2$ is a tensor product of two Azumaya algebras of rank 4, then A has an orthogonal involution σ, namely $\sigma_1 \otimes \sigma_2$, σ_i denoting the standard involutions of A_i, with $(Pf_{\sigma_1 \otimes \sigma_2}(A), \tilde{\sigma})$ trivial as a discriminant module. The following is a converse to this fact.

Theorem 5.2 *Let A be an Azumaya algebra of rank 16 over a ring R in which 2 is invertible. Suppose A has an orthogonal involution σ such that $(Pf(A), \tilde{\sigma})$ is trivial as a discriminant module. Then A is isomorphic as an algebra with involution to a tensor product of rank 4 Azumaya algebras with involution the tensor product of the standard involutions.*

Proof. Let λ: $(Pf(A), \tilde{\sigma}) \xrightarrow{\sim} (R, 1)$ be an isometry. By 3.1, there exists a map p: $\mathrm{Alt}^\sigma(A) \to \mathrm{Alt}^\sigma(A)$ such that $x\,p(x) = p(x)\,x = Pf_\lambda(x)$ and $p^2 = 1$. Let

$$\mathrm{Alt}^\sigma(A)^+ = \{x + p(x) \mid x \in \mathrm{Alt}^\sigma(A)\}$$
$$\mathrm{Alt}^\sigma(A)^- = \{x - p(x) \mid x \in \mathrm{Alt}^\sigma(A)\}.$$

It may be verified, using the identities $x\,p(x) = p(x)\,x = Pf_\lambda(x)$ and $p^2(x) = x$, $x \in \mathrm{Alt}^\sigma(A)$, that $\mathrm{Alt}^\sigma(A)^+$ and $\mathrm{Alt}^\sigma(A)^-$ are mutually orthogonal for the form Pf_λ and $\mathrm{Alt}^\sigma(A) = \mathrm{Alt}^\sigma(A)^+ \oplus \mathrm{Alt}^\sigma(A)^-$. Let n^+ and n^- denote the restictions of Pf_λ to $\mathrm{Alt}^\sigma(A)^+$ and $\mathrm{Alt}^\sigma(A)^-$ respectively. Then,

$$(\mathrm{Alt}^\sigma(A), Pf_\lambda) \simeq (\mathrm{Alt}^\sigma(A)^+, n^+) \perp (\mathrm{Alt}^\sigma(A)^-, n^-).$$

We verify that $\mathrm{Alt}^\sigma(A)^+$ and $\mathrm{Alt}^\sigma(A)^-$ are rank 3 modules by going over to a splitting. Since $2 \in R^*$, there exists a faithfully flat splitting $S \otimes A \simeq M_4(S)$ for A such that $1 \otimes \sigma$ transports to $x \mapsto x^t$ on $M_4(S)$. Clearly, $p = p_1$ (see Sect. 3) and $S \otimes \mathrm{Alt}^\sigma(A)^+ = \{x + p_1(x), x \in \mathrm{Alt}_4(S)\}$, $S \otimes \mathrm{Alt}^\sigma(A)^- = \{x - p_1(x), x \in \mathrm{Alt}_4(S)\}$ are rank 3 free modules over S. Since the Arf invariant of $(\mathrm{Alt}^\sigma(A), Pf_\lambda)$ is trivial, $\mathrm{disc}(\mathrm{Alt}^\sigma(A), Pf_\lambda) \cong (R, -1)$, disc denoting the unsigned discriminant. In particular, $\mathrm{disc}(\mathrm{Alt}^\sigma(A)^+, n^+) \simeq (R, -1) \otimes \mathrm{disc}(\mathrm{Alt}^\sigma(A)^-, n^-)$. In view of ([KOS]), any rank 3 quadratic space over R is similar to (B', n), where B is a rank 4 Azumaya algebra, B' its submodule of trace zero elements and n the reduced norm. Thus there exist Azumaya algebras B_1, B_2 and a discriminant module (I, d) such that

$$(\mathrm{Alt}^\sigma(A)^+, n^+) \simeq (I, d) \otimes (B_1', n)$$
$$(\mathrm{Alt}^\sigma(A)^-, n^-) \simeq (I, d) \otimes (B_2', -n)$$

(noting that $\mathrm{disc}(B_i', n) \simeq (R, 1)$). Thus,

$$(\mathrm{Alt}^\sigma(A), Pf_\lambda) \simeq (I, d) \otimes Q(B_1, B_2),$$

where $Q(B_1, B_2)$ is the Albert form of the pair (B_1, B_2) and by 3.4 the algebras A and $B_1 \otimes B_2$ are isomorphic. The fact that there exists an isomorphism of algebras with involution is a consequence of the following proposition. The notations are the same as in the proof of 5.2.

Proposition 5.3 *Let R be a ring in which 2 is invertible. Let (A, σ), (B, τ) be Azumaya algebras of rank 16 with involutions of orthogonal type. If the discriminant modules $(Pf(A), \tilde{\sigma})$ and $(Pf(B), \tilde{\tau})$ are trivial, the following properties are equivalent:*

1) *A and B are isomorphic as algebras with involution.*

2) *There exists a similitude $f: (\mathrm{Alt}^\sigma(A), Pf_{\lambda_1}) \simeq (\mathrm{Alt}^\tau(B), Pf_{\lambda_2}) \otimes (I, d)$ such that*

$$f(\mathrm{Alt}^\sigma(A)^+) = \mathrm{Alt}^\tau(B)^+ \otimes I \quad \text{and} \quad f(\mathrm{Alt}^\sigma(A)^-) = \mathrm{Alt}^\tau(B)^- \otimes I.$$

Proof. 1)$\Rightarrow$2): Let $\varphi: (A, \sigma) \simeq (B, \tau)$ be an isomorphism of algebras with involution, Then φ restricts to an isomorphism $\varphi: \mathrm{Alt}^\sigma(A) \xrightarrow{\sim} \mathrm{Alt}^\tau(B)$ as R-modules and gives rise to an isometry $\tilde{\varphi}: (Pf(A), \tilde{\sigma}) \xrightarrow{\sim} (Pf(B), \tilde{\tau})$. Let $\lambda_1: (Pf(A), \tilde{\sigma}) \xrightarrow{\sim} (R, 1)$ be an isometry. We set $\lambda_2 = \lambda_1 \circ \tilde{\varphi}^{-1}$. Then φ is an isometry of $(\mathrm{Alt}^\sigma(A), Pf_{\lambda_1})$ with $(\mathrm{Alt}^\tau(B), Pf_{\lambda_2})$. Let p_σ, p_τ respectively, be the map p defined in 3.1 for $\mathrm{Alt}^\sigma(A)$, $\mathrm{Alt}^\tau(B)$ respectively. After a scaling, $p_\tau = \pm \varphi p_\sigma \varphi^{-1}$. Thus we get $\varphi(\mathrm{Alt}^\sigma(A)^+) = \mathrm{Alt}^\tau(B)^+$, $\varphi(\mathrm{Alt}^\sigma(A)^-) = \mathrm{Alt}^\tau(B)^-$.

2)$\Rightarrow$1): The similitude f induces an isomorphism $C_0(f)$: $C_0(\mathrm{Alt}^\sigma(A), Pf_{\lambda_1}) \xrightarrow{\sim} C_0(\mathrm{Alt}^\tau(B), Pf_{\lambda_2})$ which satisfies $C_0(f)(xy) = i(f(x) \otimes f(y))$, $x, y \in \mathrm{Alt}^\sigma(A)$, where i: $C_0(\mathrm{Alt}^\tau(B), Pf_{\lambda_2}) \otimes I \otimes C_0(\mathrm{Alt}^\tau(B), Pf_{\lambda_2}) \otimes I \to C_0(\mathrm{Alt}^\tau(B), Pf_{\lambda_2})$ is the map $i(x \otimes \alpha \otimes y \otimes \beta) = xyd(\alpha\beta)$. We have $C_0(\mathrm{Alt}^\sigma(A)) = A \times A$, $C_0(\mathrm{Alt}^\tau(B)) = B \times B$. We claim that the composite

$$f^*: A \xrightarrow{\Delta} A \times A \xrightarrow{C_0(f)} B \times B \xrightarrow{\pi} B,$$

where Δ is the diagonal map and π is one of the projections, is an isomorphism respecting the involutions. To prove this, we may assume (by localizing) that $(I, d) = (R, \mu)$ so that f: $\mathrm{Alt}^\sigma(A) \xrightarrow{\sim} \mathrm{Alt}^\tau(B)$ is a similitude with multiplier $\mu \in R^\bullet$. The image of $\mathrm{Alt}^\sigma(A) \hookrightarrow M_2(A)$, $x \mapsto \begin{pmatrix} 0 & x \\ p_\sigma(x) & 0 \end{pmatrix}$ (see 3.3) generates $M_2(A)$. Since

$$\mathrm{Alt}^\sigma(A) = \mathrm{Alt}^\sigma(A)^+ \oplus \mathrm{Alt}^\sigma(A)^-$$

and $p_\sigma(x) = x$ if $x \in \mathrm{Alt}^\sigma(A)^+$, $p_\sigma(x) = -x$ if $x \in \mathrm{Alt}^\sigma(A)^-$, an easy verification shows that $\Delta(A)$ is generated by products xy with x, y both in $\mathrm{Alt}^\sigma(A)^+$ or $\mathrm{Alt}^\sigma(A)^-$. Since $f(\mathrm{Alt}^\sigma(A)^+) = \mathrm{Alt}^\tau(B)^+$ and $f(\mathrm{Alt}^\sigma(A)^-) = \mathrm{Alt}^\tau(B)^-$ and $C_0(f)(xy) = \mu^{-1} f(x) f(y)$, it follows that $C_0(f)$ maps $\Delta(A)$ to $\Delta(B)$. The map $C_0(f)$ respects the standard involution on the even Clifford algebras and the involutions on $C_0(\mathrm{Alt}^\sigma(A))$, $C_0(\mathrm{Alt}^\tau(B))$ respectively, are given by $(a, b) \mapsto (\sigma(b), \sigma(a))$, $(a, b) \mapsto (\tau(b), \tau(a))$ respectively. Thus f^* respects the involutions. □

The next results are immediate consequences of 5.2. In view of the example given in [ART] (see the Introduction), they are of interest even for central simple algebras over fields.

Corollary 5.4 *Let R be a ring in which 2 is invertible and let A be an Azumaya algebra of rank 16 which has an involution σ of orthogonal type. There exists a σ-invariant rank 4 Azumaya subalgebra of A if and only if $(Pf(A), \tilde{\sigma})$ is a trivial discriminant module.*

Proof. If $(Pf(A), \tilde{\sigma})$ is trivial, there exists a σ-invariant rank 4 Azumaya subalgebra in view of 5.2. Conversely, if there exists such an invariant subalgebra A_1, then $A \simeq A_1 \otimes A_2$, σ is the tensor product of involutions σ_1, σ_2 on A_1, A_2 and the claim follows from 2.2. □

Corollary 5.5 *Let R be a ring in which 2 is invertible and let A be an Azumaya algebra of rank 16 which has an involution σ of orthogonal type. Assume that σ has an invariant Azumaya subalgebra of rank 4. Let u be a unit of A such that $\sigma(u) = u$. The involution $\tau = \mathrm{Int}\, u \circ \sigma$ admits a τ-invariant rank 4 Azumaya subalgebra if and only if $n_A(u)$ is a square in R.*

Proof. The claim follows from 5.4 and 2.1. □

Remark 1 Let K be a field of characteristic not 2, with separable closure $\bar{K}$ and Galois group G. Let O_n be the orthogonal group of the diagonal form $\langle 1, \ldots, 1 \rangle$ and let PO_n be the corresponding projective group. Let n be even. It is known that $H^1(G, PO_n)$ classifies central simple K-algebras of dimension n^2 with an involution of orthogonal type. Central simple K-algebras of dimension n^2 with an involution of orthogonal type of trivial pfaffian discriminant are classified by $H^1(G, PSO_n)$ (see [KS]).

Remark 2 We give an example of a domain R, in which 2 is invertible, which admits an Azumaya algebra of rank 16 whose class is of order 2 in $\mathrm{Br}(R)$, but which is not a tensor product of rank 4 Azumaya algebras. Let $R = K[X, Y]$, K being a field of characteristic $\neq 2$ which admits a rank 6 anisotropic quadratic space q_0 of discriminant -1 (i.e. the Arf invariant of q_0 is trivial). In view of [P], there exists an indecomposable quadratic space q of rank 6 over R whose reduction modulo (X, Y) is isometric to q_0. Thus $\mathrm{disc}\, q = -1$. Let C_0 be the Clifford algebra of q and e an idempotent generating the centre of C_0. Then $C_0 e$ is a rank 16 Azumaya algebra over R, which is of order 2 in $\mathrm{Br}(R)$. We claim that $C_0 e$ is not isomorphic to a tensor product of rank 4 Azumaya algebras. Suppose, indeed, that $C_0 e \tilde{\to} A_1 \otimes A_2$, A_i rank 4 Azumaya algebras. By Theorem (9.4) of [KPS], q would be similar to the Albert form $Q(A_1, A_2)$ which is an orthogonal sum of spaces of rank 3, a contradiction to the indecomposability of q. We observe that, in contrast, if $R = K[X]$, char $K \neq 2$, every Azumaya algebra A over R whose class is of order 2 in $\mathrm{Br}(R)$, being extended from K, is a tensor product of rank 4 Azumaya algebras.

Remark 3 Let K be a field of characteristic $\neq 2$ and let $L = K(\sqrt{-d})$ be a quadratic field extension of K. Let D be a rank 16 central division algebra over L whose class in $\mathrm{Br}(L)$ is of order 2. To D we have associated the similarity class $Q(D)$ of a 6-dimensional form. We claim that D is extended from a division algebra over K if and only if the similarity class $Q(D)$ is extended from K. Let q be in the class $Q(D)$ such that $\lambda q \tilde{\to} L \otimes \langle a_0, \ldots, a_5 \rangle$ for some $\lambda \in L$ and

$a_0, \ldots, a_5 \in K^*$. If $\Pi_i a_i \equiv -1$ in K^*/K^{*2}, then $\langle a_0, \ldots, a_5\rangle$ is similar to a form $(\mathrm{Alt}^\sigma(H_1 \otimes H_2), Pf)$, H_i, $i=1, 2$ being quaternion algebras over K, and hence $D \tilde{\to} L \otimes H_1 \otimes H_2$ by our classification. If $\Pi_i a_i \not\equiv -1$ in L, then $\Pi_i a_i \equiv d$ in K^*/K^{*2}, since $\Pi_i a_i \equiv -1$ in L. Then $\langle a_0, \ldots, a_4, -da_5\rangle$ has trivial Arf invariant and its similarity class over L is also $Q(D)$. The proof can be completed as in the case $\Pi_i a_i \equiv -1$. The converse is clear. We remark that the statement is analogous to the statement for rank 4 central division algebras that the algebra is extended if and only if the reduced norm is extended.

6 The group of similitudes

Let (M, q) be a quadratic space over R. We recall that a similitude of (M, q) with multiplier $\mu \in R^*$ is an R-linear isomorphism $f: M \tilde{\to} M$ such that $q(f(x)) = \mu q(x)$ for all $x \in M$. A similitude f induces an automorphism $C_0(f)$ of $C_0(M, q)$ such that $C_0(f)(xy) = \mu^{-1} f(x) f(y)$ for $x, y \in M$ (see Sect. 1). We call a similitude f *special* if $C_0(f)$ induces the identity map on the centre Z of $C_0(M, q)$. We denote by $GO(M, q)$ the group of similitudes of (M, q) and by $GO_+(M, q)$ the subgroup of special similitudes. In this section, we shall compute the group $GO_+(M, q)$ of a rank 6 quadratic space (M, q) over any commutative ring R with $\mathrm{Pic}(R)$ trivial. We also indicate the computation of this group if the rank of M is equal to 4.

Let (M, q) be a quadratic space over R of rank 6 and with trivial Arf invariant. Let $A = C_0(M, q)e$, where e is an idempotent of $C_0(M, q)$ generating its centre over R and let $P = C_1 f$, where $f = 1 - e$. We have an isomorphism φ: $A \otimes_R A \tilde{\to} \mathrm{End}_R(P)$ defined by $\varphi(a \otimes b)(x) = axt(b)$, τ denoting the standard involution on $C(M, q)$ ([KPS]). Let $Pf(P)$ denote the pfaffian defined with respect to φ. There exists an isomorphism λ: $Pf(P) \tilde{\to} R$ such that (M, q) is isometric to $(\mathrm{Alt}(P), Pf_\lambda)$ under the map θ: $M \to \mathrm{Alt}(P)$ given by $\theta(x) = xf$ (see [KPS]).

Lemma 6.1 *Let $\eta \in R^*$, $a \in A^*$. The map $\psi_{(\eta, a)} = \eta\varphi(a \otimes a)$ restricted to $\mathrm{Alt}(P)$ is a special similitude with multiplier $\eta^2 n(a)$, n denoting the reduced norm of A.*

Proof. To prove that $\psi_{(\eta, a)}$ is a special similitude, with multiplier $\eta^2 n(a)$, we may replace R by a faithfully flat extension. We therefore assume, without loss of generality, that $A = \mathrm{End}_R(V)$ with V free, $P = V \otimes V$, φ: $\mathrm{End}_R(V) \otimes \mathrm{End}_R(V) \tilde{\to} \mathrm{End}_R(V \otimes V)$ the canonical map, $\mathrm{Alt}(P) = \wedge^2 V$ and $\{e_i\}$, $1 \leqq i \leqq 4$ a basis of V such that $e_1 \wedge e_2 \wedge e_3 \wedge e_4 = \lambda^{-1}(1) \in Pf(P) = \wedge^4 V$. Clearly, $Pf\psi_{(\eta, a)} = \eta^2 n(a)\, Pf$ where $n(a)$ is the determinant (reduced norm) of $a \in \mathrm{End}_R(V)$. Thus $\psi_{(\eta, a)}$ is a similitude with multiplier $\eta^2 n(a)$. We show now that it is special. Let $\mu = \eta^2 n(a)$ and let g: $M \to M$ be the similitude $\theta^{-1}\psi_{(\eta, a)}\theta$ with multiplier μ, θ being the isometry of (M, q) with $(\mathrm{Alt}(P), Pf_\lambda)$ defined earlier. The map $M \to C_1(M, q)$ defined by $x \mapsto g(x)f + \mu^{-1} g(x)e$ extends, by the universal property of the Clifford algebra, to an isomorphism $\tilde{g}$ of $C(M, q)$ whose restriction to $C_0(M, q)$ is $C_0(g)$. We shall show that $C_0(g)$ is identity on $Z = Re \times Rf$. We first verify that $\tilde{g} = \mathrm{Int}(a + \eta^{-1}\tau(a)^{-1})$. Writing elements of $C(M, q)$ as matrices

$$\begin{pmatrix} C_0 e & C_1 f \\ C_1 e & C_0 f \end{pmatrix}$$

we need to check that

$$\begin{pmatrix} a & 0 \\ 0 & \eta^{-1}\tau(a)^{-1} \end{pmatrix}\begin{pmatrix} 0 & xf \\ xe & 0 \end{pmatrix}\begin{pmatrix} a^{-1} & 0 \\ 0 & \eta\tau(a) \end{pmatrix}=\begin{pmatrix} 0 & g(x)f \\ \mu^{-1}g(x)\,e & 0 \end{pmatrix}.$$

By definition, we have $g(x)f=\eta axf\tau(a)$ and by the following Lemma 6.2, we get $\eta^{-1}\tau(a)^{-1}xea^{-1}=\eta^{-2}n(a)^{-1}g(x)\,e$. Thus $\tilde{g}=\mathrm{Int}(a+\eta^{-1}\tau(a)^{-1})$, whose restriction $C_0(g)$ to the centre $Re\times Rf$ is identity. □

Lemma 6.2 *Let $g\colon M\xrightarrow{\sim}M$ be an R-linear isomorphism and $\eta\in R^\cdot$, $a\in A^\cdot=C_0\,e^\cdot$ be such that for all $x\in M$, $g(x)f=\eta a(xf)\,\tau(a)$. Then, for all $x\in M$,*

$$g(x)\,e=\eta n(a)\,\tau(a)^{-1}\,xea^{-1},$$

where n denotes the reduced norm on A.

Proof. It suffices to prove this in a splitting. We therefore assume $M=\mathrm{Alt}_4(R)$ with the form given by λpf, $\lambda\in R^\cdot$. The Clifford algebra is $M_2(M_4(R))$ and the embedding $M\to C_1(M,q)$ is given for $x\in M$ by

$$x\mapsto\begin{pmatrix} 0 & x \\ \lambda p_1(x) & 0 \end{pmatrix}$$

(where p_1 is as in Sect. 3). Let $a=\begin{pmatrix} v & 0 \\ 0 & 0 \end{pmatrix}\in A=C_0\,e$, $v\in M_4(R)$. Then $g(x)\,f=\begin{pmatrix} 0 & \eta vxv^t \\ 0 & 0 \end{pmatrix}$ and

$$g(x)\,e=\begin{pmatrix} 0 & 0 \\ p_1(\eta vxv^t) & 0 \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ \eta\det(v)\,(v^t)^{-1}p_1(xv^{-1}vxv^t) & 0 \end{pmatrix}$$

so that $g(x)\,e=\eta n(a)\,\tau(a)^{-1}xea^{-1}$. This proves the lemma. □

We thus have a homomorphism

$$\psi\colon R^\cdot\times A^\cdot\to GO_+(M,q)$$

given by $\psi(\eta,a)=\theta^{-1}\psi_{(\eta,a)}\theta$ (we recall that $\theta(x)=xf$ for $x\in M$). We claim that the kernel of this homomorphism is the subgroup $\{(\eta^2,\eta^{-1})\mid\eta\in R^\cdot\}$. To prove this, it suffices to show that the condition $\eta\varphi(a\otimes a)=1_P$ implies that a is central in A. This again can be proved by going over to a faithfully flat extension and assuming that $A=C_0\,e=\mathrm{End}_R(V)$ with V free. Let $\{e_i\}$, $1\leqq i\leqq 4$, be a basis of V such that $e_1\wedge e_2\wedge e_3\wedge e_4=\lambda^{-1}(1)\in Pf(P)$. If u is the matrix of $a\in\mathrm{End}_R(V)$ with respect to this basis, $\eta\varphi(a\otimes a)\colon \mathrm{Alt}(V\otimes V)\to\mathrm{Alt}(V\otimes V)$ is given by $x\mapsto\eta uxu^t$, $x\in\mathrm{Alt}_4(R)$. If $\eta uxu^t=x$, $\forall x\in\mathrm{Alt}_4(R)$, u is central in $\mathrm{End}_R(V)$ is a consequence of the following

Lemma 6.3 *Let $u\in Gl_4(R)$ and $\eta\in R^\cdot$ be such that $x=\eta uxu^t$ for all x in $\mathrm{Alt}_4(R)$. Then $u\in R^\cdot$.*

Proof. Let $u=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}$, α, β, γ, $\delta\in M_2(R)$. Choosing $x=\begin{pmatrix}\varepsilon & 0\\ 0 & 0\end{pmatrix}$ with $\varepsilon=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$, the condition $x=\eta uxu^t$ yields $\eta\alpha\varepsilon\alpha^t=\varepsilon$ and $\gamma=0$. Similarly, setting $x=\begin{pmatrix}0 & 0\\ 0 & \varepsilon\end{pmatrix}$, we get $\eta\delta\varepsilon\delta^t=\varepsilon$ and $\beta=0$. Taking $x=\begin{pmatrix}0 & \varepsilon\\ \varepsilon & 0\end{pmatrix}$, we get $\alpha=\delta$. Finally, writing $x=\eta uxu^t$ for $x=\begin{pmatrix}0 & \beta\\ -\beta^t & 0\end{pmatrix}$, β arbitrary in $M_2(R)$, we see that α is a scalar.

Theorem 6.4 *Let (M, q) be a quadratic space of rank 6 with trivial Arf invariant, let e be an idempotent which generates the centre of C_0 and let $A=C_0 e$. There exists an exact sequence*

$$1\to(R^*\times A^*)/\{(\eta^2,\eta^{-1}),\eta\in R^*\}\xrightarrow{\bar{\psi}} GO_+(M,q)\xrightarrow{\nu}\mathrm{Pic}(R).$$

Proof. The map $\bar{\psi}$ is induced by ψ: $R^*\times C_0 e^*\to GO_+(M, q)$ as defined above and is injective by 6.3. Let $g\in GO_+(M, q)$ with μ as the multiplier. The map $M\to C(M, q)$ given by $x\mapsto g(x)f+\mu^{-1}g(x)e$ extends to an automorphism $\tilde{g}$ of $C(M, q)$. Let $I_{\tilde{g}}=\{x\in C(M, q)\mid\tilde{g}(x)y=yx,\ \forall y\in C(M, q)\}$. Since $C(M, q)$ is an Azumaya algebra, $I_{\tilde{g}}$ is an invertible R-module and $\tilde{g}$ is inner if and only if the class $[I_{\tilde{g}}]$ in $\mathrm{Pic}(R)$ is trivial. We put $\nu(g)=[I_{\tilde{g}}]$. Assume that this class is trivial, so that there exists $u\in C(M, q)$ such that $\tilde{g}=\mathrm{Int}\,u$. Since the restriction of $\tilde{g}$ to $C_0(M, q)$ is identity on its centre $Re\times Rf$, $ue=eu$ and $uf=fu$. This implies that $u\in C_0(M, q)^*$. Let $u=a+b$, $a\in C_0 e$, $b\in C_0 f$. We have $uxu^{-1}=g(x)f+\mu^{-1}g(x)e$ for $x\in M$. Since $\tau(g(x)f+\mu^{-1}g(x)e)=g(x)f+\mu^{-1}g(x)e$, $\tau(uxu^{-1})=uxu^{-1}$ for $x\in M$; i.e. $\tau(u)u$ commutes with M and hence is central in $C(M, q)$. Let $\tau(u)u=\eta\in R^*$. Then $(\tau(b)+\tau(a))(a+b)=\tau(b)a+\tau(a)b=\eta$ with $\tau(b)a\in C_0 e$, $\tau(a)b\in C_0 f$. Thus, $\tau(b)a=\eta e$, $\tau(a)b=\eta f$. Further, $(a+\eta\tau(a)^{-1}f)x=(g(x)f+\mu^{-1}g(x)e)(a+\eta\tau(a)^{-1}f)$; i.e. $g(x)f=\eta^{-1}ax\tau(a)f+\eta^{-1}axf\tau(a)$. Since $g(x)f=(\theta g\theta^{-1})(xf)=\psi_{(\eta^{-1},a)}(xf)$ for all $x\in M$, it follows that $g=\psi(\eta^{-1}, a)$. □

We next compute the group $GO_+(M, q)$ if (M, q) is a rank 6 quadratic space with arbitrary Arf invariant. Let Z be the centre of C_0 and let τ denote the restriction of the standard involution τ of $C(M, q)$ to C_0. Then τ is an involution of second kind on the rank 16 Azumaya algebra C_0 over Z. In particular τ induces the (unique) nontrivial R-linear involution of the quadratic algebra Z. We have an isomorphism of Z-algebras $Z\otimes Z\xrightarrow{\sim} Z\times Z$ given by $x\otimes y\mapsto(xy, x\tau(y))$. It follows that $Z\otimes(M, q)$ is a quadratic space of rank 6 over Z with trivial Arf invariant. Thus $Z\otimes(M, q)$ is isometric to a pfaffian arising from a 2-torsion datum. We now describe this datum. We view C_1 as a Z-module through the left action of Z. Then φ: $C_0\otimes_Z C_0\to\mathrm{End}_Z(C_1)$ given by $\varphi(a\otimes b)(x)=axt(b)$ is an isomorphism and we claim that (C_0, C_1, φ) is the required datum. Let Pf: $\mathrm{Alt}(C_1)\to Pf(C_1)$ be the pfaffian of the datum (C_0, C_1, φ). The multiplication in C induces an isomorphism $Z\otimes M\xrightarrow{\sim}\mathrm{Alt}(C_1)$ and it is shown in [KPS] that there exists an isomorphism λ: $Pf(C_1)\xrightarrow{\sim} R$ such that

$$Z\otimes(M, q)\xrightarrow{\sim}(\mathrm{Alt}(C_1), Pf_\lambda).$$

Thus there exists a (Galois) descent map δ: $(\mathrm{Alt}(C_1), Pf_\lambda) \tilde{\to} (\mathrm{Alt}(C_1), Pf_\lambda)$ such that the descended quadratic space $\{x \in C_1 \mid \delta(x) = x\}$ is (M, q) (see [KPS] for details). In view of the computation of GO_+ for the trivial Arf invariant spaces, we have a homomorphism

$$\psi\colon\ Z^* \times C_0^* \to GO_+(\mathrm{Alt}(C_1), Pf_\lambda).$$

For any Azumaya Z-algebra B with an involution τ of the second kind, let $GU(B) = \{b \in B^* \mid \tau(b)\, b \in Z^*\}$.

Lemma 6.5 *Let $\eta \in Z^*$, $b \in C_0^*$. Then $\psi_{(\eta, b)}$ belongs to $GO_+(M, q)$ if and only if $b \in GU(C_0)$ and $\bar{\eta}\eta^{-1} = n(b)(\tau(b)\, b)^{-2}$, n denoting the reduced norm on C_0, bar denoting the effect of the non-trivial automorphism σ_0 of Z over R.*

Proof. The map $\psi_{(\eta, b)}$ belongs to $GO_+(M, q)$ if and only if it commutes with the descent δ on $(\mathrm{Alt}(C_1), Pf_\lambda)$ (noting that the descent of a special similitude is automatically a special similitude). Clearly, it suffices to prove the lemma after a base change S/R with S faithfully flat over R. Thus, we assume, without loss of generality, that $C_0 = \mathrm{End}_Z(V)$, V free over Z, $C_1 = V \otimes_Z V$. Let $\{e_i\}$ be an ordered basis of V such that $e_1 \wedge e_2 \wedge e_3 \wedge e_4 = \lambda^{-1}(1) \in \wedge^4 V = Pf(\mathrm{Alt}(V \otimes_Z V))$. We identify $\mathrm{End}_Z(V)$ with $M_4(Z)$ and $\mathrm{Alt}(V \otimes_Z V)$ with $\mathrm{Alt}_4(Z)$ through the choice of this basis. We denote by $x \mapsto \bar{x}$ the extension entrywise of σ_0 to $M_4(Z)$. Let τ on $\mathrm{End}_Z(V)$ be given by $x \mapsto u^{-1}\bar{x}^t u$, $u \in \mathrm{GL}_4(Z)$. The descent map δ on $\mathrm{Alt}_4(Z)$ is, up to a scalar, the map $x \mapsto \bar{u}^{t^{-1}} \overline{p_1(x)} \bar{u}^{-1}$. The map $\psi_{(\eta, b)}$ is given by $\psi_{(\eta, b)}(x) = \eta b x b^t$, $x \in \mathrm{Alt}_4(Z)$. The condition $\psi_{(\eta, b)}\delta = \delta\psi_{(\eta, b)}$ is equivalent to the following two conditions:
1) The element $\gamma = u^{-1}\bar{b}^t u b$ is central; i.e. $\tau(b)\, b \in Z^*$ (that γ is central is a consequence of 6.3 since, for all $x \in \mathrm{Alt}_4(Z)$), $\gamma x \gamma^t = \rho x$, for some $\rho \in Z$.
2) $(\tau(b)\, b)^{-2} n(b) = \bar{\eta}\eta^{-1}$.
This proves the lemma. □

Let H be the subgroup of $Z^* \times GU(C_0)$ defined by

$$H = \{(\eta, b) \in Z^* \times GU(C_0) \mid \bar{\eta}\eta^{-1} = n_B(b)(\tau(b)\, b)^{-2}\}.$$

Theorem 6.6 *Let (M, q) be a quadratic space of rank 6 with Arf invariant Z. The association $(\eta, b) \mapsto \psi_{(\eta, b)}$ yields a homomorphism ψ: $H \to GO_+(M, q)$ such that the induced sequence*

$$1 \to H/\{(\eta^2, \eta^{-1}), \eta \in Z^*\} \xrightarrow{\bar{\psi}} GO_+(M, q) \to \mathrm{Pic}(Z)$$

is exact.

Proof. This theorem is immediate from 6.5 and 6.4. □

We now indicate a method of computation of the group of similitudes of rank 4 quadratic spaces. Let (M, q) be a quadratic space of rank 4 and Arf invariant Z over R. The map $Z \otimes M \to C_1(M, q)$, $(\zeta, x) \mapsto \zeta x$ is an isomorphism and (M, q) is in fact a descent of the reduced norm on C_1 as a C_0-module, the descent map being given by τ, the restriction of the standard involution of $C(M, q)$ to C_1. There is a map

$$R^* \times C_0^* \to GO_+(M, q)$$

given by $(\eta, b) \mapsto \psi_{(\eta, b)}$, where $\psi_{(\eta, b)}(x) = \eta b x \tau(b)$. The multiplier of the map $\psi_{(\eta, b)}$ is $\eta^2 n(b)\, \overline{n(b)}$, where n is the reduced norm of C_0 as a Z-algebra. This map is surjective if $\mathrm{Pic}(Z)$ is trivial and has kernel $\{(\mu\bar{\mu}, \mu^{-1}), \mu \in Z^*\}$. We thus have,

Theorem 6.7 *Let (M, q) be a rank 4 quadratic space over R with Arf invariant Z. Then the sequence*

$$1 \to (R^* \times C_0^*)/\{\mu\bar{\mu}, \mu^{-1}), \mu \in Z^*\} \to GO_+(M, q) \to \mathrm{Pic}(Z)$$

is exact.

Theorem 6.8 *Let (M, q) have trivial Arf invariant and e an idenpotent in the centre of $C_0(M, q)$ generating the centre over R and $A = C_0(M, q)e$. Then the sequence*

$$1 \to (A^* \times A^*)/\{(\eta, \eta^{-1}), \eta \in R^*\} \to GO_+(M, q) \to \mathrm{Pic}(R)$$

is exact.

We do not give the details of proofs of 6.7 and 6.8.

Remark. In Theorem 6.4 and Theorem 6.6, the image in $\mathrm{Pic}(R)$, resp. $\mathrm{Pic}(Z)$, of the last map of the exact sequence, consists, in fact, of elements of torsion 4 and, in 6.7 and 6.8, of elements of order 2. This follows from Theorem 3.1, p. 111, of [KO]. We also observe that the above sequences are parts of cohomology sequences induced by exact sequences of sheaves.

Acknowledgements. The first author thanks the Tata Institute for the hospitality accorded during the preparation of this paper.

References

[ART] Amitsur, S.A., Rowen, L.H., Tignol, J.P.: Division algebras of degree 4 and 8 with involution. Isr. J. Math. **33**, 133–148 (1979)

[A] Albert, A.A.: Normal division algebras of degree four over an algebraic field. Trans. Am. Math. Soc. **34**, 363–372 (1932)

[D] Dieudonné, J.: Les extensions quadratiques des corps non commutatifs et leurs applications. Acta Math. **87**, 83–146 (1952)

[J] Jacobson, N.: Some applications of Jordan norms to involutorial simple associative algebras. Adv. Math **48**, 1–15 (1983)

[K] Kneser, M.: Composition of binary quadratic forms. J. Number Theory **15**, 406–413 (1982)

[Kn] Knus, M.-A.: Pfaffians and quadratic forms. Adv. Math. **71**, 1–20 (1988)

[KO] Knus, M.-A., Ojanguren, M.: Théorie de la descente et algèbres d'Azumaya. (Lect. Notes Math., vol. 389) Springer: Berlin Heidelberg New York 1974

[KOS] Knus, M.-A., Ojanguren, M., Sridharan, R.: Quadratic forms and Azumaya algebras. J. Reine Angew. Math. **303/304**, 231–248 (1978)

[KPS] Knus, M.-A., Parimala, R., Sridharan, R.: A classification of rank 6 quadratic spaces via pfaffians. J. Reine Angew. Math. **398**, 187–218 (1989)

[KS] Knus, M.-A., Schofield, A.: Aspects of the pfaffian (to appear)

[P] Parimala, R.: Indecomposable quadratic spaces over the affine plane. Adv. Math. **62**, 1–6 (1986)

[S] Saltman, D.J.: Azumaya algebras with involutions, J. Algebra **52**, 526–539 (1978)

Math. Z. 206, 605–615 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Line bundle convexity of pseudoconvex domains in complex manifolds

Karen R. Pinney

Department of Mathematics, University of Kentucky, Lexington, KY 40506-0027, USA

Received December 28, 1989; in final form May 14, 1990

1 Introduction

The classical Levi problem in several complex variables states that a domain in $\mathbb{C}^n$ is pseudoconvex if and only if it is a domain of holomorphy, or, equivalently, if it is holomorphically convex. It is clear that for domains in a manifold, holomorphic convexity implies pseudoconvexity, and Grauert showed that strongly pseudoconvex domains are holomorphically convex [6]. However, pseudoconvex domains in a manifold are in general not holomorphically convex. Grauert has given as a counterexample a domain in an abelian variety which is pseudoconvex but which supports only constant holomorphic functions [14]. In fact, there are examples of domains $\Omega \Subset M$ ($\dim M > 2$) which are pseudoconvex and almost everywhere strongly pseudoconvex but which support only constant holomorphic functions [5]. One can prove holomorphic convexity of pseudoconvex domains $\Omega \subset M$ under various conditions on M: if M is Stein, then any pseudoconvex $\Omega \Subset M$ is Stein [2], and if M is a Kähler manifold with positive sectional curvature, then any pseudoconvex $\Omega \Subset M$ is holomorphically convex [3]. Narasimhan and Fornaess have studied analogous and related questions for complex spaces [4, 12, 13].

In this paper we restrict our attention to compact Kähler manifolds M on which there is a positive line bundle L. These are biholomorphic to smooth projective algebraic varieties, by Kodaira's Embedding Theorem. A pseudoconvex domain $\Omega \subset M$ is not necessarily holomorphically convex, as shown by the above examples. We define Ω to be L-convex if for every infinite discrete sequence $\{x_\nu\} \subset \Omega$ there is a holomorphic section σ of L over Ω such that $|\sigma(x_\nu)|$ is unbounded, where the norm of σ is measured in a hermitian metric on L. This is a natural generalization of holomorphic convexity.

In this paper we prove that if Ω has C^2 boundary, then it is convex with respect to a high enough tensor power of L.

Denote the dimension of M by n. In Sect. 2, we reduce the problem to the solution of the equation $\sum_{i=1}^{n} g_i \otimes t_i = s$, where each t_i is a section of L over M which vanishes at a point $P \in \partial\Omega$ and s is a section of $\otimes^{r+1} L$ over M which does not vanish at P. If one can find holomorphic sections g_i of $\otimes^r L$ over

Ω which solve the equation, then at least one of the sections g_i must be unbounded in a neighborhood of P.

In Sects. 3 and 5, we set up a Koszul complex and weighted L^2 Čech cohomology groups, then use these and the $\bar{\partial}$ methods of Hörmander to solve the equation locally in Sect. 6. In Sect. 7, we use $\bar{\partial}$ methods to obtain a global solution from the local solutions. In particular, we use an L^2-Dolbeault vanishing theorem of Demailly. As this theorem applies only to complete Kähler manifolds, we must construct a complete Kähler metric on Ω. Together with a theorem of Ohsawa, this construction shows that for a relatively compact domain with C^2 boundary in a Kähler manifold, the existence of a complete Kähler metric is equivalent to pseudoconvexity. This is done is Sect. 4.

Napier [11] has recently used line bundle convexity to study the holomorphic convexity of coverings of smooth projective surfaces.

I would like to thank my thesis advisor, R. Narasimhan, for suggesting this problem and for excellent teaching and guidance over the last four years.

2 Main theorem and preliminaries

A domain $\Omega \subset \mathbb{C}^n$ is pseudoconvex if and only if every point $x \in \partial\Omega$ has a neighborhood U_x such that $U_x \cap \Omega$ is pseudoconvex. One makes an analogous definition for domains in complex manifolds:

Definition. Let Ω be a domain in a complex manifold M. Then Ω is said to be pseudoconvex if every point $x \in \partial\Omega$ has a coordinate neighborhood U_x such that $U_x \cap \Omega$ is pseudoconvex.

In [5], Grauert introduced the notion of convexity with respect to sections of a line bundle. We make the following weaker definition.

Definition. Let X be a complex manifold, E a holomorphic vector bundle on X and h a hermitian metric on E. Then an open set $\Omega \subset X$ is called E-convex if for every infinite discrete sequence $\{x_\nu\}$ of points in Ω, there is a holomorphic section $\sigma \in \Gamma(\Omega, E)$ such that $|\sigma(x_\nu)|_h$ is unbounded.

Remark. If (E, h) is the trivial line bundle with its usual metric, then E-convexity is equivalent to holomorphic convexity.

Remark. If $\Omega \Subset X$, then E-convexity is independent of the metric h as long as h is a metric on E over X.

If L is a line bundle, denote the line bundle $\otimes^r L$ by L^r.

Theorem 2.1 *Suppose that M is a compact complex manifold and that L is a positive holomorphic line bundle on M. If $\Omega \subset M$ is a pseudoconvex domain with C^2 boundary, then there is a positive integer r_0 such that Ω is L^r-convex for $r \geqq r_0$.*

Conversely, suppose that Ω is K-convex for some line bundle $K \to M$. Because K is locally trivial over the intersection of Ω with small coordinate balls, Ω is locally pseudoconvex, thus pseudoconvex.

Let M, Ω, and L be as in Theorem 2.1. By the Kodaira Embedding Theorem and Chow's Theorem, we may suppose that M is an algebraic subvariety in

$\mathbb{C}P^N$ (some large N) and that L is the hyperplane bundle H restricted to M (replacing L by some positive tensor power of itself).

Let $\{x_\nu\}$ be an infinite discrete sequence in Ω. Because $\bar{\Omega}$ is compact, $\{x_\nu\}$ must have an accumulation point on the boundary of Ω, say at $P \in \partial\Omega$. Passing to a subsequence, we may assume that $\lim_{\nu \to \infty} x_\nu = P$. Choose holomorphic sections $t_1, \ldots, t_n \in \Gamma(\mathbb{C}P^N, H)$ such that $Z = \{x: t_i(x) = 0, i = 1, \ldots, n\}$ intersects M transversely and $P \in Z$. Note that this implies that if we consider the sections t_i as holomorphic functions via local trivializations of L, then $\{t_i\}$ is a set of local coordinates for M in a neighborhood of each point in $Z \cap M$. If d is the degree of M, then Z has exactly d points of intersection with M.

Let r be an integer, $r \geqq d-2$. We now construct a holomorphic section $s \in \Gamma(\mathbb{C}P^N, H^{r+1})$ which vanishes at every point of $Z \cap \Omega$ but which does not vanish at P.

For each $z_j \in Z \cap \Omega$, let H_j be a hyperplane through z_j which does not contain P. Let $s_j \in \Gamma(\mathbb{C}P^N, H)$ be a section which vanishes exactly on H_j, and let $s' \in \Gamma(\mathbb{C}P^N, H)$ be any section not vanishing at P. Let

$$s = \Big(\bigotimes_{z_j \in Z \cap \Omega} s_j\Big) \otimes \overbrace{s' \otimes \ldots \otimes s'}^{r+1-|Z \cap \Omega| \text{ times}}.$$

We see that $r \geqq d-2$ is large enough for this construction. In order to prove Theorem 2.1, we put additional constraints on r in Sect. 7.

If one can solve the equation

$$\sum_{i=1}^{n} g_i \otimes t_i = s \tag{2.1}$$

on Ω with $g_i \in \Gamma(\Omega, L')$, then one of the sections g_i must be unbounded on $\{x_\nu\}$ and the theorem is proved.

3 Koszul complex

To solve equation (2.1), we first find local solutions, then use $\bar{\partial}$ methods to patch the local solutions together. Both steps use the sheaf sequence constructed below.

For a fixed r, define the bundle E_q over M by

$$E_q = L^{r-q+1} \otimes 1^{\binom{n}{q}},$$

where $1^{\binom{n}{q}}$ is a trivial bundle of rank $\binom{n}{q}$. Each stalk of $1^{\binom{n}{q}}$ is isomorphic to $\wedge^q \mathbb{C}^n$. Let $\mathscr{L}^q$ denote the sheaf of germs of holomorphic sections of L^q and let $\mathscr{E}_q$ denote the sheaf of germs of holomorphic sections of E_q.

Define bundle maps $T_q: E_q \to E_{q-1}$ as follows. (We define T_q in exactly the same way for $\wedge^q \mathbb{C}^n$-valued functions.) Let u be a local section of E_q over a

set on which L is trivial, $\{e_I\}_{|I|=q}$ the usual basis of $\wedge^q \mathbb{C}^n$, and $\{e_J\}_{|J|=q-1}$ the basis of $\wedge^{q-1}\mathbb{C}^n$. Write $u = \sum_{|I|=q}{}' u_I e_I$. Then we define $T_q u = \sum_{|J|=q-1}{}' (T_q u)_J e_J$, where

$$(T_q u)_J = \sum_{i=1}^{n} u_{Ji} \otimes t_i.$$

T_q induces a map from $\mathscr{E}_q$ to $\mathscr{E}_{q-1}$. This map is also denoted T_q.
Consider the sequence

$$0 \longrightarrow \mathscr{E}_n \xrightarrow{T_n} \mathscr{E}_{n-1} \xrightarrow{T_{n-1}} \cdots \xrightarrow{T_2} \mathscr{E}_1 \xrightarrow{T_1} \mathscr{E}_0. \tag{3.1}$$

Proposition 3.1 *Sequence* (3.1) *is an exact sequence of sheaves.*

Proof. Let $x \in M$. Using a trivialization for L in a neighborhood of x and localizing at x, (3.1) becomes

$$0 \longrightarrow \mathscr{O}_x \xrightarrow{T_n} \wedge^{n-1} \mathscr{O}_x^n \xrightarrow{T_{n-1}} \cdots \xrightarrow{T_3} \wedge^2 \mathscr{O}_x^n \xrightarrow{T_2} \mathscr{O}_x^n \xrightarrow{T_1} \mathscr{O}_x. \tag{3.2}$$

The sequence (3.2) is the Koszul complex associated to the ring $\mathscr{O}_x$ and the sequence $((t_1)_x, \ldots, (t_n)_x)$. This is a regular sequence in $\mathscr{O}_x$. A theorem from commutative algebra then implies that the complex (3.2) is exact [10, p. 135]. □

Remark. We extend T_v to a map

$$T_v : E_v \otimes T^{*(p,q)} M \to E_{v-1} \otimes T^{*(p,q)} M$$

by setting $T_v(u \otimes \alpha) = (T_v u) \otimes \alpha$. Because the sections t_i are holomorphic, T_v commutes with the $\bar{\partial}$ operator.

Let $\mathscr{Z}_q$ denote the kernel of the sheaf map T_q. Notice that

$$\mathscr{Z}_{n-1} \overset{T_n}{\simeq} \mathscr{L}^{r-n+1} = \mathscr{E}_n.$$

For any cover $\mathscr{V} = \{V_A\}$ of Ω, we may define maps on Čech cochains

$$T_q : C^p(\mathscr{V}, \mathscr{E}_q) \to C^p(\mathscr{V}, \mathscr{Z}_{q+1})$$

by setting $T_q((u_A)) = (T_q(u_A))$. T_q commutes with the coboundary operators δ. Note that $s \in H^0(\Omega, \mathscr{E}_0)$ and that equation (2.1) may be written

$$T_1 g = s, \quad g \in H^0(\Omega, \mathscr{E}_1). \tag{3.3}$$

To prove Theorem 2.1, we will first solve equation (3.3) locally; i.e. we will find $v \in C^0(\Omega, \mathscr{E}_1)$ such that $T_1 v = s$. Then $[\delta v] \in H^1(\Omega, \mathscr{Z}_1)$. We have short exact sequences

$$0 \longrightarrow \mathscr{Z}_i \longrightarrow \mathscr{E}_i \xrightarrow{T_i} \mathscr{Z}_{i-1} \longrightarrow 0,$$

which give rise to long exact sequences

$$H^{i-1}(\Omega, \mathscr{E}_i) \to H^{i-1}(\Omega, \mathscr{Z}_{i-1}) \to H^i(\Omega, \mathscr{Z}_i) \to H^i(\Omega, \mathscr{E}_i).$$

Now if $H^{n-1}(\Omega, \mathscr{E}_n)$, $H^{i-1}(\Omega, \mathscr{E}_i)$ and $H^i(\Omega, \mathscr{E}_i)$ were to vanish for $i = 2, 3, \ldots, n-1$, then we would have

$$H^1(\Omega, \mathscr{Z}_1) \cong H^2(\Omega, \mathscr{Z}_2) \cong \ldots \cong H^{n-1}(\Omega, \mathscr{Z}_{n-1}) \cong H^{n-1}(\Omega, \mathscr{E}_n) = 0.$$

In this case we would have $\delta v = \delta f$, $f \in C^0(\Omega, \mathscr{Z}_1)$, and $g = v - f$ would solve equation (3.3).

For $q \geqq 1$, the Čech cohomology group $H^q(\Omega, \mathscr{E}_i)$ is isomorphic to the Dolbeault cohomology group

$$H_{\bar\partial}^{0,q}(\Omega, E_i) = \frac{\{\text{smooth } \bar\partial\text{-closed } E_i\text{-valued forms of type } (0, q) \text{ on } \Omega\}}{\bar\partial\{\text{smooth } E_i\text{-valued forms of type } (0, q-1) \text{ on } \Omega\}}.$$

The group $H_{\bar\partial}^{0,q}(\Omega, E_i)$ does not necessarily vanish; however, a vanishing theorem of Demailly implies that, under certain curvature conditions, $\bar\partial$-closed E_i-valued forms which are L^2 on Ω are $\bar\partial$-exact [1]. In order to apply this theorem, we need to construct a complete Kähler metric on Ω and to keep track of weighted L^2 bounds on cochains.

4 Construction of a complete Kähler metric

Theorem 4.1 *Let X be a Kähler manifold with Kähler metric ds_x^2, $\Omega \Subset X$ a relatively compact domain with C^2 boundary. Then Ω supports a complete Kähler metric with C^∞ coefficients if and only if it is pseudoconvex.*

Proof. In [15], Ohsawa proved that if U is a domain with C^1 boundary in a complex manifold, and if U has a complete Kähler metric, then U is pseudoconvex.

To prove the converse, suppose that Ω is pseudoconvex. We will construct a complete Kähler metric with smooth coefficients on Ω.

Let $\sigma \in C^2(\bar\Omega)$ be a defining function for Ω. The (1,1)-form $i\partial\bar\partial(-\log(-\sigma))$ is bounded below on Ω (see [3, 7, 16]). In our case we may see this directly by writing σ locally as $h \cdot r$, where h is positive and C^2 up to the boundary of Ω and r = distance to $\partial\Omega$ in a local Euclidean metric, and using the fact that $-\log$ (distance to the boundary) is a plurisubharmonic function for pseudoconvex domains in $\mathbb{C}^n$. Let ω_X be the (1,1)-form associated to ds_X^2. For $C > 0$ large enough, $i\partial\bar\partial(-\log(-\sigma)) + C\omega_X$ defines a Kähler metric on Ω with continuous coefficients. We may see that it is a complete metric by making a local C^∞ change of coordinates such that $\rho(z) = \operatorname{Re} z_1$ and comparing our metric

to the complete metric $\frac{1}{(\mathrm{Re}\, z_1)^2}(d(\mathrm{Re}\, z_1))^2 + C ds_X^2$. We use Whitney's Approximation Theorem [17] to approximate σ sufficiently closely in the uniform C^2 topology with respect to ds_X^2 by a function $\rho \in C^\infty(\Omega) \cap C(\bar{\Omega})$ with $\rho|_{\partial\Omega} = 0$. The (1,1)-form $i\partial\bar{\partial}(-\log(-\rho)) + C\omega_X$ then defines a complete Kähler metric with C^∞ coefficients. □

5 Čech cohomology with weighted L^2 bounds

In this section we define weighted L^2 norms on Čech cochains with values in a sheaf. Let X be any complex manifold with volume form V_X, E a holomorphic vector bundle over X, H a hermitian metric on E, and $\mathscr{E}$ the sheaf of germs of holomorphic sections of E. Let $\mathscr{V} = \{V_\alpha\}$ be a cover of Ω by open sets and ϕ a positive, continuous function on X.

For a multiindex $A = \alpha_0 \alpha_1 \ldots \alpha_p$, let $V_A = V_{\alpha_0} \cap V_{\alpha_1} \cap \ldots \cap V_{\alpha_p}$. Define $C_k^p(\mathscr{V}, \mathscr{E})$ to be the set of all Čech cochains $v \in C^p(\mathscr{V}, \mathscr{E})$ such that for all A,

$$\int_{V_A} |v_A|_H^2 \phi^k V_X < \infty.$$

Let $\delta: C^p(\mathscr{V}, \mathscr{E}_q) \to C^{p+1}(\mathscr{V}, \mathscr{E}_q)$ denote the usual Čech coboundary operator. Let $Z_k^p(\mathscr{V}, \mathscr{E}) = \{v \in C_k^p(\mathscr{V}, \mathscr{E}): \delta v = 0\}$. Since $\delta(C_k^p(\mathscr{V}, \mathscr{E})) \subset Z_k^{p+1}(\mathscr{V}, \mathscr{E})$, one may define the cohomology group

$$H_k^p(\mathscr{V}, \mathscr{E}) = \frac{Z_k^p(\mathscr{V}, \mathscr{E})}{\delta(C_k^{p-1}(\mathscr{V}, \mathscr{E}))}.$$

We now return to the notation of Sects. 2 and 3. Once for all we fix a C^∞ bounded exhaustion function $\rho: \Omega \to [-1, 0)$ such that $i\partial\bar{\partial}(-\log(-\rho)) + C\omega_M$ defines a C^∞ complete Kähler metric on Ω.

Fix a finite open cover $\mathscr{U}$ of Ω as follows. Around each $x_\alpha \in Z \cap \partial\Omega$, fix a coordinate neighborhood $W_\alpha \subset M$ such that

a) L is trivial on $\bar{W}_\alpha$
b) Considering each t_i as a holomorphic function via a trivialization of L, $\{t_i\}$ gives local coordinates on $\bar{W}_\alpha$
c) $C ds_M^2 = L(\psi_\alpha)$ on $\bar{W}_\alpha$, $\psi_\alpha \in C^\infty(\bar{W}_\alpha)$
d) $W_\alpha \cap \Omega$ is pseudoconvex and bounded (in $\mathbb{C}^n$).

Around each $x_\alpha \in Z \cap \Omega$, fix a coordinate neighborhood $W_\alpha \Subset \Omega$ such that conditions a)–c) above hold and

e) W_α is a coordinate polydisc in the coordinates $\{t_i\}$.

Also, let the sets W_α be small enough so that they are disjoint. This is possible since $Z \cap M$ is finite.

Now cover the rest of $\bar{\Omega}$ with a finite number of coordinate neighborhoods W_α such that a), c), and d) hold. The sets should be small enough so that $Z \cap \overline{W_\alpha} \cap \overline{W_\beta} = \emptyset$ for all $\alpha \neq \beta$.

Let $U_\alpha = W_\alpha \cap \Omega$ and $\mathcal{U} = \{U_\alpha\}$. This cover of Ω remains fixed throughout the rest of the paper. The coordinates on U_α and the trivialization of L on U_α are the restrictions of those on W_α and are also fixed throughout the paper.

We will need to shrink the cover $\mathcal{U}$ twice. For those sets W_α which contain a point $x_\alpha \in Z$, fix coordinate polydiscs $W_\alpha'' \Subset W_\alpha' \Subset W_\alpha$ also centered at x_α, such that if we set $W_\alpha' = W_\alpha$ and $W_\alpha'' = W_\alpha$ for W_α not containing a point of Z, then $\Omega \subset \cup_\alpha W_\alpha''$. Let $U_\alpha' = W_\alpha' \cap \Omega$, $U_\alpha'' = W_\alpha'' \cap \Omega$, $\mathcal{U}' = \{U_\alpha'\}$, and $\mathcal{U}'' = \{U_\alpha''\}$.

Fix $\phi_\alpha \in C_0^\infty(W_\alpha')$, $0 \leqq \phi_\alpha \leqq 1$, such that $\sum_\alpha \phi_\alpha \equiv 1$ on Ω.

Fix a hermitian metric h_0 on L (over M) with positive curvature. This induces metrics on L^{r-q+1} and on $E_q = L^{r-q+1} \otimes 1^{\binom{n}{q}}$. Let V_M be the volume form associated to the metric ds_M^2. We define the groups $C_k^p(\mathcal{V}, \mathcal{E}_q)$, $Z_k^p(\mathcal{V}, \mathcal{E}_q)$, and $H_k^p(\mathcal{V}, \mathcal{E}_q)$ using h_0, V_M, and the weight $-\rho$. Note that δ commutes with T_q.

Since sections of the sheaf $\mathcal{Z}_q$ are in particular sections of the bundle E_q, one can define in a similar manner a weighted L^2 norm for the cochains $C^p(\mathcal{V}, \mathcal{Z}_q)$, again using h_0, V_M, and $-\rho$. As above, one can define the groups $C_k^p(\mathcal{V}, \mathcal{Z}_q)$, $Z_k^p(\mathcal{V}, \mathcal{Z}_q)$, and $H_k^p(\mathcal{V}, \mathcal{Z}_q)$.

Because the sections t_i are bounded, $T_q(C_k^p(\mathcal{V}, \mathcal{E}_q)) \subset C_k^p(\mathcal{V}, \mathcal{Z}_{q+1})$, so there is an induced map $T_q: H_k^p(\mathcal{V}, \mathcal{E}_q) \to H_k^p(\mathcal{V}, \mathcal{Z}_{q+1})$.

Remark. By restriction, we have maps

$$\imath: C_k^p(\mathcal{U}, \mathcal{E}_q) \to C_k^p(\mathcal{U}', \mathcal{E}_q)$$
$$\imath: C_k^p(\mathcal{U}, \mathcal{Z}_q) \to C_k^p(\mathcal{U}', \mathcal{Z}_q).$$

Here $\mathcal{U}$ and $\mathcal{U}'$ may be replaced by $\mathcal{U}'$ and $\mathcal{U}''$, respectively.

6 Local solution

The following theorem of Hörmander gives a solution to the $\bar\partial$ problem for $\mathbb{C}$-valued forms on a bounded pseudoconvex domain D in $\mathbb{C}^n$. It will be of central importance in proving Theorem 2.1. For ω a (p, q)-form on D, $|\omega(x)| = \langle \omega(x), \omega(x) \rangle^{\frac{1}{2}}$ denotes the norm induced by the Euclidean metric on $\mathbb{C}^n$ and dV the Euclidean volume form.

Theorem 6.1 ([9], Theorem 2.2.3). *Let D be as above. Let $\delta =$ diameter of D and let ϕ be a plurisubharmonic function on D. For every (p, q)-form f on D, $q > 0$, such that $\bar\partial f = 0$ and*

$$\int_D |f|^2 e^{-\phi} dV < \infty,$$

there exists a $(p, q-1)$-form u on D such that $\bar\partial u = f$ and

$$q \int_D |u|^2 e^{-\phi} dV \leqq e\delta^2 \int_D |f|^2 e^{-\phi} dV.$$

We may lift a section u of $\mathscr{Z}_i$ to a C^∞ local section v' of $\mathscr{E}_{i+1}$ such that $T_{i+1}u=v'$. Theorem 6.1 may then be used to replace v' with a holomorphic local lifting v. Furthermore, given a control on the growth of u, we have a control on the growth of v. More precisely, we obtain the following two lemmas.

Lemma 6.2 *Let U be a pseudoconvex domain in $\mathbb{C}^n$, ϕ a plurisubharmonic function on U, $t_1, \ldots, t_n$ holomorphic functions on U such that $\left(\sum_{i=1}^{n} |t_i|^2\right)^{\frac{1}{2}} \geqq Ce^{-\phi}$ on U, $C>0$, and $\frac{\partial t_1}{\partial z_j}$ is bounded on U for all i, j. Then if u is a holomorphic section of $\wedge^q \mathcal{O}^n$ over U $(q \geqq 0)$ and $T_q u=0$ (if $q \geqq 1$) and*

$$\int_U |u|^2 e^{-k\phi} dV < \infty,$$

then there exists a holomorphic section v of $\wedge^{q+1}\mathcal{O}^n$ over U such that $T_{q+1}v=u$ and

$$\int_U |v|^2 e^{-(k+4(n-q))\phi} dV < \infty.$$

If $\inf_{z\in U} \sum |t_i(z)|>0$ on $\bar{U}$, then the conclusion holds with the weight $e^{-k\phi}$.

Proof. The lemma is proved by downward induction on q. One explicitly constructs a C^∞ solution v', then uses Theorem 6.1 to solve a $\bar{\partial}$-problem. Cf. Theorem 7 of [8]. □

Lemma 6.3 *Let U be a polydisc centered at 0 in $\mathbb{C}^n$, $(t_1, \ldots, t_n)$ the usual coordinates on U, and u a holomorphic section of $\wedge^q \mathcal{O}^n$ over U such that*

$$\begin{aligned} T_q u &= 0 \quad q \geqq 1 \\ u(0) &= 0 \quad q = 0. \end{aligned}$$

Then there exists a holomorphic section v of $\wedge^{q+1} \mathcal{O}^n$ over U such that $T_{q+1}v=u$.

Proof. Consider the ring $\mathscr{A}$ of holomorphic functions on U. The sequence $(t_1, \ldots, t_n)$ is a regular sequence in $\mathscr{A}$, so the Koszul complex associated to $\mathscr{A}$ and to $(t_1, \ldots, t_n)$ is exact. This proves the lemma for $q \geqq 1$. The result is immediate for $q=0$. □

Return to the notation of Sect. 5. Applying Lemmas 6.2 and 6.3 to $U=U_A$ or $U=U'_A$ with $\phi=-\log(-\rho)+\psi_\alpha$ and with $\{t_i\}$ as constructed in Sect. 2 yields

Corollary 6.4 a) *If $u \in C^0_k(\mathscr{U}, \mathscr{E}_0)$ and $u_\alpha(x)=0$ for all $x \in U_\alpha \cap Z$, then there exists $v \in C^0_{k+4n}(\mathscr{U}', \mathscr{E}_1)$ such that $T_1 v = \imath u$.*
b) *If $u \in C^0_k(\mathscr{U}, \mathscr{Z}_q)$, $q \geqq 1$, then there exists $v \in C^0_{k+4(n-q)}(\mathscr{U}, \mathscr{E}_{q+1})$ such that $T_{q+1}v = \imath u$.*
c) *If $u \in C^p_k(\mathscr{U}', \mathscr{Z}_q)$, $q \geqq 1$, $p \geqq 1$, then there exists $v \in C^p_k(\mathscr{U}', \mathscr{E}_{q+1})$ such that $T_{q+1}v = u$.*

$\mathscr{U}'$ may be replaced by $\mathscr{U}''$ and $\mathscr{U}$ by $\mathscr{U}'$.

The section s constructed in Sect. 2 is an element of $C_0^0(\mathscr{U}, \mathscr{E}_0)$, so Corollary 6.4 implies that there is a local solution to (3.3), i.e., a cochain $v \in C_{4n}^0(\mathscr{U}', \mathscr{E}_1)$ such that $T_1 v = s$.

7 Global solution

We define connecting homomorphisms on the groups $H_k^q(\mathscr{U}, \mathscr{Z}_q)$ in the usual way, using Corollary 6.4 to verify that these are well-defined:

$$D_1 : H_k^1(\mathscr{U}', \mathscr{Z}_1) \to H_{k+4(n-1)}^2(\mathscr{U}'', \mathscr{Z}_2)$$
$$D_q : H_k^q(\mathscr{U}'', \mathscr{Z}_q) \to H_k^{q+1}(\mathscr{U}'', \mathscr{Z}_{q+1}) \quad \text{for } q = 2, \ldots, n.$$

The main tool used in patching together the local solutions obtained in Sect. 6 is the following theorem, which is due to Demailly [1].

Theorem 7.2 *Let X be a complete Kähler manifold, equipped with Kähler metric g which is not necessarily complete. Let ω be the associated (1,1)-form. Let E be a positive, C^2 hermitian line bundle on X and let $R(h)$ denote the curvature form of a hermitian metric h on E. Let $|\ |_{h,g}$ be the pointwise norm induced by h and g on differential forms with values in E. If for some $c > 0$,*

$$\operatorname{Ric} g + R(h) \geqq c \cdot \omega, \tag{7.1}$$

then for every $(0, q)$-form f on X with values in $E (q \geqq 1)$ such that $\bar{\partial} f = 0$ and

$$\int_X |f|_{h,g}^2 dV_g < \infty,$$

there exists a $(0, q-1)$-form u on X with values in E such that $\bar{\partial} u = f$ and

$$\int_X |u|_{h,g}^2 dV_g < \infty.$$

We will apply this theorem with $X = \Omega$, $g = ds_M^2|_\Omega$, and $E = L^{r-q+1}$. Although $ds_M^2|_\Omega$ is not complete, Ω has a complete Kähler metric, namely $i\partial\bar{\partial}(-\log(-\rho)) + C\omega_M$. Let h_0 be any metric on L whose curvature form is positive. We set $h = (-\rho)^l h_0^{r-q+1}$, l a positive number. Because $i\partial\bar{\partial}(-\log(-\rho))$ is bounded below, (7.1) is satisfied for r large enough. This is not quite enough to ensure the vanishing of $H_l^i(\mathscr{U}, \mathscr{Z}_i)$. However, the following lemma tells us that if we have the appropriate inequality on curvature, if $v \in H_k^q(\mathscr{U}, \mathscr{Z}_q)$ and $D_q v$ is a coboundary with L^2 bounds, then v is the coboundary of some cochain w, w having L^2 bounds which may be weaker than those on v.

Lemma 7.3 *For any $k \geqq 0$ and any integer $1 \leqq q \leqq n-2$, let*

$$l = \begin{cases} k+4n-4 & \text{if } q = 1, \\ k & \text{if } q \geqq 2. \end{cases}$$

Let r be an integer large enough so that, on Ω,

$$(r-q) R(h_0) > -\operatorname{Ric}(ds_M^2) - l \cdot i\partial\bar{\partial}(-\log(-\rho)).$$

Then if $v \in Z_k^q(\mathcal{U}'', \mathcal{Z}_q)$ *(resp.* $v \in Z_k^1(\mathcal{U}', \mathcal{Z}_1)$ *if* $q=1$*) and if there exists a* q*-cochain* $\tau \in C_l^q(\mathcal{U}'', \mathcal{Z}_{q+1})$ *such that* $D_q[v]=[\delta\tau]$*, then there exists a* $(q-1)$*-cochain* $w \in C_l^{q-1}(\mathcal{U}'', \mathcal{Z}_q)$ *such that* $v=\delta w$.

Proof. Using Corollary 6.4 and the hypothesis on $D_q[v]$, we obtain $\theta \in Z_l^q(\mathcal{U}', \mathscr{E}_{q+1})$ such that $T_{q+1}\theta = v$. We patch the local sections θ_B together using the partition of unity $\{\phi_\alpha\}$ to obtain a $C^\infty(q-1)$-cochain η_1 with $\delta\eta_1=\theta$. We patch together the sections $(\bar\partial\eta_1)_A$ in the same way to obtain a $C^\infty(q-2)$-cochain η_2 with values in $E_{q+1} \otimes T^{*(0,1)}M$. We iterate the procedure to obtain $\bar\partial\eta_q$, a holomorphic 0-cocycle with values in $E_{q+1} \otimes T^{*(0,q)}M$. In other words, $\bar\partial\eta_q$ is a $\bar\partial$-closed, global $(0, q)$-form on Ω with values in E_{q+1}. We apply Theorem 7.2, using the metric $h=(-\rho)^l h_0^{r-q+1}$, to find a global $(0, q-1)$-form μ such that $\bar\partial\mu=\bar\partial\eta_q$. Set $\lambda_{q-1}=\eta_q-\mu$ and use Theorem 6.1 with the weight $-\log((-\rho)^l e^{-l\psi_\alpha})$ to obtain χ_{q-2} such that $\bar\partial\chi_{q-2}=\lambda_{q-1}$. Set $\lambda_{q-2}=\eta_{q-1}-\delta\chi_{q-2}$ and iterate the procedure, keeping track of L^2 bounds throughout, to obtain $\lambda_0 \in C_l^{q-1}(\mathcal{U}'', \mathscr{E}_{q+1})$ such that $\bar\partial\lambda_0=0$ and $\delta\lambda_0=\theta$. Then $w=T_{q+1}\lambda_0$ satisfies the conclusion of the lemma. □

Lemma 7.4 *For any* $k \geqq 0$*, let* r *be an integer large enough so that, on* Ω*,*

$$(r-n+1)R(h_0) > -\mathrm{Ric}(ds_M^2) - k \cdot i\partial\bar\partial(-\log(-\rho)).$$

Then $H_k^{n-1}(\mathcal{U}'', \mathcal{Z}_{n-1})=0$.

Proof. This may be proved in the same way as Lemma 7.3, with the difference that for $v \in Z_k^{n-1}(\mathcal{U}'', \mathcal{Z}_{n-1})$, $D_{n-1}[v]$ must equal zero in $H_k^n(\mathcal{U}'', \mathcal{Z}_{n-1})$, so the cocycle θ has an L^2 bound with the weight $(-\rho)^k$. Arguing as for Lemma 7.3, v can then be shown to be the coboundary of an L^2 cochain with the same weight. □

We are now able to prove Theorem 2.1.

Proof of theorem 2.1. The section s constructed in Sect. 2 is an element of $C_0^0(\mathcal{U}, \mathscr{E}_0)$. By Corollary 6.4, there exists $v \in C_{4n}^0(\mathcal{U}', \mathscr{E}_1)$ such that $T_1 v = \imath s$. Then $\delta v \in Z_{4n}^1(\mathcal{U}', \mathscr{E}_1)$. Define $y_1 = D_1[\delta v]$ and $y_i = D_i(y_{i-1})$ for $i=2, \ldots, n-2$. We have $y_i \in H_{8n-4}^{i+1}(\mathcal{U}'', \mathcal{Z}_{i+1})$. By Lemma 7.4, if $r \in \mathbb{Z}$ is large enough so that

$$(r-n+1)R(h_0) > -\mathrm{Ric}\, ds_M^2 - (8n-4)i\partial\bar\partial(-\log(-\rho)), \tag{7.2}$$

then $y_{n-2}=[\delta z_{n-2}]$ for a cochain $z_{n-2} \in C_{8n-4}^{n-2}(\mathcal{U}'', \mathcal{Z}_{n-1})$. From (7.2) it follows that

$$(r-i)R(h_0) > -\mathrm{Ric}\, ds_M^2 - (8n-4)i\partial\bar\partial(-\log(-\rho))$$

for $i=n-2, n-3, \ldots, 1$. Using Lemma 7.3 to iterate, we find that then $y_i=[\delta z_i]$ for a cochain $z_i \in C_{8n-4}^i(\mathcal{U}'', \mathcal{Z}_{i+1})$, and $\delta v = \delta z_0$, $z_0 \in C_{8n-4}^0(\mathcal{U}'', \mathcal{Z}_1)$. Let $g=v-w$. Then $g \in Z_{8n-4}^0(\mathcal{U}'', \mathscr{E}_1)$, $\delta g=0$, and $T_1 g = s$. Thus g solves equation (3.3), which proves Theorem 2.1. □

Remark. The solution g has a weighted L^2 bound:

$$\int_\Omega |g|_h^2 (-\rho)^{8n-4} V_M < \infty.$$

Remark. This proof requires the following lower bounds on the integer r in Theorem 2.1. These bounds depend on the dimension of M, on the degree d of M as an algebraic variety, on the Ricci curvature of M, and on the negativity of $i\partial\bar{\partial}(-\log(-\rho))$:

(i) $r \geqq d-2$

(ii) $(r-i)R(h_0) > -\mathrm{Ric}\, ds_M^2 - (8n-4)i\partial\bar{\partial}(-\log(-\rho))$.

References

1. Demailly, J.-P.: Estimations L^2 pour l'operateur $\bar{\partial}$ d'un fibré vectoriel holomorphe semi-positif au-dessus d'une variété Kahlérienne complète. Ann. Scient. Éc. Norm. Sup. **15**, 4^e série 457–511 (1982)
2. Docquier, F., Grauert, H.: Levisches Problem und Rungescher Satz für Teilgebiete Steinscher Mannigfaltigkeiten. Math. Ann. **140**, 94–123 (1960)
3. Elencwajg, G.: Pseudo-convexité locale dans les variétés kahlériennes. Ann. Inst. Fourier (Grenoble) **25**, **2**, 295–314 (1975)
4. Fornaess, J., Narasimhan, R.: The Levi problem on complex spaces with singularities. Math. Ann. **248**, 47–72 (1980)
5. Grauert, H.: Bemerkenswerte pseudokonvexe Mannifaltigkeiten. Math. Z. **81**, 377–391 (1963)
6. Grauert, H.: On Levi's problem and the imbedding of real-analytic manifolds. Ann. of Math. **68**, **2**, 460–472 (1958)
7. Greene, R.E., Wu, H.: On Kähler manifolds of positive bisectional curvature and a theorem of Hartogs. Abh. Math. Sem. Univ. Hamburg **47**, 171–185 (1978)
8. Hörmander, L.: Generators for some rings of analytic functions. Bull. Am. Math. Soc. **73**, 943–949 (1967)
9. Hörmander, L.: L^2 estimates and existence theorems for the $\bar{\partial}$ operator. Acta Math. **113**, 89–152 (1965)
10. Matsumura, H.: Commutative Algebra, New York: Benjamin 1970
11. Napier, T.: Convexity properties of coverings of smooth projective varieties. Math. Ann. **286**, 433–479 (1990)
12. Narasimhan, R.: The Levi problem for complex spaces. Math. Ann. **142**, 355–365 (1961)
13. Narasimhan, R.: The Levi problem for complex spaces II. Math. Ann. **146**, 195–216 (1962)
14. Narasimhan, R.: The Levi problem in the theory of functions of several complex variables. ICM, Stockholm 385–388 (1962)
15. Ohsawa, T.: On complete Kähler domains with C^1-boundary. Publ. RIMS **16**, 929–940 (1980)
16. Takeuchi, A.: Domaines pseudoconvexes sur les variétés Kählériennes. J. Math. Kyoto Univ. **6–3**, 323–357 (1967)
17. Whitney, H.: Analytic extensions of differentiable functions defined in closed sets. Trans. Am. Math. Soc. **36**, 63–89 (1934)

Math. Z. 206, 617–622 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

On galois automorphisms of the fundamental group of the projective line minus three points

Hiroaki Nakamura
Department of Mathematics, Faculty of Science, University of Tokyo,
Hongo, Tokyo, 113, Japan

Received March 14, 1990; in final form July 9, 1990

Let k be a number field of finite degree over $\mathbb{Q}$, and $\pi_1(U)$, $\pi_1(U_{\bar{k}})$ the etale profinite fundamental groups of $U := \mathbb{P}^1_k - \{0, 1, \infty\}$ and $U_{\bar{k}} := U \otimes \bar{k}$ respectively. Fix an odd prime ℓ. The basic object to be studied in this paper is

$$\pi_1'(U) := \pi_1(U)/J_\ell,$$

where J_ℓ denotes the kernel of the canonical surjection of $\pi_1(U_{\bar{k}})$ to its maximal pro-ℓ quotient $\pi_1^{\text{pro-}\ell}$. We shall regard it canonically as a group extension over the absolute Galois group $G_k = \mathrm{Gal}(\bar{k}/k)$:

$$1 \longrightarrow \pi_1^{\text{pro-}\ell} \longrightarrow \pi_1'(U) \xrightarrow{p_{U/k}} G_k \longrightarrow 1 \quad \text{(exact)}.$$

Let $\mathrm{Aut}_{G_k}\pi_1'(U)$ denote the group of all the continuous group automorphisms α of $\pi_1'(U)$ satisfying the Galois compatibility condition $p_{U/k} \circ \alpha = p_{U/k}$. Since each inner automorphism of $\pi_1'(U)$ coming from an element of $\pi_1^{\text{pro-}\ell}$ satisfies also this compatibility, the inner automorphism group $\mathrm{Inn}\,\pi_1^{\text{pro-}\ell}$ is considered to be a closed subgroup of $\mathrm{Aut}_{G_k}\pi_1'(U)$. Our main purpose is then to prove the following

Theorem 1

$$\frac{\mathrm{Aut}_{G_k}\pi_1'(U)}{\mathrm{Inn}\,\pi_1^{\text{pro-}\ell}} \simeq \mathrm{Aut}_k(U).$$

Note that $\mathrm{Aut}_k(U)$ is nothing but the symmetric group of the 3rd degree permuting the 3-point set $\{0, 1, \infty\}$. Generally speaking, each k-automorphism of U determines only one coset of $\mathrm{Aut}_{G_k}\pi_1'(U)$ by $\mathrm{Inn}\,\pi_1^{\text{pro-}\ell}$. Hence the point is to show that every Galois compatible group automorphism of $\pi_1'(U)$ must actually come from one of the six geometric k-automorphisms of U.

It should also be noted that our formulation follows a special case of Grothendieck's "anabelian" Tate conjecture [4], except that it is in the pro-ℓ version. For the proof of Theorem 1, three concrete tools are needed: Ihara's

anti-1-cocycle of pro-braids, a family of Belyi liftings, and a certain nonabelian weight filtration. After explaining each of them in suitable styles respectively, we shall complete the proof at the last section.

1 Ihara's anti-1-cocycle

In this section, we review part of the Ihara theory in the special case of 3-point ramification and give an application of it which shall be used later.

In the following, we denote $\pi_1^{\text{pro-}\ell}$ simply by π_1. By Grothendieck's comparison theorem, we can suitably choose generators x, y, $z(\in\pi_1)$ of inertia groups over 0, 1, ∞ respectively to give a presentation of π_1 as follows:

$$\pi_1=\pi_1^{\text{pro-}\ell}=\langle x, y, z \mid xyz=1\rangle_{\text{pro-}\ell\text{ completion}}.$$

This is isomorphic to the free pro-ℓ group of rank 2. Let Λ denote the completed group algebra of π_1 with coefficient ring $\mathbb{Z}_\ell$: $\Lambda=\mathbb{Z}_\ell[[\pi_1]]$. We naturally identify π_1 with a subgroup of $\Lambda^\times$, and put $u=x-1$, $v=y-1$, $w=z-1$. Then Λ can be regarded as the noncommutative power series ring of two variables over $\mathbb{Z}_\ell$:

$$\Lambda=\mathbb{Z}_\ell[[\pi_1]]=\mathbb{Z}_\ell[[v, w]]_{\text{n.c.}}.$$

We define a subgroup $\operatorname{Brd}\pi_1$ of $\operatorname{Aut}\pi_1$ by

$$\operatorname{Brd}\pi_1=\left\{f\in\operatorname{Aut}\pi_1 \;\middle|\; \begin{array}{l}\exists s\in\pi_1, \exists t\in\pi_1(2), \exists a\in\mathbb{Z}_\ell^\times \text{ such that}\\ f(x)=sx^as^{-1}, f(y)=ty^at^{-1}, f(z)=z^a\end{array}\right\}.$$

Here $\pi_1(2)$ is the closure of the commutator subgroup of π_1. In the above definition, it is easy to see that $a\in\mathbb{Z}_\ell^\times$ and $t\in\pi_1(2)$ are uniquely determined by $f\in\operatorname{Brd}\pi_1$. Therefore we shall write them as a_f, t_f respectively.

Let $\operatorname{Brd}\pi_1$ act on $\Lambda=\mathbb{Z}_\ell[[\pi_1]]$ canonically as a subgroup of $\operatorname{Aut}\pi_1$. In [5], Ihara introduced an anti-1-cocycle

$$\Psi\colon \operatorname{Brd}\pi_1\to\Lambda^\times$$
$$(\Psi_{fg}=f(\Psi_g)\cdot\Psi_f \quad \text{for} \quad f, g\in\operatorname{Brd}\pi_1)$$

for investigating Galois representations in towers of coverings of $\mathbb{P}^1$ in a more general situation by using mainly vaiables x and y. In this paper, however, we instead adopt y and z for our later use; namely we define Ψ_f for $f\in\operatorname{Brd}\pi_1$ by

$$\Psi_f:=1+\left(\frac{\partial t_f}{\partial z}\right)\cdot w.$$

Recall that for each $\lambda\in\Lambda$, the free differentials $\dfrac{\partial\lambda}{\partial y}$, $\dfrac{\partial\lambda}{\partial z}(\in\Lambda)$ are defined by the following formula:

$$\lambda=s(\lambda)+\frac{\partial\lambda}{\partial y}\cdot v+\frac{\partial\lambda}{\partial z}\cdot w,$$

where $s\colon\Lambda\to\mathbb{Z}_\ell$ denotes the augmentation map.

Let $\pi_1=\pi_1(1)\supset\pi_1(2)\supset\pi_1(3)\supset\ldots$ be the lower central series of π_1 and denote the kernel of the canonical surjection

$$\Lambda\to\mathbb{Z}_\ell[[\pi_1/\pi_1(N)]]$$

by I_N $(N\geqq 1)$. Here, in particular, $I=I_1$ is the augmentation ideal of Λ. We note that I^N is the two sided ideal of the power series in Λ having no term of degree less than N, and that $I^N\supsetneqq I_N$ for $N\geqq 2$.

Lemma 2 *Let G be a subgroup of* Brd π_1 *and assume that there exist relatively prime integers m, n $(\geqq 2)$ and elements g, h_1, $h_2(\in G)$ such that*
(i) $a_g\in\mathbb{Z}_\ell^\times$ is nontorsion; (ii) $a_{h_1}=a_{h_2}=1$, and

$$\Psi_{h_1}-1\in(I_2+I^m)-(I_2+I^{m+1}),\ \Psi_{h_2}-1\in(I_2+I^n)-(I_2+I^{n+1}).$$

Then G has trivial centralizer in Brd π_1.

Proof. Let $f\in$Brd π_1 centralize G. First we shall show $a_f=1$. Since $fh_i=h_if$, we have $f(\Psi_{h_i})\cdot\Psi_f=h_i(\Psi_f)\cdot\Psi_{h_i}$ $(i=1, 2)$. Then the commutativity of the ring Λ/I_2 and the triviality of the action of h_i on $\pi_1/\pi_1(2)$ for $i=1, 2$ give

$$(*)\qquad f(\Psi_{h_i})=\Psi_{h_i}\quad \bmod I_2\ (i=1,2).$$

Since f acts on I^k/I^{k+1} by a_f^k-multiplication, and since I_2+I^k/I_2+I^{k+1} is torsion free for each $k\geqq 1$, $(*)$ and the assumption (ii) imply $a_f^m=1$, $a_f^n=1$. From this we conclude $a_f=1$.

Next, we shall show that $f=1$, for which it suffices to see $t_f=1$. If not, there exists an integer $N(\geqq 2)$ such that $t_f\in\pi_1(N)-\pi_1(N+1)$. By assumption on f, we have now $f(\Psi_g)\cdot\Psi_f=g(\Psi_f)\cdot\Psi_g$. If we note that $\Psi_f-1(\in I^N)$ mod I^{N+1} lies in the center of Λ/I^{N+1}, we obtain

$$g(\Psi_f)^{-1}\cdot\Psi_f\equiv f(\Psi_g)^{-1}\cdot\Psi_g\quad(\bmod I^{N+1}).$$

Since $g(\Psi_f-1)\equiv a_g^N(\Psi_f-1)$ mod I^{N+1}, the left hand side turns out to be congruent to $1-(a_g^N-1)(\Psi_f-1)=1-(a_g^N-1)\cdot\dfrac{\partial t_f}{\partial z}\cdot w$ modulo I^{N+1}. On the other hand, since f acts on Λ/I_N trivially, $f(\Psi_g)^{-1}\cdot\Psi_g$ lies in $1+I_N$. Therefore, if we expand it into a power series

$$1+\psi_N(v,w)+\psi_{N+1}(v,w)+\ldots$$

with each ψ_k being a homogeneous polynomial of v, w of degree k $(k\geqq N)$, then $\psi_N(v, w)$ must be a Lie element. (Recall that I_N is the two sided ideal of Λ generated by the $\gamma-1$ $(\gamma\in\pi_1(N))$.) Thus we obtain

$$-(a_g^N-1)\cdot\frac{\partial t_f}{\partial z}\cdot w\equiv\psi_N(v,w)\quad\bmod I^{N+1}.$$

As is easily proved by the Dynkin-Wever-Specht theorem ([7]), every nontrivial Lie element of degree $\geqq 2$ cannot be (right-) divisible by either v or w. Therefore

we get $\frac{\partial t_f}{\partial z}\cdot w \equiv 0 \pmod{I^{N+1}}$. But then, by definition, we must have $t_f - 1 \equiv \frac{\partial t_f}{\partial y}\cdot v \pmod{I^{N+1}}$. This gives us a desired contradiction because $t_f - 1 \pmod{I^{N+1}}$ is now a nontrivial Lie element. □

2 Family of Belyi liftings

In the following, we denote (the closure of) the cyclic subgroups of π_1 generated by x, y, z respectively by $I_0 = \langle x\rangle$, $I_1 = \langle y\rangle$, $I_\infty = \langle z\rangle$. For each cyclic subgroup J of π_1 which is conjugate to I_1, define a subgroup $B(I_\infty, J)$ of $\pi_1'(U)$ by

$$B(I_\infty, J) := \left\{ \beta \in \pi_1'(U) \;\middle|\; \begin{array}{l} \exists s \in \pi_1, \exists t \in \pi_1(2) \quad \text{such that} \\ \beta I_\infty \beta^{-1} = I_\infty, \beta J \beta^{-1} = tJt^{-1}, \beta I_0 \beta^{-1} = sI_0 s^{-1} \end{array} \right\}.$$

Proposition 3 *(i) $B(I_\infty, J)$ is a subgroup of the normalizer D_∞ of I_∞ in $\pi_1'(U)$, which is isomorphic to G_k via the* map $p_{U/k}$. *We have semidirect product decomposition:*

$$D_\infty = I_\infty \rtimes B(I_\infty, J),$$
$$\pi_1'(U) = \pi_1 \rtimes B(I_\infty, J).$$

(ii) For each $\gamma \in \pi_1$, $B(I_\infty, I_1) = B(I_\infty, \gamma I_1 \gamma^{-1})$ if and only if $\gamma \in \pi_1(2)\langle y\rangle$.

Remark. We call $B(I_\infty, J)$ the Belyi lifting determined by the vector $I_\infty \to J$. In fact, the case $J = I_1$ was introduced by Belyi in [1].

Proof. Since I_∞ is known to be selfnormalizing in π_1, $B(I_\infty, J) \cap \pi_1 = \{1\}$. Therefore, for proving (i), it suffices to show that for every element $\sigma \in G_k$, there exists $\beta \in B(I_\infty, J)$ such that $p_{U/k}(\beta) = \sigma$. Take any β' from $p_{U/k}^{-1}(\sigma)$. Then the conjugation by β' sends I_∞, J, I_0 to their π_1-conjugates respectively. If we note that $\pi_1/\pi_1(2) \simeq \langle z\rangle \oplus \langle y\rangle$, multiplying β' by a certain element of π_1 we can easily obtain the desired β. For proving (ii), let $\beta \in B(I_\infty, I_1)$. Then, by simple calculation, we see that $\beta \in B(I_\infty, \gamma I_1 \gamma^{-1})$ implies $\beta\gamma\beta^{-1}\gamma^{-1} \in \pi_1(2)\langle y\rangle$. Let χ denote the cyclotomic character of $B(I_\infty, \gamma I_1 \gamma^{-1}) \simeq G_k$. Then $\beta\gamma\beta^{-1} \equiv \gamma^{\chi(\beta)} \bmod \pi_1(2)$. Therefore the condition $B(I_\infty, I_1) = B(I_\infty, \gamma I_1 \gamma^{-1})$ must yield $\gamma^{\chi(\beta)-1} \in \langle y\rangle \bmod \pi_1(2)$, where $\chi(\beta)$ ranges over an open subgroup of $\mathbb{Z}_\ell^\times$. From this we obtain $\gamma \in \pi_1(2)\langle y\rangle$. The converse is easy. □

3 Weight filtration in π_1

Let $q_0 : \bar{\eta} = \mathrm{Spec}(\Omega) \to \mathbb{P}^1$ be a fixed generic geometric point and consider a projective system $\mathrm{Et}_\ell(U)$ consisting of the pairs (Y, q) of finite connected nonsingular coverings $h : Y \to \mathbb{P}^1_k$ unramified outside $\{0, 1, \infty\}$ with Galois closures having ℓ-power geometric degrees and the pointing morphisms $q : \bar{\eta} \to Y$ such that $q_0 = h \circ q$. Note that if Ω^{ur} denotes the subfield of Ω generated by the function fields $k(Y)$ for all $(Y, q) \in \mathrm{Et}_\ell(U)$, then the fundamental group $\pi_1'(U)$ is identified with the Galois group $\mathrm{Gal}(\Omega^{ur}/k(U))$. Let $\mathscr{S}$ be the subset of $\varprojlim_{(Y,q)} \mathrm{Hom}(\bar{\eta}, Y)$ lying above $\{0, 1, \infty\}$. Then each element s of $\mathscr{S}$ (called a "cusp") gives canoni-

cally a group embedding $\phi_s: \mathbb{Z}_\ell(1) \to \pi_1'(U)$ (see for example [3] Appendix I). The image I_s of ϕ_s is the inertia group at s, which is conjugate to one of I_0, I_1, I_∞ (and hence it is selfnormalizing in π_1). Observe here in particular that $I = I_s$ has the following two properties:

($\mathscr{A}$1) the normalizer D of I in $\pi_1'(U)$ satisfies $p_{U/k}(D) = G_k$;
($\mathscr{A}$2) the Galois character $\chi_I: G_k \simeq D/I \to \mathrm{Aut}(I) \simeq \mathbb{Z}_\ell^\times$ defined by conjugation is the same as the cyclotomic character.

Conversely, if we denote the conjugacy union of I_0 (resp. I_1, I_∞) in π_1 by $\mathscr{I}_0$ (resp. $\mathscr{I}_1$, $\mathscr{I}_\infty$), and if we put $\mathscr{I} = \mathscr{I}_0 \cup \mathscr{I}_1 \cup \mathscr{I}_\infty$, then we have a characterization of the subset $\mathscr{I}$ of π_1:

Lemma 4 *A selfnormalizing cyclic subgroup I of π_1 is contained in $\mathscr{I}$ if and only if I satisfies ($\mathscr{A}$1) and ($\mathscr{A}$2).*

Proof. We have only to show the "if" part. Suppose that a selfnormalizing subgroup I of π_1 satisfying ($\mathscr{A}$1) is not conjugate to any of I_0, I_1, I_∞. Let w be a topological generator of I, and take an open subgroup B of π_1 such that $B \cap \langle w \rangle = \langle w^\ell \rangle$. Then, since every element of $\langle w \rangle - \langle w^\ell \rangle$ generates $\langle w \rangle$, $\mathscr{I} - B$ and $\langle w \rangle - B$ are disjoint compact subsets in the Hausdroff space π_1. Therefore we can take an open normal subgroup $M (\subset B)$ of π_1 such that $(\mathscr{I} - B) \cdot M \cap (\langle w \rangle - B) \cdot M = \phi$. Define an open subgroup H of π_1 by $H = M \cdot \langle w \rangle$. Then the above construction gives

$$(*) \qquad w \notin [H, H]\, H^\ell \langle \mathscr{I} \cap H \rangle.$$

Take an open subgroup $\tilde{H}$ of $\pi_1'(U)$ such that $\tilde{H} \cap \pi_1 = H$, and let $\tilde{k}$ be such that $\mathrm{Gal}(\bar{k}/\tilde{k}) = p_{U/k}(\tilde{H})$. Then the Galois representation

$$\varphi_H: \mathrm{Gal}(\bar{k}/\tilde{k}) \simeq \tilde{H}/H \to \mathrm{Aut}(H^{ab})$$

which is canonically defined by conjugation is equivalent to the $\mathbb{Z}_\ell$-dual of the ℓ-adic Galois representation in $H^1_{et}(V_{\bar{k}}, \mathbb{Z}_\ell)$ where V is the normalization of $\mathbb{P}^1 - \{0, 1, \infty\}$ in the subfield of Ω^{ur} fixed by $\tilde{H}$ (defined over $\tilde{k}$). This Galois representation has a (in our case, torsion-free) weight filtration in the sense of Deligne [2], and (*) implies that $w (\mathrm{mod}\, [H, H])$ does not lie in the weight (-2) part of H^{ab}. This means that $I = \langle w \rangle$ cannot satisfy the condition ($\mathscr{A}$2). □

Remark. A similar lemma was proved in [8] where the pro-ℓ π_1 is replaced by the full profinite π_1 and $\mathbb{Z}_\ell(1)$ is replaced by $\hat{\mathbb{Z}}(1)$. Note that in the above pro-ℓ version, the fact that every cyclic subgroup of $\pi_1^{\text{pro-}\ell}$ is isomorphic to $\mathbb{Z}_\ell$ enables us to simplify the situation.

Corollary 5 *Every Galois compatible automorphism f of $\pi_1'(U)$ must map the set $\mathscr{I}$ to itself. In particular f induces a permutation of $\{\mathscr{I}_0, \mathscr{I}_1, \mathscr{I}_\infty\}$.* □

4 The proof of Theorem 1

Let f be an arbitrary element of $\mathrm{Aut}_{G_k} \pi_1'(U)$. Then, by Corollary 5, f permutes the subsets $\mathscr{I}_0$, $\mathscr{I}_1$, $\mathscr{I}_\infty$ of π_1. Making a composition with some suitable automorphism coming from $\mathrm{Aut}_k(U) \simeq S_3$, we may assume $f(\mathscr{I}_0) = \mathscr{I}_0$, $f(\mathscr{I}_1) = \mathscr{I}_1$, $f(\mathscr{I}_\infty) = \mathscr{I}_\infty$. Further replacing f by a composition with some element

of Inn π_1, we may assume $f(\mathscr{I}_0)=\mathscr{I}_0, f(I_1)=tI_1t^{-1}(\exists t\in\pi_1(2)), f(I_\infty)=I_\infty$. Under this situation, it suffices to show that f is the identity map of $\pi_1'(U)$. At first, note that by definition f maps $B(I_\infty, I_1)$ to $B(I_\infty, tI_1t^{-1})$. Then Proposition 3 (ii) implies that f preserves $B(I_\infty, I_1)$. Further the Galois compatibility of f assures that $B(I_\infty, I_1)$ is elementwise fixed by f. Therefore if we note that $\pi_1'(U)=\pi_1\rtimes B(I_\infty, I_1)$, all the task we have to do is reduced to showing that the restriction $f|\pi_1$ of f to π_1 is the identity map. By assumption, $f|\pi_1\in\mathrm{Brd}\,\pi_1$. Further if we denote by φ the Galois representation

$$G_k\simeq B(I_\infty, I_1)\to \mathrm{Brd}\,\pi_1$$

induced from the above semidirect product decomposition, $f|\pi_1$ must belong to the centralizer of the image $im(\varphi)$ of φ in $\mathrm{Brd}\,\pi_1$. Therefore it suffices to show that this Galois image $G=im(\varphi)$ has such elements g, h_1, h_2 that satisfy (i), (ii) of Lemma 2. But the condition (i) obviously holds for many $g=\varphi(\sigma)\in G$, because $a_{\varphi(\sigma)}=\chi(\sigma)$ for $\sigma\in G_k$ (χ: the cyclotomic character). For the condition (ii), we recall the explicit formula of Anderson, Coleman, and Ihara-Kaneko-Yukinari for the abelianization of Ihara's anti-1-cocycle $\Psi_{\varphi(\sigma)}$. It states, after fixing an identification of $\mathbb{Z}_\ell\simeq\mathbb{Z}_\ell(1)$, that $F_\sigma:=\Psi_{\varphi(\sigma)}\ (\mathrm{mod}\ I_2)$ is expressed in the commutatitve power series ring $\mathbb{Q}_\ell[[v, w]]$ by

$$F_\sigma(v, w)=\exp\sum_{\substack{m\geqq 3\\ \text{odd}}}\frac{\chi_m(\sigma)}{(\ell^{m-1}-1)}\sum_{\substack{i+j=m\\ i,j\geqq 1}}\frac{V^iW^j}{i!j!}$$

for $\sigma\in\mathrm{Gal}(\bar{k}/k(\mu_{\ell^\infty}))$. Here $V=\log(1+v)$, $W=\log(1+w)$, and the χ_m are certain Kummer characters $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}(\mu_{\ell^\infty}))\to\mathbb{Z}_\ell(m)$ considered by Soulé, Deligne. (See the articles by Deligne, Ihara in [6] for references on these materials. For the dependence of Ψ on the choice of variables u, v, w, see [5] § 4.) In particular when $\ell\neq 2$, Soulé showed that, for each odd $m\geqq 3$, χ_m has an open image in $\mathbb{Z}_\ell(m)$. This assures the existence of such elements h_1, h_2 (for example, for $m=3$, $n=5$) as in (ii) of Lemma 2. □

References

1. Belyi, G.V.: On galois extensions of a maximal cyclotomic field. Izv. Acad. Nauk SSSR **43**, 267–276 (1979), (Math. USSR-Izv. **14**, 247–256 (1980))
2. Deligne, P.: Théorie de Hodge I. Actes ICM, t. I, Paris: Gauthier-Villars 1970
3. Freitag, E., Kiehl, R.: Etale cohomology and the Weil conjecture. Berlin Heidelberg New York: Springer 1988
4. Grothendieck, A.: A letter to G. Faltings, 1983
5. Ihara, Y.: On Galois representations arising from towers of coverings of $\mathbb{P}^1\setminus\{0, 1, \infty\}$. Invent. Math. **86**, 427–459 (1986)
6. Ihara, Y., Ribet, K., Serre, J.-P. (eds.): Galois Groups over $\mathbb{Q}$. Math. Sci. Res. Inst. Pub., vol. 16. Berlin Heidelberg New York: Springer 1989
7. Magnus, W., Karrass, A., Solitar, D.: Combinatorial group theory. New York: Interscience 1966
8. Nakamura, H.: Galois rigidity of the etale fundamental groups of punctured projective lines. J. Reine angew. Math. **411**, 205–216 (1990)

Math. Z. 206, 623–650 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Non-commutative moment problems

Konrad Schmüdgen

Sektion Mathematik, Karl-Marx-Universität Leipzig. O – 7010 Leipzig, Federal Republic of Germany

Received May 28, 1990; in final form July 25, 1990

Introduction

The classical moment problem has a long and rich history. Much of its theory has been developed in contact with other mathematical disciplines such as complex function theory, orthogonal polynomials, continued fractions, operator theory, harmonic analysis and others. An excellent treatment of the classical theory is the monograph by Akhiezer [1], more recent developments and applications are summarized in the volume [16].

In this paper we are concerned with two problems, called the *generalized moment problem* (denoted by (GMP)) and the *quantum moment problem* (denoted by (QMP)). Both problems can be considered as non-commutative versions of the classical moment problem (abbreviated by (CMP)). To formulate these problems, we let G be a connected Lie group and $\mathscr{E}(\mathfrak{g})$ the universal enveloping algebra of the Lie algebra of G. Suppose f is a given linear functional on $\mathscr{E}(\mathfrak{g})$ and V is a fixed unitary representation of G.

(GMP) *When do there exist a unitary representation U of G and a cyclic C^∞-vector φ for U such that $f(x) = \langle dU(x)\varphi, \varphi\rangle$, $x \in \mathscr{E}(\mathfrak{g})$?*

(QMP) *When is f of the form $f(x) = \operatorname{tr} dV(x)\varrho$, $x \in \mathscr{E}(\mathfrak{g})$, where ϱ is an operator such that $dV(x)\varrho$ is of trace class for every $x \in \mathscr{E}(\mathfrak{g})$?*

When G is $\mathbb{R}^d$, it follows that the (GMP) is equivalent to the (CMP) and there is a one-to-one correspondence between solutions of the (GMP) and solutions of the (CMP). In this sense the (GMP) is a generalization of the (CMP) when $\mathbb{R}^d$ is replaced by a connected Lie group G. Clearly, the (QMP) is not a direct generalization of the (CMP), but it can be viewed as a non-commutative variant of it. Roughly speaking, the (QMP) is obtained when the integral is replaced by the trace and the measure by a trace class operator.

We briefly describe the content of the paper. In section 2 we give four formulations of the generalized moment problem and we prove their equivalence. Also, we define the notion of determinacy for the (GMP). In section 4 we study the quantum moment problem. The main results of this paper are

obtained in sections 3 and 5. Section 3 contains various results concerning the solvability of the (GMP). There are two general theorems which can be considered as generalizations of classical results due to Haviland, and there are two theorems for concrete Lie groups, the universal covering group of $SL(2,\mathbb{R})$ and the three dimensional Heisenberg group. Section 5 deals with functionals $f(.) = \langle dU(.)\varphi, \varphi\rangle$, where the unitary representation U in (GMP) is irreducible. For two concrete Lie groups ($SL(2,\mathbb{R})$, Heisenberg group) we begin a detailed study of such states f. Note that all problems investigated in section 5 are trivial in the classical case $G = \mathbb{R}^d$.

1 Definitions and notations

In the paper we shall use some terminology and some basic facts about unbounded representation. We refer to [24] for a thorough treatment of this subject.

Suppose A is an abstract $*$-algebra with unit, denoted by 1. Let $\mathscr{D}$ be a dense linear subspace of a Hilbert space $\mathscr{H}$. A $*$- representation of A on $\mathscr{D}$ is a homomorphism π of A into the algebra of linear operators which are defined on $\mathscr{D}$ and leave $\mathscr{D}$ invariant such that $\pi(1) = I$ and $\langle \pi(a)\varphi, \psi\rangle = \langle \varphi, \pi(a^+)\psi\rangle$, $a \in A$ and $\varphi, \psi \in \mathscr{D}$. We write $\mathscr{D}(\pi)$ for $\mathscr{D}$ and $\mathscr{H}(\pi)$ for $\mathscr{H}$. The weak commutant of $\pi(A)$ is denoted by $\pi(A)'_w$. It is the set of bounded linear operators B on $\mathscr{H}(\pi)$ satisfying $\langle B\pi(a)\varphi, \psi\rangle = \langle B\varphi, \pi(a^+)\psi\rangle$ for $a \in A$ and $\varphi, \psi \in \mathscr{D}(\pi)$. Let f be a positive linear functional on A, i.e. $f(a^+a) \geqq 0$ for all $a \in A$. The *GNS* representation of f is denoted by π_f, i.e. π_f is a closed cyclic $*$- representation of A with cyclic vector φ_f such that $f(a) = \langle \pi_f(a)\varphi_f, \varphi_f\rangle$, $a \in A$. We say that f is pure if each positive linear functional h on A such that $h(a^+a) \leqq f(a^+a)$, $a \in A$, is a multiple of f. For a subset K of A, K^* is the set of linear functionals on A which are non-negative on K.

Throughout this paper, G is a *connected* Lie group, $\mathfrak{g}$ is the Lie algebra of G and $\mathscr{E}(\mathfrak{g})$ is the complexified universal enveloping algebra of $\mathfrak{g}$. $\mathscr{E}(\mathfrak{g})$ is a $*$-algebra with involution defined by $x^+ := -x$ for $x \in \mathfrak{g}$. The unit element of G is denoted by e. By a unitary representation of G we mean a strongly continuous homomorphism U of G into the group of unitaries of some Hilbert space $\mathscr{H}(U)$. Let $\mathscr{D}^\infty(U)$ denote the space of C^∞-vectors for U and dU the associated $*$-representation of the enveloping algebra on $\mathscr{D}^\infty(U)$. As usual, $\widehat{G}$ is the unitary dual of G, i.e. the set of unitary equivalence classes of irreducible unitary representations of G. For notational simplicity we shall use the same symbol for an irreducible unitary representation of G and its unitary equivalence class. The unitary equivalence is denoted by "$\cong$". We refer to [28], section 4.4, [26], chapter 0, or to [24], chapter 10, for the basic facts on Lie group representation theory used throughout this paper.

2 The generalized moment problem

2.1 We begin with the classical multidimensional moment problem. Fix $d \in \mathbb{N}$. Let $M(\mathbb{R}^d)$ denote the set of all positive regular Borel measures μ on $\mathbb{R}^d$ such that for each $n \in \mathbb{N}_0^d$ the function t^n is in $L^1(\mathbb{R}^d, \mu)$. (We freely use multi-index notations such as $t^n := t_1^{n_1} \dots t_d^{n_d}$ for $n = (n_1, \dots, n_d) \in \mathbb{N}_0^d$ and $t = (t_1, \dots, t_d)$, where $t_j^0 := 1$). The real number $\alpha_n := \int t^n d\mu(t), n \in \mathbb{N}_0^d$, is called the n-th moment of a measure $\mu \in M(\mathbb{R}^d)$. The classical moment problem asks: When is a given (complex) d-sequence $(\alpha_n;\ n \in \mathbb{N}_0^d)$ the sequence of moments for some measure $\mu \in M(\mathbb{R}^d)$, i.e. $\alpha_n = \int t^n d\mu(t)$ for all $n \in \mathbb{N}_0^d$?

Recall that a one-to-one correspondence between (complex) d-sequences $(\alpha_n;\ n \in \mathbb{N}_0^d)$ and (complex-valued) linear functionals f on the polynomial algebra $\mathbb{C}[x_1, \dots, x_d]$ is given by $f(x^n) = \alpha_n$, $n \in \mathbb{N}_0^d$. Thus we have the following formulation of the classical moment problem which is more convenient for our purposes.

(CMP) *Given a linear functional f on the polynomial algebra $\mathbb{C}[x_1, \dots, x_d]$. When does there exist a measure $\mu \in M(\mathbb{R}^d)$ such that*

$$f(p) = \int_{R^d} p(t)\, d\mu(t) \quad \textit{for all} \quad p \in \mathbb{C}[x_1, \dots, x_d]? \tag{1}$$

In order to generalize (CMP) to arbitrary Lie groups we still need a slight reformulation of (1). For this we identify the enveloping algebra $\mathcal{E}(\mathbb{R}^d)$ of the abelian Lie algebra $\mathbb{R}^d$ with the polynomial algebra in the usual way. For every $t = (t_1, \dots, t_d) \in \mathbb{R}^d$, there is an irreducible unitary representation U_t of the Lie group $\mathbb{R}^d$ on $\mathcal{H}(U_t) := \mathbb{C}$ such that $dU_t(x_j) = t_j$, $j = 1, \dots, d$. The representations U_t exhaust the dual of $\mathbb{R}^d$. We have $dU_t(p) = p(t)$ for $p \in \mathcal{E}(\mathbb{R}^d) \equiv \mathbb{C}[x_1, \dots, x_d]$, so that (1) can be written as

$$f(p) = \int_{R^d} \langle dU_t(p)1, 1\rangle\, d\mu(t) \quad \text{for all} \quad p \in \mathcal{E}(\mathbb{R}^d). \tag{2}$$

2.2 Now suppose that G is an arbitrary connected Lie group. Our aim is to generalize (*CMP*) when $\mathbb{R}^d$ is replaced by G. We shall give four equivalent formulations $(GMP)_j$, $j = 1, 2, 3, 4$, for the generalized moment problem. The first one is directly suggested by formula (2) above.

Suppose f is a linear functional on the enveloping algebra $\mathcal{E}(\mathcal{g})$.

$(\text{GMP})_1$ When is f of the form

$$f(x) = \int_{\Lambda} \langle dU_\lambda(x)\varphi_\lambda, \varphi_\lambda\rangle\, d\mu(\lambda), \quad x \in \mathcal{E}(\mathcal{g}), \tag{3}$$

where (Λ, μ) is a measure space, $\lambda \to U_\lambda$ is a measurable field of irreducible unitary representations of G and $\lambda \to \varphi_\lambda$ is a measurable field of C^∞-vectors $\varphi_\lambda \in \mathcal{D}^\infty(U_\lambda)$?

$(\mathrm{GMP})_2$ When can f be expressed as

$$f(x) = \langle dU(x)\varphi, \varphi\rangle, \quad x \in \mathscr{E}(\mathscr{g}), \tag{4}$$

where U is a cyclic unitary representation of G with cyclic vector $\varphi \in \mathscr{D}^\infty(U)$?

$(\mathrm{GMP})_3$ When does there exist a positive definite C^∞-function F on G such that

$$f(x) = (\tilde{x}F)(e) \quad \text{for all} \quad x \in \mathscr{E}(\mathscr{g})? \tag{5}$$

For the fourth variant we assume that f is a positive linear functional on the $*$-algebra $\mathscr{E}(\mathscr{g})$, i.e. $f(x^+x) \geqq 0$ for $x \in \mathscr{E}(\mathscr{g})$. This can be done without loss of generality, since the positivity of f is necessary (but not sufficient in general!) for $(\mathrm{GMP})_j$, $j = 1, 2, 3$. (For $(\mathrm{GMP})_1$ and $(\mathrm{GMP})_2$ this is obvious and $(\mathrm{GMP})_3$ is equivalent to $(\mathrm{GMP})_2$ as shown in the next subsection.) Recall that π_f is the GNS representation obtained from f and φ_f is the corresponding cyclic vector.

$(\mathrm{GMP})_4$ When does there exist a cyclic unitary representation U of G with cyclic vector φ_f such that $\pi_f \subseteq dU$?

Remarks

1) Generalized moment problems of this kind were considered first and independently by the author [22] and P. Richter [20] and also by A.Ju. Dalezkij and Ju.S. Samoilenko (see [21, §9]).

2) It is clear that we obtain equivalent formulations of $(GMP)_2$ resp. $(GMP)_4$ if the requirements that U is cyclic and φ resp. φ_f is a cyclic vector for U are omitted. These requirements are only included for the definition of determinacy, see Subsection 2.4.

3) In the case $G = \mathbb{R}^d$ $(GMP)_1$ does not reduce directly to (CMP) with (1) replaced by (2). (Note that we have an arbitrary measure space (Λ, μ) and arbitrary C^∞-vectors φ_λ in $(GMP)_1$.) But nevertheless each problem $(\mathrm{GMP})_j$, $j = 1, 2, 3, 4$, is equivalent to (CMP) for $G = \mathbb{R}^d$. One way to see this is as follows.

From Bochner's theorem it follows that a one-to-one correspondence between measures $\mu \in M(\mathbb{R}^d)$ and positive definite C^∞-functions F on $\mathbb{R}^d$ is given by $F(t) = \int_{R^d} \exp i\langle t, s\rangle \, d\mu(s)$, $t \in \mathbb{R}^d$, cf. the proof of Lemma 2.3.2 below. Differentiating the latter, we obtain the equivalence of (CMP) and $(\mathrm{GMP})_3$ in this case. In 2.3 it will be shown that $(\mathrm{GMP})_3$ is equivalent to $(\mathrm{GMP})_j$, $j = 1, 2, 4$.

4) Each of the four (equivalent) formulations $(\mathrm{GMP})_j$ shows another perspective of the generalized moment problem. Solving $(\mathrm{GMP})_1$ means to write f as an integral over vector states with respect to irreducible unitary representations of G. This is the view point which is directly adopted from the classical moment problem. $(\mathrm{GMP})_2$ is the question when f is a vector state on the infinitesimal representation of some (cyclic) unitary representation of G. $(\mathrm{GMP})_3$ describes f in terms of positive definite functions on the group. Finally, $(\mathrm{GMP})_4$ asks when the GNS representation π_f has a G-integrable extension dU in a possibly larger Hilbert space.

2.3 In this subsection we shall prove that all four formulations of the generalized moment problem are equivalent. For this we require two lemmas. The first one is a well-known technical result from decomposition theory (see e.g. [24, p. 344]).

2.3.1 Lemma. *Suppose a unitary representation U is a direct integral $\int^{\oplus} U_\lambda \, d\mu(\lambda)$ of unitary representations U_λ. If a vector $\varphi \in \mathscr{H}(U)$ belongs to $\mathscr{D}^\infty(U)$, then we have $\varphi = (\varphi_\lambda)$ with $\varphi_\lambda \in \mathscr{D}^\infty(U_\lambda)$ μ-a.e. and*

$$\langle dU(x)\varphi, \varphi\rangle = \int \langle dU_\lambda(x)\varphi_\lambda, \varphi_\lambda\rangle \, d\mu(\lambda) \quad \text{for} \quad x \in \mathscr{E}(\mathfrak{g}). \tag{6}$$

2.3.2 Lemma. *Let U be a unitary representation of G and let $\varphi \in \mathscr{H}(U)$. If the function $F(g) := \langle U(g)\varphi, \varphi\rangle$ is in $C^\infty(G)$, then the vector φ belongs to $\mathscr{D}^\infty(U)$.*

Proof. By a result of R. Goodman (cf. [24], Theorem 10.1.9), we have $\mathscr{D}^\infty(U) = \bigcap_{k=1}^{d} \bigcap_{n=1}^{\infty} \mathscr{D}(\partial U(x_k)^n)$ for every basis $\{x_1, \ldots, x_d\}$ of $\mathfrak{g}$. Here $\partial U(x_k)$ denotes the infinitesimal generator of the one-parameter unitary group $t \to U(\exp tx_k)$. Fix $k \in \{1, \ldots, d\}$. Let $E(\lambda), \lambda \in \mathbb{R}$, be the spectral projections of the self-adjoint operator $-i\partial U(x_k)$ and let ν be the measure on $\mathbb{R}$ defined by $\nu(.) := \langle E(.)\varphi, \varphi\rangle$. Our proof is complete once we have shown that the function λ^n is in $L^2(\mathbb{R}, \nu)$ for each $n \in \mathbb{N}$, since then $\varphi \in \mathscr{D}((-i\partial U(x_k))^n)$ and hence $\varphi \in \mathscr{D}^\infty(U)$.

The assumption implies that the function $H(t) := F(\exp tx_k)$ is in $C^\infty(\mathbb{R})$. We prove by induction on $n \in \mathbb{N}_0$ that $\lambda^{2n} \in L^1(\mathbb{R}, \nu)$ and

$$H^{(2n)}(t) = \int e^{it\lambda}(i\lambda)^{2n} \, d\nu(\lambda), \quad t \in \mathbb{R}. \tag{7}$$

Since $H(t) = \langle U(\exp tx_k)\varphi, \varphi\rangle = \langle \exp t\partial U(x_k)\varphi, \varphi\rangle = \int e^{it\lambda} d\nu(\lambda)$, (7) is true for $n = 0$. Suppose the assertion is true for n. Since

$$h_k(t) := -ik(e^{ik^{-1}\lambda} - 1)\lambda^n \to \lambda^{n+1} \quad \text{as} \quad k \to \infty,$$

(7) and Fatou's lemma yield

$$\begin{aligned}
\int |\lambda^{n+1}|^2 \, d\nu(\lambda) &\leqq \lim_k \int |h_k(\lambda)|^2 \, d\nu(\lambda) \\
&= \lim_k (-1)^n k^2 \left(2H^{(2n)}(0) - H^{(2n)}(k^{-1}) - H^{(2n)}(-k^{-1}) \right) \\
&= (-1)^{n+1} H^{(2n+2)}(0) \\
&< \infty,
\end{aligned}$$

i.e. $\lambda^{2(n+1)} \in L^1(\mathbb{R}, \nu)$. Having this and differentiating (7) twice under the integral sign (by the dominated convergence theorem), (7) follows for $n + 1$. □

We begin with the equivalence of $(GMP)_1$ and $(GMP)_2$. Suppose f is of the form described in $(GMP)_1$. Since G is connected and U_λ is irreducible, the Hilbert space $\mathscr{H}(U_\lambda)$ is separable. Put $V := \int^{\oplus} U_\lambda \, d\mu(\lambda)$ and $\varphi := (\varphi_\lambda)$. Then we have $\varphi \in \mathscr{D}^\infty(V)$ (because of $\varphi_\lambda \in \mathscr{D}^\infty(U_\lambda)$) and $f(.) = \langle dV(.)\varphi, \varphi\rangle$ by (3) and (6)). Let U be the restriction of V to the closed linear span of $V(G)\varphi$. Then, obviously, $f(.) = \langle dU(.)\varphi, \varphi\rangle$ and f is as required in $(GMP)_2$.

Now let f be as stated in $(GMP)_2$. The Hilbert space $\mathscr{H}(U)$ is separable, since U is cyclic and G is connected. Therefore, U_λ can be written as a direct integral $\int^{\oplus} U_\lambda d\mu(\lambda)$ of irreducible unitary representations U_λ of G [15, 8.4]. Let

$\varphi = (\varphi_\lambda)$. Since $\varphi \in \mathscr{D}^\infty(U)$, $\varphi_\lambda \in \mathscr{D}^\infty(U_\lambda)$ μ-a.e. by Lemma 2.3.1. After changing φ_λ on a null set, all φ_λ are in $\mathscr{D}^\infty(U_\lambda)$. (3) follows at once from (4) and (6).

Next we verify the equivalence of $(\mathrm{GMP})_2$ and $(\mathrm{GMP})_3$. Suppose f is as in $(\mathrm{GMP})_2$. That $\varphi \in \mathscr{D}^\infty(U)$ means that the map $g \to U(g)\varphi$ of G into $\mathscr{H}(U)$ is C^∞. Therefore, the positive definite function $F(g) := \langle U(g)\varphi, \varphi\rangle$ on G is C^∞ and (5) follows immediately from (4). Conversely, assume that f is an in $(GMP)_3$. It is wellknown that there exists a unitary representation U of G with cyclic vector φ such that $F(g) = \langle U(g)\varphi, \varphi\rangle$, $g \in G$. Lemma 2.3.2 shows that $\varphi \in \mathscr{D}^\infty(U)$, hence (4) follows from (5).

The equivalence of $(GMP)_2$ and $(GMP)_4$ follows easily from the uniqueness of the *GNS* construction (see, for instance, [24, Theor. 8.6.4]).

2.4 Let us say that the generalized moment problem is *solvable* if f is of the form stated in $(\mathrm{GMP})_j$ for some (and then for all) $j \in \{1, 2, 3, 4\}$. By a *solution* of $(\mathrm{GMP})_2$ we mean the couple $\{U, \varphi\}$ occurring therein. Similarly, the function F in $(\mathrm{GMP})_3$ and the representation U in $(\mathrm{GMP})_4$ are called solutions of $(\mathrm{GMP})_3$ and $(\mathrm{GMP})_4$, respectively.

Another fundamental concept is that of determinacy. Let f be a linear functional on the enveloping algebra $\mathscr{E}(\mathscr{g})$ for which the (GMP) is solvable. We shall say that the generalized moment problem is determinate or that f is (GMP)-*determinate* if the couple $\{U, \varphi\}$ is uniquely determined by f up to unitary equivalence, i.e. if U_1 is another unitary representation of G with cyclic vector $\varphi_1 \in \mathscr{D}^\infty(U_1)$ satisfying $f(.) = \langle dU_1(.)\varphi_1, \varphi_1\rangle$, there exists an isometry T of $\mathscr{H}(U)$ onto $\mathscr{H}(U_1)$ such that $T\varphi = \varphi_1$ and $U = T^{-1}U_1T$. Thus f is (GMP)-determinate if for all solutions $\{U, \varphi\}$ of $(\mathrm{GMP})_2$ the function $\langle U(.)\varphi, \varphi\rangle$ on G is the same.

From this we conclude easily that the (GMP) is determinate if and only if the function F in $(GMP)_3$ is uniquely determined by f. Also, it is easily seen that f is (GMP)-determinate if and only if for any two representations U_1 and U_2 as in $(GMP)_4$ there is an isometry T of $\mathscr{H}(U_1)$ onto $\mathscr{H}(U_2)$ such that $U_1 = T^{-1}U_2T$ and $T\varphi_f = \varphi_f$.

The (GMP) is determinate if there exists a solution U of $(\mathrm{GMP})_4$ such that $\mathscr{H}(U) = \mathscr{H}(\pi_f)$ and $\pi_f^* \subseteq dU$. Indeed, then we have $dU^* \subseteq \pi_f^{**}$. Combined with the relations $dU = dU^*$ and $\pi_f^{**} \subseteq \pi_f^* \subseteq dU$, we get $\pi_f^* = dU$, and f is (GMP)-determinate by the last criterion of the preceding paragraph.

Remarks

1) Suppose $G = \mathbb{R}^d$. Then, as noted in Remark 3.) of 2.2, the (GMP) is equivalent to the (CMP), and there is a one-to-one correspondence between solutions F of $(\mathrm{GMP})_3$ and solutions μ of the (CMP). Therefore, the generalized moment problem for a linear functional f on $\mathscr{E}(\mathscr{g}) \equiv \mathbb{C}[x_1, \ldots, x_d]$ is determinate (in the sense defined above) if and only if the corresponding classical moment problem is determinate (i.e. there is unique measure $\mu \in M(\mathbb{R}^d)$ satisfying (1)).

2) If G is a type I group, G has unique decomposition properties, and hence the determinacy of the (GMP) could be also defined by the uniqueness of the representation given in $(\mathrm{GMP})_1$.

3 Solvability of the generalized moment problem

3.1 For a unitary representation U of G, we define $\mathscr{K}(U) := \{x \in \mathscr{E}(\mathscr{g}) : dU(x) \geqq 0\}$. As usual, $dU(x) \geqq 0$ means that $\langle dU(x)\varphi, \varphi\rangle \geqq 0$ for all $\varphi \in \mathscr{D}^\infty(U)$.

3.1.1 Theorem. *Let f be a linear functional on $\mathscr{E}(\mathscr{g})$. The generalized moment problem for f is solvable if and only if f is $\mathscr{K}(W)$-positive for all $W \in \widehat{G}$, i.e. $f(x) \geqq 0$ whenever $dW(x) \geqq 0$ for all irreducible unitary representations W of G.*

This theorem can be found in [22]. It is essentially based on the work of R.T. Powers [19]. Using the formulation $(GMP)_4$ and Remark 2) in 2.2, it is a direct consequence of Theorem 11.4.2 in [24]. Following the main lines of the treatment in [24, Sect. 11.4], we sketch the proof of the sufficiency part of Theorem 3.1.1. In the special situation needed here this proof is simpler than the proof in the general case covered by Theorem 11.4.2 in [24]. More importantly, this proof gives two additional results stated as Theorem 3.1.2 and Corollary 3.1.3.

3.1.2 Theorem. *Let V be a fixed unitary representation of G. If a linear functional f on $\mathscr{E}(\mathscr{g})$ is $\mathscr{K}(V)$-positive, then (GMP) is solvable and the unitary representation U in $(GMP)_2$ (or equivalently, in $(GMP)_4$) can be chosen such that $\mathscr{K}(V) \subseteq \mathscr{K}(U)$.*

Sketch of Proof. Let $\mathscr{A} := dV(\mathscr{E}(\mathscr{g}))$ and let $\mathscr{B}$ be the $*$-algebra generated by $\mathscr{A}$ and by the operators $\widetilde{V}(g) := V(g) \restriction \mathscr{D}^\infty(V)$, $g \in G$. Then $\mathscr{A}$ and $\mathscr{B}$ are $*$-algebra of operators on the invariant dense domain $\mathscr{D}^\infty(V)$ in $\mathscr{H}(V)$. Since f is $\mathscr{K}(V)$-positive by assumption, we have $\ker dV \subseteq \ker f$. Hence $h(dV(x)) := f(x)$, $x \in \mathscr{E}(\mathscr{g})$, defines, unambiguously, a linear functional on $\mathscr{A}$. Then $h(a) \geqq 0$ whenever $a \geqq 0$ and $a \in \mathscr{A}$. By Lemma 11.4.6 in [24], $\mathscr{A}$ is cofinal in $\mathscr{B}$, i.e. given $b = b^+ \in \mathscr{B}$, there is an $a \in \mathscr{A}$ such that $a \geqq 0$ and $a - b \geqq 0$. Therefore, h can be extended to a linear functional h_1 on $\mathscr{B}$ such that $h_1(b) \geqq 0$ when $b \in \mathscr{B}$ and $b \geqq 0$. Define $\pi(x) := \pi_{h_1}(dV(x))$, $x \in \mathscr{E}(\mathscr{g})$, and $U(g) := \overline{\pi_{h_1}(\widetilde{V}(g))}$, $g \in G$, where π_{h_1} is the *GNS* representation of the positive linear functional h_1 on $\mathscr{B}$. Lemma 11.4.7 in [24] says that U is a unitary representation of G such that dU is the closure $\widehat{\pi}$ of the representation π of $\mathscr{E}(\mathscr{g})$. By construction, $f(x) = h(dV(x)) = h_1(dV(x)) = \langle \pi_{h_1}(dV(x))\varphi_{h_1}, \varphi_{h_1}\rangle = \langle dU(x)\varphi_{h_1}, \varphi_{h_1}\rangle$ for $x \in \mathscr{E}(\mathscr{g})$, i.e. f is of the form required in $(GMP)_2$. It remains to check that $\mathscr{K}(V) \subseteq \mathscr{K}(U)$. Let $x \in \mathscr{K}(V)$. Then $dV(x) \geqq 0$ and hence $h_1(dV(x)) = h(dV(x)) \geqq 0$. This implies that $\pi(x) = \pi_{h_1}(dV(x)) \geqq 0$. Consequently, $\widehat{\pi}(x) = dU(x) \geqq 0$ which means that $x \in \mathscr{K}(U)$. □

The sufficiency part of Theorem 3.1.1 follows at once from Theorem 3.1.2 by letting V be the direct sum of representations $W \in \widehat{G}$.

3.1.3 Corollary. *Assume that the linear functional f in Theorem 3.1.2 is $\mathscr{K}(V)$-positive and pure. Then there exists a solution $\{U, \varphi\}$ of $(GMP)_2$ (or equivalently, a solution U of $(GMP)_4$) such that $\mathscr{K}(V) \subseteq \mathscr{K}(U)$ and U is irreducible. A similar assertion is valid also for the suffiency part of Theorem 3.1.1.*

Proof. It suffices to prove this in case of Theorem 3.1.2. Let us retain the notations from the proof of Theorem 3.1.2. Since f is pure in $\mathscr{E}(\mathscr{g})$, h is pure on $\mathscr{A}$. Hence the extension h_1 of h can be chosen also to be pure on $\mathscr{B}$. Therefore, $\pi_{h_1}(\mathscr{B})'_w = \mathbb{C} \cdot I$ [24, Coroll. 8.6.7]. Since G is connected, $U(G)' = dU(\mathscr{E}(\mathscr{g}))'$ [24,

Coroll. 10.2.18]. Moreover, $dU(\mathscr{E}(\mathscr{g}))' = \hat{\pi}(\mathscr{A})'_w = \pi(\mathscr{A})'_w$. Combining these facts with the definition of $\mathscr{B}$ we conclude easily that $U(G)' = \mathbb{C} \cdot I$, i.e. U is irreducible. □

Remarks

1) Theorem 3.1.1 is obviously a direct generalization of the following result of E.K. Haviland [11]: The classical moment problem is solvable if and only if $f(p) \geqq 0$ for all polynomials $p \in \mathbb{C}[x_1, \ldots, x_d]$ such that $p(.) \geqq 0$ on $\mathbb{R}^d$.

2) In order to give a similar interpretation of Theorem 3.1.2, we need some preliminaries. For this suppose that U_0 is a unitary representation of G and $\mathscr{V}$ is a family of unitary representations of G.

Let $\mathscr{K}(\mathscr{V}) := \bigcap_{V \in \mathscr{V}} \mathscr{K}(V)$ and let $U_{\mathscr{V}}$ denote the direct sum of the representations of $\mathscr{V}$. We shall say that U_0 is *infinitesimally weakly contained* in $\mathscr{V}$ if $\mathscr{K}(\mathscr{V}) \subseteq \mathscr{K}(U_0)$. In various aspects this notion resembles the weak containment, cf. J.M.G. Fell [5], [6].

By the *support* of the representation U_0 we shall mean the set of all $U \in \hat{G}$ for which $\mathscr{K}(U_0) \subseteq \mathscr{K}(U)$. It will be denoted by supp U_0. If $\mathscr{V}$ is a subset of $\hat{G}$, the set of all $U \in \hat{G}$ such that $\mathscr{K}(\mathscr{V}) \subseteq \mathscr{K}(U)$ is called the *closure* of $\mathscr{V}$ and denoted by $\overline{\mathscr{V}}$.

3) We recall another basic theorem of Haviland [11]. Let K be a subset of $\mathbb{R}^d$. The (CMP) admits a solution $\mu \in M(\mathbb{R}^d)$ supported by the closure of K if and only if $f(p) \geqq 0$ for all $p \in \mathbb{C}[x_1, \ldots, x_d]$ such that $p(t) \geqq 0$ for all $t \in K$.

Using the terminology from the preceding remark, Theorem 3.1.2 yields the following result which can be considered as a generalization of Haviland's theorem. Let $\mathscr{V}$ be a subset of $\hat{G}$. If f is a $\mathscr{K}(\mathscr{V})$-positive linear functional on the enveloping algebra $\mathscr{E}(\mathscr{g})$, then the (GMP) is solvable, and there is a solution U of $(GMP)_4$ such that supp $U_0 \subseteq \overline{\mathscr{V}}$. (For the proof, it suffices to apply Theorem 3.1.2 to $V := U_{\mathscr{V}}$ and use the equality $\mathscr{K}(U^{\mathscr{V}}) = \mathscr{K}(\mathscr{V})$.)

4) We continue the discussion from Remark 2). Consider the following condition:

(UC) *If* $\mathscr{K}(\mathscr{V}) \cap \mathscr{K}(\mathscr{W}) \subseteq \mathscr{K}(U)$ *for* $\mathscr{V}, \mathscr{W} \subseteq \hat{G}$ *and* $U \in \hat{G}$, *then* $\mathscr{K}(\mathscr{V}) \subseteq \mathscr{K}(U)$ *or* $\mathscr{K}(\mathscr{W}) \subseteq \mathscr{K}(U)$.

Suppose for a moment that (UC) is satisfied for the Lie group G. Let $\mathscr{C}$ be the family of sets $\overline{\mathscr{V}}$, where $\mathscr{V} \subseteq \hat{G}$. It is trivial by the definition of the closure $\overline{\mathscr{V}}$ that arbitrary intersections of members of $\mathscr{C}$ also belong to $\mathscr{C}$. Condition (UC) ensures that the union of any two sets from $\mathscr{C}$ belongs to $\mathscr{C}$. Hence there exists a unique topology on $\hat{G}$, denoted by τ_{inf}, such that the sets $\overline{\mathscr{V}}$ are precisely the closed sets relative to this topology.

We shall go one step further. Define $\mathscr{K}_n(\mathscr{V}) := \mathscr{K}(\mathscr{V}) \cap \mathscr{E}_n(\mathscr{g})$ and $\mathscr{K}_n(U) := \mathscr{K}(U) \cap \mathscr{E}_n(\mathscr{g})$ for $V \subseteq \hat{G}$ and $U \in \hat{G}$, where $\mathscr{E}_n(\mathscr{g})$ is the set of elements from $\mathscr{E}(\mathscr{g})$ whose degree does not exceed n. We now replace in the preceding the wedges $\mathscr{K}(\mathscr{V})$ and $\mathscr{K}(U)$ by $\mathscr{K}_n(\mathscr{V})$ and $\mathscr{K}_n(U)$, respectively. Let $(UC)_n$ be the condition on G and τ^n_{inf} the topology on G obtained in this way. It is trivial that $\tau^n_{\text{inf}} \subseteq \tau_{\text{inf}}$.

Of course, several questions immediately arise: Are the conditions (UC) and $(UC)_n$ for large n fulfilled? Is τ_{inf} equal to the hull-kernel topology on $\hat{G}$? Is $\tau^n_{\text{inf}} = \tau_{\text{inf}}$ for large n depending on G?

I don't know the answers for an arbitrary connected Lie group, but for a number of concrete Lie groups (including e.g. $SL(2,\mathbb{R})$, $SL(2,\mathbb{C})$ and the Heisenberg group) the answers are affirmative. Details and some related matters will be discussed in another publication.

3.2 Let $\{x_1, \ldots, x_d\}$ be a basis for the Lie algebra $\mathfrak{g}$ and let f be a positive linear functional on $\mathcal{E}(\mathfrak{g})$. We shall say that f is *analytic* (with respect to this basis) if there exists a constant $M > 0$ such that $|f(x_j^{2n})|^{1/2} \leqq M^n n!$ for all $n \in \mathbb{N}_0$ and $j = 1, \ldots, d$. The so-called FS^3-theory of integrability of Lie algebra representations gives the following result.

3.2.1 Proposition. *Suppose that the (connected) Lie group G is simply connected. If f is an analytic positive linear functional on $\mathcal{E}(\mathfrak{g})$, then the generalized moment problem is solvable and determinate.*

Proof. Since $|f(x_j^{2n})|^{1/2} = \|\pi_f(x_j)\varphi_f\|$ and f is analytic, φ_f is an analytic vector the operators $\pi_f(x_j)$, $j = 1, \ldots, d$. Hence all vectors of $\pi_f(\mathcal{E}(\mathfrak{g}))\varphi_f$ are analytic for $\pi_f(x_j)$ (cf. e.g. [24, Prop. 10.4.15]. Therefore, by the integrability theorem of M. Flato, J. Simon, H. Snellman and D. Sternheimer [7] (cf. [24, Theor. 10.5.4]), the representation π_f is exponentiable, i.e. there exists a unitary representation U of G such that $\overline{\pi_f(x_j)} = \overline{dU(x_j)}$, $j = 1, \ldots, d$. From this it follows that $\pi_f \subseteq dU$ and $\pi_f^* = dU$ ([24, Prop. 10.5.2]), so the (GMP) is solvable and determinate (cf. 2.4). □

3.3 *Example. The Universal Covering Group of $SL(2,\mathbb{R})$*
Let G denote the universal covering group of $SL(2,\mathbb{R})$. The Lie algebra $\mathfrak{g} = sl(2,\mathbb{R})$ has a basis $\{x_0, x_1, x_2\}$ satisfying the commutation relations

$$[x_0, x_1] = x_2, \quad [x_1, x_2] = -x_0, \quad [x_2, x_0] = x_1. \tag{1}$$

Set $y := -ix_0$, $x_+ := x_1 - ix_2$ and $x_- := x_1 + ix_2$. The center of the enveloping algebra $\mathcal{E}(\mathfrak{g})$ is generated by the Casimir element

$$\omega := 4(x_0^2 - x_1^2 - x_2^2) = -y^2 - 2(x_+x_- + x_-x_+).$$

3.3.1 Theorem. *Suppose f is a positive linear functional on $\mathcal{E}(\mathfrak{g})$ such that:*
(i) For each $r \in \mathbb{N}_0$, the moment sequence $(f(y^n(y^{2r}+1)); n \in \mathbb{N}_0)$ is determinate.
(ii) $|\langle \pi_f(\omega)\psi, \psi\rangle| \leqq$ const. $(\|\pi_f(y)\psi\|^2 + \|\psi\|^2)$, $\psi \in \mathcal{D}(\pi_f)$.
Then there exists a unique unitary representation U of G such that $\pi_f = dU$. The generalized moment problem for f is solvable and determinate.

Proof. For notational simplicity we abbreviate $X_j = \pi_f(x_j)$, $Y = \pi_f(y)$ and $\Omega = \pi_f(\omega)$. Since $x_j^+ = -x_j$ and hence $X_j^+ = -X_j$ for $j = 0, 1, 2$, we have

$$\|X_1\psi\|^2 + \|X_2\psi\|^2 = \|X_0\psi\|^2 + \frac{1}{4}\langle\Omega\psi, \psi\rangle, \quad \psi \in \mathcal{D}(\pi_f).$$

Combined with assumption (ii), this leads to

$$\|X_1\psi\|^2 + \|X_2\psi\|^2 \leqq c(\|Y\psi\|^2 + \|\psi\|^2), \quad \psi \in \mathcal{D}(\pi_f), \tag{2}$$

with some positive constant c. A simple induction argument based on the commutation rules (1) and on the inequality (2) shows that there exist constants $c_s > 0$ such that

$$\|X_{j_1}, \dots, X_{j_s}\psi\|^2 \leqq c_s(\|Y^s\psi\|^2 + \|\psi\|^2), \quad \psi \in \mathscr{D}(\pi_f), \tag{3}$$

for $j_1, \dots, j_s \in \{+, -, 0\}$ and $s \in \mathbb{N}$.

The main step of the proof is to show that for each $m \in \mathbb{N}$ the symmetric operator Y^m is essentially self-adjoint. For this we make use of assumption (i). Fix $m \in \mathbb{N}$ and $z \in \{i, -i\}$. Take $j, k, l \in \mathbb{N}_0$. For each $r \in \mathbb{N}_0$, there exists a measure $\mu_r \in M(\mathbb{R})$ such that

$$f(y^n(y^{2r} + 1)) = \int t^n \, d\mu_r(t) \quad \text{for} \quad n, r \in \mathbb{N}_0. \tag{4}$$

Define a measure ν_r by $d\nu_r = \frac{1}{2}(t^{2r} + 1)\, d\mu_0$. By (4), the measures μ_r and ν_r have the same moments, so $d\mu_r = d\nu_r = \frac{1}{2}(t^{2r} + 1)\, d\mu_0$ by assumption (i). Set $s := j + k + l$. Since the measure μ_{m+s} is determinate by (i), the polynomials are dense in $L^2(\mathbb{R}, \mu_{m+s})$ [1, Coroll. 2.3.3]. Hence we can find a sequence (p_n) of polynomials such that

$$p_n(t) \to ((t - j + k)^m + z)^{-1} \quad \text{in} \quad L^2(\mathbb{R}, \mu_{m+s}) \quad \text{as} \quad n \to \infty. \tag{5}$$

Recall that, by the definition of the *GNS* representation, $\|q(Y)\varphi_f\|^2 = \|\pi_f(q(y))\varphi_f\|^2 = f(q(y)^+ q(y))$ for each polynomial $q(.)$. Using these facts, in particular (3), (4) and (5), we get

$$\begin{aligned}
&\|X_+^j X_-^k Y^l \varphi_f - (Y^m + z) X_+^j X_-^k Y^l p_n(Y)\varphi_f\|^2 \\
&\quad = \|X_+^j X_-^k Y^l \{1 - ((Y - j + k)^m + z)p_n(Y)\}\varphi_f\|^2 \\
&\quad \leqq c_s \|Y^s\{1 - ((Y - j + k)^m + z)p_n(Y)\}\varphi_f\|^2 \\
&\qquad + c_s \|\{1 - ((Y - j + k)^m + z)p_n(Y)\}\varphi_f\|^2 \\
&\quad = c_s f((y^{2s} + 1)\{1 - ((y - j + k)^m + z)p_n(y)\}^+ \{1 - ((y - j + k)^m + z)p_n(y)\}) \\
&\quad = c_s \int \left|1 - ((t - j + k)^m + z)p_n(t)\right|^2 d\mu_s(t) \\
&\quad = c_s \int ((t - j + k)^{2m} + 1)\left|((t - j + k)^m + z)^{-1} - p_n(t)\right|^2 \frac{1}{2}(t^{2s} + 1)\, d\mu_0(t) \\
&\quad \leqq c_s' \int \left|((t - j + k)^m + z)^{-1} - p_n(t)\right|^2 \frac{1}{2}(t^{2(m+s)} + 1)\, d\mu_0(t) \\
&\quad = c_s' \int \left|((t - j + k)^m + z)^{-1} - p_n(t)\right|^2 d\mu_{m+s}(t) \\
&\quad \to 0 \quad \text{as} \quad n \to \infty
\end{aligned}$$

where c_s' is another positive constant. This shows that the closure of $(Y^m+z)\mathscr{D}(\pi_f)$ contains each vector $X_+^j X_-^k Y^l \varphi_f \equiv \pi_f(x_+^j x_-^k y^l)\varphi_f$, where $j,k,l \in \mathbb{N}_0$. By the Poincare-Birkhoff-Witt theorem, the linear span of these vectors coincides with $\pi_f(\mathscr{E}(\mathscr{g}))\varphi_f$, so it is dense in the underlying Hilbert space $\mathscr{H}(\pi_f)$. From this it follows that Y^m is essentially self-adjoint.

Let A denote the closure of the operator Y. Then A is a self-adjoint operator. Since the closure of Y^m is self-adjoint as just shown, we conclude that $A^m = \overline{Y^m}$, $m \in \mathbb{N}$. Formula (3) shows that the graph topology of $\pi_f(\mathscr{E}(\mathscr{g}))$ is already determined by the family of seminorms $\|Y^m.\|$, $m \in \mathbb{N}_0$. Since the GNS representation π_f is closed by definition, the preceding implies that

$$\mathscr{D}(\pi_f) = \bigcap_{m=0}^{\infty} \mathscr{D}(\overline{Y^m}) = \bigcap_{m=0}^{\infty} \mathscr{D}(A^m). \tag{6}$$

In order to complete the proof, we argue in the same way as in the proof of Theorem 9.2 in [12]. From the inequality (2) and the commutation rules (1) is follows immediately that the assumptions of Theorem 2.1 in [9] are satisfied with $\mathscr{X} := \mathscr{D}(\pi_f), \beta := \{Y, I\}$ and $\xi := \{X_0, X_1, X_2\}$. Therefore, by this theorem, every analytic vector for Y is also analytic for the operators X_j, $j = 0,1,2$. Since $\mathscr{D}(\pi_f) = \cap_m \mathscr{D}(A^m)$ by (6) and $A = \overline{Y}$ is self-adjoint, $\mathscr{D}(\pi_f)$ contains a dense set, say $\mathscr{D}_a$, of analytic vectors for the operators Y. (This can be obtained at once from the spectral theorem.) Hence each vector in $\mathscr{D}_a$ is analytic for the images $X_j = \pi_f(x_j)$ of the basis elements x_j, $j = 0,1,2$. By the FS^3 theorem ([7], [24]) used already earlier, there is a unique unitary representation U of G such that $\pi_f \subseteq dU$. Since Y^m is essentially self-adjoint, $\overline{Y^m} = (Y^m)^*$ for $m \in N$, so that $\mathscr{D}(\pi_f) = \cap_m \mathscr{D}(\pi_f(y^m)^*)$ again by (6). Therefore, the representation π_f of $\mathscr{E}(\mathscr{g})$ is self-adjoint, hence $\pi_f = dU$ ([24], Prop. 10.5.2). Thus $(GMP)_4$ is solvable and determinate (cf. 2.4). □

Remarks

Assumption (ii) was first used by P.E.T. Jorgensen [12]. Note that (ii) is trivially fulfilled if the operator $\pi_f(\omega)$ is bounded, so in particular if there exists a $\gamma \in \mathrm{R}$ such that $f((\omega-\gamma)^2) = 0$, since then $\pi_f(\omega) = \gamma \cdot I$. Such representations are called quasisimple (cf. /28/, p. 257).

3.4 *Example. The Heisenberg Group*

Let G be the three dimensional Heisenberg group. The Lie algebra $\mathscr{g}$ of G has a basis $\{x,y,z\}$ satisfying the commutation rules $[x,y] = z, [x,y] = [y,z] = 0$. The infinite dimensional representations in G are parametrized by the set $\mathbb{R} \setminus \{0\}$. The correspondence is given by $dU_\gamma(z) = i\gamma \cdot I$, $\gamma \in \mathbb{R} \setminus \{0\}$. Fix such a γ. The representation U_γ can be realized in the Hilbert space $L^2(\mathbb{R})$ such that $\mathscr{D}^\infty(U_\gamma) = \mathscr{S}(\mathbb{R})$, $dU_\gamma(ix) = i\frac{d}{dt} =: p$ and $dU_\gamma(iy) = \gamma t =: \gamma q$. Thus $dU_\gamma(\mathscr{E}(\mathscr{g}))$ is nothing but the $*$-algebra $\mathscr{A}(p,q)$ which is generated by the operators p,q and I on the domain $\mathscr{S}(\mathbb{R})$.

3.4.2 Theorem. *Let $\gamma \in \mathbb{R}, \gamma \neq 0$, and let f be a positive linear functional on $\mathscr{E}(\mathfrak{g})$ such that $f((z - i\gamma)^2) = 0$. Suppose that there exists a positive constant M such that $|f(y^{2n})|^{1/2} \leqq M^n n!$ for $n \in \mathbb{N}_0$.*
Then the generalized moment problem for f is solvable and the solution $\{U, \varphi\}$ can be chosen such that $dU(z) = i\gamma \cdot I$.

Proof. Since z belongs to the center of $\mathfrak{g}$, $f((z - i\gamma)^2) = 0$ leads to $\pi_f(z) = i\gamma \cdot I$, so that $[\pi_f(x), \pi_f(y)] = i\gamma \cdot I$. The second assumption means that the vector φ_f is an analytic vector for the operator $\pi_f(iy)$, hence all vectors from $\pi_f(\mathscr{E}(\mathfrak{g}))\varphi_f$ are analytic for $\pi_f(iy)$ ([24], Prop. 10.4.15). Having a dense set of analytic vectors, the symmetric operator $\pi_f(iy)$ has a self-adjoint closure Q by a classical result of E. Nelson. Put $W(t) := \exp itQ$, $t \in \mathbb{R}$, and $P := \pi_f(ix)$. From the commutation rules and the analyticity of the vectors of $\pi_f(\mathscr{E}(\mathfrak{g}))\varphi_f$ it follows easily that $\overline{P}W(t)\psi = W(t)(\overline{P} - t\gamma)\psi$ for $\psi \in \mathscr{D}(\overline{P})$ and $t \in \mathbb{R}$. By a recent result of R. Werner [29, Prop. 5], there exist a self-adjoint operator P_1 and a one-parameter unitary group $W_1(t)$ acting on a Hilbert space $\mathscr{H}_1 \supseteq \mathscr{H}(\pi_f)$ such that $P \subseteq P_1$, $W(t) = W_1(t) \restriction \mathscr{H}(\pi_f)$ and $P_1 W_1(t)\psi = W_1(t)(P_1 - t\gamma)\psi$ for $\psi \in \mathscr{D}(P_1)$ and $t \in \mathbb{R}$. (Note that Werner treats only the case where $\gamma = 1$, but the case of general $\gamma \neq 0$ can be reduced to the special case if we replace $W(t)$ by $\widetilde{W}(t) := W(t\gamma^{-1})$.) Define $W_2(s) := \exp isP_1$, $s \in \mathbb{R}$. Fix $t \in \mathbb{R}$. The one-parameter groups $s \to W_1(t)^{-1} W_2(s) W(t)$ and $s \to (\exp -its\gamma) W_2(s)$ have the infinitesimal generators $W_1(t)^{-1}(iP_1)W_1(t)$ and $i(P_1 - t\gamma)$, respectively. Since these two operators are equal, we conclude that $W_1(t)^{-1} W_2(s) W_1(t) = (\exp -its\gamma) W_2(s)$ for $t, s \in \mathbb{R}$. From the latter relation and the multiplication rules in the group G it follows immediately that the maps $\exp tx \to W_2(-t)$, $\exp ty \to W_1(-t)$ and $\exp tz \to \exp it\gamma$ define a unitary representation U of G. Clearly, $dU(z) = i\gamma \cdot I$. By construction, we have $\overline{\pi_f(ix)} = \overline{P} \subseteq P_1 = \overline{dU(ix)}$ and $\overline{\pi_f(iy)} = Q \subseteq \overline{dU(iy)}$. This implies that $\mathscr{D}(\pi_f) \subseteq \bigcap_{n=1}^{\infty} \mathscr{D}((\overline{dU(x)})^n) \bigcap \mathscr{D}((\overline{dU(y)})^n) = \mathscr{D}^\infty(U)$ ([24], Coroll. 10.2.12). Therefore, $\pi_f \subseteq dU$, so $(GMP)_4$ is solvable. □

Remarks
The main ingredient of the preceding proof was a dilation theorem for the generalized Weyl relation $\overline{P}W(t) = W(t)(\overline{P} + t)$. This relation was extensively studied in [13] and in [23].

4 The quantum moment problem

4.1 Suppose that U is a unitary representation of G. Let $\mathbb{B}_1(\mathscr{D}^\infty(U))_+$ denote the set of all positive self-adjoint operators ϱ on $\mathscr{H}(U)$ such that $\varrho\mathscr{H}(U) \subseteq \mathscr{D}^\infty(U)$ and the operator $dU(x)\varrho$ is of trace class for each $x \in \mathscr{E}(\mathfrak{g})$.

Suppose $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$. Define $f_{U,\varrho}(x) = tr\, dU(x)\varrho$ for $x \in \mathscr{E}(\mathfrak{g})$, where *tr* means the trace. The operator ϱ can be written as $\varrho = \sum_n \varphi_n \otimes \varphi_n$, where $(\varphi_n ; n \in \mathbb{N})$ is a sequence of vectors $\varphi_n \in \mathscr{D}^\infty(U)$. (As usual, $\varphi_n \otimes \varphi_n$ denotes the operator $\langle ., \varphi_n \rangle \varphi_n$.) For each such representation of ϱ we have

$$\sum_n \|dU(x)\varphi_n\|^2 < \infty \quad \text{for all} \quad x \in \mathscr{E}(\mathfrak{g}) \tag{1}$$

and

$$f_{U,\varrho}(x) = \sum_n \langle dU(x)\varphi_n, \varphi_n\rangle, \quad x \in \mathscr{E}(\mathfrak{g}). \tag{2}$$

In particular, we see from (2) that $f_{U,\varrho}$ is a $\mathscr{K}(U)$-positive linear functional on $\mathscr{E}(\mathfrak{g})$.

Conversely, if $(\varphi_n;\ n \in \mathbb{N})$ is a sequence from $\mathscr{D}^\infty(U)$ satisfying (1), then $\varrho := \sum_n \varphi_n \otimes \varphi_n$ defines an operator of $\mathbb{B}_1(\mathscr{D}^\infty(U))_+$. (The simple proofs of the unproven assertions can be found e.g. in [24], Sect. 5.1)

The following problem will be called the quantum moment problem.

(QMP) *Let U be a unitary representation of G and let f be a linear functional on the enveloping algebra. When does there exist an operator $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$ such that $f = f_{U,\varrho}$, i.e. $f(x) = tr\, dU(x)\varrho$ for $x \in \mathscr{E}(\mathfrak{g})$?*

This problem was first studied (in a slightly different formulation) by T. Sherman [25] and by S.L. Woronowicz [30]. Woronowicz treated the important special case where G is the Heisenberg group and U is the Schrödinger representation. Then we have $dU(\mathscr{E}(\mathfrak{g})) = \mathscr{A}(p, q)$, see Example 3.4 above. According to [30], the term "quantum moment problem" was proposed by the physisist R.S. Ingarden. We shall adopt this terminology also in the general case formulated above.

In classical mechanics the states are described by measures on the phase space and the expectation value is an integral. In quantum physics the states are usually given by density matrices and the expectation value is the trace. In this context it seems to be quite natural to consider the quantum moment problem as another non-commutative version of the classical moment problem.

Let U and f be as in (*QMP*). We say that the quantum moment problem for f is *solvable* if $f = f_{U,\varrho}$ for some $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$. Each such operator ϱ is called a *solution* of the (*QMP*). We shall say that a solvable quantum moment problem is *determinate* or that f is (QMP)-*determinate* if there is only one operator $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$ such that $f = f_{U,\varrho}$.

4.2 The next theorem is essentially due to T. Sherman [25].

4.2.1 Theorem. *Suppose that U is a unitary representation of G such that the operator $U_h := \int U(g)h(g)\, d\mu_l(g)$ is completely continuous for every function $h \in L^1(G, \mu_l)$, where μ_l is the left-invariant Haar measure of G. Left f be a linear functional on $\mathscr{E}(\mathfrak{g})$.*
Then the quantum moment problem for f is solvable if and only if f is $\mathscr{K}(U)$-positive.

For the convenience of the reader and because this result is not explicitely stated in this form in [25], we shall say some words concerning the proof of this theorem. The necessity is clear, since if $f = f_{U,\varrho}$ with $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$, then f is, of course, $\mathscr{K}(U)$-positive by (2). Conversely, suppose that f is $\mathscr{K}(U)$-positive. Let $\{x_1, \ldots, x_d\}$ be a basis for the Lie algebra $\mathfrak{g}$ and let $\Delta := x_1^2 + \cdots + x_d^2$. By the assumptions of Theorem 4.2.1, a result of E. Nelson and W.F. Stinespring [18, Theor. 4.1] applies and states that the closure of the operator $dU(\Delta - 1)$ has a completely continuous inverse. Therefore the assertion follows at once from Theorem 5 in [25] (or from [24, Theor. 5.3.8]).

Remarks
1) Recall that G is called a *CCR* group if for every irreducible unitary representation U of G the assumptions of Theorem 4.2.1 are fulfilled, i.e. U_h is completely continuous for each $h \in L^1(G, \mu_l)$. Note that G is a *CCR* group if G is semisimple or if G is nilpotent. This follows from classical results of Harish-Chandra [10] and A.A. Kirillov [14], respectively.
2) The (QMP) may be viewed as a special case of the following general problem: When is a linear functional f on an unbounded operator algebra of the form $f(.) = tr.\varrho$ for some positive trace class operator ϱ? This problem is treated in [24], Section 5.3.

4.3 Let U be a unitary representation of G and let f be a linear functional on $\mathscr{E}(\mathscr{g})$. Suppose that $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$ is a solution of the (QMP) for f, i.e. $f = f_{U,\varrho}$. We shall define a solution of the (GMP) associated with ϱ.

Let $\varrho = \sum_{n=1}^\infty \varphi_n \otimes \varphi_n$ be a representation of ϱ with $\varphi_n \in \mathscr{D}^\infty(U)$ for $n \in \mathbb{N}$. Note that such a representation is always possible, but it is not unique. We let φ_ϱ denote the vector (φ_n) in the Hilbert direct sum $\sum_{n=1}^\infty \oplus \mathscr{H}(U)$. Let U_ϱ be the restriction of the direct sum $\tilde{U} := \sum_{n=1}^\infty \oplus U$ of countably many copies of U to the closed linear span of $\tilde{U}(G)\varphi_\varrho$. Then U_ϱ is a unitary representation of G with cyclic vector φ_ϱ such that $\langle dU_\varrho(x)\varphi_\varrho, \varphi_\varrho\rangle = \langle d\tilde{U}(x)\varphi_\varrho, \varphi_\varrho\rangle = \sum_n \langle dU(x)\varphi_n, \varphi_n\rangle = tr\, dU(x)\varrho = f(x)$, $x \in \mathscr{E}(\mathscr{g})$, by (2), i.e. $\{U_\varrho, \varphi_\varrho\}$ is a solution of the (GMP). Since $\langle U_\varrho(g)\varphi_\varrho, \varphi_\varrho\rangle = \langle \tilde{U}(g)\varphi_\varrho, \varphi_\varrho\rangle = \sum_n \langle U(g)\varphi_n, \varphi_n\rangle = tr\, U(g)\varrho$ for $g \in G$, the positive definite function $\langle U_\varrho(.)\varphi_\varrho, \varphi_\varrho\rangle$ on G depends only on ϱ, but not on the special representation $\sum_n \varphi_n \otimes \varphi_n$ of the operator ϱ. That is, the couple $\{U_\varrho, \varphi_\varrho\}$ is uniquely determined up to unitary equivalence by ϱ. We call $\{U_\varrho, \varphi_\varrho\}$ the *canonical solution of the* (GMP) *associated with* ϱ.

4.3.1 Lemma. *Let $U \in \hat{G}$ and let f be a linear functional on $\mathscr{E}(\mathscr{g})$ such that the* (QMP) *is solvable. If f is* (GMP)*-determinate, then f is also* (QMP)*-determinate.*

Proof. Suppose that $f = f_{U,\varrho_1} = f_{U,\varrho_2}$ with $\varrho_1, \varrho_2 \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$. We write ϱ_j as $\varrho_j = \sum_n \varphi_{jn} \otimes \varphi_{jn}$, where $\varphi_{jn} \in \mathscr{D}^\infty(U)$ for $n \in \mathbb{N}$ and $j = 1, 2$. By the (GMP)-determinacy of f, we then have $\langle U_{\varrho_1}(g)\varphi_{\varrho_1}, \varphi_{\varrho_1}\rangle = \langle U_{\varrho_2}(g)\varphi_{\varrho_2}, \varphi_{\varrho_2}\rangle$ or equivalently $tr\, U(g)\varrho_1 = tr\, U(g)\varrho_2$ for $g \in G$. (We keep the notation of the preceding paragraph.) Since U is irreducible by assumption, the linear span of elements $U(g)$, $g \in G$, is ultraweakly dense in $\mathbb{B}(\mathscr{H}(U))$. Therefore, the preceding yields $tr\, a\varrho_1 = tr\, a\varrho_2$ for all $a \in \mathbb{B}(\mathscr{H}(U))$. Consequently, $\varrho_1 = \varrho_2$. □

5 Vector states of irreducible unitary representations

5.1 In representation theory it is, of course, of special interest when the representation U in $(\mathrm{GMP})_2$ is *irreducible*, i.e. when f is a vector state of some irreducible representation. In this section we shall be concerned with this case.

Suppose U is a unitary representation of G and $\varphi \in \mathscr{D}^\infty(U)$. We call the positive linear functional $f_{U,\varphi}(.) := \langle dU(.)\varphi, \varphi\rangle$ on the $*$-algebra $\mathscr{E}(\mathscr{g})$ a *vector state* of U. (Strictly speaking, $f_{U,\varphi}$ is not a state in general, since we may have

that $f_{U,\varphi}(1) \equiv \|\varphi\|^2 \neq 1$. Also, it would be more precise to speak about a vector state of the infinitesimal representation dU.) We shall say that a vector state $f_{U,\varphi}$ is *uniquely determined* or that $f_{U,\psi}$ is (VS)-*determinate* when the equality $f_{U,\varphi} = f_{U,\psi}$ with $\psi \in \mathscr{D}^\infty(U)$ always implies that $\psi = \lambda\varphi$, $\lambda \in \mathbb{C}$.

There are a number of natural and interesting problems concerning vector states of irreducible representations. We mention three of them. For this let $U \in \widehat{G}$ and $\varphi \in \mathscr{D}^\infty(U)$.

1) *Determinacy Problems*
Is $f_{U,\varphi}$ (GMP)-determinate? Is $f_{U,\varphi}$ (QMP)-determinate? Is $f_{U,\varphi}$ (VS)-determinate?
2) *Characterization of Vector States*
Given a positive linear functional f on $\mathscr{E}(\mathfrak{g})$, when is f of the form $f = f_{U,\varphi}$ for some vector $\varphi \in \mathscr{D}^\infty(U)$?
3) *Separation of the Dual*
Let $U \in \widehat{G}$ and $V \in \widehat{G}$. Does $\mathscr{K}(U)^* \cap \mathscr{K}(V)^* \neq \{0\}$ imply that $U \cong V$? In particular, if there are non-zero vectors $\varphi \in \mathscr{D}^\infty(U)$ and $\psi \in \mathscr{D}^\infty(V)$ such that $f_{U,\varphi} = f_{V,\psi}$, does it follow that $U \cong V$?

Of course, general results concerning 1) and 3) are easily obtained if we impose, roughly speaking, analyticity assumptions. For instance, if $\{x_1, \ldots, x_d\}$ is a basis of $\mathfrak{g}$ and if the vector $\varphi \in \mathscr{D}^\infty(U)$ is analytic for the operators $dU(x_j), j = 1, \ldots, d$, then the answers to the questions in 1.) and in 3.) are affirmative. In the next two subsections we shall study these problems under much weaker assumptions.

5.2 *Example* $SL(2,\mathbb{R})$
Let $G = SL(2,\mathbb{R})$. The Lie algebra $\mathfrak{g} = sl(2,\mathbb{R})$ has been described in Example 3.3. We retain the notation introduced therein and we set $z := 2y = -2ix_0$.

In what follows we need some facts about the irreducible unitary representations of G. (Proofs of the assertions stated here can be found e.g. in [26, Chap. 8]. For each $U \in \widehat{G}$, the spectrum of the self-adjoint operator $\overline{dU(z)}$ is a set of integers, denoted by $J(U)$, and the operator $dU(\omega)$ is a scalar multiple of the identity operator, say $\gamma(U) \cdot I$. Every $U \in \widehat{G}$ is uniquely determined by the set $J(U)$ and the number $\gamma(U)$. Each nontrivial representation $U \in \widehat{G}$ is unitarily equivalent to one of the following list. The representations of this list are mutually inequivalent.

a) *The discrete series $U_r^+, r \in \mathbb{N}$, and $U_{-r}^-, r \in \mathbb{N}$:*

$$J(U_r^+) = r + 2\mathbb{N}_0, \quad \gamma(U_r^+) = 1 - (r-1)^2$$
$$J(U_{-r}^-) = r - 2\mathbb{N}_0, \quad \gamma(U_{-r}^-) = 1 - (r-1)^2$$

b) *The principle series U_{is}^e, $s \in [0, +\infty)$, and U_{is}^0, $s \in (0, +\infty)$:*

$$J(U_{is}^e) = 2\mathbb{Z}, \quad \gamma(U_{is}^e) = 1 + s^2$$
$$J(U_{is}^0) = 2\mathbb{Z} + 1, \quad \gamma(U_{is}^0) = 1 + s^2$$

c) *The complementary series* U_s^e, $s \in (0,1)$:

$$J(U_s^e) = 2\mathbb{Z}, \quad \gamma(U_s^e) = 1 - s^2.$$

We briefly describe the infinitesimal action of these representations $U \in \widehat{G}$. The underlying Hilbert space of U is $l_2(J(U))$. Vectors of this space will be written as sequences (φ_m). We have $dU(z)(\varphi_m) = (m\varphi_m)$, $dU(x_+)(\varphi_m) = (\alpha_{m-2}\varphi_{m-2})$ and $dU(x_-)(\varphi_m) = (-\overline{\alpha_m}\varphi_{m+2})$ for $(\varphi_m) \in \mathscr{D}^\infty(U) \subseteq l_2(J(U))$. Here the α_m are complex numbers depending only on $\gamma(U)$. The concrete form of these numbers is not needed here. For our purposes it is sufficient to know that $\alpha_m \neq 0$ if $m+2 \in J(U)$ and that $|\alpha_m| \leqq \text{const.}\ |m|$ for large $|m|$.

Our main result in this subsection is

5.2.1 Theorem. *Suppose that $U \in \widehat{G}$ and that f is a $\mathscr{K}(U)$-positive linear functional on $\mathscr{E}(\mathcal{g})$ such that the moment sequence $(f(z^n);\ n \in \mathbb{N}_0)$ is determinate.*
Then the functional f is (GMP)*-determinate and there exists a unique operator $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$ such that $f(x) = f_{U,\varrho}(x) = \operatorname{tr} dU(x)\varrho$ for $x \in \mathscr{E}(\mathcal{g})$.*

For the proof of this theorem we require the following

5.2.2 Lemma. *Keep the assumptions and the notation of Theorem 5.2.1. If the couple $\{V, \varphi\}$ is a solution of the generalized moment problem* $(\text{GMP})_2$ *for f, then V is unitarily equivalent to a direct sum of finitely or countably many copies of U.*

Proof. To avoid trivial cases, we assume that $f \neq 0$ and that U is not the trivial representation. Since V is cyclic, $\mathscr{H}(V)$ is separable. Therefore, V is unitarily equivalent to a direct integral $W = \int_{\widehat{G}}^{\oplus} W_\lambda \, d\mu(\lambda)$ of primary representations W_λ of G (cf. [15, 8.4, second part of Theor. 3]). The representations W_λ are of the form $W_\lambda = U_\lambda \otimes I_\lambda$, where U_λ is a representation of the class $\lambda \in \widehat{G}$ and I_λ is the trivial representation on some Hilbert space.
For notational simplicity we identify V and W. Let U belong to $\lambda_0 \in \widehat{G}$. Our proof will be complete once we have shown that μ is supported on the singleton $\{\lambda_0\}$. Indeed, then we write the tensor product $U_{\lambda_0} \otimes I_{\lambda_0}$ as a direct sum and the assertion follows.
Set

$$\gamma_0 := \gamma(U), \quad \Gamma := \{\lambda \in \widehat{G} : \gamma(U_\lambda) = \gamma_0\}$$

and

$$\Gamma_n := \left\{\lambda \in \widehat{G} : |\gamma(U_\lambda) - \gamma_0| \geqq \frac{1}{n}\right\} \quad \text{for} \quad n \in \mathbb{N}.$$

Let $\varphi = (\varphi_\lambda)$. Since $dU(\omega - \gamma_0) = 0$ and hence $\pm(\omega - \gamma_0) \in \mathscr{K}(U)$, it follows that

$\pm f((\omega-\gamma_0)^2) \geqq 0$, so $f((\omega-\gamma_0)^2)=0$. For $n \in \mathbb{N}$, we therefore have

$$\begin{aligned} 0 &= f((\omega-\gamma_0)^2) \\ &= \|dV(\omega-\gamma_0)\varphi\|^2 \\ &= \int_{\widehat{G}} \|dW_\lambda(\omega-\gamma_0)\varphi_\lambda\|^2\, d\mu(\lambda) \\ &\geqq \int_{\Gamma_n} |\gamma(U_\lambda)-\gamma_0|^2 \|\varphi_\lambda\|^2\, d\mu(\lambda) \\ &\geqq n^{-2} \int_{\Gamma_n} \|\varphi_\lambda\|^2\, d\mu(\lambda). \end{aligned}$$

Hence $\varphi_\lambda = 0$ μ-a.e. on Γ_n. Since $G \setminus \Gamma = \bigcup_n \Gamma_n$, $\varphi_\lambda = 0$ on $G \setminus \Gamma$. Since φ is cyclic for V, $\mu(G \setminus \Gamma) = 0$ and so $V = W = \overset{\oplus}{\int} W_\lambda\, d\mu(\lambda)$.

Since $\gamma(U) = \gamma(U_\lambda)$ for $\lambda \in \Gamma$, it follows from the above list of representations that U and the elements of Γ belong either to a), to b) or to c). If $U = U_s^e$ for some $s \in [0,1)$, Γ is a singleton and the proof is complete. Assume now that $U \neq U_s^e$ for all $s \in [0,1)$. Then Γ consists of two elements λ_1 and λ_2. Put $W_j := \overset{\oplus}{\int}_{\{\lambda_j\}} W_\lambda\, d\mu(\lambda)$. Then $W = W_1 \oplus W_2$. Let $\varphi = \varphi_1 \oplus \varphi_2$ and $f_j(.) := \langle dW_j(.)\varphi_j, \varphi_j\rangle$, $j = 1, 2$. Clearly, $f = f_1 + f_2$. Since $W_{\lambda_j} = U_{\lambda_j} \otimes I_{\lambda_j}$, f_j is $\mathscr{K}(U_{\lambda_j})$-positive. Recall that f is $\mathscr{K}(U)$-positive by assumption and that $G = SL(2,\mathbb{R})$ is a CCR group. Hence, by Theorem 4.2.1, there are operators $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$ and $\varrho_j \in \mathbb{B}_1(\mathscr{D}^\infty(U_{\lambda_j}))_+$ such that $f = f_{U,\varrho}$ and $f_j = f_{U_{\lambda_j},\varrho_j}$. (For f_j this follows also directly from the construction without appealing to Theorem 4.2.1) Let $\varrho = \sum_k \psi_k \otimes \psi_k$ and $\varrho_j = \sum_k \psi_k^j \otimes \psi_k^j$, where $\psi_k \equiv (\psi_{km}) \in \mathscr{D}^\infty(U)$ and $\psi_k^j = (\psi_{km}^j) \in \mathscr{D}^\infty(U_{\lambda_j})$.

Case 1 $U = U_r^+$, $r \in \mathbb{N}$.
Let $U_{\lambda_1} = U_r^+$ and $U_{\lambda_2} = U_{-r}^-$. We have for $n \in \mathbb{N}_0$

$$\begin{aligned} \sum_k \sum_{m=0}^{\infty} (r+2m)^n |\psi_{k,r+2m}|^2 &= \sum_k \langle dU_r^+(z^n)\psi_k, \psi_k\rangle \\ &= f(z^n) = f_1(z^n) + f_2(z^n) \\ &= \sum_k \langle dU_r^+(z^n)\psi_k^1\rangle + \sum_k \langle dU_{-r}^-(z^n)\psi_k^2, \psi_k^2\rangle \\ &= \sum_k \sum_{m=0}^{\infty} (r+2m)^n |\psi_{k,r+2m}|^2 + \sum_k \sum_{m=0}^{\infty} (r-2m)^n |\psi_{k,r-2m}^2|^2. \end{aligned}$$

Since the moment sequence $(f(z^n))$ is determinate, the preceding implies that the

measures

$$\nu_1 := \sum_{m=0}^{\infty} \Big(\sum_k |\psi_{k,r+2m}|^2 \Big) \delta_{r+2m}$$

and

$$\nu_2 := \sum_{m=0}^{\infty} \Big(\Big(\sum_k |\psi^1_{k,r+2m}|^2 \Big) \delta_{r+2m} + \Big(\sum_k |\psi^2_{k,r-2m}|^2 \Big) \delta_{r-2m} \Big)$$

coincide, so that for all $m \in \mathbb{N}_0$

$$\sum_k |\psi_{k,r+2m}|^2 = \Big(\sum_k |\psi^2_{k,r}|^2 \Big) \delta_{0,m} + \sum_k |\psi^1_{k,r+2m}|^2 \tag{1}$$

and

$$\psi^2_{k,r-2m} = 0 \quad \text{for} \quad m \in \mathbb{N}. \tag{2}$$

Using (2), we get

$$\begin{aligned} \sum_k \sum_{m=0}^{\infty} |\psi_{k,r+2m}|^2 |\alpha_{r+2m}|^2 &= \sum_k \| dU_r^+(x_+)\psi_k \|^2 \\ &= f((x_+)^+ x_+) = f_1((x_+)^+ x_+) + f_2((x_+)^+ x_+) \\ &= \sum_k \| dU_r^+(x_+)\psi_k^1 \|^2 + \sum_k \| dU_{-r}^-(x_+)\psi_k^2 \|^2 \\ &= \sum_k \sum_{m=0}^{\infty} |\psi^1_{k,r+2m}|^2 |\alpha_{r+2m}|^2 + 0. \end{aligned}$$

Combined with (1), the latter gives $\psi^2_{kr} = 0$ for all k. Because of (2), we obtain $\varrho_2 = 0$, $f_2 \equiv 0$ and hence $\varphi_2 = 0$. Since φ is cyclic for V, this is only possible if $\mathscr{H}(W_2) = \{0\}$, i.e. $\mu(\{\lambda_2\}) = 0$ and μ is supported on $\{\lambda_1\}$.

Case 2 $U = U^e_{is}$, $s \in (0, +\infty)$

Let $U_{\lambda_1} = U^e_{is}$ and $U_{\lambda_2} = U^0_{is}$. For $n \in \mathbb{N}_0$, we have

$$\begin{aligned} \sum_k \sum_{m \in \mathbb{Z}} (2m)^n |\psi_{k,2m}|^2 &= \sum_k \langle dU^e_{is}(z^n)\psi_k, \psi_k \rangle \\ &= f(z^n) = f_1(z^n) + f_2(z^n) \\ &= \sum_k \langle dU^e_{is}(z^n)\psi_k^1, \psi_k^1 \rangle + \sum_k \langle dU^0_{is}(z^n)\psi_k^2, \psi_k^2 \rangle \\ &= \sum_k \sum_{m \in \mathbb{Z}} (2m)^n |\psi^1_{k,2m}|^2 + \sum_k \sum_{m \in \mathbb{Z}} (2m+1)^n |\psi^2_{k,2m+1}|^2. \end{aligned}$$

By the determinacy of the sequence $(f(z^n))$, the measures

$$\nu_3 := \sum_{m\in\mathbb{Z}} \left(\sum_k |\psi_{k,2m}|^2 \right) \delta_{2m}$$

and

$$\nu_4 := \sum_{m\in\mathbb{Z}} \left(\left(\sum_k |\psi^1_{k,2m}|^2 \right) \delta_{2m} + \left(\sum_k |\psi^2_{k,2m+1}|^2 \right) \delta_{2m+1} \right)$$

are equal. This in turn yields that $\psi^2_{k,2m+1} = 0$ for all k and m. As in case 1 this gives $\mu(\{\lambda_2\}) = 0$.

The cases $U = U^-_{-r}$ and $U = U^0_{is}$ are treated in a similar way. □

Proof of Theorem 5.2.1 Let us assume that $f \not\equiv 0$, since otherwise the assertion is trivial. Suppose that $\{V_j, \varphi^j\}$, $j = 1, 2$, are two solutions of the $(\mathrm{GMP})_2$ for f. Then $f(x) = \langle dV_j(x)\varphi^j, \varphi^j\rangle$, $x \in \mathscr{E}(\mathscr{G})$. Take $j \in \{1, 2\}$. By Lemma 5.2.2., $V_j = \sum_k \oplus U$. Let (φ^j_k) be the vector in the direct sum $\sum_k \oplus \mathscr{H}(U)$ which corresponds to φ^j under the unitary equivalence. Clearly, we have $\varrho_j := \sum_k \varphi^j_k \otimes \varphi^j_k \in \mathbb{B}_1(\mathscr{D}^\infty(V_j))_+$ and $f(.) = \sum_k \langle dV_j(.)\varphi^j_k, \varphi^j_k\rangle = tr\, dV_j(.)\varrho_j = f_{V_j,\varrho_j}(.)$. Our proof is complete once we have shown that $\varrho_1 = \varrho_2$, since then $\langle V_1(g)\varphi^1, \varphi^1\rangle = tr\, U(g)\varrho_1 = tr\, U(g)\varrho_2 = \langle V_2(g)\varphi^2, \varphi^2\rangle$, $g \in G$, and hence $\{V_1, \varphi^1\}$ is unitarily equivalent to $\{V_2, \varphi^2\}$.

We write the vectors $\varphi^j_k \in \mathscr{H}(U) \equiv l_2(J(U))$ as $\varphi^j_k = (\varphi^j_{km})$. For $j = 1, 2$ and $n \in \mathbb{N}_0$, we have

$$f(z^n) = \sum_k \langle dV_j(z^n)\varphi^j_k, \varphi^j_k\rangle = \sum_k \sum_m m^n |\varphi^j_{km}|^2, \tag{3}$$

i.e. the measure $\mu_j := \sum_m (\sum_k |\varphi^j_{km}|^2)\delta_m$ has the n-th moment $f(z^n)$. By the determinacy of the moment sequence $(f(z^n))$, it follows that $\mu_1 = \mu_2$, so

$$\beta_m := \sum_k |\varphi^1_{km}|^2 = \sum_k |\varphi^2_{km}|^2 \quad \text{for} \quad m \in J(U). \tag{4}$$

Since $f \not\equiv 0$, the set $\mathbb{I} := \{m \in J(U) : \beta_m \not\equiv 0\}$ is not empty. Let $m \in \mathbb{I}$ and $s \in \mathbb{N}$ be such that $m + 2s \in J(U)$. Then, by the possible forms of $J(U)$ listed above, $m + 2j \in J(U)$ if $j \in \mathbb{N}$, $j \leqq s$. Thus

$$\begin{aligned} f((x_+)^s z^n) &= \sum_k \langle dU(x_+)^s dU(z^n)\varphi^j_k, \varphi^j_k\rangle \\ &= \sum_k \sum_k m^n \psi^j_{km}\varphi^j_{k,m+2s}\alpha_m\alpha_{m+2}\cdots\alpha_{m+2(s-1)} \end{aligned} \tag{5}$$

for $j = 1, 2$ and $n \in \mathbb{N}_0$. Set

$$\eta_m = \sum_k \left(\varphi^1_{km}\overline{\varphi^1_{k,m+s}} - \varphi^2_{km}\overline{\varphi^2_{k,m+2s}} \right) \alpha_m \cdots \alpha_{m+2(s-1)}\, \beta_m^{-1}$$

for $m \in \mathbb{I}$ and $\eta_m = 0$ otherwise. From (4) and (5) we obtain

$$\sum_{m \in J(U)} m^n \overline{\eta_m} \beta_m = 0 \quad \text{for all} \quad n \in \mathbb{N}_0. \tag{6}$$

Since $|\alpha_m \dots \alpha_{m+2(s-1)}| \leqq \text{const.} |m|^s$ for large $|m|$ and the right-hand side of (3) is finite for arbitrary $n \in \mathbb{N}$, it follows that the sequence (η_m) is in $L^2(\mu_1)$. But μ_1 is determinate, so the polynomials are dense in $L^2(\mu_1)$ ([1]). Thus $\eta_m = 0$ for all $m \in \mathbb{I}$. Since $\alpha_m \dots \alpha_{m+2(s-1)} \neq 0$, this yields

$$\sum_k \varphi^1_{km} \overline{\varphi^1_{k,m+2s}} = \sum_k \varphi^2_{km} \overline{\varphi^2_{k,m+2s}} \tag{7}$$

for $m \in \mathbb{I}$ and $s \in \mathbb{N}$ provided that $m + 2s \in J(U)$. Replacing x_+ by x_- in the preceding reasoning, we obtain the same assertion for negative integers s. From (7) we conclude that $\varrho_1 = \varrho_2$. □

The following two corollaries contain some answers to the problems 3) and 1) formulated in 5.1.

5.2.3 Corollary. *Suppose that f is a non-zero positive linear functional on $\mathscr{E}(\mathscr{g})$ such that the moment sequence $(f(z^n);\ n \in \mathbb{N}_0)$ is determinate. Let $U \in \hat{G}$ and $V \in \hat{G}$.*

(i) *If $f \in \mathscr{H}(U)^* \cap \mathscr{H}(V)^*$, then $U \cong V$.*
(ii) *If $f = f_{U,\varrho} = f_{V,\sigma}$ with $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$ and $\sigma \in \mathbb{B}_1(\mathscr{D}^\infty(V))_+$,*

then there exists an isometry T of $\mathscr{H}(V)$ onto $\mathscr{H}(U)$ such that $T^{-1}UT = V$ and $T^{-1}\varrho T = \sigma$.

Proof. If $f \in \mathscr{H}(U)^* \cap \mathscr{H}(V)^*$, then, by Theorem 5.2.1, f is of the form assumed in (ii). Therefore, it suffices to prove (ii). Let $f = f_{U,\varrho} = f_{V,\sigma}$ be as in (ii). The proof of Lemma 5.2.2 shows that $U \cong V$. (Indeed, first we obtain $\gamma(U) = \gamma(V)$ and then we apply the second part of this proof with $f = f_{U,\varrho}$, $f_1 \equiv 0$ and $f_2 = f_{V,\sigma}$.) Let T be an isometry of $\mathscr{H}(V)$ onto $\mathscr{H}(V)$ such that $T^{-1}UT = V$. Then, obviously, $\tilde{\sigma} := T^{-1}\varrho T \in \mathbb{B}_1(\mathscr{D}^\infty(V))_+$ and $f_{V,\sigma} = f_{V,\tilde{\sigma}}$. By the uniqueness assertion of Theorem 5.2.1, $\sigma = \tilde{\sigma}$. □

Letting $U = V$ and $\varrho = \varphi \otimes \varphi$ in Corollary 5.2.3, (ii), we obtain

5.2.4 Corollary. *Let $U \in \hat{G}$ and $\varphi \in \mathscr{D}^\infty(U)$. The vector state $f_{U,\varphi}$ is uniquely determined if the moment sequence $(f_{U,\varphi}(z^n) \equiv \langle dU(z^n)\varphi, \varphi\rangle;\ n \in \mathbb{N}_0)$ is determinate.*

5.3 *Example The Heisenberg Group*

Let us retain the notation of Example 3.4. We fix a non-zero real number γ and set $U := U_\gamma$. Recall that $dU(z) = i\gamma \cdot I$, $dU(ix) = p \equiv i\frac{d}{dt}$ and $dU(iy) = \gamma q \equiv \gamma t$.

Suppose $\varphi \in \mathscr{D}^\infty(U) \equiv \mathscr{S}(\mathbb{R})$. Let $S(\varphi)$ be the set of all $t \in \mathbb{R}$ for which there exists an $r \in \mathbb{N}_0$ such that $\varphi^{(r)}(t) \neq 0$. Let $S_n(\varphi)$, $n \in \mathbb{I}$, denote the connected components of the open subset $S(\varphi)$ of $\mathbb{R}$.

5.3.1 Theorem. *Let $\varphi \in \mathscr{D}^\infty(U) \equiv \mathscr{S}(\mathbb{R})$, $\varphi \not\equiv 0$. Suppose that the moment sequence $(f_{U,\varphi}((iy)^n) \equiv \langle q^n\varphi, \varphi\rangle;\ n \in \mathbb{N}_0)$ is determinate. Then a trace class operator ϱ on $\mathscr{H}(U)$ is a solution of the (QMP) for $f_{U,\varphi}$ and U if and only if*

ϱ is of the following form: There are complex numbers $\lambda_{kn}, k \in \mathbb{N}$ and $n \in \mathbb{I}$, and functions ψ_k, $k \in \mathbb{N}$, on $\mathbb{R}$ such that $\psi_k(t) = \lambda_{kn}\varphi(t)$ for $t \in S_n(\varphi)$, $\psi_k(t) = 0$ for $t \in \mathbb{R} \setminus S(\varphi)$, $\sum_n |\lambda_{kn}|^2 = 1$ for each $k \in \mathbb{N}$ and $\varrho = \sum_k \psi_k \otimes \psi_k$.
Moreover, each solution of the (GMP) *for $f_{U,\varphi}$ is unitarily equivalent to the couple $\{U_\varrho, \varphi_\varrho\}$ (cf. 4.3), where ϱ is of described form.*

Proof. If ψ_k and ϱ are of the above form, then obviously $\psi_k \in \mathscr{S}(\mathbb{R})$ and $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$.

We prove the necessity part. For simplicity we carry out the proof only for $\gamma = 1$. Suppose that $\varrho \in \mathbb{B}_1(\mathscr{D}^\infty(U))_+$ and $f_{U,\varphi} = f_{U,\varrho}$. We write ϱ as $\varrho = \sum_k \psi_k \otimes \psi_k$, where $\psi_k \in \mathscr{D}^\infty(U)$. We shall show that the functions ψ_k are of the form stated above.

For $n \in \mathbb{N}_0$, we have

$$\begin{aligned}\int t^n |\varphi(t)|^2 \, dt &= \langle q^n\varphi, \varphi\rangle = f_{U,\varphi}((iy)^n) = f_{U,\varrho}((iy)^n) \\ &= \sum_k \langle q^n\psi_k, \psi_k\rangle \\ &= \int t^n \left(\sum_k |\psi_k(t)|^2\right) dt.\end{aligned}$$

Since the moment sequence $(\langle q^n\varphi, \varphi\rangle;\ n \in \mathbb{N}_0)$ is assumed to be determinate, it has a unique representing measure. Hence

$$|\varphi(t)|^2 = \sum_k |\psi_k(t)|^2 \quad \text{for} \quad t \in \mathbb{R}. \tag{8}$$

In particular, $\psi_k(t) = 0$ if $t \in \mathbb{R} \setminus S(\varphi)$. Let $j \in \{1, 2\}$. We have

$$\begin{aligned}\int t^n \varphi^{(j)}(t)\,\overline{\varphi(t)}\, dt &= \langle q^n(-ip)^j\varphi, \varphi\rangle \\ &= \sum_k \langle q^n(-ip)^j\psi_k, \psi_k\rangle \\ &= \int t^n \left(\sum_k \psi_k^{(j)}(t)\overline{\psi_k(t)}\right) dt,\end{aligned}$$

i.e.

$$\int t^n \overline{\eta(t)}\, d\nu(t) = 0 \quad \text{for all} \quad n \in \mathbb{N}_0, \tag{9}$$

where $\eta(t) := (\varphi^{(j)}(t)\overline{\varphi(t)} - \sum_k \psi_k^{(j)}(t)\overline{\psi_k(t)})|\varphi(t)|^{-2}$ for $t \in M(\varphi) := \{s \in \mathbb{R} : \varphi(s) \neq 0\}$ and $\eta(t) := 0$ otherwise and $d\nu(t) := |\varphi(t)|^2\, dt$. Using (8), it follows easily that $\eta \in L^2(\mathbb{R}, \nu)$. Since the moment sequence $(\langle q^n\varphi, \varphi\rangle)$ of ν is determinate, the polynomials are dense in $L^2(\mathbb{R}, \nu)$ ([1]). Thus, by (9), $\eta = 0$. Hence

$$\varphi^{(j)}(t)\overline{\varphi(t)} = \sum_k \psi_k^{(j)}(t)\overline{\psi_k(t)} \quad \text{for} \quad t \in M(\varphi) \quad \text{and} \quad j = 1, 2. \tag{10}$$

Define $\lambda_k(t) := \psi_k(t)\varphi(t)^{-1}$ for $t \in M(\varphi)$, $k \in \mathbb{N}$. By (8), we have

$$\sum_k |\lambda_k(t)|^2 = 1 \quad \text{for} \quad t \in M(\varphi). \tag{11}$$

From (10) applied with $j = 1$ and (11) we obtain that

$$\sum_k \lambda_k'(t)\overline{\lambda_k(t)} = 0 \quad \text{for} \quad t \in M(\varphi). \tag{12}$$

Recall that λ_k is C^∞ on $M(\varphi)$, since φ and ψ_k are in $S(\mathbb{R})$. From the properties of the trace class operator ϱ it is wellknown that the kernel $K(t,s) := \sum_k \psi_k(t)\overline{\psi_k(s)}$ is in $\mathscr{S}(\mathbb{R}^2)$ and that the series converges in the topology of $\mathscr{S}(\mathbb{R}^2)$. Hence K can be differentiated by interchanging summation and differentiation. Therefore, differentiating (8) twice and using (10) with $j = 2$ yields

$$|\varphi'(t)|^2 = \sum_k |\psi_k'(t)|^2, \quad t \in M(\varphi). \tag{13}$$

We have

$$\begin{aligned} |\psi_k'(t)|^2 &= |(\lambda_k\psi)'(t)|^2 \\ &= |\lambda_k'(t)|^2\,|\varphi(t)|^2 + |\lambda_k(t)|^2\,|\varphi'(t)|^2 + 2\operatorname{Re}\lambda_k'(t)\overline{\lambda_k(t)}\varphi(t)\overline{\varphi'(t)}, \quad t \in M(\varphi). \end{aligned}$$

From this identity combined with (13), (11) and (12) we conclude that

$$\sum_k |\lambda_k'(t)|^2 = 0, \quad t \in M(\varphi). \tag{14}$$

Therefore, by (14), every function $\lambda_k(t)$ is constant on each open interval contained in $M(\varphi)$.

To complete the proof, we define $\lambda_k(t)$ on the whole of $S(\varphi)$. Let $t_0 \in S(\varphi)$. Then there exists a number $r \in \mathbb{N}_0$ such that $\varphi^{(r)}(t_0) \neq 0$. From the mean value theorem it follows that there is an $\varepsilon > 0$ such that $J_- := (t_0 - \varepsilon, t_0) \subseteq M(\varphi)$ and $J_+ := (t_0, t_0 + \varepsilon) \subseteq M(\varphi)$. Recall that $\psi_k = \lambda_k\varphi$ on $J_- \cup J_+$ and that λ_k is constant on J_- and on J_+. Therefore, since $\varphi^{(r)}(t_0) \neq 0$ and $\psi_k \in \mathscr{S}(\mathbb{R})$, we have $\lambda_k(t_0 - 0) = \lambda_k(t_0 + 0)$. If $t_0 \notin M(\varphi)$, we define $\lambda_k(t_0)$ by this value. Then, by the preceding $\lambda_k(t)$ is locally constant on $S(\varphi)$. Letting λ_{kn} be the value on the connected component $S_n(\varphi)$ of $S(\varphi)$, the proof of the necessity is complete.

The last assertion follows by a similar reasoning as used in the first part of the proof of Lemma 5.2.2. Recall that in the present case $\gamma(V) = \gamma(U) \neq 0$ for $V \in \widehat{G}$ implies $V \cong U$. □

The next three corollaries follow at once from Theorem 5.3.1 and from Lemma 4.3.1. Throughout these corollaries we retain the assumptions of Theorem 5.3.1.

5.3.2 Corollary. *For $\psi \in \mathscr{S}(\mathbb{R})$, we have $f_{U,\varphi} = f_{U,\psi}$ if and only if there are complex numbers $\lambda_n, n \in \mathbb{I}$, such that $|\lambda_n| = 1$, $\psi(t) = \lambda_n\varphi(t)$ for $t \in S_n(\varphi)$, $n \in \mathbb{I}$ and $\psi(t) = 0$ for $t \in \mathbb{R} \setminus S(\varphi)$.*

5.3.3 Corollary. *The number of connected components of the set $\mathscr{S}(\varphi)$ is equal to the supremum of dim range ϱ, taken over all solutions ϱ of the (QMP) for $f_{U,\varphi}$ and U.*

5.3.4 Corollary. *The following conditions are equivalent:*

(i) $f_{U,\varphi}$ *is* (GMP)*-determinate.*
(ii) $f_{U,\varphi}$ *is* (QMP)*-determinate.*
(iii) $f_{U,\varphi}$ *is* (VS)*-determinate.*
(iv) $S(\varphi)$ *is connected.*

The preceding corollary contains an answer to problem 1). From the parametrization of $\widehat{G}$ it follows at once that the answer to 3) is always affirmative. Now we turn to problem 2). For technical reasons we prefer to consider the vector states on the algebra $\mathscr{A}(p,q) \equiv dU(\mathscr{E}(\mathscr{g}))$ rather than on the enveloping algebra. For $\varphi \in \mathscr{S}(\mathbb{R})$, let f_φ denote the vector state $\langle .\varphi, \varphi\rangle$ on $\mathscr{A}(p,q)$. The next theorem gives an abstract characterization of the vector states f_φ with φ of compact support. For this we need some terminology.

Let h_1 and h_2 be positive linear functionals on a $*$-algebra A. We say that h_1 and h_2 are *orthogonal* if the relations $h(a^+a) \leqq h_1(a^+a)$ and $h(a^+a) \leqq h_2(a^+a)$, $a \in A$, for a positive linear functional h on A always imply that $h \equiv 0$ (cf. [24], Definition 8.6.13). Now let f_1 and f_2 be positive linear functionals on $\mathscr{A}(p,q)$. We shall say that f_1 and f_2 are *q-orthogonal* and write $f_1 \perp_q f_2$ if for arbitrary numbers $k,l \in \mathbb{N}_0$ the positive linear functionals $f_1(p^k \cdot p^k) \restriction \mathbb{C}[q]$ and $f_2(p^l \cdot p^l) \restriction \mathbb{C}[q]$ on the polynomial algebra $\mathbb{C}[q]$ are orthogonal.

5.3.5 Theorem. *Let f be a positive linear functional on the $*$-algebra $\mathscr{A}(p,q)$. Then $f = f_\varphi$ with $\varphi \in C_0^\infty(\mathbb{R})$ if and only if the following two conditions are fulfilled:*

(i) *f can be written as a (possibly infinite) sum $f = \sum_{n\in\mathbb{I}} f_n$ of pure positive linear functionals f_n, $n \in \mathbb{I}$, on $\mathscr{A}(p,q)$ such that $f_n \perp_q f_m$ for $n,m \in \mathbb{I}$, $n \neq m$.*
(ii) *There is a positive number M such that $f(q^{2n}) \leqq M^{2n}$ for $n \in \mathbb{N}_0$.*

First we prove a preliminary lemma.

5.3.6 Lemma. *For $\eta, \zeta \in C_0^\infty(\mathbb{R})$, we have $f_\eta \perp_q f_\zeta$ if and only if $S(\eta) \cap S(\zeta) = \emptyset$.*

Proof. First suppose $f_\eta \perp_q f_\zeta$. We want to show that $S(\eta) \cap S(\zeta) = \emptyset$. Assume the contrary. Then there exists a $t_0 \in S(\eta) \cap S(\zeta)$. There are numbers $k,l \in \mathbb{N}_0$, $\varepsilon > 0$ and $\delta > 0$ such that $|\eta^{(k)}(t)| \geqq \delta$, $|\zeta^l(t)| \geqq \delta$ for $t \in J := (t_0 - \varepsilon, t_0 + \varepsilon)$. Take a function $\psi \in C_0^\infty(\mathbb{R})$ such that $\psi \not\equiv 0$, supp $\psi \subseteq J$ and $|\psi(t)| \leqq \delta$, $t \in J$. Then, f_ψ is obviously a positive linear functional on the polynomial algebra $\mathbb{C}[q]$ satisfying $f_\psi(a^+a) \leqq f_\eta(p^k a^+ a p^k)$ and $f_\psi(a^+a) \leqq f_\zeta(p^l a^+ a p^l)$ for $a \in \mathbb{C}[q]$. Since $f_\eta \perp_q f_\zeta$, this implies $f_\psi = 0$ which contradicts $\psi \not\equiv 0$.

Conversely, assume that $S(\eta) \cap S(\zeta) = \emptyset$. Let $k,l \in \mathbb{N}_0$. We have to show that the positive linear functionals $f_\eta(p^k \cdot p^k)$ and $f_\zeta(p^l \cdot p^l)$ on the polynomial algebra $\mathbb{C}[q]$ are orthogonal. For this let f be a positive linear functional on $\mathbb{C}[q]$

such that $f(a^+a) \leqq f_\eta(p^k a^+ a p^k)$ and $f(a^+a) \leqq f_\zeta(p^l a^+ a p^l)$ for $a \in \mathbb{C}[q]$. Then we have $|f(a)|^2 \leqq f(1)f(a^+a) \leqq f(1)f_\eta(p^k a^+ a p^k) = f(1)\int |a(t)|^2 |\eta^{(k)}(t)|^2\, dt$, i.e. f is continuous relative to the norm of $L^2(\mathbb{R}, \mu)$, where $d\mu := |\eta^{(k)}(t)|^2\, dt$. Hence there exists a function $\varphi \in L^2(\mathbb{R}, \mu)$ such that

$$f(a) = \int a(t)\varphi(t)\, |\eta^{(k)}(t)|^2\, dt, \quad a \in \mathbb{C}[q]. \tag{15}$$

Similarly, there is a function $\psi \in L^2(\mathbb{R}, \nu)$ such that

$$f(a) = \int a(t)\psi(t)\, |\varphi^{(l)}(t)|^2\, dt, \quad a \in \mathbb{C}[q], \tag{16}$$

where $d\nu := |\zeta^{(l)}(t)|^2\, dt$. Since η and ζ are in $C_0^\infty(\mathbb{R})$, there is an $\alpha > 0$ such that supp $\eta \cup$ supp $\zeta \subseteq (-\alpha, \alpha)$. From (15) and (16),

$$\int_{-\alpha}^{\alpha} a(t)\Big\{\varphi(t)\, |\eta^{(k)}(t)|^2 - \psi(t)\, |\varphi^{(l)}(t)|^2\Big\}\, dt = 0 \quad \text{for all} \quad a \in \mathbb{C}[q].$$

The function $\vartheta(t) := \varphi(t)|\eta^{(k)}(t)|^2 - \psi(t)|\zeta^{(l)}(t)|^2$ is in $L^2([-\alpha, \alpha])$, hence $\vartheta = 0$ in $L^2([-\alpha, \alpha])$. Thus $\varphi(t)|\eta^{(k)}(t)|^2 = \psi(t)|\zeta^{(l)}(t)|^2$ a.e. on $[-\alpha, \alpha]$ relative to the Lebesgue measure. Since $S(\eta) \cap S(\zeta) = \emptyset$ by assumption, the sets $\{t : \eta^{(k)}(t) \neq 0\}$ and $\{t : \zeta^{(l)}(t) \neq 0\}$ are disjoint. Therefore, the preceding equality implies that $\varphi = 0$ in $L^2(\mathbb{R}, \mu)$. Thus $f \equiv 0$ and the proof is complete. □

Proof of Theorem 5.3.5

First suppose that $f = f_\varphi$ with $\varphi \in C_0^\infty(\mathbb{R})$. (ii) is trivial. We prove (i). Let χ_m denote the characteristic function of the connected component $S_m(\varphi)$ of $S(\varphi)$. From the definition of $S(\varphi)$ it is clear that $\varphi_m := \chi_m \varphi \in C_0^\infty(\mathbb{R})$. Define $f_m := f_{\varphi_m}$, $m \in \mathbb{I}$. Obviously, $f = f_\varphi = \sum_{m \in \mathbb{I}} f_m$. For $n \neq m$, $S(\varphi_n) \cap S(\varphi_m) = \emptyset$ by the definition of φ_n and φ_m, so that $f_n \perp_q f_m$ by Lemma 5.3.6. We show that f_m is pure. For this we argue similarly as in [24], Example 8.6.15.

Let B be an operator from the weak commutant of $\pi_{f_m}(\mathscr{A}(p, q))$. It is clear that the GNS representation π_{f_m} is unitarily equivalent to the closure of the representation $a \to a \restriction \mathscr{A}(p, q)\varphi_m$. For simplicity we identify π_{f_m} with the latter. By definition, $B\pi_{f_m}(q) \subseteq \pi_{f_m}(q)^* B$. Hence $B\overline{\pi_{f_m}(q)} \subseteq \pi_{f_m}(q)^* B$. Since $\overline{\pi_{f_m}(q)} = \pi_{f_m}(q)^*$ is obviously the (self-adjoint) multiplication operator by the variable t on $L^2(S_m(\varphi))$, it follows that B is also a multiplication operator by some L^∞-function ψ on $S_m(\varphi)$. Let T be the differential operator $i\frac{d}{dt}$ with domain $C_0^\infty(S_m(\varphi))$. We prove that $BT \subseteq T^* B$. For this let $\eta \in C_0^\infty(S_m(\varphi))$. Clearly, $\eta\varphi_m^{-1} \in C_0^\infty(S_m(\varphi))$. Hence there exists a sequence (r_n) of polynomials in t such that $(\eta\varphi_m^{-1})^{(k)}(t) = \lim r_n^{(k)}(t)$ uniformly on the bounded set $S_m(\varphi)$ (recall that φ has compact support) for $k = 0, 1$. This implies $\eta - r_n\varphi_m \equiv (\eta\varphi_m^{-1} - r_n)\varphi_m \to 0$ and $T(\eta - r_n\varphi_m) = i(\eta\varphi_m^{-1} - r_n)'\varphi_m + i(\eta\varphi_m^{-1} - r_n)\varphi_m' \to 0$ in $L^2(S_m(\varphi))$ as $n \to \infty$, so $r_n\varphi_m \to \eta$ and $\pi_{f_m}(p)r_n\varphi_m = Tr_n\varphi_m \to T\eta$. Thus $\eta \in \mathscr{D}(\overline{\pi_{f_m}(p)})$ and $T\eta = \pi_{f_m}(p)\eta$, hence $T \subseteq \overline{\pi_{f_m}(p)}$ and $\pi_{f_m}(p)^* \subseteq T^*$. We have

$BT \subseteq B\overline{\pi_{f_m}(p)} \subseteq \pi_{f_m}(p)^* B \subseteq T^* B$. Since T^* is the differential operator $i\frac{d}{dt}$ acting in the distribution sense, it follows that the function $\psi(t)$ is constant on the interval $S_m(\varphi)$, say $\psi(t) = \lambda, \lambda \in \mathbb{C}$. Thus $B = \lambda \cdot I$, so the weak commutant of $\pi_{f_m}(A(p,q))$ is trivial and f_m is pure [24, Coroll. 8.6.7]. Thus the proof of (i) is complete.

Conversely, assume that f fulfills conditions (i) and (ii). Let f_m be as in (i). Let U be the unique element of $\hat{G}$ satisfying $dU(iz) = -I$. Recall that $\mathscr{A}(p,q) = dU(\mathscr{E}(\mathscr{g}))$. Define a positive linear functional h_m on $\mathscr{E}(\mathscr{g})$ by $h_m(.) := f_m(dU(.))$. Since $h_m((z-i)^2) = f_m(dU(z-i)^2) = 0$ and f_m satisfies also condition (ii), h_m fulfills the assumptions of Theorem 3.4.2. Hence there exists a unitary representation V of G with cyclic vector ψ such that $h_m(.) = \langle dV(.)\psi, \psi\rangle$ and $dV(z) = i \cdot I$. Writing V as a direct integral of primary representations [15, 8.4] and arguing in the same way as in the first two paragraphs of the proof of Lemma 5.2.2, we conclude that V is unitarily equivalent to $U \otimes I$, where I is the identity representation of G in some Hilbert space. From this and the fact that $h_m(.) = \langle dV(.)\psi, \psi\rangle$ it follows that h_m can be written as $h_m(.) = \sum_k \langle dU(.)\psi_{mk}, \psi_{mk}\rangle$ with $\psi_{mk} \in \mathscr{D}^\infty(U)$. By condition (i), f_m is pure. It is easily seen that h_m is also pure. Combined with the preceding representation of h_m, it follows that $\langle dU(.)\psi_{mk}, \psi_{mk}\rangle = \alpha_{mk} h_m(.)$ with complex numbers α_{mk}. Putting this in the representation $h_m(.) = \sum_k \langle dU(.)\psi_{mk}, \psi_{mk}\rangle$, we obtain a vector $\varphi_m \in \mathscr{D}^\infty(U)$ such that $h_m(.) = \langle dU(.)\varphi_m, \varphi_m\rangle$. Thus $f_m(dU(x)) = h_m(x) = \langle dU(x)\varphi_m, \varphi_m\rangle$ for $x \in \mathscr{E}(\mathscr{g})$, i.e. f_m is the vector state f_{φ_m} on the algebra $\mathscr{A}(p,q)$.

For $k \in \mathbb{N}_0$, we have $f_m(q^{2k}) \leqq f(q^{2k}) \leqq M^{2k}$ and

$$|f_m(q^{2k+1})|^2 \leqq f_m(q^2) f_m(q^{4k}) \leqq f(q^2) f(q^{4k}) \leqq M^2 M^{4k} = (M^{2k+1})^2,$$

i.e. $|f_m(q^n)| \leqq M^n$ for $n \in \mathbb{N}_0$. Therefore, from the theory of moments it is well-known that the unique positive regular Borel measure with moments $f_m(q^n)$, $n \in \mathbb{N}_0$, is supported on the interval $[-M, M]$. Thus supp $\varphi_m \subseteq [-M, M]$ and $\varphi_m \in C_0^\infty(\mathbb{R})$. Now suppose that $n, m \in \mathbb{I}$, $n \neq m$. Since $f_n \equiv f_{\varphi_n} \perp_q f_{\varphi_m} \equiv f_m$ by condition (i) it follows from Lemma 5.3.6 that $S(\varphi_n) \cap S(\varphi_m) = \emptyset$. Hence $\varphi(t) := \sum_{m \in \mathbb{I}} \varphi_m(t)$ is a well-defined function of $C_0^\infty(\mathbb{R})$ and

$$f = \sum_{m \in \mathbb{I}} f_m = \sum_{m \in \mathbb{I}} f_{\varphi_m} = f_\varphi. \quad \square$$

Remarks

Actually, the preceding proof gives the following assertion: A positive linear functional f on $\mathscr{A}(p,q)$ is of the form $f = f_\varphi$ with $\varphi \in \mathscr{S}(\mathbb{R})$ provided that f satisfies conditions (i) and

(ii)′ For each $m \in \mathbb{I}$, there is a positive number M_m such that

$$f_m(q^{2n})^{1/2} \leqq M_m^n n! \quad \text{for all} \quad n \in \mathbb{N}_0.$$

References

1. Akhiezer, N.I.: The classical moment problem and some related questions in analysis. Edinburgh: Oliver and Boyd 1965
2. Berg, C., Christensen, J.P.R., Ressel, P.: Harmonic analysis on semigroups. Theory of positive definite and related functions. Berlin Heidelberg New York: Springer 1984
3. Dixmier, J.: C^*-Algebras. Amsterdam: North-Holland 1977
4. Dixmier, J.: Enveloping algebras. Amsterdam: North-Holland 1976
5. Fell, J.M.G.: The dual spaces of C^*-algebras. Trans. Am. Math. Soc. **94,** 365–403 (1960)
6. Fell, J.M.G.: Weak containment and induced representations of groups. Canad. J. Math. **14,** 237–268 (1962)
7. Flato, M., Simon, J., Snellman, H., Sternheimer, D.: Simple facts about analytic vectors and integrability. Ann. Sci. l'Ecole Norm. Sup. **5,** 423–434 (1972)
8. Fuglede, B.: The multidimensional moment problem. Expo. Math. **1,** 47–65 (1983)
9. Goodman, F., Jorgensen, P.E.T.: Lie algebras of unbounded derivations. J. Funct. Anal. **52,** 369–384 (1983)
10. Harish-Chandra: Representations of a semisimple Lie group on a Banach space I. Trans. Am. Math. Soc. **75,** 185–243 (1953)
11. Haviland, E.K.: On the momentum problem for distributions in more than on dimension, I, II. Am. J. Math. **57,** 562–568 (1935) and **58,** 164–168 (1936)
12. Jorgensen, P.E.T.: Representations of differential operators on a Lie group, and conditions for a Lie algebra of operators to generate a representation of the group. J. d'Analyse Math. **43,** 251–288 (1983/84)
13. Jorgensen, P.E.T., Muhly, P.S.: Self-adjoint extensions satisfying the Weyl operator commutation relations. J. d'Analyse Math. **37,** 46–99 (1980)
14. Kirillov, A.A.: Unitary representations of nilpotent Lie groups. Russ. Math. Surv. **17,** 53–104 (1962)
15. Kirillov, A.A.: Elements of the theory of representations. Berlin Heidelberg New York: Springer 1976
16. Landau, H.J.: Moments in mathematics. Proc. Appl. Math. No. 37, Am. Math. Soc., Providence 1987
17. Nelson, E.: Analytic vectors. Ann. Math. **70,** 572–615 (1959)
18. Nelson, E., Stinespring, W.F.: Representation of elliptic operators in an enveloping algebra. Am. J. Math. **81,** 547–560 (1959)
19. Powers, R.T.: Self-adjoint algebras of unbounded operators II. Trans. Am. Math. Soc. **187,** 261–293 (1974)
20. Richter, P.: Zur Extremalzerlegung von Zuständen. Wiss. Z. KMU Leipzig, Math.-Naturw. R. **27,** 293–297 (1978)
21. Samoilenko, J.S.: Spectral theory of families of self-adjoint operators. (Russian) Naukova dumka, Kiev 1984
22. Schmüdgen, K.: Positive cones in enveloping algebras. Reports Math. Phys. **14,** 385–404 (1978)
23. Schmüdgen, K.: On the Heisenberg commutation relation I. J. Funct. Anal. **50,** 8–49 (1983)
24. Schmüdgen, K.: Unbounded operator algebras and representation theory. Berlin: Akademie-Verlag and Basel: Birkhäuser Verlag 1990
25. Sherman, T.: Positive linear functionals on *-algebras of unbounded operators. J. Math. Anal. Appl. **22,** 285–318 (1968)
26. Taylor, M.E.: Noncommutative harmonic analysis. Math. Surv. Monogr. No. 22, Am. Math. So., Providence 1986
27. Vershik, A.M., Karpushev, S.I.: Cohomology of groups in unitary representations, a neighbourhood of the identity and conditionally positive definite functions. (Russian) Mat. Sb. **119,** 521–533 (1982)
28. Warner, G.: Harmonic analysis on semi-simple Lie groups I, II. Berlin Heidelberg New York: Springer 1972

29. Werner, R.F.: Dilations of symmetric operators shifted by a unitary group. J. Funct. Anal. **92**, 166–176 (1990)
30. Woronowicz, S.L.: The quantum moment problem I, II. Reports Math. Phys. **1**, 135–145 (1970) and **1**, 175–183 (1970)
31. Jorgensen, P.E.T., Powers, R.T.: Positive elements in the algebra of the quantum moment problem. (Preprint) 1990

Math. Z. 206, 651–658 (1991)

Mathematische Zeitschrift

© Springer-Verlag 1991

Affine spheres with constant affine sectional curvature*

Luc Vrancken[1,**], **An-Min Li**[2,3,***] **and Udo Simon**[2,****]

[1] Departement Wiskunde, K.U. Leuven, Celestijnenlaan 200B, B-3030 Leuven, Belgium
[2] Fachbereich Mathematik, Technische Universität, Sekretariat MA 8-3,
Strasse des 17. Juni 135, 1000 Berlin 12, Federal Republic of Germany
[3] Department of Mathematics, Sichuan University, Chengdu, Sichuan, P.R. China

Received April 18, 1990; in final form August 9, 1990

1 Introduction

In this paper, we study locally strongly convex affine hypersurfaces in $\mathbb{R}^{n+1}$. It is well known that there exists a canonical transversal vector field called the affine normal vector field. The second fundamental form associated with this normal vector field is called the affine metric. These immersions were first studied by W. Blaschke and his coworkers ([B]) around 1920. The modern approach which is used in this paper can be found in [N].

M is called a proper (resp. improper) affine hypersphere if all the affine normals pass through a fixed point (resp. if all the affine normals are parallel). The affine hyperspheres form a very important class of affine hypersurfaces. Unfortunately, there is not yet a complete local classification of affine hyperspheres, not even in dimension 2.

In $\mathbb{R}^3$, affine spheres with constant curvature metric have been studied amongst others by J. Radon, A.M. Li, G. Penn, M. Magid, P. Ryan and U. Simon ([R], [L-P], [M-R]$_1$, [S]). They obtain a complete classification of the affine surfaces which are affine spheres and have constant curvature metric.

In higher dimensions locally strongly convex affine hyperspheres have been studied amongst others by A.M. Li and J.H. Yu. They obtained the following results.

Theorem ([L]). *Let M be a locally strongly convex proper affine hypersphere with $H<0$ and $\kappa=0$. Then M is equiaffinely equivalent to the hypersurface $x_1 x_2 \ldots x_{n+1}=1$.*

Here H denotes the affine mean curvature of M and κ denotes the scalar curvature of the affine metric.

---

* This paper gives a shorter proof of the Main Theorem which originally was proved in a preprint of the first author

** Research Assistant of the National Fund for Scientific Research Belgium
*** Supported by the Alexander von Humboldt Stiftung
**** Partially Supported by the GADGET program of the European Community

Theorem ([Y]). *Let M^3 be a locally strongly convex affine hypersphere in $\mathbb{R}^4$ with constant sectional curvature. Then M is an open part of a quadric or M is affinely equivalent with $x_1 x_2 x_3 x_4 = 1$*

A similar theorem for not locally strongly convex affine hyperspheres in $\mathbb{R}^4$ has been proved recently by M. Magid and P. Ryan in [M-R]$_2$. In this paper, we generalise the result of J.H. Yu to arbitrary dimensions by proving the following theorem.

Main theorem. *Let M^n be a locally strongly convex affine hypersphere in $\mathbb{R}^{n+1}$ with constant sectional curvature. Then M is an open part of a quadric or M is affinely equivalent with $x_1 x_2 \ldots x_{n+1} = 1$.*

2 Preliminaries

Let f: $M^n \to \mathbb{R}^{n+1}$ be an immersion of a connected differentiable manifold of dimension n into the affine space $\mathbb{R}^{n+1}$ equipped with its usual flat connection D and a parallel volume element ω and let ξ be an arbitrary local field of transversal vectors to $f(M^n)$. For any vector fields X and Y on M^n, we write

$$D_X f_*(Y) = f_*(\nabla_X Y) + h(X, Y)\xi, \tag{2.1}$$

thus defining an affine connection ∇ and a symmetric tensor of type (0, 2). We call h the second fundamental form. We can also define a volume element θ on M by

$$\theta(X_1, X_2, \ldots, X_n) = \omega(f_* X_1, \ldots, f_* X_n, \xi), \tag{2.2}$$

for any tangent vectors $X_1, \ldots, X_n$ to M^n.

If (M, ∇) is an affine manifold, we call an immersion f an affine immersion if there exists locally a transversal vector field such that (2.1) holds. If (∇, θ) is an equiaffine structure on M, i.e. if ∇ and θ satisfy $\nabla\theta = 0$, we call an affine immersion equiaffine if θ satisfies (2.2).

We say that f is nondegenerate if h is nondegenerate (and this condition is independent of the choice of ξ). In this case, it is known (see [N], [N-P]) that there is a unique choice of ξ such that the corresponding induced connection ∇, the nondegenerate metric h, and the induced volume element θ satisfy the following conditions:

(i) $\nabla\theta = 0$, thus (∇, θ) is an equiaffine structure on M^m;

(ii) $\theta = \omega_h$ (volume element given by h).

We call ∇ the induced connection and h the affine metric. We will call such an immersion a Blaschke immersion. In this paper, we will always consider Blaschke immersions. If h is positive (or negative) definite, the immersion is called locally strongly convex. Notice that if h is negative definite, we can always replace ξ by $-\xi$, thus making the new affine metric positive definite. Therefore, from now on, we will assume that M is locally strongly convex and hence that the affine metric h is positive definite. Condition (i) implies that $D_X \xi$ is

tangent to $f(M^n)$ for any tangent vector X to M^n. We define a tensor field S of type (1, 1) on M^n, called the shape operator, by

$$D_X\xi = -f_*(SX). \tag{2.3}$$

We have the following fundamental equations which relate the curvature tensor of ∇, the affine metric and the shape operator.

$$R(X, Y)Z = h(Y, Z)SX - h(X, Z)SY \quad \text{(Equation of Gauss)} \tag{2.4}$$

$$(\nabla h)(X, Y, Z) = (\nabla h)(Y, X, Z) \quad \text{(Equation of Codazzi for } h\text{)} \tag{2.5}$$

$$(\nabla_X S)Y = (\nabla_Y S)X \quad \text{(Equation of Codazzi for } S\text{)} \tag{2.6}$$

$$h(X, SY) = h(SX, Y) \quad \text{(Equation of Ricci)} \tag{2.7}$$

$$\nabla\omega_h = 0. \tag{2.8}$$

M is called an affine sphere if $S = \lambda I$. Since $\dim(M) \geqq 2$, if M an affine sphere, it then follows from (2.6) that $\lambda = H$ is constant and H is the affine mean curvature. If $H \neq 0$, we say that M is a proper affine sphere and if $H = 0$, we call M an improper affine sphere. From (2.5) it follows that the cubic form $C(X, Y, Z) = (\nabla h)(X, Y, Z)$ is symmetric in X, Y and Z. The Theorem of Berwald states that C vanishes identically if and only if M is an open part of a nondegenerate quadric.

Let $\hat{\nabla}$ denote the Levi Civita connection of the affine metric h. The difference tensor K is defined by

$$K(X, Y) = \nabla_X Y - \hat{\nabla}_X Y, \tag{2.9}$$

for vector fields X and Y on M. Notice that K is symmetric in X and Y. We also write $K_X(Y) = K(X, Y)$. From [N], we have that

$$h(K_X Y, Z) = -\tfrac{1}{2} C(X, Y, Z), \tag{2.10}$$

$$\hat{R}(X, Y)Z = \tfrac{1}{2}(h(Y, Z)SX - h(X, Z)SY + h(SY, Z)X - h(SX, Z)Y) - [K_X, K_Y]Z, \tag{2.11}$$

where $\hat{R}$ denotes the curvature tensor of $\hat{\nabla}$. Notice also that the apolarity condition implies that trace $K_X = 0$ for every tangent vector X. In the special case that $S = HI$, i.e. M is an affine sphere, this equation becomes

$$\hat{R}(X, Y)Z = H(h(Y, Z)X - h(X, Z)Y) - [K_X, K_Y]Z \tag{2.12}$$

Further if M is an affine sphere, we have from [N] that

$$(\hat{\nabla}_Y K)(X, Z) = (\hat{\nabla}_X K)(Y, Z), \tag{3.13}$$

where $(\hat{\nabla}_Y K)(X, Z) = \hat{\nabla}_Y(K(X, Z)) - K(\hat{\nabla}_Y X, Z) - K(X, \hat{\nabla}_Y Z)$.

Finally, we will use the following result about Riemannian geometry from [Der]. Let (M, h) be an n-dimensional Riemannian manifold. A Codazzi tensor A is a tensor field which is symmetric with respect to h and which statisfies the Codazzi equation $(\hat{\nabla}_X A)(Y) = (\hat{\nabla}_Y A)(X)$, where $\hat{\nabla}$ is the Levi Civita connec-

tion of h. Let $p \in M$ and let A be a Codazzi tensor defined on a neighbourhood U of p. Suppose that on U, we have the following differential decomposition, $TU = V_{\lambda_0} \oplus V_{\lambda_1}$, where λ_0, λ_1 are constant eigenvalues of A on U and the eigendistributions satisfy $\dim(V_{\lambda_0}) = 1$ and $\dim(V_{\lambda_1}) = n-1$. Then Lemma 1 of [Der] implies that

$$\hat{\nabla}_X X = 0, \quad \text{for } X \in V_{\lambda_0} \text{ with } h(X, X) = 1,$$
$$\hat{\nabla}_Y Z \in V_{\lambda_1} \quad \text{for } Y, Z \in V_{\lambda_1}.$$

Hence the de Rham decomposition theorem implies that M is a product. So, we have the following lemma.

Lemma 2.1 *Let (M, h) be an n-dimensional Riemannian manifold with constant sectional curvature c. Suppose that M admits a Codazzi tensor as defined above. Then M is flat, i.e. $c = 0$.*

3 Proof of the theorem

Throughout this section, we will assume that M is a locally strongly convex affine hypersphere with constant sectional curvature c. Hence the curvature tensor $\hat{R}$ is given by

$$\hat{R}(X, Y)Z = c(h(Y, Z)X - h(X, Z)Y).$$

Then, (2.11) becomes

$$h(K_X W, K_Y Z) - h(K_Y W, K_X Z) = (H - c)(h(Y, Z)h(X, W) - h(X, Z)h(Y, W)). \tag{3.1}$$

The affine theorema egregium states $c = J + H$ where J is the Pick invariant. Using the apolarity condition, we obtain from this and (3.1) that

$$h(K, K) = n(n-1)J.$$

Hence, since f is convex and J is constant, this implies that if K vanishes at one point of M, then K vanishes identically on M. Hence M is a quadric. Furthermore if M is not an open part of a quadric, it follows, $J > 0$. From now on, we will assume that M is not a quadric. Hence for every point p of M, we know that K does not vanish identically at p.

Let $p \in M$. Then, we choose a basis of $T_p M$ in the following way. First, we define $UM_p = \{v \in T_p M \mid h(v, v) = 1\}$. Notice that UM_p is compact. We define a function F_p on UM_p by

$$F_p(v) = h(K(v, v), v).$$

Since UM_p is compact, we know that there exists a vector u in which the function F attains its maximum value and hence u is a relative maximum of F. Then, we will use the following lemma.

Lemma 3.1 *Let g be a positive definite metric on $\mathbb{R}^k$ and let T: $\mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ be a symmetric multilinear mapping. We denote by $S^{k-1} = \{v \in \mathbb{R}^k \mid g(v, v) = 1\}$ and define a function f on S^{k-1} by $F(v) = T(v, v, v)$. Let u be a vector at which the*

function F on S^{k-1} attains an extremal value and let $w \in S^{k-1}$ such that $g(u, w)=0$. Then $T(u, u, w)=0$. Moreover, if F attains a relative maximum in u, then we have also that $T(u, u, u)-2T(u, w, w) \geqq 0$.

Proof. We may assume that $w \in S^{k-1}$ such that $g(u, w)=0$. Then we define a real function $\tilde{F}$ by $\tilde{F}(t)=F(\cos(t)u+\sin(t)w)$. Since F attains an extremal value at $t=0$, we have that $\tilde{F}(0)=0$ and if F attains a relative maximum at u, we also have that $F''(0) \leqq 0$. Computing these derivatives, using the symmetry properties of T we find that $T(u, u, w)=0$ and $T(u, u, u)-2T(u, w, w) \geqq 0$. □

Lemma 3.2 *Let M be a locally strongly convex affine hypersphere with constant sectional curvature c. We assume also that M is not an open part of a quadric. Then $H<c$. Furthermore let $p \in M$. Then there exist a neighbourhood U of p and local vector fields $E_1, E_2, \ldots, E_n$, orthonormal with respect to the affine metric and defined on U, such that*

$$K_{E_1} E_1=(n-1)\lambda_1 E_1$$

$$K_{E_1} E_i=K_{Ei} E_1=-\lambda_1 E_i,$$

where $\lambda_1>0$ and $\lambda_1^2=\dfrac{J}{n}$.

Proof. Since M is not an open part of a quadric, we know by the remark made at the beginning of this section that there is no point of M at which K is identically zero. We put $J=a^2>0$. We call $s \in \mathbb{R}$ a critical value if there exist a point q of M and a vector $v \in UM_q$ such that f_q attains a relative extremum in v and $f_q(v)=s$. We denote by A the set of all critical values. First, we will show that A is finite.

So let $q \in M$ and let v be a vector in which the function f_q attains a relative extremum. Because of (2.10) and the equation of Codazzi, it is clear that we can apply Lemma 3.1. So we have that v is an eigenvector of K_v, say with eigenvalue μ_1. We put $e_1=v$. Then because of (2.10), we know that there exist vectors $e_2, \ldots, e_n$ and numbers $\mu_2, \ldots, \mu_n$ such that

$$h(e_i, e_j)=\delta_{ij},$$

$$K_{e_1} e_j=\mu_j e_j,$$

$i, j=1, \ldots, n$. Let $k=2, \ldots, n$. Applying (3.1) for $X=e_1$, $Y=e_k$, $Z=e_1$, $W=e_k$ then gives

$$h(K_{e_1} e_k, K_{e_k} e_1)-h(K_{e_k} e_k, K_{e_1} e_1)=a^2.$$

Using the properties of K and the choice of our basis, this becomes

$$\mu_k^2-\mu_k \mu_1-a^2=0. \tag{3.2}$$

Hence, there exists an integer r, $0 \leqq r \leqq (n-1)$ such that if necessary after renumbering the basis

$$K_{e_1} e_1=\mu_1 e_1,$$

$$K_{e_1} e_k=\tfrac{1}{2}(\mu_1+\sqrt{\mu_1^2+4a^2})e_k, \quad 1<k \leqq (r+1)$$

$$K_{e_1} e_k=\tfrac{1}{2}(\mu_1-\sqrt{\mu_1^2+4a^2})e_k, \quad (r+1)<k \leqq n. \tag{3.3}$$

Since trace $K_{e_1}=0$, we therefore find that

$$(f_q(v))^2=\mu_1^2=\frac{(n-2r-1)^2}{(n-r)(r+1)}a^2. \tag{3.4}$$

Since a is constant on M and since r is a natural number satisfying $0\leqq r\leqq(n-1)$, it immediately follows that A is finite.

Since UM_q is compact, we know there exists a vector v in which the function f_q attains an absolute maximum. Since K is not identically zero at q, we must have $f_q(v)>0$. From Lemma 3.1 it follows that in this case $r=0$ and we see that $v\in UM_q$ is an absolute maximum of f_q if and only if

$$h(K_v v, v)=(n-1)\lambda_1,$$

where $\lambda_1>0$ and $\lambda_1^2=\frac{a^2}{n}$, where $a^2=J$. Since $r=0$, we also deduce in this case from (3.3) and (3.4) that $K_v u=-\lambda_1 u$ for any u which is orthogonal to v.

Next, let $p\in M$, we will show that if f_p attains an absolute maximum in v, we can extend v differentiably to a vector field V on a neighbourhood U of p such that $h(V, V)=1$ and at every point q of U, f_q attains an absolute maximum in $V(q)$. In order to do this, let $U_1, \ldots, U_n$ be an arbitrary differentiable h-orthonormal basis defined on a neighbourhood U of p such that $U_1(p)=v$. Then, we define a function γ by

$$\gamma: \mathbb{R}^n\times U\to\mathbb{R}^n: (a_1, \ldots, a_n, q)\mapsto\gamma(a_1, \ldots, a_q, q)=(b_1, \ldots, b_n),$$

where for $k=1, 2, \ldots, n$,

$$b_k=\sum_{i,j=1}^{n} a_i a_j h(K_{U_i}U_j, U_k)-(n-1)\lambda_1 a_k.$$

Using the fact that f_p attains an absolute maximum in $U_1(p)$, we then obtain that

$$\begin{aligned}\frac{\partial b_k}{\partial a_m}(1, 0, \ldots, 0, p)&=2h(K_{U_1(p)}U_m(p), U_k(p))-(n-1)\lambda_1\delta_{km}\\&=0 \quad \text{if } k\neq m\\&=(n-1)\lambda_1 \quad \text{if } k=m=1,\\&=-(n+1)\lambda_1 \quad \text{if } k=m\neq 1.\end{aligned}$$

Hence the implicit function theorem shows that there exist differentiable functions $a_1, \ldots, a_n$ defined on a neighbourhood U' of p such that the local vector field V defined by

$$V=a_1 U_1+a_2 U_2+\ldots+a_n U_n$$

satisfies $V(p)=v$ and $K_V V=(n-1)\lambda_1 V$. Let us define $\|V\|^2=h(V, V)$. Then $\|V\|$ is not zero in a neighbourhood U'' of p. Furthermore the vectorfield W defined by $W=V/\|V\|$ satisfies

$$K_W W=(((n-1)\lambda_1)/\|V\|)W.$$

Hence at every point q of U'', f_q attains a relative extremum in $W(q)$. However since the set of extremal values is finite and $((n-1)\lambda_1)/\|V\|$ changes continuously, we obtain that

$$\|V(q)\| = \|V(p)\| = 1,$$

for $q \in U''$. Then taking $E_1 = V$ and $E_2, \ldots, E_n$ locally orthonormal vector fields, orthogonal to E_1, with respect to h, completes the proof of this lemma because at every point q of U'', f_q attains an absolute maximum at $E_1(q)$.

Main theorem. *Let M be a locally stronly convex hypersphere in $\mathbb{R}^{n+1}$ which has constant sectional curvature with respect to the affine metric h. Then either M is an open part of a nondegenerate quadric or M is affinely equivalent with the affine hypersurface given by $x_1 x_2 x_3 \ldots x_{n+1} = 1$.*

Proof. Since M is an affine sphere, we know that $S = HI$. Let c denote the constant sectional curvature of M. If $H \geqq c$, then it follows from Lemma 3.2 that M is an open part of a nondegenerate quadric. Hence, we may assume that $c > H$. But in this case, Lemma 3.2 together with (2.13) implies that the tensor field $A = K_{E_1}$ is a Codazzi tensor field on (M, h) which satisfies the conditions of Lemma 2.1. Therefore, M is flat with respect to h. Hence $c = 0$ and thus $H < 0$. Applying then the theorem of A.M. Li gives that M is affinely equivalent with the affine hypersurface given by $x_1 x_2 \ldots x_{n+1} = 1$. □

Remark. The Main theorem can also proved directly, without using the theorem of A.M. Li. Using similar techniques as in Lemma 3.2, it is easy to show that one can construct a local orthonormal (with respect to h) basis $\{E_1, E_2, \ldots, E_n\}$ which satisfies

$$\begin{aligned}
&h(K_{E_i} E_j, E_k) = 0 && \text{for mutually different } i, j, k,\\
&h(K_{E_i} E_j, E_j) = 0 && \text{for } j < i,\\
&h(K_{E_i} E_j, E_j) = -\lambda_i && \text{for } j > i,\\
&h(K_{E_i} E_i, E_i) = (n-i)\lambda_i,
\end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ are positive numbers defined inductively by $\lambda_1^2 = \frac{J}{n}$, $\lambda_j^2 = \left(J + \sum_{m=1}^{j-1} \lambda_m^2\right) \Big/ (n-j+1)$. Then, it follows from (2.13) that

$$\hat{\nabla}_{E_i} E_j = 0.$$

Hence there exist local coordinates $x_1, x_2, \ldots, x_n$ on M such that $\partial_i = \mathrm{E}_i$. Hence

$$\begin{aligned}
D_{\partial_i} \partial_j &= -\lambda_i \partial_j && j > i,\\
D_{\partial_i} \partial_i &= -\sum_{j=1}^{i-1} \lambda_j \partial_j + (n-i)\lambda_i \partial_i + \xi && i > 1,\\
D_{\partial_1} \partial_1 &= (n-1)\lambda_1 \partial_1 + \xi,\\
D_{\partial_i} \xi &= -\lambda \xi.
\end{aligned}$$

Then either by solving this system of differential equations explicitly or by using the uniqueness of its solution up to an affine transformation, we find that M is affinely equivalent with the affine hypersurface given by $x_1 x_2 x_3 \ldots x_{n+1} = 1$.

Acknowledgement. The authors would like to thank Professor K. Nomizu for helpful conversations.

References

[B] Blaschke, W.: Vorlesungen über Differentialgeometrie II. Affine Differentialgeometrie. Berlin: Springer 1923

[Der] Derdzinski, A.: Classification of Certain Compact Riemannian Manifolds with Harmonic Curvature and Non-parallel Ricci Tensor. Math. Z. **172**, 273–280 (1980)

[L] Li, A.M.: Some theorems in affine differential geometry. Acta Math. Sinica, N.S. **5**, 345–354 (1989)

[L-P] Li, A.M., Penn, G.: Uniqueness theorems in affine differential geometry II. Results in Math. **13**, 308–317 (1988)

$[M\text{-}R]_1$ Magid, M., Ryan, P.: Flat affine spheres. Geom. Dedicata (to appear)

$[M\text{-}R]_2$ Magid, M., Ryan, P.: Affine 3-Spheres with Constant Affine Curvature. (preprint)

[N] Nomizu, K.: Introduction to affine differential geometry, part I. MPI/88-37, Bonn 1988

[N-P] Nomizu, K., Pinkall, U.: On the geometry of affine immersions. Math. Z. **195**, 165–178 (1987)

[R] Radon, J.: Zur Affingeometrie der Regelflächen. Leipziger Berichte **70**, 147–155 (1918)

[S] Simon, U.: Local classification of twodimensional affine spheres with constant curvature metric. Differ. Geom. Appl. (Brno) **1** (to appear)

[Y] Yu, J.H.: Affine hyperspheres with constant sectional curvature in A^4. Sichuan University (preprint)

Math. Z. 206, 659–670 (1991)

Mathematische Zeitschrift
© Springer-Verlag 1991

Composing infinitely differentiable functions

J. Appell [1], V.I. Nazarov [2] and P.P. Zabrejko [2]

[1] Fakultät für Mathematik, Universität Würzburg, Am Hubland, W-8700 Würzburg, Federal Republic of Germany
[2] Belgosuniversitet, Matematicheskij Fakultet, pl. Lenina 1, SU-220080, Soviet Union

Received May 7, 1990; in final form June 11, 1990

It is well known, and has been often used by many authors, that the composition of two *smooth* (i.e. C^∞) *functions* is again smooth and the composition of two *analytic functions* is again analytic. In the last years, the question arose, mainly in view of applications to partial differential equations [17], whether or not the same is true for various "intermediate spaces" between smooth and analytic functions. For example, the answer is positive within the class of *quasi-analytic functions*, or functions whose derivatives satisfy certain growth conditions (see below).

This problem may be reformulated more precisely as follows: given some class $E \subset C^\infty(\Omega, X)$ ($\Omega \subseteq \mathbb{R}^n$, X a Banach space), find conditions on a function $f: X \to X$ (or, more generally, $f: \Omega \times X \to X$) under which the corresponding *superposition operator*

$$Fx(t) = f(x(t)) \qquad (t \in \Omega) \tag{1}$$

or

$$Fx(t) = f(t, x(t)) \qquad (t \in \Omega) \tag{2}$$

respectively) maps E into itself. In this general form, very little is known about this problem. The first results seem to be due to M. Gevrey ([10], see also [6, 7]) who studied this problem for special function sets which now are usually called *Gevrey classes*. Similar sets (called *Roumieu classes*) have been introduced in [26]; for results on the superposition operator in these classes see [21], for applications [22]. Various other classes of smooth functions (e.g. *Beurling classes* and classes of functions with prescribed decay in all derivatives) are dealt with, for example, in Chap. 6 of [9] and Chap. 7 of [17].

In this paper, we give a unified approach to the investigation of the *acting problem* for the operators (1) and (2) in a fairly large class of infinitely differentiable functions which embraces all classes mentioned before and which plays a fundamental role in the *theory of ultradistributions*. It turns out that this problem reduces essentially to that of describing the asymptotic behaviour of

the so called *Bell polynomials* which arise naturally in combinatorics. This allows us to give acting conditions for the operators (1) and (2) which are sufficient and (under suitable additional hypotheses on f) also necessary. Moreover, we give conditions under which the operators (1) and (2) are bounded, continuous, or compact. Such conditions are indispensable for applying the main principles of nonlinear analysis, such as *degree theory*, *bifurcation theory*, or *fixed point theory*. Finally, certain subclasses of smooth functions with rapidly decreasing derivatives are dealt with in the last section.

1 Some classes of smooth functions

Let $\mu=(M_0, M_1, \ldots, M_n, \ldots)$ $(M_0=1)$ be some increasing sequence of positive numbers, and let X be a Banach space. Given $L>0$, we denote by $R_\mu(L, X)$ the set of all infinitely differentiable functions x on $[a, b]$ with values in X for which the norm

$$\|x|R_\mu(L, X)\| = \sup_{k\geq 0} \frac{\|x^{(k)}|C\|}{L^k M_k}$$

$(C=C([a, b], X))$ is finite. The set $R_\mu(L, X)$ is a Banach space [26] and is usually called (special) *Roumieu class*.

It is not hard to see that the imbedding $R_\mu(L, X)\subseteq R_\mu(L', X)$ is continuous for $L\leqq L'$; in case $X=\mathbb{R}^N$, this imbedding is even compact [26]. The set

$$R_\mu(0, X)= \bigcap_{0<L<\infty} R_\mu(L, X), \tag{3}$$

equipped with the topology of the projective limit, is called *Beurling class*; likewise, the set

$$R_\mu(\infty, X)= \bigcup_{0<L<\infty} R_\mu(L, X), \tag{4}$$

equipped with the topology of the inductive limit, is called (general) *Roumieu class*. The space $R_\mu(0, X)$ is a Fréchet space for any Banach space X; the space $R_\mu(\infty, X)$ is a Montel space for $X=\mathbb{R}^N$ (i.e. every bounded closed set in $R_\mu(\infty, \mathbb{R}^N)$ is compact).

Now we consider some examples of sequences $\mu=(M_0, M_1, \ldots)$ which generate special Roumieu classes $R_\mu(L, X)$. The first important example is

$$M_n=(n!)^s \qquad (0<s<\infty) \tag{5}$$

which leads to the well-known *Gevrey classes* $G^s=G^s(L, X)$. For $0<s<1$, the class G^s consists of *entire functions* on $[a, b]$ of order at most $(1-s)^{-1}$ (see [10]). For $s=1$, G^s is the set of all *analytic functions* on $[a, b]$. For $s>1$, G^s contains more than analytic functions. In general, the problem is interesting whether or not G^s contains only *quasianalytic functions* (recall that a smooth function x is called quasi-analytic if the fact that x is flat at some point $t_0\in[a, b]$, i.e. $x^{(k)}(t_0)=0$ for $k=0, 1, 2, \ldots$, implies that $x\equiv 0$ on $[a, b]$).

As was already shown by A. Borel in 1895, for any real sequence $\xi_0, \xi_1, \xi_2, \ldots$ one can construct a smooth function x on $[a, b]$ such that $x^{(k)}(a) = \xi_k$ $(k=0, 1, \ldots)$. The solution of this "interpolation problem" is, of course, not unique, since there exist flat functions $x \not\equiv 0$. However, if one imposes appropriate growth conditions on the sequence $(M_0, M_1, M_2, \ldots)$, then $x \equiv 0$ is the only flat function satisfying

$$\max_{a \leqq t \leqq b} |x^{(k)}(t)| \leqq L^k M_k \qquad (k=0, 1, 2, \ldots), \tag{6}$$

i.e. the space $R_\mu(L, X)$ consists only of quasi-analytic functions. By Carleman's theorem ([6, 20]), a necessary and sufficient condition for this is

$$\sum_{n=1}^{\infty} \frac{M_n}{M_{n+1}} = \infty \quad \text{or} \quad \sum_{n=1}^{\infty} \frac{1}{\sqrt[n]{M_n}} = \infty. \tag{7}$$

Apart from (5), another interesting example of μ is

$$M_n = n^{ns} \qquad (1 < s < \infty) \tag{8}$$

which was studied e.g. in [7], and

$$M_n = H^{n^s} \qquad (1 < s < \infty) \tag{9}$$

$(H > 1$ fixed) which was studied e.g. in [9]. By (7), the space $R_\mu(L, X)$ with μ as in (8) or (9) contains also functions which are not quasi-analytic. All three examples (5), (8), and (9) will be considered in what follows.

In order to simplify the notation, we shall use the following abbreviations in the sequel for a multi-index $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_n)$ $(\alpha_j = 0, 1, 2, \ldots)$:

$$|\alpha| = \alpha_0 + \alpha_1 + \ldots + \alpha_n, \qquad \|\alpha\| = \alpha_0 + 1\alpha_1 + 2\alpha_2 + \ldots + n\alpha_n,$$
$$C_n^\alpha = n! [\alpha_0! \alpha_1! \ldots \alpha_n! (1!)^{\alpha_1} (2!)^{\alpha_2} \ldots (n!)^{\alpha_n}]^{-1}.$$

For instance, in this notation the formula for higher derivatives of a composite function $y(t) = f(t, x(t))$ becomes

$$y^{(n)}(t) = \sum_{\|\alpha\| = n} C_n^\alpha D_t^{\alpha_0} D_u^{|\alpha| - \alpha_0} f[t, x(t)] x'(t)^{\alpha_1} \ldots x^{(n)}(t)^{\alpha_n} \tag{10}$$

where D_t and D_u denote the partial derivatives of $f(t, u)$ with respect to t and u, respectively; see e.g. [8, 15, 25] or [28]. (The powers of the derivatives $x^{(k)}(t)$ are to be understood in the usual sense of multilinear forms, evaluated on the diagonal; the terms on the right-hand side of (10) are usually called Bell polynomials [24]). Formula (10) implies the well-known result that the superposition of two smooth functions is again smooth. For analytic functions, the same fact may be proved either by passing to complex domains and using the fact that the superposition of two C^1 functions is again C^1, or by applying certain majorization techniques. These techniques turn out to be useful also for studying the stability of the classes $R_\mu(L, X)$ under superpositions and will be discussed now in more detail.

2 The majorant technique

For $n=0, 1, \ldots, u>0, v>0$, and $\tau>0$, consider the sequence

$$\psi_n(u, v, \tau)=\sum_{\|\alpha\|=n} C_n^{\alpha} M_{|\alpha|} M_n^{-1} M_1^{\alpha_1} \ldots M_n^{\alpha_n} u^{\alpha_0} v^{|\alpha|-\alpha_0} \tau^{\|\alpha\|-|\alpha|+\alpha_0}, \tag{11}$$

and denote by W_μ the set of all $(u, v, \tau)\in\mathbb{R}_+^3$ for which the sequence (11) is bounded:

$$W_\mu=\{(u, v, \tau): \psi(u, v, \tau)=\sup_{n=0, 1, \ldots} \psi_n(u, v, \tau)<\infty\}. \tag{12}$$

The problem arises to characterize those sequences μ for which the set W_μ is non-empty. Two classes of such μ will be given in the following two lemmas.

Lemma 1 *Suppose that the sequence $\mu=(M_0, M_1, \ldots)$ satisfies the estimate*

$$C_{i+j}^i M_i M_j \leqq (i+j) M_{i+j-1} \tag{13}$$

$(i, j=1, 2, \ldots)$. Then the set

$$Q_0=\{(u, v, \tau): u, v, \tau>0, u\tau+v+\tau\leqq 1\}$$

is contained in W_μ.

Proof. Given $i_1, i_2, \ldots, i_k\in\mathbb{N}$, by a repeated application of (13) we obtain

$$M_{i_1} M_{i_2} \ldots M_{i_k} \leqq \frac{i_1! i_2! \ldots i_k!}{(i_1+\ldots+i_k-k+1)!} M_{i_1+\ldots+i_k-k+1}.$$

hence

$$M_{|\alpha|} M_1^{\alpha_1} \ldots M_n^{\alpha_n} \leqq (n!)^{-1} M_{n|\alpha|} (1!)^{\alpha_1} \ldots (n!)^{\alpha_n}. \tag{14}$$

By the well known combinatorial identity (see [24])

$$\sum_{n, k} \frac{k!}{\beta_1! \ldots \beta_n!}=C_{n-1}^{k-1}$$

$(\beta=(\beta_1, \ldots, \beta_n), |\beta|=k, \|\beta\|=n)$ and (14) we get for $\beta_1=\alpha_0+\alpha_1, \beta_2=\alpha_2, \ldots, \beta_n=\alpha_n$

$$\begin{aligned}
\psi_n(u, v, \tau) &\leqq \sum_{\|\alpha\|=n} \frac{|\alpha|!}{\alpha_0! \alpha_1! \ldots \alpha_n!} u^{\alpha_0} v^{|\alpha|-\alpha_0} \tau^{\|\alpha\|-|\alpha|+\alpha_0} \\
&= \sum_{\|\alpha\|=n} \sum_{\beta_1=0}^{n} \sum_{\gamma=0}^{\beta_1} \frac{(\gamma+\alpha_1)!}{\gamma! \alpha_1!} \frac{|\alpha|!}{(\gamma+\alpha_1)! \alpha_2! \ldots \alpha_n!} \cdot u^{\gamma} \left(\frac{v}{\tau}\right)^{\alpha_1} \left(\frac{v}{\tau}\right)^{|\alpha|-\gamma-\alpha_1} \tau^n \\
&= \sum_{\|\beta\|=n} \frac{|\beta|!}{\beta_1! \ldots \beta_n!} \left(u+\frac{v}{\tau}\right)^{\beta_1} \left(\frac{v}{\tau}\right)^{|\beta|-\beta_1} \tau^n \\
&\leqq \sum_{\|\beta\|=1}^{n} C_{n-1}^{|\beta|-1} \left(u+\frac{v}{\tau}\right)^{|\beta|} \tau^n \leqq (u\tau+v+\tau)^n
\end{aligned}$$

which implies that $Q_0\subseteq W_\mu$. □

We remark that the sequence (5) satisfies the estimate (13) for $s \geqq 1$. This follows from the inequality

$$(i!\, j!)^{s-1} \leqq [(i+j-1)!]^{s-1} \qquad (s \geqq 1).$$

Lemma 1 immediately implies the following

Lemma 2 *Suppose that the sequence $\mu=(M_0, M_1, \ldots)$ satisfies the estimate*

$$C_{i+j}^{i} M_i M_j \leqq (i+j) H^{i+j-1} M_{i+j-1} \tag{15}$$

$(i, j=1, 2, \ldots)$ for some $H>1$. Then the set

$$Q_H = \{(u, v, \tau): u, v, \tau > 0, (u\tau + v + \tau) H \leqq 1\}$$

is contained in W_μ. □

An example of a sequence μ satisfying (15) is (9). Moreover, the sequence (5) satifies (15) for $0<s<1$, provided we take $H \leqq 2^{1-s}$. This follows from the inequality

$$\left[\frac{(i+j-1)!}{i!\, j!}\right]^{1-s} \leqq H^{i+j-1} \leqq 2^{(1-s)(i+j-1)} \qquad (0<s<1).$$

The conditions (13) and (15), which seem rather artificial, have a natural analytic meaning: they are sufficient for the class $R_\mu(L, X)$ to be invariant under composition (i.e. under the operator (1) with $f \in R_\mu(L, X)$). The problem of finding invariance conditions which are both necessary and sufficient has been open for a long time; this problem has been solved in a certain sense in a recent paper by Ider [11].

3 Sufficient acting conditions

In this section we are concerned with conditions on the function f: $[a, b] \times X \to X$ under which the corresponding superposition operator (2) acts between two Roumieu spaces $R_\mu(L_1, X)$ and $R_\mu(L_2, X)$. To this end, we shall assume throughout that f is smooth and satisfies

$$\| D_t^i D_u^j f(t, u) | X \| \leqq \gamma A^i B^j M_{i+j} \tag{16}$$

for $i, j=0, 1, \ldots$, $a \leqq t \leqq b$, and $\|u | X\| \leqq r$, where the constants A, B and γ depend, in general, on $r>0$. By formula (10), condition (16) implies that the functions satisfying (6) are invariant under composition with f; in fact, a more precise statement is true:

Theorem 1 *Suppose that the sequence μ is such that*

$$\left(\frac{A}{L}, Br\tau, \tau\right) \in W_\mu$$

for some fixed constants A, B, r, τ, $L>0$. *Assume that the function* f *satisfies* (16), *and let*

$$\rho=\gamma\psi\left(\frac{A}{L}, Br\tau, \tau\right)$$

with ψ *as in* (12). *Then the operator* (2) *maps the ball* $\|x\,|\,R_\mu(L, X)\|\leqq r$ *of* $R_\mu(L, X)$ *into the ball* $\|y\,|\,R_\mu(L/\tau, X)\|\leqq\rho$ *of* $R_\mu(L/\tau, X)$.

Proof. From the formula (10) for the derivatives of $y=Fx$ we conclude that, for $a\leqq t\leqq b$ and $\|x(t)\,|\,X\|\leqq r$,

$$\begin{aligned}\|y^{(n)}(t)\,|\,X\|&\leqq\sum_{\|\alpha\|=n}C_n^\alpha\gamma A^{\alpha_0}B^{\alpha_1+\ldots+\alpha_n}M_{|\alpha|}(rLM_1)^{\alpha_1}\ldots(rLM_n)^{\alpha_n}\\&\leqq\sum_{\|\alpha\|=n}C_n^\alpha\gamma A^{\alpha_0}B^{|\alpha|-\alpha_0}r^{|\alpha|-\alpha_0}L^{n-\alpha_0}M_{|\alpha|}M_1^{\alpha_1}\ldots M_n^{\alpha_n}\\&=\gamma\sum_{\|\alpha\|=n}C_n^\alpha\left(\frac{A}{L}\right)^{\alpha_0}(Br\tau)^{|\alpha|-\alpha_0}\tau^{n-|\alpha|+\alpha_0}M_{|\alpha|}M_1^{\alpha_1}\ldots M_n^{\alpha_n}\left(\frac{L}{\tau}\right)^n\\&\leqq\gamma\psi\left(\frac{A}{L}, Br\tau, \tau\right)\left(\frac{L}{\tau}\right)^n M_n=\rho\left(\frac{L}{\tau}\right)^n M_n.\quad\square\end{aligned}$$

The crucial point in Theorem 1 is, of course, to find elements in W_μ; this can be done by means of Lemma 1 and 2. For instance, combining Theorem 1 and Lemma 2 yields the following

Theorem 2 *Suppose that the sequence* μ *satisfies condition* (15), *and the function* f *satisfies* (16). *Let* $K=H(Br+1+A/L)$ *and*

$$\rho=\gamma\psi\left(\frac{A}{L}, \frac{Br}{K}, \frac{1}{K}\right)$$

with ψ *as in* (12). *Then the operator* (2) *maps the ball* $\|x\,|\,R_\mu(L, X)\|\leqq r$ *of* $R_\mu(L, X)$ *into the ball* $\|y\,|\,R_\mu(KL, X)\|\leqq\rho$ *of* $R_\mu(KL, X)$. $\square$

4 Necessary acting conditions

As the preceding theorems show, sufficient conditions for the operator F to act between two spaces $R_\mu(L_1, X)$ and $R_\mu(L_2, X)$ may be obtained by combining a growth condition (like (12)) on the sequence μ with a growth condition (like (16)) on the function f. In this section we shall prove a certain converse of this. First of all, we need a technical auxiliary result.

Lemma 3 *Let* f *be a smooth function on* $[a, b]\times X$, *and suppose that the corresponding superposition operator* (2) *maps the set of all linear functions*

$$x_{k,n}(t)=\frac{k+1}{n+1}(t-t_0)+x_0\qquad(k=0, 1, \ldots, n)\tag{17}$$

($t_0\in[a, b]$, $x_0\in X$ *fixed with* $\|x_0\,|\,X\|<r$) *into the ball* $\|y\,|\,R_\mu(L, X)\|\leqq\gamma$ *of* $R_\mu(L, X)$. *Then* f *satisfies* (16) *for* $i, j=0, 1, \ldots, a\leqq t\leqq b$, *and* $\|u\,|\,X\|\leqq r$, *where* $A=B=2eL$.

Proof. By hypothesis, $\|y_{k,n}|R_\mu(L, X)\| \leqq \gamma$ for $y_{k,n} = F x_{k,n}$; by (10) we have

$$y_{k,n}^{(n)}(t_0) = \sum_{\|\alpha\|=n} C_n^\alpha D_t^{\alpha_0} D_u^{|\alpha|-\alpha_0} f(t_0, x_0)\, x'^{\alpha_1}_{k,n}(t_0) \tag{18}$$

$$= \sum_{i=0}^{n} C_n^i D_t^{n-i} D_u^i f(t_0, x_0) \left[\frac{k+1}{n+1}\right]^i \quad (k=0, \ldots, n).$$

This system of linear equations for $D_t^n f(t_0, x_0), D_t^{n-1} D_u f(t_0, x_0), \ldots, D_u^n f(t_0, x_0)$ has a unique solution for each right-hand side, since the determinant of its coefficients is

$$\Delta_n = C_n^0 C_n^1 \ldots C_n^n \bar{\Delta}_n, \tag{19}$$

where

$$\bar{\Delta}_n = \begin{vmatrix} 1 & \theta_1 & \ldots & \theta_1^{n-1} & \theta_1^n \\ 1 & \theta_2 & \ldots & \theta_2^{n-1} & \theta_2^n \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & \theta_n & \ldots & \theta_n^{n-1} & \theta_n^n \\ 1 & 1 & \ldots & 1 & 1 \end{vmatrix}$$

$\left(\theta_i = \dfrac{i}{n+1}\right)$ is a Vandermonde determinant. Thus,

$$D_t^{n-i} D_u^i f(t_0, x_0) = \Delta_n^{-1} \Delta_{n,i}, \tag{20}$$

where $\Delta_{n,i} = C_n^0 C_n^1 \ldots C_n^{i-1} C_n^{i+1} \ldots C_n^n \bar{\Delta}_{n,i}$ with

$$\bar{\Delta}_{n,i} = \begin{vmatrix} 1 & \theta_1 & \ldots & \theta_1^{i-1} & y_{0,n}^{(n)}(t_0) & \theta_1^{i+1} & \ldots & \theta_1^n \\ 1 & \theta_2 & \ldots & \theta_2^{i-1} & y_{1,n}^{(n)}(t_0) & \theta_2^{i+1} & \ldots & \theta_2^n \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & \ldots & 1 & y_{n,n}^{(n)}(t_0) & 1 & \ldots & 1 \end{vmatrix}.$$

By evaluating this determinant with the usual product rule (see e.g. [19, 23]), a straightforward but cumbersome calculation leads to the estimate

$$\|D_t^{n-i} D_u^i f(t_0, x_0)|X\| \leqq \sum_{k=0}^{n} C_n^k \frac{(n+1)^{i+1}}{(i+1)!} \gamma L^n M_n = \frac{(n+1)^{i+1}}{(i+1)!} \gamma (2L)^n M_n$$

$$\leqq \frac{(n+1)^{n+1}}{(n+1)!} \gamma (2L)^n M_n \leqq \frac{(n+1)^{n+1} e^{n+1}}{(n+1)^{n+1} \sqrt{2\pi(n+1)}} \gamma (2L)^n M_n$$

$$\leqq \gamma (2eL)^n M_n,$$

where we used the fact that

$$\|y_{k,n}^{(n)}|C\| \leqq L^n M_n \|y_{k,n}|R_\mu(L, X)\| \leqq L^n M_n \gamma.$$

This proves the assertion. □

We pass now to the Beurling and Roumieu classes $R_\mu(0, X)$ and $R_\mu(\infty, X)$ defined in (3) and (4). To simplify the terminology, we shall call an operator F *weakly bounded* in $R_\mu(\infty, X)$ if F maps any bounded set in $R_\mu(L, X)$ into a bounded subset of $R_\mu(L', X)$ for some $L' \geqq L$.

Theorem 3 *Let $W_\mu \neq \emptyset$. The superposition operator* (2) *acts in the Roumieu class $R_\mu(\infty, X)$ and is weakly bounded if and only if* (16) *holds for some $A>0$, $B>0$, and $\gamma>0$.*

Proof. Let first $x \in R_\mu(\infty, X)$; by definition (4), we have then $x \in R_\mu(L, X)$ and $\|x|R_\mu(L, X)\| = r$ for some $L>0$ and $r>0$. Applying (16) to this r, we get

$$\|D_t^i D_u^j f(t, x(t))|X\| \leqq \gamma A^i B^j M_{i+j}$$

for some $\gamma>0$. By assumption, we find $(u, v, \tau) \in W_\mu$. If we choose $L' \geqq \max\{L, A/u\}$ and $\tau' \leqq \min\{\tau, v/Br\}$, we have, by Theorem 1, $\|Fx|R_\mu(L'/\tau', X)\| \leqq \rho$ with $\rho = \gamma\psi(u, v, \tau)$, and hence F maps $R_\mu(\infty, X)$ into itself and is weakly bounded.

Conversely, suppose that F acts in $R_\mu(\infty, X)$ and is weakly bounded. Since the set Q of all affine functions $x(t) = \alpha(t-t_0) + x_0 (0 \leqq \alpha \leqq 1, t_0 \in [a, b], \|x_0|X\| \leqq r)$ contains all functions of the form (17) and is bounded in each space $R_\mu(L, X)$, the hypotheses of Lemma 3 are satisfied, and hence (16) holds. □

Now we give an analogous result for the superposition operator F in the Beurling class (3). Recall that an operator F is called *bounded* if F maps bounded sets into bounded sets; a nonlinear operator may be bounded without being continuous, or continuous without being bounded.

Theorem 4 *Let $W_\mu \neq \emptyset$. The superposition operator* (2) *acts in the Beurling class $R_\mu(0, X)$ and is bounded if and only if, for each $A>0$ and $B>0$,* (16) *holds for some $\gamma>0$.*

Proof. Since the parameter τ' in the proof of Theorem 3 does not depend on L', the operator F actually maps the space $R_\mu(0, X)$ into itself. Moreover, since any bounded set in $R_\mu(0, X)$ is trivially bounded in $R_\mu(L, X)$ for each $L>0$, the hypothesis on (16) implies that that F is also bounded in $R_\mu(0, X)$. This proves the "if" part of Theorem 4. To prove the "only if" part, we remark that the set Q mentioned in the proof of Theorem 3 is of course bounded in $R_\mu(0, X)$, and thus it suffices to repeat the reasoning of the "only if" part of Theorem 3. □

5 Continuity and compactness

Suppose that E is some space of functions over an interval $[a, b]$ which is mapped by the superposition operator (2) into itself. In this situation, the problem of deciding whether or not the operator (2) is continuous on E is rather delicate and essentially depends on intrinsic properties of the space E involved. For example, it is well known (see e.g. [4]) that, whenever F maps the space $C = C([a, b], \mathbb{R})$ into itself, F is always continuous. On the other hand, F may map the Hölder space $C^\alpha = C^\alpha([0, 1], \mathbb{R})$, say, into itself without being continu-

ous; just consider the function $f(t, u)=\min\{u, 1\}$ which generates a discontinuous (at $x_0(t)=t^\alpha$, for example) superposition operator F in C^α [5]. Finally, the function

$$f(t, u)=\begin{cases} 0 & \text{if } u<0, \\ 3t^{-1}u^2-2t^{-3/2}u^3 & \text{if } 0\leqq u\leqq t^{1/2}, \\ 1 & \text{if } u>t^{1/2}. \end{cases}$$

generates a superposition operator F which maps the space $C^1=C^1([0, 1], \mathbb{R})$ into itself, but is not continuous [16]; observe that in this case even the function f is discontinuous at (0, 0), and hence F does not act in the space C.

Concerning the compactness of F, the situation is even worse. In fact, in many well-known function spaces E the operator (2) is compact only if F is constant, i.e. the function f does not depend on u; this degeneracy occurs not only in the space C [1], but also in the Lebesgue space L_p [12], as well as in many other spaces of measurable functions arising in applications [3]. This is, of course, rather disappointing when one needs compactness to apply classical techniques of nonlinear analysis (e.g. Schauder's fixed point principle).

In contrast to this, the following two theorems show that F is always continuous and compact in the space $R_\mu(0, X)$ and $R_\mu(\infty, X)$ in case $X=\mathbb{R}^N$.

Theorem 5 *Suppose that $W_\mu\neq\emptyset$ and* (16) *holds for some $A>0$, $B>0$, and $\gamma>0$. Then the superposition operator* (2) *maps the Roumieu class $R_\mu(\infty, \mathbb{R}^N)$ into itself and is bounded, continuous, and compact.*

Proof. The fact that F acts in $R_\mu(\infty, \mathbb{R}^N)$ was already proved in Theorem 3. Let $Q\subset R_\mu(\infty, \mathbb{R}^N)$ be bounded. Since $R_\mu(\infty, \mathbb{R}^N)$ is a regular inductive limit (see [26, 29]), we have $Q\subset R_\mu(L, \mathbb{R}^N)$ for some $L>0$, and hence $F(Q)$ is bounded in $R_\mu(L', \mathbb{R}^N)$ for some $L'\geqq L$, by the weak boundedness of F (Theorem 3). This shows that the operator (2) is bounded in $R_\mu(\infty, \mathbb{R}^N)$.

The compactness of F follows, of course, from the fact that bounded sets and precompact sets in $R_\mu(\infty, \mathbb{R}^N)$ coincide. The essential part of Theorem 5 is the continuity assertion which may be proved as follows.

Again by the regularity of the inductive limit $R_\mu(\infty, \mathbb{R}^N)$, continuity is equivalent to sequential continuity [29]; moreover, it suffices to prove the continuity of F at zero, under the hypothesis that $f(t, 0)\equiv 0$ (otherwise one may pass from the function f to the function $f(t, x_0(t)+u)-f(t, x_0(t))$). Now, if x_n is a sequence in $R_\mu(\infty, \mathbb{R}^N)$ converging to zero, by the regularity of $R_\mu(\infty, \mathbb{R}^N)$ and the boundedness of $\{x_n\}$ we have $x_n\in R_\mu(L, \mathbb{R}^N)$ for some $L>0$. Since F is weakly bounded in $R_\mu(\infty, \mathbb{R}^N)$, the set $\{Fx_n\}$ is bounded in $R_\mu(L', \mathbb{R}^N)$ for some $L'\geqq L$. Moreover, $\{Fx_n\}$ is precompact in $R_\mu(L'', \mathbb{R}^N)$ for $L''>L'$, and hence has an accumulation point y^*. Since the function f is continuous on $[a, b]\times\mathbb{R}^N$, the operator F is continuous in the space $C=C([a, b], \mathbb{R}^N)$, and hence Fx_n converges to zero in C. But this implies that y^* is zero, and therefore F is continuous at zero, considered as an operator in $R_\mu(\infty, \mathbb{R}^N)$. □

In rather the same way we obtain the following result from Theorem 4.

Theorem 6 *Suppose that $W_\mu\neq\emptyset$ and, for each $A>0$ and $B>0$,* (16) *holds for some $\gamma>0$. Then the superposition operator* (2) *maps the Beurling class $R_\mu(0, \mathbb{R}^N)$ into itself and is bounded, continuous, and compact.*

6 The space $R_\mu^0(L, X)$

The results of the previous section show that, whenever F acts in the class $R_\mu(0, \mathbb{R}^N)$ or $R_\mu(\infty, \mathbb{R}^N)$ and is bounded, then F is automatically continuous and compact. These classes are, of course, locally convex spaces which are not normable (this follows, for instance, from the fact that bounded sets are precompact). On the other hand, the most advanced theory on the superposition operator in function spaces is formulated in the Banach space setting; thus the question arises what can be said about the analytical "behaviour" of F in the Banach spaces $R_\mu(L, X)$ themselves, rather than their inductive limit as $L \to \infty$, or projective limit as $L \to 0$.

It turns out that, apart from $R_\mu(L, X)$, the space $R_\mu^0(L, X)$ of all functions $x \in R_\mu(L, X)$ for which

$$\frac{\|x^{(k)}|C\|}{L^k M_k} \to 0 \qquad (k \to \infty) \tag{21}$$

is useful for studying the superposition operator; for instance, these spaces have the advantage that they allow a simple explicit description of compact subsets:

Lemma 4 *A bounded closed set $A \subset R_\mu^0(L, X)$ is compact if and only if* (21) *holds uniformly in $x \in A$.*

Proof. Let

$$\alpha(A) = \overline{\lim_{k \to \infty}} \sup_{x \in A} L^{-k} M_k^{-1} \|x^{(k)}|C\|, \tag{22}$$

and suppose first that A is compact. Given $\varepsilon > 0$, we may find a finite ε-net $\{z_1, \ldots, z_m\} \subset A$ for A. Fix $x \in A$, and choose z_j such that $\|x - z_j|R_\mu(L, X)\| \leqq \varepsilon$. Since $z_j \in R_\mu^0(L, X)$, we have $\|z_j^{(k)}|C\| \leqq \varepsilon L^k M_k$ for $k \geqq k_0$, hence $\|x^{(k)}|C\| \leqq 2\varepsilon L^k M_k$ for these k; this shows that $\alpha(A) = 0$ for compact A.

Conversely, suppose that $\alpha(A) = 0$, i.e. for $\varepsilon > 0$ we find $\bar{k} \in \mathbb{N}$ such that $\|x^{(k)}|C\| \leqq \varepsilon L^k M_k$ for $k \geqq \bar{k}$. Being a bounded closed set in $R_\mu(L, X)$, A is certainly compact in each of the spaces $C^k = C^k([a, b], X)$, by the Arzelà-Ascoli theorem. Consequently, we find a finite set $\{z_1, \ldots, z_m\} \subset A$ such that, given $x \in A$, we have

$$\|x - z_j|C^{\bar{k}}\| = \max_{k \leqq \bar{k}} \|x^{(k)} - z_j^{(k)}|C\| \leqq \varepsilon \max_{k \leqq \bar{k}} L^k M_k$$

for some z_j. Since $z_j \in R_\mu^0(L, X)$, we have also that $\|z_j^{(k)}|C\| \leqq \varepsilon L^k M_k$ for $k \geqq \bar{k}$. Altogether,

$$\begin{aligned} \|x - z_j|R_\mu(L, X)\| &= \sup_{k \geqq 0} L^{-k} M_k^{-1} \|x^{(k)} - z_j^{(k)}|C\| \\ &\leqq \max_{k \leqq \bar{k}} L^{-k} M_k^{-1} \|x^{(k)} - z_j^{(k)}|C\| + \sup_{k \geqq \bar{k}} L^{-k} M_k^{-1} \|x^{(k)} - z_j^{(k)}|C\| \\ &\leqq \varepsilon + 2\varepsilon, \end{aligned}$$

and hence $\{z_1, \ldots, z_m\}$ is a finite 3ε-net for A in $R_\mu(L, X)$; this shows that A is compact if $\alpha(A) = 0$. □

We remark that (22) gives an explicit formula for the *Hausdorff measure of noncompactness* (see e.g. [27]) of the set A in $R_\mu(L, X)$; for details see [2].

It is easy to see that the imbedding

$$R_\mu^0(L, X) \subseteq R_\mu(L, X) \subseteq R_\mu^0(L+\varepsilon, X) \tag{23}$$

holds for every $\varepsilon > 0$. Moreover, Lemma 4 shows that the second imbedding in (23) is compact.

As a consequence of Lemma 4, one may obtain continuity and compactness results for superposition operators mapping the class $R_\mu(L, X)$ into the subclass $R_\mu^0(L, X)$. Thus, if the operator (1) maps the space R_μ of all smooth functions x with $\|x^{(k)}|C\| = O(L^k M_k)$ into the subspace R_μ^0 of all functions $y \in R_\mu$ satisfying $\|y^{(k)}|C\| = o(L^k M_k)$, F is continuous. Such results have analogues in other well known function spaces. For instance, if the operator (1) maps the *Hölder space* $C^\alpha = C^\alpha([a, b], \mathbb{R})$ of all continuous functions x on $[a, b]$ with $|x(s)-x(t)| = O(|s-t|^\alpha)$ into the "little" Hölder space $C^{\alpha,0} = C^{\alpha,0}([a, b], \mathbb{R})$ of all $y \in C^\alpha$ satisfying $|y(s)-y(t)| = o(|s-t|^\alpha)$, then F is automatically continuous [5]. The same phenomenon holds if the operator (2) maps an *ideal space of measurable functions* X (see e.g. [30]) into the subspace X^0 of all $x \in X$ with absolutely continuous norm [14]. By this we mean that

$$\|\chi_D x|X\| \to 0 \qquad (\text{mes } D \to 0),$$

where χ_D denotes the characteristic function of a measurable set $D \subseteq [a, b]$. For example [13], any superposition operator which acts in an Orlicz space L_M with $M \in \Delta_2$ is automatically continuous.

Acknowledgement. The authors are grateful to the referee for pointing their attention to the paper [11].

References

1. Appell, J.: Implicit functions, nonlinear integral equations, and the measure of noncompactness of the superposition operators. J. Math. Anal. Appl. **83**, 1, 251–263 (1981)
2. Appell, J., De Pascale, E.: Lipschitz and Darbo conditions for the superposition operator in some non-ideal spaces of differentiable functions, Annali Mat. Pura Appl. (to appear)
3. Appell, J., Massabò, I., Vignoli, A., Zabrejko, P.P.: Lipschitz and Darbo conditions for the superposition operator in ideal spaces. Annali Mat. Pura Appl. **152**, 123–137 (1988)
4. Appell, J., Zabrejko, P.P.: Nonlinear superposition operators. Cambridge: Cambridge University Press 1990
5. Berkolajko, M.Z.: On a nonlinear operator acting in generalized Hölder spaces. (Russian), Voronezh. Gos. Univ. Trudy Sem. Funk. Anal. **12**, 96–104 (1969)
6. Dzhanashiya, G.A.: On the Carleman problem for Gevrey function classes (Russian), Doklady Akad. Nauk SSSR **145**, 2, 259–262 (1962) [=Soviet Math., Doklady **3**, 969–972 (1962)]
7. Dzhanashiya, G.A.: On the superposition of two functions from Gevrey function classes. (Russian), Soobsh. Akad. Nauk Gruz. SSR **33**, 257–262 (1964)
8. Fraenkel, L.E.: Formulae for high derivatives of composite functions. Math. Proc. Cambridge Philos. Soc. **83**, 159–165 (1978)
9. Friedman, A.: Generalized functions and partial differential equations. Englewood Cliffs, New Jersey: Prentice-Hall, 1963
10. Gevrey, M.: Sur la nature analytique des solutions des équations aux dérivées partielles. Ann. Ecole Norm. Sup. Paris **35**, 129–190 (1918)

11. Ider, M.: On the superposition of functions in Carleman classes. Bull. Austral. Math. Soc. **39**, 471–476 (1989)
12. Krasnosel'skij, M.A.: Topological methods in the theory of nonlinear integral equations. (Russian), Moscow: Gostekhizdat 1956 [Engl. transl.: New York: McMillan 1964]
13. Krasnosel'skij, M.A., Rutitskij, Y.B.: Convex functions and Orlicz spaces. (Russian), Moscow: Fizmatgiz 1958 [Engl. transl.: Noordhoff, Groningen 1961]
14. Krasnosel'skij, M.A., Rutitskij, Y.B., Sultanov, R.M.: On a nonlinear operator acting in a space of abstract functions. (Russian), Izv. Akad. Nauk Azerb. SSR **3**, 15–21 (1959)
15. Krasnosel'skij, M.A., Vajnikko, G.M., Zabrejko, P.P., Rutitskij, Y.B., Stetsenko, V.Y.: Approximate solutions of operator equations. (Russian), Moscow: Nauka 1969 [Engl. transl.: Groningen: Noordhoff 1972]
16. Lichawski, K., Matkowski, J., Miś, J.: Locally defined operators in the space of differentiable functions. Univ. Bielsko-Biała 1988 (preprint)
17. Lions, J.-L., Magenes, E.: Problèmes aux limites non homogènes et aplications III. Paris: Dunod 1970
18. Mandelbrojt, S.: Séries adhérentes, régularisation de suites, applications. Paris: Gauthier-Villars 1952
19. Mishina, A.P.: Propuryanov, I.V.: Higher algebra. (Russian), Moscow: Nauka 1965
20. Mityagin, B.S.: An indefinitely differentiable function with the values of its derivatives given at a point. (Russian), Doklady Akad. Nauk SSSR **138**, 2, 289–292 (1961) [=Soviet Math., Doklady **2**, 594–597 (1961)
21. Nazarov, V.I.: The superposition operator in Roumieu spaces of infinitely differentiable functions. (Russian), Izv. Akad. Nauk BSSR **5**, 22–28 (1984)
22. Nazarov, V.I.: A nonlinear differential equation of first order in Roumieu spaces. (Russian), Doklady Akad. Nauk BSSR **28**, 780–783 (1984)
23. Northcott, D.G.: Multilinear algebra, Cambridge: Cambridge University Press 1984
24. Riordan, J.: Combinatorial identities. New York: Wiley 1967
25. Roman, S.: The formula of Faà di Bruno. Am. Math. Monthly **87**, 805–809 (1980)
26. Roumieu, C.: Sur quelques extensions de la notion de distributions. Ann. Ecole Norm. Sup. Paris **77**, 47–121 (1960)
27. Sadovskij, B.N.: Limit-compact and condensing operators. (Russian), Uspekhi Mat. Nauk **27**, 1, 81–146 (1972) [=Russ. Math. Surv. **27**, 85–155 (1972)]
28. Shilov, G.Y.: Differentiability of functions in linear spaces. (With an appendix by P.P. Zabrejko). (Russian), Yaroslav. Gos. Univ. 6–120 (1978)
29. Silva, J.S.: Su certe classi di spazi localmente convessi importanti per le applicazioni. Roma: Rend. Mat. Univ. **14**, 388–410 (1955)
30. Zabrejko, P.P.: Ideal function spaces I (Russian), Vestnik: Yaroslav. Gos. Univ. **8**, 12–52 (1974)