SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Inventiones Mathematicae
Verlag: Springer
Jahr: 1991
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN356556735_0105
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN356556735_0105

Übergeordnetes Werk

Werk Id: PPN356556735
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN356556735
OPAC: http://opac.sub.uni-goettingen.de/DB=1/PPN?PPN=356556735

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

Inventiones mathematicae

Volume 105 1991

Editores

M. Berger, Bures-sur-Yvette **J.-M. Bismut,** Orsay **J.-M. Bony,** Palaiseau
A. Borel, Princeton **J. Coates,** Cambridge **A. Connes,** Bures-sur-Yvette
B. Mazur, Cambridge, Mass. **Y. Meyer,** Paris **J.W. Morgan,** New York
R. Remmert, Münster

Editores quondam

J.F. Adams (1974–1984) V.I. Arnold (1966–1978) H. Bauer (1966–1978)
B.J. Birch (1966–1978) E. Bombieri (1969–1979) P. Cartier (1966–1975)
P. Gabriel (1966–1967) A. Haefliger (1966–1968) L. Hörmander (1976–1984)
A.A. Kirillov (1978–1991) N.H. Kuiper (1966–1974) B. Malgrange (1966–1976)
Y. Manin (1966–1978) J.K. Moser (1974–1982) J-P. Serre (1967–1979, 1982–1988)
J. Sjöstrand (1984–1991) T.A. Springer (1976–1986) J. Tits (1966–1975)
E. Vesentini (1966–1973) S.-T. Yau (1980–1989)

Springer International

Inventiones mathematicae

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 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 0020-9910, 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.

Urheberrecht. Voraussetzung für die Einreichung eines Manuskriptes an die Redaktion der Zeitschrift ist, daß die Arbeit noch nicht publiziert oder an anderer Stelle zur Publikation eingereicht wurde. Ferner wird vorausgesetzt, daß die Publikation von allen beteiligten Autoren einer Arbeitsgruppe genehmigt und daß die Arbeit, wenn sie zur Publikation angenommen wurde, nicht an anderer Stelle in gleicher Form, ohne die Genehmigung des Copyright-Inhabers, publiziert wird, entweder in derselben oder einer anderen Sprache.
Die Zeitschrift sowie alle in ihr enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen schriftlichen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden. Jeder Autor, der Deutscher ist oder ständig in der Bundesrepublik Deutschland oder Berlin (West) lebt oder Bürger Österreichs, der Schweiz oder eines Staates der Europäischen Gemeinschaft ist, kann unter bestimmten Voraussetzungen an der Ausschüttung der Bibliotheks- und Fotokopietantiemen teilnehmen. Nähere Einzelheiten können direkt von der Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestr. 49, D-8000 München 2, eingeholt werden.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in dieser Zeitschrift berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, daß solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürfen.

Printers: Universitätsdruckerei H. Stürtz AG, Würzburg

© Springer International 1991

Contents

Apanasov, B.N.: Nonstandard uniformized conformal structures on hyperbolic manifolds 137

Aschbacher, M., Segev, Y.: The uniqueness of groups of type J_4 589

Colliot-Thélène, J.-L., Raskind, W..: Groupe de Chow de codimension deux des variétés définies sur un corps de nombres: un théorème de finitude pour la torsion 221

Costa, C.J.: Classification of complete minimal surfaces in R^3 with total curvature 12π 273

Delorme, P.: Coefficients généralisés de séries principales sphériques et distributions sphériques sur $G_{\mathbb{C}}/G_{\mathbb{R}}$ 305

Ecker, K., Huisken, G.: Interior estimates for hypersurfaces moving by mean curvature 547

Elkies, N.D., Odlyzko, A.M., Rush, J.A.: On the packing densities of superballs and other bodies 613

Ferry, S.C., Weinberger, S..: Curvature, tangentiality, and controlled topology 401

Gelbart, S.S., Ragowski, J.D.: L-functions and Fourier-Jacobi coefficients for the unitary group U(3) 445

Gersten, S.M., Short, H.: Small cancellation theory and automatic groups: Part II 641

Ginzburg, D.: A Rankin-Selberg integral for the adjoint representation of GL_3 571

Habegger, N., Saper, L.: Intersection cohomology of cs-spaces and Zeeman's filtration 247

Holschneider, M., Tchamitchian, P.: Pointwise analysis of Riemann's "nondifferentiable" function 157

Huisken, G. → Ecker, K.

Jerison, D.: Prescribing harmonic measure on convex domains 375

Kawamata, Y.: On the length of an extremal rational curve 609

Kirby, R., Melvin, P.: The 3-manifold invariants of Witten and Reshetikhin-Turaev for sl(2, C) 473

Knop, F.: Nichtlinearisierbare Operationen halbeinfacher Gruppen auf affinen Räumen 217

Li, P., Tam, L.: The heat equation and harmonic maps of complete manifolds 1

Lions, P.L., Perthame, B.: Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system 415

Maldeghem, H. van → Thas, J.A.

Melvin, P. → Kirby, R.

Moscovici, H., Stanton, R.J.: R-torsion and zeta functions for locally symmetric manifolds 185

Moussu, R., Roche, C.: Théorie de Hovanskii et problème de Dulac 431

Nowicki, T., Strien, S. van: Invariant measures exist under a summability condition for unimodal maps 123

Odlyzko, A.M. → Elkies, N.D.

Payne, S.E. → Thas, J.A.

Perthame, B. → Lions, P.L.

Ragowski, J.D. → Gelbart, S.S.

Raskind, W.. → Colliot-Thélène, J.-L.

Roche, C. → Moussu, R.

Rush, J.A. → Elkies, N.D.

Saper, L. → Habegger, N.

Schneider, P., Stuhler, U.: The cohomology of p-adic symmetric spaces 47

Segev, Y. → Aschbacher, M.

Short, H. → Gersten, S.M.

Silverman, J.H.: Rational points on K 3 surfaces: A new canonical height 347

Stanton, R.J. → Moscovici, H.

Strien, S. van → Nowicki, T.

Stuhler, U. → Schneider, P.

Tam, L. → Li, P.

Tchamitchian, P. → Holschneider, M.

Thas, J.A., Payne, S.E., Maldeghem, H. van: Half Moufang implies Moufang for finite generalized quadrangles 153

Weinberger, S.. → Ferry, S.C.

Wilson, J.S.: Finite presentations of pro-p groups and discrete groups 177

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

Inventiones mathematicae

1-3 TD

Vol. 105 Fasc. 1 1991

2 A 18 3

P. Li, L. Tam	The heat equation and harmonic maps of complete manifolds	1
P. Schneider, U. Stuhler	The cohomology of p-adic symmetric spaces	47
T. Nowicki, S. van Strien	Invariant measures exist under a summability condition for unimodal maps	123
B.N. Apanasov	Nonstandard uniformized conformal structures on hyperbolic manifolds	137
J.A. Thas, S.E. Payne, H. Van Maldeghem	Half Moufang implies Moufang for finite generalized quadrangles	153
M. Holschneider, P. Tchamitchian	Pointwise analysis of Riemann's „nondifferentiable" function	157
J.S. Wilson	Finite presentations of pro-p groups and discrete groups	177
H. Moscovici, R.J. Stanton	R-torsion and zeta functions for locally symmetric manifolds	185
F. Knop	Nichtlinearisierbare Operationen halbeinfacher Gruppen auf affinen Räumen	217

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

Springer International

222 Invent math ISSN 0020-9910 INVMBH 105(1) 1–220 (1991) July 1991
Printed on acid-free paper

Eingg. 03. Juli

X 20

Inventiones mathematicae

This journal is published at frequent intervals to bring out new contributions to mathematics It is a policy of the journal to publish papers within four months of acceptance.

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 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 0020-9910, 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.

Pour permettre une publication plus rapide, le journal «Inventiones mathematicae» paraît par fascicules séparés réunis ensuite en volumes La rédaction s'efforce de publier les travaux reçus dans les quatre mois suivant leur acceptation.

Droits d'auteur

Ne peuvent être acceptés que des articles n'étant pas simultanément remis pour publication ailleurs et n'ayant jamais été publiés, ni dans leur pays d'origine, ni à l'étranger. L'auteur s'engage à ne pas soumettre son manuscrit ailleurs après qu'il a été accepté. Dès l'acceptation d'un manuscrit et sa publication par la maison d'édition, le droit d'impression en toutes langues et pour tous pays, y compris le droit de reproduction photographique, en microforme ou par tout autre moyen, est acquis par la maison d'édition.

Die Zeitschrift erscheint, um eine rasche Publikation zu ermöglichen, in einzelnen Heften, die zu Bändern vereinigt werden. Es ist das Ziel der Zeitschrift, Arbeiten innerhalb von 4 Monaten nach Annahme zu publizieren.

Urheberrecht

Voraussetzung für die Einreichung eines Manuskriptes an die Redaktion der Zeitschrift ist, daß die Arbeit noch nicht publiziert oder an anderer Stelle zur Publikation eingereicht wurde. Ferner wird vorausgesetzt, daß die Publikation von allen beteiligten Autoren einer Arbeitsgruppe genehmigt und daß die Arbeit, wenn sie zur Publikation angenommen wurde, nicht an anderer Stelle in gleicher Form, ohne die Genehmigung des Copyright-Inhabers, publiziert wird, entweder in derselben oder einer anderen Sprache.

Die Zeitschrift sowie alle in ihr enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen schriftlichen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen

Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden.

Jeder Autor, der Deutscher ist oder ständig in der Bundesrepublik Deutschland oder Berlin (West) lebt oder Bürger Österreichs, der Schweiz oder eines Staates der Europäischen Gemeinschaft ist, kann unter bestimmten Voraussetzungen an der Ausschüttung der Bibliotheks- und Fotokopietantiemen teilnehmen. Nähere Einzelheiten können direkt von der Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestr. 49, D-8000 München 2, eingeholt werden.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in dieser Zeitschrift berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, daß solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Manuscripts in duplicate should be submitted to

Les manuscrits en deux exemplaires doivent être envoyés à

Manuskripte in doppelter Anfertigung erbeten an

Reinhold Remmert
Berliner Strasse 7
W-4540 Lengerich, FRG

or to any of the editors
ou à un éditeur
oder an einen der Herausgeber

Invent. math. 105, 1–46 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

The heat equation and harmonic maps of complete manifolds

Peter Li* and Luen-fai Tam**

Department of Mathematics, University of California, Irvine, CA 92717, USA

Oblatum 19-III-1990

Table of contents

0 Introduction 1
1 Mean-value inequalities 4
2 Gradient estimates 11
3 Short time solutions of the parabolic harmonic map equation 17
4 Harmonic maps: $K = 0$ and $\kappa = 0$ 23
5 Harmonic maps: $\lambda(M) > 0$ and $\kappa = 0$ 28
6 Harmonic maps: Small total energy 39
References 45

0 Introduction

The primary goal of this article is to understand when one can solve the harmonic map equation between two complete noncompact manifolds. Specifically, we would like to obtain sufficient conditions to ensure that a C^1 map, h, with bounded energy density between two complete noncompact manifolds can be deformed smoothly to a smooth harmonic map by solving the parabolic equation for harmonic maps. The assumption on the energy density of h is natural for most of our applications. In fact, it is known that even the standard scalar heat equation on $\mathbf{R}^m$ does not have unique solution if the initial data grows too rapidly as a function of $\mathbf{R}^m$.

Let the metrics of two complete noncompact manifolds M^m and N^n be given locally by $ds_M^2 = g_{ij}dx^i dx^j$ and $ds_N^2 = h_{\alpha\beta}du^\alpha du^\beta$, respectively, then the energy density function of u is defined by

$$e(u)(x) = g^{ij}(x)\frac{\partial u^\alpha}{\partial x^i}(x)\frac{\partial u^\beta}{\partial x^j}(x)h_{\alpha\beta}(u(x)) ,$$

* Research supported by an NSF grant and a John Simon Guggenheim Fellowship.
** Research partially supported by the Earmarked Grant for Research, Hong Kong.

and the total energy of u is given by

$$E(u) = \int_M e(u)(x)\,dx\ .$$

The harmonic map equation from M into N, which is the Euler-Lagrange equation for critical points of the total energy functional, can be written as

$$\Delta u^\alpha(x) = -\Gamma^\alpha_{\beta\gamma}(u(x))\frac{\partial u^\beta}{\partial x^i}(x)\frac{\partial u^\gamma}{\partial x^j}(x)\,g^{ij}(x)$$

for all α, where the $\{\Gamma^\alpha_{\beta\gamma}\}$ are the Christoffel symbols of N. The corresponding parabolic equation defined on $M \times [0, \infty)$, with initial condition h is given by

$$\left(\Delta - \frac{\partial}{\partial t}\right)u^\alpha(x,t) = -\Gamma^\alpha_{\beta\gamma}(u(x,t))\frac{\partial u^\beta}{\partial x^i}(x,t)\frac{\partial u^\gamma}{\partial x^j}(x,t)\,g^{ij}(x)$$

$$u(x,0) = h(x)\ . \tag{0.1}$$

When M and N are compact without boundary and N has nonpositive sectional curvature, Eells-Sampson in [E-S] proved that any C^1 map from M into N can be deformed to a harmonic map by solving (0.1). The analogous version for compact manifolds with boundary was proved by Hamilton in [Hm]. Mitteau in [M] also proved that for compact manifolds M and N without boundary, there exists a constant $\varepsilon > 0$ depending on M and N such that any C^1 harmonic map from M to N, with energy density bounded above by ε, can be smoothly deformed via (0.1) to a smooth harmonic map. We would like to point out that these theorems can also be proven by considering the elliptic harmonic maps equation directly using techniques in calculus of variations. The reader should refer to the article of Schoen in [S] in regard to this approach. For other related topics concerning harmonic maps we would like to point out the survey articles of [E-L1], [E-L2], [J1], and [J2].

In [S-Y1], Schoen-Yau proved that if M is complete and noncompact and N has nonpositive sectional curvature, then any C^1 map from M into N with finite total energy can be deformed to a harmonic map with finite total energy. Jost-Yau in their articles [J-Y1] and [J-Y 2] applied the above results to study the rigidity of locally symmetric spaces with finite volume. Recently, Liao and the second author gave another proof of Schoen-Yau's theorem by solving the parabolic equation (0.1) in [Lo-T].

To this date, not much is known for a general noncompact manifold M, if we do not impose the finite total energy condition on the initial map. Struwe and Chen-Struwe in [St] and [C-St] studied equations (0.1) for the case when M is $\mathbf{R}^m$ and the initial map is C^1 with bounded energy density. A year ago, Choi-Treibergs in [C-T1] gave a 1-parameter family of examples of harmonic maps from the hyperbolic 2-space $\mathbf{H}^2$ into itself which have the property that they map the geometric boundary of $\mathbf{H}^2$ into itself. In [C-T2], they also constructed examples of harmonic maps from higher dimensional manifolds which are diffeomorphic to $\mathbf{R}^n$ into $\mathbf{H}^n$. These harmonic maps were obtained as the Gauss map of constant mean-curvature hypersurfaces of Minkowski space. Aviles-Choi-Micallef in [A-C-M] also proved that any continuous map from the geometric boundary of a strongly negatively curved Hadamard manifold into bounded convex ball can be extended uniquely to a harmonic map. Recently, Shi-Yau in [Sh-Y] solved (0.1), using a different argument than ours, for the case when M is noncompact with

nonnegative Ricci curvature, N is compact with nonpositive curvature, and the initial map has bounded energy density. This result is also a consequence of our program which will be discussed in §4. We would also like to point out that some of these theorems were reproved by Kendall [K] using probabilistic approach.

In the first two sections of this paper, we will deal with the heat equation for the Laplacian on functions. These estimates are not only essential for the purpose of solving (0.1), they are also interesting in their own rights. In §1, we will first prove two versions of the mean-value inequality for nonnegative subsolutions of the heat equation. The first version (Theorem 1.1.) can be viewed as the parabolic analogue of the mean-value inequality for nonnegative subharmonic functions derived in [L-S] by Schoen and the first author. The main point of this mean-value inequality is that the dependency of the constant is explicit in terms of the lower bound of the Ricci curvature and the radius of the ball alone. In the second section, we will derive estimates for solutions of the homogeneous and inhomogeneous heat equations for functions on a complete manifold.

The third section §3 is devoted to the question of short time existence of solutions for (0.1). We will prove that if M is noncompact with Ricci curvature bounded from below and N is an arbitrary noncompact manifold, then (0.1) can be solved on $M \times [0, \varepsilon)$ for any initial map h with bounded energy density and $h(M)$ being bounded in N.

In §4, §5, and §6 we will consider the questions of long time existence and of convergence to a harmonic map as time approaches infinity. Theorems 4.1, and 4.1′ are the noncompact, bounded energy density anologue of the long time existence theorem of Eells-Sampson. Both theorems deal with the case when M is noncompact with Ricci curvature bounded from below, N is noncompact with nonpositive sectional curvature, and the initial map h is of bounded energy density. Under these assumptions, these theorems assert that there is a unique solution of (0.1) on $M \times [0, \infty)$. In the event if the Ricci curvature of M is nonnegative, then the solution of the parabolic harmonic map equation obtained by Theorem 4.1′ in fact converges to a harmonic map (see Theorem 4.3) provided that there exists a point $p \in M$ such that some sequence $u(p, t_\nu)$ is convergent in N.

In §5, we consider the situation which is basically opposite to the assumption that M has nonnegative Ricci curvature. We will assume that the Ricci curvature of M is bounded from below by $-K \leqq 0$, the volume of the unit ball centered at any point in M is uniformly bounded away from 0, and the lower bound of the spectrum $\lambda(M)$ is positive. If the target manifold N has nonpositive sectional curvature, then for any initial map with bounded energy density and tension field in $L^p \cap L^\infty$ for some $p > 2$, it can be deformed to a harmonic map through (0.1). Applying this general existence theorem to hyperbolic spaces, we will prove in Theorem 5.4 that given any C^3 map from the infinity sphere $\mathbf{S}^{m-1}$ of $\mathbf{H}^m$ to the infinity sphere $\mathbf{S}^{n-1}$ of $\mathbf{H}^n$ whose energy density is nonvanishing, there exists a harmonic from $\mathbf{H}^m$ to $\mathbf{H}^n$ which realizes the given boundary map. In Theorem 5.5, we prove that if in addition $m = n = 2$ and the boundary map is a diffeomorphism, then the harmonic map is in fact a diffeomorphism. Recently, K. Akutagawa [A] has independently proved the case $m = n = 2$ but assuming the boundary map is C^4 instead. His argument used the fact that harmonic maps are conformally invariant when the domain manifold is of dimension 2.

In Theorem 6.1, we will prove a noncompact version of Mitteau's theorem. We assume that M is noncompact with Ricci curvature bounded from below by $-K \leqq 0$ and N is complete with section curvature bounded from above by $\kappa \geqq 0$.

Suppose that the volume of unit balls in M is uniformly bounded from below by $b > 0$. Then for any $\alpha > 0$, there exists a constant $\varepsilon > 0$ depending only on m, K, κ, b, and α, such that for any C^1 map h which has its energy density bounded from above by α, its total energy bounded from above by ε, and $h(M)$ is bounded, then (0.1) has a solution on $M \times [0, \infty)$. Moreover, if there exists a point $p \in M$ and a sequence $t_v \to \infty$ such that $u(p, t_v)$ converges in N, then the sequence of maps $u(\cdot, t_v)$ converges uniformly on compact subsets to a harmonic map. The assumption that $h(M)$ is bounded in N can be deleted if N satisfy a retraction assumption which we refer to as condition (A). Theorem 6.5 states that if M has nonnegative Ricci curvature and has volume of geodesic balls grows like $V_p(r) \geqq C_5 r^2$, then there exists a constant $\varepsilon > 0$ depending only on κ, C_5, and α, such that any harmonic map whose energy density is bounded from above by α and whose total energy is bounded from above by ε must be a constant map. Finally, we will give an example of a nonconstant harmonic map which has small total energy satisfying the hypotheses of Theorem 6.1.

Acknowledgement. The first author would like to thank R. Schoen for many valuable discussions concerning Theorem 5.4.

1 Mean-value inequalities

In this section we will prove two versions of the mean-value inequality for positive subsolutions of the heat equation on a manifold. The first (Theorem 1.1) is a parabolic version of the mean-value inequality for positive subharmonic functions derived in [L-S], while the second (Theorem 1.2) is a slight modification of the first version with the initial data taken into account. We will also apply this to study the question of uniqueness of L^p solutions for the heat equation.

Theorem 1.1 *Let M^m be a complete noncompact Riemannian manifold with boundary. Let $x_0 \in M$ and $R > 0$ be such that the geodesic ball $B(2R) = B_{x_0}(2R)$, centered at x_0 of radius $2R$, does not intersect the boundary of M. Suppose g is a nonnegative subsolution of the heat equation on $B(2R) \times [0, T]$ for some $0 < T \leqq \frac{1}{4} R^2$. If $-K \leqq 0$ is the lower bound of the Ricci curvature of $B(2R)$, then for $p > 0$, there exists positive constants C_{14} and C_{15} depending only on m and p such that for any $0 < \tau < T$, $0 < \delta < \frac{1}{2}$, and $0 < \eta < \frac{1}{2}$,*

$$\sup_{B((1-\delta)R) \times [\tau, T]} g^p \leqq C_{14} \frac{V(K, 2R)}{V(R)} (R\sqrt{K} \coth(R\sqrt{K}) + 1) \exp(C_{15} K T)$$

$$\times \left(\frac{1}{\delta R} + \frac{1}{\sqrt{\eta \tau}} \right)^{m+2} \int_{(1-\eta)\tau}^{T} ds \int_{B(R)} g^p(y, s)\, dy ,$$

where $V(K, r)$ is the volume of the geodesic ball of radius r in the model space of constant sectional curvature $-\dfrac{K}{m-1}$, and $V(r)$ is the volume of the geodesic ball of radius r centered at x_0.

Proof. For any $0 < \delta, \eta < \frac{1}{4}$, let $\tilde{\phi}$ and $\tilde{\psi}$ be two smooth functions on $[0, \infty)$ so that $0 \leqq \tilde{\phi}, \tilde{\psi} \leqq 1$ and

$$\tilde{\phi}(s) = \begin{cases} 1, & 0 \leqq s \leqq 1 - \delta; \\ 0, & s \geqq 1, \end{cases}$$

$$\tilde{\psi}(s) = \begin{cases} 0, & 0 \leqq s \leqq 1 - 2\eta; \\ 1, & s \geqq 1 - \eta. \end{cases}$$

We can choose $\tilde{\phi}$ and $\tilde{\psi}$ such that,

$$-\frac{C_1}{\delta} \leqq \frac{\tilde{\phi}'}{\tilde{\phi}^{1/2}} \leqq 0\,,$$

$$\tilde{\phi}'' \geqq -\frac{C_1}{\delta^2}\,,$$

and

$$0 \leqq \tilde{\psi}' \leqq \frac{C_1}{\eta}\,,$$

for some constant $C_1 > 0$ which is independent of δ and η.

Let $r(x)$ denote the distance function from x_0, we defined $\phi(x) = \tilde{\phi}\left(\frac{r(x)}{R}\right)$, and $\psi(t) = \tilde{\psi}\left(\frac{t}{\tau}\right)$. Since g is a nonnegative subsolution of the heat equation, for $p > 1$, we have

$$\begin{aligned}\left(\Delta - \frac{\partial}{\partial t}\right)(\phi(x)\psi(t)g^p(x,t)) &\geqq (\Delta\phi)\psi g^p + 2p\psi g^{p-1}\langle \nabla\phi, \nabla g\rangle \\ &\quad + p(p-1)\phi\psi g^{p-2}|\nabla g|^2 - \phi\psi' g^p \\ &\geqq (\Delta\phi)\psi g^p - \frac{p}{p-1}\psi g^p \frac{|\nabla\phi|^2}{\phi} - \phi\psi' g^p\,.\end{aligned}$$

On the other hand, a standard comparison theorem in Riemannian geometry asserts that

$$\left(\Delta\phi - \frac{p}{p-1}\frac{|\nabla\phi|^2}{\phi}\right) \geqq -C_2\left(\frac{R\sqrt{K}\coth(R\sqrt{K}) + 1}{\delta^2 R^2}\right), \tag{1.1}$$

for some constant C_2 depending only on p and m.

If $H(x, y, t)$ is the heat kernel of M, then for $x \in B((1-2\delta)R)$ and $\tau < t < T$, there exists a constant C_3 depending only on m and p so that,

$$\begin{aligned} g^p(x,t) &= g^p(x,t)\phi(x)\psi(t) \\ &= -\int_0^t ds \int_M H(x, y, t-s)\left(\Delta - \frac{\partial}{\partial t}\right)(g^p(y,s)\phi(y)\psi(s))\,dy \end{aligned}$$

$$(1.2)\quad = -\int_{(1-2\eta)\tau}^{t} ds \int_{B(R)-B((1-\delta)R)} H(x,y,t-s)\left(\Delta - \frac{\partial}{\partial t}\right)(g^p(y,s)\phi(y)\psi(s))\,dy$$

$$\leqq C_3\left\{\left(\frac{R\sqrt{K}\coth(R\sqrt{K})+1}{\delta^2 R^2}\right)\int_{(1-2\eta)\tau}^{T} ds \int_{B(R)-B((1-\delta)R)} H(x,y,t-s)\right.$$

$$\left. g^p(y,s)\,dy + \frac{1}{\eta\tau}\int_{(1-2\eta)\tau}^{(1-\eta)\tau} ds \int_{B(R)} H(x,y,t-s)g^p(y,s)\,dy\right\}.$$

Here we have used (1.1) and the definitions of ϕ and ψ.

In [L-Y] (see Theorem 3.3), Li-Yau derived local estimates on fundamental solutions of parabolic Schrödinger equations. By setting the potential to be identically zero and the sets $S_1 = B_y(\sqrt{t-s})$ and $S_2 = B_x(\sqrt{t-s})$, for any $x, y \in B(R)$ and $t - s \leqq \frac{1}{4}R^2$, we have

$$H(x,y,t-s) \leqq C_4 V_x^{-1/2}(\sqrt{t-s})\, V_y^{-1/2}(\sqrt{t-s})$$
$$\exp\left(-C_5\frac{r^2(x,y)}{t-s} + C_6 K(t-s)\right),$$

where C_4, C_5 and C_6 are constants depending only on m, $V_x(\sqrt{t-s})$ is the volume of the geodesic ball of radius $\sqrt{t-s}$ centered at x, and $r(x,y)$ is the distance between x and y.

On the other hand by the volume comparison theorem, for $x \in B(R)$ and $0 < s < t \leqq T \leqq \frac{1}{4}R^2$,

$$\frac{V_{x_0}(R)}{V_x(\sqrt{t-s})} \leqq \frac{V_x(2R)}{V_x(\sqrt{t-s})} \leqq \frac{V(K,2R)}{V(K,\sqrt{t-s})} \leqq C_7(t-s)^{-m/2}V(K,2R),$$

for some constant C_7 depending only on m. Hence

$$(1.3)\quad H(x,y,t-s) \leqq C_8\frac{V(K,2R)}{V(r)}(t-s)^{-m/2}\exp\left(-C_5\frac{r^2(x,y)}{t-s} + C_6K(t-s)\right)$$

$$\leqq C_9(\delta R)^{-m}\frac{V(K,2R)}{V(R)}\exp(C_6K(t-s)),$$

for $x \in B((1-2\delta)R)$, $y \in B(R) - B((1-\delta)R)$ and $0 < s \leqq t \leqq T \leqq \frac{1}{4}R^2$, where C_8 and C_9 are constants depending only on m.

Also for $t > \tau$, and $s \leqq (1-\eta)\tau$, we can find a constant C_{10} depending only on m so that

$$(1.4)\qquad H(x,y,t-s) \leqq C_{10}(\eta\tau)^{-m/2}\exp(C_6K(t-s))\,.$$

Combining (1.2)–(1.4), we conclude that

$$(1.5)\quad \sup_{B((1-2\delta)R)\times[\tau,T]} g^p \leqq C_{11}\frac{V(K,2R)}{V(R)}(R\sqrt{K}\coth(R\sqrt{K})+1)\exp(C_{12}KT)$$
$$\times\left(\frac{1}{\delta R}+\frac{1}{\sqrt{\eta\tau}}\right)^{m+2}\int_{(1-2\eta)\tau}^{T} ds \int_{B(R)} g^p(y,s)\,dy\,,$$

for $0 < \tau < T \leqq \frac{1}{4}R^2$, $0 < \delta, \eta \leqq \frac{1}{4}$ and $p > 1$, where C_{11} and C_{12} depend only on m and p. This completes the proof of the theorem for the case when $p > 1$ if we replace 2δ by δ and 2η by η.

In order to prove the case when $p \leqq 1$, we can proceed as in [L-S]. Take $p = 2$ in (1.5), we have

$$(1.6)\quad \sup_{B((1-\delta)R)\times[\tau,\,T]} g^2 \leqq C_{11}\frac{V(K,2R)}{V(R)}(R\sqrt{K}\coth(R\sqrt{K})+1)\exp(C_{12}KT) \times\left(\frac{1}{\delta R}+\frac{1}{\sqrt{\eta\tau}}\right)^{m+2}\int_{(1-\eta)\tau}^{T} ds\int_{B(R)} g^2(y,s)\,dy\,,$$

for $0 < \delta, \eta < \frac{1}{2}$, and $T \leqq \frac{1}{4}R^2$.

For $k \geqq 0$ and $l \geqq 1$, let

$$M(k,l) = \sup_{B((1-\delta_k)R)\times[\tau_l,\,T]} g^2\,,$$

where $\delta_k = 2^{-k}\delta$, and $\tau_l = (1-\eta\sum_{i=1}^{l}2^{-i})\tau$.

Inequality (1.6) implies that

$$M(k,l) \leqq C_{11}\frac{V(K,2R)}{V(R)}(R\sqrt{K}\coth(R\sqrt{K})+1)\exp(C_{12}KT) \times\left(\frac{2^{k+1}}{\delta R}+\frac{2^{(l+1)/2}}{\sqrt{\eta\tau}}\right)^{m+2}(M(k+1,l+1))^{\lambda}\int_{(1-\eta)\tau}^{T} ds\int_{B(R)} g^p(y,s)\,dy\,,$$

where $\lambda = 1-\dfrac{p}{2}$.

Let $\tilde{M}(k) = M(k, 2k+1)$, then

$$\tilde{M}(k) \leqq A(2^{m+2})^{k+1}\cdot(\tilde{M}(k+1))^{\lambda}\,,$$

where

$$A = C_{11}\frac{V(K,2R)}{V(R)}(R\sqrt{K}\coth(R\sqrt{K})+1)\exp(C_{12}KT)\left(\frac{1}{\delta R}+\frac{1}{\sqrt{\eta\tau}}\right)^{m+2} \times\int_{(1-\eta)\tau}^{T} ds\int_{B(R)} g^p(y,s)\,dy\,.$$

Iterating as in [L-S], we get

$$\tilde{M}(0) \leqq C_{13}A^{1/(1-\lambda)}\,,$$

where C_{13} is a constant depending only on m and p. From this, it is easy to conclude that the theorem is also true for $0 < p \leqq 1$.

Using a similar argument, taking the initial data into consideration, we can obtain another form of the mean-value inequality for positive subsolutions of the heat equation.

Theorem 1.2 *With the same notations and the same assumptions as in Theorem* 1.1, *for any* $p > 0$ *and* $\varepsilon > 0$, *there exists positive constants* C_{16} *and* C_{17} *depending only on* m *and* p *with* C_{16} *depending also on* ε, *such that for any* $0 < \delta < \frac{1}{2}$,

$$\sup_{B((1-\delta)R)\times[0,T]} g^p \leqq C_{16}\left\{\frac{1}{(\delta R)^{m+2}}\frac{V(K,2R)}{V(R)}(R\sqrt{K}\coth(R\sqrt{K})+1)\right.$$
$$\left.\exp(C_{17}KT)\right\}\times\int_0^T ds\int_{B(R)} g^p(y,s)\,dy + ((1+\varepsilon)S(R))^p,$$

where $S(r) = \sup_{B(r)} g(\cdot, 0)$.

Proof. Let ϕ be a cut-off function with support in $B(R)$, then for $p > 1$ and for $x \in B(R)$ and $0 < t < T$, we have

$$\begin{aligned} g^p(x,t) &= g^p(x,t)\phi(x) \\ &= -\int_0^t ds\int_M H(x,y,t-s)\left(\Delta - \frac{\partial}{\partial s}\right)(g^p(y,s)\phi(y))\,dy \\ &\quad + \int_M H(x,y,t)g^p(y,0)\phi(y)\,dy\,. \end{aligned}$$

Hence, using a similar argument as before, for $p > 1$, $T \leqq \frac{1}{4}R^2$ and $0 < \delta < \frac{1}{2}$, we have

$$\sup_{B((1-\delta)R)\times[0,T]} g^p \leqq \frac{C_{18}}{(\delta R)^{m+2}}\frac{V(K,2R)}{V(R)}(R\sqrt{K}\coth(R\sqrt{K})+1)\exp(C_{19}KT)$$
$$\times\int_0^T ds\int_{B(R)} g^p(y,s)\,dy + (S(R))^p,$$

for some constants C_{18} and C_{19} depending only on m and p, where we have used the fact that $H(x,y,t)$ is positive and $\int_M H(x,y,t)\,dy = 1$.

Now consider the case when $0 < p \leqq 1$. For $k \geqq 0$, let $\delta_k = 2^{-k}\delta$, and let $M(k)$ to be the supremum of g^2 on $B((1-\delta_k)R)\times[0,T]$. If $\lambda = 1 - \frac{p}{2}$, then

$$M(k) \leqq A2^{(m+2)(k+1)}(M(k+1))^\lambda + (S(R))^2\,, \tag{1.6}$$

for all $k \geqq 0$, where

$$A = \frac{C_{18}}{(\delta R)^{m+2}}\frac{V(K,2R)}{V(R)}(R\sqrt{K}\coth(R\sqrt{K})+1)\exp(C_{19}KT)\int_0^T ds\int_{B(R)} g^p(y,s)\,dy\,.$$

If $M(0) \leqq ((1+\varepsilon)S(R))^2$, then

$$\sup_{B((1-\delta)R)\times[0,T]} g^p \leqq ((1+\varepsilon)S(R))^p\,, \tag{1.7}$$

and the theorem is proved. We may assume that $M(0) > ((1+\varepsilon)S(R))^2$. Since $M(k)$ is nondecreasing in k, so $M(k) > ((1+\varepsilon)S(R))^2$ for all k. By (1.6), we have

$$M(k) \leqq \frac{(1+\varepsilon)^2}{\varepsilon(2+\varepsilon)}A2^{(m+2)(k+1)}(M(k+1))^\lambda\,.$$

Iterating as in the proof of Theorem 1.1 and taking into account of (1.7), the result follows.

Now we apply Theorem 1.2 to prove some uniqueness results for L^p solutions of the heat equation on manifolds.

Theorem 1.3 *Let M^m be a complete Riemannian manifold satisfying either one of the following conditions:*

(a) M has nonnegative Ricci curvature; or

(b) The Ricci curvature of M is bounded below by $-K \leqq 0$ and $\inf_M V_x(1) > 0$ where $V_x(1)$ is the volume of the geodesic ball of radius 1 with center at x.

If $g(x, t)$ is a nonnegative subsolution of the heat equation on $M \times [0, T]$ such that $g \in L^p(M \times [0, T])$ for some $p > 0$, then

$$\sup_{M \times [0, T]} g \leqq \sup_{x \in M} g(x, 0) .$$

In particular, any $L^p(M \times [0, T])$ solution of the heat equation is uniquely determined by its initial data.

Proof. First we remark that for $p \geqq 1$, the theorem was proved by the first author in [L]. In fact, if $p > 1$, then we do not need any curvature assumption on M.

Now let g be a nonnegative subsolution of the heat equation. If M satisfies (a), then by Theorem 1.2, for a fixed point $x_0 \in M$ and for any $\varepsilon > 0$, there exists a constant C_{20} depending only on m, p and ε, such that for all $R > \sqrt{T}$,

$$\sup_{B_{x_0}(1/2R) \times [0, T]} g^p \leqq \frac{C_{20}}{R^2 V_{x_0}(R)} \int_0^T ds \int_{B_{x_0}(R)} g^p(y, s)\, dy + ((1 + \varepsilon)S)^p ,$$

where $S = \sup_{x \in M} g(x, 0)$. A theorem of Yau in [Y1] shows that

$$V_{x_0}(R) \geqq C_{21} R ,$$

for some constant $C_{21} > 0$. By letting $R \to \infty$ and then $\varepsilon \to 0$, we conclude that

$$\sup_{M \times [0, T]} g \leqq \sup_{x \in M} g(x, 0) .$$

Suppose M satisfies (b). By the semigroup property of $g(x, t)$, we may assume that $T < \frac{1}{4}$. Taking $R = 1$ in Theorem 1.2, for any $\varepsilon > 0$ there exists a constant $C_{22} > 0$, not depending on x and t, such that

$$g^p(x, t) \leqq C_{22} \int_0^T ds \int_{B_x(1)} g^p(y, s)\, dy + ((1 + \varepsilon)S)^p ,$$

for all $x \in M$ and $0 \leqq t \leqq T$. Here we have used the fact that $\inf_M V_x(1) > 0$. Since $g \in L^p(M \times [0, T])$, we conclude that

$$\limsup_{x \to \infty} g(x, t) \leqq (1 + \varepsilon) \sup_{x \in M} g(x, 0)$$

uniformly on $t \in [0, T]$. By the maximum principle and by letting $\varepsilon \to 0$, the result follows.

Theorem 1.4 *Let M^m be a complete noncompact Riemannian manifold. There exists a constant $\delta(m) > 0$ depending only on m, such that if for some point $x_0 \in M$, the Ricci curvature satisfies*

$$\mathrm{Ric}_M(x) \geqq -\delta(m) r_0^{-2}(x)$$

whenever the distance from x to x_0, $r_0(x)$, is sufficiently large, then any nonnegative subsolution $g \in L^p(M \times [0, T])$, for some $p > 0$, of the heat equation on $M \times [0, T]$ must satisfy

$$\sup_{M \times [0, T]} g \leqq \sup_{x \in M} g(x, 0) .$$

Proof. Let $\beta > \dfrac{2}{2^{-m} - 1} > 1$ be a fixed number. For any positive integer k, let us define

$$R_k = 1 + 2\beta + \beta^2 + \ldots + 2\beta^{k-1} + \beta^k = 2 \sum_{i=0}^{k} \beta^i - 1 - \beta^k .$$

By the proof of Theorem 2.5 in [L-S], there exists $\delta(m) > 0$ depending only on m, such that if the Ricci curvature of M satisfies the condition of the theorem then there exists a constant $C_{23} > 0$ depending only on m so that for all $x \in \partial B_{x_0}(R_k)$

$$V_x\left(\frac{\beta^k}{4}\right) \geqq C_{23} \left(\frac{\beta^m}{(\beta + 2)^m - \beta^m}\right)^k V_{x_0}(1) \tag{1.8}$$

for all k.

Let the lower bound of the Ricci curvature of $B_x(r)$ be $-K(x, r) \leqq 0$. Let $g \in L^p(M \times [0, T])$ be a nonnegative subsolution of the heat equation on $M \times [0, T]$. By Theorem 1.2, given $\varepsilon > 0$, for all $x \in \partial B_{x_0}(R_k)$, $k > 0$, and $0 \leqq t \leqq T$, we have

$$g^p(x, t) \leqq C_{24} \frac{1}{\beta^{k(m+2)}} \times \frac{V_x(K(x, \frac{1}{2}\beta^k), \frac{1}{2}\beta^k)}{V_x(\frac{1}{4}\beta^k)}$$

$$\times \left(\frac{1}{2}\beta^k \sqrt{K\left(x, \frac{1}{2}\beta^k\right)} \coth\left(\frac{1}{2}\beta^k \sqrt{K\left(x, \frac{1}{2}\beta^k\right)}\right) + 1\right) + ((1 + \varepsilon)S)^p \tag{1.9}$$

for some constant C_{24} depending only on $\int_0^T \int_M g^p$, m, p, K, T, and ε, where $S = \sup_{x \in M} g(x, 0)$.

By the assumption on the Ricci curvature, for k sufficiently large, we have

$$K\left(x, \frac{1}{2}\beta^k\right) \leqq C_{25} \left(2 \sum_{i=0}^{k} \beta^i - 1 - \beta^k\right)^{-2} \leqq (C_{26}\beta^k)^{-2} \tag{1.10}$$

for some constants C_{25}, $C_{26} > 0$ depending only on m, and

$$V_x\left(K\left(x, \frac{1}{2}\beta^k\right), \frac{1}{2}\beta^k\right) \leqq C_{27} \int_0^{1/2\beta^k} \left(C_{26}\beta^k \sinh\left(\frac{t}{C_{26}\beta^k}\right)\right)^{m-1} dt$$

$$\leqq C_{28}\beta^{mk} \tag{1.11}$$

for some C_{27} and C_{28} depending only on m. From (1.8)–(1.11), it is easy to see that

$$\lim_{k\to\infty} \sup_{\partial B_{x_0}(R_k)\times[0,\,T]} g \leqq (1+\varepsilon)S\,.$$

The theorem then follows from the maximum principle and by letting $\varepsilon \to 0$.

2 Gradient estimates

In this section, we will derive various estimates for solutions of the homogeneous and the inhomogeneous heat equations on a complete manifold. We will also estimate the rate of convergence of any solution of the heat equation to the initial data as $t \to 0$. Other than their application to harmonic maps, these results are also interesting in their own right.

Lemma 2.1 *Let M^m be a complete noncompact Riemannian manifold with Ricci curvature bounded from below by $-K$. Let $H(x, y, t)$ be the heat kernel of M. For any $T > 0$, there exists a constant $C_1(m, K, T)$ so that for $0 < t < T$ and $p, q \in M$,*

$$\int_M |H(p, y, t) - H(q, y, t)|\, dy \leqq C_1\left(\frac{r(p,q)}{\sqrt{t}}\right),$$

where $r(p, q)$ is the distance between p and q. When $K = 0$, the constant C_1 can be taken to depend only on m alone.

Proof. Let $1 < \alpha < 4$ be fixed. For $p, q \in M$ with $p \neq q$ and $\delta > 0$ let us consider

$$\begin{aligned}(2.1)\qquad \int_M |H(p, y, t) - H(q, y, t)|\, dy &\leqq \int_M |H(q, y, (1+\delta)t) - H(q, y, t)|\, dy \\ &\quad + \int_M |H(p, y, t) - H(q, y, (1+\delta)t|\, dy \\ &= I + II\,.\end{aligned}$$

To estimate I, we use the Harnack inequality of [L-Y] which asserts that

$$H(q, y, t) \leqq (1+\delta)^{m\alpha/2} \exp(A\delta t) H(q, y, (1+\delta)t)\,,$$

where $A > 0$ is a constant depending only on m, K and α. Hence

$$\begin{aligned} I &\leqq \int_M |(1+\delta)^{m\alpha/2} \exp(A\delta t) H(q, y, (1+\delta)t) - H(q, y, t)|\, dy \\ &\quad + \int_M |\{(1+\delta)^{m\alpha/2} \exp(A\delta t) - 1\} H(q, y, (1+\delta)t)|\, dy \\ &= \int_M \{(1+\delta)^{m\alpha/2} \exp(A\delta t) H(q, y, (1+\delta)t) - H(q, y, t)\}\, dy \\ &\quad + \{(1+\delta)^{m\alpha/2} \exp(A\delta t) - t\} \int_M H(q, y, (1+\delta)t)\, dy\,.\end{aligned}$$

Using the facts that $H(x, y, t) > 0$ and $\int_M H(x, y, t)\, dy = 1$ for all x and t (see [Y2] and [K-L]), this inequality can be written as

$$(2.2)\qquad I \leqq 2((1+\delta)^{m\alpha/2} \exp(A\delta t) - 1)\,.$$

Similarly, by the Harnack inequality of [L-Y], we have

$$H(p, y, t) \leqq (1+\delta)^{m\alpha/2} \exp\left(A\delta t + \frac{\alpha}{4\delta t} r^2(p, q)\right) H(q, y, (1+\delta)t) .$$

Hence

$$\begin{aligned} II &\leqq \int_M |H(p, y, t) - (1+\delta)^{m\alpha/2} \exp(A\delta t) H(q, y, (1+\delta)t)| \, dy \\ &\quad + \int_M |((1+\delta)^{\alpha/4\delta t} r^2(p, q) - 1) H(q, y, (1+\delta)t)| \, dy \\ &\leqq 2\left((1+\delta)^{m\alpha/2} \exp\left(A\delta t + \frac{\alpha}{4\delta t} r^2(p, q)\right) - 1\right) . \end{aligned} \tag{2.3}$$

Combining (2.1)–(2.3), we have

$$\int_M |H(p, y, t) - H(q, y, t)| \, dy \leqq 4\left\{(1+\delta)^{m\alpha/2} \exp\left(A\delta t + \frac{\alpha}{4\delta t} r^2\right) - 1\right\}, \tag{2.4}$$

where $r = r(p, q)$. Let us first assume that $\dfrac{r}{\sqrt{t}} \leqq 1$. By setting $\delta = \dfrac{r\sqrt{\alpha}}{2\sqrt{t}}$, we have $\delta^2 \leqq \dfrac{\alpha}{4} < 1$, because $1 < \alpha < 4$. For $0 < t < T$,

$$\begin{aligned} (1+\delta)^{m\alpha/2} \exp\left(A\delta t + \frac{\alpha}{4\delta t} r^2\right) - 1 &\leqq (1+\delta)^{m\alpha/2} \exp((AT+1)\delta) - 1 \\ &\leqq C_2(m, \alpha, K, T) \times \delta \\ &= C_2(m, \alpha, K, T)\left(\frac{r}{\sqrt{t}}\right), \end{aligned}$$

where we have used the fact that $\delta \leqq 1$. Combining this with (2.4), we have

$$\int_M |H(p, y, t) - H(q, y, t)| \, dy \leqq C_3(m, \alpha, K, T)\left(\frac{r}{\sqrt{t}}\right) . \tag{2.5}$$

On the other hand if $\dfrac{r(p, q)}{\sqrt{t}} \geqq 1$, then

$$\int_M |H(p, y, t) - H(q, y, t)| \, dy \leqq 2 \leqq 2\left(\frac{r(p, q)}{\sqrt{t}}\right) .$$

The proof of the theorem is now complete. If $K = 0$, then the constant A in the proof can be taken to be 0, and the constant C_1 will depend only on m.

Theorem 2.2 *Let $f: M \times [0, T) \to \mathbf{R}$ be a bounded function, where M satisfies the hypotheses of Theorem* 2.1. *Let $H(x, y, t)$ be the heat kernel of M, and*

$$u(x, t) = -\int_0^t d\tau \int_M H(x, y, t-\tau) f(y, \tau) \, dy .$$

Then u is Lipschitz in the space variable. More precisely, for any $T > 0$, there exists a constant $C_4 = C_4(m, K, T)$ such that

$$\sup_{\substack{x, x' \in M,\ x \neq x' \\ \tau \in (0, t)}} \frac{|u(x, \tau) - u(x', \tau)|}{r(x, x')} \leqq C_4 \left(\sup_{M \times (0, t)} |f| \right) t^{1/2} . \tag{2.6}$$

When $K = 0$, the constant C_4 can be taken to depend on m alone.

Proof. By Theorem 2.1, for x and $x' \in M$, we have

$$\begin{aligned} |u(x, t) - u(x', t)| &\leqq \int_0^t d\tau \int_M |H(x, y, t - \tau) - H(x', y, t - \tau)| \, |f(y, \tau)| \, dy \\ &\leqq \left(\sup_{M \times (0, t)} |f| \right) \int_0^t d\tau \int_M |H(x, y, t - \tau) - H(x', y, t - \tau)| \, dy \\ &\leqq C_5(m, K, T) \left(\sup_{M \times (0, t)} |f| \right) \int_0^t \frac{r(x, x')}{\sqrt{t - \tau}} d\tau \\ &= 2C_5 r(x, x') \left(\sup_{M \times (0, t)} |f| \right) t^{1/2} . \end{aligned}$$

One can also obtain a similar estimate for the homogeneous heat equation.

Theorem 2.3 *Let M be a complete noncompact manifold as in Theorem 2.1. Suppose $u(x, t)$ is a solution of the heat equation on M with bounded initial data h. Then for any $T > 0$, there exists $C_6(m, K, T)$ such that for $0 < t < T$ and $x \in M$,*

$$|\nabla u(x, t)| \leqq C_6 \left(\sup_M |h| \right) t^{-1/2} .$$

When $K = 0$, the constant C_6 can be taken to depend on m alone.

Proof. For $x, x' \in M$ and $0 < t < T$, by Theorem 2.1, we have

$$\begin{aligned} |u(x, t) - u(x', t)| &\leqq \int_M |H(x, y, t) - H(x', y, t)| h(y) \, dy \\ &\leqq C_6(m, K, T) \left(\sup_M |h| \right) r(x, x') t^{-1/2} , \end{aligned}$$

and the result follows.

Proposition 2.4 *Let M be a complete noncompact Riemannian manifold with Ricci curvature bounded below by $-K$. Let h be a bounded C^1 function on M so that $\sup_M |\nabla h| < \infty$. If*

$$u(x, t) = \int_M H(x, y, t) h(y) \, dy ,$$

then

$$\sup_M |\nabla u(\cdot, t)| \leqq \exp(Kt) \sup_M |\nabla h| ,$$

for all $t > 0$.

Proof. Let us first assume that h has compact support. Then

$$d_x u(x, t) = \int_M H_1(x, y, t)\, dh(y)\,, \tag{2.7}$$

where $H_1(x, y, t)$ is the fundamental solution for the heat equation on 1-forms. By [D-L] and the assumption on h, we conclude that

$$\sup_{M \times (0, t)} |\nabla u| < \infty\,, \tag{2.8}$$

for all $t > 0$, and

$$\lim_{t \to 0} \nabla u(x, t) = \nabla h(x)\,. \tag{2.9}$$

Since u satisfies the heat equation on $M \times (0, \infty)$, Bochner's formula implies that

$$\left(\Delta - \frac{\partial}{\partial t}\right) |\nabla u| \geqq -K |\nabla u|\,.$$

Hence by considering $\exp(-Kt)|\nabla u|(x, t)$, using (2.8), (2.9) and the proof of Theorem 1 in [K-L] (see also [Lo-T]), one concludes that

$$\sup_M \exp(-Kt)|\nabla u|(\cdot, t) \leqq \sup_M |\nabla h|\,.$$

This establishes the proposition for compactly supported h.

In general, let h be a bounded C^1 function with $\sup_M |\nabla h| < \infty$. One can find a sequence of compactly supported smooth functions, h_ν, which are uniformly bounded, converging uniformly on compact subsets of M to h, and

$$\limsup_{\nu \to \infty} \sup_M |\nabla h_\nu| \leqq \sup_M |\nabla h|\,. \tag{2.10}$$

Let $u_\nu(x, t) = \int_M H(x, y, t) h_\nu(y)\, dy$. Then the sequence u_ν is uniformly bounded and are solutions of the heat equation. The Schauder estimates implies that, by passing to a subsequence, u_ν together with their gradients converge uniformly on compact subsets of M. However for each fixed x, $\lim_{\nu \to \infty} u_\nu(x, t) = u(x, t)$. Hence the sequence u_ν converges to u in C^1. Using the result for compactly supported functions and (2.10), the proposition is proved.

Proposition 2.5 *Let M be a complete noncompact manifold and $f: M \times [0, T) \to \mathbf{R}$ is a bounded function. Suppose $H(x, y, t)$ is the minimal fundamental solution of the heat equation of M obtained by taking compact exhaustion. If*

$$u(x, t) = -\int_0^t d\tau \int_M H(x, y, t - \tau) f(y, \tau)\, dy\,,$$

then

$$\sup_{M \times (0, t)} |u| \leqq t \sup_{M \times (0, t)} |f|\,. \tag{2.11}$$

Proof. This follows from the facts that $H(x, y, t) > 0$ and $\int_M H(x, y, t)\, dy \leqq 1$.

Lemma 2.6 *Let M^m be a complete noncompact Riemannian manifold with Ricci curvature bounded from below by $-K \leqq 0$. Let $H(x, y, t)$ be the heat kernel of M. For any $0 < \alpha \leqq 1$ and $T > 0$, there exists a constant $C_7(\alpha, m, K, T)$ such that for $0 < t < T$, we have*

$$\int_M H(x, y, t) r^\alpha(x, y)\, dy \leqq C_7 t^{\alpha/2}\,, \tag{2.12}$$

where $r(x, y)$ is the distance between x and y. When $K = 0$, the constant C_7 can be taken to depend on α and m alone.

Proof. Let $B_x(r)$ be the geodesic ball of radius r will center at x and let $V_x(r)$ be its volume. The integral

$$\begin{aligned}\int_M H(x, y, t) r^\alpha(x, y)\, dy &= \int_{B_x(2\sqrt{t})} H(x, y, t) r^\alpha(x, y)\, dy \\ &\quad + \int_{M - B_x(2\sqrt{t})} H(x, y, t) r^\alpha(x, y)\, dy \\ &\leqq (2\sqrt{t})^\alpha + \int_{M - B_x(2\sqrt{t})} H(x, y, t) r^\alpha(x, y)\, dy\,,\end{aligned} \tag{2.13}$$

where we have used the fact that $H(x, y, t) > 0$ and $\int_M H(x, y, t)\, dy \equiv 1$.

The upper estimates of the heat kernel of [L-Y] implies that, for $0 < t < T$, there exists a constant $C_8(K, T)$ such that

$$H(x, y, t) \leqq C_8 V_x^{-1/2}(\sqrt{t}) V_y^{-1/2}(\sqrt{t}) \exp\left(-\frac{ar^2(x, y)}{t}\right),$$

where $a = \frac{2}{9}$. The volume comparison theorem asserts that, for $r = r(x, y) \geqq 2\sqrt{t}$,

$$V_x(\sqrt{t}) \leqq V_y(r + \sqrt{t}) \leqq V_y(\sqrt{t}) \frac{V(K, r + \sqrt{t})}{V(K, \sqrt{t})}\,,$$

where $V(K, s)$ is the volume of the geodesic ball of radius s in the simply connected space form of constant $-\dfrac{K}{m-1}$ curvature. Hence

$$\begin{aligned}\int_{M - B_x(2\sqrt{t})} H(x, y, t) r^\alpha(x, y)\, dy &\leqq C_8 \int_{2\sqrt{t}}^{\infty} \left(\frac{V(K, r + \sqrt{t})}{V(K, \sqrt{t})}\right)^{1/2} \exp\left(-\frac{ar^2}{t}\right) r^\alpha A_x(r)\, dr \\ &\leqq C_8 V_x^{-1}(\sqrt{t}) \int_{2\sqrt{t}}^{\infty} \left(\frac{V(K, 2r)}{V(K, \sqrt{t})}\right)^{1/2} \\ &\quad \times \exp\left(-\frac{ar^2}{t}\right) r^\alpha A_x(r)\, dr\,,\end{aligned} \tag{2.14}$$

where $A_x(r)$ is the area of $\partial B_x(r)$.

Integrating by parts gives

$$
\begin{aligned}
I &= \int_{2\sqrt{t}}^{\infty} \left(\frac{V(K,2r)}{V(K,\sqrt{t})}\right)^{1/2} \exp\left(-\frac{ar^2}{t}\right) r^{\alpha} A_x(r)\, dr \\
&= \left(\frac{V(K,2r)}{V(K,\sqrt{t})}\right)^{1/2} \exp\left(-\frac{ar^2}{t}\right) r^{\alpha} V_x(r) \Bigg]_{2\sqrt{t}}^{\infty} \\
&\quad - \left(\int_{2\sqrt{t}}^{\infty} V_x(r) \left\{ \alpha r^{\alpha-1} \left(\frac{V(K,2r)}{V(K,\sqrt{t})}\right)^{1/2} \exp\left(-\frac{ar^2}{t}\right) \right.\right. \\
&\quad + r^{\alpha} \frac{A(K,2r)}{(V(K,2r)V(K,\sqrt{t}))^{1/2}} \exp\left(-\frac{ar^2}{t}\right) - \frac{2ar^{1+\alpha}}{t} \left(\frac{V(K,2r)}{V(K,\sqrt{t})}\right)^{1/2} \\
&\quad \left.\left. \times \exp\left(-\frac{ar^2}{t}\right) \right\} dr \right) \\
&\leqq \int_{2\sqrt{t}}^{\infty} \frac{2ar^{1+\alpha}}{t} \left(\frac{V(K,2r)}{V(K,\sqrt{t})}\right)^{1/2} \exp\left(-\frac{ar^2}{t}\right) V_x(r)\, dr\,,
\end{aligned}
$$

where we have used the volume comparison theorem and the fact that

$$V(K,r) \leqq C_9 \exp((m-1)\sqrt{K}r)\,.$$

Letting $s = \dfrac{r}{\sqrt{t}}$, and using the volume comparison theorem again,

$$
\begin{aligned}
I &\leqq t^{\alpha/2} \int_2^{\infty} 2\alpha s^{1+\alpha} \left(\frac{V(K,2s\sqrt{t})}{V(K,\sqrt{t})}\right)^{1/2} \exp(-as^2) V_x(s\sqrt{t})\, ds \\
&\leqq 2\alpha t^{\alpha/2} V_x(\sqrt{t}) \int_2^{\infty} s^{1+\alpha} \left(\frac{V(K,2s\sqrt{t})}{V(K,\sqrt{t})}\right)^{3/2} \exp(-as^2)\, ds\,.
\end{aligned} \tag{2.15}
$$

However, for $s \geqq 2$ and $0 < t < T$,

$$
\begin{aligned}
\frac{V(K,2s\sqrt{t})}{V(K,\sqrt{t})} &= \frac{\int_0^{2s\sqrt{t}} (\sinh(\sqrt{K}\tau)^{m-1}\, d\tau}{\int_0^{\sqrt{t}} (\sinh(\sqrt{K}\tau))^{m-1}\, d\tau} \\
&\leqq C_{10} s^m \exp(C_{11} s)\,,
\end{aligned} \tag{2.16}
$$

where C_{10} and C_{11} are constants depending only on m, K, and T. Inequality (2.12) now follows from (2.13)–2.16).

Theorem 2.7 *Let M be as in Theorem* 2.6. *Suppose h is a bounded continuous function on M, such that*

$$[h]_{\alpha, M} = \sup_{\substack{x, x' \in M \\ x \neq x'}} \frac{|h(x) - h(x')|}{r^{\alpha}(x, x')} < \infty\,, \text{ for some } 0 < \alpha \leqq 1\,.$$

For any $T > 0$, there exists a constant $C_{12}(\alpha, m, K, T)$ so that for $0 < t < T$, if

$$u(x,t) = \int_M H(x,y,t) h(y)\, dy\,,$$

then

$$|u(x, t) - h(x)| \leqq C_{12}[h]_{\alpha, M} t^{\alpha/2} .$$

When $K = 0$, the constant C_{12} can be taken to depend on α and m alone.

Using Theorem 2.3, Theorem 2.7 and the semigroup property of the heat equation, the next theorem can be proved similarly.

Theorem 2.8 *Let M be a complete noncompact Riemannian manifold with Ricci curvature bounded below by $-K \leqq 0$. Let $u(x, t)$ be a solution of the heat equation on $M \times [0, T)$ with initial condition h. Then there exists a constant $C_{13}(m, K, T)$ such that for $x \in M$ and $0 < t < t' < T$,*

$$|u(x, t') - u(x, t)| \leqq C_{13}\left(\sup_M |h|\right)\left(\frac{t' - t}{t}\right)^{1/2} .$$

When $K = 0$, the constant C_{13} can be taken to depend on m alone.

3 Short time solutions of the parabolic harmonic map equation

We will now apply the results in §1 and §2 to study harmonic maps on noncompact manifolds. Let M^m and N^n be complete Riemannian manifolds with metrics $g_{ij}dx^i dx^j$ and $h_{\alpha\beta}du^\alpha du^\beta$, respectively. During the course of this discussion we will always implicitly assume that M is noncompact. Suppose $h: M \to N$ is a C^1 map so that $h(M)$ is bounded in N, then there exists an open neighborhood N' of $h(M)$ with compact closure so that N' can be embedded into $\mathbf{R}^q$ isometrically for some q. If necessary, by choosing a smaller neighborhood, we may assume that there is a bounded tubular neighborhood $\tilde{N}$ of N' in $\mathbf{R}^q$ exists. Let $\pi: \tilde{N} \to N'$ be the nearest point projection denoted by $\pi = (\pi^1, \ldots, \pi^q) = (\pi^A)_{1 \leqq A \leqq q}$. By choosing an even smaller N', we may assume that π can be extended smoothly to the whole $\mathbf{R}^q$ so that each π^A is compactly supported. Hence, in particular, π^A, $\pi^A_B = \dfrac{\partial \pi^A}{\partial z^B}$, $\pi^A_{BC} = \dfrac{\partial^2 \pi^A}{\partial z^B \partial z^C}$, etc. are bounded, where $z = (z^A)$ are the standard coordinates of $\mathbf{R}^q$. Let us consider the following system of differential equations for a map $u: M \times [0, T) \to \mathbf{R}^q$ with $u = (u^A)$, given by

$$\left(\Delta - \frac{\partial}{\partial t}\right)u^A = \pi^A_{BC}(u)\frac{\partial u^B}{\partial x^i}\frac{\partial u^C}{\partial x^j}g^{ij} \tag{3.1}$$

$$= \pi^A_{BC}(u)\langle \nabla u^B, \nabla u^C \rangle$$

on $M \times (0, T)$ for $A = 1, \ldots, q$, satisfying the initial condition

$$u(x, 0) = h(x), \qquad x \in M . \tag{3.2}$$

Here and from here on, the summation convention is taken on repeated indices.

Lemma 3.1 *Suppose the image of u lies in N', then u satisfies the parabolic equation for harmonic maps from $M \times (0, T) \to N$, if and only if u satisfies* (3.1).

Proof. See [E-S, p. 140].

Lemma 3.2 *Suppose there exists a point $p \in M$ and a constant $C > 0$ so that $V_p(r) \leqq \exp(C(1 + r^2))$ for all r. Let u be a solution of (3.1) which is continuous on $M \times [0, T)$ with $u(x, 0) \in N$ for all $x \in M$. Suppose $u(x, t) \in \tilde{N}$ on $M \times [0, T)$, then $u(x, t) \in N$ for all $(x, t) \in M \times [0, T)$.*

Proof. For $z \in \mathbf{R}^q$, let us define $\rho : \mathbf{R}^q \to \mathbf{R}^q$ by $\rho(z) = z - \pi(z)$. The function $\rho(u(x, t))$ is bounded on $M \times [0, T)$ because $\tilde{N}$ is bounded. Consider

$$\varphi(x, t) = \sum_{A=1}^{q} (\rho^A(u(x, t))^2$$

$$= \text{the square of the distance from } u(x, t) \text{ to } N'.$$

By [E-S, p. 141], we have

$$\left(\Delta - \frac{\partial}{\partial t}\right)\varphi(x, t) = 2 \sum_{A} |\nabla \rho^A(u(x, t))|^2$$

$$\geqq 0 .$$

Since $\varphi \geqq 0$ is bounded and $\varphi(x, 0) = 0$, the proof of Theorem 1 in [K-L] (also see [Lo-T]) implies that $\varphi \equiv 0$ on $M \times (0, T)$. This establishes the lemma.

In what follows, we will prove short time existence for the initial value problem of (3.1) and (3.2).

Theorem 3.3 *Let M be a complete noncompact Riemannian manifold with Ricci curvature bounded from below by $-K \leqq 0$. Given $h \in C^1(M, N)$ so that $h(M)$ is bounded and the energy density of h is also bounded. Then there exists $T_0 > 0$, such that (3.1) and (3.2) has a solution on $M \times [0, T_0)$.*

Proof. We will follow the argument of [E-S] on successive approximation. Let us set

$$\Lambda = \sup_{\mathbf{R}^q, A, B, C, D} (|\pi^A_{BC}|, |\pi^A_{BCD}|) .$$

The hypotheses on h imply that $\sup_M |h^A| < \infty$, and $\sup_M |\nabla h^A| < \infty$ for $1 \leqq A \leqq q$. For $\nu = -1, 0, 1, 2, \ldots$, let us define inductively a sequence of maps $u^\nu : M \times [0, 1) \to \mathbf{R}^q$ as follows:

$$u^{-1, A} = 0 ; \tag{3.3}$$

$$u^{0, A}(x, t) = \int_M H(x, y, t) h^A(y)\, dy ; \tag{3.4}$$

and

$$u^{\nu, A}(x, t) = -\int_0^t d\tau \int_M H(x, y, t - \tau) F^{\nu - 1, A}(y, \tau)\, dy + u^{0, A}(x, t) \tag{3.5}$$

for $\nu \geqq 1$. Here $H(x, y, t)$ is the heat kernel of M, and

$$F^{\nu, A} = \pi^A_{BC}(u^\nu) \langle \nabla u^{\nu, B}, \nabla u^{\nu, C} \rangle$$

for $v \geqq -1$, $A = 1, \ldots, q$. The function u^{-1} is introduced for the sake of convenience. Note that $F^{-1} = 0$. Obviously, u^{-1} and u^0 are well-defined. Also it is easy to see that u is smooth on $M \times (0, 1)$.

For $0 < t < 1$, let us consider the function

$$p_v(t) = \sup_{M \times (0,\, t)} \left(\sum_A |\nabla u^{v,\, A}|^2 \right)^{1/2} . \tag{3.6}$$

Clearly p_v is nondecreasing in t, and there is a constant $C_1(m, q)$ so that

$$\sup_{M \times (0,\, t)} |F^{v,\, A}| \leqq C_1 p_v^2(t) . \tag{3.7}$$

In order to prove that the functions u^v are well-defined, it suffices to show that $p_v(t) < \infty$ for all v and for all $0 < t < 1$. Let us first consider the case when $v = 0$. By Proposition 2.4 and the definition of u^0, we have

$$p_0(t) \leqq C_2(q) \exp(2Kt) \sup_{M,\, A} |\nabla h^A| , \tag{3.8}$$

where $C_2(q)$ is a constant depending only on q. Hence u^1 is well-defined and is smooth on $M \times (0, 1)$. Suppose u^v is defined, $p_{v-1}(t) < \infty$ for $0 < t < 1$, and u^v is smooth on $M \times (0, 1)$. Theorem 2.2 implies that for $0 < t < 1$,

$$\begin{aligned} \sup_{M \times (0,\, t)} |\nabla u^{v,\, A}| &\leqq C_3(m, K) t^{1/2} \sup_{M \times (0,\, t)} |F^{v-1,\, A}| + p_0(t) \\ &\leqq C_4(m, q, K, \Lambda) t^{1/2} p_{v-1}^2(t) + p_0(t) . \end{aligned}$$

Hence,

$$p_v(t) \leqq C_5(m, q, K, \Lambda) t^{1/2} p_{v-1}^2(t) + p_0(t) . \tag{3.9}$$

By induction hypothesis, we conclude that u^{v+1} is well-defined and is smooth on $M \times (0, t)$.

Now choose $0 < T_1 < 1$ such that

$$C_2 C_5 T_1^{1/2} \exp(2K) \sup_{A,\, M} |\nabla h^A| \leqq \frac{1}{4} . \tag{3.10}$$

By (3.8), we have

$$C_5 T_1^{1/2} p_0(t) \leqq \frac{1}{4} . \tag{3.11}$$

If $C_5 T_1^{1/2} p_{v-1}(t) \leqq \frac{1}{2}$ on $(0, T_1)$, then (3.9) implies that for $0 < t < T_1$,

$$\begin{aligned} C_5 T_1^{1/2} p_v(t) &\leqq (C_5 T_1^{1/2} p_{v-1}(t))^2 + C_5 T_1^{1/2} p_0(t) \\ &\leqq \frac{1}{4} + \frac{1}{4} \\ &\leqq \frac{1}{2} . \end{aligned}$$

Hence one can prove inductively that for all $v \geqq 0$,

$$C_5 T_1^{1/2} p_v(t) \leqq \frac{1}{2} \tag{3.12}$$

on $(0, T_1)$. In particular, $p_v(t)$ are uniformly bounded on $(0, T_1)$. Also by the definition of u^v, inequality (3.7), and the fact that h is bounded, we conclude that u^v are uniformly bounded on $M \times (0, T_1)$. Therefore the following functions X_v and $\bar{X}_v$ are well-defined

$$\begin{aligned} X_v(t) = &\sup_M \sum_A |u^{v,A}(\cdot, t) - u^{v-1,A}(\cdot, t)| \\ &+ \sup_M \left(\sum_A |\nabla u^{v,A}(\cdot, t) - \nabla u^{v-1,A}(\cdot, t)|^2 \right)^{1/2} \end{aligned} \tag{3.13}$$

and

$$\bar{X}_v(t) = \sup_{0<\tau<t} X_v(\tau)\,, \tag{3.14}$$

for $0 < t < T_1$. Now

$$\begin{aligned} F^{v,A} - F^{v-1,A} &= \pi^A_{BC}(u^v) \langle \nabla u^{v,B}, \nabla u^{v,C} \rangle - \pi^A_{BC}(u^{v-1}) \langle \nabla u^{v-1,B}, \nabla u^{v-1,C} \rangle \\ &= (\pi^A_{BC}(u^v) - \pi^A_{BC}(u^{v-1})) \langle \nabla u^{v,B}, \nabla u^{v,C} \rangle \\ &\quad + \pi^A_{BC}(u^{v-1}) \langle \nabla u^{v,B} - \nabla u^{v-1,B}, \nabla u^{v,C} \rangle \\ &\quad + \pi^A_{BC}(u^{v-1}) \langle \nabla u^{v-1,B}, \nabla u^{v,C} - \nabla u^{v-1,C} \rangle\,. \end{aligned}$$

Applying the mean-value theorem to

$$\pi^A_{BC}(u^v) - \pi^A_{BC}(u^{v-1})\,,$$

we have

$$\sup_{M \times (0,t)} |F^{v,A} - F^{v-1,A}| \leqq C_6 \bar{X}_v(t)(p_v^2(t) + p_v(t) + p_{v-1}(t))\,,$$

where $C_6 = C_6(q, \Lambda)$. Inequality (3.12) asserts that for $0 < t < T_1$,

$$\sup_{M \times (0,t)} |F^{v,A} - F^{v-1,A}| \leqq C_7 \bar{X}_v(t)\,, \tag{3.15}$$

where $C_7 = C_7(m, K, q, \Lambda)$. On the other hand, Proposition 2.4 implies that for $0 < t < T_1$ and for $x \in M$,

$$\begin{aligned} |u^{v+1,A}(x, t) - u^{v,A}(x, t)| &\leqq t \sup_{M \times (0,t)} |F^{v,A} - F^{v-1,A}| \\ &\leqq C_7 t \bar{X}_v(t)\,. \end{aligned} \tag{3.16}$$

Using (3.15) and Theorem 2.2,

$$\begin{aligned} |\nabla u^{v+1,A}(x, t) - \nabla u^{v,A}(x, t)| &\leqq C_8(m, K) t^{1/2} \sup_{M \times (0,t)} |F^{v,A} - F^{v-1,A}| \\ &\leqq C_9(m, q, K, \Lambda) t^{1/2} \bar{X}_v(t)\,. \end{aligned} \tag{3.17}$$

Combining (3.16), (3.17) and the definition of $\bar{X}_\nu$, we have

$$\bar{X}_{\nu+1}(t) \leqq C_{10}(m, q, K, \Lambda) t^{1/2} \bar{X}_\nu(t) , \tag{3.18}$$

for $0 < t < T_1$.

Let us now choose $0 < T_0 < 1$ to be sufficiently small so that $C_{10} T_0 < 1$. If $0 < t < T_0$, then by induction we conclude that

$$\begin{aligned} \bar{X}_{\nu+1}(t) &\leqq (C_{10} t^{1/2})^\nu \bar{X}_0(t) \\ &\leqq (C_{10} T_0^{1/2})^\nu \bar{X}_0(t) \\ &\leqq C_{11}(C_{10} T_0^{1/2})^\nu \left(\sup_{M, A} |h^A| + \sup_{M, A} |\nabla h^A| \right) , \end{aligned} \tag{3.19}$$

where C_{11} is a constant depending only on m, K and q. Here we have used Proposition 2.4, and the fact that $|u^{0, A}| \leqq \sup_M |h^A|$ for all A.

Since $C_{10} T_0^{1/2} < 1$, the sequence of functions $\sum_{\nu=1}^{\infty} X_\nu(t)$ converges uniformly on $(0, T_0)$. Hence $u^{\nu, A}$ and $\nabla u^{\nu, A}$ converges uniformly on $M \times (0, T_0)$. Let $u = \lim_{\nu \to \infty} u^\nu$. Then ∇u^A exists and is the uniform limit of $\nabla u^{\nu, A}$. By (3.15) it is easy to see that $F^{\nu, A}$ also converges uniformly to $F^A = \pi^A_{BC}(u) \langle \nabla u^B, \nabla u^C \rangle$ on $M \times (0, T_0)$ for all A. Hence on $M \times (0, T_0)$, we have

$$u^A(x, t) = - \int_0^t d\tau \int_M H(x, y, t - \tau) F^A(u, \tau) dy + \int_M H(x, y, t) h^A(y) dy . \tag{3.20}$$

Note that each u^ν is smooth on $M \times (0, T_0)$, and satisfies $\left(\Delta - \dfrac{\partial}{\partial t} \right) u^{\nu, A} = F^{\nu-1, A}$ for $A = 1, \ldots, q$. By (3.5), (3.7) and (3.12) it is easy to see that F^ν and u^ν are uniformly bounded on $M \times (0, T_0)$. By [L-S-U, p. 211 Theorem 11.1], for any compact subset $\Omega \subset M$ and $0 < t_1 < t_2 < T_0$, there exists $C_{12} > 0$ and $0 < \beta < 1$ independent of ν and A, so that

$$|\nabla u^{\nu, A}(x, t) - \nabla u^{\nu, A}(x', t')| \leqq C_{12}((r(x, x'))^\beta + |t - t'|^{\beta/2})$$

for all $x, x' \in \Omega$ and for all $t_1 < t, t' < t_2$. Letting $\nu \to \infty$, we have

$$|\nabla u^A(x, t) - \nabla u^A(x', t')| \leqq C_{12}((r(x, x'))^\beta + |t - t'|^{\beta/2}) .$$

Hence by (3.20) one can conclude that u satisfies (3.1) in $M \times (0, T_0)$. It is easy to see that $\lim_{t \to 0} u^A(x, t) = h^A(x)$ and the proof of the theorem is completed.

Theorem 3.4 *Let M and N be complete Riemannian manifolds where M is noncompact with Ricci curvature bounded below by $-K \leqq 0$. Let $h \in C^1(M, N)$ be a map with bounded energy density such that $h(M)$ is also bounded. Then there exists $T_0 > 0$ and a map u which satisfies the parabolic equation for harmonic maps on $M \times (0, T_0)$ with initial data h.*

Proof. We embed a bounded nieghborhood N' of $h(M)$ as described in the beginning of this section. By Theorem 3.3, one can solve (3.1) and (3.2) for short

time. Theorem 3.4 then follows from Proposition 2.5, Theorem 2.7, Lemma 3.1 and Lemma 3.2.

Remark 3.1 From the proof of the theorem, it is easy to see that T_0 depends only on M, N, a neighborhood of $h(M)$ and the bound on the energy density of h. Furthermore, the energy density and the image of the solution u is bounded on $M \times [0, T_0)$ by a constant depending only on the known quantities mentioned above. The limit $\lim_{t \to 0} u(x, t) = h(x)$ is uniform on M, and also by (3.20), (2.6) and (2.9) we have $\lim_{t \to 0} e(u)(x, t) = e(h)(x)$ uniformly on compact subsets.

We will now prove a uniqueness theorem for the parabolic equation of harmonic maps.

Theorem 3.5 *Let M and N be complete Riemannian manifolds where M is noncompact. Suppose there exists a point $p \in M$ and a constant $a > 0$ such that $V_p(r) \leqq \exp(a(1 + r^2))$ for all $r > 0$. Let u_1 and u_2 be two solutions of the parabolic equation for harmonic maps from $M \times (0, T)$ to N with the same initial condition. Suppose the images and the energy densities of u_1 and u_2 are bounded on $M \times [0, T)$, then $u_1 = u_2$.*

Proof. By assumptions, there exists a bounded open set N' which contains the images of u_1 and u_2. Using the method of embedding N' in $\mathbf{R}^q$ described in the beginning of this section, Lemma 3.1 implies that u_1 and u_2 satisfy (3.1) in $M \times (0, T)$. Let

$$\varphi = \sum_A (u_1^A - u_2^A)^2 .$$

Then φ satisfies

$$\left(\Delta - \frac{\partial}{\partial t}\right)\varphi = 2\sum_A (u_1^A - u_2^A)\left(\Delta - \frac{\partial}{\partial t}\right)(u_1^A - u_2^A) + 2\sum_A |\nabla(u_1^A - u_2^A)|^2$$

$$= 2 \sum_{A,B,C} (u_1^A - u_2^A)(\pi_{BC}^A(u_1)\langle \nabla u_1^B, \nabla u_1^C\rangle - \pi_{BC}^A(u_2)\langle \nabla u_2^B, \nabla u_2^C\rangle) \tag{3.21}$$

$$+ 2\sum_A |\nabla u_1^A - u_2^A)|^2 .$$

As in the proof of Theorem 3.3, one can compute and show that

$$\left(\Delta - \frac{\partial}{\partial t}\right)\varphi \geqq -C_{13}\left(\sum_A |u_1^A - u_2^A|\right)\left[\sum_B (|u_1^B - u_2^B| + |\nabla(u_1^B - u_2^B)|)\right]$$

$$+ 2\sum_A |\nabla(u_1^A - u_2^A)|^2$$

$$\geqq -C_{14}\varphi ,$$

where C_{13} and C_{14} are constants depending only on q, the upper bounds of $|\pi_{BC}^A|$ and $|\pi_{BCD}^A|$ for $1 \leqq A, B, C, D \leqq q$, and the upper bounds of the energy densities of u_1 and u_2. Since φ is bounded and $\varphi(x, 0) \equiv 0$, by [K-L, Theorem 1], we conclude that $\varphi \equiv 0$ on $M \times [0, T)$. Hence u_1 and u_2 are equal.

4 Harmonic maps: $K = 0$ and $\kappa = 0$

In this section we will show that if the upper bound of the sectional curvature κ of N is 0, then equation (0.1) possesses a smooth solution for all time. If in addition, the Ricci curvature of M is nonnegative, then one can deform the initial map to a harmonic map through the solution of the parabolic equation.

Theorem 4.1 *Let M be a complete Riemannian manifold with Ricci curvature bounded from below by $-K \leqq 0$, and let N be a complete Riemannian manifold with nonpositive sectional curvature. Let $h \in C^1(M, N)$ be such that $h(M)$ is bounded in N and $\sup_M e(h) < \infty$. Then (0.1) has a unique smooth solution u defined on $M \times (0, \infty)$ with $u(x, 0) = h(x)$, which satisfies the properties that:*

(a) $u(M \times [0, T])$ is bounded for all $T > 0$; and

(b) $\sup_{M \times [0, T]} e(u) < \infty$, for all ∞.

Proof. Let T^* be the supremum of those T_1 such that (0.1) has a solution u on $M \times [0, T_1)$ satisfying (a) and (b) for all $T < T_1$. By Theorem 3.4 and Remark 3.5, we have $T^* > 0$. Let us assume that $T^* < \infty$. The assumption that N has nonpositive curvature implies that (see [E-S]) the energy density satisfies

$$\left(\Delta - \frac{\partial}{\partial t}\right) e(u) \geqq -2Ke(u) \tag{4.1}$$

on $M \times (0, T^*)$. By Remark 3.5, we have

$$\lim_{t \to 0} e(u(x, t)) = e(h)(x) \tag{4.2}$$

uniformly on compact sets of M. Inequality (4.1) asserts that the function

$$\exp(-2Kt)e(u(x, t))$$

is a subsolution of the heat equation for functions. By (4.2) and the maximum principle in [L-K] (see also [Lo-T]) we have

$$\exp(-2Kt)e(u(x, t)) \leqq \sup_M e(h) .$$

Hence

$$\sup_{M \times [0, T^*)} e(u) \leqq \exp(2KT^*) \sup_M e(h) . \tag{4.3}$$

On the other hand, by the definition of T^*, $u\left(M \times \left[0, \frac{T^*}{2}\right]\right)$ is in a bounded open set N' of N. Let $\delta > 0$ be the injectivity radius of N'. By Theorem 3.6, Remark 3.5, and the proof of Theorem 3.3, we conclude that as $t \to 0$, $u(x, t)$ converges uniformly on M to $h(x)$. Therefore, there exists $t_0 > 0$ such that $d(u(x, t), h(x)) < \delta$ for all $x \in M$ and $0 < t < t_0$, where d is the distance function of N.

Consider the map $v(x, t) = u(x, t + t_0)$ for $0 \leqq t < T^* - t_0$. Clearly, v also satisfies (0.1) on $M \times [0, T^* - t_0)$ with initial condition $v(x, 0) = u(x, t_0)$. For $x \in M$, let $\sigma(x, s)$ for $0 \leqq s \leqq 1$, be the unique minimal geodesic joining $v(x, 0)$ to

$u(x, 0)$. For $0 < t < T^* - t_0$, let $\sigma(x, t, \cdot)$ be the geodesic parametrized by arclength joining $v(x, t)$ to $u(x, t)$ which is homotopic to the curve

$$f(x, s) = \begin{cases} v(x, t, t - s) , & 0 \leqq s \leqq t ; \\ \sigma(x, s - t) , & t \leqq s \leqq t + 1 ; \\ u(x, s - t - 1) , & t + 1 \leqq s \leqq 2t + 1 . \end{cases}$$

Since N has nonpositive curvature, $\sigma(x, t, \cdot)$ is unique. Let $\rho(x, t)$ be the length of $\sigma(x, t, \cdot)$. Then ρ^2 is smooth on $M \times (0, T^* - t_0)$ (see [S-Y3]) and the function $\varphi = (\rho^2 + 1)^{1/2}$ must satisfy

$$\left(\Delta - \frac{\partial}{\partial t} \right) \varphi \geqq 0$$

and

$$|\nabla \varphi|^2 \leqq 2(e(u) + e(v))$$

on $M \times (0, T^* - t_0)$. Hence by (4.3), the fact that M has at most exponential volume growth and Lemma 3.1 in [Lo-T], we have

$$\sup_M \varphi(\cdot, t) \leqq \sup_M \varphi(\cdot, 0) \leqq (\delta^2 + 1)^{1/2}$$

for all $0 < t < T^* - t_0$. In particular

$$\sup_{M \times [0, T^* - t_0)} d(v, u) \leqq \sup_{M \times [0, T^* - t_0)} \rho \leqq \delta .$$

Since $u\left(M \times \left[0, T^* - \frac{t_0}{2} \right] \right)$ is bounded, by the definition of v and the above inequality, we conclude that $u(M \times [0, T^*))$ is bounded. Using this fact together with (4.3), we can extend u as a solution of (0.1) beyond T^* because of Theorem 3.3 and Theorem 3.6. This contradicts the assumption that $T^* < \infty$. Hence T^* must be ∞.

In fact, we can remove the restriction that the image of the initial data is bounded. First we need the following lemma.

Lemma 4.2 *Let N^n be a complete simply connected manifold with nonpositive curvature. Let p be a fixed point. For any $R > 0$ there exists a smooth map ϕ_R from N to N such that $\phi_R|_{B_p(R)} =$ identity, $\phi_R(M) \subset B_p(2R)$, and $\sup_M e(\phi_R) \leqq C_1$ for some constant C_1 which is independent of R.*

Proof. Note that N is diffeomorphic to $\mathbf{R}^n$ under the exponential map at p. In geodesic polar coordinates at p, the metric on N can be expressed as

$$ds^2 = dr^2 + g_{\alpha\beta}(r, \theta) d\theta^\alpha d\theta^\beta ,$$

where $(\theta^2, \ldots, \theta^n)$ is the local coordinates on $\mathbf{S}^{n-1}$. Here and below, the Greek indices ranges from 2 to n.

For any $R > 0$, let $\tilde{\phi}$ be a real valued function on $[0, \infty)$, such that $0 \leqq \tilde{\phi} \leqq 1$, $\tilde{\phi} = 1$ on $[0, R]$, $\tilde{\phi} = 0$ on $[2R, \infty)$, and $|\tilde{\phi}'| \leqq \dfrac{C_2}{R}$ for some C_2 which is independent of R. Now if x has coordinates (r, θ), we define $\phi_R(x)$ to be the point with coordinates $(r\phi_R(r), \theta)$. Then $\phi_R = identity$ on $B(R)$ and $\phi_R(M) \subset B(2R)$. It remains to estimate the energy density of ϕ_R. Direct computation shows that

$$\frac{\partial}{\partial r} g_{\alpha\beta} = 2h_{\alpha\beta} ,$$

where $(h_{\alpha\beta})$ is the second fundamental form of $\partial B(r)$ with respect to $\dfrac{\partial}{\partial r}$. Since N is simply connected and has nonpositive curvature, therefore $\partial B(r)$ is strictly convex and $(h_{\alpha\beta})$ is positive definite. Hence for all $\xi \in \mathbf{R}^n$ and for all $r \leqq r'$,

$$g_{\alpha\beta}(r, \theta)\xi^\alpha \xi^\beta \leqq g_{\alpha\beta}(r', \theta)\xi^\alpha \xi^\beta , \tag{4.4}$$

It is clear that $e(\phi_R) = 1$ on $B(R)$. For x with $r = r(x) \geqq R$, let us choose θ such that $g_{\alpha\beta}(r, \theta) = \delta_{\alpha\beta}$. Then the energy density of ϕ_R is

$$\begin{aligned} e(\phi_R)(x) &= \left(\frac{\partial}{\partial r} (r\tilde{\phi}(r)) \right)^2 + g^{\alpha\beta}(r, \theta) g_{\alpha\beta}(\phi_R(r), \theta) \\ &\leqq (1 + C_2)^2 + \sum_\alpha g_{\alpha\alpha}(\phi_R(r), \theta) \\ &\leqq (1 + C_2)^2 + \sum_\alpha g_{\alpha\alpha}(r, \theta) \\ &\leqq (1 + C_2)^2 + (n - 1) , \end{aligned}$$

where we have used (4.4) and the fact that $\phi_R(r) \leqq r$ for all r. This completes the proof of the lemma.

Theorem 4.1 *Let M and N be as in Theorem* 4.1. *For any $h \in C^1(M, N)$ with bounded energy density, there exists a unique solution u of* (0.1) *on $M \times (0, \infty)$ such that $u(x, 0) = h(x)$ and $\sup_{M \times [0, T]} e(u) < \infty$ for all $T > 0$.*

Proof. Let us first assume that N is simply connected. Let y_0 be a fixed point in N and let $R_\nu \to \infty$. Denote $\Omega_\nu = B_{y_0}(R_\nu)$. By Lemma 4.2, there exists smooth maps g_ν from N to itself such that $g_\nu|_{\Omega_\nu} = identity$, $g_\nu(N)$ is compact, and $\sup_{\nu, N} e(g_\nu) < \infty$. Consider $h_\nu = g_\nu \circ h \in C^1(M, N)$. Then there exists a constant C_3, independent of ν such that

$$\sup_M e(h_\nu) \leqq C_3 \tag{4.5}$$

for all ν. Since $h_\nu(M)$ is bounded, by Theorem 4.1 there exists a solution u_ν of (0.1) on $M \times (0, \infty)$. From the proof of Theorem 4.1 and (4.5), it is easy to see that for all $T > 0$ and for all ν, there exists a constant C_4 not depending on ν such that

$$\sup_{M \times [0, T]} e(u_\nu) \leqq C_4 . \tag{4.6}$$

The proof of the main lemma in [H] and the fact that N has nonpositive curvature implies that

$$\left(\Delta - \frac{\partial}{\partial t}\right)|(u_v)_t|^2 \geqq 0 .$$

Let $p \in M$ be a point. As in [Lo-T], there exists a constant C_5 depending only on m such that for t, $R > 0$,

$$\int_0^t \int_{B_p(R)} |(u^v)_t|^2(y, \tau)\, dy\, d\tau \leqq E_v(0, 2R) + \frac{C_5}{R^2}\int_0^t E_v(\tau, 2R)\, d\tau , \tag{4.7}$$

where $E_v(t, R) = \int_{B_p(R)} e(u_v)(y, t)\, dy$.

On the other hand because $|(u^v)_t|^2$ is a nonnegative subsolution for the heat equation, the mean-value inequality Theorem 1.1 gives

$$|(u^v)_t|^2(p, T) \leqq C(K, 1, R, T) \int_{B_p(R)\times[0, T]} |(u^v)_t|^2(y, s)\, dy\, ds$$

for $0 < T \leqq \frac{1}{2}R^2$, where

$$C(K, 1, R, T) = C_6 \frac{V(K, 2R)}{V_p(R)}(R\sqrt{K}\coth(R\sqrt{K}) + 1)$$
$$\times \exp(C_7 KT)(R^{-m-2} + T^{-(m/2)-1})$$

for some constants C_6 and C_7 depending only on m. Hence by setting $R = 2\sqrt{T}$, this implies

$$|(u^v)_t|^2(p, T) \leqq C_8 T^{-(m/2)-1} \frac{V(K, 4\sqrt{T})}{V_p(K, 2\sqrt{T})} \int_{B_p(2\sqrt{T})\times[0,T]} |u_t|^2(y, s)\, dy\, ds ,$$

where C_8 is a constant depending only on m and the upper bound of T. Together with (4.6), we have

$$|(u^v)_t|^2(p, T) \leqq C_8 T^{-(m/2)-1} \frac{V(K, 4\sqrt{T})}{V_p(K, 2\sqrt{T})}\left(E_v(0, 4\sqrt{T}) + \frac{C_5}{2T}\int_0^T E_v(\tau, 4\sqrt{T})\, d\tau\right).$$

Therefore, given a bounded domain Ω in M and $T > 0$, there exists a constant depending only on m, T, Ω and $\sup_{M\times[0, T]} e(u_v)$, such that for $0 < t \leqq T$ and $x \in \Omega$, we obtain the estimates

$$|(u_v)_t|(x, t) \leqq C_9 t^{-1/2} ,$$

and

$$\operatorname{dist}(u_v(x, t), h_v(x)) \leqq 2C_9 t^{1/2} . \tag{4.8}$$

By (4.6) and (4.7) and $g_v = \textit{identity}$ on Ω_v, we conclude that for any $T > 0$ and $R > 0$ there exists a bounded set $N(T, R) \subset N$ such that

$$u_v(B_p(R)\times[0, T]) \subset N(T, R) \tag{4.9}$$

for sufficiently large v.

From (4.6) and (4.9), it is easy to deduce that if we pass to a subsequence, u_ν will converge to a map u which is a solution of (0.1) on $M \times (0, \infty)$ and by (4.6) and (4.8) we can see that $u(x, 0) = h(x)$. The convergence is uniform on $B_p(R) \times [0, T]$ for all $0 < T < \frac{1}{4}R^2$ together with their first and second derivatives in the space variable and their first derivative in the time variable. Hence u is a solution of (0.1) which satisfies the condition of the theorem. Uniqueness of u follows from Theorem 2.3 in [Lo-T].

In general, let $\bar{M}$ and $\bar{N}$ be the universal coverings of M and N and let π_M and π_N be the projections from $\bar{M}$ to M and from $\bar{N}$ to N respectively. Then one can find h^* from $\bar{M}$ to $\bar{N}$ such that $h \circ \phi_M = \phi_N \circ h^*$ and that h^* is invariant under deck transformations of $\bar{M}$. Note that h^* has bounded energy density. Therefore by the previous argument, there is a solution u^* of the parabolic equation for harmonic maps from $\bar{M} \times (0, \infty)$ to $\bar{N}$ with initial data h^*. Define $u(x, t) = \pi_N(u^*(\bar{x}, t))$, where $\pi_M(\bar{x}) = x$. In order to show that $u(x, t)$ is well-defined and is a solution of (0.1) on $M \times (0, \infty)$ with initial data h, it is sufficient to show that $u^*(\bar{x}, t) = u^*(\sigma(\bar{x}), t)$ for all deck transformations σ on $\bar{M}$. However, this follows from Theorem 2.3 in [Lo-T] and the fact that h^* is invariant under deck transformations.

Theorem 4.3 *Let M^m be a complete noncompact Riemannian manifold with nonnegative Ricci curvature, and let N^n be a complete Riemannian manifold with nonpositive sectional curvature. Let $h \in C^1(M, N)$ with bounded energy density. Then (0.1) has a unique solution u on $M \times (0, \infty)$ with initial condition $u(\cdot, 0) = h$, which satisfies $\sup_{M \times [0, T]} e(u) < \infty$ for all $T > 0$. Furthermore, if there exists a point $p \in M$ and a sequence $t_\nu \to \infty$ such that $u(p, t_\nu)$ converges in N (for example when N is compact), then by passing to a subsequence, $u(\cdot, t_\nu)$ converges uniformly on compact subsets together with their first and second derivatives to a harmonic map u_∞.*

Proof. The existence of u is guaranteed by Theorem 4.3. Since the Ricci curvature of M is nonnegative and the sectional curvature of N is nonpositive, Bochner's formula in [E-S] implies that $e(u)$ is a subsolution of the heat equation for functions on M. Theorem 3.1 asserts that $e(u)$ is bounded on $M \times [0, T]$ for all $T < \infty$. Hence by the maximum principle of [K-L], we have

$$\sup_{M \times [0, \infty)} e(u) \leqq \sup_M e(h) < \infty \,. \tag{4.10}$$

Suppose there exist $p \in M$ and $t_\nu \to \infty$ such that $u(p, t_\nu)$ converges. By (4.7), (4.10), the fact that M has nonnegative Ricci curvature, and the volume comparison theorem, we have

$$\lim_{\substack{R \to \infty \\ t \to \infty}} \frac{1}{tV(B)} \int_0^t d\tau \int_{B_p(R)} |u_t|^2(y, \tau)\, d\tau = 0 \,, \tag{4.11}$$

where $V(R)$ is the volume of $B_p(R)$. On the other hand, as in the proof of the main lemma in [H], since N has nonpositive sectional curvature, we have

$$\left(\Delta - \frac{\partial}{\partial t} \right) |u_t|^2 \geqq 0 \,.$$

Hence by Theorem 1.1 and (4.11), a subsequence of $|u_t|^2(\cdot, t_\nu)$ converges to 0, uniformly on compact sets of M.

The Hölder estimates for the gradients, the Schauder estimates for the parabolic equations (see [L-S-U]), (4.10), the fact that u satisfies (0.1), and that $u(p, t_\nu)$ converges imply that the sequence of maps $v_\nu(x, t) = u(x, t_\nu + t)$ converges together with their first and second derivatives in the space variable and the first derivative of the time variable uniformly on $B_p(R) \times [0, 1]$ for all $R > 0$. Since each v_ν satisfies (0.1), and for all x, $|u_t(x, t_\nu)| \to 0$ as $\nu \to \infty$, the theorem follows.

5 Harmonic maps: $\lambda(M) > 0$ and $\kappa = 0$

In this section we would like to discuss the question of existence of harmonic maps from a manifold M whose spectrum has a positive lower bound. We also assume that the sectional curvature of N is nonpositive. In particular we will study the boundary valued problem at infinity for harmonic maps between two hyperbolic spaces.

Lemma 5.1 *Let M be a complete noncompact manifold with Ricci curvature bounded from below by $-K \leqq 0$ and $\inf_{x \in M} V_x(1) = b > 0$. Assume that the lower bound of the spectrum $\lambda(M)$ of M is positive. Let N be a complete manifold with sectional curvature bounded above by $\kappa \geqq 0$. Let u be a solution of the parabolic equation for harmonic map from $M \times [0, T)$ to N where $1 < T \leqq \infty$, and suppose $|u_t|$ is a bounded function on $M \times [0, T)$. If $\kappa > 0$, we also assume that $\sup_{M \times [0, T)} e(u) = \beta < \infty$. For any $p > 1$, there exists a constant C_1 depending only on K, λ, p, and b such that for all $1 < t < T$ and for all $x \in M$,*

$$|u_t|^2(x, t) \leqq C_1 \exp\left(-\frac{4\lambda(q-1)}{q^2} t + 2\kappa\beta t\right)\left(\int_M |u_t|^{2p}(y, 0)\, dy\right)^{1/p}, \tag{5.1}$$

where q is such that $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. We may assume that the L^p norm of $|u_t|^2(\cdot, 0)$ is finite, otherwise the result is obviously true. As before, by [H] the function $f(x, t) = \exp(-2\kappa\beta t)|u_t|^2(x, t)$ is a subsolution of the heat equation for functions on M. By the maximum principle in [K-L], we have

$$f(x, t) \leqq \int_M H(x, y, t) f(y, 0)\, dy \tag{5.2}$$

for all $x \in M$ and for all $0 < t < T$.

Let us first estimate the L^q norm of the heat kernel H on M with $q > 1$. For any smooth compactly supported function ϕ, for any $0 < \varepsilon < 1$, and for a real number δ, we have

$$\begin{aligned} \frac{\partial}{\partial t} \int_M \phi^2(y) H^q(x, y, t)\, dy &= q \int_M \phi^2(y) H^{q-1}(x, y, t) \frac{\partial H}{\partial t}(x, y, t)\, dy \\ &= q \int_M \phi^2(y) H^{q-1}(x, y, t) \Delta H(x, y, t)\, dy \end{aligned}$$

$$= -q(q-1)\int_M \phi^2(y)H^{q-2}(x,y,t)|\nabla H(x,y,t)|^2\,dy$$

$$-2q\int_M \phi(y)H^{q-1}(x,y,t)\langle \nabla\phi(y), \nabla H(x,y,t)\rangle\,dy$$

$$= -\frac{4(q-1)}{q}\int_M \phi^2(y)|\nabla H^{q/2}(x,y,t)|^2\,dy$$

$$-4\int_M \phi(y)H^{q/2}(x,y,t)\langle \nabla\phi(y), \nabla H^{q/2}(x,y,t)\rangle\,dy$$

$$= -\frac{4(q-1)}{q}\Bigg(\int_M \phi^2(y)|\nabla H^{q/2}(x,y,t)|^2\,dy$$

$$+\left(\frac{q}{q-1}-2\delta\right)\int_M \phi(y)H^{q/2}(x,y,t)$$

$$\langle \nabla\phi(y), \nabla H^{q/2}(x,y,t)\rangle\,dy$$

$$+2\delta\int_M \phi(y)H^{q/2}(x,y,t)\langle \nabla\phi(y), \nabla H^{q/2}(x,y,t)\rangle\,dy\Bigg).$$

Using the inequality

$$\left|2\delta\int_M \phi(y)H^{q/2}(x,y,t)\langle\nabla\phi(y), \nabla H^{q/2}(x,y,t)\rangle\,dy\right|$$

$$\leqq \varepsilon\int_M \phi^2(y)|\nabla H^{q/2}(x,y,t)|^2\,dy + \frac{\delta^2}{\varepsilon}\int_M |\nabla\phi(y)|^2H^q(x,y,t)\,dy,$$

we conclude that

$$\frac{\partial}{\partial t}\int_M \phi^2(y)H^q(x,y,t)\,dy \leqq -\frac{4(q-1)}{q}\Bigg((1-\varepsilon)\int_M \phi^2(y)|\nabla H^{q/2}(x,y,t)|^2\,dy$$

$$+\left(\frac{q}{q-1}-2\delta\right)\int_M \phi(y)H^{q/2}(x,y,t)\langle\nabla\phi(y), \nabla H^{q/2}(x,y,t)\rangle\,dy$$

$$-\frac{\delta^2}{\varepsilon}\int_M |\nabla\phi(y)|^2H^q(x,y,t)\,dy\Bigg).$$

Let us now choose δ such that $2(1-\varepsilon) = \frac{q}{q-1} - 2\delta$, the identity

$$\int_M \phi^2(y)|\nabla H^{q/2}(x,y,t)|^2\,dy + 2\int_M \phi(y)H^{q/2}(x,y,t)\langle\nabla\phi(y), \nabla H^{q/2}(x,y,t)\rangle\,dy$$

$$= \int_M |\nabla(\phi(y)H^{q/2}(x,y,t))|^2\,dy - \int_M |\nabla\phi(y)|^2H^q(x,y,t)\,dy$$

and the fact that the lower bound of the spectrum λ is positive imply that

$$\begin{aligned}\frac{\partial}{\partial t}\int_M \phi^2(y)H^q(x,y,t)\,dy &\leqq -\frac{4(q-1)}{q}\Bigg((1-\varepsilon)\int_M |\nabla(\phi(y)H^{q/2}(x,y,t))|^2\,dy\\ &\quad -\left(1-\varepsilon-\frac{\delta^2}{\varepsilon}\right)\int_M |\nabla\phi(y)|^2 H^q(x,y,t)\,dy\Bigg)\\ &\leqq -\frac{4(q-1)}{q}\Bigg(\lambda(1-\varepsilon)\int_M \phi^2(y)H^q(x,y,t)\,dy\\ &\quad +\left(1-\varepsilon-\frac{\delta^2}{\varepsilon}\right)\int_M |\nabla\phi(y)|^2 H^q(x,y,t)\,dy\Bigg).\end{aligned}$$

However, by the estimate of $H(x,y,t)$ in [L-Y], since M has Ricci curvature bounded from below by $-K$ and $\inf_{x\in M} V_x(1) = b > 0$, $H(x,y,t)$ is uniformly bounded from above as a function of y. Together with $\int_M H(x,y,t)\,dy = 1$, this implies that the integral $\int_M H^q(x,y,t)\,dy$ must be finite for $q > 1$. Hence if we choose the cut-off function to be identically 1 on $B_x(R)$, identically 0 on $M\backslash B_x(2R)$, and satisfies $|\nabla\phi|^2 \leqq \frac{3}{R^2}$, the above inequality will approach

$$\frac{\partial}{\partial t}\int_M H^q(x,y,t)\,dy \leqq \lambda(1-\varepsilon)\frac{4(q-1)}{q}\int_M H^q(x,y,t)\,dy$$

as $R\to\infty$. Taking $\varepsilon\to 0$, this yields the estimate

$$\frac{\partial}{\partial t}\int_M H^q(x,y,t)\,dy \leqq -\frac{4\lambda(q-1)}{q}\int_M H^q(x,y,t)\,dy\,.$$

Integrating from 1 to t, and using the estimate of the heat kernel in [L-Y], we have

$$\begin{aligned}\int_M H^q(x,y,t)\,dy &\leqq \exp\left(-\frac{4\lambda(q-1)}{q}(t-1)\right)\int_M H^q(x,y,1)\,dy\\ &\leqq \exp\left(-\frac{4\lambda(q-1)}{q}(t-1)\right)\left(\sup_{y\in M} H^{q-1}(x,y,1)\right)\int_M H(x,y,1)\,dy\\ &\leqq C_2\exp\left(-\frac{4\lambda(q-1)}{q}(t-1)\right),\end{aligned}$$

where C_2 is a constant depending only on K, q, and b.

By (5.1), (5.2) and the Hölder inequality we have for $1 < t < T$, and for $x\in M$,

$$\begin{aligned}f(x,t) &\leqq \left(\int_M H^q(x,y,t)\,dy\right)^{1/q}\left(\int_M f^p(y,0)\,dy\right)^{1/p}\\ &\leqq C_3\exp\left(-\frac{4\lambda(q-1)}{q^2}t\right)\left(\int_M f^p(y,0)\,dy\right)^{1/p},\end{aligned}$$

where $q > 1$ is such that $\frac{1}{p} + \frac{1}{q} = 1$, and C_3 is a constant depending only on K, λ, p, and b. The result now follows from the definition of f.

Recall that the tension field $\tau(u)$ of a map u from M to N is a vector field along u. In local coordinates of M and N, using previous notations,

$$\tau(u)^\alpha = \Delta u^\alpha + \Gamma^\alpha_{\beta\gamma} \frac{\partial u^\beta}{\partial x^i} \frac{\partial u^\gamma}{\partial x^j} g^{ij} .$$

Theorem 5.2 *Let M^m be a complete manifold with Ricci curvature bounded below by $-K \leqq 0$, and $\inf_{x \in M} V_x(1) = b > 0$. Assume that the lower bound of the spectrum $\lambda(M)$ of M is positive. Let N^n be a complete manifold with nonpositive curvature. Let u be a solution of the parabolic equation for harmonic maps from $M \times [0, \infty)$ to N^n. Suppose that the square of the norm of the tension field of the initial data of u is bounded and in $L^p(M)$ for some $p > 1$ and suppose that the energy density of u is also bounded on $M \times [0, T]$ for all $T < \infty$. Then $u(\cdot, t)$ converges to a harmonic map u_∞ as $t \to \infty$ with the following properties: (i) $\sup_M e(u) < \infty$; (ii) $\sup_{x \in M} \operatorname{dist}(u_\infty(x), u(x, 0)) < \infty$; and (iii) if in addition that the norm of the tension field of the initial data tends to 0 as $x \to \infty$ then $\lim_{x \to \infty} \operatorname{dist}(u_\infty(x), u(x, 0)) = 0$.*

Proof. Since the square of the norm of the tension field of the initial data is bounded and in $L^p(M)$ for some $p > 1$, the equation implies that $|u_t|^2(\cdot, 0)$ is bounded and in $L^p(M)$. By Theorem 1.2, the fact that $|u_t|^2$ is a subsolution of the heat equation, the assumption on the volume of geodesic balls of radius 1, that $e(u)$ is bounded on $M \times [0, T]$ for finite T, and (4.8), we conclude that $|u_t|^2$ is bounded on $M \times [0, T]$ for all finite T. By Lemma 5.1 implies that $|u_t|^2(x, t)$ decays exponentially in t. Therefore as $t \to \infty$, $u(x, t)$ converges uniformly on M to a map u_∞. In order to prove that u_∞ is a harmonic map and satisfies (i), it suffices to prove that there exists a constant $C_4 > 0$ such that

$$e(u)(x, t) \leqq C_4$$

for all x and for all $t \geqq 1$. As in [Lo-T], if ϕ is a smooth function with compact support in M then we have

$$\frac{d}{dt} \int_M \phi^2 e(u) \leqq 4 \left(\int_M |\nabla \phi|^2 |u_t|^2 \right)^{1/2} \left(\int_M \phi^2 e(u) \right)^{1/2} .$$

Using the fact that

$$|u_t|^2(x, t) \leqq C_5 \exp(-at) \tag{5.3}$$

for some constants $C_5 > 0$, $a > 0$ and for all x and t, we have

$$\frac{d}{dt} \int_M \phi^2 e(u) \leqq 4\sqrt{C_5} \exp\left(-\frac{1}{2} at \right) \left(\int_M |\nabla \phi|^2 \right)^{1/2} \left(\int_M \phi^2 e(u) \right)^{1/2} .$$

From this it is easy to see that

$$\left(\int_M \phi^2 e(\cdot, t) \right)^{1/2} \leqq \frac{4\sqrt{C_5}}{a} \left(\int_M |\nabla \phi|^2 \right)^{1/2} + \left(\int_M \phi^2 e(\cdot, 0) \right)^{1/2} .$$

Using a suitable ϕ we see that $\int_{B_x(1)} e(u)(\cdot, t)$ is uniformly bounded by a constant independent of x and t. By the mean-value inequality, we conclude that $e(u)(x, t)$ is uniformly bounded by a constant C_4 independent of x and t for $t \geqq 1$.

It remains to prove that u_∞ satisfies (ii) and (iii). By (5.3) it is easy to see that

$$\sup_{x \in M} \operatorname{dist}(u_\infty(x), (x, 0)) < \infty .$$

Suppose the norm of the tension field of the initial data tends to 0 as $x \to \infty$, then

$$\lim_{x \to \infty} |u_t(x, 0)| = 0 .$$

Since the L^p norm of $|u_t(\cdot, t)|^2$ is less than or equal to the L^p norm of $|u_t(\cdot, 0)|^2$, hence for $\infty > T > 0$ the L^p norm of $|u_t|^2$ on $M \times [0, T]$ is finite. Since $|u_t|^2$ is a subsolution of the heat equation, therefore by Theorem 1.2 we conclude that

$$\lim_{x \to \infty} \sup_{t \in [0, T]} |u_t(x, t)| = 0 .$$

Now

$$d_N(u_\infty(x), u(x, 0)) \leqq \int_0^\infty |u_t(x, t)| = \int_0^T |u_t| + \int_T^\infty |u_t(x, t)| .$$

Given $\varepsilon > 0$, there exists $T > 0$, such that

$$\int_T^\infty |u_t(x, t)| \leqq \varepsilon .$$

For a fixed T, there exists $R > 0$ such that if $d_M(x) \geqq R$, then

$$\int_0^T |u_t(x, t) \leqq \varepsilon ,$$

where $d_M(x)$ is the distance from x to some fixed point. Hence if $d_M(x) \geqq R$, then

$$d_N(u_\infty(x), u(x, 0)) \leqq 2\varepsilon .$$

This complete the proof of the theorem.

As an application of the above theorem, let us consider the boundary value problem at infinity from $M = \mathbf{H}^m$ to $N = \mathbf{H}^n$, the hyperbolic spaces of dimension m and n with constant curvature -1. Let us first introduce some notations. Both M and N can be taken to be the unit balls in $\mathbf{R}^m$ and $\mathbf{R}^n$ with the Poincaré metric. The ideal boundaries of M and N are identified with $\mathbf{S}^{m-1}$ and $\mathbf{S}^{n-1}$, respectively.

Lemma 5.3 *Let f be a C^3 map from $\mathbf{S}^{m-1}$ into $\mathbf{S}^{n-1}$ so that the energy density of f with respect to the standard metrics on $\mathbf{S}^{m-1}$ and $\mathbf{S}^{n-1}$ is nowhere zero. Then f can be extended to a map from M to N so that the energy density is bounded, and the square of the norm of the tension field tends to 0 as $x \to \infty$ and is in $L^k(M)$ for all $k > \dfrac{m-1}{2}$.*

Proof. On M, the metric is

$$ds_M^2 = 4(1 - \rho^2)^{-2} g_{ij} dx^i dx^j ,$$

where $g_{ij} = \delta_{ij}$ and $\rho^2 = \sum_{i=1}^{m} (x^i)^2 < 1$. On N, we denote the metric in terms of polar coordinates by

$$ds_N^2 = 4(1 - r^2)^{-2}(dr^2 + r^2 h_{pq} d\theta^p d\theta^q),$$

where $r^2 = \sum_{\alpha=1}^{n} (u^\alpha)^2 < 1$. The indices p and q ranges from 2 to n, and $h_{pq} d\theta^p \theta^q$ is the standard metric on $\mathbf{S}^{n-1}$. The Christoffel symbols on N are given by

$$\Gamma_{11}^1 = \frac{2r}{1 - r^2},$$

$$\Gamma_{1q}^1 = 0,$$

$$\Gamma_{11}^s = 0$$

$$\Gamma_{pq}^1 = -\frac{r(1 + r^2) h_{pq}}{1 - r^2},$$

$$\Gamma_{1q}^s = \frac{(1 + r^2)\delta_{sq}}{r(1 - r^2)},$$

and

$$\Gamma_{pq}^s = \tilde{\Gamma}_{pq}^s$$

for $2 \leqq s, p, q \leqq n$. The terms $\tilde{\Gamma}_{pq}^s$ are the Christoffel symbols on $\mathbf{S}^{n-1}$ with respect to the standard metric.

For a map u which is given by $r = r(x^1, \ldots, x^m)$, and $\theta^p = \theta^p(x^1, \ldots, x^m)$, let us compute the component of the tension field $\tau(u)$. Direct computation shows that

$$\begin{aligned}\tau(u)^1 &= \Delta r + \frac{2r|\nabla r|^2}{1 - r^2} - \frac{r(1 + r^2)}{1 - r^2} \sum_{p,q=2}^{n} h_{pq} \langle \nabla\theta^p, \nabla\theta^q \rangle \\ &= \frac{(1 - \rho^2)^2}{4} \Delta_0 r + \frac{1}{4}\Bigg(2(m - 2)(1 - \rho^2)\rho \frac{\partial r}{\partial \rho} \\ &\quad + \frac{r(1 - \rho^2)^2 (2|\nabla_0 r|^2 - (1 + r^2) \sum_{p,q=2}^{n} h_{pq} \langle \nabla_0 \theta^p, \nabla_0 \theta^q \rangle)}{1 - r^2} \Bigg). \end{aligned} \tag{5.4}$$

For $s \geqq 2$, we have

$$\begin{aligned}\tau(u)^s &= \Delta\theta^s + \frac{2(1 + r^2)}{r(1 - r^2)} \langle \nabla r, \nabla\theta^s \rangle + \sum_{p,q=2}^{n} \tilde{\Gamma}_{pq}^s \langle \nabla\theta^p, \nabla\theta^q \rangle \\ &= \frac{(1 - \rho^2)^2}{4}\Bigg(\Delta_0 \theta^s + \sum_{p,q=2}^{n} \tilde{\Gamma}_{pq}^s \langle \nabla_0 \theta^p, \nabla_0 \theta^q \rangle \Bigg) \\ &\quad + \frac{1 - \rho^2}{2}\Bigg((m - 2)\rho \frac{\partial \theta^s}{\partial \rho} + \frac{(1 + r^2)(1 - \rho^2) \langle \nabla_0 r, \nabla_0 \theta^s \rangle}{r(1 - r^2)} \Bigg). \end{aligned} \tag{5.5}$$

Here, ∇ denotes the gradient on M, ∇_0 denotes the Euclidean gradient, and Δ_0 denotes the Euclidean Laplacian operator. We have used the fact that the hyperbolic Laplacian is given by

$$\Delta = \frac{1}{4}\Bigg((1 - \rho^2)^2 \Delta_0 + 2(m - 2)(1 - \rho^2)\rho \frac{\partial}{\partial \rho} \Bigg).$$

Let us extend f near infinity in the following way. Denote

$$f(\eta^2, \ldots, \eta^m) = (f^2(\eta^2, \ldots, \eta^m), \ldots, f^n(\eta^2, \ldots, \eta^m)),$$

where $(\eta^2, \ldots, \eta^m)$ and $(\theta^2, \ldots, \theta^n)$ are coordinates on $\mathbf{S}^{m-1}$ of the domain manifold and $\mathbf{S}^{n-1}$ on the target manifold, respectively. Let $e(f)$ be the energy density of f with respect to the standard metrics on $\mathbf{S}^{m-1}$ and $\mathbf{S}^{n-1}$ which is nowhere zero. Define an extension u of f to M by

$$\theta^p(\rho, \eta^2, \ldots, \eta^m) = f^p(\eta^2, \ldots, \eta^m)$$

and

$$r(\rho, \eta^2, \ldots, \eta^m) = 1 - (1 - \rho)\phi(\eta^2, \ldots, \eta^m),$$

where

$$\phi = \sqrt{\frac{e(f)}{m-1}}.$$

This extension is well defined providing ρ is close to 1.

It is easy to see that $\Delta_0 r$ and $\sum_{p,q=2}^{n} \tilde{\Gamma}^s_{pq} \langle \nabla_0 \theta^p, \nabla_0 \theta^q \rangle$ are bounded if ρ is close to 1. We also have,

$$\frac{1-\rho}{1-r} = \frac{1}{\phi}, \tag{5.6}$$

$$\frac{\partial r}{\partial \rho} = \phi, \tag{5.7}$$

$$|\nabla_0 r|^2 = \phi^2 + \frac{(1-\rho)^2}{\rho^2} |\tilde{\nabla}\phi|^2, \tag{5.8}$$

$$\langle \nabla_0 r, \nabla_0 \theta^s \rangle = \frac{1-\rho}{\rho^2} \langle \tilde{\nabla}\phi, \tilde{\nabla} f^s \rangle, \tag{5.9}$$

and

$$\sum_{p,q=2}^{n} h_{pq} \langle \nabla_0 \theta^p, \nabla_0 \theta^q \rangle = \frac{e(f)}{\rho^2}, \tag{5.10}$$

where $\tilde{\nabla}$ denotes the gradient on $\mathbf{S}^{m-1}$.

By (5.6) and the fact that ϕ is nowhere zero, there exists a constant $C_7 > 0$ such that

$$C_7^{-1}(1-r) \leqq (1-\rho) \leqq C_7(1-r). \tag{5.11}$$

Hence

$$\begin{aligned}\tilde{\tau}(u)^1 &= 2(m-2)\frac{\partial r}{\partial \rho} + \frac{1-\rho^2}{1-r^2}\left(2|\nabla_0 r|^2 - (1+r^2)\sum_{p,q=2}^{n} h_{pq}\langle \nabla_0\theta^p, \nabla_0\theta^q\rangle\right) \\ &= 2(m-2)\phi + \frac{1-\rho^2}{1-r^2}\left(2\phi^2 + \frac{2(1-\rho)^2}{\rho^2}|\tilde{\nabla}\phi|^2 - \frac{(1+r^2)e(f)}{\rho^2}\right)\end{aligned}$$

$$= 2(m-2)\phi + \frac{2(1+\rho)}{1+r}\phi - \frac{(n-1)(1+\rho)(1+r^2)}{\rho^2(1+r)}\phi$$

$$+ \frac{2(1-\rho^2)(1-\rho)^2}{\rho^2(1-r^2)}|\tilde{\nabla}\phi|^2$$

$$= 2\left(\frac{1+\rho}{1+r} - 1\right)\phi + \left(2 - \frac{(1+\rho)(1+r^2)}{\rho^2(1+r)}\right)(n-1)\phi$$

$$+ \frac{2(1-\rho^2)(1-\rho)^2}{\rho^2(1-r^2)}|\tilde{\nabla}\phi|^2$$

$$= 2\frac{\rho - r}{1+r}\phi + \frac{(2\rho+1)(\rho-1) + (\rho-r)(\rho+r+\rho r) - \rho^2(1-r)}{\rho^2(1+r)}(n-1)\phi$$

$$+ \frac{2(1-\rho^2)(1-\rho)^2}{\rho^2(1-r^2)}|\tilde{\nabla}\phi|^2 \tag{5.12}$$

$$= O(1-\rho)$$

as $\rho \to 1$, where we have used the definition of r, (5.11), and the fact that ϕ is nowhere zero.

Similarly, for $n \geqq s \geqq 2$,

$$\tilde{\tau}(u)^s = (m-2)\frac{\partial \theta^s}{\partial \rho} + \frac{(1+r^2)(1-\rho^2)}{1-r^2}\langle \nabla_0 r, \nabla_0 \theta^s \rangle$$

$$= -\frac{(1+r^2)(1-\rho^2)(1-\rho)}{\rho^2(1-r^2)}\langle \tilde{\nabla}\phi, \tilde{\nabla} f^s \rangle$$

$$= O(1-\rho) \tag{5.13}$$

as $\rho \to 1$, where we have used (5.11) and the fact that f is in C^3.

It is easy to see that there is a constant C_8, such that

$$|\tau(u)^\alpha - \frac{1-\rho^2}{4}\tilde{\tau}(u)^\alpha| \leqq C_8(1-\rho)^2,$$

for $1 \leqq \alpha \leqq n$. Combining this with (5.12) and (5.13), we conclude that

$$|\tau(u)^\alpha| \leqq C_9(1-\rho)^2$$

for some constant C_9 and for ρ close to 1. Hence by (5.11), the norm of the tension field

$$\|\tau(u)\| = 2(1-r^2)^{-1}|\tau(u)|$$

of the extension is in $L^k \cap L^\infty$ for $k > m-1$.

The energy density is given by

$$e(u) = \left(\frac{1-\rho^2}{1-r^2}\right)^2\left(\phi^2 + \frac{(1-\rho)^2}{\rho^2}|\tilde{\nabla}\phi|^2 + \frac{r^2}{\rho^2}e(f)\right).$$

which is bounded because of (5.11). Therefore we have extended f near infinity, and it is easy to see that one can extend f further to a map defined on M which satisfies the required conditions.

Theorem 5.4 *With the same notations as in Lemma* 5.3, *let f be a C^3 map from $\mathbf{S}^{m-1}$ to $\mathbf{S}^{n-1}$ with nowhere vanishing energy density. Then there exists a harmonic map from $\mathbf{H}^m$ into $\mathbf{H}^n$ which has the boundary map given by f at infinity.*

Proof. By Lemma 5.3, one can extend f to be a map from M to N which is also denoted by f, such that it has bounded energy density and the square of the norm of its tension field is bounded and in L^k for $k > \dfrac{m-1}{2}$. Let us first assume that f is C^∞. By Theorem 4.1, the parabolic equation for harmonic map from $M \times [0, \infty)$ to N with initial data f has a solution u. Since hyperbolic space satisfies the conditions of Theorem 5.2, by the properties of f, $u(\cdot, t)$ converges uniformly to a harmonic map u_∞. The fact that

$$\sup_{x \in M} \operatorname{dist}(u_\infty(x), f(x)) < \infty$$

implies that u_∞ and f have the same boundary value. Since the computations in Lemma 5.3 involve at most the third derivative of f, hence we may relax the smoothness of f to C^3 by an approximation argument.

Theorem 5.5 *With the same assumptions and notations as in Theorem* 5.4, *if $m = n = 2$ and if the boundary map is a diffeomorphism, then the harmonic map obtained in the theorem is a diffeomorphism.*

Proof. Using the same notations as in Lemma 5.3 with $m = n = 2$. We identify M and N with the unit disk D in $\mathbf{R}^2$ with the Poincaré metric. Let f be a C^3 map from S^1 to S^1 so that $e(f) > 0$. Note that if we extend f as a map from D to D in two different ways to get h and h^*, and if both h and h^* satisfy the conditions in Lemma 5.3, then h and h^* can be deformed to harmonic maps u and u^* through the parabolic equation. Suppose

$$\lim_{x \to \infty} \operatorname{dist}(h(x), h^*(x)) = 0 ,$$

then the two harmonic maps will be the same by property (i) in Theorem 5.2 and the fact that $\operatorname{dist}^2(u(x), u^*(x))$ is subharmonic. Therefore let us be more precise on the extension h of f. Let ψ_1 be a smooth function from $[0, 1]$ to itself, so that $\psi_1 = 0$ on $\left[0, \dfrac{b}{2}\right]$, $\psi_1(\rho) \leqq \rho$ and $\psi_1(\rho) = \rho$ on $[b, 1]$, where $0 < b < 1$ is chosen to satisfy $\left(1 - \dfrac{b}{2}\right) e(f) < 1$. Let ψ_2 be another smooth function from $[0, 1]$ to $[0, 1]$, such that $\psi_2 = 0$ on $\left[0, \dfrac{b}{2}\right]$ and $\psi_2 = 1$ on $[b, 1]$. Let $\frac{1}{2} < a_j < 1$, and $a_j \uparrow 1$. For each j, we extend f to a map h_j from D to D by setting

$$r = r_j(\rho, \eta) = a_j \psi_2(\rho)(1 - (1 - \psi_1(\rho))\phi_j(\eta))$$

and

$$\theta = f(\eta) ,$$

where

$$\phi_j = \sqrt{\frac{1 + a_j^2}{2a_j^2} e(f)} .$$

We take h to be the map defined as above by replacing a_j by 1. It is easy to see that h_j are well-defined smooth maps. The image of $\rho \leqq \frac{b}{2}$ is the origin, the image of h_j is in the set given by $r \leqq a_j$. The boundary of D is mapped bijectively onto $r = a_j$. Moreover, for $\rho \geqq b$,

$$r_j = a_j(1 - (1 - \rho)\phi_j) .$$

In what follows, the constants C_i are independent of j. Consider $b \leqq \rho < 1$, using the same notation as in the proof of Lemma 5.3,

$$\begin{aligned}
\tilde{\tau}(h_j)^1 &= \frac{1 - \rho^2}{1 - r_j^2}(2|\nabla_0 r_j|^2 - (1 + r_j^2)e(f)^2) \\
&= \frac{1 - \rho^2}{1 - r_j^2}\left(2a_j^2\phi_j^2 + \frac{2a_j^2(1 - \rho)^2}{\rho^2}|\tilde{\nabla}\phi_j|^2 - \frac{(1 + r_j^2)e(f)}{\rho^2}\right) \\
&= 2\frac{1 - \rho_2}{1 - r_j^2}\left(\frac{a_j^2(1 - \rho)^2|\tilde{\nabla}\phi_j|^2}{\rho^2} + \frac{e(f)(a_j^2(1 - \rho)^2\phi_j^2 - 2a_j^2(1 - \rho)\phi_j)}{\rho^2}\right) \\
&\quad + \frac{1 - \rho^2}{1 - r_j^2}\left(2a_j^2\phi_j^2 - \frac{e(f)(1 + a_j^2)}{\rho^2}\right) \\
&= 2\frac{1 - \rho^2}{1 - r_j^2}\left(\frac{a_j^2(1 - \rho)^2|\tilde{\nabla}\phi_j|^2}{\rho^2} + \frac{f'^2(a_j^2(1 - \rho)^2\phi_j^2 - 2a_j^2(1 - \rho)\phi_j)}{\rho^2}\right) \\
&\quad + \frac{1 - \rho^2}{1 - r_j^2} \times \frac{2a_j^2\phi_j^2(\rho^2 - 1)}{\rho^2} ,
\end{aligned} \tag{5.14}$$

where we have used the definition of ϕ_j. Since for $\rho \geqq b$,

$$\tau(h_j)^1 = \frac{(1 - \rho^2)^2}{4}\Delta_0 r_j + \frac{r_j(1 - \rho^2)}{4}\tilde{\tau}(h_j)^1 ,$$

there exists a constant C_6 not depending on j, such that for $\rho \geqq b$

$$|\tau(h_j)^1| \leqq C_1(1 - \rho)^2 .$$

Similarly, one can show that there exists a constant C_7 not depending on j such that

$$|\tau(h_j)^2| \leqq C_7(1 - \rho)^2$$

for $\rho \geqq b$. Hence there exists a constant C_8 not depending on j, such that for $\rho \geqq b$,

$$\|\tau(h_j)\|^2 \leqq C_8(1 - \rho)^2 ,$$

where $\|\tau(h_j)\|$ is the norm of the tension field $\tau(h_j)$. Also one can check that, there is C_9, such that,

$$\|\tau(h_j)\| \leqq C_9$$

for $\rho \leqq b$. Let $p > 1$ be fixed, there exists C_{10} not depending on j, such that

$$\int_M \|\tau(h_j)\|^{2p} \leqq C_{10} . \tag{5.14}$$

Also there is a constant C_{11} such that

$$\sup_M e(h_j) \leqq C_{11} . \tag{5.15}$$

Therefore by Theorem 5.2, we can solve the parabolic equation for harmonic maps with initial data h_j. If we denote the solution to be u_j, then letting $t \to \infty$ we obtain a harmonic map $u_{j,\infty}$. As in the proof of Theorem 5.2, noting that the constants in (5.14) and (5.15) do not depend on j, we have

$$|(u_j)_t|^2(x, t) \leqq C_{12} \exp(-at) , \tag{5.16}$$

where C_{12} and $a > 0$ are independent of j. As before, one argues that

$$\sup_{M \times [0, \infty)} e(u_j) \leqq C_{13}$$

and

$$\operatorname{dist}(u_{j,\infty}(x), h_j(x)) \leqq C_{14}$$

for all $x \in M$. Therefore we have two kinds of convergence. Namely, by passing to a subsequence if necessary, noting that $h_j \to h$ as $j \to \infty$ uniformly on compact set, and by uniqueness of the parabolic equation,

$$\lim_{j \to \infty} u_j = u ,$$

where u is the solution of the parabolic equation with initial data h. Also

$$\lim_{j \to \infty} u_{j,\infty} = v ,$$

where v is a harmonic map. The convergence are uniform on compact sets, together with the first and second derivatives in the space variable, and the first derivative in the time variable for the functions u_j. We want to show that v is exactly the harmonic map u_∞ we get in Theorem 5.4. It suffices to show that $\operatorname{dist}(v(x), u_\infty(x)) \to 0$ as $x \to \infty$. Note that

$$\operatorname{dist}(u_{j,\infty}(x), h_j(x)) \leqq \int_0^\infty |(u_j)_t| \, dt . \tag{5.18}$$

By (5.16), given $\varepsilon > 0$ we can find $T > 0$ which is independent of j so that

$$\int_T^\infty |(u_j)_t| \, dt \leqq \varepsilon .$$

For a fixed x, $u_j(x, t)$ tends to $u(x, t)$ uniformly on $[0, T]$. Hence by letting $j \to \infty$ in (5.18), we obtain

$$\operatorname{dist}(v(x), h(x)) \leqq \varepsilon + \int_0^T |u_t| \, dt \ .$$

Therefore $\operatorname{dist}(v(x), h(x))$ tends to 0 as $x \to \infty$, and $v = u_\infty$.

Each $u_{j,\infty}$ can be considered as a harmonic map with boundary data h_j which is a homeomorphism onto the set $\{r = a_j\}$. By [S-Y2], u_j is a diffeomorphism in D and the Jacobian of u_j is nonzero everywhere on D, hence the Jacobian of u_∞ is either nonnegative everywhere, or nonpositive everywhere. We claim that u_∞ is a diffeomorphism. For $0 < \rho_0 < 1$, denote

$$D_{\rho_0} = \{x | \rho(x) < \rho_0\} \ ,$$

and for $0 < r_0 < 1$, denote

$$D_{r_0} = \{u | r(u) < r_0\} \ .$$

By (5.17) and the definitions of h_j, there exists $0 < \rho_1 < 1$ such that

$$u_{j,\infty}(D_{\rho_1}) \supset D_{1/2} \tag{5.19}$$

for all j. Also for any $0 < \rho_2 < 1$, there exists $0 < r_1 < 1$ such that

$$u_{j,\infty}(D_{\rho_2}) \subset D_{r_1}$$

for all j. Hence by Theorem 7.1 in [J2, p. 80], for any $\rho_1 \leqq \rho < 1$, there is a constant $C_{14} > 0$ which is independent of j such that for all $x \in D_\rho$,

$$|J_j(x)| \geqq C_{14} \ ,$$

where J_j is the Jacobian of $u_{j,\infty}$. Letting $j \to \infty$ and then $\rho \to 1$ we conclude that $|J(x)| > 0$ for all $x \in D$, where J is the functional determinant of u_∞. As in the proof of (5.19), one can show that u_∞ is onto. By (5.14) and the fact that u_∞ is a local diffeomorphism, for any point q in D the set $u_\infty^{-1}(q)$ is finite and u_∞ is a covering map. Hence u_∞ must be a diffeomorphism because D is simply connected.

6 Harmonic maps: Small total energy

In this section we would like to generalize a theorem of Mitteau [M], which states that if M and N are compact manifolds, then every $h \in C^1(M, N)$ with sufficiently small energy density can be deformed through the parabolic equation to a harmonic map. Note that in the compact case, a map with small energy density implies that the total energy is also small.

Theorem 6.1 *Let M^m be a complete noncompact Riemannian manifold with Ricci curvature bounded from below by $-K$ for $K \geqq 0$. Suppose there exists a constant $b > 0$ such that $\inf_{x \in M} V_x(1) = b > 0$. Let N^n be a complete Riemannian manifold with sectional curvature bounded from above by $\kappa > 0$. For any $\alpha > 0$, there exists $\varepsilon > 0$, depending only on m, K, b, κ, and α, such that if $h \in C^1(M, N)$ satisfying $\sup_M e(h) \leqq \alpha$, $E(h) \leqq \varepsilon$, and that $h(M)$ is bounded, then (0.1) has a solution u on*

$M \times [0, \infty)$ with initial condition $u(x, 0) = h(x)$. If in addition there exists a point $p \in M$ and a sequence $t_\nu \to \infty$ such that $u(p, t_\nu)$ converges in N (for example when N is compact), then by passing to a subsequence, $u(\cdot, t_\nu)$ converges uniformly on compact subsets of M to a harmonic map u_∞.

Proof. Let $h \in C^1(M, N)$ be such that $\sup_M e(h) \leqq \alpha$. In view of Theorem 3.3 and Remark 3.5, to prove long time existence of solution of the parabolic equation it suffices to show that if u is a solution of (0.1) on $M \times [0, T_0)$ satisfying the properties that $\sup_{M \times [0, t)} e(u) < \infty$, and $u(M \times [0, t))$ is bounded for all $0 < t < T_0$, then there exists $\varepsilon > 0$, depending only on m, K, b, κ, and α, such that if $E(h) < \varepsilon$ then $\sup_{M \times [0, T_0]} e(u) \leqq \beta$ for some $\beta > 0$ independent of T_0 and $u(M \times [0, T_0))$ is bounded.

The Bochner formula for the energy density $e(u(x, t))$ of u asserts that

$$\left(\Delta - \frac{\partial}{\partial t}\right) e \geqq -2Ke - 2\kappa e^2 . \tag{6.1}$$

For $\delta > 0$ sufficiently small, let us consider the function

$$g(t) = \frac{K}{\kappa}\left(\frac{1}{1 - 2\kappa e^{A + 2Kt}} - 1\right),$$

with $A = \log\left(\dfrac{\alpha + \delta}{2K + 2\kappa(\alpha + \delta)}\right)$. Obviously, g satisfies the differential equation

$$g' = 2Kg + 2\kappa g^2 ,$$

with initial condition $g(0) = \alpha + \delta$. The function $f(x, t) = g(t) - e(u(x, t))$ satisfies the differential inequality

$$\left(\Delta - \frac{\partial}{\partial t}\right) f \leqq -2Kf - 2\kappa f(g + e) ,$$

with $f(x, 0) \geqq \delta$. We claim that f must be positive on $M \times [0, t]$, for all $t < T_1 = \dfrac{-(A + \log 2\kappa)}{2K}$. If not, let $T < T_1$ be the first t such that $\inf_M f(x, T) = 0$. Using the facts that $f \geqq 0$ on $M \times [0, T]$ and g is an increasing function on $[0, T_1)$, we have

$$\left(\Delta - \frac{\partial}{\partial t}\right) f \leqq -2Kf - 2\kappa g(0) f .$$

This implies that the function $F = e^{-2(K + \kappa g(0))t} f$ satisfies

$$\left(\Delta - \frac{\partial}{\partial t}\right) F \leqq 0 ,$$

with initial condition $F(x, 0) \geqq \delta$. Hence $F \geqq \delta$ on $M \times [0, T]$, which is a contradiction. Therefore we may conclude that there exists a $t_0 > 0$ such that $e(u(x, t)) \leqq 2\alpha$, for all $x \in M$ and $t \in [0, t_0]$ whenever $u(x, t)$ is defined.

To estimate $e(u(x, t_1))$ for $t_1 \geqq t_0$, let $s = \sup_{M\times[0,t_1]} e(u(x, t))$. By the above discussion, we may assume that the value s is achieved at some $t_2 \in (t_0, t_1]$. Inequality (6.1) can be rewritten as

$$\left(\Delta - \frac{\partial}{\partial t}\right) e \geqq -2Ke - 2s\kappa e ,$$

on $M \times (0, t_1]$. Hence the function $h(x, t) = e^{-2(K+s\kappa)t} e(u(x, t))$ must satisfy

$$\left(\Delta - \frac{\partial}{\partial t}\right) h \geqq 0 .$$

For $t_3 \in (t_2 - t_0, t_2)$, let us solve the heat equation

$$\left(\Delta - \frac{\partial}{\partial t}\right) \phi = 0 ,$$

with

$$\phi(x, 0) = h(x, t_3) .$$

By the maximum principle of [K-L], $\phi(x, t_2 - t_3) \geqq h(x, t_2)$. Now

$$\begin{aligned} \phi(x, t_2 - t_3) &= \int_M H(x, y, t_2 - t_3)\phi(y, 0)\, dy \\ &\leqq \sup_{y\in M} H(x, y, t_2 - t_3) \int_M h(y, t_3)\, dy \\ &\leqq \sup_{y\in M} H(x, y, t_2 - t_3)\varepsilon e^{-2(K+s\kappa)t_3} , \end{aligned} \tag{6.2}$$

where we have used the fact (see [Lo-T]) that the total energy is a non-increasing function of t.

On the other hand, the upper estimate of the heat kernel in [L-Y] asserts that

$$\begin{aligned} H(x, y, t) &\leqq H^{-1/2}(x, x, t) H^{-1/2}(y, y, t) \\ &\leqq C_1 V_x(\sqrt{t})^{-1/2} V_y(\sqrt{t})^{-1/2} \exp(C_2 K t) , \end{aligned}$$

for some constants $C_1, C_2 > 0$ depending only on m. Substituting this into (6.2), the volume comparison theorem and the assumption on the volume of unit balls implies that there exists a constant $\gamma > 0$ which depends only on $t_2 - t_3$, b, and m, such that

$$\phi(x, t_2 - t_3) \leqq \gamma\varepsilon e^{-2(K+s\kappa)t_3} .$$

Hence

$$s e^{-2(K+s\kappa)(t_2-t_3)} \leqq \gamma\varepsilon .$$

One observes that the function $\psi(s) = s e^{-2(K+s\kappa)(t_2-t_3)}$ has the properties that $\psi(0) = 0$, $\psi(s) \to 0$ as $s \to \infty$, and ψ has exactly one maximum at $s_0 = (2\kappa(t_2 - t_3))^{-1}$ with maximum value at $\psi_0 = (2\kappa(t_2 - t_3))^{-1} \exp\{-2K(t_2 - t_3) - 1\}$. If we first choose t_3 sufficiently closed to t_2 such that $2\alpha \leqq s_0$, then for $\varepsilon < \dfrac{\psi_0}{\gamma}$ we may conclude that $s < s_0$. This proves the energy estimate with $\beta = s_0$.

In order to prove that $u(M \times [0, T_0))$ is bounded, let us first point out that Theorem 2.4 of [Lo-T] asserts that for $0 \leqq t < T < T_0$, we have

$$E(T) + 2\int_t^T d\tau \int_M |u_t|^2(x, \tau)\, dx \leqq E(t)\,, \tag{6.3}$$

where $E(t) = \int_M e(u)(\cdot, t)$. As in the proof of the main lemma in [H], we also have

$$\left(\Delta - \frac{\partial}{\partial t}\right)|u_t|^2 \geqq -2\kappa s_0 |u_t|^2 \tag{6.4}$$

on $M \times (0, T_0)$. Hence the function $g(x, t) = \exp(-2\kappa s_0 t)|u_t|^2(x, t)$ is a subsolution of the heat equation for functions on M. Let $R^2 \geqq T_0 + 1$ be fixed. By Theorem 1.1, for $\dfrac{T_0}{2} \leqq t < T_0$ and for $x \in M$, there exists a constant C_2 which is independent of x and t such that

$$g(x, t) \leqq \frac{C_3}{t V_x(R)} \int_0^t d\tau \int_{V_x(R)} |u_t|^2(y, \tau)\, dy\,. \tag{6.5}$$

Combining this with (6.4) and the assumption that $\inf_{x \in M} V_x(1) = b$, we conclude that

$$\sup_{M \times \left[\frac{T_0}{2}, T_0\right)} g < \infty\,.$$

Hence

$$\sup_{M \times \left[\frac{T_0}{2}, T_0\right)} |u_t|^2 < \infty\,.$$

From the fact that $u\left(M \times \left[\dfrac{T_0}{2}, T_0\right)\right)$ is bounded, the above estimate imply that $u(M \times [0, T_0))$ is also bounded in N. Hence we have long time existence.

Since the energy density of u is uniformly bounded, in order to prove that if $u(p, t_\nu)$ converges for some p and some sequence t_ν, it remains to prove that $|u_t| \to 0$ as $t \to \infty$. Using (6.3) and (6.5), as in the proof of (6.5), one can show that there exists a constant C_4 depending only on κ and β which is the upper bound of $e(u)$ on $M \times [0, \infty)$ such that for all $t > 1$ and $x \in M$

$$|u_t|^2(x, t) \leqq \frac{C_4}{V_x(4)}(E(t-1) - E(t))\,.$$

Since $E(t)$ is non-increasing and is bounded, $\lim_{t \to \infty} E(t)$ has a limit. So $E(t-1) - E(t)$ tends to zero as $t \to \infty$. This implies that $|u_t|^2(x, t)$ tends to zero as $t \to \infty$, and the theorem follows.

In some cases we can remove the restriction that the image of the initial data is bounded.

Definition 6.2. A complete Riemannian manifold N is said to *satisfy condition* (A), if there exists a compact exhaustion $\{\Omega_\nu\}$ of N and C^∞ maps g_ν from N into itself such that $g_\nu|_{\Omega_\nu} = \text{identity}$, $g_\nu(M)$ is compact, and $\sup_{\nu, N} e(g_\nu) < \infty$.

Obviously if N is compact then it satisfies condition (A). By Lemma 4.3, if N has nonpositive curvature then it satisfies condition (A). Also, if $N = N_1 \times N_2$ with N_1 and N_2 satisfying condition (A), then so does N. In particular, by [C-G], if N is the universal covering of a compact manifold with nonnegative Ricci curvature then N satisfies condition (A).

Theorem 6.3 *Theorem* 6.1 *is still true without assuming that* $h(M)$ *is bounded, provided that* N *satisfies condition* (*A*). *The constant* ε *also depends on the compact exhaustion* $\{\Omega_\nu\}$ *and the maps* g_ν.

Proof. Since N satisfies condition (A), there exists a compact exhaustion $\{\Omega_\nu\}$ of N and C^∞ maps g_ν from N into itself such that $g_\nu|_{\Omega_\nu} = identity$, $g_\nu(M)$ is compact, and

$$\sup_{\nu, N} e(g_\nu) = C_5 < \infty \ . \tag{6.6}$$

As in the proof of Theorem 4.1′, let $h_\nu = g_\nu \circ h \in C^1$. Note that $h_\nu(M)$ is bounded. Suppose $\sup_M e(h) \leqq \alpha$, noting that $e(h_\nu)(x) \leqq C_5 e(h)(x)$ by (6.6), so

$$\sup_{\nu, M} e(h_\nu) \leqq C_5 \alpha \ .$$

Hence by Theorem 6.1, there exists a constant $\varepsilon > 0$ such that if $E(h_\nu) \leqq C_5 \varepsilon$ then (0.1) has a solution u_ν on $M \times (0, \infty)$ and $u_\nu(x, 0) = h_\nu(x)$. Suppose $E(h) \leqq \varepsilon$ then $E(h_\nu) \leqq C_5 \varepsilon$, and the solutions u_ν exist. From the proof of Theorem 6.1 we see that

$$\sup_{M \times [0, \infty)} e(u_\nu) \leqq \beta < \infty \tag{6.7}$$

for some $\beta > 0$ for all ν. By [H], we see that

$$\left(\Delta - \frac{\partial}{\partial t} \right) |(u_\nu)_t|^2 \geqq -2\kappa\beta |(u_\nu)_t|^2 \ .$$

Hence the function $\exp(-2\kappa\beta)|(u_\nu)_t|^2(x, t)$ is a subsolution of the heat equation for functions on M. As in the proof of Theorem 4.1′, one can show that given a bounded domain Ω in M and $T > 0$ there exists a constant C_6 which is independent of ν such that

$$|(u_\nu)_t|(x, t) \leqq C_6 t^{-1/2} \tag{6.8}$$

for $x \in \Omega$ and $0 < t \leqq T$. From (6.7) and (6.8), we conclude that passing to a subsequence, u_ν converges to a solution u of (0.1) in $M \times (0, \infty)$. Since $g_\nu|_{\Omega_\nu} = identity$, by the definition of h_ν, the fact that $u_\nu(x, 0) = h_\nu(x)$, and (6.8), we conclude that $u(x, 0) = h(x)$. Since the convergence of u_ν together with their derivatives is uniform on compact subsets of $M \times (0, \infty)$, it is easy to see that $\sup_{M \times [0, \infty)} e(u) \leqq \beta$ and $E(u(\cdot, t)) \leqq C_5 \varepsilon$, where we have used the fact that the total energy is nonincreasing in t. Hence, as in the proof of Theorem 6.1, if there exists a point $p \in M$ and a sequence $t_j \to \infty$ such that $u(p, t_j)$ converges in N, then by passing to a subsequence, $u(\cdot, t_j)$ converges uniformly on compact subsets of M to a harmonic map u_∞. This completes the proof of the theorem.

Remark 6.4 From the proof of the Theorem 6.1 and Theorem 6.3, it is easy to see that if $\kappa = 0$, then ε can be any finite number. This recovers one of the result in

[Lo-T], that any map from a complete noncompact manifold to a compact manifold with nonpositive curvature can be deformed to a harmonic map through the parabolic equation, for the case that the domain manifold has Ricci curvature is bounded below.

In [S-Y1], it was shown that any harmonic map with finite total energy from a complete manifold with nonnegative Ricci curvature to a complete manifold with nonpositive sectional curvature must be constant. When M is $\mathbf{R}^m$, Sealey [Se] proved that any harmonic map with finite total energy into an arbitrary manifold N must be constant provided that $m \geqq 3$. Struwe [St] showed that any harmonic map with bounded energy density from $\mathbf{R}^m$ to a compact manifold is constant if the total energy is small enough. In view of Sealey's theorem, Struwe's result is only interesting when $m = 2$. We would like to point out that Sealey's result does not hold for general complete manifolds with nonnegative Ricci curvature. The following theorem is a generalization of these results.

Theorem 6.5 *Let M be a complete noncompact manifold with nonnegative Ricci curvature. Suppose there exists a constant $C_7 > 0$ such that $V_x(r) \geqq C_7 r^2$ for all $x \in M$ and for all $r \geqq 1$. Let N be a complete manifold with sectional curvature bounded above by $\kappa \geqq 0$. Given $\alpha > 0$ there exists $\varepsilon > 0$ depending only on κ, C_7 and α such that if u is a harmonic map from M to N with $\sup_M e(u) \leqq \alpha$ and $E(u) < \varepsilon$ then u is constant.*

Proof. Let $\sup_M e(u) = \beta$. As in [H], we have

$$\Delta e(u) \geqq -2\kappa(e(u))^2 \geqq -2\kappa\beta e(u) .$$

Let us consider $e(x) = e(u)(x)$ as a function of x and t, which is independent of time t, then the function $\exp(-2\kappa\beta t)e(x)$ is a subsolution of the heat equation for functions on M. By the maximum principle, the estimate of the heat kernel in [L-Y], and the assumption on the volume growth of M, we have for all $t \geqq 1$ and for all $x \in M$

$$\begin{aligned} \exp(-2\kappa\beta t)e(x) &= \int_M H(x, y, t)e(y)\,dy \\ &\leqq C_8 V_x^{-1}(\sqrt{t})E(u) \\ &\leqq \frac{C_8}{C_7} t^{-1} E(u) , \end{aligned}$$

where C_8 is an absolute constant. Hence

$$\beta t \exp(-2\kappa\beta t) \leqq \frac{C_8}{C_7} E(u) ,$$

for $t \geqq 1$. Suppose $\alpha \geqq \beta > 0$. If we let $t = \dfrac{\alpha}{\beta} \geqq 1$ in the above inequality then we will have a contradiction if $E(u)$ is small.

Example. We would like to give an example to show that in general the harmonic map obtained in Theorem 6.1 is not constant. More precisely, in the notations in Theorem 6.1, we can find K, κ, b and α, such that given $\varepsilon > 0$, there exists complete surfaces M and N and a harmonic map from M to N, with the following properties:

(1) the curvature of M is bounded below by $-K$, $\inf_{x\in M} V_x(1) > b > 0$, and the curvature of N is bounded above by κ; (2) $\sup_M e(u) \leqq \alpha$, $E(u) < \varepsilon$ and u is not constant.

Let M be a complete surface which is diffeomorphic to a torus minus a point so that its curvature near infinity is zero. Note that $\inf_{x\in M} V_x(1) > 0$. Let $\mathbf{S}^1$ be a circle, with radius not necessarily equal to 1. The one point compactification $\bar{M}$ of M is diffeomorphic to a torus. For $h \in C(\bar{M}, N)$, h can be deformed to a smooth map which is constant near infinity of M. Hence, we may assume h has this property. By Theorem 6.1 and Remark 6.4, h can be deformed through the parabolic equation to a harmonic map u_∞. Note that the energy density of u_∞ is bounded, and $E(u_\infty)$ is finite. We claim that for some h, u_∞ is not constant. Otherwise, we would have $H_1(\bar{M}, \mathbf{Z}) = 0$ which is a contradiction. This proves the claim. If we let the radius of the circle $\mathbf{S}^1$ goes to zero, then $E(u_\infty)$ can be made arbitrarily small. However, we can embed $\mathbf{S}^1$ into a dumb bell shape surface N so $\mathbf{S}^1$ is a geodesic and the curvature of N is bounded above by a constant independent of the radius of the circle. The composition of u_∞ with the embedding clearly is a harmonic map of M into N satisfying properties (1) and (2).

References

[A] Akutagawa, K.: Harmonic diffeomorphisms of the hyperbolic plane. preprint

[A-C-M] Avilés, P., Choi, H.I., Micallef, M.: Boundary behavior of harmonic maps on non-smooth domains and complete negatively curved manifolds. preprint

[C-T1] Choi, H.I., Treibergs, A.: New examples of harmonic diffeomorphisms of hyperbolic plane. Manuscr. Math. **62**, 249–256 (1988)

[C-T2] Choi, H.I., Treibergs, A.: Gauss map of spacelike constant mean curvature hypersurfaces of Minkowski space. to appear in J. Diff. Geom.

[C-G] Cheeger, J., Gromoll, D.: The splitting theorem for manifolds of nonnegative Ricci curvature. J. Diff. Geom. **6**, 119–128 (1971)

[C-St] Chen, Y., Struwe, M.: Existence and partial regularity results for the heat flow for harmonic maps. preprint

[D-L] Donnelly, H., Li, P.: Heat equation and compactifications of complete Riemannian manifolds. Duke Math. J. **51**, 667–673 (1984)

[E-L1] Eells, J., Lemaire, L.: A report on harmonic maps. Bull. Lond. Math. Soc. **10**, 1–68 (1978)

[E-L2] Eells, J., Lemaire, L.: Another report on harmonic maps. Bull. Lond. Math Soc. **20**, 385–524 (1988)

[E-S] Eells, J., Sampson, J.: Harmonic mappings of Riemannian manifolds. Am. J. Math., **86**, 109–160 (1964)

[Hm] Hamilton, R.: Harmonic maps of manifolds with boundary. Lect. Notes Math., Vol. 471. Berlin Heidelberg New York: Springer 1975

[H] Hartman, P.: On homotopic harmonic maps. Can. J. Math. **19**, 673–687 (1967)

[J1] Jost, J.: Harmonic mappings between Riemannian manifolds. Proc. Cent. Math. Anal. Aust. Natl. Univ. (1983)

[J2] Jost, J.: Harmonic mappings between surfaces. Lect. Notes Math., vol. 1062. Berlin Heidelberg New York: Springer 1984

[J-Y1] Jost, J., Yau, S.T.: A strong rigidity of locally symmetric complex manifolds of rank one and finite volume. Math. Ann. **275**, 291–304 (1986)

[J-Y2] Jost, J., Yau, S.T.: On the rigidity of certain discrete groups and algebraic varieties. Math. Ann. **278**, 481–486 (1987)

[K] Kendall, W.: Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. Proc. London Math. Soc. (3) **61**, 371–406 (1990)

[K-L] Karp, K., Li, P.: The heat equation on complete Riemannian manifolds. unpublished

[L-S-U] Ladyzenskaja, O.A., Solonnikov, V.A., Ural'ceva, N.N.: Linear and quasilinear equations of parabolic type. AMS Transl. Math. Monographs, vol. 23. Providence 1968

[L] Li, P.: Uniqueness of L^1 solutions for the Laplace equation and the heat equation on Riemannian manifolds. J. Diff. Geom. **20**, 447–457 (1984)

[L-S] Li, P., Schoen, R.: L^p and mean value properties of subharmonic functions on Riemannian manifolds. Acta Math. **153**, 279–301 (1984)

[L-Y] Li, P., Yau, S.-T.: On the parabolic kernel of the Schrödinger operator. Acta Math. **156**, 153–201 (1986)

[Lo-T] Liao, G., Tam, L.-F.: On the heat equation for harmonic maps from non-compact manifolds. to appear in Pacific J. Math

[M] Mitteau, J.-C.: Sur les applications harmoniques. J. Diff. Geom. **9**, 41–54 (1974)

[S] Schoen, R.: Analytic aspect of the harmonic map problem. MSRI Publ. **2**, 321–358 (1984)

[S-Y1] Schoen, R., Yau, S.-T.: Harmonic maps and the topology of stable hypersurfaces and manifolds of nonnegative Ricci curvature. Comm. Math. Helv. **39**, 333–341 (1976)

[S-Y2] Schoen, R., Yau, S.-T.: On univalent harmonic maps between surfaces. Invent. Math. **44**, 265–278 (1978)

[S-Y3] Schoen, R., Yau, S.-T.: Compact group actions and the topology of manifolds with nonpositive curvature. Topology, **18**, 361–380 (1979)

[Se] Sealey, H.C.J.: Some conditions ensuring the vanishing of harmonic differential forms with applications to harmonic maps Yang-Mills theory. Math. Proc. Camb. Phil. Soc. **91**, 441–452 (1982)

[Sh-Y] Shi, W., Yau, S.-T.: Harmonic maps on complete noncompact Riemannian manifolds, preprint

[St] Struwe, M.: On the evolution of harmonic maps in higher dimensions. J. Diff. Geom. **28**, 485–502 (1988)

[Y1] Yau, S.-T.: Some function-theoretic properties of complete Riemannian manifolds and their applications to geometry. Indiana Univ. Math. J. **25**, 659–670 (1976)

[Y2] Yau, S.-T.: On the heat kernel of a complete Riemannian manifold. J. Math. Pure Appl. **57**, 191–201 (1978)

Invent. math. 105, 47–122 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

The cohomology of *p*-adic symmetric spaces

P. Schneider[1] **and U. Stuhler**[2]

[1]Mathematisches Institut der Universität zu Köln, Weyertal 86-90, W-5000 Köln 41, Federal Republic of Germany
[2]Fachbereich Mathematik Universität-6HS Wüppertal, Gaußstr. 20, W-5600 Wuppertal 1, Federal Republic of Germany

Oblatum 27-VI-1990

In 1972 Mumford [37] exhibited a whole class of algebraic curves over a local field K which can be uniformized by a p-adic "upper half plane". This was taken up by Drinfeld in his study of the moduli varieties of elliptic modules [16]. He showed that the spaces $\Omega^{(d+1)}$ which arise from projective d-space over K by removing all rational hyperplanes have a natural rigid analytic structure. Later on [17] he gave a new construction of $\Omega^{(d+1)}$ as a moduli space for formal groups. As an application he obtained a new proof of a result of Cherednik describing explicitly the p-adic uniformization of certain Shimura curves. Mustafin [39] proved that for arbitrary $d \geqq 1$ the quotients of $\Omega^{(d+1)}$ by cocompact discrete groups are projective algebraic varieties. Rapoport and Zink [41] used the p-adic uniformization theory in their study of certain higher dimensional Shimura varieties.

It seems quite natural to view the spaces $\Omega^{(d+1)}$ as p-adic analogs of real symmetric spaces. But there is at least one essential difference. Although $\Omega^{(d+1)}$ is analytically simply connected [40] it still has many etale coverings and is far from being cohomologically trivial. Drinfeld in [16] computed explicitly the first etale cohomology group of $\Omega^{(2)}$. And in [17] he constructed a natural family of equivariant etale coverings of $\Omega^{(d+1)}$ expressing the fascinating hope that their cohomology realizes all discrete series representations of $\mathrm{GL}_{d+1}(K)$.

In the first part of this paper we compute the cohomology of $\Omega^{(d+1)}$ for arbitrary d in any good cohomology theory. Our main result is that in degrees $\leqq d$ the cohomology realizes certain explicitly described admissible representations of $\mathrm{GL}_{d+1}(K)$; it vanishes in degrees $> d$. The specific cohomology theory only enters into the formula through its coefficients $A := H^0(\mathit{point})$. For example the highest cohomology of $\Omega^{(d+1)}$ is the A-dual of the Steinberg representation. In case of etale cohomology the Galois action turns out to be pure and even by powers of the cyclotomic character.

In the second part we then use this result in order to study the cohomology of quotient varieties $X_\Gamma := \Gamma \backslash \Omega^{(d+1)}$ by cocompact discrete groups $\Gamma \subseteq \mathrm{PGL}_{d+1}(K)$. We obtain that X_Γ has primitive cohomology only in middle degree. Furthermore this middle cohomology has a filtration the subquotients of which are pure and have dimensions which depend in a very simple way on a single invariant. This is the multiplicity of the Steinberg representation in the representation induced from

the trivial character on Γ. Of course, this filtration has to be Deligne's monodromy filtration but we cannot prove this.

In the last part we develop a general formalism for constructing natural resolutions with good finiteness properties for smooth $\mathrm{GL}_{d+1}(K)$-modules. This is needed as a second essential ingredient in the computation of the cohomology of the varieties X_Γ.

For the convenience of the reader we now describe the content of the individual Paragraphs in more detail. In the first Paragraph we discuss the rigid analytic structure of $\Omega^{(d+1)}$. We introduce certain neighbourhoods of a rational hyperplane in projective space over K and study the geometry of intersections of those neighbourhoods. The reason for considering them is that by the local compactness of the field K already finitely many of them contain all infinitely many rational hyperplanes. One then can try to imitate classical techniques for computing the singular cohomology of a complex vector space with finitely many hyperplanes removed (compare [10]). We also prove that $\Omega^{(d+1)}$ is a Stein-space in the strong sense of Kiehl.

The second Paragraph contains general considerations from homological algebra. We introduce the notion of an abstract cohomology theory on the category of smooth rigid analytic varieties over K which is equipped axiomatically with four basic properties. The essential ones among them are the homotopy invariance and the validity of the usual formula for the cohomology of projective space. Those four properties will suffice to compute the cohomology of $\Omega^{(d+1)}$. Using results of Kiehl we show that the de Rham cohomology in characteristic 0 has the required properties. We then give a careful definition of the etale topology in the rigid analytic context. Here the necessary properties beyond any doubt hold true, too, but unfortunately there exist no published proofs for the two essential ones. At least there is unpublished work of Gabber which seems to lead to the wanted results. In the second half of the Paragraph we establish a general spectral sequence which allows to compute the cohomology of a pair $U \subseteq X$ in terms of the cohomology of the pairs $U_{i_0} \cup \ldots \cup U_{i_r} \subseteq X$ whenever U is the intersection of the U_i's.

In the third Paragraph we will use this spectral sequence in order to express the cohomology of $\Omega^{(d+1)}$ in terms of the relative cohomology of the projective space with respect to the complement of the intersection of finitely many hyperplane neighbourhoods. In turns out that our axioms together with the simple geometry of those hyperplane neighbourhoods lead to an explicit computation of the E_1-terms in this spectral sequence. A closer inspection of the d_1-differential then shows that the lines in the spectral sequence are closely related to the simplicial cohomology of certain generalized topological Tits buildings $\mathscr{T}_\bullet^{(s)}$ for $\mathrm{GL}_{d+1}(K)$. We show that the $\mathscr{T}_\bullet^{(s)}$ have cohomology only in highest degree which implies the degeneration of our spectral sequence. From this it is easy to deduce our first main result (Theorem 1) that

$$H^s(\Omega^{(d+1)}) = \begin{cases} A & \text{if } s = 0\ , \\ \mathrm{Hom}_{\mathbb{Z}}(\tilde{H}^{s-1}(|\mathscr{T}_\bullet^{(s)}|, \mathbb{Z}), A) & \text{if } 1 \leqq s \leqq d\ , \\ 0 & \text{if } s > d\ . \end{cases}$$

The building $\mathscr{T}_\bullet^{(d)}$ together with a computation of its cohomology already appears in [3].

In the fourth Paragraph we give two different explicit descriptions of the reduced cohomology groups $\tilde{H}^{s-1}(|\mathscr{T}_\bullet^{(s)}|, \mathbb{Z})$. Extending a technique in [3] we first construct certain filtrations on these groups the subquotients of which can be identified naturally with spaces of locally constant functions with compact support on cells in flag manifolds. In the case of de Rham cohomology we obtain in this way a natural map — a kind of p-adic Mellin transform — which associates with any closed s-form on $\Omega^{(d+1)}$ a K-valued distribution on the $s/2 \cdot (2d+1-s)$-dimensional affine space over K (Corollary 6). Secondly we identify $H^s(\Omega^{(d+1)})$ with an explicitly defined subspace of the space of s-dimensional cochains on the Bruhat–Tits building $\mathscr{B}\mathscr{T}$ of $\mathrm{SL}_{d+1}(K)$ (Corollary 17). In the case $s = d$ the resulting subspace is the space of harmonic cochains in the sense of [19].

The fifth Paragraph contains the computation of the cohomology of the quotient varieties X_Γ. Of course we use a Hochschild–Serre type spectral sequence for the covering $\Omega^{(d+1)} \to X_\Gamma$. The E_2-terms can be transformed into Ext-groups in the category of smooth $\mathrm{PGL}_{d+1}(K)$-representations. In order to compute these Ext-groups we generalize ideas of Casselman [12, 13] which he used in his representation theoretic proof of the Garland vanishing theorem [19]. There are two additional features: One is that for the representations which appear in the E_2-terms we need resolutions with nice finiteness properties with respect to the Γ-action. Their construction is postponed to Paragraph 6 (Proposition 16). Second we establish a general cohomological duality (Duality Theorem) in the category of smooth representations in which the role of the dualizing object is played by the Steinberg representation. This result resembles very much the Borel/Serre duality for discrete groups [3] but it does not seem to be a formal consequence of it. The final results (Theorems 4 and 5) were already explained above. In the rest of the Paragraph we deduce from our computation of $H^1_{DR}(\Omega^{(2)})$ a proof of the p-adic Shimura isomorphism between modular forms and group cohomology (a corrected version of the corresponding statement in [44]). We do this by a careful analysis of certain p-adic representations of $\mathrm{SL}_2(K)$. In particular, we study the nonexistence of Γ-invariant continuous linear forms on Morita's p-adic principal series of $\mathrm{SL}_2(K)$ (Theorem 6). As another application of this we will see the surprising fact that a cocompact discrete subgroup $\Gamma \subseteq \mathrm{SL}_2(K)$ has no nonvanishing automorphic forms of weight 1 (Corollary 13). A much more direct proof of the p-adic Shimura isomorphism was given by de Shalit [46]; he also pointed out that the original form of the statement in [44] does not hold.

The Bernstein–Borel–Matsumoto theory of smooth representations generated by their Iwahori-fixed vectors (see, e.g., [2]) would provide us rather easily with (non-canonical) resolutions of the type which we needed in Paragraph 5 if it would be sufficient for our purposes to work with $\mathbb{Q}$-coefficients. But the $\mathrm{GL}_{d+1}(K)$-modules we have to study have $\mathbb{Z}$- or at least $\mathbb{Z}[1/p]$-coefficients. The theory unfortunately breaks down in this situation. Instead we start in the last Paragraph with the observation that any smooth $\mathrm{GL}_{d+1}(K)$-module in a completely natural way gives rise to a coefficient system on the Bruhat–Tits building $\mathscr{B}\mathscr{T}$. The chain complex of this coefficient system is a good candidate for the required resolution. For coefficient systems of a certain type on an arbitrary contractible simplicial complex $\mathscr{K}$ we develop a criterion which reduces the homological triviality of the coefficient system to the question whether certain subcomplexes of $\mathscr{K}$ (defined in terms of the coefficient system) are contractible (Proposition 1). In the concrete situation on $\mathscr{B}\mathscr{T}$ to which we want to apply this criterion we explicitly describe the respective subcomplexes as unions of certain apartments in $\mathscr{B}\mathscr{T}$. This makes it

possible to apply the geodesic action on $\mathscr{BT}$ to show their contractibility. In this way we prove the chain complexes in question to be exact (Theorems 7 and 8).

Acknowledgements. We are especially grateful to P. Deligne. He pointed out an erroneous argument in our proof of §3 Proposition 5 and he suggested the present version of the proof. We also want to thank O. Gabber and M. Rapoport for several interesting conversations.

1 The rigid analytic space $\Omega^{(d+1)}$

Once and for all we fix a local field K, that is a field K which is locally compact with respect to a non-trivial non-archimedean valuation $|\ |$. We always will assume the valuation $|\ |$ to be normalized. Let $p > 0$ be the characteristic of the finite residue class field of K, o be the valuation ring of K, and $\hat{\bar{K}}$ be the completion of a fixed algebraic closure $\bar{K}$ of K. We also fix a natural number d and denote by $\mathbb{P}^d_{/K}$, as usual, the d-dimensional projective space viewed as a K-analytic variety.

If $\mathscr{H}$ is the set of all K-rational hyperplanes in $\mathbb{P}^d_{/K}$ then our object of interest is

$$\Omega^{(d+1)} := \mathbb{P}^d_{/K} \setminus \bigcup_{H \in \mathscr{H}} H\ .$$

Proposition 1 *$\Omega^{(d+1)}$ is an admissible open subset and consequently an open analytic subvariety of $\mathbb{P}^d_{/K}$.*

We indicate two proofs and explicitly give a third one.

(A) The first proof is taken from [16]. As this will be of use later on, we give some of the ideas behind it. Of central importance is the close relationship between $\Omega^{(d+1)}$ and the Bruhat–Tits building $\mathscr{BT}$ of the group $\mathrm{SL}_{d+1}(K)$.

Definition ([9]) $\mathscr{BT}$ is the simplicial complex whose vertices are the similarity classes $[L]$ of o-lattices in the vector space K^{d+1} and whose q-simplices are given by families $\{[L_0], \ldots, [L_q]\}$ of similarity classes such that

$$L_0 \subsetneq L_1 \subsetneq \ldots \subsetneq L_q \subsetneq \pi^{-1} L_0$$

where π is a uniformizing element in K.

Remarks.

(i) $\mathscr{BT}$ is contractible.

(ii) $\mathrm{GL}_{d+1}(K)$ in a natural way acts on $\mathscr{BT}$.

(iii) ([21]) The topological space of similarity classes of real norms on K^{d+1} can be identified $\mathrm{GL}_{d+1}(K)$ — equivariantly with the geometric realization $|\mathscr{BT}|$ of the simplicial complex $\mathscr{BT}$.

On the other hand, any point $z = [z_0 : \ldots : z_d] \in \Omega^{(d+1)}(\hat{\bar{K}})$ defines a similarity class of norms $\| \ \|_{\rho(z)}$ by

$$\|w\|_{\rho(z)} := \left| \sum_{i=0}^{d} w_i z_i \right| \qquad \text{for } w = (w_0, \ldots, w_d) \in K^{d+1}\ .$$

In this way we obtain a $\mathrm{GL}_{d+1}(K)$-equivariant map

$$\rho\colon \Omega^{(d+1)} \to |\mathscr{BT}|\ .$$

With the help of this map Drinfel'd constructs an explicit family $(U_i)_{i\in I}$ of open affinoid subvarieties of $\mathbb{P}^d_{/K}$ such that

a. $\bigcup_{i\in I} U_i = \Omega^{(d+1)}$ and

b. any K-morphism $f: Y \to \mathbb{P}^d$ from a K-affinoid variety into $\mathbb{P}^d_{/K}$ with $f(Y) \subseteq \Omega^{(d+1)}$ factorizes through some U_i.

This in particular shows (compare [6] 9.1.4 Prop. 2) that $\Omega^{(d+1)}$ is an admissible open subspace of $\mathbb{P}^d_{/K}$.

(B) In the second proof one constructs a formal scheme $\hat{\Omega}^{(d+1)}$ over Spf (o) of which $\Omega^{(d+1)}$ is the "generic fibre". For details see [17, 37, 39, 30, 42, 32].

(C) The third proof proceeds again by constructing a certain explicit family of admissible open subvarieties of $\mathbb{P}^d_{/K}$ with the properties a. and b. above. Precisely this covering of $\Omega^{(d+1)}$ later on will be the main technical tool in our computations. First we introduce the following convention: If not indicated otherwise any coordinate representation $z = [z_0 : \ldots : z_d]$ of a point $z \in \mathbb{P}^d(\hat{\bar{K}})$ is assumed to be unimodular, i.e., such that $|z_i| \leqq 1$ for $0 \leqq i \leqq d$ and $|z_i| = 1$ for some i. Similarly, for any hyperplane $H \in \mathscr{H}$, we let $l_H \in L_0^*$ with $L_0 := o^{d+1}$ always be a unimodular vector (determined up to a unit in o) such that $H(\hat{\bar{K}})$ is its zero set

$$H(\hat{\bar{K}}) = \{z \in \mathbb{P}^d(\hat{\bar{K}}) : l_H(z) = 0\} .$$

Definition. If $\varepsilon > 0$ is a rational number the set

$$H(\varepsilon) := \{z \in \mathbb{P}^d(\hat{\bar{K}}) : |l_H(z)| \leqq \varepsilon\}$$

is called the *ε-neighbourhood* of the hyperplane $H \in \mathscr{H}$.

Because of our above convention this definition obviously is independent of the particular choices of l_H and of the coordinate representation for z. For simplicity, we let $H(\varepsilon)$ also denote the subset of $\mathbb{P}^d_{/K}$ which we get by identifying (over K) conjugate points. Usually ε will be of the form $\varepsilon = |\pi|^n$ for some $n \in \mathbb{N}$ where π is a fixed uniformizing element of o.

Definition. Two hyperplanes $H, H' \in \mathscr{H}$ are called *congruent* mod(π^n) for some $n \in \mathbb{N}$ if appropriate representing vectors l_H and $l_{H'}$ satisfy the congruence

$$l_H \equiv l_{H'} \bmod \pi^n L_0^* .$$

If $\mathscr{H}_n$ denotes the set of equivalence classes of hyperplanes $H \in \mathscr{H}$ mod(π^n) then we have $\mathscr{H}_n = \mathbb{P}(L_0^*/\pi^n L_0^*)$ and $\mathscr{H} = \varprojlim \mathscr{H}_n$.

Lemma 2 *For two hyperplanes $H, H' \in \mathscr{H}$ we have $H(|\pi|^n) = H'(|\pi|^n)$ if and only if H and H' are congruent* mod(π^n).

Proof. Since l_H and $l_{H'}$ are unimodular they induce surjective linear maps

$$\bar{l}_H \qquad \text{and} \qquad \bar{l}_{H'}: L_0/\pi^n L_0 \to o/\pi^n o .$$

If we assume that $H(|\pi|^n) = H'(|\pi|^n)$ then $\bar{l}_H$ and $\bar{l}_{H'}$ have the same kernel. This implies that $\bar{l}_{H'} = \alpha \cdot \bar{l}_H$ for some $\alpha \in (o/\pi^n o)^\times$. The reverse implication easily follows from the inequality $|l_H(z)| \leqq \max\{|l_H(z) - l_{H'}(z)|, |l_{H'}(z)|\}$.

Lemma 3

$$\bigcup_{H\in\mathscr{H}_n} H(|\pi|^n) \supseteq \bigcup_{H\in\mathscr{H}} H\,.$$

Proof. Clear from Lemma 2.

We now consider the subsets

$$\Omega_n := \Omega(|\pi|^n) := \mathbb{P}^d_{/K} \setminus \bigcup_{H\in\mathscr{H}_n} H(|\pi|^n)$$

in $\Omega^{(d+1)}$. They are admissible open in $\mathbb{P}^d_{/K}$. Each Ω_n is a finite intersection of subsets of the form $\mathbb{P}^d_{/K}\setminus H(|\pi|^n)$. It therefore suffices to show that those subsets are admissible open. But up to isomorphism $\mathbb{P}^d_{/K}\setminus H(|\pi|^n)$ is an open polydisc in the affine d-space $\mathbb{P}^d_{/K}\setminus H$. The family $\{\mathbb{P}^d_{/K}\setminus H(|\pi|^n)\}_{n\in\mathbb{N}}$ even is an admissible covering of $\mathbb{P}^d_{/K}\setminus H$. Therefore, if $f\colon Y\to\mathbb{P}^d$ is any K-morphism from a K-affinoid variety Y into $\mathbb{P}^d$ such that $f(Y)\subseteq\mathbb{P}^d\setminus H$ then there exists a $n(H)\in\mathbb{N}$ such that $f(Y)\subseteq\mathbb{P}^d\setminus H(|\pi|^{n(H)})$. If we apply this to a morphism f such that $f(Y)\subseteq\Omega^{(d+1)}$ we see that

$$f(Y)\subseteq\mathbb{P}^d\setminus\bigcup_{H\in\mathscr{H}} H(|\pi|^{n(H)})\,.$$

Because of Lemma 2 the sets $\{H'\in\mathscr{H} : H'\subseteq H(|\pi|^{n(H)})\}$ are open in $\mathscr{H} = \varprojlim \mathscr{H}_n$. Since $\mathscr{H}$ is compact we conclude that we find finitely many hyperplanes $H_1,\ldots,H_r\in\mathscr{H}$ and numbers $n_1,\ldots,n_r\in\mathbb{N}$ such that

$$\bigcup_{H\in\mathscr{H}} H\subseteq H_1(|\pi|^{n_1})\cup\ldots\cup H_r(|\pi|^{n_r})\,.$$

As a consequence, for $n := \max n_j$, we get that

$$\bigcup_{H\in\mathscr{H}} H\subseteq\bigcup_{H\in\mathscr{H}_n} H(|\pi|^n)\subseteq\bigcup_j H_j(|\pi|^{n_j})$$

which, in particular, means that $f(Y)\subseteq\Omega_n$. We thus have established Proposition 1 and have shown that the increasing sequence of admissible open subvarieties Ω_n forms an admissible covering of $\Omega^{(d+1)}$.

Proposition 4 *$\Omega^{(d+1)}$ is a Stein-space.*

Proof. (See [29] §2 for the notion of Stein-space.) Similarly as before one shows that the increasing sequence of subsets

$$\bar{\Omega}_n := \{z\in\mathbb{P}^d(\hat{\bar{K}}) : |l_H(z)|\geqq|\pi|^n \text{ for all } H\in\mathscr{H}\}$$

in $\Omega^{(d+1)}$ forms an admissible covering by open affinoid subvarieties. For any pair H, H' in $\mathscr{H}$ we have the analytic function

$$f_{H,H'} := \frac{l_H}{l_{H'}} \qquad \text{on } \Omega^{(d+1)}\,.$$

For each n we choose a set $\bar{\mathscr{H}}_n$ of representatives for the equivalence classes of hyperplanes in $\mathscr{H}_{n+1}$ in such a way that it contains the coordinate hyperplanes

$H_i = \{z_i = 0\}$ for $0 \leqq i \leqq d$. It is then easy to see that

$$\begin{aligned}\bar{\Omega}_n &= \{z \in \Omega^{(d+1)} : |f_{H,H'}(z)| \leqq |\pi|^{-n} \text{ for all } H, H' \in \bar{\mathscr{H}}_n\} \\ &= \{z \in \Omega^{(d+1)} : |f_{H_i,H'}(z)| \leqq |\pi|^{-n} \text{ for all } 0 \leqq i \leqq d \text{ and } H' \in \bar{\mathscr{H}}_n\} \\ &= \{z \in \bar{\Omega}_{n+1} : |\pi^{n+1} f_{H,H'}(z)| \leqq |\pi| \text{ for all } H, H' \in \bar{\mathscr{H}}_{n+1}\} .\end{aligned}$$

The last description shows that in order to establish our assertion it suffices to prove that the functions $\pi^n f_{H,H'}$ for $H, H' \in \bar{\mathscr{H}}_n$ form a system of affinoid generators of $\mathscr{O}(\bar{\Omega}_n)$. This will be done by explicitly determining $\mathscr{O}(\bar{\Omega}_n)$. Define the affinoid K-algebra A_n to be the free Tate algebra over K in the indeterminates $T_{H,H'}$ for $H, H' \in \bar{\mathscr{H}}_n$ divided by the (closed) ideal generated by

$$T_{H,H} - \pi^n \qquad \text{for } H \in \bar{\mathscr{H}}_n ,$$

$$T_{H,H'} \cdot T_{H',H''} - \pi^n T_{H,H''} \qquad \text{for } H, H', H'' \in \bar{\mathscr{H}}_n , \quad \text{and}$$

$$T_{H,H_j} - \sum_{i=0}^{d} \lambda_i T_{H_i,H_j} \qquad \text{if } l_H(z) = \sum_{i=0}^{d} \lambda_i z_i \quad \text{for } H \in \bar{\mathscr{H}}_n \quad \text{and} \quad 0 \leqq j \leqq d .$$

We then have the K-morphisms

$$\phi_n : \bar{\Omega}_n \to \mathrm{Sp}(A_n) \quad \text{given by} \quad \begin{matrix} A_n \to \mathscr{O}(\bar{\Omega}_n) \\ T_{H,H'} \mapsto \pi^n f_{H,H'} \end{matrix}$$

and

$$\tilde{\psi}_n : \mathrm{Sp}(A_n) \to \mathbb{P}^d$$

$$x \mapsto [T_{H_0,H_j}(x) : \ldots : T_{H_d,H_j}(x)] \text{ (not necessarily unimodular)}$$

the latter being independent of the particular choice of $0 \leqq j \leqq d$. We leave it to the reader to check that the image of $\tilde{\psi}_n$ is contained in $\bar{\Omega}_n$ so that it factorizes through a K-morphism $\psi_n : \mathrm{Sp}(A_n) \to \bar{\Omega}_n$ and furthermore that ψ_n and ϕ_n are inverse to each other.

For later computations it is necessary to look more closely at the geometric nature of the admissible open subvarieties in $\mathbb{P}^d_{/K}$ of the form

$$\mathbb{P}^d_{/K} \backslash (H_0(|\pi|^n) \cap \ldots \cap H_r(|\pi|^n))$$

where $H_0, \ldots, H_r \in \mathscr{H}$ are finitely many hyperplanes. We consider the o-module

$$M := \sum_{i=0}^{r} o l_{H_i} \subseteq L_0^* .$$

Obviously we have

$$H_0(|\pi|^n) \cap \ldots \cap H_r(|\pi|^n) = \{z \in \mathbb{P}^d(\hat{\bar{K}}) : |l(z)| \leqq |\pi|^n \text{ for all } l \in M\} .$$

By the elementary divisor theorem we find a basis $l_0, \ldots, l_d$ in L_0^* such that $\pi^{\alpha_0} l_0, \ldots, \pi^{\alpha_m} l_m$ with appropriate integers $0 \leqq m \leqq d$ and $0 \leqq \alpha_0 \leqq \ldots \leqq \alpha_m$ is a basis of M. Therefore, up to a linear transformation, we can assume that

$$H_0(|\pi|^n) \cap \ldots \cap H_r(|\pi|^n) = \{[z_0 : \ldots : z_d] \in \mathbb{P}^d(\hat{\bar{K}}) : |\pi^{\alpha_i} z_i| \leqq |\pi|^n \text{ for } 0 \leqq i \leqq m\} .$$

Since M contains a unimodular vector we must have $\alpha_0 = 0$. On the other hand, the condition $|\pi^{\alpha_i} z_i| < |\pi|^n$ is automatically fulfilled if $\alpha_i \geqq n$. Writing $\beta_i := n - \alpha_i$ we can state the following result.

Lemma 5 *There are integers* $0 \leqq s \leqq d$ *and* $n = \beta_0 \geqq \beta_1 \geqq \ldots \geqq \beta_s > 0$ *such that, up to K-linear isomorphism,*

$$H_0(|\pi|^n) \cap \ldots \cap H_r(|\pi|^n) = \{[z_0 : \ldots : z_d] \in \mathbb{P}^d(\hat{\bar{K}}) : |z_i| \leqq |\pi|^{\beta_i} \text{ for } 0 \leqq i \leqq s\} .$$

The integer s in the above assertion can be intrinsically characterized in the following way: Put

$$\bar{M} := \sum_{i=0}^{r} (o/\pi^n o) l_{H_i} \subseteq L_0^*/\pi^n L_0^* .$$

The minimal number of generators of $\bar{M}$ as an o-module is called the rank of $\bar{M}$. We have

$$\operatorname{rank} \bar{M} = s + 1 .$$

Proposition 6 *We have a (in the rigid sense) locally trivial fibration*

$$\mathbb{P}^d \backslash (H_0(|\pi|^n) \cap \ldots \cap H_r(|\pi|^n)) \to \mathbb{P}^s$$

over K with fibers open polydiscs in $\mathbf{A}^{d-s}$.

Proof. By Lemma 5 it suffices to consider the case

$$X := \mathbb{P}^d_{/K} \backslash \{z : |z_i| \leqq |\pi|^{\beta_i} \text{ for } 0 \leqq i \leqq s\} .$$

The projection

$$\begin{aligned} pr : \quad X \quad &\to \quad \mathbb{P}^s \\ [z_0 : \ldots : z_d] &\mapsto [z_0 : \ldots : z_s] \text{ (not necessarily unimodular)} \end{aligned}$$

is well-defined because at least for one i with $0 \leqq i \leqq s$ we have $|z_i| > |\pi|^{\beta_i}$. The subsets

$$U_j := \left\{ [z_0 : \ldots : z_s] \in \mathbb{P}^s(\hat{\bar{K}}) : \frac{|z_j|}{|\pi|^{\beta_j}} \geqq \frac{|z_i|}{|\pi|^{\beta_i}} \text{ for } 0 \leqq i \leqq s \right\}$$

for $0 \leqq j \leqq s$ form an admissible K-affinoid covering of $\mathbb{P}^s_{/K}$. We now consider, for a fixed j, the K-morphism

$$pr^{-1}(U_j) \xrightarrow{(pr, \phi)} U_j \times \mathbf{A}^{d-s}$$

where ϕ is defined by

$$\phi([z_0 : \ldots : z_d]) := \left(\frac{z_{s+1}}{z_j}, \ldots, \frac{z_d}{z_j} \right) .$$

The above argument for pr being well-defined also shows that on $pr^{-1}(U_j)$ we have $\dfrac{|z_j|}{|\pi|^{\beta_j}} > 1$. Therefore the image of ϕ is contained in the open polydisc

$$D_j := \{(w_{s+1}, \ldots, w_d) \in \mathbf{A}^{d-s}(\hat{\bar{K}}) : |w_t| < |\pi|^{-\beta_j} \text{ for } s < t \leqq d\} .$$

We claim that

$$(pr, \phi) : pr^{-1}(U_j) \to U_j \times D_j$$

is a K-analytic isomorphism. But an inverse morphism is given by

$$([z_0 : \ldots : z_s], (w_{s+1}, \ldots, w_d)) \mapsto [z_0 : \ldots : z_s : z_j w_{s+1} : \ldots : z_j w_d]$$

where we may assume that the projective coordinates on the right hand side (but not necessarily the ones on the left hand side) are unimodular. If $|z_i| = 1$ for some $0 \leqq i \leqq s$ we trivially have $|z_i| > |\pi|^{\beta_i}$ and if $|z_j w_t| = 1$ for some $s < t \leqq d$ we get $|z_j| = |w_t|^{-1} > |\pi|^{\beta_j}$. In any case we see that the right hand side lies in X. q.e.d.

We like to consider such fibrations as in Proposition 6 as "homotopy equivalences". At least they should induce isomorphisms in any reasonable cohomology theory.

2 Abstract cohomology theory

Since our later computations will be valid in at least two different interesting cohomology theories we proceed in an axiomatic way. Let $\mathscr{V}$ be the category of smooth separated (rigid) analytic varieties over K equipped with a fixed Grothendieck topology which we assume to be finer than the analytic topology. Let $\mathscr{F}$ be an object in the derived category $D^{\geqq 0}(\mathscr{V})$ of complexes of sheaves on $\mathscr{V}$ in nonnegative degrees. For any variety X in $\mathscr{V}$ we put

$$H^*(X) := H^*(X, \mathscr{F}) ;$$

if $U \subseteq X$ is an (admissible) open subvariety we also will use the relative cohomology

$$H^*(X, U) := H^*(X, U; \mathscr{F}) .$$

We recall that relative cohomology of the pair (X, U) is the derived functor of the functor "sections on X which vanish on U".

Remark. In algebraic geometry relative cohomology usually is denoted by $H^*_Z(X, \bullet)$ with $Z := X \backslash U$. Since in our context Z rarely is a closed subvariety of X the above notation seems to be more appropriate.

These groups constitute our "abstract" cohomology theory. In order to make this theory interesting we require the following properties to be fulfilled:

(I) (Homotopy invariance)
If D denotes the 1-dimensional open unit disc then, for any affinoid variety X in $\mathscr{V}$, the projection $X \times D \to X$ induces an isomorphism

$$H^*(X) \xrightarrow{\cong} H^*(X \times D)$$

in cohomology.

(II) (Product structure)
There are homomorphisms

$$\cup : \mathscr{F} \overset{L}{\underset{\mathbb{Z}}{\otimes}} \mathscr{F} \to \mathscr{F} \qquad \text{and} \qquad e : \mathbb{Z} \to \mathscr{F}$$

in $D^+(\mathscr{V})$ such that $\cup$ is associative and (graded) commutative with unit e.

(III) (Cohomology of the point)
We have $H^s(\mathrm{Sp}(K)) = 0$ for $s \geqq 1$. Furthermore, the ring

$$A := H^0(\mathrm{Sp}(K))$$

is artinian.

(IV) (Cohomology of projective space)
We have $H^s(\mathbb{P}^n_{/K}) = 0$ for odd s or $s > 2n$. Furthermore, there is a homomorphism

$$c\colon \mathbf{G}_m[-1] \to \mathscr{F}$$

in $D^{\geqq 0}(\mathscr{V})$ such that, for $0 \leqq s \leqq n$, the map

$$\pi^*(\ ) \cup \xi^s : A = H^0(\mathrm{Sp}(K)) \xrightarrow{\cong} H^{2s}(\mathbb{P}^n_{/K})$$

is an isomorphism where $\pi\colon \mathbb{P}^n_{/K} \to \mathrm{Sp}(K)$ is the structure morphism and ξ is the image of the canonical line bundle, i.e.,

$$\begin{array}{ccc} H^1(\mathbb{P}^n_{/K}, \mathscr{O}^\times) & \xrightarrow{c} & H^2(\mathbb{P}^n_{/K}) \\ \mathscr{O}(1) & \mapsto & \xi \end{array}.$$

Our abstract cohomology theory consequently takes values in the category of A-modules. For future reference let us state two further simple consequences of these properties.

Lemma 1 *Any locally trivial fibration $Y \to X$ in $\mathscr{V}$ with fibers open polydiscs in affine space induces an isomorphism $H^*(X) \xrightarrow{\cong} H^*(Y)$ in cohomology.*

Proof. This is a well-known formal consequence of the property I above.

Lemma 2 *Let $m \leqq n$ be natural numbers and let α be a K-linear automorphism of $\mathbb{P}^n$. For the morphism $f\colon \mathbb{P}^m \to \mathbb{P}^n$ given by $[z_0 : \ldots : z_m] \mapsto \alpha([z_0 : \ldots : z_m : 0 : \ldots : 0])$ and any $0 \leqq s \leqq m$ we have the commutative diagram*

$$\begin{array}{ccc} A & \xrightarrow{\cong} & H^{2s}(\mathbb{P}^n_{/K}) \\ \| & & \downarrow f^* \\ A & \xrightarrow{\cong} & H^{2s}(\mathbb{P}^m_{/K}) \end{array}$$

where the horizontal arrows represent the isomorphisms in property IV above.

Proof. f^* respects the canonical line bundles.

There are two basic examples for such a cohomology theory. The first one is the de Rham cohomology. We assume that our base field K has characteristic 0 and we equip $\mathscr{V}$ with the analytic topology. Let $\mathscr{F}$ be the complex

$$\Omega^\bullet : \mathscr{O} \xrightarrow{d} \Omega^1 \xrightarrow{d} \Omega^2 \xrightarrow{d} \ldots$$

of sheaves of holomorphic differential forms. The de Rham cohomology of a

variety X in $\mathscr{V}$ is defined to be the hypercohomology

$$H^*_{DR}(X) := H^*(X, \Omega^\bullet) .$$

It has the four properties which we have required above:

The product structure is given by the usual exterior multiplication of differential forms. The cohomology of the point is obvious. The cohomology of projective space can be computed as in [23] (last paragraph of §7.1) using in addition the GAGA-principle [28] that analytic and algebraic coherent sheaf cohomology of $\mathbb{P}^n$ are equal.

For the homotopy invariance we imitate the argument in [23] (Prop. 7.1). If X is affinoid then X and $X \times D$ have trivial coherent sheaf cohomology so that in each case the de Rham cohomology is computed by the complex of global sections of $\Omega^\bullet$. We therefore have to show that

$$\Omega^\bullet(X) \to \Omega^\bullet(X \times D)$$

is a quasi-isomorphism. The injectivity on cohomology groups as well as on the level of complexes is obvious since the projection $X \times D \to X$ has a section. It remains to establish the surjectivity on cohomology groups. Each $\omega \in \Omega^*(X \times D)$ is of the form

$$\omega = \sum_{i \geqq 0} \omega_i T^i + \sum_{i \geqq 0} \eta_i T^i dT \qquad \text{with } \omega_i \in \Omega^*(X), \quad \eta_i \in \Omega^{*-1}(X)$$

$$\text{such that } \limsup \sqrt[i]{\|\omega_i\|} \leqq 1, \limsup \sqrt[i]{\|\eta_i\|} \leqq 1 .$$

Observing that

$$\limsup \sqrt[i]{\frac{\|\eta_{i-1}\|}{i}} \leqq 1 \quad \text{and}$$

$$d\left(\sum_{i \geqq 1} \frac{\eta_{i-1}}{i} T^i\right) = \sum_{i \geqq 1} \frac{d\eta_{i-1}}{i} T^i + \sum_{i \geqq 0} \eta_i T^i dT$$

we see that modulo exact forms any ω has the form

$$\omega = \sum_{i \geqq 0} \omega_i T^i \qquad \text{with } \omega_i \in \Omega^*(X), \quad \limsup \sqrt[i]{\|\omega_i\|} \leqq 1 .$$

We then have

$$d\omega = \sum_{i \geqq 0} d\omega_i \cdot T^i + \sum_{i \geqq 1} \omega_i \cdot i \cdot T^{i-1} dT .$$

If ω is closed it follows that $\omega = \omega_0 \in \Omega^*(X)$.

Proposition 3 *We have*

$$H^*_{DR}(\Omega^{(d+1)}) = \{\omega \in \Omega^*(\Omega^{(d+1)}) : d\omega = 0\}/d\Omega^{*-1}(\Omega^{(d+1)}) .$$

Proof. $\Omega^{(d+1)}$ is a Stein-space according to §1 Proposition 4. But Theorem B ([29] Satz 2.4) says that Stein-spaces have trivial coherent sheaf cohomology.

The second example is the etale cohomology over $\hat{\bar{K}}$ with coefficients in a finite ring whose order is prime to the characteristic of the residue class field of K.

Definition. A morphism $f: X \to Y$ of K-analytic varieties is called *etale* if, for any $x \in X$, the induced homomorphism of local rings $\mathcal{O}_{Y, f(x)} \to \mathcal{O}_{X, x}$ is flat and unramified.

Over an algebraically closed base field this definition is equivalent to the one given in [18] V.3. Due to the fact that the image of an etale morphism need not to be an admissible open subvariety the notion of an etale covering of analytic varieties is not as straightforward as the corresponding notion in algebraic geometry.

Remark. If $f: X \to Y$ is a flat morphism between K-affinoid varieties then $f(X)$ is a finite union of K-affinoid open subvarieties of Y and, in particular, is an admissible open subvariety of Y.

Proof. [32] 3.4.8.

A family of etale morphisms $f_i: X_i \to Y$ between K-affinoid varieties is called an etale covering if the $f_i(X_i)$ form an admissible covering of Y. Because of the above Remark it is clear that in this way the category of K-affinoid varieties is equipped with a Grothendieck topology. If Y is an arbitrary K-analytic variety, $Y_j \xrightarrow{\subseteq} Y$ is an admissible covering by K-affinoid open subvarieties, and $X_{ij} \to Y_j$ for each j is an etale covering of K-affinoid varieties then we call the family $X_{ij} \to Y$ a special etale covering. An arbitrary family of etale morphisms $f_i: X_i \to Y$ in the category of K-analytic varieties is called an etale covering if it can be refined to a special etale covering. This defines a Grothendieck topology on the category of K-analytic varieties (compare SGA 3 exp. IV §6.2) which restricts to the previously defined topology on the subcategory of K-affinoid varieties. We call it as well as its restriction to $\mathscr{V}$ the etale topology and denote by $\mathscr{V}_{et}$ the corresponding site.

Let $\bar{\mathscr{V}}$ be the category of smooth separated analytic varieties over $\bar{K}$, and let A be a fixed finite ring of order m prime to the characteristic of K viewed as a constant sheaf on $\bar{\mathscr{V}}_{et}$. The extension of ground field functor $X \mapsto \bar{X}$ ([6] 9.3.6) induces a morphism of sites $e: \bar{\mathscr{V}}_{et} \to \mathscr{V}_{et}$. The complex $\mathscr{F}$ on $\mathscr{V}_{et}$ then is defined to be the total direct image

$$\mathscr{F} := \mathrm{Re}_* A$$

under e of the sheaf A; the corresponding cohomology of a variety X in $\mathscr{V}$ is

$$H^*(X) = H^*_{et}(X, \mathscr{F}) = H^*_{et}(\bar{X}, A) .$$

Among our four properties the cohomology of the point is obvious and the product structure is induced by the ring multiplication on A. The other two properties are consequences of general results of O. Gabber (unpublished). In the formula for the cohomology of projective space one has

$$H^{2s}_{et}(\mathbb{P}^n_{/\bar{K}}, A) = A(s) \qquad \text{for } 0 \leqq s \leqq n$$

where $A(s)$ is the s-th Tate twist of A. For simplicity we use a once and for all fixed primitive m-th root of unity in $\bar{K}$ in order to identify $A(s)$ and A. The homotopy invariance only holds under the assumption that the order of A is prime to the residue class field characteristic of K.

Apart from the axiomatically given properties and their consequences we will need two results from general cohomology theory. The first result concerns the behaviour of cohomology with respect to certain direct limits of varieties. Let X be

a variety in $\mathscr{V}$ and let $U \subseteq X$ be an open subvariety which possesses an admissible covering by an increasing family

$$\ldots \subseteq U_n \subseteq U_{n+1} \subseteq \ldots \qquad \text{for } n \in \mathbb{N}$$

of open subvarieties.

Proposition 4 *There is a natural exact sequence*

$$0 \to \varprojlim {}^{(1)} H^{*-1}(X, U_n) \to H^*(X, U) \to \varprojlim H^*(X, U_n) \to 0 .$$

Proof. (See [27] §§1–2 for the basic properties of the derived functors of $\varprojlim$.) We first show that, for an arbitrary injective sheaf I on $\mathscr{V}$, we have

$$H^0(X, U; I) = \varprojlim H^0(X, U_n; I) \quad \text{and} \quad \varprojlim {}^{(r)} H^0(X, U_n; I) = 0 \quad \text{for } r \geqq 1 .$$

The sequence of projective systems (with respect to n)

$$0 \to H^0(X, U_n; I) \to H^0(X, I) \to H^0(U_n, I) \to 0$$

is exact since the right arrows are surjective by the injectivity of I (SGA 4 V 4.7). Furthermore, the sheaf property of I implies $H^0(U, I) = \varprojlim H^0(U_n, I)$. Therefore in the projective limit we get the sequence

$$0 \to H^0(X, U; I) \to H^0(X, I) \to H^0(U, I) \to 0$$

which is exact for the same reason as before. From that follows the first part of our claim and also the second part once we observe that the projective systems $\{H^0(X, I)\}_n$ and $\{H^0(U_n, I)\}_n$ are acyclic, the first one trivially and the second one because of the surjectivity of the transition maps.

Now let $\mathscr{F} \to\!\!\!\!\to I^\bullet$ be an injective resolution. We consider the two hypercohomology spectral sequences

$$E_1^{r,s} = \varprojlim {}^{(s)} H^0(X, U_n; I^r) \Rightarrow R^{r+s} \varprojlim (H^0(X, U_n; I^\bullet)) \quad \text{and}$$

$$E_2^{r,s} = \varprojlim {}^{(r)} H^s(H^0(X, U_n; I^\bullet)) \Rightarrow R^{r+s} \varprojlim (H^0(X, U_n; I^\bullet)) .$$

The first one, by the above established facts, degenerates and gives

$$H^s(H^0(X, U; I^\bullet)) = H^s(\varprojlim H^0(X, U_n; I^\bullet)) = R^s \varprojlim (H^0(X, U_n; I^\bullet)) .$$

Because of $\varprojlim^{(r)} = 0$ for $r \geqq 2$ the second one then splits into the short exact sequences in our assertion.

Corollary 5 *If the A-modules in the projective system $\{H^{s-1}(X, U_n)\}_n$ are finitely generated then we have $H^s(X, U) = \varprojlim H^s(X, U_n)$.*

Proof. Use [27] Cor. 7.2.

For the second result let $U_1, \ldots, U_m \subseteq X$ be a finite family of open subvarieties of the variety X in $\mathscr{V}$ and put $U := U_1 \cap \ldots \cap U_m$. We then want to construct a strongly convergent spectral sequence

$$(*) \qquad E_1^{r,s} = \bigoplus_{1 \leqq i_0, \ldots, i_{-r} \leqq m} H^s(X, U_{i_0} \cup \ldots \cup U_{i_{-r}}) \Rightarrow H^{r+s}(X, U) .$$

This is based on the following observation about simplicial abelian groups. Let $G_1, \ldots, G_m$ be a finite family of subgroups of some abelian group G. We then consider the simplicial abelian group

$$\bigoplus_i G_i \;\longleftarrow\!\!\!\!\!\!\leftarrow\; \bigoplus_{i_0, i_1} G_{i_0} \cap G_{i_1} \;\longleftarrow\!\!\!\!\!\!\leftarrow\!\!\!\!\!\!\leftarrow\; \bigoplus_{i_0, i_1, i_2} G_{i_0} \cap G_{i_1} \cap G_{i_2} \;\longleftarrow\!\!\!\!\!\!\leftarrow\!\!\!\!\!\!\leftarrow\!\!\!\!\!\!\leftarrow \ldots$$

with the obvious face and degeneracy maps. Let $C(G_1, \ldots, G_m)$ denote the associated (homological) complex of abelian groups (where the differential is given by the alternating sum of the face maps).

Proposition 6 *Suppose that*

$$\left(\sum_{i \in V} G_i\right) \cap \left(\bigcap_{j \in W} G_j\right) = \sum_{i \in V} \left(G_i \cap \left(\bigcap_{j \in W} G_j\right)\right)$$

holds true for all subsets $V, W \subseteq \{1, \ldots, m\}$. *Then* $C(G_1, \ldots, G_m)$ *is an acyclic resolution of* $\sum_i G_i$.

Proof. We first establish that the subcomplex

$$C^+(G_1, \ldots, G_m): \bigoplus_i G_i \leftarrow \bigoplus_{i_0 < i_1} G_{i_0} \cap G_{i_1} \leftarrow \bigoplus_{i_0 < i_1 < i_2} G_{i_0} \cap G_{i_1} \cap G_{i_2} \leftarrow \ldots$$

is an acyclic resolution of $\sum_i G_i$. For that we consider the short exact sequence of augmented strict-simplicial abelian groups

$$\begin{array}{ccccccccc}
0 \to & \sum\limits_{i \neq 1} G_i & \to & \sum\limits_{i} G_i & \to & G_1/G_1 \cap \sum\limits_{i \neq 1} G_i & \to 0 \\
 & \uparrow & & \uparrow & & \uparrow & \\
0 \to & \bigoplus\limits_{i \neq 1} G_i & \to & \bigoplus\limits_{i} G_i & \to & G_1 & \to 0 \\
 & \uparrow\uparrow & & \uparrow\uparrow & & \circ\uparrow\uparrow & \\
0 \to & \bigoplus\limits_{1 < i_0 < i_1} G_{i_0} \cap G_{i_1} & \to & \bigoplus\limits_{i_0 < i_1} G_{i_0} \cap G_{i_1} & \to & \bigoplus\limits_{1 < i_1} G_1 \cap G_{i_1} & \to 0 \\
 & \uparrow\uparrow\uparrow & & \uparrow\uparrow\uparrow & & \circ\uparrow\uparrow\uparrow & \\
0 \to & \bigoplus\limits_{1 < i_0 < i_1 < i_2} G_{i_0} \cap G_{i_1} \cap G_{i_2} & \to & \bigoplus\limits_{i_0 < i_1 < i_2} G_{i_0} \cap G_{i_1} \cap G_{i_2} & \to & \bigoplus\limits_{1 < i_1 < i_2} G_1 \cap G_{i_1} \cap G_{i_2} & \to 0 \\
 & \uparrow\uparrow\uparrow\uparrow & & \uparrow\uparrow\uparrow\uparrow & & \circ\uparrow\uparrow\uparrow\uparrow & .
\end{array}$$

Since the left vertical arrows in the last column represent, as indicated, the zero maps this can be rewritten as a short exact sequence of augmented complexes

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 \to & \sum\limits_{i \neq 1} G_i & \to & \sum\limits_{i} G_i & \to & G_1/G_1 \cap \sum\limits_{i \neq 1} G_i & \to 0 \\
 & \uparrow & & \uparrow & & \uparrow & & & \\
0 \to & C^+(G_2, \ldots, G_m) & \to & C^+(G_1, \ldots, G_m) & \to & \left\{\begin{array}{c} G_1 \\ \uparrow \\ C^+(G_1 \cap G_2, \ldots, G_1 \cap G_m) \end{array}\right\} & \to 0.
\end{array}$$

We now argue by induction with respect to m. There is nothing to prove for $m = 1$. Obviously, with $G_1, \ldots, G_m$, also the families $G_2, \ldots, G_m$ and $G_1 \cap G_2, \ldots, G_1 \cap G_m$ fulfill the assumption of our assertion. Therefore, by induction, we can assume that $C^+(G_2, \ldots, G_m)$, resp. $C^+(G_1 \cap G_2, \ldots, G_1 \cap G_m)$, is an acyclic resolution of $\sum_{i \neq 1} G_i$, resp. $\sum_{i \neq 1} G_1 \cap G_i = G_1 \cap \sum_{i \neq 1} G_i$. The above short exact sequence then shows that also $C^+(G_1, \ldots, G_m)$ is an acyclic resolution of $\sum_i G_i$.

It remains to prove that the inclusion $C^+(G_1, \ldots, G_m) \subseteq C(G_1, \ldots, G_m)$ is a homotopy equivalence. In case $G = G_1 = \ldots = G_m = \mathbb{Z}$ we view $\mathbb{Z}N_\bullet = C(\mathbb{Z}, \ldots, \mathbb{Z})$ as the complex associated with the free abelian group on the simplicial set

$$N \;\underset{\longleftarrow}{\overset{\longleftarrow}{\longleftarrow}}\; N \times N \;\underset{\longleftarrow}{\overset{\longleftarrow}{\longleftarrow}}\; N \times N \times N \;\underset{\longleftarrow}{\overset{\longleftarrow}{\longleftarrow}}\; \ldots, \quad N := \{1, \ldots, m\}\,.$$

It is well-known (compare, for example, [6] p. 322/323) that the homomorphisms

$$\begin{aligned} \mathbb{Z}N_q &\to \mathbb{Z}N_q \\ (i_0, \ldots, i_q) &\mapsto \begin{cases} (\operatorname{sgn}\pi)\,(i_{\pi(0)}, \ldots, i_{\pi(q)}) & \text{if there is a permutation} \\ & \pi \text{ such that } i_{\pi(0)} < \ldots < i_{\pi(q)}\,, \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

form an endomorphism of the complex $\mathbb{Z}N_\bullet$ which is homotopic to the identity; furthermore the homotopy $\mathbb{Z}N_\bullet \to \mathbb{Z}N_{\bullet+1}$ can be chosen in such a way that

$$(i_0, \ldots, i_q) \mapsto \sum_{j_0, \ldots, j_q \in \{i_0, \ldots, i_q\}} c_{j_0 \ldots j_q}^{i_0 \ldots i_q} \cdot (i_0, j_0, \ldots, j_q)$$

with appropriate integers $c_{\ldots}^{\ldots} \in \mathbb{Z}$. Coming back to our general situation we define an endomorphism ϕ of $C(G_1, \ldots, G_m)$ by

$$g \in G_{i_0} \cap \ldots \cap G_{i_q} \mapsto \begin{cases} (\operatorname{sgn}\pi) \cdot g \in G_{i_{\pi(0)}} \cap \ldots \cap G_{i_{\pi(q)}} & \text{if there is a permutation} \\ & \pi \text{ such that} \\ & i_{\pi(0)} < \ldots < i_{\pi(q)}\,, \\ 0 & \text{otherwise} \end{cases}$$

and a homotopy h on $C(G_1, \ldots, G_m)$ by

$$g \in G_{i_0} \cap \ldots \cap G_{i_q} \mapsto \bigoplus_{j_0, \ldots, j_q \in \{i_0, \ldots, i_q\}} c_{j_0 \ldots j_q}^{i_0 \ldots i_q} \cdot g \in G_{i_0} \cap G_{j_0} \cap \ldots \cap G_{j_q}\,.$$

It is clear that h is a homotopy between ϕ and id; furthermore we have $\phi \mid C^+(G_1, \ldots, G_m) = id$ and $im(\phi) = C^+(G_1, \ldots, G_m)$. q.e.d.

We are going to use this general result in the following particular situation.

Lemma 7 *For any injective sheaf I on $\mathscr{V}$ we have (where all groups are considered as subgroups of $I(X)$):*

i. $\sum_i H^0(X, U_i; I) = H^0(X, U; I)$;

ii. $(\sum_{i \in V} H^0(X, U_i; I)) \cap (\bigcap_{j \in W} H^0(X, U_j; I)) = \sum_{i \in V} (H^0(X, U_i; I) \cap (\bigcap_{j \in W} H^0(X, U_j; I)))$ *for any two subsets $V, W \subseteq \{1, \ldots, m\}$.*

Proof. i. Assume first that $m = 2$ and let $s \in I(X)$ be a section with $s \mid U_1 \cap U_2 = 0$. By the sheaf property we find a section $s_1' \in I(U_1 \cup U_2)$ such that $s_1' \mid U_1 = 0$ and $s_1' \mid U_2 = s \mid U_2$, and by the injectivity of I the section s_1' extends to a section $s_1 \in I(X)$. If we put $s_2 := s - s_1$ then we obviously have $s_2 \mid U_2 = 0$. The general case follows by induction.

ii. Because of $\bigcap_{j \in W} H^0(X, U_j; I) = H^0(X, \bigcup_{j \in W} U_j; I)$ we can assume that, say, $W = \{1\}$. But, using i., we compute

$$\left(\sum_{i \in V} H^0(X, U_i; I)\right) \cap H^0(X, U_1; I) = H^0\left(X, \bigcap_{i \in V} U_i; I\right) \cap H^0(X, U_i; I)$$

$$= H^0\left(X, \left(\bigcap_{i \in V} U_i\right) \cup U_1; I\right) = H^0\left(X, \bigcap_{i \in V} (U_i \cup U_1); I\right)$$

$$= \sum_{i \in V} H^0(X, U_i \cup U_1; I) = \sum_{i \in V} (H^0(X, U_i; I) \cap H^0(X, U_1; I)) .$$

q.e.d.

Now, let $\mathscr{F} \xrightarrow{\sim} I^\bullet$ be an injective resolution. We then have the augmented double complex

$$\begin{array}{c} H^0(X, U; I^\bullet) \\ \uparrow \\ C(H^0(X, U_1; I^\bullet), \ldots, H^0(X, U_m; I^\bullet)) \end{array}$$

in which the columns are acyclic according to Proposition 6 and Lemma 7. Consequently, the homology of the total complex of this double complex is equal to the homology of the augmenting complex which is $H^*(X, U)$. On the other hand, the homology of the r-th row in the double complex, for trivial reasons, is equal to $\oplus_{i_0, \ldots, i_r} H^*(X, U_{i_0} \cup \ldots \cup U_{i_r})$. Therefore $(*)$ simply is the second spectral sequence of this double complex. It is strongly convergent since, as we have seen in the proof of Proposition 6, the above double complex is homotopy equivalent to the subcomplex $C^+(H^0(X, U_1; I^\bullet), \ldots, H^0(X, U_m; I^\bullet))$ whose r-th row is zero for $r \geqq m$.

3 The cohomology of $\Omega^{(d+1)}$—Connection to the Tits building

In this Paragraph we want to compute the abstract cohomology of $\Omega^{(d+1)}$ in terms of the cohomology of certain p-adically topologized simplicial complexes which naturally arise in the theory of the Tits building for GL_{d+1}. Because of the relative cohomology sequence

$$\ldots \to H^s(\mathbb{P}^d) \to H^s(\Omega^{(d+1)}) \to H^{s+1}(\mathbb{P}^d, \Omega^{(d+1)}) \to H^{s+1}(\mathbb{P}^d) \to \ldots$$

and our axiomatic knowledge of the cohomology of $\mathbb{P}^d$ we equivalently have to determine the relative cohomology $H^*(\mathbb{P}^d, \Omega^{(d+1)})$. In §1 we have seen that the open subvarieties

$$\Omega_n = \mathbb{P}^d \setminus \bigcup_{H \in \mathscr{H}_n} H(|\pi|^n)$$

form an increasing admissible covering of $\Omega^{(d+1)}$. Later on we will see that we can apply §2 Corollary 5 and calculate $H^*(\mathbb{P}^d, \Omega^{(d+1)})$ as the projective limit of the groups $H^*(\mathbb{P}^d, \Omega_n)$. But Ω_n can be written as an intersection

$$\Omega_n = \bigcap_{H \in \mathscr{H}_n} U(H; n)$$

of finitely many open subvarieties in $\mathbb{P}^d$ of the form

$$U(H; n) := \mathbb{P}^d \backslash H(|\pi|^n) .$$

Therefore the spectral sequence $(*)$ which we have constructed in §2 is at our disposal:

$$E_1^{-r,s} = \bigoplus_{\substack{(H_0, \ldots, H_r) \\ \in \mathscr{H}_n^{r+1}}} H^s(\mathbb{P}^d, U(H_0; n) \cup \ldots \cup U(H_r; n)) \Rightarrow H^{s-r}(\mathbb{P}^d, \Omega_n) .$$

We see that the central problem is to understand this spectral sequence. For that it is useful to introduce the notation

$$\mathrm{rk}(H_0, \ldots, H_r) := \mathrm{rank}\, \bar{M}$$

where

$$\bar{M} := \sum_{i=0}^{r} (o/\pi^n o) l_{H_i} \subseteq L_0^*/\pi^n L_0^* .$$

We observe that $\mathrm{rk}(H_0, \ldots, H_r) \geqq 1$ since the l_{H_i} are unimodular.

Lemma 1 $H^s(\mathbb{P}^d, U(H_0; n) \cup \ldots \cup U(H_r; n))$

$$= \begin{cases} A & \text{if } s \text{ is even with } \mathrm{rk}(H_0, \ldots, H_r) \leqq \dfrac{s}{2} \leqq d , \\ 0 & \text{otherwise} . \end{cases}$$

Proof. According to §1 Lemma 5 and Proposition 6 we have, after a suitable K-linear automorphism of $\mathbb{P}^d$, the diagram

$$\begin{array}{ccc} \mathbb{P}^d & \xleftarrow{\cong} & U(H_0; n) \cup \ldots \cup U(H_r; n) = \mathbb{P}^d \backslash \bigcap_{i=0}^{r} H_i(|\pi|^n) \\ & & \uparrow \\ & & \Big\downarrow \mathrm{pr} \\ & & \mathbb{P}^m \end{array}$$

with $m := \mathrm{rk}(H_0, \ldots, H_r) - 1$ where pr is the projection $[z_0 : \ldots : z_d] \mapsto [z_0 : \ldots : z_m]$ and the dotted arrow represents the section $[z_0 : \ldots : z_m] \mapsto [z_0 : \ldots : z_m : 0 : \ldots : 0]$; we furthermore know that pr is a locally trivial fibration with polydiscs as fibers. By §2 Lemma 1 pr induces an isomorphism in cohomology. It then follows from §2 Lemma 2 that the restriction map

$$H^s(\mathbb{P}^d) \to H^s(U(H_0; n) \cup \ldots \cup U(H_r; n))$$

is an isomorphism for $s < 2\mathrm{rk}(H_0, \ldots, H_r)$ and is surjective and the zero map otherwise. Using the relative cohomology sequence this is easily translated into our assertion. q.e.d.

Inserting this into our spectral sequence we get

$$E_1^{-r,s} = \begin{cases} \bigoplus_{\substack{(H_0,\ldots,H_r)\in\mathscr{H}_n^{r+1} \\ \mathrm{rk}(H_0,\ldots,H_r)\leqq s/2}} A & \text{if } s \text{ is even and } 2 \leqq s \leqq 2d\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

In order to keep track of the d_1-differential in this spectral sequence we introduce the simplicial sets $Y_\bullet^{(n,k)}$, for $1 \leqq k \leqq d$, given by

$$Y_r^{(n,k)} := \text{set of all } (H_0,\ldots,H_r)\in\mathscr{H}_n^{r+1} \text{ such that } \mathrm{rk}(H_0,\ldots,H_r) \leqq k$$

with face, resp. degeneracy, maps given by omitting, resp. doubling, one hyperplane in a tuple. If $C(Y_\bullet^{(n,k)}, A)$ denotes the chain complex on $Y_\bullet^{(n,k)}$ with coefficients in the abelian group A (viewed as a cohomological complex in negative degrees) we have

$$E_1^{-r,s} = \begin{cases} C(Y_r^{(n,s/2)}, A) & \text{if } s \text{ is even and } 2 \leqq s \leqq 2d\ , \\ 0 & \text{otherwise}\ ; \end{cases}$$

furthermore, the d_1-differential is induced from the chain differentials in the complexes $C(Y_\bullet^{(n,k)}, A)$. The corresponding E_2-spectral sequence therefore reads

$$E_2^{-r,s} = \left.\begin{cases} H_r(Y_\bullet^{(n,s/2)}, A) & \text{if } s \text{ is even and} \\ & 2 \leqq s \leqq 2d \\ 0 & \text{otherwise} \end{cases}\right\} \Rightarrow H^{s-r}(\mathbb{P}^d, \Omega_n)\ .$$

Remarks.

(1) Because of $Y_r^{(n,k)} = \mathscr{H}_n^{r+1}$ for $r < k$ we have

$$H_r(Y_\bullet^{(n,k)}, \mathbb{Z}) = 0 \qquad \text{for } 0 < r < k-1\ .$$

(2) Since the $C(Y_\bullet^{(n,k)}, A)$ are complexes of finitely generated modules over the artinian ring A the spectral sequence shows that the relative cohomology groups $H^s(\mathbb{P}^d, \Omega_n)$ are finitely generated A-modules.

Instead of trying to compute the E_2-terms further we will pass at this point to the projective limit with respect to n. It is easy to check that there is a homomorphism of spectral sequences

$$\begin{array}{ccc} H_r(Y_\bullet^{(n+1,s/2)}, A) \text{ (resp. } 0) & \Rightarrow & H^{s-r}(\mathbb{P}^d, \Omega_{n+1}) \\ \downarrow & & \downarrow \\ H_r(Y_\bullet^{(n,s/2)}, A) \text{ (resp. } 0) & \Rightarrow & H^{s-r}(\mathbb{P}^d, \Omega_n) \end{array}$$

where the left arrow is induced from the obvious map of simplicial sets $Y_\bullet^{(n+1,k)} \to Y_\bullet^{(n,k)}$ and the right arrow is induced from the inclusion $\Omega_n \subsetneq \Omega_{n+1}$. According to the above Remark we therefore have a projective system of spectral sequences of finitely generated modules over the artinian ring A so that passing to the projective limit still gives a spectral sequence ([27] Cor. 7.2). The same Remark together with §2 Corollary 5 implies that in the abutment we get

$$\varprojlim H^*(\mathbb{P}^d, \Omega_n) = H^*(\mathbb{P}^d, \Omega^{(d+1)})\ .$$

The limit spectral sequence consequently reads

$$E_2^{-r,s} = \varprojlim H_r(Y_\bullet^{(n,s/2)}, A) \text{ (resp. } 0) \Rightarrow H^{s-r}(\mathbb{P}^d, \Omega^{(d+1)}) .$$

In the next step we will express the E_2-terms above in terms of the simplicial profinite sets

$$Y_\bullet^{(k)} := \varprojlim Y_\bullet^{(n,k)} .$$

For any tuple $(H_0, \ldots, H_r) \in \mathscr{H}^{r+1}$ we have

$$\operatorname{rank}\left(\sum_{i=0}^{r} (o/\pi^n o)\ell_{H_i}\right) \leqq k \text{ for all } n \in \mathbb{N}$$

if and only if

$$\dim_K\left(\sum_{i=0}^{r} Kl_{H_i}\right) \leqq k .$$

We therefore can interprete $Y_\bullet^{(k)}$ as the simplicial profinite set given by

$$Y_r^{(k)} = \text{set of all } (H_0, \ldots, H_r) \in \mathscr{H}^{r+1} \text{ such that}$$

$$\dim_K\left(\sum_{i=0}^{r} Kl_{H_i}\right) \leqq k$$

with face, resp. degeneracy, maps given by omitting, resp. doubling, one hyperplane in a tuple; the topology on $Y_r^{(k)}$ is induced from the obvious topology on $\mathscr{H} = \mathbb{P}((K^{d+1})^*) = \mathbb{P}^d(K)$ so that $Y_r^{(k)}$ is a closed subset of $\mathscr{H}^{r+1}$.

Let $|Y_\bullet^{(k)}|$ be the topological realization of $Y_\bullet^{(k)}$. We recall that $|Y_\bullet^{(k)}|$ is defined to be the quotient of $\bigcup_{r \geq 0} Y_r^{(k)} \times \Delta_r$ with respect to the equivalence relation given by the face and degeneracy maps; here the topology on $Y_r^{(k)} \times \Delta_r$ is the product of the profinite topology on $Y_r^{(k)}$ and the usual topology on the topological standard r-simplex Δ_r (compare [45] §1).

Lemma 2 *There are natural exact sequences*

$$0 \to \operatorname{Ext}_{\mathbb{Z}}^1(H^{*+1}(|Y_\bullet^{(k)}|, \mathbb{Z}), A) \to \varprojlim H_*(Y_\bullet^{(n,k)}, A)$$

$$\to \operatorname{Hom}_{\mathbb{Z}}(H^*(|Y_\bullet^{(k)}|, \mathbb{Z}), A) \to 0 .$$

Proof. (Cohomology of a topological space always is (constant) sheaf cohomology!) By the universal coefficient theorem ([47] Th. 5.5.12) we have natural exact sequences

$$0 \to \operatorname{Ext}_{\mathbb{Z}}^1(H^{*+1}(Y_\bullet^{(n,k)}, \mathbb{Z}), A) \to H_*(Y_\bullet^{(n,k)}, A) \to \operatorname{Hom}_{\mathbb{Z}}(H^*(Y_\bullet^{(n,k)}, \mathbb{Z}), A) \to 0 .$$

Since all groups involved are finitely generated modules over the artinian ring A passing to the projective limit gives the exact sequences

$$0 \to \operatorname{Ext}^1\left(\varinjlim H^{*+1}(Y_\bullet^{(n,k)}, \mathbb{Z}), A\right) \to \varprojlim H_*(Y_\bullet^{(n,k)}, A)$$

$$\to \operatorname{Hom}_{\mathbb{Z}}\left(\varinjlim H^*(Y_\bullet^{(n,k)}, \mathbb{Z}), A\right) \to 0$$

([27] Th. 4.2 and Cor. 7.2). It remains to show that we have

$$H^*(|Y_\bullet^{(k)}|, \mathbb{Z}) = \varinjlim H^*(Y_\bullet^{(n,k)}, \mathbb{Z}) .$$

Using the standard spectral sequence

$$H^s(Y_r^{(k)}, \mathbb{Z}) \Rightarrow H^{s+r}(|Y_\bullet^{(k)}|, \mathbb{Z})$$

([45] Prop. 5.1 together with [47] Th. 6.6.5) and analogous spectral sequences for the $Y_\bullet^{(n,k)}$ this follows from the continuity property of cohomology ([47] p. 319) which implies that

$$H^*(Y_r^{(k)}, \mathbb{Z}) = \varinjlim H^*(Y_r^{(n,k)}, \mathbb{Z})$$

holds true for each $r \geqq 0$.

Remark. Let $S_\bullet$ be an arbitrary simplicial profinite set. If $C^\infty(S_r, \mathbb{Z})$ denotes the group of $\mathbb{Z}$-valued locally constant functions on S_r then we have the (cohomological) complex

$$\ldots \to C^\infty(S_r, \mathbb{Z}) \to C^\infty(S_{r+1}, \mathbb{Z}) \to \ldots$$

where the differential is given as the alternating sum of the maps induced from the face maps in $S_\bullet$. Since

$$H^s(S_r, \mathbb{Z}) = \begin{cases} C^\infty(S_r, \mathbb{Z}) & \text{if } s = 0 , \\ 0 & \text{if } s > 0 \end{cases}$$

by the continuity property of cohomology the same spectral sequence as in the above proof shows that the complex $C^\infty(S_\bullet, \mathbb{Z})$ computes the cohomology of $|S_\bullet|$, i.e., that

$$H^*(|S_\bullet|, \mathbb{Z}) = H^*(C^\infty(S_\bullet, \mathbb{Z})) .$$

We now define simplicial profinite sets $\mathscr{T}_\bullet^{(k)}$, for $1 \leqq k \leqq d$, by

$$\mathscr{T}_r^{(k)} := \text{set of all flags } W_0 \subseteq W_1 \subseteq \ldots \subseteq W_r \text{ of } K - \text{subspaces}$$

$$\text{in } (K^{d+1})^* \text{ such that } 1 \leqq \dim_K W_i \leqq k$$

with face, resp. degeneracy, maps given by omitting, resp. doubling, one subspace in a flag; furthermore, the topology on $\mathscr{T}_r^{(k)}$ is given in the following way: The group $\mathrm{GL}_{d+1}(K)$ in a natural way acts on $\mathscr{T}_r^{(k)}$ and the stabilizer of a flag $\tau \in \mathscr{T}_r^{(k)}$ is a parabolic subgroup $P_\tau(K)$. We take the finest topology on $\mathscr{T}_r^{(k)}$ whose restriction on each orbit

$$\mathrm{GL}_{d+1}(K) \cdot \tau = \mathrm{GL}_{d+1}(K)/P_\tau(K)$$

is the obvious topology on the right hand side (induced from the valuation on K). (Remember that $\mathscr{T}_\bullet^{(d)}$ considered only as a simplicial set is the *Tits building* of GL_{d+1} over K. This has to be well distinguished from the Bruhat–Tits building $\mathscr{BT}$ which we introduced earlier. In case $d = 1$, for example, $\mathscr{BT}$ is the familiar tree whereas the Tits building is just a set of points which by the way can be viewed as the set of ends of the tree. An excellent introductory text to the theory of

buildings is [8].) We will show that $|Y_\bullet^{(k)}|$ and $|\mathscr{T}_\bullet^{(k)}|$ have naturally isomorphic cohomology.

Lemma 3 *For any subspace* $W \subseteq (K^{d+1})^*$ *such that* $1 \leqq \dim_K W \leqq k$ *let* $Y_\bullet^{(k)}(W) \subseteq Y_\bullet^{(k)}$, *resp.* $\mathscr{T}_\bullet^{(k)}(W) \subseteq \mathscr{T}_\bullet^{(k)}$, *be the simplicial profinite subset defined by*

$$Y_r^{(k)}(W) := \{(H_0, \ldots, H_r) \in Y_r^{(k)} : \sum_{i=0}^{r} Kl_{H_i} \subseteq W\}, \text{ resp.}$$

$$\mathscr{T}_r^{(k)}(W) := \{(W_0 \subseteq \ldots \subseteq W_r) \in \mathscr{T}_r^{(k)} : W \subseteq W_0\} .$$

Then the augmented complexes

$$\mathbb{Z} \to C^\infty(Y_\bullet^{(k)}(W), \mathbb{Z}) \text{ and } \mathbb{Z} \to C^\infty(\mathscr{T}_\bullet^{(k)}(W), \mathbb{Z})$$

are acyclic.

Proof. This is a standard fact in the context of simplicial sets. We only have to observe in addition that the maps

$$Y_{r-1}^{(k)}(W) \to Y_r^{(k)}(W)$$

$$(H_0, \ldots, H_{r-1}) \mapsto (H, H_0, \ldots, H_{r-1})$$

where $H \in \mathscr{H}$ is some fixed hyperplane such that $l_H \in W$ and

$$\begin{aligned} \mathscr{T}_{r-1}^{(k)}(W) &\to \mathscr{T}_r^{(k)}(W) \\ (W_0 \subseteq \ldots \subseteq W_{r-1}) &\mapsto (W \subseteq W_0 \subseteq \ldots \subseteq W_{r-1}) \end{aligned}$$

are continuous so that they also induce contracting homotopies in our profinite setting.

In order to relate $Y_\bullet^{(k)}$ and $\mathscr{T}_\bullet^{(k)}$ we introduce the bisimplicial profinite set $Z_{\bullet\bullet}^{(k)}$ defined by

$$Z_{rs}^{(k)} := \{(W_0 \subseteq \ldots \subseteq W_r; H_0, \ldots, H_s) \in \mathscr{T}_r^{(k)} \times Y_s^{(k)} : \sum_{i=0}^{s} K\ell_{H_i} \subseteq W_0\}$$

together with the obvious face and degeneracy maps. It will follow from Lemma 3 and a simplified version of the base change formalism (compare [20] II.4.17) that the projection maps

$$\begin{array}{ccc} & & Y_\bullet^{(k)} \\ & & \uparrow \\ \mathscr{T}_\bullet^{(k)} & \leftarrow & Z_{\bullet\bullet}^{(k)} \end{array}$$

induce cohomology isomorphisms between the corresponding topological realizations.

Remark. Let $T \subseteq S$ be a closed subset in a profinite set S. We have:

(i) The restriction map $C^\infty(S, \mathbb{Z}) \to C^\infty(T, \mathbb{Z})$ is surjective;

(ii) if $f: S' \to S$ is a continuous map between profinite sets and if U runs through the compact open neighbourhoods of T in S then

$$C^\infty(f^{-1}(T), \mathbb{Z}) = \varinjlim C^\infty(f^{-1}(U), \mathbb{Z}) .$$

Proof. i. Since the compact open subsets in S form a base of the topology it is easy to see that for any compact open subset V in T there is a compact open subset U in S such that $U \cap T = V$. ii. From i. we see that

$$C^\infty(f^{-1}(T), \mathbb{Z}) = \varinjlim C^\infty(U', \mathbb{Z})$$

where U' runs through the compact open neighbourhoods of $f^{-1}(T)$ in S'. But any U' contains a $f^{-1}(U)$ (compare [20] p. 202).

Lemma 4 *Let $f_\bullet : S_\bullet \to T$ be a map from a simplicial profinite set $S_\bullet$ into a (constant simplicial) profinite set T. If the augmented complexes*

$$\mathbb{Z} \xrightarrow{f_0^*} C^\infty(f_\bullet^{-1}(t), \mathbb{Z}) \qquad \text{for } t \in T$$

are acyclic then the augmented complex

$$C^\infty(T, \mathbb{Z}) \xrightarrow{f_0^*} C^\infty(S_\bullet, \mathbb{Z})$$

is acyclic, too.

Proof. We consider the complex of sheaves on T

$$Z \to f_{0*}\mathbb{Z} \to \ldots \to f_{r*}\mathbb{Z} \to \ldots$$

From the above Remark together with [20] II.3.3 Cor. 1 we conclude that the sheaves $f_{r*}\mathbb{Z}$ are soft. Part ii. of the above Remark also shows that

$$\mathbb{Z} \to C^\infty(f_\bullet^{-1}(t), \mathbb{Z})$$

is the associated complex of stalks in t. Our complex therefore is an acyclic complex of soft sheaves. According to [20] II Th. 3.5.4 the associated complex of global sections

$$C^\infty(T, \mathbb{Z}) \to C^\infty(S_\bullet, \mathbb{Z})$$

then is acyclic, too.

Proposition 5 *For any $1 \leqq k \leqq d$ we have a natural isomorphism*

$$H^*(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}) \cong H^*(|Y_\bullet^{(k)}|, \mathbb{Z}) .$$

Proof. Consider the biaugmented double complex

$$\begin{array}{ccc} & & C^\infty(Y_\bullet^{(k)}, \mathbb{Z}) \\ & & \downarrow \\ C^\infty(\mathscr{T}_\bullet^{(k)}, \mathbb{Z}) & \to & C^\infty(Z_{\bullet\bullet}^{(k)}, \mathbb{Z}) . \end{array}$$

Because of Lemma 3 we can apply Lemma 4 to each row and each column of this double complex and we get that all rows and columns are acyclic. Therefore both augmentation maps are quasi-isomorphisms.

Remark. Define a third simplicial profinite set $X_\bullet^{(k)}$ by

$$X_r^{(k)} := \text{set of all flags } M_0 \subseteq M_1 \subseteq \ldots \subseteq M_r \text{ of } o\text{-submodules in } L_0^*$$
$$\text{such that } M_0 \text{ contains an unimodular vector and } \operatorname{rank} M_r \leqq k .$$

The cohomology of $|X_\bullet^{(k)}|$ also is naturally isomorphic to the cohomology of $|Y_\bullet^{(k)}|$.

Proposition 6

(*i*) $\varprojlim H_r(Y_\bullet^{(n,k)}, A) = \mathrm{Hom}_{\mathbb{Z}}(H^r(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}), A)$;

(*ii*) $H^r(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}) = 0 \qquad$ *for* $r \neq 0, k-1$;

(*iii*) $H^0(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}) = \begin{cases} C^\infty(\mathbb{P}((K^{d+1})^*), \mathbb{Z}) & \text{if } k = 1 \\ \mathbb{Z} & \text{if } k > 1 \end{cases}$.

Proof. Let $N\mathscr{T}_r^{(k)}$ denote the open and closed subset of "nondegenerate" flags $W_0 \subsetneqq \ldots \subsetneqq W_r$ in $\mathscr{T}_r^{(k)}$. By the cosimplicial version of the normalization theorem (compare [31] VIII. 6) the inclusion of complexes

$$C^\infty(N\mathscr{T}_\bullet^{(k)}, \mathbb{Z}) \subseteq C^\infty(\mathscr{T}_\bullet^{(k)}, \mathbb{Z})$$

is a homotopy equivalence. But we have $N\mathscr{T}_r^{(k)} = \phi$ for $r \geqq k$ which implies $H^r(C^\infty(\mathscr{T}_\bullet^{(k)}, \mathbb{Z})) = 0$ and consequently $H^r(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}) = 0$ for $r \geqq k$. On the other hand we already know from the Remark after Lemma 1 that $H_r(Y_\bullet^{(n,k)}, \mathbb{Z}) = 0$ for $0 < r < k-1$ and $n \in \mathbb{N}$. By the universal coefficient theorem ([47] Th. 5.5.3) we then also have $H^r(Y_\bullet^{(n,k)}, \mathbb{Z}) = 0$ for $0 < r < k-1$ and $n \in \mathbb{N}$. Using Proposition 5 and the argument in the proof of Lemma 2 we see that $H^r(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}) = H^r(|Y_\bullet^{(k)}|, \mathbb{Z}) = \varinjlim H^r(Y_\bullet^{(n,k)}, \mathbb{Z}) = 0$ for $0 < r < k-1$. This establishes our second assertion. In order to prove the first assertion we consider the exact sequence

$$0 \to \mathrm{Ext}^1_{\mathbb{Z}}(H^{r+1}(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}), A) \to \varprojlim H_r(Y_\bullet^{(n,k)}, A)$$

$$\to \mathrm{Hom}_{\mathbb{Z}}(H^r(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z}), A) \to 0$$

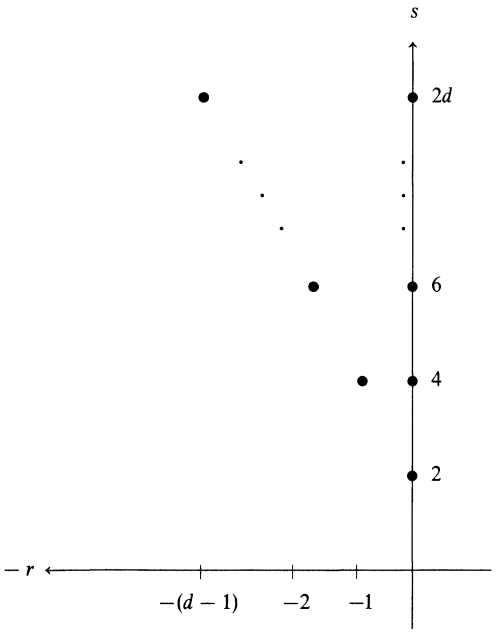
which arises from Lemma 2 and Proposition 5. By what we have just seen, for $r \neq k-2$, the first term vanishes and, for $r = k-2$, even the middle term vanishes. Finally, our third assertion is clear since $|\mathscr{T}_\bullet^{(k)}|$ obviously is connected for $k > 1$ and since, on the other hand, we have $N\mathscr{T}_0^{(1)} = \mathscr{T}_0^{(1)} = \mathbb{P}((K^{d+1})^*)$. q.e.d.

In the next Paragraph we will compute the groups $H^{k-1}(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z})$. Here we go back to our spectral sequence

$$\varprojlim H_r(Y_\bullet^{(n,s/2)}, A) \text{ (resp. } 0) \Rightarrow H^{s-r}(\mathbb{P}^d, \Omega^{(d+1)}) .$$

Because of Proposition 6 we can rewrite it in the form

$$E_2^{-r,s} = \left\{ \begin{array}{ll} \mathrm{Hom}_{\mathbb{Z}}(H^r(|\mathscr{T}_\bullet^{(s/2)}|, \mathbb{Z}), A) & \text{if } s \text{ is even,} \\ & 2 \leqq s \leqq 2d, \\ & \text{and } r = 0 \text{ or } \frac{s}{2} - 1 \\ 0 & \text{otherwise} \end{array} \right\} \Rightarrow H^{s-r}(\mathbb{P}^d, \Omega^{(d+1)}) .$$



Lemma 7 *The composed map*

$$E_2^{0,s} \to H^s(\mathbb{P}^d, \Omega^{(d+1)}) \to H^s(\mathbb{P}^d),$$

where the first arrow is the edge homomorphism in the above spectral sequence, is an isomorphism for $s > 2$ and is surjective for $s = 2$.

Proof. We can assume that s is even with $2 \leqq s \leqq 2d$ since otherwise both terms are zero. The edge homomorphism

$$E_1^{0,s} = \bigoplus_{H \in \mathscr{H}_n} H^s(\mathbb{P}^d, U(H; n)) \to H^s(\mathbb{P}^d, \Omega_n)$$

in our original spectral sequence is, of course, the natural homomorphism induced by the inclusions $\Omega_n \subseteq U(H; n)$. In the proof of Lemma 1 we have seen that the natural map

$$\begin{array}{ccc} \bigoplus_{H \in \mathscr{H}_n} H^s(\mathbb{P}^d, U(H; n)) & \longrightarrow & H^s(\mathbb{P}^d) \\ \| & & \| \\ \bigoplus_{H \in \mathscr{H}_n} A & \xrightarrow{\Sigma} & A \end{array}$$

can be identified with the sum homomorphism. If π denotes the set of connected components of $|Y_\bullet^{(n,s/2)}|$ we therefore get the following commutative diagram

$$\begin{array}{ccccccc} E_1^{0,s} & \longrightarrow & E_2^{0,s} & \longrightarrow & H^s(\mathbb{P}^d, \Omega_n) & \longrightarrow & H^s(\mathbb{P}^d) \\ \| & & \| & & & & \| \\ & & H_0(Y_\bullet^{(n,s/2)}, A) & & & & \\ & & \| & & & & \\ \bigoplus_{H \in \mathscr{H}_n} A & \xrightarrow[C]{\bigoplus \sum_{H \in C}} & \bigoplus_{C \in \pi} A & & \xrightarrow{\qquad\qquad} & & A \\ & & & & \sum & & \end{array} .$$

It remains to observe that $|Y_\bullet^{(n,s/2)}|$ is connected for $s > 2$. q.e.d.

This Lemma implies that $E_2^{0,s} = E_\infty^{0,s}$ is canonically a direct summand of $H^s(\mathbb{P}^d, \Omega^{(d+1)})$. Since, on the other hand, there can be no differentials between terms on the line $s = 2 - 2 \cdot (-r)$ we see that our spectral sequence degenerates and gives canonical isomorphisms

$$H^s(\mathbb{P}^d, \Omega^{(d+1)}) = \begin{cases} E_2^{-(s-2),2(s-1)} \oplus E_2^{0,s} & \text{if } s > 2 \text{ is even} , \\ E_2^{-(s-2),2(s-1)} & \text{if } s > 1 \text{ is odd} , \\ E_2^{0,2} & \text{if } s = 2 , \\ 0 & \text{if } s = 0, 1 . \end{cases}$$

In order to pass to the cohomology of $\Omega^{(d+1)}$ we use the relative cohomology sequence which, in the light of our axiom about the cohomology of $\mathbb{P}^d$, breaks up into exact sequences

$$0 \to H^{2t-1}(\Omega^{(d+1)}) \to H^{2t}(\mathbb{P}^d, \Omega^{(d+1)}) \to H^{2t}(\mathbb{P}^d) = A \text{ (resp. 0 for } t > d)$$
$$\to H^{2t}(\Omega^{(d+1)}) \to H^{2t+1}(\mathbb{P}^d, \Omega^{(d+1)}) \to 0 .$$

Inserting the above table into these sequences and applying Lemma 7 once more we derive

$$H^s(\Omega^{(d+1)}) = \begin{cases} E_2^{-(s-1),2s} & \text{if } s \geqq 2 , \\ \ker(E_2^{0,2} \to H^2(\mathbb{P}^d)) & \text{if } s = 1 , \\ H^0(\mathbb{P}^d) & \text{if } s = 0 . \end{cases}$$

We finally have established our first main result.

Theorem 1

$$H^s(\Omega^{(d+1)}) = \begin{cases} A & \text{if } s = 0 , \\ \ker(\mathrm{Hom}_{\mathbf{Z}}(C^\infty(\mathbb{P}((K^{d+1})^*), \mathbb{Z}), A) \to A) & \text{if } s = 1 , \\ \mathrm{Hom}_{\mathbf{Z}}(H^{s-1}(|\mathscr{T}_\bullet^{(s)}|, \mathbb{Z}), A) & \text{if } 2 \leqq s \leqq d , \\ 0 & \text{if } s > d . \end{cases}$$

Remarks. (1) As we have seen in the proof of Lemma 7 the map $\mathrm{Hom}_{\mathbf{Z}}(C^\infty(\mathbb{P}((K^{d+1})^*), \mathbb{Z}), A) \to A$ in the above statement is given by evaluation on the constant function on $\mathbb{P}((K^{d+1})^*)$ with value 1.

(2) The isomorphism in Theorem 1 is equivariant with respect to the natural actions of $\mathrm{GL}_{d+1}(K)$ on both sides. This is not entirely obvious since in its construction we have used unimodular coordinates in order to define the Ω_n. But if we fix a $g \in \mathrm{GL}_{d+1}(K)$ and choose an $a \geqq 0$ such that $\pi^a L_0 \subseteq gL_0 \subseteq \pi^{-a} L_0$ then it is easy to see that g induces a homomorphism from the spectral sequence for Ω_{n+2a} into the corresponding spectral sequence for Ω_n. Therefore $\mathrm{GL}_{d+1}(K)$ acts on the limit spectral sequence (after Prop. 6).

(3) If the cohomology theory is etale cohomology then the isomorphism in Theorem 1 is $\underline{\mathrm{Aut}}(\bar{K} \mid K)$-equivariant if the left hand side is replaced by the Tate twist $H^s_{et}(\Omega^{(d+1)}, A(s))$.

4 The cohomology of $\Omega^{(d+1)}$—Distributions and harmonic cochains

In this Paragraph we will discuss two ways of computing the cohomology groups $H^{k-1}(|\mathscr{T}^{(k)}_\bullet|, \mathbb{Z})$. As a consequence we will obtain two explicit expressions for the cohomology of $\Omega^{(d+1)}$ one in terms of distributions on flag manifolds and another one in terms of generalized harmonic cochains on the Bruhat–Tits building.

We have seen in the proof of §3 Proposition 6 that there is a canonical exact sequence

$$C^\infty(N\mathscr{T}^{(k)}_{k-2}, \mathbb{Z}) \xrightarrow{d} C^\infty(N\mathscr{T}^{(k)}_{k-1}, \mathbb{Z}) \to H^{k-1}(|\mathscr{T}^{(k)}_\bullet|, \mathbb{Z}) \to 0$$

where d is the alternating sum of the maps induced from the face maps

$$N\mathscr{T}^{(k)}_{k-1} \overset{\rightarrow}{\underset{\rightarrow}{\vdots}} N\mathscr{T}^{(k)}_{k-2} .$$

In order to reinterpret this map in purely group theoretic terms we have to recall very briefly the theory of parabolic subgroups. Set $G := \mathrm{GL}_{d+1}(K)$ and let $e_1^*, \ldots, e_{d+1}^*$ be the standard basis of $(K^{d+1})^*$. For every subset $I \subseteq \Delta := \{1, \ldots, d\}$ we have the standard flag

$$\tau_I := \left(\sum_{i=i_r+1}^{d+1} Ke_i^* \subsetneqq \ldots \subsetneqq \sum_{i=i_1+1}^{d+1} Ke_i^* \subsetneqq \sum_{i=i_0+1}^{d+1} Ke_i^* \right)$$

$$= \left(\sum_{i=1}^{i_0} Ke_i \subsetneqq \sum_{i=1}^{i_1} Ke_i \subsetneqq \ldots \subsetneqq \sum_{i=1}^{i_r} Ke_i \right)^*$$

where $\Delta \backslash I = \{i_0 < i_1 < \ldots < i_r\}$ and its stabilizer P_I in G. Then the map

$$\begin{array}{ccc} \text{subsets of } \Delta & \to & \text{parabolic subgroups of } G \\ & & \text{which contain } P_\phi \\ I & \mapsto & P_I \end{array}$$

is an inclusion preserving bijection. Furthermore, for $0 \leqq r < k \leqq d$, we have the homeomorphism

$$\begin{array}{ccc} \bigcup^{\bullet}_{\substack{|\Delta \backslash I| = r+1 \\ \Delta \backslash I \subseteq \{d+1-k, \ldots, d\}}} G/P_I & \xrightarrow{\sim} & N\mathscr{T}^{(k)}_r \\ gP_I & \mapsto & g(\tau_I) . \end{array}$$

Of particular interest for us are the homeomorphisms

$$G/P_{\{1,\ldots,d-k\}} \xrightarrow{\sim} N\mathscr{T}^{(k)}_{k-1}$$

and, if $k > 1$,

$$\bigcup_{d-k<\iota\leqq d}^{\bullet} G/P_{\{1,\ldots,d-k,\iota\}} \xrightarrow{\sim} N\mathscr{T}^{(k)}_{k-2} .$$

The face maps $N\mathscr{T}^{(k)}_{k-1} \to N\mathscr{T}^{(k)}_{k-2}$ obviously correspond under those identifications to the projections

$$G/P_{\{1,\ldots,d-k\}} \to G/P_{\{1,\ldots,d-k,\jmath\}} \subseteq \bigcup_{d-k<\iota\leqq d}^{\bullet} G/P_{\{1,\ldots,d-k,\iota\}} .$$

Lemma 1

$$\tilde{H}^{k-1}(|\mathscr{T}^{(k)}_{\bullet}|, \mathbb{Z}) = C^{\infty}(G/P_{\{1,\ldots,d-k\}}, \mathbb{Z}) \Big/ \sum_{i=d-k+1}^{d} C^{\infty}(G/P_{\{1,\ldots,d-k,\iota\}}, \mathbb{Z}) .$$

Proof. ($\tilde{H}^{\bullet}(\ )$ denotes reduced cohomology.) Obvious.

In order to further analyze the quotient group on the right hand side in the above Lemma we use the theory of the Bruhat decomposition. Let $W \subseteq G$ be the subgroup of permutation matrices which we identify with the symmetric group on the $d+1$ letters $1, \ldots, d+1$. The injection

$$\begin{array}{ccc} \Delta & \to & W \\ i & \mapsto & s_i := (i \;\; i+1) \end{array}$$

defines a set of generators for W; if we speak about the length of an element in W this is always meant with respect to this set of generators. For $I \subseteq \Delta$ we set

$$W_I := \text{subgroup of } W \text{ generated by } \{s_i : i \in I\} .$$

We then have

$$P_I = P_\phi W_I P_\phi$$

and the Bruhat decomposition

$$G/P_I = \bigcup_{w \in W/W_I}^{\bullet} C_I(w)$$

of G/P_I as a disjoint union of the subsets

$$C_I(w) := P_\phi w P_I / P_I .$$

Lemma 2 *In each coset wW_I there is a unique element $\tilde{w}$ of minimal length; it is also the unique element in wW_I such that the projection map $C_\phi(\tilde{w}) \to C_I(\tilde{w})$ is injective (and then even a homeomorphism).*

Proof. [4] Prop. (3.9) and (3.16).

Let $W^I \subseteq W$ denote the subset of those elements w which are of minimal length in their coset wW_I. The Bruhat decomposition then reads

$$G/P_I = \bigcup_{w \in W^I}^{\bullet} C_I(w) .$$

In terms of permutations we can explicitly describe W^I as being the set of permutations w such that $w(i) < w(i+1)$ for all $i \in I$ ([7] Chap. IV§1 Exerc. 3 and 4). We put

$$V^I := W^I \setminus \bigcup_{i \in \Delta \setminus I} W^{I \cup \{i\}} .$$

Using Lemma 2 we see that

V^I is the set of all $w \in W^I$ such that none of the projection maps $C_I(w) \to C_{I \cup \{i\}}(w)$ for $i \in \Delta \setminus I$ is a homeomorphism.

The topological properties of the Bruhat decomposition are described by the Bruhat order on W. We write $w \to tw$ for $w \in W$ and any transposition $t \in W$ such that $\text{length}(tw) = \text{length}(w) + 1$. The Bruhat order $\leqq$ is defined to be the transitive closure of $\to$ (compare [24] I §6).

Proposition 3 *For any $w \in W^I$ the closure of $C_I(w)$ in G/P_I is*

$$\overline{C_I(w)} = \bigcup_{\substack{v \in W^I \\ v \leqq w}} C_I(v) .$$

Proof. In case $I = \phi$ the assertion is shown in [4] Cor. 3.15. The general case can easily be derived from that using that the projection maps $G/P_\phi \to G/P_I$ are closed and that $w \leqq ww'$ holds true for all $w \in W^I$ and $w' \in W_I$.

We now fix an enumeration $W^I = \{w_1, w_2, \ldots, w_N\}$ of the elements in W^I in such a way that

$$a \leqq b \qquad \text{if } w_a \leqq w_b .$$

On the group $C^\infty(G/P_I, \mathbb{Z})$ we then have the descending filtration

$$C^\infty(G/P_I, \mathbb{Z}) = F_I^0 \supseteq F_I^1 \supseteq \ldots \supseteq F_I^N = \{0\}$$

defined by

$$F_I^a := \{f \in C^\infty(G/P_I, \mathbb{Z}) : f | C_I(w_\alpha) = 0 \text{ for all } 1 \leqq \alpha \leqq a\} .$$

From Proposition 3 we easily deduce that, for any $0 \leqq a < b \leqq N$, the restriction of functions induces an isomorphism

$$F_I^a / F_I^b = C_c^\infty\left(\bigcup_{a < \beta \leqq b} C_I(w_\beta), \mathbb{Z} \right).$$

Here $C_c^\infty(S, \mathbb{Z})$ denotes the group of $\mathbb{Z}$-valued locally constant functions with compact support on the topological space S. In the following we want to examine more closely the filtration $\bar{F}_I^\bullet$ on the quotient group

$$C^\infty(G/P_I, \mathbb{Z}) \Big/ \sum_{i \in \Delta \setminus I} C^\infty(G/P_{I \cup \{i\}}, \mathbb{Z})$$

which is induced by $F_I^\bullet$.

Proposition 4 *(i) If $w_a \in W^I \setminus V^I$ then $\bar{F}_I^a = \bar{F}_I^{a-1}$;*
(ii) if $w_\beta \in V^I$ for all $a < \beta \leqq b$ then $\bar{F}_I^a / \bar{F}_I^b = C_c^\infty(\bigcup_{a < \beta \leqq b} C_I(w_\beta), \mathbb{Z})$.

Proof. (i) Let f be any function in F_I^{a-1} and put

$$f_0 := f | C_I(w_a) \in C_c^\infty(C_I(w_a), \mathbb{Z}) .$$

Since w_a is not in V^I there exists an $i \in \Delta \backslash I$ such that the projection map induces a homeomorphism $C_I(w_a) \xrightarrow{\sim} C_{I\cup\{i\}}(w_a)$. Define $g_0 \in C_c^\infty(C_{I\cup\{i\}}(w_a), \mathbb{Z})$ to be the function which corresponds to f_0 under the induced isomorphism

$$C_c^\infty(C_{I\cup\{i\}}(w_a), \mathbb{Z}) \xrightarrow{\cong} C_c^\infty(C_I(w_a), \mathbb{Z}) .$$

According to Proposition 3 the set $C_{I\cup\{i\}}(w_a)$ is open in the closed subset $\bigcup_{1\leqq\alpha\leqq a} C_{I\cup\{i\}}(w_\alpha)$ of $G/P_{I\cup\{i\}}$. We therefore find a function $g \in C^\infty(G/P_{I\cup\{i\}}, \mathbb{Z})$ such that

$$g \mid \bigcup_{1\leqq\alpha\leqq a} C_{I\cup\{i\}}(w_\alpha) = \text{extension by zero of } g_0 .$$

By construction we have $f - g \in F_I^a$.

(*ii*) We have to show that

$$F_I^a \cap \sum_{i\in\Delta\backslash I} C^\infty(G/P_{I\cup\{i\}}, \mathbb{Z}) \subseteq F_I^b .$$

Let f be any function in F_I^a which can be written as a sum

$$f = \sum_{i\in\Delta\backslash I} f_i$$

of functions $f_i \in C^\infty(G/P_{I\cup\{i\}}, \mathbb{Z})$. In a first step we show that then f also can be written as a sum

$$f = \sum_{i\in\Delta\backslash I} g_i$$

of functions $g_i \in F_I^a \cap C^\infty(G/P_{I\cup\{i\}}, \mathbb{Z})$. By induction we may assume that the f_i already lie in F_I^{a-1}. For any $i \in \Delta\backslash I$ such that $w_a \notin W^{I\cup\{i\}}$ we find an $\alpha_i \leqq a-1$ such that

$$C_{I\cup\{i\}}(w_a) = C_{I\cup\{i\}}(w_{\alpha_i}) .$$

Therefore for those i the function $g_i := f_i$ must be contained in F_I^a. If there is at most one $i \in \Delta\backslash I$ with $w_a \in W^{I\cup\{i\}}$ the sum relation implies that the corresponding f_i also is contained in F_I^a so that we again define $g_i := f_i$. Otherwise there are $i, j \in \Delta\backslash I$, $i \neq j$, such that $w_a \in W^{I\cup\{i\}} \cap W^{I\cup\{j\}} = W^{I\cup\{i,j\}}$. The projection map then induces a homeomorphism

$$C_{I\cup\{i\}}(w_a) \xrightarrow{\sim} C_{I\cup\{i,j\}}(w_a)$$

and similarly as in the proof of part i. we find a function

$$h \in F_I^{a-1} \cap C^\infty(G/P_{I\cup\{i,j\}}, \mathbb{Z})$$

such that $f_i - h \in F_I^a$. Replacing f_i by $g_i := f_i - h$ and f_j by $f_j + h$ we get a new sum representation for f. It is clear that in this manner we can construct inductively a sum representation for f of the wanted form. In the second step it remains to observe that $F_I^a \cap C^\infty(G/P_{I\cup\{i\}}, \mathbb{Z}) \subseteq F_I^b$ for any $i \in \Delta\backslash I$. By assumption we namely have that, for any $a < \beta \leqq b$,

$$C_{I\cup\{i\}}(w_\beta) = C_{I\cup\{i\}}(w_\alpha)$$

with an appropriate $\alpha \leqq a$ depending on i and β.

Corollary 5 *$H^{k-1}(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z})$ is $\mathbb{Z}$-free.*

Proof. According to the above Proposition the group in question has a finite filtration the quotients of which are of the form $C_c^\infty(S, \mathbb{Z})$ with some locally

compact totally disconnected and metrizable space S. Such groups are $\mathbb{Z}$-free as is shown in [3] 2.2.

Corollary 6 *For* $0 \leqq s \leqq d$ *define* $l(s) := \dfrac{(2d+1-s)\cdot s}{2}$. *There is a natural surjective homomorphism*

$$H^s(\Omega^{(d+1)}) \to \mathrm{Dist}(\mathbf{A}^{l(s)}(K), A)$$

where the right hand side denotes the group of A*-valued distributions on* $\mathbf{A}^{l(s)}(K)$*; for* $s = 0$ *or* d *this even is an isomorphism.*

Proof. (We recall that a distribution is a finitely additive function on the family of compact open subsets.) Let w_I denote the unique element of maximal length in W_I. For any enumeration of W^I we then have $w_N = w_\Delta w_I$. Therefore the subspace $\bar{F}_I^{N-1}$ in $C^\infty(G/P_I, \mathbb{Z})/\sum_{i\in\Delta\setminus I} C^\infty(G/P_{I\cup\{i\}}, \mathbb{Z})$ is defined independently of the particular choice of the enumeration. Since furthermore $w_\Delta w_I \in V^I$ the above Proposition implies that $\bar{F}_I^{N-1} = C_c^\infty(C_I(w_N), \mathbb{Z})$ is a direct summand. On the other hand it is well-known (compare [4]) that if U_I denotes the unipotent radical of P_I and U_I^- its transpose then the map

$$U_I^- \to C_I(w_\Delta w_I)$$

$$u \mapsto w_\Delta u P_I / P_I$$

is a homeomorphism and that

$$U_I = \mathbf{A}^{l(I)}(K)$$

with $l(I) := \mathrm{length}(w_\Delta w_I) = \dfrac{d(d+1)}{2} - \mathrm{length}(w_I)$. Combining these facts and dualizing we obtain a natural epimorphism

$$\mathrm{Hom}_{\mathbb{Z}}\left(C^\infty(G/P_I, \mathbb{Z}) \Big/ \sum_{i\in\Delta\setminus I} C^\infty(G/P_{I\cup\{i\}}, \mathbb{Z}), A\right) \to \mathrm{Dist}(\mathbf{A}^{l(I)}(K), A)\ .$$

In case $I = \phi$ we have $V^I = \{w_\Delta\}$ so that, by the above Proposition again, this map even is an isomorphism. For $I = \{1, \ldots, d-s\}$ the left hand side is canonically isomorphic to $H^s(\Omega^{(d+1)})$ and $\mathrm{length}(w_I) = \dfrac{(d-s)(d-s+1)}{2}$.

In particular, if our cohomology theory is the de Rham cohomology we have constructed in this way surjections

$$\begin{array}{ccc} \text{closed } s\text{-forms} & \longrightarrow & K\text{-valued distributions} \\ \text{on } \Omega^{(d+1)} & & \text{on } \mathbf{A}^{l(s)}(K) \end{array}$$

for $0 \leqq s \leqq d$ which can be viewed as some kind of p-adic Mellin transforms. An explicit version of this map in case $d = s = 1$ was given in [44].

Remarks. (1) In case $k = d$ our Lemma 1 and Proposition 4 already are contained in [3] §§1–3. In fact, the topological realization of our simplicial profinite set $\mathscr{T}_\bullet^{(d)}$ appears as space Y_t in loc. cit. Our proof of Proposition 4 is a slight extension of the method proposed in 3.6 of loc. cit.

(2) The filtration on $H^{k-1}(|\mathscr{T}_\bullet^{(k)}|, \mathbb{Z})$ which corresponds to $\bar{F}^\bullet_{\{1,\ldots,d-k\}}$ depends on an enumeration of $W^{\{1,\ldots,d-k\}}$. There seems to be no distinguished choice of

such an enumeration which leads to a filtration of minimal length. Consider, for example, the case $d = 3$ and $k = 2$. The set $W^{\{1\}}$ then has 12 elements. In any enumeration of $W^{\{1\}}$ which is compatible with the Bruhat order the first 8 elements lie in $W^{\{1\}} \setminus V^{\{1\}}$. The remaining 4 elements are

$$\begin{array}{ccccc} (24) & \to & (124) & \to & (1324) \\ & & \nearrow & & \searrow \\ & & (13)(24) & & \end{array}$$

of which the three upper ones form the set $V^{\{1\}}$. We see that either $w_{10} \notin V^{\{1\}}$ or $w_{11} \notin V^{\{1\}}$. This means that our construction gives two different filtrations of minimal length 2 on $H^1(|\mathscr{T}_\bullet^{(2)}|, \mathbb{Z})$.

We now turn to our second computation in terms of generalized harmonic cochains. Let $B \subseteq \mathrm{GL}_{d+1}(o)$ denote the standard Iwahori subgroup, i.e., the subgroup of all matrices which are upper triangular modulo π, and, for simplicity, put $P := P_\phi$. Let χ be the characteristic function of the compact open subset $BP/P \subseteq G/P$. We will study the homomorphism of G-modules

$$\begin{aligned} H : C_c^\infty(G/B, \mathbb{Z}) &\to C^\infty(G/P, \mathbb{Z}) \\ \varphi &\mapsto \varphi * \chi := \sum_{g \in G/B} \varphi(g) \cdot g(\chi) \ . \end{aligned}$$

The theory of Bernstein, Borel, and Casselman of representations generated by its Iwahori fixed vectors cannot be applied since we are working with integral coefficients. Instead we observe that the properties of H obviously reflect properties of the family $\{ gBP/P : g \in G \}$ of compact open subsets in G/P so that we are going to explore this family. Define the semigroup T^{++} by

$$T^{++} := \left\{ \begin{pmatrix} t_1 & & 0 \\ & \ddots & \\ 0 & & t_{d+1} \end{pmatrix} \in G : 1 \geqq |t_1| \geqq \ldots \geqq |t_{d+1}| \right\} .$$

In particular, we put

$$t := \begin{pmatrix} 1 & & & 0 \\ & \pi & & \\ & & \ddots & \\ 0 & & & \pi^d \end{pmatrix} \quad \text{and} \quad y_j := \left.\begin{pmatrix} 1 & & & & & 0 \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \pi & & \\ & & & & \ddots & \\ 0 & & & & & \pi \end{pmatrix}\right\} j$$

for $0 \leqq j \leqq d$.

Proposition 7 *If the sets byBP and $\tilde{b}y\tilde{B}P$ with $b, \tilde{b} \in \mathrm{GL}_{d+1}(o)$ and $y \in T^{++}$ are not disjoint then they are equal and we have by $B = \tilde{b}yB$ (and, in particular, $bB = \tilde{b}\tilde{B}$).*

Proof. Obviously it suffices to treat the case that *byBP* and *yBP* are not disjoint. From the Iwahori decomposition

$$B = (B \cap U^-)(B \cap P) \qquad \text{with } U^- := w_\Delta U_\phi w_\Delta$$

we deduce that

$$byBP = by(B \cap U^-)y^{-1}P \quad \text{and}$$

$$yBP = y(B \cap U^-)y^{-1}P .$$

Because of

$$y(B \cap U^-)y^{-1} \subseteq B \cap U^-$$

we then have $bBP \cap BP \neq \phi$ so that there exist $b_0, b_1 \in B$ and $p \in P$ with $bb_0 = b_1 p$. This implies $p \in \mathrm{GL}_{d+1}(o) \cap P = B \cap P$ and hence $b \in B$. Again by the Iwahori decomposition we find a $b' \in B \cap U^-$ such that $b^{-1}b' \in B \cap P$ and we see using

$$y^{-1}(B \cap P)y \subseteq B \cap P$$

that

$$byBP = b'yBP .$$

Our assumption becomes

$$b'y(B \cap U^-)y^{-1}P \cap y(B \cap U^-)y^{-1}P \neq \phi .$$

Since b' and $y(B \cap U^-)y^{-1}$ are contained in $B \cap U^-$ already the intersection $b'y(B \cap U^-)y^{-1} \cap y(B \cap U^-)y^{-1}$ must be nonempty which means that $b' \in y(B \cap U^-)y^{-1}$, resp. that $b \in y(B \cap U^-)y^{-1}(B \cap P)$. Our assertions are easily derived from this.

Proposition 8 *Any compact open subset in G/P can be written, for any $n \geqq 0$ big enough, as a finite disjoint union of subsets of the form bt^nBP/P with $b \in \mathrm{GL}_{d+1}(o)$.*

Proof. We have

$$\begin{aligned} t^nBP &= t^n(B \cap U^-)t^{-n}P \\ &= \bigcup_x xt^{n+1}(B \cap U^-)t^{-n-1}P \\ &= \bigcup_x xt^{n+1}BP \end{aligned}$$

where x runs through the left cosets of $t^{n+1}(B \cap U^-)t^{-n-1}$ in $t^n(B \cap U^-)t^{-n}$. In the light of the above Proposition we therefore only have to show that the sets bt^nBP/P for $b \in \mathrm{GL}_{d+1}(o)$ and $n \geqq 0$ form a basis of the topology of G/P. Because of the Iwasawa decomposition $G = \mathrm{GL}_{d+1}(o) \cdot P$ it furthermore suffices to show that the sets t^nBP/P for $n \geqq 0$ form a fundamental system of neighbourhoods of the trivial coset. But we have

$$t^nBP = t^n(B \cap U^-)t^{-n}P \subseteq B^{(n)}P$$

where the compact open subgroups

$$B^{(n)} := \{b \in B : b \equiv 1 \bmod \pi^n\}$$

in G form a fundamental system of neighbourhoods of the unit element.

Corollary 9 *$C^\infty(G/P, \mathbb{Z})$ as a G-module is generated by the characteristic function χ; in particular, the homomorphism H is surjective.*

In order to determine the kernel of H we recall that $C_c^\infty(B\backslash G/B, \mathbb{Z})$ is an associative ring with unit (the Hecke ring of B) via the convolution product

$$\varphi * \psi := \sum_{g \in G/B} \varphi(g) \cdot g(\psi) .$$

It acts via convolution from the right on $C_c^\infty(G/B, \mathbb{Z})$. We denote by $\mathscr{A}$ the subring which is generated by the characteristic functions χ_y of the double cosets ByB for $y \in T^{++}$. More generally let χ_M denote the characteristic function of a subset $M \subseteq G$. The following result is well-known but we have not found an appropriate reference.

Lemma 10 *$\mathscr{A}$ is a polynomial ring over $\mathbb{Z}$ in the variables $\chi_{y_0}, \ldots, \chi_{y_d}$; we have $\chi_y * \chi_{\tilde{y}} = \chi_{y\tilde{y}}$ for $y, \tilde{y} \in T^{++}$.*

Proof. By definition we have

$$\chi_y * \chi_{\tilde{y}} = \sum_{g \in ByB/B} g(\chi_{\tilde{y}}) = \sum_{g \in ByB/B} \chi_{gB\tilde{y}B} \cdot$$

Using once more the formulae

$$B = (B \cap U^-)(B \cap P) \quad \text{and}$$

$$y^{-1}(B \cap P)y \subseteq B \cap P, \qquad y(B \cap U^-)y^{-1} \subseteq B \cap U^-$$

we get

$$ByB = (B \cap U^-)yB = \overset{\bullet}{\bigcup_x} xy\,B$$

where x runs through the left cosets of $y(B \cap U^-)y^{-1}$ in $B \cap U^-$ (the union is disjoint). We also get

$$\begin{aligned} xy\,B\tilde{y}B &= xy(B \cap U^-)(B \cap P)\tilde{y}B = xy(B \cap U^-)\tilde{y}B \\ &= x(y(B \cap U^-)y^{-1})y\tilde{y}B \subseteq By\tilde{y}B\,. \end{aligned}$$

Combining these computations we obtain

$$\chi_y * \chi_{\tilde{y}} = \sum_x \chi_{xyB\tilde{y}B} \quad \text{and}$$

$$By\tilde{y}B = \bigcup_x xy\,B\tilde{y}B\,.$$

Next we have to show that the above union is disjoint. Assume therefore that

$$xyB\tilde{y}B \cap yB\tilde{y}B \neq \phi\,.$$

We then find $b_0, b_1 \in y(B \cap U^-)y^{-1}$ such that $xb_0y\tilde{y}B = b_1y\tilde{y}B$. Since xb_0 and b_1 are contained in $B \cap U^-$ it follows that

$$xb_0(y\tilde{y}(B \cap U^-)\tilde{y}^{-1}y^{-1}) = b_0(y\tilde{y}(B \cap U^-)\tilde{y}^{-1}y^{-1})$$

and hence that $x \in y(B \cap U^-)y^{-1}$. Therefore the above union is disjoint and we have established that

$$\chi_y * \chi_{\tilde{y}} = \chi_{y\tilde{y}} \cdot$$

But then

$$\begin{aligned} \mathbb{Z}[X_0, \ldots, X_d] &\to \mathscr{A} \\ X_j &\mapsto \chi_{y_j} \end{aligned}$$

defines a surjective ring homomorphism. It is also injective since, by the Cartan decomposition of G, for any $y \in T^{++}$ there are unique integers $n_0, \ldots, n_d \geqq 0$ such that

$$ByB = By_0^{n_0} \cdot \ldots \cdot y_d^{n_d} B\,.$$

q.e.d.

In the following we always view $\mathbb{Z}$ as an $\mathscr{A}$-module through the ring homomorphism

$$\mathscr{A} \to \mathbb{Z}$$

$$\chi_{y_j} \mapsto 1 .$$

Proposition 11 *H induces a G-isomorphism $C_c^\infty(G/B, \mathbb{Z}) \otimes_{\mathscr{A}} \mathbb{Z} \xrightarrow{\cong} C^\infty(G/P, \mathbb{Z})$.*

Proof. $C_c^\infty(G/B, \mathbb{Z})$ as a G-module is generated by χ_1. In order to see that H factorizes through $C_c^\infty(G/B, \mathbb{Z}) \otimes_{\mathscr{A}} \mathbb{Z}$ it therefore suffices to show that

$$H(\chi_1 * (\chi_y - \chi_1)) = H(\chi_y - \chi_1) = 0 \qquad \text{for any } y \in T^{++} .$$

According to Proposition 7 the union in

$$ByBP = \bigcup_{g \in ByB/B}^{\bullet} gBP$$

is disjoint. We get

$$H(\chi_y - \chi_1) = \left(\sum_{g \in ByB/B} \chi_{gBP} \right) - \chi_{BP} = \chi_{ByBP} - \chi_{BP} .$$

It remains to notice that

$$ByBP = By(B \cap U^-) y^{-1} P = BP .$$

The induced map

$$(*) \qquad C_c^\infty(G/B, \mathbb{Z}) \underset{\mathscr{A}}{\otimes} \mathbb{Z} \to C^\infty(G/P, \mathbb{Z})$$

clearly is surjective. For the proof of its injectivity we need the following fact.

Lemma 12 *For any $g \in G$ there is a $y \in T^{++}$ such that $gByB \subseteq \mathrm{GL}_{d+1}(o) T^{++} B$.*

Proof. Let T be the subgroup of diagonal matrices in G and let $N(T)$ denote its normalizer. Because of the Cartan decomposition $G = \mathrm{GL}_{d+1}(o) TB$ we can assume $g \in T$. As in any generalized Tits system (compare [26]) we then find finitely many $g_\alpha \in N(T)$ such that

$$gBhB \subseteq \bigcup_\alpha Bg_\alpha hB \qquad \text{for any } h \in N(T) .$$

Since $N(T) = WT$ we can write $g_\alpha = w_\alpha y_\alpha$ with $w_\alpha \in W$ and $y_\alpha \in T$. Choosing a $y \in T^{++}$ such that $y_\alpha y \in T^{++}$ for all α we then have

$$gByB \subseteq \bigcup_\alpha Bw_\alpha y_\alpha yB \subseteq \mathrm{GL}_{d+1}(o) T^{++} B .$$

q.e.d.

Because of

$$\chi_{gB} + \chi_{gB} * (\chi_y - \chi_1) = \chi_{gB} * \chi_y = \chi_{gByB}$$

the above Lemma says that the composed map

$$C_c^\infty(\mathrm{GL}_{d+1}(o) T^{++} B/B, \mathbb{Z} \xrightarrow{\subseteq} C_c^\infty(G/B, \mathbb{Z}) \longrightarrow C_c^\infty(G/B, \mathbb{Z}) \underset{\mathscr{A}}{\otimes} \mathbb{Z}$$

is surjective. Given $y \in T^{++}$ and $n \geqq 0$ big enough we find a $\tilde{y} \in T^{++}$ such that $y\tilde{y} = t^n$. In the proof of Lemma 10 we have seen that $yB\tilde{y}B \subseteq Bt^nB$ so that

$$\chi_{yB} + \chi_{yB} * (\chi_{\tilde{y}} - \chi_1) = \chi_{yB\tilde{y}B} \in C_c^\infty(Bt^nB/B, \mathbb{Z}) .$$

Therefore even the composed map

$$\bigcup_{n \geqq 0} C_c^\infty(\mathrm{GL}_{d+1}(o)t^nB/B, \mathbb{Z}) \xrightarrow{\subseteq} C_c^\infty(G/B, \mathbb{Z}) \longrightarrow C_c^\infty(G/B, \mathbb{Z}) \otimes_{\mathscr{A}} \mathbb{Z}$$

is surjective. But as an immediate consequence of Proposition 7 we have the following result which then implies the injectivity of (*).

Lemma 13 *For any $y \in T^{++}$ the restriction of H to the subspace $C_c^\infty(\mathrm{GL}_{d+1}(o)yB/B, \mathbb{Z})$ is injective.*

In the next step we want to identify the preimages under the map H of the G-submodules $C^\infty(G/P_I, \mathbb{Z})$ in $C^\infty(G/P, \mathbb{Z})$. In $C_c^\infty(G/B, \mathbb{Z})$ we have the G-submodules $C_c^\infty(G/B_I, \mathbb{Z})$ where the subgroups B_I in G are given by the inclusion preserving bijection

$$\begin{aligned} \text{subsets of } \Delta &\to \text{subgroups of } \mathrm{GL}_{d+1}(o) \\ &\qquad \text{which contain } B \\ I &\mapsto B_I := BW_IB \end{aligned}$$

Any parahoric subgroup of G is conjugate to some (possibly several) B_I.

Lemma 14 *i. $B_I = (B \cap U_I^-)(B_I \cap P_I)$ with $U_I^- :=$ transpose of U_I;*
ii. $B_IP_I = BP_I = B_IP$.

Proof. i. In case $I = \phi$ this is the Iwahori decomposition of B which also holds in the form

$$B = (B \cap U_I^-)(B \cap P_I) \qquad \text{for any } I \subseteq \Delta .$$

In the general case we first notice that, for any $w \in W$, the Iwahori decomposition together with the fact that

$$w(B \cap U^-) \subseteq Bw$$

implies that

$$BwB = Bw(B \cap P) .$$

We then compute

$$\begin{aligned} B_I &= \bigcup_{w \in W_I} BwB = \bigcup_{w \in W_I} (B \cap U_I^-)(B \cap P_I)w(B \cap P) \\ &= (B \cap U_I^-)(B_I \cap P_I) . \end{aligned}$$

ii. The first equality immediately follows from the assertion i. For the second equality we use the fact that

$$wPs_i \subseteq BwP \cup Bws_iP$$

holds true for any $w \in W$ and $i \in \Delta$. This can be shown by adapting the proof in [7] Chap. IV §2.2. We then see that

$$P_I = PW_IP \subseteq BW_IP \subseteq B_IP$$

and therefore that $B_IP_I = B_IP$.

Proposition 15 *H induces a surjective G-homomorphism $C_c^\infty(G/B_I, \mathbb{Z}) \to C^\infty(G/P_I, \mathbb{Z})$; its kernel as a G-module is generated by the functions $\chi_{By_jB_I} - \chi_{B_I}$ for $0 \leqq j \leqq d$.*

Proof. By Lemma 14 ii and Proposition 7 we have

$$B_I P_I = B_I P = \dot{\bigcup_{g \in B_I/B}} gBP$$

where the union is disjoint. We therefore obtain

$$H(\chi_{B_I}) = \sum_{g \in B_I/B} \chi_{gBP} = \chi_{B_I P_I}$$

which implies $H(C_c^\infty(G/B_I, \mathbb{Z})) \subseteq C^\infty(G/P_I, \mathbb{Z})$ since $C_c^\infty(G/B_I, \mathbb{Z})$ as a G-module is generated by χ_{B_I}. In order to establish the surjectivity we introduce the semigroup

$$T_I^{++} := \left\{ \begin{pmatrix} t_1 & & 0 \\ & \ddots & \\ 0 & & t_{d+1} \end{pmatrix} \in T^{++} : |t_i| = |t_{i+1}| \text{ for } i \in I \right\}$$

and $t_I := \prod_{j \in \Delta \setminus I} y_j \in T_I^{++}$.

Exactly the same arguments as in the proof of the Propositions 7 and 8 based now on Lemma 14 i and on the obvious inclusions

$$y(B \cap U_I^-) y^{-1} \subseteq B \cap U_I^- \quad \text{and} \quad y^{-1}(B_I \cap P_I) y \subseteq B_I \cap P_I \qquad \text{for } y \in T_I^{++}$$

then lead to the following generalization of those Propositions.

Proposition 8' *i. If the sets by $B_I P_I$ and $\tilde{b}y B_I P_I$ with $b, \tilde{b} \in \mathrm{GL}_{d+1}(o)$ and $y \in T_I^{++}$ are not disjoint then they are equal and we have $byB_I = \tilde{b}yB_I$ and $bB_I = \tilde{b}B_I$;*

ii. any compact open subset in G/P_I can be written, for any $n \geqq 0$ big enough, as a finite disjoint union of subsets of the form $bt_I^n B_I P_I / P_I$ with $b \in \mathrm{GL}_{d+1}(o)$.

Corollary 9' *$C^\infty(G/P_I, \mathbb{Z})$ as a G-module is generated by $H(\chi_{B_I}) = \chi_{B_I P_I}$.*

Since $C_c^\infty(G/B_I, \mathbb{Z})$ is not a right $\mathscr{A}$-submodule of $C_c^\infty(G/B, \mathbb{Z})$ the computation of the kernel becomes slightly more complicated. Let $M \subseteq C_c^\infty(G/B_I, \mathbb{Z})$ denote the G-submodule generated by $\chi_{By_jB_I} - \chi_{B_I}$ for $0 \leqq j \leqq d$. Using Lemma 14 i we see that with x running through the left cosets of $y_j(B \cap U_I^-) y_j^{-1}$ in $B \cap U_I^-$ we have

$$H(\chi_{B_I}) = \chi_{B_I P_I} = \chi_{(B \cap U_I^-)P_I} = \sum_x \chi_{xy_j(B \cap U_I^-)y_j^{-1} P_I}$$

$$= \sum_x \chi_{xy_j B_I P_I} = H\left(\sum_x \chi_{xy_j B_I} \right).$$

Lemma 16 *$y_j^{-1}(B \cap P_I) y_j \subseteq B_I \cap P_I$ for $0 \leqq j \leqq d$.*

Proof. Exercise.

From Lemma 14 i and Lemma 16 we now deduce that

$$B \cap y_j B_{I y_j^{-1}} = (B \cap U_I^-)(B \cap P_I) \cap [y_j(B \cap U_I^-) y_j^{-1}][y_j(B_I \cap P_I) y_j^{-1}]$$
$$= y_j(B \cap U_I^-) y_j^{-1}(B \cap P_I)$$

and consequently that

$$B/B \cap y_j B_I y_j^{-1} = (B \cap U_I^-)/y_j(B \cap U_I^-) y_j^{-1} .$$

Since By_jB_I is the disjoint union of the cosets $x'y_jB_I$ where x' runs through the left cosets of $B \cap y_jB_I y_j^{-1}$ in B we obtain

$$H(\chi_{B_I}) = H\left(\sum_x \chi_{xy_jB_I}\right) = H\left(\sum_{x'} \chi_{x'y_jB_I}\right) = H(\chi_{By_jB_I}) .$$

Therefore M is contained in the kernel of H. On the other hand we get

$$\begin{aligned} \chi_{yB_I} &= \chi_{yBy_jB_I} - y(\chi_{By_jB_I} - \chi_{B_I}) \\ &\in C_c^\infty(Byy_jB_I/B_I, \mathbb{Z}) + M \qquad \text{for any } y \in T^{++} \end{aligned}$$

since $yBy_jB_I \subseteq Byy_jB_I$. It follows inductively that

$$\chi_{B_I} \in C_c^\infty(ByB_I/B_I, \mathbb{Z}) + M \qquad \text{for any } y \in T^{++}$$

which by Lemma 12 implies that the projection map

$$C_c^\infty(\mathrm{GL}_{d+1}(o) T^{++} B_I/B_I, \mathbb{Z}) \to C_c^\infty(G/B_I, \mathbb{Z})/M$$

is surjective. Let us fix a $y \in T^{++}$ for the moment. For any $z = wy_jw^{-1}$ with $w \in W_I$ we have

$$\begin{aligned} \chi_{(B \cap U_I^-)zB_I} - \chi_{B_I} &= \chi_{w(B \cap U_I^-)y_jB_I} - \chi_{B_I} \\ &= w(\chi_{By_jB_I} - \chi_{B_I}) \in M ; \end{aligned}$$

it suffices to observe that $w^{-1}(B \cap U_I^-)w = B \cap U_I^-$ and that $By_jB_I = (B \cap U_I^-) y_jB_I$ by Lemma 16. This shows that

$$\begin{aligned} \chi_{yB_I} &= \chi_{y(B \cap U_I^-)zB_I} - y(\chi_{(B \cap U_I^-)zB_I} - \chi_{B_I}) \\ &\in C_c^\infty((B \cap U_I^-) yzB_I/B_I, \mathbb{Z}) + M . \end{aligned}$$

But for any $n \geqq 0$ big enough we find a sequence $z_1, \ldots, z_a$ of elements in the set $\{wy_jw^{-1} : 0 \leqq j \leqq d, w \in W_I\}$ such that

$$yz_1, yz_1z_2, \ldots, yz_1 \cdot \ldots \cdot z_a \in T^{++} \qquad \text{and} \qquad yz_1 \cdot \ldots \cdot z_a = t_I^n .$$

We therefore obtain inductively that

$$\chi_{yB_I} \in C_c^\infty((B \cap U_I^-) t_I^n B_I/B_I, \mathbb{Z}) + M .$$

This means that even the projection map

$$\bigcup_{n \geqq 0} C_c^\infty(\mathrm{GL}_{d+1}(o) t_I^n B_I/B_I, \mathbb{Z}) \to C_c^\infty(G/B_I, \mathbb{Z})/M$$

is surjective. But by Proposition 8'i the restriction of H to $C_c^\infty(\mathrm{GL}_{d+1}(o) t_I^n B_I/B_I, \mathbb{Z})$ is injective. Therefore the kernel of H restricted to $C_c^\infty(G/B_I, \mathbb{Z})$ is contained in M. This finishes the proof of Proposition 15.

Corollary 17 *For $0 \leqq s \leqq d$ there is a $\mathrm{GL}_{d+1}(K)$-equivariant isomorphism*

$$H^s(\Omega^{(d+1)}) = \mathrm{Hom}_{\mathbb{Z}}(C_c^\infty(G/B_I, \mathbb{Z})/R_I, A)$$

where $I = \{1, \ldots, d-s\}$ *and where* R_I *is the* $\mathrm{GL}_{d+1}(K)$*-submodule generated by*

$$\chi_{By_jB_I} - \chi_{B_I} \quad \text{for } 0 \leqq j \leqq d \qquad \text{and} \qquad \chi_{B_Is_iB_I} + \chi_{B_I} \quad \text{for } d-s < i \leqq d\,.$$

Proof. The only necessary additional observation is that for subsets of Δ of the form $I = \{1, \ldots, d-s\}$ we have $W_{I\cup\{i\}} = W_I \cup W_I s_i W_I$. Quite generally $B_{I\cup\{i\}}$ is the disjoint union of the double cosets $B_I w B_I$ with w running through $W_I \backslash W_{I\cup\{i\}} / W_I$. It then follows that for $I = \{1, \ldots, d-s\}$ and $d-s < i \leqq d$ the relation $\chi_{B_{I\cup\{i\}}} = \chi_{B_Is_iB_I} + \chi_{B_I}$ holds true.

Remarks. (1) In case $I = \phi$ Proposition 15 gives a natural isomorphism

$$C_c^\infty(G/B, \mathbb{Z})/R_\phi = \tilde{H}^{d-1}(|\mathscr{T}_\bullet^{(d)}|, \mathbb{Z})\,.$$

A straightforward computation shows that the left hand side is nothing else than the highest cohomology with compact support $H_c^d(|\mathscr{BT}|, \mathbb{Z})$ of the Bruhat–Tits building $\mathscr{BT}$. In this way we obtain a new proof of (part of) a result of Borel/Serre ([3] Th. 5.6) which avoids the use of their compactification theory for $\mathscr{BT}$.

(2) It is clear that the space on the right hand side of the statement of Corollary 17 can be viewed as a space of certain cochains on the Bruhat–Tits building $\mathscr{BT}$. In case $I = \phi$ it is precisely the space of harmonic cochains in the sense of [19].

(3) In the case $s = d = 1$ with etale cohomology as the underlying cohomology theory Corollary 17 was proved by Drinfeld ([16]) in an entirely different manner. If K has characteristic 0 his proof even works for arbitrary coefficients.

5 The cohomology of quotient varieties

Let $\Gamma \subseteq PGL_{d+1}(K)$ be a cocompact discrete subgroup which we always assume to act without fixed points on $\Omega^{(d+1)}$. We want to apply our results to the study of the cohomology of the quotient

$$X_\Gamma := \Gamma \backslash \Omega^{(d+1)}\,.$$

Theorem 2 *X_Γ in a natural way is a proper and smooth (rigid) analytic variety over K. The projection map $pr\colon \Omega^{(d+1)} \to X_\Gamma$ is an etale covering.*

The proof which we will sketch in the following is extracted from [16]. For torsionfree Γ a corresponding result in the context of formal schemes is shown in [39]. X_Γ becomes a G-ringed space (in the terminology of [6]) over K in the following way: The topology is the quotient topology, i.e., a subset $U \subseteq X_\Gamma$ is called admissible open if $pr^{-1}(U)$ is admissible open in $\Omega^{(d+1)}$ and similarly for admissible coverings; the structure sheaf on X_Γ is given by

$$\mathcal{O}_{X_\Gamma}(U) := \mathcal{O}_{\Omega^{(d+1)}}(pr^{-1}(U))^\Gamma\,.$$

Obviously, $pr\colon \Omega^{(d+1)} \to X_\Gamma$ then is a morphism of G-ringed spaces over K. In order to analyze the situation more closely we use the open affinoid subvarieties $U_\sigma^a \subseteq \Omega^{(d+1)}$, for any simplex σ in $\mathscr{BT}$ and any rational number $0 < a < 1$, which were constructed in [16] Prop. 6.1 and which have the following properties:

(1) The U_σ^a, for fixed a and varying σ, form an admissible covering of $\Omega^{(d+1)}$ whose nerve is the barycentric subdivision of $\mathscr{BT}$;

(2) $g(U_\sigma^a) = U_{g\sigma}^a$ for any $g \in PGL_{d+1}(K)$;

(3) $U_\sigma^a \Subset_K U_\sigma^b$ for $0 < a < b < 1$ (for the notation see [6] 9.6.2).

Drinfeld first constructs subsets $V_\sigma^a \subseteq |\mathscr{BT}|$ with corresponding properties and then defines $U_\sigma^a := \rho^{-1}(V_\sigma^a)$ where $\rho\colon \Omega^{(d+1)} \to |\mathscr{BT}|$ is the map which we described in the first Paragraph.

Let us fix σ and a for a moment. The subgroup $\Gamma_\sigma := \{g \in \Gamma : g\sigma = \sigma\}$ is finite; furthermore σ and $g\sigma$ for $g \in \Gamma \backslash \Gamma_\sigma$ never are neighbouring vertices in the barycentric subdivision of $\mathscr{BT}$. By 1) and 2) this implies that the open affinoid subsets $U_{g\sigma}^a = g(U_\sigma^a)$ for $g \in \Gamma/\Gamma_\sigma$ are pairwise disjoint. Put

$$U_{\Gamma_\sigma}^a := \bigcup_{g \in \Gamma/\Gamma_\sigma}^{\bullet} U_{g\sigma}^a = \bigcup_{g \in \Gamma/\Gamma_\sigma}^{\bullet} g(U_\sigma^a)$$

and

$$X_{\Gamma,\sigma}^a := pr(U_{\Gamma_\sigma}^a) .$$

Using 1) and the fact that $\mathscr{BT}$ is locally finite we find for any simplex τ in $\mathscr{BT}$ finitely many $g_1, \ldots, g_r \in \Gamma$ such that

$$U_\tau^a \cap U_{\Gamma\sigma}^a = \bigcup_{i=1}^{r} (U_\tau^a \cap U_{g_i\sigma}^a) .$$

This shows that $U_{\Gamma\sigma}^a$ is admissible open in $\Omega^{(d+1)}$ and consequently that $X_{\Gamma,\sigma}^a$ is admissible open in X_Γ. Moreover we see that pr induces an isomorphism of G-ringed spaces

$$\Gamma_\sigma \backslash U_\sigma^a \xrightarrow{\cong} X_{\Gamma,\sigma}^a .$$

But by [16] Prop. 6.3 (or [6] 6.3.3) the left hand side is an affinoid variety over K. Since Γ is cocompact there are simplices $\sigma_1, \ldots, \sigma_m$ such that any simplex in $\mathscr{BT}$ is a Γ-translate of one of them. The $X_{\Gamma,\sigma_1}^a, \ldots, X_{\Gamma,\sigma_m}^a$ then form an admissible covering of X_Γ by affinoid varieties over K; the intersections $X_{\Gamma,\sigma_i}^a \cap X_{\Gamma,\sigma_j}^a$ are easily checked to be affinoid, too. Therefore it follows that X_Γ is a separated analytic variety over K. By 3) and [16] Prop. 6.4 we have $X_{\Gamma,\sigma}^a \Subset_K X_{\Gamma,\sigma}^b$ for $0 < a < b < 1$ which implies that X_Γ is proper. Finally our assumption that Γ acts without fixed points on $\Omega^{(d+1)}$ guarantees that pr is an etale covering and that X_Γ is smooth.

We also obtain that

$$\begin{array}{ccc} \coprod_{g \in \Gamma} \Omega^{(d+1)} & \xrightarrow{\cong} & \Omega^{(d+1)} X_\Gamma \Omega^{(d+1)} \\ z \text{ in component } g & \mapsto & (gz, z) \end{array}$$

is an isomorphism (compare [38] Prop. 2 on p. 70). If Γ is torsionfree so that the subgroups Γ_σ are trivial we see that pr even is an analytic covering in the sense that the natural sequence

$$\mathscr{G}(X_\Gamma) \to \mathscr{G}(\Omega^{(d+1)}) \rightrightarrows \mathscr{G}(\Omega^{(d+1)} X_\Gamma \Omega^{(d+1)})$$

is exact for any sheaf $\mathscr{G}$ in the analytic topology.

Lemma 1 (*i*) *For any injective sheaf $\mathscr{G}$ on $\mathscr{V}$ the Γ-module $\mathscr{G}(\Omega^{(d+1)})$ is injective;*

(*ii*) *if either the given topology is finer than the etale topology or Γ is torsionfree then we have $\mathscr{G}(X_\Gamma) = \mathscr{G}(\Omega^{(d+1)})^\Gamma$ for any sheaf $\mathscr{G}$ on $\mathscr{V}$.*

Proof. (i) The functor

$$\begin{aligned} \text{sheaves on } \mathscr{V} &\to \Gamma\text{-modules} \\ \mathscr{G} &\mapsto \mathscr{G}(\Omega^{(d+1)}) \end{aligned}$$

respects injective objects since it has an exact left adjoint functor $M \to \mathscr{M}$ which is given in the following way: For any variety Y in $\mathscr{V}$ put

$$M(Y) := \left(\bigotimes_{x \in \Omega^{(d+1)}(Y)} M \right) \text{ modulo the subgroup generated}$$

by all elements of the form

$$(m \text{ in component } x) - (gm \text{ in component } gx)$$

with $g \in \Gamma$.

This is a presheaf on $\mathscr{V}$; $\mathscr{M}$ is defined to be the associated sheaf.

(ii) This follows easily from the two observations preceding this Lemma (compare [33] II.1.4).

Proposition 2 *We have the spectral sequence*

$$H^r(\Gamma, H^s(\Omega^{(d+1)})) \Rightarrow H^{r+s}(X_\Gamma)$$

in each of the following cases:

1) *Γ is torsionfree*;
2) *the given topology on $\mathscr{V}$ is finer than the etale topology*;
3) *the given cohomology theory is de Rham cohomology.*

Proof. In the cases 1) and 2) this is an immediate consequence of Lemma 1. But then also the assertion in the case 3) follows since de Rham cohomology can be computed on the small etale site on X_Γ. This can be seen in much the same way as the corresponding statement in algebraic geometry: [33] III.3.7 and II.1.6 based on [32] 4.1.8 and on the following fact.

Lemma 3 *For any etale morphism $f: X \to Y$ of K-analytic varieties we have*

$$f^{-1}\Omega^1_{Y/K} \underset{f^{-1}\mathscr{O}_Y}{\otimes} \mathscr{O}_X = \Omega^1_{X/K} .$$

Proof. Obviously there is a natural map $f^{-1}\Omega^1_{Y/K} \otimes \mathscr{O}_X \to \Omega^1_{X/K}$. Since both sides are coherent $\mathscr{O}_X$-modules it can be checked to be an isomorphism locally in the neighbourhood of each point of X. Furthermore we may assume that K is algebraically closed. The fibers of any etale morphism between K-affinoid varieties are finite. As the analytic topology is Hausdorff it follows that f is locally injective at each point. But then SGA 1 exp. I Cor. 4.4 and [6] 7.3.3 Prop. 5 imply that f even is locally an isomorphism.

Let us now assume that our cohomology theory has the properties I)–IV) in Paragraph 2 and is such that the spectral sequence in Proposition 2 is at our disposal. Preserving the notations introduced in Paragraph 4 we put, for any subset $I \subseteq \Delta$ and any abelian group M,

$$\begin{aligned} V_I(M) &:= C^\infty(G/P_I, M) \Big/ \sum_{i \in \Delta \setminus I} C^\infty(G/P_{I \cup \{i\}}, M) \\ &= V_I(\mathbb{Z}) \otimes M . \end{aligned}$$

It is known ([5] X.4.6 and 4.11) that in case M is a field of characteristic 0 the $V_I(M)$ are the irreducible constituents of the $PGL_{d+1}(K)$-module $C^\infty(G/P, M)$. Our main result says that $H^s(\Omega^{(d+1)}) = 0$ for $s > d$ and that

$$\begin{aligned} H^s(\Omega^{(d+1)}) &= \mathrm{Hom}_{\mathbb{Z}}(V_{\{1,\ldots,d-s\}}(\mathbb{Z}), A) \\ &= \mathrm{Hom}_A(V_{\{1,\ldots,d-s\}}(A), A) \qquad \text{for } 0 \leqq s \leqq d\,. \end{aligned}$$

Since $V_{\{1,\ldots,d-s\}}(A)$ is a free A-module by §4 Corollary 5 we get

$$\begin{aligned} H^r(\Gamma, H^s(\Omega^{(d+1)})) &= H^r(\Gamma, \mathrm{Hom}_A(V_{\{1,\ldots,d-s\}}(A), A)) \\ &= \mathrm{Ext}^r_{A[\Gamma]}(V_{\{1,\ldots,d-s\}}(A), A) \qquad \text{for } 0 \leqq s \leqq d \end{aligned}$$

so that the spectral sequence in Proposition 2 becomes

$$E_2^{r,s} = \left\{ \begin{matrix} \mathrm{Ext}^r_{A[\Gamma]}(V_{\{1,\ldots,d-s\}}(A), A) & \text{if } 0 \leqq s \leqq d \\ 0 & \text{otherwise} \end{matrix} \right\} \Rightarrow H^{r+s}(X_\Gamma)\,.$$

We therefore have the task of studying the groups $\mathrm{Ext}^r_{A[\Gamma]}(V_I(A), A)$.

Proposition 4 (*i*) *$V_I(\mathbb{Z})$ has a projective resolution by finitely generated free $\mathbb{Z}[\Gamma]$-modules;*

(*ii*) *$\mathrm{Ext}^r_{\mathbb{Z}[\Gamma]}(V_I(\mathbb{Z}), \mathbb{Z})$ is finitely generated for any $r \geqq 0$;*

(*iii*) *there is a natural exact sequence*

$$0 \to \mathrm{Ext}^r_{\mathbb{Z}[\Gamma]}(V_I(\mathbb{Z}), \mathbb{Z}) \underset{\mathbb{Z}}{\otimes} A \to \mathrm{Ext}^r_{A[\Gamma]}(V_I(A), A) \to \mathrm{Tor}_{\mathbb{Z}}(\mathrm{Ext}^{r+1}_{\mathbb{Z}[\Gamma]}(V_I(\mathbb{Z}), \mathbb{Z}), A) \to 0$$

for any $r \geqq 0$.

Proof. (i) The proof of this result is quite complicated and will be given in the last Paragraph (Proposition 16). (ii) This is an immediate consequence of the first assertion. (iii) Let $F_\bullet \to V_I(\mathbb{Z})$ be a projective resolution by finitely generated free $\mathbb{Z}[\Gamma]$-modules. Then $F_\bullet \otimes_{\mathbb{Z}} A \to V_I(A)$ is a projective resolution by finitely generated free $A[\Gamma]$-modules since $V_I(\mathbb{Z})$ is $\mathbb{Z}$-free by the argument in the proof of §4 Corollary 5. Therefore $\mathrm{Ext}^*_{\mathbb{Z}[\Gamma]}(V_I(\mathbb{Z}), \mathbb{Z})$, resp. $\mathrm{Ext}^*_{A[\Gamma]}(V_I(A), A)$, can be computed from the complex $\mathrm{Hom}_{\mathbb{Z}[\Gamma]}(F_\bullet, \mathbb{Z})$, resp. $\mathrm{Hom}_{A[\Gamma]}(F_\bullet \otimes_{\mathbb{Z}} A, A)$. The assertion follows now from the universal coefficient theorem applied to the complex

$$\mathrm{Hom}_{A[\Gamma]}\left(F_\bullet \underset{\mathbb{Z}}{\otimes} A, A\right) = \mathrm{Hom}_{\mathbb{Z}[\Gamma]}(F_\bullet, \mathbb{Z}) \underset{\mathbb{Z}}{\otimes} A\,.$$

This Proposition in particular implies that we may pass in the corresponding spectral sequences for $H^*_{et}(\bar{X}_\Gamma, \mathbb{Z}/l^\nu\mathbb{Z})$ with $l \neq p$ to the projective limit with respect to ν and obtain the Galois equivariant spectral sequence

$$E_2^{r,s} = \left\{ \begin{matrix} \mathrm{Ext}^r_{\mathbb{Q}_l[\Gamma]}(V_{\{1,\ldots,d-s\}}(\mathbb{Q}_l), \mathbb{Q}_l)(-s) & \\ & \text{if } 0 \leqq s \leqq d \\ 0 & \text{otherwise} \end{matrix} \right\} \Rightarrow H^{r+s}_{et}(\bar{X}_\Gamma, \mathbb{Q}_l)$$

for $\mathbb{Q}_l$-adic cohomology. In the following we will compute the groups $\mathrm{Ext}^r_{A[\Gamma]}(V_I(A), A)$ under the assumptions that

— Γ is a discrete cocompact subgroup in $PGL_{d+1}(K)$, and
— A is a field of characteristic 0.

Because of the above Proposition this amounts to the computation of

$$\dim_{\mathbb{C}} \mathrm{Ext}^r_{\mathbb{C}[\Gamma]}(V_I(\mathbb{C}), \mathbb{C}) .$$

Since Γ is cocompact the $PGL_{d+1}(K)$-representation

$$\mathrm{Ind}_\Gamma := C^\infty(PGL_{d+1}(K)/\Gamma, \mathbb{C})$$

is admissible. By Shapiro's lemma ([13] A.8) we have

$$\mathrm{Ext}^*_{\mathbb{C}[\Gamma]}(V_I(\mathbb{C}), \mathbb{C}) = \mathrm{Ext}^*(V_I(\mathbb{C}), \mathrm{Ind}_\Gamma)$$

where Ext* (without a subscript) always denotes the Ext-functor on the category of smooth $PGL_{d+1}(K)$-representations. In order to understand these Ext-groups on the right hand side we use the ideas in the proof of the Garland–Casselman theorem in [12], [13], and [5]. The representation Ind_Γ is unitary and therefore completely reducible (compare [12] and [11] 2.1.14). In addition the admissibility implies that only finitely many of its irreducible constituents can have a nonzero vector fixed under the Iwahori subgroup B. We therefore obtain a decomposition

$$\mathrm{Ind}_\Gamma \cong V_0 \oplus V_1 \oplus \ldots \oplus V_m$$

into admissible unitary representations V_j such that $V_0^B = 0$ and V_j is irreducible with $V_j^B \neq 0$ for $1 \leqq j \leqq m$. In particular we get

$$\mathrm{Ext}^*(V_I(\mathbb{C}), \mathrm{Ind}_\Gamma) \cong \bigoplus_{j=0}^{m} \mathrm{Ext}^*(V_I(\mathbb{C}), V_j) .$$

Proposition 5 *Let V and V' be smooth $PGL_{d+1}(K)$-representations such that $V^B = 0$ but V' is generated as a $PGL_{d+1}(K)$-module by $(V')^B$. Then we have*

$$\mathrm{Ext}^*(V', V) = 0 .$$

Proof. Consider any Yoneda extension of smooth $PGL_{d+1}(K)$-representations

$$E: 0 \to V = E_0 \to E_1 \to \ldots \to E_{r+1} = V' \to 0 .$$

In any E_j we have the subrepresentation $\hat{E}_j$ generated by E_j^B. In this way we obtain the commutative diagram

$$\begin{array}{ccccccccc} 0 & \to & V = E_0 & \to & E_1 & \to & \ldots & \to & E_{r+1} = V' \to 0 \\ & & \uparrow \subseteq & & \uparrow \subseteq & & & & \| \\ 0 & \to & 0 = \hat{E}_0 & \to & \hat{E}_1 & \to & \ldots & \to & \hat{E}_{r+1} = V' \to 0 . \end{array}$$

If the lower row would be exact, too, then obviously the extension class of E would be trivial. That this indeed holds true is a consequence of the following fact: If a smooth $PGL_{d+1}(K)$-representation is generated by its Iwahori fixed vectors then all of its subrepresentations have the same property. This is proved in [2] 4.4 with B replaced by a principal congruence subgroup. But the same proof works for B if one takes into account the following additional observation: Let $J \subseteq \Delta$ be any subset and let L_J be the standard Levi component of P_J, i.e., the intersection of P_J with its transpose; for any $w \in (W^J)^{-1}$ we then have

$$wBw^{-1} \cap L_J = B \cap L_J .$$

Proposition 6 *If V is an irreducible (admissible) $PGL_{d+1}(K)$-representation such that $V^B \neq 0$ and $V| \cong V_J(\mathbb{C})$ for all $J \subseteq \Delta$ then* $\mathrm{Ext}^*(V_I(\mathbb{C}), V) = 0$.

Proof. Let L denote the subgroup of diagonal matrices in $PGL_{d+1}(K)$, and let $\delta: P \to \mathbb{C}^\times$ be the modulus character. Our assumptions guarantee that there is an unramified character $\chi: L \to \mathbb{C}^\times$ with

$$\chi \neq (w\delta^{-1/2})\delta^{1/2} =: \delta_w \qquad \text{for all } w \in W$$

such that V is a constituent of the unramified principal series representation

$$\mathrm{Ind}(\chi) := \{f \in C^\infty(G, \mathbb{C}) : f(hg) = \chi(h)f(g) \text{ for all } h \in P, g \in G\}$$

where we view χ as a character of P in the obvious way ([5] X.3.2). By an easy induction argument (compare the proof of X.4.3 in [5]) it suffices to prove that

$$\mathrm{Ext}^*(V_I(\mathbb{C}), \mathrm{Ind}(\chi)) = 0 .$$

From [13] A.12 and 1.3 we get

$$\begin{aligned} \mathrm{Ext}^*(V_I(\mathbb{C}), \mathrm{Ind}(\chi)) &= \mathrm{Ext}^*_L(V_I(\mathbb{C})_U, \chi) \\ &= \bigoplus_{w \in W^I} \mathrm{Ext}^*_L(\delta_w, \chi) ; \end{aligned}$$

here U denotes the unipotent radical of P and Ext^*_L is the Ext-functor on the category of smooth L-representations. Because of $\delta_w \neq \chi$ the same argument as in [5] IX.1.9 shows that those last groups vanish.

The following is the main result of Casselman ([13] 2.1 or [5] XI.4.5).

Proposition 7 *If $V_J(\mathbb{C})$ is a constituent of* Ind_Γ *for some $J \subseteq \Delta$ then $J = \phi$ or Δ.*

If we apply these three Propositions to our above decomposition of Ind_Γ then we obtain

$$\begin{aligned} &\mathrm{Ext}^*(V_I(\mathbb{C}), \mathrm{Ind}_\Gamma) \\ &\quad = \mathrm{Ext}^*(V_I(\mathbb{C}), \mathbb{C}) \oplus [\mathrm{Hom}_G(V_\phi(\mathbb{C}), \mathrm{Ind}_\Gamma) \otimes \mathrm{Ext}^*(V_I(\mathbb{C}), V_\phi(\mathbb{C}))] . \end{aligned}$$

Observe that $\mathrm{Hom}_G(V_\Delta(\mathbb{C}), \mathrm{Ind}_\Gamma) = \mathbb{C}$.

Proposition 8 $\mathrm{Ext}^r(V_I(\mathbb{C}), \mathbb{C}) \cong \begin{cases} \mathbb{C} & \text{if } r = \#\Delta \backslash I , \\ 0 & \text{otherwise} . \end{cases}$

Proof. Put $\bar{I} := \{d + 1 - i : i \in I\}$. The considerations on p. 915 in [13] imply that $V_{\bar{I}}(\mathbb{C})$ is isomorphic to the smooth contragredient of $V_I(\mathbb{C})$. By [13] A.11 we then have

$$\mathrm{Ext}^*(V_I(\mathbb{C}), \mathbb{C}) \cong \mathrm{Ext}^*(\mathbb{C}, V_{\bar{I}}(\mathbb{C})) .$$

But those last groups were computed by Casselman ([13] A.13 or [5] X.4.7).

In order to determine the remaining terms we need the following general duality statement. Let $H_*(\bullet) = H_*(PGL_{d+1}(K), \bullet)$ denote the homology functor on the category of smooth $PGL_{d+1}(K)$-representations (see [13] p. 925).

Duality theorem. *There is a natural isomorphism*

$$\mathrm{Ext}^{d-*}(V_\phi(\mathbb{C}), \bullet) \cong H_*(\bullet) .$$

Proof. We put

$$\mathscr{H} := \text{space of all } \mathbb{C}\text{-valued locally constant functions with compact support on } PGL_{d+1}(K) .$$

Via convolution (fixing once and for all a Haar measure on $PGL_{d+1}(K)$) $\mathscr{H}$ is an associative ring which acts in a natural way on any smooth $PGL_{d+1}(K)$-representation V. The space $\mathscr{H}$ itself is a smooth representation via left translations. More generally we have the smooth representation $\mathscr{H} \otimes_{\mathbb{C}} V$ with $PGL_{d+1}(K)$ acting only on the first factor. The map

$$\rho_V : \mathscr{H} \underset{\mathbb{C}}{\otimes} V \to V$$

$$\varphi \otimes v \mapsto \varphi v$$

then is a $PGL_{d+1}(K)$-equivariant epimorphism. By a result of Blanc (see [13] A.4) $\mathscr{H} \otimes_{\mathbb{C}} V$ is a projective object in the category of smooth representations. The maps $\rho_V, \rho_{\ker(\rho V)}, \ldots$ therefore constitute a functorial projective resolution of V which we will use later on. We also observe that the map $\rho_{\mathbb{C}} \otimes id_V$ induces a natural $\mathbb{C}$-linear isomorphism

$$H_0\left(\mathscr{H} \underset{\mathbb{C}}{\otimes} V\right) \cong V .$$

In order to understand the groups $\mathrm{Ext}^*(V_\phi(\mathbb{C}), V)$ we use the following explicit resolution of $V_\phi(\mathbb{C})$. Let $\mathscr{BT}_{(q)}$ be the set of oriented q-simplices of $\mathscr{BT}$; for $\sigma \in \mathscr{BT}_{(q)}$ with $q \geqq 1$ we denote by $\bar{\sigma}$ that oriented simplex with the same underlying simplex as σ but with the opposite orientation. Let $C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C})$ denote the space of $\mathbb{C}$-valued oriented q-cochains with finite support on $\mathscr{BT}$. On the one hand we can view $C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C})$ also as the space of oriented q-chains so that we have the boundary map

$$\partial_q : C_c^{or}(\mathscr{BT}_{(q+1)}, \mathbb{C}) \to C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C}) .$$

Because of the contractibility of $\mathscr{BT}$ the complex $(C_c^{or}(\mathscr{BT}_{(\bullet)}, \mathbb{C}), \partial_\bullet)$ is a resolution of $\mathbb{C}$. On the other hand since $\mathscr{BT}$ is locally finite the coboundary map restricts to a map

$$d^q : C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C}) \to C_c^{or}(\mathscr{BT}_{(q+1)}, \mathbb{C}) .$$

By [3] 3.3 and 5.6 we have the exact sequence of smooth representations

$$0 \to C_c^{or}(\mathscr{BT}_{(0)}, \mathbb{C}) \xrightarrow{d^0} \ldots \xrightarrow{d^{d-1}} C_c^{or}(\mathscr{BT}_{(d)}, \mathbb{C}) \to V_\phi(\mathbb{C}) \to 0 .$$

We want to show that $C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C})$ is a projective representation. Fix a $\sigma \in \mathscr{BT}_{(q)}$ and define $\omega_\sigma \in C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C})$ by

$$\omega_\sigma(\tau) := \begin{cases} +1 & \text{if } \tau = \sigma , \\ -1 & \text{if } \tau = \bar{\sigma} , \\ 0 & \text{otherwise} ; \end{cases}$$

let $C(\omega_\sigma) \subseteq C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C})$ be the subrepresentation generated by ω_σ. Obviously $C_c^{or}(\mathscr{BT}_{(q)}, \mathbb{C})$ is a finite direct sum of subrepresentations of the form $C(\omega_\sigma)$ so that

it suffices to show that $C(\omega_\sigma)$ is projective. Let $B_\sigma \subseteq G$ denote the stabilizer of σ; the image of B_σ in $PGL_{d+1}(K)$ is a compact open subgroup. In case there is no $g \in G$ such that $g\sigma = \bar{\sigma}$ we have

$$C(\omega_\sigma) \cong C_c^\infty(G/B_\sigma, \mathbb{C}) .$$

Since $\mathrm{Hom}_G(C_c^\infty(G/B_\sigma, \mathbb{C}), V) = V^{B_\sigma}$ is an exact functor on smooth $PGL_{d+1}(K)$-representations V ([13] App.) we see that $C(\omega_\sigma)$ is projective. Let us now assume that there is a $h \in G$ such that $h\sigma = \bar{\sigma}$. Then h normalizes B_σ with $h^2 \in B_\sigma$; we have

$$C(\omega_\sigma) \cong \{\chi \in C_c^\infty(G/B_\sigma, \mathbb{C}) : \chi(gh) = -\chi(g) \text{ for all } g \in G\}$$

and consequently

$$\mathrm{Hom}_G(C(\omega_\sigma), V) \cong (V^{B_\sigma})^{h=-1}$$

which again is an exact functor. Therefore $C(\omega_\sigma)$ is projective in this case, too. Using this explicit projective resolution of $V_\phi(\mathbb{C})$ we see that the groups $\mathrm{Ext}^*(V_\phi(\mathbb{C}), V)$ can be computed as the homology groups of the complex

$$\mathrm{Hom}_G((C_c^{or}(\mathscr{B}\mathscr{T}_{(\bullet)}, \mathbb{C}), d^\bullet), V) .$$

As an immediate consequence we obtain that

$$\mathrm{Ext}^r(V_\phi(\mathbb{C}), V) = 0 \text{ for } r > d . \tag{1}$$

For $V = \mathscr{H}$ a straightforward computation shows that

$$\mathrm{Hom}_G((C_c^{or}(\mathscr{B}\mathscr{T}_{(\bullet)}, \mathbb{C}), d^\bullet), \mathscr{H}) \xrightarrow{\cong} (C_c^{or}(\mathscr{B}\mathscr{T}_{(\bullet)}, \mathbb{C}), \partial_\bullet)$$

$$f \mapsto (\sigma \mapsto f(\omega_\sigma)(1))$$

is an isomorphism of complexes. This implies, more generally, that

$$\mathrm{Hom}_G\Big((C_c^{or}(\mathscr{B}\mathscr{T}_{(\bullet)}, \mathbb{C}), d^\bullet), \mathscr{H} \underset{\mathbb{C}}{\otimes} V\Big) \cong \mathrm{Hom}_G((C_c^{or}(\mathscr{B}\mathscr{T}_{(\bullet)}, \mathbb{C}), d^\bullet), \mathscr{H}) \underset{\mathbb{C}}{\otimes} V$$

$$\cong (C_c^{or}(\mathscr{B}\mathscr{T}_{(\bullet)}, \mathbb{C}), \partial_\bullet) \underset{\mathbb{C}}{\otimes} V .$$

Since $\mathscr{B}\mathscr{T}$ is contractible the last complex is a resolution of V which in particular means that

$$\mathrm{Ext}^r\Big(V_\phi(\mathbb{C}), \mathscr{H} \underset{\mathbb{C}}{\otimes} V\Big) = 0 \qquad \text{for } r < d . \tag{2}$$

From the facts (1) and (2) we deduce (compare [22] I.7.4) that $\mathrm{Ext}^*(V_\phi(\mathbb{C}), \bullet)$ is the left derived functor of $\mathrm{Ext}^d(V_\phi(\mathbb{C}), \bullet)$. It remains to exhibit a natural isomorphism

$$\mathrm{Ext}^d(V_\phi(\mathbb{C}), V) \cong H_0(V) .$$

For that we consider the natural transformation

$$\mathrm{Hom}_G(C_c^{or}(\mathscr{B}\mathscr{T}_{(0)}, \mathbb{C}), V) \to H_0(V) = V/\ldots$$

$$f \mapsto f(\omega_\sigma) \bmod \ldots$$

where σ is some fixed vertex of $\mathscr{BT}$. Since G acts transitively on $\mathscr{BT}_{(0)}$ this map actually is independent of the choice of σ. It is surjective since it identifies with the projection map

$$\mathrm{Hom}_G(C_c^{or}(\mathscr{BT}_{(0)}, \mathbb{C}), V) = V^{B_\sigma} \xrightarrow{pr} H_0(V)$$

and since the functor $V \mapsto V^{B_\sigma}$ is exact.

Next we claim that

$$\mathrm{im}\, d^0 \subseteq \ker(C_c^{or}(\mathscr{BT}_{(1)}, \mathbb{C}) \xrightarrow{pr} H_0(C_c^{or}(\mathscr{BT}_{(1)}, \mathbb{C})))$$

holds true. We have $d^0(\omega_\sigma) = \sum_\tau \omega_{\langle \tau, \sigma \rangle}$ where the sum ranges over all vertices τ which are neighbours of σ; $\langle \tau, \sigma \rangle$ then denotes the corresponding oriented 1-simplex. For any τ there is a $g \in G$ such that $\tau = g\sigma$; then also $g^{-1}\sigma$ is a neighbouring vertex of σ. We get

$$pr(\omega_{\langle \tau, \sigma \rangle}) = pr(\omega_{\langle g\sigma, \sigma \rangle}) = pr(\omega_{\langle \sigma, g^{-1}\sigma \rangle}) = -pr(\omega_{\langle g^{-1}\sigma, \sigma \rangle})$$

and consequently

$$pr(\omega_{\langle \tau, \sigma \rangle} + \omega_{\langle g^{-1}\sigma, \sigma \rangle}) = 0 .$$

If $g^{-1}\sigma = \tau$ then already $pr(\omega_{\langle \tau, \sigma \rangle}) = 0$ must hold. We see that $pr(d^0(\omega_\sigma)) = 0$.

Therefore our above natural transformation vanishes on the image of $\mathrm{Hom}_G(d^0, V)$ and factorizes through a surjective natural transformation

$$\mathrm{Ext}^d(V_\phi(\mathbb{C}), V) \to H_0(V) .$$

In order to establish injectivity it suffices to consider the case $V = \mathscr{H}$ where both sides are easily checked to be 1-dimensional over $\mathbb{C}$.

Remark. Similar considerations as in the above proof show that

$$\mathrm{Ext}^r(\mathbb{C}, \mathscr{H}) \cong \begin{cases} V_\phi(\mathbb{C}) & \text{if } r = d , \\ 0 & \text{otherwise} . \end{cases}$$

Proposition 9

$$\mathrm{Ext}^r(V_I(\mathbb{C}), V_\phi(\mathbb{C})) \cong \begin{cases} \mathbb{C} & \text{if } r = \#I , \\ 0 & \text{otherwise} . \end{cases}$$

Proof. By [13] A.11 and the above duality we have

$$\begin{aligned} \mathrm{Ext}^*(V_I(\mathbb{C}), V_\phi(\mathbb{C})) &\cong \mathrm{Ext}^*(V_\phi(\mathbb{C}), V_{\bar{I}}(\mathbb{C})) \\ &\cong H_{d-*}(V_{\bar{I}}(\mathbb{C})) . \end{aligned}$$

By [13] A.10 the $\mathbb{C}$-dual of this last group is isomorphic to $\mathrm{Ext}^{d-*}(\mathbb{C}, V_I(\mathbb{C}))$ so that we again are reduced to Casselman's computation.

Definition. $\mu(\Gamma) :=$ multiplicity of the Steinberg representation $V_\phi(\mathbb{C})$ in Ind_Γ.

Altogether we have proved now the following result.

Theorem 3 *If $\Gamma \subseteq \mathrm{PGL}_{d+1}(K)$ is a cocompact discrete subgroup and A is a field of characteristic 0 then we have*

$$\mathrm{Ext}^r_{A[\Gamma]}(V_I(A), A) \cong \begin{cases} A & \text{if } \#I = d - r \neq \frac{d}{2}, \\ A^{\mu(\Gamma)} & \text{if } \#I = r \neq \frac{d}{2}, \\ A^{\mu(\Gamma)+1} & \text{if } \#I = r = \frac{d}{2}, \\ 0 & \text{otherwise}. \end{cases}$$

In particular we obtain that in our spectral sequence nonvanishing $E_2^{r,s}$-terms only occur on the lines $r = s$ and $r + s = d$. If d is even then all differentials in the spectral sequence automatically must be zero. The only interesting cohomology group of X_Γ is $H^d(X_\Gamma)$. Let

$$H^d(X_\Gamma) = F^0 H^d(X_\Gamma) \supseteq F^1 H^d(X_\Gamma) \supseteq \ldots \supseteq F^{d+1} H^d(X_\Gamma) = 0$$

be the filtration such that

$$F^r H^d(X_\Gamma)/F^{r+1} H^d(X_\Gamma) = E_\infty^{r,d-r}.$$

Corollary 10 *Assume that $A = H^0(Sp(K))$ is a field of characteristic 0. If d is even we have*

$$H^s(X_\Gamma) \cong \begin{cases} A & \text{if } 0 \leqq s \leqq 2d, \quad s \neq d, s \text{ even}, \\ A^{(d+1)\mu(\Gamma)+1} & \text{if } s = d, \\ 0 & \text{otherwise} \end{cases}$$

and

$$F^r H^d(X_\Gamma)/F^{r+1} H^d(X_\Gamma) \cong \begin{cases} A^{\mu(\Gamma)} & \text{if } 0 \leqq r \leqq d, \quad r \neq \frac{d}{2}, \\ A^{\mu(\Gamma)+1} & \text{if } r = \frac{d}{2}. \end{cases}$$

If d is odd we have

$$(d+1)\mu(\Gamma) - 2 \leqq \dim_A H^d(X_\Gamma) \leqq (d+1)\mu(\Gamma).$$

In $\mathbb{Q}_l$-adic cohomology the spectral sequence always degenerates for reasons of weight. In case $d = 1$ the next result was proved in [48] by a different method.

Theorem 4 *For $l \neq p$ we have*

$$H^s_{et}(\bar{X}_\Gamma, \mathbb{Q}_l) \cong \begin{cases} \mathbb{Q}_l\left(-\frac{s}{2}\right) & \text{if } 0 \leqq s \leqq 2d, \quad s \neq d, s \text{ even}, \\ 0 & \text{if } s \neq d \text{ odd or } s > 2d \end{cases}$$

and

$$F^r H^d_{et}(\bar{X}_\Gamma, \mathbb{Q}_l)/F^{r+1} H^d_{et}(\bar{X}_\Gamma, \mathbb{Q}_l) \cong \begin{cases} \mathbb{Q}_l(r-d)^{\mu(\Gamma)} & \text{if } 0 \leqq r \leqq d, \quad r \neq \frac{d}{2}, \\ \mathbb{Q}_l\left(-\frac{d}{2}\right)^{\mu(\Gamma)+1} & \text{if } r = \frac{d}{2}. \end{cases}$$

In de Rham cohomology the spectral sequence degenerates as well, and we can use it in order to strengthen a result of Mustafin ([39] 4.1.II).

Theorem 5 *If K is of characteristic 0 we have*

$$H^d_{DR}(X_\Gamma) \cong \begin{cases} K^{(d+1)\mu(\Gamma)} & \text{if } d \text{ is odd}, \\ K^{(d+1)\mu(\Gamma)+1} & \text{if } d \text{ is even} \end{cases}$$

and

$$H^s(X_\Gamma, \Omega^j_{X_\Gamma}) \cong \begin{cases} K & \text{if } s = j \neq \frac{d}{2}, \\ 0 & \text{if } s \neq j, \quad s + j \neq d. \end{cases}$$

Proof. By [39] 4.1.I the analytic variety X_Γ is algebraizable to a projective variety over K. The GAGA-principle ([28]) then implies that the coherent and the de Rham cohomology of X_Γ are equal to the corresponding algebraic cohomology groups. But since K has characteristic 0 we know in the algebraic context that the de Rham spectral sequence for X_Γ degenerates and furthermore that the strong Lefschetz theorem holds. From the first fact we deduce that

$$\dim_K H^s_{DR}(X_\Gamma) = \sum_{j=0}^{s} \dim_K H^{s-j}(X_\Gamma, \Omega^j)$$

and from the second that

$$\dim_K H^{2s}_{DR}(X_\Gamma) \geqq \dim_K H^s(X_\Gamma, \Omega^s) \geqq 1.$$

Because of this last inequality our above spectral sequence has to degenerate so that $\dim_K H^s_{DR}(X_\Gamma) = 0$ for $s \neq d$ odd, $= 1$ for $s \neq d$ even, and is as asserted for $s = d$.

For both, $\mathbb{Q}_l$-adic and de Rham cohomology we have a natural nilpotent monodromy operator on $H^d(X_\Gamma)$ ([15] 1.7.2 and [25]). We expect that our filtration $F^\bullet H^d(X_\Gamma)$ in both cases is the associated monodromy filtration ([15] 1.6) and that $F^\bullet H^d_{DR}(X_\Gamma)$ is opposite to the Hodge–de Rham filtration.

In the last part of this Paragraph we will deduce from Theorem 1 a proof of (a slightly weakened version of) the p-adic Shimura isomorphism stated in [44] p. 226. We assume from now on that K is of characteristic 0 and, for simplicity, we put

$$O := \mathcal{O}(\Omega^{(2)}) \qquad \text{and} \qquad \Omega := \Omega^1(\Omega^{(2)}).$$

In [44] p. 225 there was constructed a $\mathrm{SL}_2(K)$-equivariant homomorphism

$$I: \Omega/dO \to C_{har}(\mathscr{B}\mathscr{T}, K)$$

where on the right hand side we have the space of K-valued harmonic 1-cochains

on the Bruhat–Tits tree $\mathscr{BT}$. It is an exercise to check that under the natural isomorphism

$$C_{har}(\mathscr{BT}, K) = \ker(\mathrm{Hom}_{\mathbb{Z}}(C^{\infty}(\mathbb{P}^1(K), \mathbb{Z}), K) \to K)$$
$$\mu \mapsto \mu(1)$$

(compare [44] Lemma on p. 226) the map I corresponds to the isomorphism

$$H^1_{DR}(\Omega^{(2)}) = \ker(\mathrm{Hom}_{\mathbb{Z}}(C^{\infty}(\mathbb{P}^1(K), \mathbb{Z}), K) \to K)$$

given in Theorem 1. We therefore have the exact sequence of $\mathrm{SL}_2(K)$-modules

$$0 \to K \xrightarrow{\subseteq} O \xrightarrow{d} \Omega \xrightarrow{I} C_{har}(\mathscr{BT}, K) \to 0 .$$

More generally, tensoring with the natural $\mathrm{SL}_2(K)$-representation

$$\mathrm{Sym}^n := n\text{-th symmetric power of } K^2$$

we obtain the exact sequence of $\mathrm{SL}_2(K)$-modules

$$0 \to \mathrm{Sym}^n \to O \underset{K}{\otimes} \mathrm{Sym}^n \to \Omega \underset{K}{\otimes} \mathrm{Sym}^n \to C_{har}(\mathscr{BT}, \mathrm{Sym}^n) \to 0$$

for any $n \geqq 0$.

Let $x \in O$ denote the coordinate function

$$x : \Omega^{(2)}(\hat{\hat{K}}) \to \hat{\hat{K}}$$
$$[z_0 : z_1] \mapsto \frac{z_0}{z_1} .$$

The O-module Ω is free with generator dx. Also let $e_1 = (1, 0)$, $e_2 = (0, 1)$ be the standard basis of K^2. We consider the filtration

$$O \underset{K}{\otimes} \mathrm{Sym}^n = F^0 \supseteq F^1 \supseteq \ldots \supseteq F^{n+1} = 0$$

which is defined by

$$F^q := \text{the } O\text{-submodule generated by}$$
$$(xe_1 + e_2)^q e_1^{n-q-j} e_2^j \qquad \text{for } 0 \leqq j \leqq n - q ;$$

it induces a corresponding filtration

$$\Omega \underset{K}{\otimes} \mathrm{Sym}^n = \Omega \underset{O}{\otimes} F^0 \supseteq \Omega \underset{O}{\otimes} F^1 \supseteq \ldots \supseteq \Omega \underset{O}{\otimes} F^{n+1} = 0 .$$

(This is inspired by [1].) Because of

$$g(xe_1 + e_2) = (-\gamma x + \alpha)^{-1} \cdot (xe_1 + e_2) \qquad \text{for } g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(K)$$

both filtrations are $\mathrm{SL}_2(K)$-invariant. $O \otimes_K \mathrm{Sym}^n$ is a free O-module with basis $e_1^{n-q} e_2^q$ for $0 \leqq q \leqq n$. On the other hand from $e_1^i e_2 = (xe_1 + e_2) e_1^i - x e_1^{i+1}$ we deduce inductively that

$$(xe_1 + e_2)^q e_1^{n-q-j} e_2^j \in O(xe_1 + e_2)^q e_1^{n-q} + F^{q+1} \qquad \text{for } 1 \leqq j \leqq n - q .$$

Therefore the elements $(xe_1 + e_2)^q e_1^{n-q}$ for $0 \leqq q \leqq n$ also generate $O \otimes_K \mathrm{Sym}^n$ as an O-module and then have to be an O-basis, too. This shows that

$$O \to F^q/F^{q+1}$$
$$f \mapsto f(xe_1 + e_2)^q e_1^{n-q} \bmod F^{q+1}$$

is an O-module isomorphism for $0 \leqq q \leqq n$. Using the formula

$$\alpha e_1 + \gamma e_2 = (-\gamma x + \alpha)e_1 + \gamma(xe_1 + e_2)$$

we compute

$$\begin{aligned} g[f(xe_1 + e_2)^q e_1^{n-q}] &= g(f)\cdot(-\gamma x + \alpha)^{-q}\cdot(xe_1 + e_2)^q\cdot(\alpha e_1 + \gamma e_2)^{n-q} \\ &= g(f)\cdot(-\gamma x + \alpha)^{-q}\cdot(xe_1 + e_2)^q\cdot[(-\gamma x + \alpha)e_1 + \gamma(xe_1 + e_2)]^{n-q} \\ &\equiv g(f)\cdot(-\gamma x + \alpha)^{n-2q}\cdot(xe_1 + e_2)^q e_1^{n-q} \bmod F^{q+1} \end{aligned}$$

for $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(K)$. This means that the above isomorphism becomes an O-module and a $\mathrm{SL}_2(K)$-module isomorphism

$$\Theta_q : O[n-2q] \to F^q/F^{q+1}$$
$$f \mapsto f(xe_1 + e_2)^q e_1^{n-q} \bmod F^{q+1}$$

if, for any $m \in \mathbb{Z}$, $O[m]$ denotes O with the twisted $\mathrm{SL}_2(K)$-action

$$\pi_m(g)f := g(f)\cdot(-\gamma x + \alpha)^m \qquad \text{for } g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

Similarly we have an O-$\mathrm{SL}_2(K)$-isomorphism

$$\Theta'_q : O[n-2q-2] \to \Omega \underset{O}{\otimes} F^q/\Omega \underset{O}{\otimes} F^{q+1}$$
$$f \mapsto dx \otimes f(xe_1 + e_2)^q e_1^{n-q} \bmod \ldots .$$

The identity

$$(*) \qquad \begin{aligned} &(d \otimes \mathrm{id})(f(xe_1 + e_2)^q e_1^{n-q}) \\ &= dx \otimes qf(xe_1 + e_2)^{q-1} e_1^{n-q+1} + df \otimes (xe_1 + e_2)^q e_1^{n-q} \end{aligned}$$

for $1 \leqq q \leqq n$ implies

$$(d \otimes \mathrm{id})(f(xe_1 + e_2)^q e_1^{n-q}) \equiv dx \otimes qf(xe_1 + e_2)^{q-1} e_1^{n-q+1} \bmod \Omega \underset{O}{\otimes} F^q .$$

This is the commutativity of the diagram

$$\begin{array}{ccc} O[n-2q] & \xrightarrow{q\bullet} & O[n-2q] \\ \Big\downarrow{\scriptstyle \Theta_q} & & \Big\downarrow{\scriptstyle \Theta'_{q-1}} \\ F^q/F^{q+1} & \xrightarrow{d\otimes \mathrm{id}} & \Omega \underset{O}{\otimes} F^{q-1}/\Omega \underset{O}{\otimes} F^q \end{array}$$

for $1 \leqq q \leqq n$. As a consequence we obtain that $d \otimes \mathrm{id}$ induces an isomorphism

$$F^1 \xrightarrow{\cong} \underset{K}{\Omega \otimes} \mathrm{Sym}^n / \underset{O}{\Omega \otimes} F^n$$

so that we have in particular the $\mathrm{SL}_2(K)$-invariant decomposition

$$\underset{K}{\Omega \otimes} \mathrm{Sym}^n = (d \otimes \mathrm{id})(F^1) \oplus \left(\underset{O}{\Omega \otimes} F^n \right).$$

We see that our original exact sequence can be simplified to

$$0 \to \mathrm{Sym}^n \to \underset{K}{O \otimes} \mathrm{Sym}^n / F^1 \to \underset{O}{\Omega \otimes} F^n \to C_{har}(\mathscr{BT}, \mathrm{Sym}^n) \to 0 .$$

In order to compute the map in the middle we observe that the above identity $(*)$ also implies

$$dx \otimes f(xe_1 + e_2) e_1^{n-q} \equiv -\frac{1}{q+1} df \otimes (xe_1 + e_2)^{q+1} e_1^{n-(q+1)} \bmod (d \otimes \mathrm{id})(F^1)$$

for $0 \leqq q < n$. Inductively we get

$$dx \otimes f e_1^n \equiv (-1)^n \frac{1}{n!} dx \otimes f^{(n)}(xe_1 + e_2)^n \bmod (d \otimes \mathrm{id})(F^1) .$$

Here, as usual, we put $f^{(n)} := (f^{(n-1)})'$ where $df = f'dx$. In particular, the diagram

$$\begin{array}{ccc} \underset{K}{O \otimes} \mathrm{Sym}^n / F^1 & \longrightarrow & \underset{O}{\Omega \otimes} F^n \\ \Big\uparrow \Theta_0 & & \Big\uparrow \Theta_n' \\ O[n] & \longrightarrow & O[-n-2] \\ & f \mapsto (-1)^n \frac{1}{n!} f^{(n+1)} & \end{array}$$

is commutative. In $O[n]$ we have the $\mathrm{SL}_2(\mathrm{K})$-invariant subspace P_n of all polynomials in x of degree $\leqq n$; obviously it is the kernel of the map $f \mapsto f^{(n+1)}$. Thus we finally arrive at the exact sequence of $\mathrm{SL}_2(K)$-modules

$$0 \to P_n \xrightarrow{\subseteq} O[n] \to O[-n-2] \to C_{har}(\mathscr{BT}, \mathrm{Sym}^n) \to 0$$

$$f \mapsto f^{(n+1)} \qquad f \mapsto (I \otimes \mathrm{id})(dx \otimes f(xe_1 + e_2)^n) .$$

We now fix a cocompact discrete subgroup $\Gamma \subseteq \mathrm{SL}_2(K)$. Let $\mathscr{M}_{n+2}(\Gamma)$ denote the space of all K-rational automorphic forms of weight $n+2$ for Γ, i.e., the space of all $f \in O$ such that

$$f \circ g = f \cdot (\gamma x + \delta)^{n+2} \qquad \text{for all } g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma .$$

Theorem (p-adic Shimura isomorphism). *The map*

$$I_n : \mathscr{M}_{n+2}(\Gamma) \to H^0(\Gamma, C_{har}(\mathscr{BT}, \mathrm{Sym}^n))$$
$$f \mapsto (I \otimes \mathrm{id})(dx \otimes f(xe_1 + e_2)^n)$$

is an isomorphism.

Proof. Because of

$$\mathscr{M}_{n+2}(\Gamma) = O[-n-2]^\Gamma$$

the above exact sequence shows that the injectivity of the map I_n is equivalent to the vanishing of $(O[n]/P_n)^\Gamma$. This vanishing will be established later on as a consequence of more general considerations (see Corollary 12). For the moment we will take it as granted. It then remains to prove that the two K-vector spaces in the assertion have the same finite dimension. By passing to an appropriate normal subgroup of finite index we may assume that Γ is a free group of rank r. According to [44] we then have

$$\dim_K \mathscr{M}_{n+2}(\Gamma) = \begin{cases} r & \text{if } n = 0 \ , \\ (n+1)(r-1) & \text{if } n \geqq 1 \ . \end{cases}$$

On the other hand the space of harmonic 1-cochains on $\mathscr{BT}$ is the linear dual of the Steinberg representation:

$$C_{har}(\mathscr{BT}, K) = \mathrm{Hom}_K(V_\phi(K), K) \ .$$

Since Sym^n carries a natural $\mathrm{SL}_2(K)$-invariant nondegenerate bilinear form we obtain

$$\begin{aligned} H^0(\Gamma, C_{har}(\mathscr{BT}, \mathrm{Sym}^n)) &= H^0(\Gamma, \mathrm{Hom}_K(V_\phi(K), \mathrm{Sym}^n)) \\ &= H^0\Big(\Gamma, \mathrm{Hom}_K\Big(V_\phi(K) \underset{K}{\otimes} \mathrm{Sym}^n, K\Big)\Big) \\ &= \mathrm{Hom}_K\Big(H_0\Big(\Gamma, V_\phi(K) \underset{K}{\otimes} \mathrm{Sym}^n\Big), K\Big) \\ &= \mathrm{Hom}_K(H^1(\Gamma, \mathrm{Sym}^n), K) \ . \end{aligned}$$

Here the last identity is a special case of the Borel–Serre duality for discrete groups ([3] 6.2). Since Γ is free a standard Euler–Poincaré characteristic computation gives

$$\dim_K H^1(\Gamma, \mathrm{Sym}^n) - \dim_K H^0(\Gamma, \mathrm{Sym}^n) = (n+1)\cdot(r-1) \ .$$

But we have $H^0(\Gamma, \mathrm{Sym}^n) = 0$ for $n \geqq 1$.

In order to deal with the remaining problem of the vanishing of $(O[n]/P_n)^\Gamma$ we will analyze the p-adic principal series representations which were introduced by Morita. We fix a locally analytic character $\chi : K^\times \to K^\times$ and put

$$\begin{aligned} \mathscr{L}_\chi := \ & K\text{-vector space of all locally analytic} \\ & \text{functions } \Phi : K \times K \backslash (0,0) \to K \text{ such that} \\ & \Phi(z^\cdot, z^\cdot) = \chi(z)\Phi(.,.) \text{ for all } z \in K^\times \ . \end{aligned}$$

The group $\mathrm{SL}_2(K)$ acts on $\mathscr{L}_\chi$ by

$$g(\Phi)(x, y) := \Phi(\delta x - \beta y, \ -\gamma x + \alpha y) \qquad \text{for } g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

Furthermore $\mathscr{L}_\chi$ in a natural way is a topological vector space. For the definition which is a little bit technical we refer to [35]. The above $\mathrm{SL}_2(K)$-action is continuous. This is the p-adic principal series studied by Morita [35]. In order to analyze the continuous linear forms on $\mathscr{L}_\chi$ we "cover" it in the following way by K-Banach spaces. Let $m \geqq 1$ be minimal such that

$$\chi \text{ is analytic on } \{z \in K^\times : |z - 1| \leqq |\pi|^m\}.$$

The congruence subgroup

$$B_m := \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(o) : |\alpha - 1|, |\delta - 1| \leqq |\pi|^m, |\beta| \leqq |\pi|^{m-1}, |\gamma| \leqq |\pi| \right\}$$

acts on the disk

$$\Delta := \{z \in K : |z| \leqq |\pi|^{m-1}\}$$

by fractional linear transformations. It also acts isometrically on the Banach space $\mathscr{A}_\chi$ of K-analytic functions on Δ by

$$g(f)(z) := \chi(-\gamma z + \alpha) \cdot f\left(\frac{\delta z - \beta}{-\gamma z + \alpha}\right) \qquad \text{for } g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

(In case $K = \mathbb{Q}_p$ the B_m-representation $\mathscr{A}_\chi$ is studied in [43].) We then have the injective continuous B_m-equivariant map

$$\mathscr{A}_\chi \to \mathscr{L}_\chi$$

$$f \mapsto \Phi_f(x, y) := \begin{cases} \chi(y) f\left(\frac{x}{y}\right) & \text{if } |x| \leqq |\pi^{m-1} y|, \\ 0 & \text{otherwise} \end{cases}$$

(compare [35] p. 198) which for simplicity we will view as an inclusion. It is not too difficult to see (loc. cit.) that

$$\sum_{g \in \mathrm{SL}_2(K)} g\mathscr{A}_\chi = \mathscr{L}_\chi \tag{1}$$

holds true. Now let $\| \ \|$ denote the norm on the K-Banach space $\mathscr{A}_\chi$. For any continuous linear form $L : \mathscr{L}_\chi \to K$ we can define a kind of generalized operator norm

$$\|L\|(g) := \sup\left\{ \frac{|L(g(\Phi_f))|}{\|f\|} : f \in \mathscr{A}_\chi \setminus \{0\} \right\}.$$

Since B_m acts isometrically on $\mathscr{A}_\chi$ we obtain in this way a norm function

$$\|L\| : \mathrm{SL}_2(K)/B_m \to \mathbb{R}_{\geqq 0}$$

$$gB_m \mapsto \|L\|(g).$$

The property (1) implies that

$$L = 0 \qquad \text{if and only if } \|L\| = 0 \,. \tag{2}$$

It is also clear that the map $L \mapsto \|L\|$ is $\mathrm{SL}_2(K)$-equivariant in the obvious sense.

Lemma 11 *If $L : \mathscr{L}_\chi \to K$ is a Γ-invariant continuous linear form then the norm function $\|L\|$ is bounded.*

Proof. We only have to note that by the cocompactness of Γ the set $\Gamma\backslash \mathrm{SL}_2(K)/B_m$ is finite.

In order to describe the growth of $\|L\|$ we let $R \subseteq o$ be a fixed set of representatives for the cosets in $\pi^{m-1}o/\pi^{m+1}o$. Put

$$h_\beta := \begin{pmatrix} \pi & \pi^{-1}\beta \\ 0 & \pi^{-1} \end{pmatrix} = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \pi & 0 \\ 0 & \pi^{-1} \end{pmatrix} \in \mathrm{SL}_2(K)$$

and

$$\Delta_\beta := \{z \in K : |z - \beta| \leqq |\pi|^{m+1}\} \qquad \text{for } \beta \in R \,.$$

We have

$$\Delta = \overset{\bullet}{\bigcup_{\beta \in R}} \Delta_\beta \qquad \text{and} \qquad \Delta_\beta = h_\beta(\Delta)$$

(viewing h_β as a fractional linear transformation). For any $f \in \mathscr{A}_\chi$ we define

$$f_\beta := (f|\Delta_\beta) \circ h_\beta \in \mathscr{A}_\chi \,.$$

It is clear that

$$\|f_\beta\| \leqq \|f\| \,.$$

On the other hand a straightforward computation shows that

$$\chi(\pi) \cdot \Phi_f = \sum_{\beta \in R} h_\beta(\Phi_{f_\beta}) \,.$$

For any given continuous linear form L on $\mathscr{L}_\chi$ we therefore obtain

$$|\chi(\pi)| \cdot \frac{|L(\Phi_f)|}{\|f\|} \leqq \max_\beta \frac{|L(h_\beta(\Phi_{f_\beta}))|}{\|f_\beta\|} \,.$$

Since f was arbitrary this implies

$$|\chi(\pi)| \cdot \|L\|(1) \leqq \max_\beta \|L\|(h_\beta) \,.$$

Replacing L by $L(g\cdot)$ we even see that

$$|\chi(\pi)| \cdot \|L\|(g) \leqq \max_\beta \|L\|(gh_\beta) \tag{3}$$

holds true for any $g \in \mathrm{SL}_2(K)$.

Theorem 6 *If $|\chi(\pi)| > 1$ then there is no nonzero Γ-invariant continuous linear form on $\mathscr{L}_\chi$.*

Proof. By Lemma 11 any Γ-invariant continuous linear form L on $\mathscr{L}_\chi$ has a bounded norm function $\|L\|$. But the assumption $|\chi(\pi)| > 1$ together with (3) then implies that $\|L\| = 0$. Because of (2) the linear form L itself has to be zero.

Corollary 12 *For $n \geqq 0$ we have $(O[n]/P_n)^\Gamma = 0$.*

Proof. The character $z \mapsto z^{-n-2}$ fulfills the assumption of Theorem 6. On the other hand it is shown in [34] p. 292 and [36] p. 394 that $O[n]/P_n$ can be identified $\mathrm{SL}_2(K)$-equivariantly with the topological dual of $\mathscr{L}_{z^{-n-2}}$.

Corollary 13. *There are no nonzero automorphic forms of weight 1 for Γ.*

Proof. In loc. cit. Morita also shows that $O[-1]$ is $\mathrm{SL}_2(K)$-equivariantly isomorphic to the topological dual of $\mathscr{L}_{z^{-1}}$.

By working with the tree of Γ instead of $\mathscr{BT}$ the p-adic Shimura isomorphism can be obtained in an analogous way for any finitely generated discrete subgroup in $\mathrm{SL}_2(K)$ with an infinite limit set. A much more direct proof is given in [46]. It was pointed out by de Shalit that contrary to what is claimed in [44] in general not every cohomology class in $H^1(\Gamma, \mathrm{Sym}^n)$ contains a harmonic cochain; nevertheless this holds true if $n = 0$ and as is shown in [46] for any $n \geqq 0$ if Γ is arithmetic. Still another proof is to be found in [49].

Remark. Assume that Γ acts without fixed points on $\Omega^{(2)}$. It follows from $(O/K)^\Gamma = (dO)^\Gamma = 0$ that the filtration $F^\bullet H^1_{DR}(X_\Gamma)$ which we introduced earlier is opposite to the de Rham filtration. This confirms in the case of $\Omega^{(2)}$ our general expectation which we explained after Theorem 5. In particular we have the natural decomposition

$$H^1_{DR}(X_\Gamma) = H^1(\Gamma, K) \oplus H^0(X_\Gamma, \Omega^1) .$$

Using the isomorphism $I_0 : H^0(X_\Gamma, \Omega^1) = M_2(\Gamma) \xrightarrow{\cong} H^0(\Gamma, C_{har}(\mathscr{BT}, K)) = H^1(\Gamma, K)$ we can rewrite this as a natural isomorphism

$$H^1_{DR}(X_\Gamma) = H^1(\Gamma, K) \oplus I_0^{-1}(H^1(\Gamma, K)) .$$

6 Resolutions of $\mathrm{GL}_{d+1}(K)$-modules

In a rather general context we study in the following the homology of certain coefficient systems on a simplicial complex. This is then applied to construct the resolutions of our $\mathrm{GL}_{d+1}(K)$-modules $V_I(\mathbb{Z})$ which we needed in the last Paragraph.

Let $\mathscr{K}$ be a finite-dimensional simplicial complex, say, of dimension d. In order to make notations simpler we equip $\mathscr{K}$ with a fixed orientation and we let denote $[\tau : \sigma]$ the corresponding incidence numbers; nothing will really depend on this choice. A coefficient system $\underline{A}$ on $\mathscr{K}$ consists of

— abelian groups A_σ for each simplex σ of $\mathscr{K}$, and

— homomorphisms $r_\sigma^{\sigma'} : A_{\sigma'} \to A_\sigma$ for each pair $\sigma \subseteq \sigma'$ of simplices of $\mathscr{K}$ such that $r_\sigma^\sigma = \mathrm{id}$ and $r_\sigma^{\sigma''} = r_\sigma^{\sigma'} \circ r_{\sigma'}^{\sigma''}$ whenever $\sigma \subseteq \sigma' \subseteq \sigma''$.

Any such coefficient system gives rise to a homological complex: Let $\mathscr{K}_q$ denote the set of all q-simplices of $\mathscr{K}$. We have the boundary map

$$\partial : \bigoplus_{\tau \in \mathscr{K}_{q+1}} A_\tau \to \bigoplus_{\sigma \in \mathscr{K}_q} A_\sigma$$

$$(a_\tau)_\tau \mapsto \left(\sum_{\tau \supseteq \sigma} [\tau : \sigma] \cdot r_\sigma^\tau(a_\tau) \right)_\sigma .$$

A standard computation then shows that

$$\bigoplus_{\tau\in\mathscr{K}_d} A_\tau \xrightarrow{\partial} \ldots \xrightarrow{\partial} \bigoplus_{\sigma\in\mathscr{K}_0} A_\sigma$$

is a complex. For example, if $\underline{A}$ is the constant coefficient system given by $\mathbb{Z}$ we obtain in this way the usual chain complex of $\mathscr{K}$.

We are interested in coefficient systems of the following form: Let T be a fixed profinite set and suppose that for each 0-simplex $\sigma\in\mathscr{K}_0$ there is given a continuous surjection $T\to T_\sigma$ onto a finite set T_σ. For an arbitrary simplex $\tau=\{\sigma_0,\ldots,\sigma_q\}\in\mathscr{K}_q$ we put

$$T_\tau := T_{\sigma_0}\coprod_T \ldots \coprod_T T_{\sigma_q}.$$

For any pair $\sigma\subseteq\tau$ of simplices of $\mathscr{K}$ we then have a commutative diagram of continuous surjections

$$\begin{array}{ccc} T & \longrightarrow & T_\sigma \\ & \searrow \quad \swarrow & \\ & T_\tau & \end{array}.$$

It is clear that the groups

$$C(T_\sigma,\mathbb{Z}) := \text{all functions } T_\sigma\to\mathbb{Z}$$

form a coefficient system on $\mathscr{K}$. In fact all the groups $C(T_\sigma,\mathbb{Z})$ can be viewed as subgroups of $C^\infty(T,\mathbb{Z})$ so that the transition maps of this coefficient system become inclusion maps. Furthermore the corresponding complex is augmented in a natural way

$$\bigoplus_{\tau\in\mathscr{K}_d} C(T_\tau,\mathbb{Z}) \xrightarrow{\partial} \ldots \xrightarrow{\partial} \bigoplus_{\sigma\in\mathscr{K}_0} C(T_\sigma,\mathbb{Z}) \to C^\infty(T,\mathbb{Z})$$
$$(\varphi_\sigma)_\sigma \mapsto \sum_\sigma \varphi_\sigma .$$

We want to establish a sufficient condition under which this complex is an acyclic resolution of $C^\infty(T,\mathbb{Z})$.

First we observe that for any simplex σ we have the simplicial profinite set

$$T_\bullet^{(\sigma)} : \ldots \;\substack{\longrightarrow\\ \longrightarrow\\ \longrightarrow\\ \longrightarrow}\; T\times_{T_\sigma} T\times_{T_\sigma} T \;\substack{\longrightarrow\\ \longrightarrow\\ \longrightarrow}\; T\times_{T_\sigma} T \;\substack{\longrightarrow\\ \longrightarrow}\; T$$

together with a continuous map

$$T_\bullet^{(\sigma)}\to T_\sigma$$

to the constant simplicial (pro)finite set T_σ. We immediately note that the induced cohomological complex

$$0\to C(T_\sigma,\mathbb{Z})\to C^\infty(T_0^{(\sigma)},\mathbb{Z})\to C^\infty(T_1^{(\sigma)},\mathbb{Z})\to\ldots$$

is exact. This follows from §3 Lemma 4 since the fiber of $T_\bullet^{(\sigma)}\to T_\sigma$ in a fixed $t\in T_\sigma$ is the simplicial profinite set

$$\ldots \;\substack{\longrightarrow\\ \longrightarrow\\ \longrightarrow}\; T_t\times T_t\times T_t \;\substack{\longrightarrow\\ \longrightarrow\\ \longrightarrow}\; T_t\times T_t \;\substack{\longrightarrow\\ \longrightarrow}\; T_t$$

where T_t is the fiber of $T\to T_\sigma$ in t. A simplicial profinite set of this form namely is cohomologically trivial as we have recalled already in the proof of §3 Lemma 3.

Each surjection $T_\sigma \to T_\tau$ for $\sigma \subseteq \tau$ extends in an obvious way to a map of augmented simplicial profinite sets

$$\begin{array}{ccc} T_\bullet^{(\sigma)} & \longrightarrow & T_\sigma \\ \downarrow & & \downarrow \\ T_\bullet^{(\tau)} & \longrightarrow & T_\tau \ ; \end{array}$$

similarly each surjection $T \to T_\sigma$ extends to a map $T_\bullet \to T_\bullet^{(\sigma)}$ of the constant simplicial profinite set $T_\bullet$ into $T_\bullet^{(\sigma)}$ such that

$$\begin{array}{ccc} & T_\bullet^{(\sigma)} & \\ \nearrow & & \searrow \\ T_\bullet & \longrightarrow & T_\sigma \end{array}$$

is commutative. By passing to functions we have corresponding commutative diagrams between coefficient systems on $\mathscr{K}$. All this can be expressed by saying that we have the biaugmented double complex

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 \to & \bigoplus_{\tau \in \mathscr{K}_d} C(T_\tau, \mathbb{Z}) & \to \ldots \to & \bigoplus_{\sigma \in \mathscr{K}_0} C(T_\sigma, \mathbb{Z}) & \dashrightarrow & C^\infty(T, \mathbb{Z}) & \to 0 & & \\
 & \downarrow & & \downarrow & & \| & & & \\
0 \to & \bigoplus_{\tau \in \mathscr{K}_d} C^\infty(T, \mathbb{Z}) & \to \ldots \to & \bigoplus_{\sigma \in \mathscr{K}_0} C^\infty(T, \mathbb{Z}) & \dashrightarrow & C^\infty(T, \mathbb{Z}) & \to 0 & & \\
 & \downarrow & & \downarrow & & \downarrow 0 & & & \\
0 \to & \bigoplus_{\tau \in \mathscr{K}_d} C^\infty(T \times T, \mathbb{Z})_{T_\tau} & \to \ldots \to & \bigoplus_{\sigma \in \mathscr{K}_0} C^\infty(T \times T, \mathbb{Z})_{T_\sigma} & \dashrightarrow & C^\infty(T \times T, \mathbb{Z})_{T} & \to 0 & & \\
 & \downarrow & & \downarrow & & \| & & & \\
0 \to & \bigoplus_{\tau \in \mathscr{K}_d} C^\infty(T \times T \times T, \mathbb{Z})_{T_\tau T_\tau} & \to \ldots \to & \bigoplus_{\sigma \in \mathscr{K}_0} C^\infty(T \times T \times T, \mathbb{Z})_{T_\sigma T_\sigma} & \dashrightarrow & C^\infty(T \times T \times T, \mathbb{Z})_{T T} & \to 0 & & \\
 & \downarrow & & \downarrow & & \downarrow 0 & & & \\
 & \vdots & & \vdots & & \vdots & , & &
\end{array}$$

resp.

$$\begin{array}{ccc} \bigoplus_{\sigma \in \mathscr{K}_\bullet} C(T_\sigma, \mathbb{Z}) & \to & C^\infty(T, \mathbb{Z}) \\ \downarrow & & \downarrow \\ \bigoplus_{\sigma \in \mathscr{K}_\bullet} C^\infty(T_\bullet^{(\sigma)}, \mathbb{Z}) & \to & C^\infty(T_\bullet, \mathbb{Z}) \ . \end{array}$$

As already noted the columns are exact, resp. the perpendicular arrows in the

second diagram are quasi-isomorphisms. Therefore let us fix some $m \geqq 0$ and consider the line

$$0 \to \bigoplus_{\tau \in \mathscr{K}_d} C^\infty(T_m^{(\tau)}, \mathbb{Z}) \to \ldots \to \bigoplus_{\sigma \in \mathscr{K}_0} C^\infty(T_m^{(\sigma)}, \mathbb{Z}) \to C^\infty(T, \mathbb{Z}) \to 0 .$$

In order to sheafify this sequence we put $T^{m+1} := T \times \ldots \times T$ and let denote $j^{(\sigma)}$: $T_m^{(\sigma)} \to T^{m+1}$ the inclusion map and $\delta : T \to T^{m+1}$ the diagonal map. We then have the obvious complex of sheaves

$$(*) \qquad 0 \to \bigoplus_{\tau \in \mathscr{K}_d} j_*^{(\tau)} \mathbb{Z} \to \ldots \to \bigoplus_{\sigma \in \mathscr{K}_0} j_*^{(\sigma)} \mathbb{Z} \to \delta_* \mathbb{Z} \to 0$$

on T^{m+1} of which our line is the complex of global sections (the global section functor on a compact space commutes with arbitrary direct sums of sheaves!). Since the sheaves in $(*)$ are soft it is sufficient for the exactness of our line that $(*)$ is an exact complex of sheaves (compare the argument in the proof of §3 Lemma 4); the latter can be checked stalkwise. For $\underline{t} = (t_0, \ldots, t_m) \in T^{m+1}$ we have

$$(j_*^{(\sigma)} \mathbb{Z})_{\underline{t}} = \begin{cases} \mathbb{Z} & \text{if } \underline{t} \in T_m^{(\sigma)} , \\ 0 & \text{otherwise} . \end{cases}$$

This suggests the following definition:

$\mathscr{K}^{(t_0, \ldots, t_m)} :=$ the subcomplex of $\mathscr{K}$ consisting of all simplices σ such that $t_0, \ldots, t_m$ are not all identified under the map $T \to T_\sigma$.

Equivalently, σ is a simplex of $\mathscr{K}^{(t_0, \ldots, t_m)}$ if and only if $\underline{t} \notin T_m^{(\sigma)}$ if and only if $(j_*^{(\sigma)} \mathbb{Z})_{\underline{t}} = 0$. We make now the additional assumption that the natural map $T \to \prod_{\sigma \in \mathscr{K}_0} T_\sigma$ is injective. Then $\mathscr{K}^{(t_0, \ldots, t_m)}$ is empty if and only if $t_0 = \ldots = t_m$ if and only if $(\delta_* \mathbb{Z})_{\underline{t}} \neq 0$. We obtain that the complex of stalks of $(*)$ in $\underline{t}$ is the relative augmented chain complex of the pair $\mathscr{K}^{(t_0, \ldots, t_m)} \subseteq \mathscr{K}$. Therefore if $\mathscr{K}$ and all its non-empty subcomplexes $\mathscr{K}^{(t_0, \ldots, t_m)}$ are contractible the complexes $(*)$ are exact. Going back to our biaugmented double complex we then see that all its lines apart possibly from the first one are exact. But since the columns are exact anyway the first line has to be exact, too. We sum up this discussion in the following criterion.

Proposition 1 *Assume that the map $T \to \prod_{\sigma \in \mathscr{K}_0} T_\sigma$ is injective and that $\mathscr{K}$ and all its non-empty subcomplexes $\mathscr{K}^{(t_0, \ldots, t_m)}$ for $t_i \in T$ are contractible. Then the complex*

$$0 \to \bigoplus_{\tau \in \mathscr{K}_d} C(T_\tau, \mathbb{Z}) \to \ldots \to \bigoplus_{\sigma \in \mathscr{K}_0} C(T_\sigma, \mathbb{Z}) \to C^\infty(T, \mathbb{Z}) \to 0$$

is exact.

We want to apply this result in the concrete situation where the underlying simplicial complex is the Bruhat–Tits building $\mathscr{B}\mathscr{T}$ and where the given profinite set is $T := G/P_I$ for some fixed subset $I \subseteq \Delta = \{1, \ldots, d\}$. Let σ be a simplex in $\mathscr{B}\mathscr{T}$. Then

$$B_\sigma := \{ g \in G : g\sigma = \sigma \text{ and } \det g \in o^\times \}$$

is a compact open subgroup in G; it has a unique maximal normal pro-p-subgroup U_σ which itself is compact open. The set

$$T_\sigma := U_\sigma \backslash T$$

is finite. We have the continuous projection $T \to T_\sigma$ and

$$C(T_\sigma, \mathbb{Z}) = C^\infty(G/P_I, \mathbb{Z})^{U_\sigma} .$$

It is also obvious that $U_\sigma \subseteq U_\tau$ if $\sigma \subseteq \tau$.

Lemma 2 *If $\sigma_0, \ldots, \sigma_q$ are the vertices of the simplex σ in $\mathscr{BT}$ then U_σ is generated by $U_{\sigma_0} \cup \ldots \cup U_{\sigma_q}$.*

Proof. Let τ be that maximal simplex in $\mathscr{BT}$ for which $B_\tau = B$ is the standard Iwahori subgroup considered in §4. If $\tau_0, \ldots, \tau_d$ are the vertices of τ appropriately enumerated we have

$$B_{\tau_j} = y_j \mathrm{GL}_{d+1}(o) y_j^{-1} .$$

Using the notations introduced in §4 we observe that

$$B_{\tau_0} \cap B_{\tau_j} = B_{\Delta \setminus \{j\}} \quad \text{and} \tag{1}$$

$$B \cap U_{\Delta \setminus \{j\}} \subseteq y_j U_{\tau_0} y_j^{-1} = U_{\tau_j} . \tag{2}$$

By conjugation we may assume that $\sigma \subseteq \tau$ and $\sigma_0 = \tau_0, \sigma_1 = \tau_{j_1}, \ldots, \sigma_q = \tau_{j_q}$. From (1) we then deduce that

$$B_\sigma = B_J \qquad \text{with } J := \Delta \setminus \{j_1, \ldots, j_q\} .$$

Furthermore one easily checks that

$$U_\sigma = U_{\sigma_0} \cdot (B \cap U_J) .$$

By (2) we therefore are reduced to prove that

$$B \cap U_J \subseteq U_{\sigma_0} \cdot (B \cap U_{\Delta \setminus \{j_1\}}) \cdot \ldots \cdot (B \cap U_{\Delta \setminus \{j_q\}}) .$$

Reducing mod π this amounts to the statement that the unipotent radical of a fixed parabolic subgroup is equal to the product of the unipotent radicals of the maximal proper parabolic subgroups containing the given one. This is a well-known fact (compare [4] 3.2) which for GL_{d+1} can be seen easily using elementary matrix transformations.

The Lemma says that

$$T_\sigma = T_{\sigma_0} \coprod_T \ldots \coprod_T T_{\sigma_q}$$

so that we are in the situation which was discussed above.

Lemma 3 *The map $T \to \prod_{\sigma \in \mathscr{BT}_0} T_\sigma$ is injective.*

Proof. Since points in the same fiber of the map in question cannot be separated by functions in the image of the homomorphism

$$\bigoplus_{\sigma \in \mathscr{BT}_0} C(T_\sigma, \mathbb{Z}) \to C^\infty(T, \mathbb{Z})$$

it suffices to show that the latter is surjective. According to §4 Corollary 9' the G-module $C^\infty(T, \mathbb{Z})$ is generated by the characteristic function $\chi_{B_I P_I}$. If $\sigma_0 \in \mathscr{BT}_0$ is a vertex such that $U_{\sigma_0} \subseteq B_I$ then $\chi_{B_I P_I} \in C^\infty(T, \mathbb{Z})^{U_{\sigma_0}} = C(T_{\sigma_0}, \mathbb{Z})$. Therefore the above G-module homomorphism contains a generator of the right hand side in its image and consequently has to be surjective.

The Bruhat–Tits building is contractible. Therefore it remains to deal with the last assumption in Proposition 1 that all the non-empty subcomplexes $\mathscr{B}\mathscr{T}^{(t_0,\ldots,t_m)}$ of $\mathscr{B}\mathscr{T}$ are contractible. This requires a more elaborate argument. We will first describe $\mathscr{B}\mathscr{T}^{(t_0,\ldots,t_m)}$ as the union of an explicitly given family of apartments in $\mathscr{B}\mathscr{T}$. As a convenient technical tool we interpret the sets T_σ in terms of lattices. Recall that a simplex σ in $\mathscr{B}\mathscr{T}$ is a family $\sigma = \{[L_0], \ldots, [L_q]\}$ of similarity classes of o-lattices in K^{d+1} such that

$$L_0 \subsetneqq L_1 \subsetneqq \ldots \subsetneqq L_q \subsetneqq \pi^{-1} L_0 =: L_{q+1} .$$

It is clear that U_σ is the subgroup of all elements in G which fix each L_j and induce the identity on each L_{j+1}/L_j. We now introduce

$$\mathscr{T}(L_0, \ldots, L_q) := \text{set of all sequences } (X_0, \ldots, X_q) \text{ of } o\text{-lattices such that } L_j \subseteq X_j \subseteq L_{j+1} \text{ for all } 0 \leqq j \leqq q ;$$

it is partially ordered by

$$(X_0, \ldots, X_q) \leqq (Y_0, \ldots, Y_q) \quad \text{if } X_j \subseteq Y_j \quad \text{for all } 0 \leqq j \leqq q .$$

Furthermore we call

$$\dim(X_0, \ldots, X_q) := \sum_{j=0}^{q} \dim_{o/\pi o} X_j/L_j$$

the dimension of the element $(X_0, \ldots, X_q) \in \mathscr{T}(L_0, \ldots, L_q)$. Assume that $\Delta \backslash I = \{i_0 < i_1 < \ldots < i_r\}$. A flag $\xi_0 \leqq \ldots \leqq \xi_r$ in $\mathscr{T}(L_0, \ldots, L_q)$ is called of type I if $\dim \xi_v = i_v$ for all $0 \leqq v \leqq r$. It is rather obvious that the map

$$\mathscr{T}(L_0) \to \mathscr{T}(L_0, \ldots, L_q)$$

$$X \mapsto (L_0 + X \cap L_1, L_1 + X \cap L_2, \ldots, L_q + X \cap L_{q+1})$$

is order and dimension preserving. Also the group U_σ acts on $\mathscr{T}(L_0)$ in an order and dimension preserving way.

Proposition 4 *The above map induces a bijection*

$$U_\sigma \backslash \{\textit{flags of type } I \textit{ in } \mathscr{T}(L_0)\} \xrightarrow{\sim} \{\textit{flags of type } I \textit{ in } \mathscr{T}(L_0, \ldots, L_q)\} .$$

Proof. To check surjectivity is an easy exercise. The injectivity amounts to the following statement in linear algebra: Let V be a finite dimensional vector space over some field equipped with a flag $0 = V_0 \subsetneqq V_1 \subsetneqq \ldots \subsetneqq V_q \subsetneqq V_{q+1} = V$ of subspaces; furthermore let $U_0 \subseteq \ldots \subseteq U_r \subseteq V$ and $U'_0 \subseteq \ldots \subseteq U'_r \subseteq V$ be two other flags of subspaces such that $V_j + U_v \cap V_{j+1} = V_j + U'_v \cap V_{j+1}$ for all $0 \leqq j \leqq q$ and $0 \leqq v \leqq r$. Then there exists a $g \in \mathrm{GL}(V)$ which fixes each V_j, induces the identity on each V_{j+1}/V_j and fulfills $g(U_v) = U'_v$ for all $0 \leqq v \leqq r$. By induction with respect to q this is immediately reduced to another slightly more general statement: Let $V_1 \subseteq V$ be a subspace and let $U_0 \subseteq \ldots \subseteq U_r \subseteq V$ and $U'_0 \subseteq \ldots \subseteq U'_r \subseteq V$ be two flags of subspaces such that $U_v \cap V_1 = U'_v \cap V_1$ and $\bar{g}(V_1 + U_v/V_1) = V_1 + U'_v/V_1$ for all $0 \leqq v \leqq r$ and some fixed element $\bar{g} \in \mathrm{GL}(V/V_1)$. Then there exists a $g \in \mathrm{GL}(V)$ which induces the identity on V_1 and $\bar{g}$ on V/V_1 and fulfills $g(U_v) = U'_v$ for all $0 \leqq v \leqq r$. The proof of this is left to the reader.

Similarly as before a flag $W_0 \subseteq \ldots \subseteq W_r$ of K-subspaces in K^{d+1} is called of type I if $\dim_K W_v = i_v$ for all $0 \leqq v \leqq r$. We have the bijection

$$T = G/P_I \xrightarrow{\sim} \text{set of all flags of type } I \text{ in } K^{d+1}$$

$$gP_I \mapsto g(\tau_I^*)$$

where

$$\tau_I^* := \left(\sum_{i=1}^{i_0} Ke_i \subseteq \sum_{i=1}^{i_1} Ke_i \subseteq \ldots \subseteq \sum_{i=1}^{i_r} Ke_i \right).$$

Corollary 5 *The map*

$$U_\sigma \backslash T \xrightarrow{\sim} \text{set of all flags of type } I \text{ in } \mathscr{T}(L_0, \ldots, L_q)$$

$$(W_0 \subseteq \ldots \subseteq W_r) \bmod U_\sigma \mapsto (\xi_0 \leqq \ldots \leqq \xi_r)$$

with $\xi_v := (L_0 + W_v \cap L_1, \ldots, L_q + W_v \cap L_{q+1})$ *is a bijection.*

Proof. We have

$$L_j + (L_0 + W_v \cap \pi^{-1} L_0) \cap L_{j+1} = L_j + W_v \cap L_{j+1} .$$

Therefore by Proposition 4 we are reduced to treat the case $q = 0$. Surjectivity is easy to check again. For the injectivity let $t = (W_0 \subseteq \ldots \subseteq W_r)$ and $t' = (W_0' \subseteq \ldots \subseteq W_r')$ be two flags in T such that $L_0 + W_v \cap \pi^{-1} L_0 = L_0 + W_v' \cap \pi^{-1} L_0$ for all $0 \leqq v \leqq r$. Let $P_t \subseteq G$ be that parabolic subgroup which fixes t. Since B_σ is a maximal compact subgroup we find by the Iwasawa decomposition $G = B_\sigma P_t$ a $g \in B_\sigma$ such that $g(t) = t'$. On the other hand let $\underline{P}_{\bar{t}} \subseteq \mathrm{GL}(\pi^{-1} L_0 / L_0) = B_\sigma / U_{\bar{\sigma}}$ be that parabolic subgroup which fixes $\bar{t} := (\bar{W}_0 \subseteq \ldots \subseteq \bar{W}_r)$ with $\bar{W}_v := L_0 + W_v \cap \pi^{-1} L_0 / L_0$. Our assumption implies $g \bmod U_\sigma \in \underline{P}_{\bar{t}}$. Since any element in $\underline{P}_{\bar{t}}$ can be lifted to $B_\sigma \cap P_t$ we find a $h \in B_\sigma \cap P_t$ such that $g \bmod U_\sigma = h \bmod U_\sigma$. We then have $gh^{-1} \in U_\sigma$ with $gh^{-1}(t) = t'$.

In the following we always will make the above identifications and view elements in T, resp. T_σ, as flags of subspaces, resp. lattices. As already mentioned our aim is to describe the subcomplexes $\mathscr{B}\mathscr{T}^{(t_0, \ldots, t_m)}$ in terms of apartments in $\mathscr{B}\mathscr{T}$. We therefore have to recall briefly this notion. Any basis $v_1, \ldots, v_{d+1}$ of the vector space K^{d+1} determines an apartment $\mathscr{A}$ in $\mathscr{B}\mathscr{T}$ which is the full subcomplex generated by all vertices of the form

$$[\pi^{a_1} o v_1 + \ldots + \pi^{a_{d+1}} o v_{d+1}] \qquad \text{with } a_1, \ldots, a_{d+1} \in \mathbb{Z} .$$

Its topological realization $|\mathscr{A}|$ is a d-dimensional affine space. The same basis also determines an apartment $\partial\mathscr{A}$ in the topological Tits building $\mathscr{T}_\bullet^{(d)}$ which is the full (and finite) subcomplex generated by all vertices of the form

$$Kv_{\alpha_1} + \ldots + Kv_{\alpha_t} \qquad \text{with } \phi \neq \{\alpha_1, \ldots, \alpha_t\} \subsetneq \{1, \ldots, d+1\} .$$

Its topological realization $|\partial\mathscr{A}|$ is a $(d-1)$-sphere. In the Borel–Serre compactification of $|\mathscr{B}\mathscr{T}|$ (see [3]) $|\partial\mathscr{A}|$ is the boundary of $|\mathscr{A}|$. Since we don't need this fact in the moment we by definition call $\partial\mathscr{A}$ the boundary of $\mathscr{A}$.

Lemma 6 *Let* $W', W'' \in T$ *where* $\# \Delta \backslash I = 1$. *Any apartment* $\mathscr{A}$ *in* $\mathscr{B}\mathscr{T}$ *such that* W' *and some* 1*-dimensional subspace* $W \subseteq W'' \backslash W'$ *are vertices in the boundary* $\partial\mathscr{A}$ *is contained in* $\mathscr{B}\mathscr{T}^{(W', W'')}$.

Proof. Fix W and $\mathscr{A}$ as in the statement and let $\sigma = \{[L_0], \ldots, [L_q]\}$ with $L_0 \subseteq L_1 \subseteq \ldots \subseteq L_q \subseteq \pi^{-1}L_0 = L_{q+1}$ be a simplex in $\mathscr{A}$. We then find a basis $v_1, \ldots, v_{d+1}$ of K^{d+1} such that

$$W' = Kv_1 + \ldots + Kv_{k-1} \qquad \text{for some } 2 \leqq k \leqq d+1\,,$$

$$W = Kv_k, \qquad \text{and} \qquad L_0 = ov_1 + \ldots + ov_{d+1}\,.$$

In order to prove the assertion it suffices to find an $0 \leqq j_0 \leqq q$ such that

$$L_{j_0} + W \cap L_{j_0+1} \nsubseteq L_{j_0} + W' \cap L_{j_0+1}\,.$$

Put

$$j_0 := \min\{0 \leqq j \leqq q : \pi^{-1}v_k \in L_{j+1}\} - 1\,.$$

Since $\pi^{-1}v_k \notin L_0$ and $\pi^{-1}v_k \in L_{q+1}$ we have $0 \leqq j_0 \leqq q$. By definition $\pi^{-1}v_k$ is contained in $W \cap L_{j_0+1}$. On the other hand it is easy to check that $\pi^{-1}v_k \notin L_{j_0} + W' \cap L_{j_0+1}$.

Lemma 7 *Let $t = (W_0 \subseteq \ldots \subseteq W_r) \in T$ and let $W' \subseteq K^{d+1}$ be a subspace with $\dim_K W' = \dim_K W_r$. Then $\mathscr{BT}^{(W_r, W')}$ is contained in the union of all apartments in $\mathscr{BT}$ which have t and some 1-dimensional subspace $W \subseteq W' \setminus W_r$ as simplices in their boundary.*

Proof. Let $\sigma = \{[L_0], \ldots, [L_q]\}$ with $L_0 \subseteq L_1 \subseteq \ldots \subseteq L_q \subseteq \pi^{-1}L_0 = L_{q+1}$ be a simplex in $\mathscr{BT}^{(W_r, W')}$. We then have

$$(L_0 + W_r \cap L_1, \ldots, L_q + W_r \cap L_{q+1}) \neq (L_0 + W' \cap L_1, \ldots, L_q + W' \cap L_{q+1})$$

but

$$\dim(L_0 + W_r \cap L_1, \ldots, L_q + W_r \cap L_{q+1}) = \dim_K W_r = \dim_K W'$$
$$= \dim(L_0 + W' \cap L_1, \ldots, L_q + W' \cap L_{q+1})\,;$$

therefore there has to be a $0 \leqq j_0 \leqq q$ such that

$$L_{j_0} + W' \cap L_{j_0+1} \nsubseteq L_{j_0} + W_r \cap L_{j_0+1}\,.$$

We now consider the flag

$$L_0 \subseteq L_0 + W_0 \cap L_1 \subseteq \ldots \subseteq L_0 + W_r \cap L_1 \subseteq$$

$$L_1 \subseteq \ldots$$

$$\vdots$$

$$L_{j_0} \subseteq L_{j_0} + W_0 \cap L_{j_0+1} \subseteq \ldots \subseteq L_{j_0} + W_r \cap L_{j_0+1} \subsetneq L_{j_0} + W_r \cap L_{j_0+1} + W'$$

$$\cap L_{j_0+1} \subseteq L_{j_0+1} \subseteq \ldots$$

$$\vdots$$

$$L_q \subseteq L_q + W_0 \cap L_{q+1} \subseteq \ldots \subseteq L_q + W_r \cap L_{q+1} \subseteq$$

$$L_{q+1}$$

viewed as a flag of subspaces in $\pi^{-1}L_0/L_0$. We fix a $o/\pi o$-basis $\bar{v}_1, \ldots, \bar{v}_{d+1}$ of $\pi^{-1}L_0/L_0$ such that

— any subspace in the above flag is generated by some of the $\bar{v}_\alpha$'s, and
— there is an $1 \leqq \alpha_0 \leqq d+1$ with

$$\bar{v}_{\alpha_0} \in (L_{j_0} + W' \cap L_{j_0+1}/L_0) \backslash (L_{j_0} + W_r \cap L_{j_0+1}/L_0) .$$

Although some of the inclusions in the above flag might be equalities for any fixed $\bar{v}_\alpha$ precisely one of the following cases holds true:

$$\bar{v}_\alpha \in L_j + W_v \cap L_{j+1}/L_0 \text{ and } \notin L_j + W_{v-1} \cap L_{j+1}/L_0 \text{ for some } 0 \leqq j \leqq q \text{ and } 1 \leqq v \leqq r, \text{ or} \tag{1}$$

$$\bar{v}_\alpha \in L_j + W_0 \cap L_{j+1}/L_0 \text{ and } \notin L_j/L_0 \text{ for some } 0 \leqq j \leqq q, \text{ or} \tag{2}$$

$$\bar{v}_\alpha \in L_{j+1}/L_0 \text{ and } \notin L_j + W_r \cap L_{j+1}/L_0 \text{ for some } 0 \leqq j \leqq q \text{ and } \alpha \neq \alpha_0, \text{ or} \tag{3}$$

$$\bar{v}_\alpha = \bar{v}_{\alpha_0} . \tag{4}$$

In cases (1) and (2) we find

$$v_\alpha \in W_v \cap L_{j+1} \qquad \text{such that } v_\alpha \equiv \bar{v}_\alpha \bmod L_j ;$$

in case (3) we find

$$v_\alpha \in L_{j+1} \qquad \text{such that } v_\alpha \equiv \bar{v}_\alpha \bmod L_0 ;$$

finally in case (4) we find

$$v_{\alpha_0} \in W' \cap L_{j_0+1} \qquad \text{such that } v_{\alpha_0} \equiv \bar{v}_{\alpha_0} \bmod L_{j_0} .$$

The $v_1, \ldots, v_{d+1}$ constructed in this way form a K-basis of K^{d+1} such that any of the subspaces $W_0, \ldots, W_r$ is generated by some of the v_α's. Put $W := Kv_{\alpha_0}$; we then have $W \subseteq W' \backslash W_r$. Therefore the apartment $\mathscr{A}$ in $\mathscr{BT}$ determined by this basis has t and W as simplices in its boundary. It also contains σ since $L_0 = \pi o v_1 + \ldots + \pi o v_{d+1}$ and any L_j/L_0 is generated by some of the $v_\alpha \bmod L_0$.

Proposition 8 *Let $t_0, \ldots, t_m \in T$ such that $t_\mu = (W_0^{(\mu)} \subseteq \ldots \subseteq W_r^{(\mu)})$. Then $\mathscr{BT}^{(t_0, \ldots, t_m)}$ is equal to the union of all apartments in $\mathscr{BT}$ which have, for some $0 \leqq v \leqq r$ and some $1 \leqq \mu \leqq m$, the flag $(W_0^{(0)} \subseteq \ldots \subseteq W_v^{(0)})$ and some 1-dimensional subspace $W \subseteq W_v^{(\mu)} \backslash W_v^{(0)}$ as simplices in their boundary.*

Proof. This is an immediate consequence of the previous two Lemmata once we observe that

$$\begin{aligned} \mathscr{BT}^{(t_0, \ldots, t_m)} &= \bigcup_{1 \leqq \mu \leqq m} \mathscr{BT}^{(t_0, t_\mu)} \\ &= \bigcup_{1 \leqq \mu \leqq m} \bigcup_{0 \leqq v \leqq r} \mathscr{BT}^{(W_v^{(0)}, W_v^{(\mu)})} \end{aligned}$$

holds true.

All apartments which occur in the statement of Proposition 8 have the common vertex $W_0^{(0)}$ in their boundary. This enables us to explore the geodesic action of $W_0^{(0)}$ on $\mathscr{BT}$ for our purpose. Any proper subspace $0 \subsetneqq V \subsetneqq K^{d+1}$ defines a simplicial self-map

$$\begin{aligned} \gamma_V : \mathscr{BT} &\to \mathscr{BT} \\ [L] &\mapsto [\pi^{-1} L \cap V + L] \end{aligned}$$

called the geodesic action. In order to see its simplicial nature assume that $L_0 \subseteq L_1 \subseteq \pi^{-1}L_0$ are o-lattices such that $[\pi^{-1}L_0 \cap V + L_0] = [\pi^{-1}L_1 \cap V + L_1]$. Because of $\pi^{-1}L_0 \cap V + L_0 \subseteq \pi^{-1}L_1 \cap V + L_1 \subseteq \pi^{-1}(\pi^{-1}L_0 \cap V + L_0)$ this means that

$$\pi^{-1}L_1 \cap V + L_1 = \pi^{-1}L_0 \cap V + L_0 \quad \text{or} \quad = \pi^{-1}(\pi^{-1}L_0 \cap V + L_0)\,.$$

In the first case we obtain that

$$\begin{aligned}\dim_K V &= \dim_{o/\pi o}\pi^{-1}L_0 \cap V + L_0/L_0 \\ &= \dim_{o/\pi o}(\pi^{-1}L_1 \cap V + L_1/L_1) + \dim_{o/\pi o}L_1/L_0 \\ &= \dim_K V + \dim_{o/\pi o}L_1/L_0\end{aligned}$$

and similarly in the second case that

$$\dim_K V = \dim_K V + \dim_{o/\pi o}\pi^{-1}L_0/L_1.$$

This implies $[L_0] = [L_1]$.

Remark. (i) $\{[L], \gamma_V[L]\}$ is a 1-simplex in $\mathscr{BT}$;

(ii) $[L]$ is a vertex in $\mathscr{BT}^{(W', W'')}$ if and only if $\gamma_{w'}[L] \neq \gamma_{w''}[L]$.

Lemma 9 *If the subcomplex $\mathscr{K}$ of $\mathscr{BT}$ is a union of apartments which have V as a vertex in their boundary then we have:*

i. $\gamma_V(\mathscr{K}) \subseteq \mathscr{K}$;

ii. the map $|\gamma_V|$ induced by γ_V on the topological realization $|\mathscr{K}|$ is homotopic to the identity.

Proof. i. We can assume that $\mathscr{K} = \mathscr{A}$ is a single apartment and it suffices to show that with any vertex $[L]$ in $\mathscr{A}$ also $\gamma_V[L]$ lies in $\mathscr{A}$. But we find a basis $v_1, \ldots, v_{d+1}$ of K^{d+1} such that

$$V = Kv_1 + \ldots + Kv_k \qquad \text{for some } 1 \leqq k \leqq d$$

$$\text{and } L = ov_1 + \ldots + ov_{d+1}\,.$$

Then $[\pi^{-1}L \cap V + L] = [\pi^{-1}ov_1 + \ldots + \pi^{-1}ov_k + ov_{k+1} + \ldots + ov_{d+1}]$ obviously lies in $\mathscr{A}$, too.

ii. On $|\mathscr{BT}|$ we have the metric d which restricted to an apartment is the Euclidean metric on the affine space. Furthermore any two points $x, y \in |\mathscr{BT}|$ are joined by a unique geodesic $[xy]$ which lies in every apartment containing x and y. For any real number $0 \leqq t \leqq 1$ let $tx + (1-t)y$ be the unique point $z \in [xy]$ such that $d(x, z) = (1-t)d(x, y)$. The argument in the proof of part i. shows that with any point $x \in |\mathscr{K}|$ the whole geodesic $[x|\gamma_V|(x)]$ is contained in $|\mathscr{K}|$. Therefore the map

$$\begin{aligned}|\mathscr{K}| \times [0, 1] &\to |\mathscr{K}| \\ (x, t) &\mapsto tx + (1-t)|\gamma_V|(x)\end{aligned}$$

is well-defined and provides the required homotopy once it turns out to be continuous. But it is the restriction of the composed map

$$\begin{array}{ccccc} |\mathscr{BT}| \times [0,1] & \to & |\mathscr{BT}| \times |\mathscr{BT}| \times [0,1] & \to & |\mathscr{BT}| \\ (x, t) & \mapsto & (x, |\gamma_V|(x), t) & & \\ & & (x, y, t) & \mapsto & tx + (1-t)y\,; \end{array}$$

here the first arrow is continuous since $|\gamma_V|$ is continuous and the second arrow is continuous according to [9] (2.5.15).

Lemma 10 *Let $\mathscr{K}$ be a subcomplex of $\mathscr{BT}$ with the following two properties:*
— $\mathscr{K}$ is a union of apartments which have V as a vertex in their boundary, and
— for any simplex σ in $\mathscr{BT}$ there is a $m \in \mathbb{N}$ such that $\gamma_V^m(\sigma)$ lies in $\mathscr{K}$.
Then $\mathscr{K}$ is contractible.

Proof. By a theorem of Whitehead it suffices to show that $\pi_n(|\mathscr{K}|, x) = 0$ for all $n \geqq 0$ and $x \in |\mathscr{K}|$. Let $f: (S^n, s_0) \to (|\mathscr{K}|, x)$ be a continuous map. Since $\mathscr{BT}$ is contractible we find a base point preserving homotopy

$$F: S^n \times [0, 1] \to |\mathscr{BT}| \qquad \text{with } F(.,0) = x \quad \text{and} \quad F(.,1) = f .$$

The image of F being compact is contained in the topological realization of a finite subcomplex of $\mathscr{BT}$. Our assumptions then guarantee that there is a $m \in \mathbb{N}$ such that the image of $|\gamma_V|^m \circ F$ is contained in $|\mathscr{K}|$. This means that the homotopy class of $|\gamma_V|^m \circ f$ in $\pi_n(|\mathscr{K}|, y)$ with $y := |\gamma_V|^m(x)$ is trivial. But according to our first assumption and the previous Lemma the map $|\gamma_V|^m : \pi_n(|\mathscr{K}|, x) \to \pi_n(|\mathscr{K}|, y)$ is a bijection. Therefore already the homotopy class of f was trivial.

Lemma 11 *Let $W \subseteq K^{d+1}$ be a 1-dimensional subspace such that $V \cap W = \{0\}$. For any simplex σ in $\mathscr{BT}$ there is a $m \in \mathbb{N}$ such that $\gamma_V^m(\sigma)$ is contained in an apartment of $\mathscr{BT}$ which has V and W as vertices in its boundary.*

Proof. The simplex σ is of the form $\sigma = \{[L_0], \ldots, [L_q]\}$ with $L_0 \subseteq L_1 \subseteq \ldots \subseteq L_q \subseteq \pi^{-1} L_0 = L_{q+1}$. For any $m \in \mathbb{N}$ we put $L_j^{(m)} := \pi^{-m} L_j \cap V + L_j$ so that we have

$$\gamma_V^m(\sigma) = \{[L_0^{(m)}], \ldots, [L_q^{(m)}]\} .$$

In order to prove the assertion it suffices to find a $m \in \mathbb{N}$ and a $0 \leqq j \leqq q$ such that

$$L_j^{(m)} + W \cap L_{j+1}^{(m)} \nsubseteq L_j^{(m)} + V \cap L_{j+1}^{(m)} .$$

Then the required apartment can be constructed by exactly the same procedure as in the proof of Lemma 7. It is easiest to deduce a contradiction from the assumption that

$$L_j^{(m)} + W \cap L_{j+1}^{(m)} \subseteq L_j^{(m)} + V \cap L_{j+1}^{(m)} \qquad \text{for all } m \in \mathbb{N} \quad \text{and} \quad 0 \leqq j \leqq q .$$

Since $\dim_K W = 1$ we know from Corollary 5 that given $m \in \mathbb{N}$ there is precisely one j such that $L_j^{(m)} \subsetneqq L_j^{(m)} + W \cap L_{j+1}^{(m)}$. By the pigeonhole principle there must be then one j which we fix from now on and some infinite subset $M \subseteq \mathbb{N}$ such that

$$L_j^{(m)} \subsetneqq L_j^{(m)} + W \cap L_{j+1}^{(m)} \subseteq L_j^{(m)} + V \cap L_{j+1}^{(m)} \qquad \text{for all } m \in M .$$

Inserting the definition of $L_j^{(m)}$ we obtain

$$\pi^{-m} L_j \cap V + L_j \subsetneqq (\pi^{-m} L_j \cap V + L_j) + W \cap (\pi^{-m} L_{j+1} \cap V + L_{j+1})$$
$$\subseteq \pi^{-m} L_{j+1} \cap V + L_j \qquad \text{for all } m \in M .$$

This means that for any $m \in M$ there is an element

$$w_m \in W \cap (\pi^{-m} L_{j+1} \cap V + L_{j+1})$$

which is of the form

$$w_m = \pi^{-m} v_m + l_m \qquad \text{with } v_m \in L_{j+1} \cap V \quad \text{and} \quad l_m \in L_j$$

and which is not contained in $\pi^{-m} L_j \cap V + L_j$. The latter implies that $v_m \notin L_j \cap V$. In the identity

$$\pi^m w_m = v_m + \pi^m l_m$$

all three terms viewed as sequences in m lie in compact subsets of K^{d+1} so that we may assume replacing M by a smaller subset that all three sequences are even convergent. Now the $\pi^m l_m \in \pi^m L_j$ obviously converge to 0; similarly the $\pi^m w_m \in W \cap (L_{j+1} \cap V + \pi^m L_{j+1})$ converge to an element in $W \cap V = \{0\}$, that is to 0. But the v_m lie in the closed set $(L_{j+1} \cap V) \backslash (L_j \cap V)$ so that they cannot converge to 0. This is the wanted contradiction.

Proposition 12 *The subcomplexes $\mathscr{B}\mathscr{T}^{(t_0, \ldots, t_m)}$ in $\mathscr{B}\mathscr{T}$ are contractible whenever the elements $t_0, \ldots, t_m \in T$ are not all equal.*

Proof. The flag t_μ is of the form $t_\mu = (W_0^{(\mu)} \subseteq \ldots \subseteq W_r^{(\mu)})$. Since not all the $t_0, \ldots, t_m$ are equal we find a smallest $0 \leqq \nu \leqq r$ such that

$$W_\nu^{(0)} \neq W_\nu^{(\mu)} \qquad \text{for some } 1 \leqq \mu \leqq m .$$

We fix a 1-dimensional subspace $W \subseteq W_\nu^{(\mu)} \backslash W_\nu^{(0)}$. From Lemma 6 and Proposition 8 we then deduce that the subcomplex $\mathscr{B}\mathscr{T}^{(t_0, \ldots, t_m)}$
— is a union of apartments in $\mathscr{B}\mathscr{T}$ which have $V := W_\nu^{(0)}$ as vertex in their boundary, and
— contains all apartments which have V and W as vertices in their boundary.
In this situation Lemma 11 says that $\mathscr{B}\mathscr{T}^{(t_0, \ldots, t_m)}$ fulfills the assumptions of Lemma 10 and therefore is contractible.

We now have established all the assumptions of Proposition 1 for $\mathscr{B}\mathscr{T}$ and the profinite set $T = G/P_I$ so that Proposition 1 gives the following result.

Theorem 7 *The complex*

$$0 \to \bigoplus_{\tau \in \mathscr{B}\mathscr{T}_d} C^\infty(G/P_I, \mathbb{Z})^{U_\tau} \to \ldots \to \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_0} C^\infty(G/P_I, \mathbb{Z})^{U_\sigma}$$
$$\to C^\infty(G/P_I, \mathbb{Z}) \to 0$$

is exact for any subset $I \subseteq \Delta$.

From this Theorem we want to deduce a similar result for the G-modules $V_I(\mathbb{Z})$. We start with the simple observation that any smooth G-module A (by which we mean an abelian group A with smooth G-action) gives rise to a coefficient system $\underline{A}$ on the Bruhat–Tits building $\mathscr{B}\mathscr{T}$ in which $A_\sigma := A^{U_\sigma}$ and the transition maps are the obvious inclusions. The corresponding augmented homological complex reads

$$0 \to \bigoplus_{\tau \in \mathscr{B}\mathscr{T}_d} A^{U_\tau} \to \ldots \to \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_0} A^{U_\sigma} \to A \to 0 .$$

For $A = C^\infty(G/P_I, \mathbb{Z})$ this is exactly the situation which we have discussed above. Clearly the coefficient system $\underline{A}$ is functorial in the smooth G-module A. We

therefore have, for any complex $\ldots \to A_0 \to \ldots \to A_m \to \ldots$ of smooth G-modules, the double complex

$$\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \bigoplus_{\tau \in \mathscr{B}\mathscr{T}_d} A_m^{U_\tau} & \to \ldots \to & \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_0} A_m^{U_\sigma} & \to & A_m & \to & 0 \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
 & & \vdots & & \vdots & & \vdots & & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \to & \bigoplus_{\tau \in \mathscr{B}\mathscr{T}_d} A_0^{U_\tau} & \to \ldots \to & \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_0} A_0^{U_\sigma} & \to & A_0 & \to & 0 \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
 & & \vdots & & \vdots & & \vdots & &
\end{array}$$

Since the U_σ are pro-p-groups the above functor $A \mapsto \underline{A}$ is exact on smooth $G - \mathbb{Z}[1/p]$-modules. Some of the complications which we have to overcome in the following are caused by our insisting in $\mathbb{Z}$-coefficients. By different techniques one can show that the above homological complex is exact for all irreducible smooth $G - \mathbb{Q}$-modules A for which $A^{U_\sigma} \neq \{0\}$ for some vertex $\sigma \in \mathscr{B}\mathscr{T}_0$. We hope to come back to this in another paper.

The obvious idea to treat the G-modules $V_I(\mathbb{Z})$ is to resolve them by the G-modules $C^\infty(G/P_J, \mathbb{Z})$ for $J \supseteq I$. For some of the subsets $I \subseteq \Delta$ we have done this implicitly already in §3 Proposition 6 by investigating certain topological Tits buildings. This has to be generalized appropriately. Since $V_\Delta(\mathbb{Z}) = C^\infty(G/P_\Delta, \mathbb{Z}) = \mathbb{Z}$ we may assume in the following that $I \subsetneq \Delta$. Define $\mathscr{T}_\bullet^I$ to be the simplicial profinite set of all flags $W_0 \subseteq \ldots \subseteq W_r$ of subspaces in K^{d+1} such that $\dim_K W_\nu \in \Delta \backslash I$ for all $0 \leqq \nu \leqq r$. Furthermore let $N\mathscr{T}_r^I \subseteq \mathscr{T}_r^I$ denote the open and closed subset of "nondegenerate" flags $W_0 \subsetneq \ldots \subsetneq W_r$; $N\mathscr{T}_r^I$ consists of all flags of type $J \supseteq I$ with $\# \Delta \backslash J = r + 1$. (Neglecting that in §3 we worked in the dual space we have $\mathscr{T}_\bullet^{(k)} = \mathscr{T}_\bullet^{\{k+1,\ldots,d\}}$.)

Proposition 13 *The reduced cohomology* $\tilde{H}^r(|\mathscr{T}_\bullet^I|, \mathbb{Z})$ *vanishes for* $r \neq d - 1 - \# I$.

Proof. Put $k := d - \# I$. Since $N\mathscr{T}_r^I = \phi$ for $r \geqq k$ it is clear that $\tilde{H}^r(|\mathscr{T}_\bullet^I|, \mathbb{Z}) = 0$ for $r \geqq k$ (compare the proof of §3 Proposition 6). We will prove the vanishing for $r < k - 1$ by induction with respect to d. The case $d = 1$ is trivial. For general d we use the spectral sequence of a certain double complex. Set

$$Y_\bullet^I := \text{simplicial profinite set of all sequences } (H_0, \ldots, H_r) \text{ of 1-dim. subspaces } H_i \subseteq K^{d+1} \text{ such that } \dim_K \sum_{i=0}^r H_i \leqq j \text{ for some } j \in \Delta \backslash I$$

and

$$Z_{\bullet\bullet}^I := \text{bisimplicial profinite set of all } (W_0 \subseteq \ldots \subseteq W_r; H_0, \ldots, H_s) \in \mathscr{T}_r^I \times Y_s^I$$

$$\text{such that } \sum_{i=0}^s H_i \subseteq W_0 .$$

From the obvious projection maps

$$\begin{array}{ccc} & & Y_\bullet^I \\ & & \uparrow \\ \mathscr{T}_\bullet^I & \leftarrow & Z_{\bullet\bullet}^I \end{array}$$

we obtain the biaugmented double complex

$$\begin{array}{ccc} & & C^\infty(Y_\bullet^I, \mathbb{Z}) \\ & & \downarrow \\ C^\infty(\mathscr{T}_\bullet^I, \mathbb{Z}) & \to & C^\infty(Z_{\bullet\bullet}^I, \mathbb{Z}) \,. \end{array}$$

The same argument as in §3 Proposition 5 shows that the horizontal augmentation map is a quasi-isomorphism. Therefore the second spectral sequence of this double complex reads

$$E_1^{r,s} = H^s(C^\infty(Z_{\bullet r}^I, \mathbb{Z})) \Rightarrow H^{r+s}(\mathrm{Tot}\, C^\infty(Z_{\bullet\bullet}^I, \mathbb{Z})) = H^{r+s}(|\mathscr{T}_\bullet^I|, \mathbb{Z}) \,.$$

We claim the following partial exactness property of the perpendicular augmentation map: For $0 \leqq r \leqq k-1$ the sequence

$$0 \to C^\infty(Y_r^I, \mathbb{Z}) \to C^\infty(Z_{0r}^I, \mathbb{Z}) \to \ldots \to C^\infty(Z_{k-1-r,r}^I, \mathbb{Z})$$

is exact.

According to the argument in the proof of §3 Lemma 4 (i.e., [20] Chap. II) this can be checked on the fibers of the projection map

$$f_\bullet : Z_{\bullet r}^I \to Y_r^I \,.$$

Fix $h = (H_0, \ldots, H_r) \in Y_r^I$ and put $W := \sum_{i=0}^r H_i$ and $m := \dim_K W$; we have $m \leqq r+1$. The fiber $f_\bullet^{-1}(h)$ is the simplicial profinite set of all flags $W_0 \subseteq \ldots \subseteq W_r$ such that $\dim_K W_v \in \Delta \backslash I$ and $W \subseteq W_0$.

There are two cases to distinguish. First let us assume that $m \in \Delta \backslash I$. Then W is a vertex in $f_\bullet^{-1}(h)$ which therefore is cohomologically trivial by the argument in §3 Lemma 3. On the other hand if $m \in I$ then

$$f_\bullet^{-1}(h) \to \mathscr{T}(K^{d+1}/W)_\bullet^{I'}$$

$$(W_0 \subseteq \ldots \subseteq W_r) \mapsto (W_0/W \subseteq \ldots \subseteq W_r/W)$$

is an isomorphism of simplicial profinite sets where the right hand side is the topological Tits building of all flags in K^{d+1}/W whose dimensions are contained in $\{1, \ldots, d-m\} \backslash I'$ with $I' := \{i - m : i \in I \text{ and } i > m\}$. By the induction hypothesis the reduced cohomology of $|\mathscr{T}(K^{d+1}/W)_\bullet^{I'}|$ vanishes in degrees $< d - m - 1 - \# I'$. Because of $m \in I$ we have $\# I' < \# I$. We conclude that the reduced cohomology of $|f_\bullet^{-1}(h)|$ vanishes in degrees $< k-1-r \leqq k-m = d-m-\# I \leqq d-m-1-\# I'$. This establishes our claim. In terms of the above spectral sequence it means that

$$E_1^{r,s} = 0 \qquad \text{for } r+s < k-1 \quad \text{and} \quad s \neq 0 \,,$$

$$E_1^{r,0} = C^\infty(Y_r^I, \mathbb{Z}) \quad \text{for } r < k-1, \quad \text{and} \quad C^\infty(Y_{k-1}^I, \mathbb{Z}) \subseteq E_1^{k-1,0} \,.$$

Since certainly $d - \# I \leqq j$ for some $j \in \Delta \backslash I$ we have

$$Y_r^I = \underbrace{Y_0^I \times \ldots \times Y_0^I}_{r+1\text{-times}} \qquad \text{for } r \leqq k-1 \,.$$

The contracting argument in the proof of §3 Lemma 3 then implies that

$$\begin{array}{ccccccc} 0 \to \mathbb{Z} \to & C^\infty(Y_0^I, \mathbb{Z}) & \to \ldots \to & C^\infty(Y_{k-2}^I, \mathbb{Z}) & \to & C^\infty(Y_{k-1}^I, \mathbb{Z}) \\ \| & \| & & \| & & \cap \\ 0 \to \mathbb{Z} \to & E_1^{0,0} & \to \ldots \to & E_1^{k-2,0} & \to & E_1^{k-1,0} \end{array}$$

is exact. Altogether we obtain that $E_2^{r,s} = 0$ for $0 < r + s < k - 1$ and $E_2^{0,0} = \mathbb{Z}$ which in the abutment gives the vanishing of $\tilde{H}^r(|\mathscr{T}_\bullet^I|, \mathbb{Z})$ for $0 \leqq r < k - 1$. q.e.d.

As already discussed several times we have

$$C^\infty(N\mathscr{T}_r^I, \mathbb{Z}) = \bigoplus_{\substack{I \subseteq J \subseteq \Delta \\ \# \Delta \setminus J = r+1}} C^\infty(G/P_J, \mathbb{Z}) .$$

Therefore Proposition 13 can be reformulated by saying that

$$0 \to \mathbb{Z} \to C^\infty(N\mathscr{T}_0^I, \mathbb{Z}) \to \ldots \to C^\infty(N\mathscr{T}_{d-1-\#I}^I, \mathbb{Z}) \to V_I(\mathbb{Z}) \to 0$$

is an exact sequence of smooth G-modules. In the corresponding diagram

$$\begin{array}{ccc} 0 & & 0 \\ \uparrow & & \uparrow \\ \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_\bullet} V_I(\mathbb{Z})^{U_\sigma} & \longrightarrow & V_I(\mathbb{Z}) \\ \uparrow & & \uparrow \\ \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_\bullet} C^\infty(N\mathscr{T}_{d-1-\#I}^I, \mathbb{Z})^{U_\sigma} & \longrightarrow & C^\infty(N\mathscr{T}_{d-1-\#I}, \mathbb{Z}) \\ \uparrow & & \uparrow \\ \vdots & & \vdots \\ \uparrow & & \uparrow \\ \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_\bullet} C^\infty(N\mathscr{T}_0^I, \mathbb{Z})^{U_\sigma} & \longrightarrow & C^\infty(N\mathscr{T}_0^I, \mathbb{Z}) \\ \uparrow & & \uparrow \\ \bigoplus_{\sigma \in \mathscr{B}\mathscr{T}_\bullet} \mathbb{Z} & \longrightarrow & \mathbb{Z} \\ \uparrow & & \uparrow \\ 0 & & 0 \end{array}$$

all horizontal arrows apart possibly from the first one are quasi-isomorphisms by Theorem 7. Our aim is to prove that the first arrow is a quasi-isomorphism, too. For this it is sufficient to check that all the left hand columns are exact. Let us fix a simplex $\sigma = \{[L_0], \ldots, [L_q]\}$ in $\mathscr{B}\mathscr{T}$ with $L_0 \subseteq \ldots \subseteq L_q \subseteq \pi^{-1}L_0 = L_{q+1}$. Because of

$$C^\infty(N\mathscr{T}_r^I, \mathbb{Z})^{U_\sigma} = C(U_\sigma \backslash N\mathscr{T}_r^I, \mathbb{Z})$$

we deduce from Corollary 5 that

$$C^\infty(N\mathscr{T}_0^I, \mathbb{Z})^{U_\sigma} \to \ldots \to C^\infty(N\mathscr{T}_{d-1-\#I}^I, \mathbb{Z})^{U_\sigma}$$

is the cochain complex of the flag complex associated with the poset

$$\mathscr{T}^I(\sigma) := \{\xi \in \mathscr{T}(L_0, \ldots, L_q) : \dim \xi \in \Delta \setminus I\} .$$

We only remark that $\mathscr{T}^I(\sigma)$ up to canonical isomorphism does not depend on the choice of $L_0, \ldots, L_q$. In the following we will not distinguish notationally between a poset and its associated flag complex.

Proposition 14 *The reduced cohomology* $\tilde{H}^r(|\mathscr{T}^I(\sigma)|, \mathbb{Z})$ *vanishes for* $r \neq d - 1 - \# I$.

Proof. This can be proved in a completely analogous way as Proposition 13. Therefore we only explain how to translate the notions we have used in that argument into the present context. First of all it is convenient to work in a slightly more general setting. Suppose we are given finitely many $o/\pi o$-vector spaces $\bar{L}_0, \ldots, \bar{L}_q$ such that

$$\dim_{o/\pi o} \bar{L}_0 + \ldots + \dim_{o/\pi o} \bar{L}_q = d + 1 .$$

We then have the poset

$$\mathscr{T}(\bar{L}_0, \ldots, \bar{L}_q) := \text{set of all sequences } (\bar{X}_0, \ldots, \bar{X}_q) \text{ where } \bar{X}_j \text{ is an } o/\pi o\text{-subspace of } \bar{L}_j$$

partially ordered by

$$(\bar{X}_0, \ldots, \bar{X}_q) \leqq (\bar{Y}_0, \ldots, \bar{Y}_q) \quad \text{if } \bar{X}_j \subseteq \bar{Y}_j \qquad \text{for all } 0 \leqq j \leqq q .$$

Any element $(\bar{X}_0, \ldots, \bar{X}_q) \in \mathscr{T}(\bar{L}_0, \ldots, \bar{L}_q)$ has the dimension

$$\dim(\bar{X}_0, \ldots, \bar{X}_q) := \sum_{j=0}^{q} \dim_{o/\pi o} \bar{X}_j .$$

Furthermore we may define the sum of any two such elements as

$$(\bar{X}_0, \ldots, \bar{X}_q) + (\bar{Y}_0, \ldots, \bar{Y}_q) := (\bar{X}_0 + \bar{Y}_0, \ldots \bar{X}_q + \bar{Y}_q) .$$

For any proper subset $I \subsetneqq \Delta$ we consider the poset

$$\mathscr{T}^I(\bar{L}_0, \ldots, \bar{L}_q) := \{\xi \in \mathscr{T}(\bar{L}_0, \ldots, \bar{L}_q) : \dim \xi \in \Delta \backslash I\} .$$

The assertion now is proved simultaneously for all these posets by induction with respect to d. The argument is exactly the same as in the proof of Proposition 13; actually it can be somewhat simplified since no profinite topology is involved.

Proposition 15 $\tilde{H}^{d-1-\#I}(|\mathscr{T}^I(\sigma)|, \mathbb{Z}) = \tilde{H}^{d-1-\#I}(|\mathscr{T}^I_\bullet|, \mathbb{Z})^{U_\sigma} = V_I(\mathbb{Z})^{U_\sigma}$.

Proof. We will show that the sequence

$$\left(\bigoplus_{i \in \Delta \backslash I} C^\infty(G/P_{I \cup \{i\}}, \mathbb{Z}) \right)^{U_\sigma} \to C^\infty(G/P_I, \mathbb{Z})^{U_\sigma} \to V_I(\mathbb{Z})^{U_\sigma} \to 0$$

is exact. Since U_σ is a pro-p-group taking U_σ-invariants is exact if we deal with $\mathbb{Z}[1/p]$-modules. Therefore our sequence certainly is exact after tensoring with $\mathbb{Z}[1/p]$. By conjugation we may assume that $B_\sigma = B_J$ for some $J \subseteq \Delta$ is one of the standard parahoric subgroups considered in §4. If σ_0 denotes the vertex $\sigma_0 = [o^{d+1}]$ we then have

$$U_{\sigma_0} \subseteq U_\sigma \subseteq B_\sigma = B_J \subseteq B_{\sigma_0} = B_\Delta = \mathrm{GL}_{d+1}(o) .$$

Let $\bar{P}_I$, resp. $\bar{U}_I$, be the standard parabolic subgroups (containing the upper triangular matrices), resp. their unipotent radicals, in the group $\bar{G} := \mathrm{GL}_{d+1}(o)/U_{\sigma_0} = \mathrm{GL}_{d+1}(o/\pi o)$; they are defined in the same way as the P_I and U_I in §4. It is clear that

$$\bar{P}_I = B_I/U_{\sigma_0}, \qquad \bar{P}_J = B_\sigma/U_{\sigma_0}, \qquad \text{and} \qquad \bar{U}_J = U_\sigma/U_{\sigma_0} .$$

Using the Iwasawa decomposition $G = B_\Delta P_\phi$ we can rewrite our sequence in the following way:

$$\left(\bigoplus_{i \in \Delta \setminus I} C(\bar{G}/\bar{P}_{I \cup \{i\}}, \mathbb{Z}) \right)^{\bar{U}_J} \to C(\bar{G}/\bar{P}_I, \mathbb{Z})^{\bar{U}_J} \to (V_I(\mathbb{Z})^{U_{\sigma_0}})^{\bar{U}_J} \to 0 .$$

A drastically simplified version of §4 Proposition 4 (no topology is involved!) tells us that the subgroup

$$C\left(\bigcup_{w \in V^I}^{\bullet} \bar{P}_\phi w \bar{P}_I/\bar{P}_I, \mathbb{Z} \right)^{\bar{U}_J} = C\left(\bigcup_{w \in V^I}^{\bullet} \bar{U}_J \backslash \bar{P}_\phi w \bar{P}_I/\bar{P}_I, \mathbb{Z} \right)$$

in

$$C(\bar{G}/\bar{P}_I, \mathbb{Z})^{\bar{U}_J} = C(\bar{U}_J \backslash G/\bar{P}_I, \mathbb{Z})$$

is a complement of the image of the left hand term in the above sequence, i.e., of the subgroup

$$\sum_{i \in \Delta \setminus I} C(\bar{U}_J \backslash \bar{G}/\bar{P}_{I \cup \{i\}}, \mathbb{Z}) .$$

We recall that

$$V^I = W^I \setminus \bigcup_{i \in \Delta \setminus I} W^{I \cup \{i\}}$$

is the set of all permutations $w \in W$ which are of minimal length in their coset wW_I but not in any coset $wW_{I \cup \{i\}}$ for $i \in \Delta \setminus I$. In this way we are reduced to prove that

$$C\left(\bigcup_{w \in V^I}^{\bullet} \bar{P}_\phi w \bar{P}_I/\bar{P}_I, \mathbb{Z} \right) \to V_I(\mathbb{Z})^{U_{\sigma_0}}$$

is bijective. Both sides are $\mathbb{Z}$-free, the left side obviously and the right side since $V_I(\mathbb{Z})$ is $\mathbb{Z}$-free by the argument in the proof of §4 Corollary 5. We already know that the map is bijective after tensoring with $\mathbb{Z}[1/p]$. Therefore it is at least injective with torsion cokernel. For its surjectivity it then suffices that the cokernel of the map

$$C\left(\bigcup_{w \in V^I}^{\bullet} \bar{P}_\phi w \bar{P}_I/\bar{P}_I, \mathbb{Z} \right) \to V_I(\mathbb{Z})$$

is torsionfree. This in turn certainly is the case if the dual map

$$\mathrm{Hom}_{\mathbb{Z}}(V_I(\mathbb{Z}), \mathbb{Z}) \to \mathrm{Hom}_{\mathbb{Z}}\left(C\left(\bigcup_{w \in V^I}^{\bullet} \bar{P}_\phi w \bar{P}_I/\bar{P}_I, \mathbb{Z} \right), \mathbb{Z} \right)$$

is surjective. We observe that this still is a map of B_ϕ-modules. For any $h \in G$ we have the Dirac distribution

$$\delta_h : C^\infty(G/P_I, \mathbb{Z}) \to \mathbb{Z}$$
$$\psi \mapsto \psi(h) .$$

The action of G on the Dirac distributions is described by the formula

$$^g(\delta_h) = \delta_{gh} .$$

We now consider the linear form

$$\delta_I := \sum_{w \in W_{\Delta \setminus I}} (-1)^{l(w)} \delta_w$$

on $C^\infty(G/P_I, \mathbb{Z})$. Fix an $i \in \Delta \setminus I$ and let $\psi \in C^\infty(G/P_{I \cup \{i\}}, \mathbb{Z})$ be some function. From $\psi(\,.\, s_i) = \psi$ and from the decomposition

$$W_{\Delta \setminus I} = (W_{\Delta \setminus I} \cap W^{\{i\}}) \dot{\bigcup} (W_{\Delta \setminus I} \cap W^{\{i\}}) s_i$$

we deduce

$$\begin{aligned} \delta_I(\psi) &= \sum_{w \in W_{\Delta \setminus I}} (-1)^{l(w)} \psi(w) \\ &= \sum_{w \in W_{\Delta \setminus I} \cap W^{\{i\}}} (-1)^{l(w)} \psi(w) + \sum_{w \in W_{\Delta \setminus I} \cap W^{\{i\}}} (-1)^{l(ws_i)} \psi(ws_i) \\ &= 0 \,. \end{aligned}$$

Therefore δ_I lies in $\mathrm{Hom}_{\mathbb{Z}}(V_I(\mathbb{Z}), \mathbb{Z})$. For any $v \in V^I$ the restriction of the linear form ${}^v(\delta_I)$ to $C(\bigcup^{\bullet}_{w \in V^I} \bar{P}_\phi w \bar{P}_I / \bar{P}_I, \mathbb{Z})$ is equal to

$$\sum_{\substack{w \in W_{\Delta \setminus I} \\ vw \in V^I W_I}} (-1)^{l(w)} \delta_{vw} = \sum_{v' \in V^I} m(v, v') \cdot \delta_{v'}$$

where

$$m(v, v') := \sum_{\substack{w \in W_{\Delta \setminus I} \\ vw \in v' W_I}} (-1)^{l(w)} \,.$$

Because of $W_{\Delta \setminus I} \cap W_I = \{1\}$ the pair $(w, w') \in W_{\Delta \setminus I} \times W_I$ in the identity

$$vw = v'w'$$

is uniquely determined by the pair $(v, v') \in V^I \times V^I$. This implies that $m(v, v') \in \{0, \pm 1\}$ and that $m(v, v) = 1$. Since v' is the element of minimal length in its coset $v' W_I$ we have $v' \leqq v' w'$ ($\leqq$ denotes the Bruhat order). Also any $v \in V^I$ is the unique element of maximal length in its coset $v W_{\Delta \setminus I}$ so that $vw \leqq v$. Together we obtain

$$v' \leqq v \qquad \text{if } m(v, v') \neq 0 \,.$$

We see that the matrix formed by the coefficients $m(v, v')$ is integrally invertible. Therefore an appropriate integral linear combination of the ${}^{v'}(\delta_I)$ with $v' \in V^I$ restricts to δ_v on $C(\bigcup^{\bullet}_{w \in V^I} \bar{P}_\phi w \bar{P}_I / \bar{P}_I, \mathbb{Z})$. Since the δ_v for $v \in V^I$ generate

$$\mathrm{Hom}_{\mathbb{Z}}\left(C\left(\bigcup^{\bullet}_{w \in V^I} \bar{P}_\phi w \bar{P}_I / \bar{P}_I, \mathbb{Z} \right), \mathbb{Z} \right)$$

as a B_ϕ-module we have established the wanted surjectivity. q.e.d.

We know now that the columns in our initial diagram are exact and as already explained this implies the result we were aiming at.

Theorem 8 *The complex*

$$0 \to \bigoplus_{\tau \in \mathscr{F}_d} V_I(\mathbb{Z})^{U_\tau} \to \ldots \to \bigoplus_{\sigma \in \mathscr{F}_0} V_I(\mathbb{Z})^{U_\sigma} \to V_I(\mathbb{Z}) \to 0$$

is exact for all subsets $I \subseteq \Delta$.

We want to complete this discussion by relating $\mathscr{T}^I(\sigma)$ to that simplicial complex which most naturally is associated with the simplex σ. This is the combinatorial Tits building $\mathscr{T}_\sigma$ of the finite reductive group B_σ/U_σ which also can be interpreted as the link of the simplex σ in $\mathscr{B}\mathscr{T}$. It is convenient to put ourselves into the slightly more general framework of those simplicial complexes which were introduced in the proof of Proposition 14. We repeat that, given $o/\pi o$-vector spaces $\bar{L}_0, \ldots, \bar{L}_q$ such that

$$\dim_{o/\pi o} \bar{L}_0 + \ldots + \dim_{o/\pi o} \bar{L}_q = d + 1 ,$$

we have the poset

$$\mathscr{T}(\bar{L}_0, \ldots, \bar{L}_q) := \text{set of all sequences } (\bar{X}_0, \ldots, \bar{X}_q)$$
$$\text{where } \bar{X}_j \text{ is an } o/\pi o\text{-subspace of } \bar{L}_j$$

with the dimension function

$$\dim(\bar{X}_0, \ldots, \bar{X}_q) := \sum_{j=0}^{q} \dim_{o/\pi o} \bar{X}_j .$$

If we put

$$\bar{L} := \bar{L}_0 \oplus \ldots \oplus \bar{L}_q$$

then we may also consider the poset

$$\mathscr{T}[\bar{L}_0, \ldots, \bar{L}_q] := \text{set of all } o/\pi o\text{-subspaces } \bar{X} \subseteq \bar{L}$$
$$\text{such that } \bar{X} = (\bar{X} \cap \bar{L}_0) \oplus \ldots \oplus (\bar{X} \cap \bar{L}_q)$$

ordered by inclusion. It is clear that

$$\mathscr{T}[\bar{L}_0, \ldots, \bar{L}_q] \xrightarrow{\sim} \mathscr{T}(\bar{L}_0, \ldots, \bar{L}_q)$$
$$\bar{X} \mapsto (\bar{X} \cap \bar{L}_0, \ldots, \bar{X} \cap \bar{L}_q)$$

is a dimension preserving isomorphism of posets. On the other hand we fix a semisimple element $s \in \mathrm{GL}(\bar{L})$ such that

$$\mathrm{GL}(\bar{L}_0) \times \ldots \times \mathrm{GL}(\bar{L}_q) = \text{centralizer of } s \text{ in } \mathrm{GL}(\bar{L}) .$$

One then easily checks that $\mathscr{T}[\bar{L}_0, \ldots, \bar{L}_q]$ is the subposet of fixed points of s in $\mathscr{T}[\bar{L}]$:

$$\mathscr{T}[\bar{L}_0, \ldots, \bar{L}_q] = \mathscr{T}[\bar{L}]^s = \mathscr{T}(\bar{L})^s .$$

For any proper subset $I \subsetneq \Delta$ we had considered the poset

$$\mathscr{T}^I(\bar{L}_0, \ldots, \bar{L}_q) := \{\xi \in \mathscr{T}(\bar{L}_0, \ldots, \bar{L}_q) : \dim \xi \in \Delta \backslash I\} .$$

Obviously the $\mathscr{T}^I(\sigma)$ are of this form. Using the above isomorphism we obtain

$$\mathscr{T}^I(\bar{L}_0, \ldots, \bar{L}_q) \cong \{\bar{X} \in \mathscr{T}^\phi(\bar{L})^s : \dim_{o/\pi o} \bar{X} \in \Delta \backslash I\} .$$

The combinatorial Tits building of the group $\mathrm{GL}(\bar{L})$, resp. $\mathrm{GL}(\bar{L}_0) \times \ldots \times \mathrm{GL}(\bar{L}_q)$, by definition is $\mathscr{T}^\phi(\bar{L})$, resp. the join $\mathscr{T}^\phi(\bar{L}_0) * \ldots * \mathscr{T}^\phi(\bar{L}_q)$. In [14] (7.4) it is shown that there is a $\mathrm{GL}(\bar{L}_0) \times \ldots \times \mathrm{GL}(\bar{L}_q)$-equivariant homeomorphism

$$\mathrm{Susp}^q |\mathscr{T}^\phi(\bar{L}_0) * \ldots * \mathscr{T}^\phi(\bar{L}_q)| \cong |\mathscr{T}^\phi(\bar{L})^s|$$

where Susp^q denotes the q-fold suspension. We see in particular that for a q-simplex σ in $\mathscr{B}\mathscr{T}$ the topological realizations $|\mathscr{T}^I(\sigma)|$ of our simplicial complexes $\mathscr{T}^I(\sigma)$ are

B_σ/U_σ-equivariantly homeomorphic to closed subspaces in $\mathrm{Susp}^q|\mathscr{T}_\sigma|$. For general I it seems rather complicated to find a description of those subspaces which is intrinsic in terms of $\mathscr{T}_\sigma$. For $I = \phi$ we of course have $|\mathscr{T}^\phi(\sigma)| \cong \mathrm{Susp}^q|\mathscr{T}_\sigma|$. This latter remark shows for example that the term $V_\phi(\mathbb{Z})^{U_\sigma}$ in the resolution of Theorem 8 for the Steinberg module $V_\phi(\mathbb{Z})$ is the Steinberg module of the finite reductive group B_σ/U_σ (this can also be seen directly from the computations in the proof of Proposition 15).

The resolutions as stated in the Theorems 7 and 8 have the defect not to be G-equivariant. The reason is that in the above considerations we have fixed an orientation of $\mathscr{BT}$ and this cannot be done in a G-invariant way. But this is not a serious problem. Recall that an ordered q-simplex of $\mathscr{BT}$ is a sequence $(\sigma_0, \ldots, \sigma_q)$ of vertices such that $\{\sigma_0, \ldots, \sigma_q\}$ is a q-simplex in $\mathscr{BT}$. Two such ordered q-simplices are called equivalent if they differ by an even permutation of the vertices; the corresponding equivalence classes are called oriented q-simplices and are denoted by $\langle \sigma_0, \ldots, \sigma_q \rangle$. Let $\mathscr{BT}_{(q)}$ be the set of all oriented q-simplices of $\mathscr{BT}$. If now A is a smooth G-module we put

$$C_c^{or}(\mathscr{BT}_{(q)}, A) := \text{group of all maps } \omega : \mathscr{BT}_{(q)} \to A$$

such that

— ω has finite support,

— $\omega(\langle \sigma_0, \ldots, \sigma_q \rangle) \in A^{U_{\{\sigma_0, \ldots, \sigma_q\}}}$,

— $\omega(\langle \sigma_{\iota(0)}, \ldots, \sigma_{\iota(q)} \rangle) = \mathrm{sgn}(\iota) \cdot \omega(\langle \sigma_0, \ldots, \sigma_q \rangle)$

for any permutation ι.

The group G acts on $C_c^{or}(\mathscr{BT}_{(q)}, A)$ via

$$(g\omega)(\langle \sigma_0, \ldots, \sigma_q \rangle) := g(\omega(\langle g^{-1}\sigma_0, \ldots, g^{-1}\sigma_q \rangle)) .$$

With respect to the boundary map

$$\partial : C_c^{or}(\mathscr{BT}_{(q+1)}, A) \to C_c^{or}(\mathscr{BT}_{(q)}, A)$$

$$\omega \mapsto \left(\langle \sigma_0, \ldots, \sigma_q \rangle \mapsto \sum_{\substack{\{\sigma, \sigma_0, \ldots, \sigma_q\} \\ \in \mathscr{BT}_{q+1}}} \omega(\langle \sigma, \sigma_0, \ldots, \sigma_q \rangle) \right)$$

we then have the augmented complex of smooth G-modules

$$C_c^{or}(\mathscr{BT}_{(d)}, A) \xrightarrow{\partial} \ldots \xrightarrow{\partial} C_c^{or}(\mathscr{BT}_{(0)}, A) \to A$$

$$\omega \mapsto \sum_{\sigma \in \mathscr{BT}_{(0)}} \omega(\sigma) .$$

It is rather clear that this complex is isomorphic to our original complex

$$\bigoplus_{\tau \in \mathscr{BT}_d} A^{U_\tau} \to \ldots \to \bigoplus_{\sigma \in \mathscr{BT}_0} A^{U_\sigma} \to A .$$

Proposition 16 *Let $\Gamma \subseteq \mathrm{PGL}_{d+1}(K)$ be a cocompact discrete subgroup. Any of the G-modules $C^\infty(G/P_I, \mathbb{Z})$ and $V_I(\mathbb{Z})$ for $I \subseteq \Delta$ has a projective resolution by finitely generated free $\mathbb{Z}[\Gamma]$-modules; if Γ is torsionfree then there exists such a resolution of length $\leqq d$.*

Proof. Let A denote one of these G-modules. By the argument in the proof of §4 Corollary 5 we know that A is $\mathbb{Z}$-free so that the $\mathbb{Z}$-modules A^{U_σ} are finitely generated and free. Let us first assume that Γ is torsionfree. Then Γ acts freely on $\mathscr{B}\mathscr{T}$ with a finite number of orbits in the q-simplices for each q. Therefore the $C_c^{or}(\mathscr{B}\mathscr{T}_{(q)}, A)$ are finitely generated free $\mathbb{Z}[\Gamma]$-modules and the above constructed resolution has the required properties. In general, Γ contains a normal subgroup Γ' of finite index which is torsionfree ([19] 2.7). Let $B_\bullet := \otimes_{\mathbb{Z}}^{\bullet+1} \mathbb{Z}[\Gamma/\Gamma']$ denote the (unnormalized) bar resolution of the trivial Γ/Γ'-module $\mathbb{Z}$. If now $C_\bullet$ is a resolution of the required type for A as a Γ/Γ'-module then $B_\bullet \otimes_{\mathbb{Z}} C_\bullet$ has the required properties for A as a Γ-module.

References

1. Báyer, P., Neukirch, J.: On Automorphic Forms and Hodge Theory. Math. Ann. **257**, 137–155 (1981)
2. Bernstein, I.N., Zelevinskii, A.V.: Representations of the group GL(n, F) where F is a non-archimedean local field. Russ. Math. Surv. **31**(3), 1–68 (1976)
3. Borel, A., Serre, J.-P.: Cohomologie d'immeubles et de groupes S-arithmétiques. Topology **15**, 211–232 (1976)
4. Borel, A., Tits, J.: Compléments à l'article: "Groupes réductifs". Publ. Math. IHES **41**, 253–276 (1972)
5. Borel, A., Wallach, N.: Continuous cohomology, discrete subgroups, and representations of reductive groups. Ann. Math. Stud. 94. Princeton University Press 1980
6. Bosch, S., Güntzer, U., Remmert, R.: Non-Archimedean analysis. Berlin Heidelberg New York: Springer 1984
7. Bourbaki, N.: Groupes et algèbres de Lie, Chap. 4–6. Paris: Masson 1981
8. Brown, K.S.: Buildings. Berlin Heidelberg New York: Springer 1989
9. Bruhat, F., Tits, J.: Groupes réductifs sur un corps local. I. Données radicielles valuées. Publ. Math. IHES **41**, 5–252 (1972)
10. Cartier, P.: Les arrangements d'hyperplane: Un chapitre de géométrie combinatoire. Sém. Bourbaki 1980/81, exp. 561. Lect. Notes Math. vol. 901. Berlin Heidelberg New York: Springer 1981
11. Casselman, W.: Introduction to the theory of admissible representations of $\wp$-adic reductive groups (Preprint)
12. Casselman, W.: On a p-adic vanishing theorem of Garland. Bull. Soc. AMS **80**, 1001–1004 (1974)
13. Casselman, W.: A new non-unitarity argument for p-adic representations. J. Fac. Sci. Univ. Tokyo **28**, 907–928 (1981)
14. Curtis, C.W., Lehrer, G.I., Tits, J.: Spherical Buildings and the Character of the Steinberg Representation. Invent. Math. **58**, 201–210 (1980)
15. Deligne, P.: La conjecture de Weil II. Publ. Math. IHES **52**, 137–252 (1980)
16. Drinfel'd, V.G.: Elliptic modules. Math. USSR Sbornik **23**, 561–592 (1974)
17. Drinfel'd, V.G.: Coverings of p-adic symmetric regions. Funct. Anal. Appl. **10**, 107–115 (1976)
18. Fresnel, J., van der Put, M.: Géométrie analytique rigide et applications. Boston: Birkhäuser 1981
19. Garland, H.: p-adic curvature and the cohomology of discrete subgroups of p-adic groups. Ann. Math. **97**, 375–423 (1973)
20. Godement, R.: Topologie algébrique et théorie des faisceaux. Paris: Hermann 1964
21. Goldmann, O., Iwahori, N.: The space of $\wp$-adic norms. Acta Math. **109**, 137–177 (1963)
22. Hartshorne, R.: Residues and duality. Lect. Notes Math., Vol. 20. Berlin Heidelberg New York: Springer 1966
23. Hartshorne, R.: On the de Rham cohomology of algebraic varieties. Publ. Math. IHES **45**, 5–99 (1975)
24. Hiller, H.: Geometry of Coxeter groups. Boston: Pitman 1982
25. Hyodo, O.: On the de Rham–Witt complex attached to a semistable family. Preprint 1988

26. Iwahori, N.: Generalized Tits system (Bruhat decomposition) on $\wp$-adic semisimple groups. In: Algebraic groups and discontinuous subgroups. Proc. Symp. Pure Math. **9**, 71–83 (1966)
27. Jensen, C.U.: Les foncteurs dérivés de $\varprojlim$ et leurs applications en théorie des modules. Lect. Notes Math., vol. 254. Berlin Heidelberg New York: Springer 1972
28. Kiehl, R.: Der Endlichkeitssatz für eigentliche Abbildungen in der nichtarchimedischen Funktionentheorie. Invent. Math. **2**, 191–214 (1967)
29. Kiehl, R.: Theorem A und Theorem B in der nichtarchimedischen Funktionentheorie. Invent. Math. **2**, 256–273 (1967)
30. Kurihara, A.: Construction of p-adic unit balls and the Hirzebruch proportionality. Am. J. Math. **102**, 565–648 (1980)
31. MacLane, S.: Homology. Berlin Heidelberg New York: Springer 1975
32. Mehlmann, F.: Ein Beweis für einen Satz von Raynaud über flache Homomorphismen affinoider Algebren. Schriftenr. Math. Inst. Univ. Münster, 2. Serie, Heft 19. Münster: 1981
33. Milne, J.S.: Etale cohomology. Princeton University Press 1980
34. Morita, Y.: Analytic representations of SL_2 over a $\wp$-adic number field, II. Automorphic forms of several variables, Taniguchi Symp. 1983. Progr. Math. **46**, 282–297 (1984)
35. Morita, Y.: Analytic representations of SL_2 over a $\wp$-adic number field, III. Automorphic forms and number theory. Adv. Stud. Pure Math. **7**, 185–222 (1985)
36. Morita, Y., Schikhof, W.: Duality of projective limit spaces and inductive limit spaces over a nonspherically complete nonarchimedean field. Tôhoku Math. J. **38**, 387–397 (1986)
37. Mumford, D.: An analytic construction of degenerating curves over complete local rings Compos. Math. **24**, 129–174 (1972)
38. Mumford, D.: Abelian varieties. Oxford University Press 1974
39. Mustafin, G.A.: Nonarchimedean uniformization. Math. USSR Sbornik **34**, 187–214 (1978)
40. van der Put, M.: A note on p-adic uniformization. Proc. Kon. Ned. Akad. Wet. A **90**, 313–318 (1987)
41. Rapoport, M.: On the bad reduction of Shimura varieties. In: Clozel, L., Milne, J.S. (Eds.) Automorphic forms, Shimura varieties, and L-functions II. Perspect. Math. 11, pp. 253–321. Boston: Academic Press 1990
42. Raynaud, M.: Geometrie analytique rigide d'apres Tate, Kiehl. Bull. Soc. Math. France, Mémoire **39–40**, 319–327 (1974)
43. Robert, A.: Représentations p-adiques irréductibles de sous-groupes ouverts de $SL_2(\mathbb{Z}_p)$. C.R. Acad. Sci. Paris **298**, 237–240 (1984)
44. Schneider, P.: Rigid-analytic L-transforms. In: Number Theory, Noordwijkerhout 1983. Lect. Notes Math., vol. 1068, pp. 216–230. Berlin Heidelberg New York: Springer 1984
45. Segal, G.: Classifying spaces and spectral sequences. Publ. Math. IHES **34**, 105–112 (1968)
46. de Shalit, E.: Eichler cohomology and periods of modular forms on p-adic Schottky groups. J. Reine angew. Math. **400**, 3–31 (1989)
47. Spanier, E.H.: Algebraic topology. New York: McGraw-Hill 1966
48. Stuhler, U.: Über die Kohomologie einiger arithmetischer Varietäten I. Math. Ann. **273**, 685–699 (1986)
49. Teitelbaum, J.: Values of p-adic L-functions and a p-adic Poisson kernel. Invent. Math. **101**, 395–410 (1990)

SGA 1 Grothendieck, A.: Revêtements etales et groupe fondamental. Lect. Notes Math., vol. 224. Berlin Heidelberg New York: Springer 1971
SGA 3 Demazure, M., Grothendieck, A.: Schémas en Groupes I. Lect. Notes Math., vol. 151. Berlin Heidelberg New York: Springer 1970
SGA 4 Artin, M., Grothendieck, A., Verdier, J.L.: Théorie des topos et cohomologie etale des schémas. Lect. Notes Math., vol. 270. Berlin Heidelberg New York: Springer 1972

Invent. math. 105, 123–136 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Invariant measures exist under a summability condition for unimodal maps

Tomasz Nowicki[1,*] **and Sebastian van Strien**[2]

[1] Mathematics Department, Warsaw University, Poland
[2] Mathematics Department, TU Delft, The Netherlands

Oblatum 7-V-1990 & 19-XI-1990

Summary. For unimodal maps with negative Schwarzian derivative a sufficient condition for the existence of an invariant measure, absolutely continuous with respect to Lebesgue measure, is given. Namely the derivatives of the iterations of the map in the (unique) critical value must be so large that the sum of (some root of) the inverses is finite.

1 Introduction and statement of results

The aim of this paper is to introduce a new, very weak, condition which guarantees the existence of an invariant probability measure, which is absolutely continuous with respect to the Lebesgue measure (acim in short), for unimodal maps of the interval. We believe this condition is so weak that it is even equivalent to the existence of acim's.

It is well known that there are three possibilities for such a map;

- f has a periodic attractor;
- there exist arbitrarily small intervals which are mapped into themselves by some iterate of f;
- the non-wandering set of f contains intervals.

In the last case such a map may have an acim. The purpose of such a measure is to describe the statistical properties of orbits: the frequency with which a trajectory falls into a set is given by the measure of this set. For several years it was believed that such maps have automatically an acim, but as was shown by Johnson [Jo] this is not true.

In general, the existence of an absolutely continuous measure of some interval map is related to the amount of expansion this map has. Indeed, if a map $f\colon [0, 1] \to [0, 1]$ is everywhere expanding (and therefore not smooth) then it has an acim [L-Y]. However, if the map has a critical point there is no universal

* Partially supported by the NWO grant.

expansion. Nevertheless by some analytical means one can estimate the counterplay between the contraction ruled by the derivative near the critical point and the expansion ruled by the derivative near the critical value [Ja, M, Sz, C-E, N1, N2, K1, K2, N-S, Str2]. In these results the expansion near the critical value was assumed to be exponential.

In what follows we prove that some weaker expansion is sufficient, and we conjecture that a condition similar in nature to ours is also necessary.

We shall deal with C^3 maps f of the interval $[0, 1]$ into itself with negative Schwarzian derivative, i.e., $Sf = f'''/f' - 3/2(f''/f')^2 < 0$. The maps f we deal with are unimodal. By this we mean that there is a unique $c \in [0, 1]$ such that $Df > 0$ on $[0, c)$ and $Df < 0$ on $(c, 1]$, where Df denotes the derivative of f. This point c is called the critical point of f. We say that the critical point c has order l if there are constants O_1, O_2 so that

$$O_1|x - c|^{l-1} \leqq |Df(x)| \leqq O_2|x - c|^{l-1} \tag{NF}$$

As usual let f^n be the n-th iterate of f and let $c_1 = f(c)$. Furthermore denote the Lebesgue measure of a measurable set I by $|I|$.

Main Theorem *Suppose that f is unimodal, C^3, has negative Schwarzian derivative and that the critical point of f is of order $l \geqq 1$. Moreover assume that the growth-rate of $|Df^n(c_1)|$ is so fast that*

$$\sum_{n=0}^{\infty} |Df^n(c_1)|^{-1/l} < \infty$$

holds. Then f has a unique absolutely continuous invariant probability measure μ which is ergodic and of positive entropy. Furthermore there exists a positive constant K such that

$$\mu(A) \leqq K|A|^{1/l},$$

for any measurable set $A \subset (0, 1)$. Finally, the density ρ of the measure μ with respect to the Lebesgue measure is a $L^{\tau-}$ function where $\tau = l/(l-1)$, $L^{\tau-} = \bigcup_{1 \leqq t < \tau} L^t$ and $L^t = \{\rho \in L^1 : \int |\rho|^t dx < \infty\}$.

It is not hard to show that there exist many parameters a for which the quadratic map $f(x) = ax(1 - x)$ satisfies the assumption of this theorem and not the condition that $|Df^n(c_1)|$ grows exponentially (which was introduced in [C-E]). So the condition from this paper is much weaker than the well-known Collet-Eckmann condition. Benedicks and Young [B-Y] proved the existence of acim's for maps with a non-flat critical point for which $|Df^n(c_1)|$ is at least $e^{\alpha\sqrt{n}}$ and for which moreover the distance of $f^n(c_1)$ to c is at least of the form $e^{-\alpha n}$. Clearly our result implies theirs. Moreover, we think that our proof is simpler than theirs.

Since the measure μ from this theorem is absolutely continuous, $\lambda_\mu := \int \log|Df| \, d\mu$ is strictly positive and since μ is ergodic this implies

$$\lim_{n \to \infty} \frac{1}{n} \log|Df^n(x)| = \lambda_\mu > 0 \quad \text{for } \mu - \text{almost all } x .$$

Of course the estimate $\mu(A) \leqq K|A|^{1/l}$ shows that the poles of the invariant measure μ are at most of the form $|x - x_0|^{1/l-1}$. It is not hard to show that any absolutely continuous invariant probability measure has a pole of this order at the critical values $f^n(c)$, $n \geqq 1$, and therefore this estimate is optimal. Even for maps for

which $|Df^n(c_1)|$ grows exponentially this result is new (the results in [C-E] and [N-S] only give some bounds for the order of the poles). Notice that the density of the invariant measure is always a $L^{\tau-}$ function where $\tau = l/(l-1)$, independently of the size of $\sum_{n=0}^{\infty} |Df^n(c_1)|^{-1/l} < \infty$. We conjecture that these maps either have an absolutely continuous probability invariant measure with a $L^{\tau-}$ density or do not have a finite absolutely continuous invariant measure at all. Finally it is a pleasure to thank Gerhard Keller, A.O. Lopes and P. Thieullen for some very useful comments.

2 A reformulation of the Main Theorem and an outline of its proof

In [BL1] and [BL2] it is shown that any unimodal map with negative Schwarzian derivative is ergodic (w.r.t. to the Lebesgue measure) and that any absolutely continuous invariant probability measure μ has positive metric entropy. More precisely, as was shown in [BL1] and [BL2] any forward invariant set of positive Lebesgue measure has the critical point as a density point. From this the ergodicity of any absolutely continuous invariant measure μ follows immediately. Furthermore if the entropy of the measure μ were zero, then f would be μ-almost everywhere invertible. But this would imply that the support of μ could have at most density $1/2$ at the critical point of f, a contradiction with the previous statement. Therefore, in order to prove the Main Theorem it is enough to establish the existence of an absolutely continuous invariant probability measure μ.

In order to prove the existence of this invariant measure we will use the strategy of [N-S]. Usually invariant measures are constructed by considering iterations of the Perron-Frobenius operator. This operator associates to the density of a measure ν the density of $f_* \nu$. Of course $f_* \nu$ will have poles at the critical values of f even if ν does not. Therefore in order to show that iterations of the Perron-Frobenius operator (i.e. the densities $f_*^n \nu$) have a nice limit density, one has to choose a good 'topology' on a space of densities with infinitely many poles. In some cases one chooses L^p spaces, in other cases spaces with weighted norms.

Rather than to look at the densities of $f_*^n \nu$, in [N-S] it was proposed to compare the measures $f_*^n \nu$ with the Lebesgue measure. More precisely, using general arguments one can show that f has an absolutely continuous invariant probability measure provided that for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any measurable set A with $|A| < \delta$ one has that $|f^{-n}(A)| < \varepsilon$ for all $n > 0$. In fact in this paper we will prove the following more precise statement: there exists a constant K such that for every n and every measurable set A,

$$|f^{-n}(A)| < K|A|^{1/l}. \tag{1}$$

Let us first explain why (1) implies that f has an absolutely continuous invariant probability measure with a $L^{\tau-}$ density where $\tau = l/(l-1)$. For simplicity assume that $|I| = 1$ and let λ be the Lebesgue measure on I. Define $\lambda_n(A) = |f^{-n}(A)|$ (which is nothing but the probability measure $f_*^n \lambda$ from above when λ is the Lebesgue measure) and let $\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} \lambda_i$, i.e., $\mu_n(A) = \frac{1}{n} \sum_{i=0}^{n-1} |f^{-i}(A)|$. Since the space of probability measures on I is compact (with respect to the weak topology), there exists a sequence $n_i \to \infty$ and a probability measure μ such that μ_{n_i} converges weakly to μ. From the definition of μ_n it follows easily that μ is invariant, $\mu(f^{-1}(A)) = \mu(A)$, and from (1) one has that $\mu(A) \leqq K|A|^{1/l}$ for each measurable set A. Hence μ is absolutely continuous. Let ρ be the density of μ with respect to the

Lebesgue measure, i.e., $\mu(A) = \int_A \rho(x)dx$. Informally speaking, the inequality $\mu(A) \leqq K|A|^{1/l}$ implies that the poles of the density can be at worse of the form $x^{1/l-1}$: if $d\mu \approx x^{1/l-1}dx$ then integrating gives $\int_0^\varepsilon x^{1/l-1}dx = \varepsilon^{1/l}$. So one expects the density to be in the space $L^{\tau-}$ where $\tau = l/(l-1)$. Let us make this argument precise. Take $t \geqq 1$ and $C_k = \{x; k \leqq \rho^t(x) \leqq k+1\}$ and $D_k = \bigcup_{l=k}^{\infty} C_l = \{x; \rho^t(x) \geqq k\}$. Since $\mu(A) \leqq K|A|^{1/l}$, $k^{1/t}|D_k| \leqq \int_{D_k} \rho dx = \mu(D_k) \leqq K \cdot |D_k|^{1/l}$ and therefore

$$|D_k| \leqq K' \cdot k^{-\frac{l}{(l-1)t}}.$$

Hence

$$\int \rho^\tau dx \leqq \sum_{k=0}^{\infty} (k+1)|C_k| = 1 + \sum_{k=0}^{\infty} |D_k|,$$

i.e.,

$$\int \rho^\tau dx \leqq 1 + \sum_{k=0}^{\infty} K' \cdot k^{-\frac{l}{(l-1)t}} < \infty$$

whenever $t < \tau = l/(l-1)$. This shows that $\rho \in L^{\tau-}$.

One of the main results in [N-S] was to show that (1) can be deduced from the following: there exists a constant K' such that for any n and every $\varepsilon > 0$,

$$|f^{-n}(c_1 - \varepsilon, c_1)| < K'\varepsilon^{1/l} \tag{2}$$

where l is the order of the critical point of f. Because of the non-flatness condition at the critical point this is equivalent to: there exists a constant K'' such that for every $n > 0$ and every $\varepsilon > 0$

$$|f^{-n}(c - \varepsilon, c + \varepsilon)| < K''\varepsilon.$$

From all this it follows that the Main Theorem can be deduced from

Theorem A *Suppose that f is unimodal, C^3, has negative Schwarzian derivative and that the critical point of f is of order $l \geqq 1$. Moreover assume that*

$$\sum_{n=0}^{\infty} |Df^n(c_1)|^{-1/l} < \infty$$

holds. Then there exists a constant $K < \infty$ such that for each $\varepsilon > 0$,

$$|f^{-n}(c - \varepsilon, c + \varepsilon)| < K\varepsilon. \tag{3}$$

Let us say a few words about the proof of inequality (3). The main idea in [N-S] was to show that $f^{-n}(c - \varepsilon, c + \varepsilon)$ is contained in the union of sets of the form

$$f^{-(n-k)}\left(c - \frac{\varepsilon}{|Df^k(c_1)|^{1/l}}, c + \frac{\varepsilon}{|Df^k(c_1)|^{1/l}}\right),$$

where k runs over all integers such that $n - k > \log(\varepsilon)$. In this way it was possible to prove inductively that there exists $\delta > 0$ such that $|f^{-n}(c - \varepsilon, c + \varepsilon)| \leqq \varepsilon^\delta$ using the exponential growth of $|Df^k(c_1)|$.

In this paper a more refined version of this strategy is chosen. It is proved that each component of $f^{-n}(c - \varepsilon, c + \varepsilon)$ is either contained in or at least can be compared in size (this process we will call 'sliding') with a set of the form

$$f^{-(n-k)}\left(c - \frac{\varepsilon}{|Df^k(c_1)|^{1/l}}, c + \frac{\varepsilon}{|Df^k(c_1)|^{1/l}}\right).$$

Using this and the summability condition, inequality (3) will then be proved by induction.

3 Some notation and some prerequisites

Throughout this paper we will mean by (α, β) the interval with endpoints α and β irrespective of their ordering.

In this paper we assume that f has negative Schwarzian derivative, i.e., that $Sf = f'''/f' - 3/2\,(f''/f')^2 < 0$. From this it easily follows $Sf^n < 0$ and that $|Df^n|$ has no positive local minima. More generally such maps f have the following properties. A proof of the first property can be found in for example [M.S.1]. The second property was proved first used in [Str1], but see also [Str2], [M.S.2], [M.M.S.].

i (The cross-ratio is expanded) *For every k and every $\alpha < \beta < \gamma < \delta$ if $Df^k_{|(\alpha,\delta)} \neq 0$ then*

$$\frac{|f^k(I)|}{|I|}\frac{|f^k(T)|}{|T|} > \frac{|f^k(L)|}{|L|}\frac{|f^k(R)|}{|R|}, \tag{EC}$$

where $L = (\alpha, \beta)$, $I = (\beta, \gamma)$, $R = (\gamma, \delta)$, $T = (\alpha, \delta)$.

ii (Koebe Lemma) *For every $\tau > 0$ there exists a constant $\mathcal{C} = \mathcal{C}(\tau, KL) > 0$ such that for every k and every $\alpha < x < \beta$ with $Df^k_{|(\alpha,\beta)} \neq 0$ if $|f^k(\alpha, x)| > \tau |f^k(\alpha, \beta)|$ then*

$$|Df^k(x)| > \mathcal{C}(\tau, KL)|Df^k(\beta)|\,. \tag{KL}$$

In the proof we shall use the following convention. The symbol $\mathcal{C}$ will describe various finite and positive universal constants (i.e., they are independent on the iterate of f). All the estimates in this paper are based on the non-flatness condition, the fact that the cross-ratio is expanded or on the Koebe Lemma. $\mathcal{C}(NF)$ denotes a constant which is based on the non-flatness condition and $\mathcal{C}(\tau, KL)$ is the constant from the Koebe Lemma.

4 Branches which will be 'slided' later on

Let f^n be monotone on an interval I and assume that $f^n(I) = (c - \varepsilon, c + \varepsilon)$. Let T be the largest interval containing I on which f^n is monotone and lable the endpoints α and β of T so that $|f^n(\alpha) - c| \leqq |f^n(\beta) - c|$. Denote the endpoints of I by γ and δ so that either $\alpha < \gamma < \delta < \beta$ or $\alpha > \gamma > \delta > \beta$. In this section we will assume that

$$|f^n(\alpha) - f^n(\gamma)| \geqq 2\varepsilon\,. \tag{*}$$

Let

$$A_0 = f^n(\alpha, \gamma),\ I_0 = f^n(I),\ R_0 = f^n(\delta, \beta)\,.$$

By (*) one has

$$|R_0| \geqq |A_0| \geqq |I_0|\,. \tag{**}$$

Later on we shall show that if $|R_0|$ is not too large compared to $|A_0|$ then the set $I_0 = f^n(I)$ can be 'slided'. In this section we will show that if I_0 cannot be 'slided' at least some smaller iterate $f^{k_s}(I)$ of I can be 'slided'. If $|A_0 \cup I_0| \geqq |R_0|$ then set

$s = 0$ and we are finished. Otherwise we shall define inductively a finite sequence of intervals $T^i = [\alpha_i, \alpha_{i-1}]$ and integers n_i as follows. Let $n_0 = n$, $\alpha_0 = \alpha$, $\alpha_{-1} = \beta$ and $T^0 = [\alpha_0, \alpha_{-1}] = [\alpha, \beta]$ (i.e. $T^0 = T$). By maximality of T one can choose n_1 such that $0 < n_1 < n$ and $f^{n_1}(\alpha_0) = c$. Now choose α_1 such that T^1 is the maximal interval of the form $T^1 = [\alpha_1, \alpha_0]$ which contains T^0 and on which f^{n_1} is monotone (of course one may have $T^1 = T$). Now assume that n_{i-1} and $T^{i-1} = [\alpha_{i-2}, \alpha_{i-1}]$ are defined. Then simply define $n_i < n_{i-1}$ such that $f^{n_i}(\alpha_{i-1}) = c$, and let T^i be the maximal interval of the form $[\alpha_i, \alpha_{i-1}]$ which contains $T^{i-1} = [\alpha_{i-2}, \alpha_{i-1}]$ and on which f^{n_i} is monotone. It follows that for $i \geqq 2$, T^i and T^{i-1} have precisely one common boundary point and that

$$I \subset T^0 \subset \ldots \subset T^i .$$

Let us now define the integers k_i and intervals I_i, R_i, A_i, L_i as follows:

$$k_i = n_i - n_{i+1}, I_i = f^{n_i}(I) ,$$

$$R_i = f^{n_i}(\alpha_{i-1}, \gamma) \backslash I_i, A_i = f^{n_i}(\alpha_i, \gamma) \backslash I_i, L_i = f^{n_i}(\alpha_{i-2}, \gamma) \backslash I_i .$$

In other words R_i is the component of $f^{n_i}(T^i \backslash I)$ which contains c and A_i is the other component. Furthermore L_i is contained in A_i and

$$f^{k_i}(I_{i+1}) = I_i, f^{k_i}(R_{i+1}) = A_i, f^{k_i}(L_{i+1}) = R_i ,$$

for all $i = 0, \ldots, s-1$. We stop the construction at $i = s$, when

$$|A_s \cup I_s| \geqq |R_s| . \tag{***a}$$

In particular

$$|A_i \cup I_i| \leqq |R_i|, \quad \text{for } i = 0, 1, \ldots, s-1 . \tag{***b}$$

The main result of this section is the following

Proposition 4.1 *There exist constants K', K'' such that*

$$|I_s| \leqq K' \frac{|f^n(I)|}{\prod_{i=0}^{s-1} K'' |Df^{k_i}(c_1)|^{1/l}} .$$

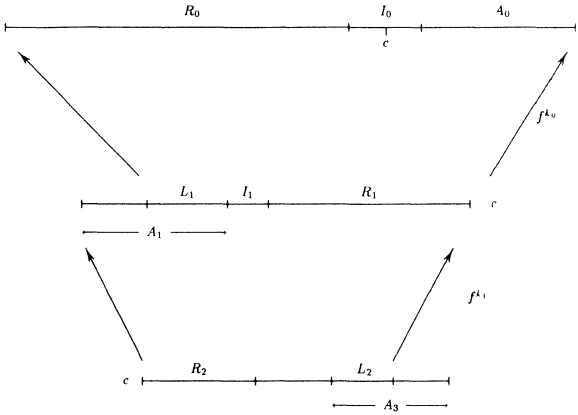


Fig. 1. The intervals R_i, I_i, L_i and A_i

In order to prove this proposition we need some lemmas:

Lemma 4.2

$$|I_s| \leqq |I_1| \frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \prod_{i=1}^{s-1} \frac{|R_{i+1}|}{|R_i|} .$$

Proof. By (EC) we have

$$\frac{|f^{k_1}(L_{i+1})|}{|L_{i+1}|} \frac{|f^{k_1}(R_{i+1})|}{|R_{i+1}|} < \frac{|f^{k_1}(I_{i+1})|}{|I_{i+1}|} \frac{|f^{k_1}(L_{i+1} \cup I_{i+1} \cup R_{i+1})|}{|L_{i+1} \cup I_{i+1} \cup R_{i+1}|} .$$

By definition of the sequences of intervals A, L, I, R this is equivalent to

$$|I_{i+1}| \leqq |I_i| \frac{|L_{i+1}|}{|A_i|} \frac{|R_i \cup I_i \cup A_i|}{|R_{i+1} \cup I_{i+1} \cup L_{i+1}|} \frac{|R_{i+1}|}{|R_i|} .$$

By induction we get

$$|I_s| \leqq |I_1| \frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \times \prod_{i=1}^{s-1} \frac{|R_{i+1}|}{|R_i|} \times \prod_{i=2}^{s-1} \left(\frac{|L_i|}{|A_i|} \frac{|R_i \cup I_i \cup A_i|}{R_i \cup I_i \cup L_i|} \right) \times \frac{|L_s|}{|L_s \cup I_s \cup R_s|} .$$

The last factor is clearly less than 1, and we can say the same about the last but one factor because these terms are of the form $\dfrac{l(a+w)}{a(l+w)}$ and because $l(a+w) < a(l+w)$ for positive l, a, w and $l < a$. □

Lemma 4.3 *Assume that f^k is a diffeomorphism on (c, w) and that for some $z \in (c, w)$ one has $f^k(z) = c$ and $|f^k(c, z)| < \tau |f^k(z, w)|$ for some $\tau \in (0, 1)$. Then*

$$\frac{|f^k(c, z)|}{|(c, z)|} \geqq \mathcal{O}(\tau) |Df^k(c_1)|^{1/l} .$$

Proof. Using the chain-rule, the non-flatness condition, (KL) and $f^k(z) = c$ one has

$$\begin{aligned} |Df^k(c_1)| &= |Df(f^{k-1}(c_1))| |Df^{k-1}(c_1)| \\ &\leqq \mathcal{O}(NF) |f^{k-1}(c_1) - c|^{l-1} |Df^{k-1}(c_1)| \\ &\leqq \mathcal{O}(NF) \mathcal{O}(KL, \tau) |f^{k-1}(c_1) - c|^{l-1} \frac{|f^{k-1}(f(c, z))|}{|f(c, z)|} \\ &= \mathcal{O}(NF) \mathcal{O}(KL, \tau) \frac{|f^k(c, z)|^l}{|f(c, z)|} . \end{aligned}$$

Using again the non-flatness condition gives the required estimate. □

Lemma 4.4 *Assume that f^k is a diffeomorphism on (c, z), that $f^k(z) = c$ and that for some $y \in (c, z)$ one has $|f^k(c, y)| < \tau |f^k(y, z)|$ for some $\tau \in (0, 1)$. Then*

$$\frac{|f^k(y, z)|}{|(c, y)|} \geqq \mathcal{O}(\tau) |Df^k(c_1)|^{1/l} .$$

Proof. From the non-flatness condition and since $|f^k(y, z)| \geqq \dfrac{1}{\tau} |f^k(c, y)|$ one has

$$\left(\frac{|f^k(y, z)|}{|c, y|} \right)^l \geqq \mathcal{O}(NF) \frac{|f^k(y, z)|^{l-1} |f^k(y, z)|}{|f(c, y)|} \geqq \frac{\mathcal{O}(NF)}{\tau} |f^k(y, z)|^{l-1} \frac{|f^k(c, y)|}{|f(c, y)|} .$$

Since $|f^k(c,y)| < \tau |f^k(y,z)|$ one gets from (KL) that the last factor is at least $\mathcal{O}(\tau, KL)|Df^{k-1}(c_1)|$. Moreover one has $|f^k(y,z)| \geqq \frac{1}{1+\tau}|f^k(c,z)| = \frac{1}{1+\tau}|f^k(c)-c|$. From all this

$$\left(\frac{|f^k(y,z)|}{|(c,y)|}\right)^l \geqq \mathcal{O}|f^k(c)-c|^{l-1}|Df^{k-1}(c_1)| .$$

By the non-flatness condition $|f^k(c)-c|^{l-1} \geqq \mathcal{O}(NF)|Df(f^k(c))|$, and therefore the lemma follows. □

Lemma 4.5 *If $s > 0$ then there exists a constant $K < \infty$ such that*

$$|I_1|\frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \leqq K \frac{|f^n(I)|}{|Df^{k_0}(c_1)|^{1/l}},$$

$$\frac{|f^{k_0}(R_1)|}{|R_1|} \geqq \frac{1}{K}|Df^{k_0}(c_1)|^{1/l},$$

$$\frac{|R_i|}{|R_{i+1}|} \geqq \frac{1}{K}|Df^{k_i}(c_1)|^{1/l} \text{ for } i = 1, \ldots, s-1$$

Proof. By (EC) we have

$$|I_1|\frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \leqq |f^{k_0}(I_1)|\frac{|R_1|}{|f^{k_0}(R_1)|}\frac{|f^{k_0}(A_1 \cup I_1 \cup R_1)|}{|f^{k_0}(A_1)|} \leqq$$

$$\leqq |f^n(I)|\frac{|R_1|}{|f^{k_0}(R_1)|}\left(1+\frac{|I_0 \cup A_0|}{|f^{k_0}(L_1)|}\right)$$

$$\leqq |f^n(I)|\frac{|R_1|}{|f^{k_0}(R_1)|}\left(1+\frac{|I_0 \cup A_0|}{|R_0|}\right).$$

By $(**)$ this gives

$$|I_1|\frac{|A_1 \cup I_1 \cup R_1|}{|A_1|} \leqq |f^n(I)|\frac{|R_1|}{|f^{k_0}(R_1)|}\cdot 3 .$$

It is therefore enough to prove that $|f^{k_0}(R_1)|/|R_1| > \mathcal{O}|Df^{k_0}(c_1)|^{1/l}$. So let us prove this. One has $f^{k_0}(R_1) = A_0$. Let $R_1' \subset R_1 \cup I_1$ be the smallest interval containing R_1 such that $f^{k_0}(R_1')$ contains c. We want to apply Lemma 4.3 by taking (z,w) to be the interval T_1 and (c,z) the interval R_1'. Since $f^{k_0}(T_1 \setminus R_1') \supset f^{k_0}(L_1) = R_0$ and since $s > 0$ we get from $(***\text{b})$,

$$|f^{k_0}(T_1 \setminus R_1')| \geqq |R_0| \geqq |A_0 \cup I_0| \geqq |f^{k_0}(R_1')| .$$

From all this it follows that we can apply Lemma 4.3 and get that

$$|f^{k_0}(R_1')|/|R_1'| > \mathcal{O}|Df^{k_0}(c_1)|^{1/l} .$$

But since $|f^{k_0}(R_1')| = |A_0| + \varepsilon < 2|A_0| = 2|f^{k_0}(R_1)|$, this implies

$$\frac{|f^{k_0}(R_1)|}{|R_1|} \geqq \frac{1}{2}\frac{|f^{k_0}(R_1')|}{|R_1'|} > \frac{\mathcal{O}}{2}|Df^{k_0}(c_1)|^{1/l} .$$

So the first two statements of the lemma follow.

To prove the third statement let as before $R'_{i+1} \subset R_{i+1} \cup I_{i+1}$ be the smallest interval containing R_{i+1} such that $f^{k_i}(R'_{i+1})$ contains c. Because the construction did not stop at i, $|f^{k_i}(L_{i+1})| = |R_i| \geqq |A_i \cup I_i| \geqq |f^{k_i}(R'_{i+1})|$. Therefore one can apply Lemma 4.4 and we get

$$\frac{|R_i|}{|R_{i+1}|} \geqq \frac{|R_i|}{|R'_{i+1}|} = \frac{|f^{k_i}(L_{i+1})|}{|R'_{i+1}|} \geqq \mathcal{O}|Df^{k_i}(c_1)|^{1/l} .$$

This proves the third statement of the lemma. □

Proof of Proposition 4.1 This follows immediately from Lemmas 4.2, 4.5. □

5 Estimates

In this section we shall prepare the estimates for the preimages of the intervals around the critical point c. So consider the set $E(\delta) = (c - \delta, c + \delta)$ and its preimages $E_n(\delta) = f^{-n}(E(\delta))$. For a given $\varepsilon > 0$ we shall subdivide the collection of components of $E_n(\varepsilon)$ into three subcollections.

Let σ be some positive number and let $\varepsilon \in (0, \frac{1}{2}\sigma)$. We define $\nu(\sigma)$ as $\inf\{k > 0 : |f^k(c) - c| < \sigma\}$. Clearly $\nu(\sigma)$ is monotone and $\nu(\sigma)$ tends to infinity as $\sigma \to 0$. Later on we shall choose σ appropriately.

Let I be a component of $E_n(\varepsilon)$. Suppose that $I \subset I' \subset I''$, where I' is a component of $E_n(2\varepsilon)$ and I'' is a component of $E_n(\sigma)$. If $Df^n_{|I''} \neq 0$ then I belongs to the collection $\mathcal{R}_n$. If $I \notin \mathcal{R}_n$ but $Df^n_{|I'} \neq 0$ then I belongs to the collection $\mathcal{S}_n$. All the other components form the collection $\mathcal{T}_n$.

5.1 The collection $\mathcal{R}_n$, the regular case

If $I \in \mathcal{R}_n$, then f^n is a diffeomorphism on I and there exists $\gamma \in I$ such that $f^n(\gamma) = c$. Let (α, β) be the maximal interval containing I on which f^n is a diffeomorphism. By definition of $\mathcal{R}_n$ we have $|f^n(\alpha, \gamma)|, |f^n(\beta, \gamma)| \geqq \sigma$. Therefore we can use (KL) on I and obtain:

Proposition 5.1 *There exists a constant K_R such that for $\varepsilon < \sigma/2$ and any regular component I as above, one has*

$$\frac{|I|}{|(\alpha, \beta)|} \leqq K_R \frac{\varepsilon}{\sigma}.$$

Proof. By (KL) one has $|Df^n(x)| \leqq \mathcal{O}(KL)|Df^n(y)|$ for any $x \in (\alpha, \beta)$ and $y \in I$. Therefore

$$\frac{\sigma}{|(\alpha, \beta)|} \leqq \frac{|f^n(\alpha, \beta)|}{|(\alpha, \beta)|} \leqq \mathcal{O}\frac{|f^n(I)|}{|I|} \leqq \mathcal{O}\frac{\varepsilon}{|I|} . \qquad \square$$

Corollary. *For $I \in \mathcal{R}_n$ let $\Delta_n(I)$ be the maximal interval on which f^n is a diffeomorphism. By the previous proposition we obtain*

$$\sum_{I \in \mathcal{R}_n} |I| \leqq \sum_{I \in \mathcal{R}_n} \frac{|I|}{|\Delta_n(I)|} |\Delta_n(I)| \leqq K_R \frac{\varepsilon}{\sigma} \sum_{I \in \mathcal{R}_n} |\Delta_n(I)| \leqq K_R \frac{\varepsilon}{\sigma} .$$

5.2 The collection $\mathscr{S}_n$, the case to slide

Let $I \in \mathscr{S}_n$. Then there exists $s \geqq 0$ and a sequence $n_s < n_{s-1} < \ldots < n_0 = n$ as in section 4 such that (in the terminology of that section) $|A_s \cup I_s| \geqq |R_s|$. Since for each $i = 0, \ldots, s$, f^{n_i} is a diffeomorphism on $T^i \supset I$, since $f^{n_i}(T^i) \supset R_i$ and since R_i contains c, there exists an interval $J^i \subset T^i$ such that $G_i = f^{n_i}(J^i)$ contains c and such that

$$|I| = |J^i| .$$

In other words by choosing an appropriate $x_i \in [x_i', c] = R_i \cup I_i$ one can assure that the preimage J^i of $G_i = [x_i, c]$ has the same size as I. Let $G = G_s$ and $J = J^s$. This process we call sliding. Note that $f^{n_i}(J^i)$ contains c. Because $|A_s \cup I_s| \geqq |R_s|$ we can use (KL) and obtain

$$\frac{|f^{n_s}(J)|}{|J|} \leqq \mathcal{O}(KL) \frac{|f^{n_s}(I)|}{|I|}$$

and therefore $|G| \leqq \mathcal{O}|I_s|$. Therefore by Proposition 4.1

$$|G| \leqq \frac{\mathcal{O}K'|f^n(I)|}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{1/l}} \leqq \frac{\mathcal{O}K'2\varepsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{1/l}} .$$

So for each such component I of $E_n(\varepsilon)$, there exists an interval J as above such that $|I| = |J|$ and therefore such that $|I|$ is at most the size of the part of the f^{n_s}-preimage of

$$\left(c - \frac{\mathcal{O}K'2\varepsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{1/l}}, c + \frac{\mathcal{O}K'2\varepsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{1/l}} \right)$$

that is contained in T^s. Now even for a given sequence of $n_s < n_{s-1} < \ldots < n_0 = n$, there may be several such components I in T^s. Even worse, some of these may give the same interval J (or at least overlapping intervals). But for every given sequence of $n_s < n_{s-1} < \ldots < n_0 = n$, there exist at most 2^s different components I of $E_n(\varepsilon)$ of type $\mathscr{S}_n$, such that the corresponding intervals J overlap. Indeed, at the first step of the construction two intervals I and $\tilde{I}$ can only slide onto overlapping intervals J^1 and $\tilde{J}^1$ if there is precisely one turning point of f^{n_1} between these two intervals. Similarly at the i-th step two intervals J^{i-1} and $\tilde{J}^{i-1}$ can only slide onto overlapping intervals J^i and $\tilde{J}^i$ if there is precisely one turning point of f^{n_i} between these two intervals. So at each step the number of intervals $I \in \mathscr{S}_n$ which correspond to overlapping intervals J can at most double. Thus we get

$$\sum_{I \in \mathscr{S}_n} |I| \leqq \sum_{\sum_{j=0}^{s-1} k_j \leqq n} 2^s |f^{-n_s}(c - |G|, c + |G|)|. \qquad (*)$$

Lemma 5.2 *There exists $\sigma_0 > 0$ such that $k_0, \ldots, k_{s-1} \geqq \nu(\sigma)$ for each $\sigma \in (0, \sigma_0)$.*

Proof. Choose $\sigma_0 > 0$ so small that for each $\sigma \in (0, \sigma_0)$ and each $k > \nu(\sigma)$ one has $|Df^k(c_1)|^{1/l} > 2K$ where K is the constant from Lemma 4.5. Because $\nu(\sigma) \to \infty$ as $\sigma \to 0$ and $|Df^k(c_1)| \to \infty$ as $k \to \infty$ this is possible.

By definition of k_0, $f^{k_0}(c)$ is contained in the closure of A_0. Since $I \in \mathscr{S}_n$, at least one critical value of $f^n | T_0$ is cotained in $(c - \sigma, c + \sigma)$. As $T_0 = [\alpha, \beta]$ and as $|f^n(\alpha) - c| \leqq |f^n(\beta) - c|$, this implies $f^n(\alpha) = f^{k_0}(c) \in (c - \sigma, c + \sigma)$. Hence $k_0 \geqq \nu(\sigma)$ and, observing the definition of A_0, $A_0 \subset (c - \sigma, c + \sigma)$.

Next notice that if for some $i = 1, \ldots, s-1$ one has

$$|R_i \cup I_i \cup A_i| \leqq \sigma \text{ implies } k_i \geqq v(\sigma) . \qquad (**)$$

Indeed, c is contained in the closure of $R_i \cup I_i \cup A_i$, this interval has at most length σ and $f^{k_i}(c)$ is contained in A_i. Therefore $k_i \geqq v(\sigma)$ follows from the definition of $v(\sigma)$.

So let us show by induction that for $\sigma \in (0, \sigma_0)$, $|R_i \cup I_i \cup A_i| \leqq \sigma$ for $i = 1, \ldots, s-1$. So assume $s \geqq 2$ and let us first prove that this inequality holds for $i = 1$. From Lemma 4.5, $|f^{k_0}(R_1)| \geqq K|Df^{k_0}(c_1)|^{1/l}|R_1|$. Therefore

$$|R_1 \cup I_1 \cup A_1| \leqq 2|R_1| \leqq 2K \frac{|f^{k_0}(R_1)|}{|Df^{k_0}(c_1)|^{1/l}}$$

$$= 2K \frac{|A_0|}{|Df^{k_0}(c_1)|^{1/l}} \leqq |A_0| ,$$

where the last inequality holds provided $\sigma \in (0, \sigma_0)$ because $k_0 \geqq v(\sigma)$, and $|Df^k(c_1)| > 2K$ for $k \geqq v(\sigma)$. Since $|A_0| \leqq \sigma$, the induction assertion is proved for $i = 1$. Similarly, we get for $i < s$, using the third inequality of Lemma 4.5, that

$$|R_i \cup I_i \cup A_i| \leqq 2|R_i| \leqq 2K \frac{|R_{i-1}|}{|Df^{k_{i-1}}(c_1)|^{1/l}} .$$

From the inductive assumption we know that $|R_{i-1} \cup I_{i-1} \cup A_{i-1}| \leqq \sigma$ and from $(**)$ this implies $k_{i-1} > v(\sigma)$ and so we get again that for $\sigma \in (0, \sigma_0)$ sufficiently small that

$$|R_i \cup I_i \cup A_i| \leqq |R_{i-1}| \leqq |R_{i-1} \cup I_{i-1} \cup A_{i-1}| \leqq \sigma . \qquad \square$$

Therefore one has $k_j \geqq v(\sigma)$. Using this, $(*)$ and the estimate for $|G|$ one gets the following

Proposition 5.3 *There exists a constant K_s such that for $\sigma \in (0, \sigma_0)$,*

$$\sum_{I \in \mathscr{T}_n} |I| \leqq \sum_{\substack{\{k_j \geqq v(\sigma)\} \\ \sum_{j=0}^{s-1} k_j \leqq n}} 2^s \left| f^{-n_s}\left(E\left(K_S \frac{\varepsilon}{\prod_{j=1}^{s-1} K'' |Df^{k_j}(c_1)|^{1/l}} \right) \right) \right| . \qquad \square$$

5.3 The collection $\mathscr{T}_n$, the case to transport

We shall use the idea from [N-S], which is to reduce the estimation of the n-th preimage $I \in \mathscr{T}_n$ to the estimation of some k-th preimage, with $k < n$.

Let $I \in \mathscr{T}_n$ and let I' (resp. I'') be the component of $E_n(2\varepsilon)$ (resp. $E_n(\sigma)$) containing I. By definition f^n has at least one critical point in $I' \supset I$.

Since f^n has a critical point in I' there exists an integer $k < n$ such that $c \in f^k(I')$. Let k be the largest such integer. For simplicity we say that I belongs to the subcollection $\mathscr{T}_n^k$ of $\mathscr{T}_n$. From the properties just stated one has

$$f^{n-k}(c) \in (c - 2\varepsilon, c + 2\varepsilon) .$$

Since $c \notin f^i(I')$ for $i = k+1, \ldots, n-1$, f^{n-k-1} is clearly a diffeomorphism on $f^{k+1}(I')$.

Proposition 5.4 *There exists a constant K_T such that for every n*

$$\sum_{I\in\mathcal{T}_n}|I|=\sum_k\sum_{I\in\mathcal{T}_n^k}|I|\leqq\sum_k\left|f^{-k}\left(E\left(\frac{K_T\varepsilon}{|Df^{n-k}(c_1)|^{1/l}}\right)\right)\right|.$$

Proof. Let $f^{k+1}(I')=(x,c_1]\supset f^{k+1}(I)$. As we saw f^{n-k-1} is a diffeomorphism in (x,c_1). Moreover $f^n(I)\subset(c-\varepsilon,c+\varepsilon)$, $f^{n-k-1}(x)=c\pm 2\varepsilon$ and $f^n(I')\subset(c-2\varepsilon,c+2\varepsilon)$. Therefore one gets from the Koebe Lemma immediately that

$$\frac{|f^n(I)|}{|f^{k+1}(I)|}\geqq\mathcal{O}(KL)|Df^{n-k-1}(c_1)|\,.$$

Hence

$$|f^{k+1}(I)|\leqq\mathcal{O}\frac{\varepsilon}{|Df^{n-k-1}(c_1)|}\,.$$

From the non-flatness condition this gives

$$|f^k(I)|\leqq\mathcal{O}(NF)|f^{k+1}(I)|^{1/l}\leqq\mathcal{O}\left(\frac{\varepsilon}{|Df^{n-k-1}(c_1)|}\right)^{1/l}.$$

Since $f^{n-k-1}(c_1)\in(c-2\varepsilon,c+2\varepsilon)$, the non-flatness condition implies that

$$\left(\frac{\varepsilon}{|Df^{n-k-1}(c_1)|}\right)^{1/l}=\left(\frac{\varepsilon^l}{|\varepsilon^{l-1}Df^{n-k-1}(c_1)|}\right)^{1/l}\leqq\mathcal{O}\frac{\varepsilon}{|Df^{n-k}(c_1)|^{1/l}}\,,$$

i.e.,

$$|f^k(I)\leqq\mathcal{O}\frac{\varepsilon}{|Df^{n-k}(c_1)|^{1/l}}\,.$$

Since $f^n(I)\subset(c-\varepsilon,c+\varepsilon)$ it follows that there exists a constant K_T such that

$$I\subset f^{-k}(f^k(I))\subset f^{-k}E\left(\frac{K_T\varepsilon}{|Df^{n-k}(c_1)|^{1/l}}\right).$$

The proposition follows. □

6 The proof of the Main Theorem

Let σ be fixed so small that for every n,

$$\sum_{\substack{\{k_j\geqq v(\sigma)\},\,1\leqq j\leqq s\\ \sum k_j\leqq n}}3K_S\prod_{j=0}^{s-1}|K''Df^{k_j}(c_1)/2|^{-1/l}\leqq 1\,,$$

and

$$\sum_{k>v(\sigma)}3K_T|Df^k(c_1)|^{-1/l}\leqq 1\,.$$

This is possible by the summability condition, since $v(\sigma)$ tends to infinity as $\sigma\to 0$ and because of the following

Lemma 6.1 *Suppose that $d_k\geqq 0$ and $\sum_{k=0}^{\infty}d_k<\infty$. Then for any $\eta,\psi>0$ there exists a v_0 such that*

$$P=\sum_n\prod_{\substack{\{k_j\geqq v_0\}\\ \sum_{j=0}^{s-1}k_j\leqq n}}(\eta d_{k_j})<\psi\,.$$

Proof. Consider $S_{k_0} = \sum_{k>k_0} \eta d_k$. Then both S_{k_0} and $S = \sum_{s>0} S_{k_0}^s$ tend to zero when k_0 tends to infinity. Clearly $P \leqq S$ which proves the assertion. □

We shall now prove Theorem A in the following formulation:

Theorem B *For any n and $\varepsilon \leqq \sigma/2$*

$$|f^{-n}(c-\varepsilon, c+\varepsilon)| \leqq 3K_R \frac{\varepsilon}{\sigma}.$$

Proof. With the notations from the previous section we have

$$f^{-n}(c-\varepsilon, c+\varepsilon) = \bigcup_{I\in\mathscr{R}_n} I \cup \bigcup_{I\in\mathscr{S}_n} I \cup \bigcup_{I\in\mathscr{T}_n} I,$$

and

$$|f^{-n}(c-\varepsilon, c+\varepsilon)| \leqq \sum_{I\in\mathscr{R}_n} |I| + \sum_{I\in\mathscr{S}_n} |I| + \sum_{I\in\mathscr{T}_n} |I|.$$

Therefore

$$|f^{-n}(c-\varepsilon, c+\varepsilon)| \leqq K_R \frac{\varepsilon}{\sigma} +$$

$$+ \sum_{\substack{\{k_j \geqq \nu(\sigma)\}, 0\leqq j\leqq s-1 \\ \sum k_j \leqq n}} 2^s |f^{-n_s}\left(E\left(K_S \frac{\varepsilon}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{1/l}}\right)\right)| +$$

$$+ \sum_{k\geqq\nu(\sigma)} |f^{-k}\left(E\left(K_T \frac{\varepsilon}{|Df^{n-k}(c_1)|^{1/l}}\right)\right)|.$$

We shall apply the induction. For n small only the first term is non-zero and the assertion of the theorem is true. Suppose that it is true for any $\varepsilon < \sigma/2$ and any $n < N$. Then by the choice of σ, the above formula and the induction assumption we have

$$|f^{-N}(c-\varepsilon, c+\varepsilon)| \leqq K_R \frac{\varepsilon}{\sigma} +$$

$$+ \sum_{\substack{\{k_j \geqq \nu(\sigma)\}, 0\leqq j\leqq s-1 \\ \sum k_j \leqq N}} 2^s K_S \frac{3K_R\varepsilon/\sigma}{\prod_{j=0}^{s-1} K''|Df^{k_j}(c_1)|^{1/l}}$$

$$+ \sum_{\nu(\sigma)\leqq k\leqq N} K_T \frac{3K_R\varepsilon/\sigma}{|Df^{N-k}(c_1)|^{1/l}} \leqq 3K_R \frac{\varepsilon}{\sigma},$$

which completes the proof. □

References

[B-Y] Benedicks, M., Young, L.S.: Absolutely continuous invariant measures and random perturbations for certain one-dimensional maps. Preprint (1990)

[BL1] Blokh, A.M., Lyubich, M.Yu.: Attractors of maps of the interval. Funct. Anal. Appl. **21**, 70–71 (1987) (in Russian)

[BL2] Blokh, A.M., Lyubich, M.Yu.: Measurable dynamics of S-unimodal maps of the interval. Preprint (1989)

[C-E] Collet, P., Eckmann, J.-P.: Positive Liapounov exponents and absolute continuity for maps of the interval. Ergodic Theory Dyn. Syst. **3**, 13–46 (1983)

[Ja] Jakobson, M.: Absolutely continuous measures for one-parameter families of one-dimensional maps. Comm. Math. Phys. **81**, 39–88 (1981)
[Jo] Johnson, S.: Singular measures without restrictive intervals. Comm. Math. Phys. **110**, 655–659 (1987)
[K1] Keller, G.: Invariant measures and Lyapounov exponents for S-unimodal maps. Preprint (1987)
[K2] Keller, G.: Exponents, attractors, and Hopf decompositions for interval maps. Preprint (1988)
[L.Y] Lasota, A., Yorke, J.: On the existence of invariant measures for piecewise monotonic transformations. Trans. A.M.S. **186**, 481–488 (1973)
[M.S.1] de Melo, W., van Strien, S.: A structure theorem in one-dimensional dynamics, Ann. Math. **129**, 519–546 (1989)
[M.S.2] de Melo, W., van Strien, S.: One-dimensional dynamics. to appear
[M.M.S.] Martens, M., de Melo, W., van Strien, S.: Julia-Fatou-Sullivan theory for real one-dimensional dynamics. Acta Math. (To appear)
[M] Misiurewicz, M.: Absolutely continuous measure for certain maps of an interval. Publ. I.H.E.S. **53**, 17–51 (1981)
[N1] Nowicki, T.: Symmetric S-unimodal properties and positive Liapounov exponents, Ergodic Theory Dyn. Syst. **5**, 611–616 (1985)
[N2] Nowicki, T.: A positive Liapounov exponent of the critical value of S-unimodal maps implies uniform hyperbolicity. Ergodic Theory Dyn. Syst. **8**, 425–435 (1988)
[N-S] Nowicki, T., van Strien, S.J.: Absolutely continuous invariant measures for C^2 maps satisfying the Collet-Eckmann conditions. Invent. Math. **93**, 619–635 (1988)
[Str1] van Strien, S.J.: On the creation of horseshoes. Lect. Notes Math. **898**, 316–351 (1981)
[Str2] van Strien, S.J.: Hyperbolicity and invariant measures for general C^2 interval maps satisfying the Misiurewicz condition. Commun. Math Phys. **128**, 437–496 (1990)
[Sz] Szlenk, W.: Some dynamical properties of certain differentiable mappings. Bol. Soc. Mat. Mex., II Ser. **24**, 57–87 (1979)

Note added in proof

It is not hard to show that if a map f as above has an absolutely continuous invariant probability measure μ with $\mu(A) \leqslant k.\,|A|^{1/l}$ for all measurable sets A then the summability condition is also satisfied. We conjecture that a weighted summability condition (where the weights are related to the position of $f^i(c_1)$) is necessary for the existence of absolutely continuous invariant probability measures.

Invent. math. 105, 137–152 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Nonstandard uniformized conformal structures on hyperbolic manifolds

Boris N. Apanasov*

Institute of Mathematics, Siberian branch of Acad Sci USSR, Universitetskii Pr. 4, Novosibirsk, USSR 630090 and
Institut Mittag-Leffler, The Royal Swedish Academy of Sciences, Auravägen 17, S-182 62 Djursholm, Sweden

Oblatum 17-III-1990 & 26-XI-1990

Summary. The main question which the paper discusses is:

Is the space $\mathscr{C}(M^n)$ of uniformized conformal structures of a finite volume (or closed) hyperbolic n-manifold M^n (or its Teichmüller space) connected or not?

For the surface case ($n = 2$) the answer is well known to be in the affirmative. By contrast, for the case $n = 3$, we describe herein some exotic conformal structures on a closed hyperbolic 3-manifold M which are uniformized but cannot be approximated by structures on M obtained from the distinguished conformal (= hyperbolic) structure on M by any of the presently known deformations, meaning by bendings or stampings of M along totally geodesic submanifolds.

1 Introduction

Let $\mathbf{H}^n$ be the subspace

$$\{(x_0, \ldots, x_n) \in \mathbf{R}^{n+1} | q(x_0, \ldots, x_n) = -x_0^2 + x_1^2 + \ldots + x_n^2 = -1, x_0 > 0\} .$$

The quadratic form q restricts to give a positive definite form on each tangent space of $\mathbf{H}^n$ and consequently, endows $\mathbf{H}^n$ with a Riemannian metric. We call this Riemannian manifold hyperbolic n-space. It has constant sectional curvature -1 and is homogeneous. Its group of isometries is the real linear subgroup $SO(n, 1)$ (the component of the identity) of matrices in $SL_{n+1}(\mathbf{R})$ preserving the form q and $\mathbf{H}^n$. A hyperbolic n-manifold M is a complete Riemannian manifold locally isometric to $\mathbf{H}^n$. In fact, such a manifold M is isometric to $\mathbf{H}^n/G_M$, where $G_M \cong \pi_1(M)$ is some discrete, torsion-free subgroup of $SO(n, 1)$ determined by M up to conjugation in $SO(n, 1)$.

* This paper was written while the author was visiting Institut Mittag-Leffler, which he thanks for its hospitality. Financial support from Institut Mittag-Leffler is gratefully acknowledged. The author is thankful to Bruce Palka for his help with the English language and to Karin Lindberg for her excellent typesetting.

Now let us fix an n-manifold M. By a conformal structure (= flat conformal) on M we mean a structure locally modelled on the conformal n-sphere $S^n = \mathbf{R}^n \cup \{\infty\}$, i.e. a $(S^n, \text{Möb}_n)$-structure, where Möb_n is the group of all conformal self-maps of S^n (all compositions of reflections in subspheres, i.e., Möbius transformations). It is well known (see [AP$_1$]) that $\text{Möb}_n \cong SO(n+1, 1)$. In other words, a conformal structure is a maximal atlas (U_i, φ_i) all of whose changes of charts

$$\varphi_i \cdot \varphi_j^{-1}, \text{ for } U_i \cap U_j \neq \varnothing$$

are Möbius transformations on S^n.

We call the conformal structure on M uniformized (cf. [K-P]) if there is a discrete torsion-free subgroup $G \subset \text{Möb}_n$ and a connected component Ω of the discontinuity set $\Omega(G) \subset S^n - G$ is called a Kleinian group if $\Omega(G) \neq \varnothing$ – such that M is conformally equivalent to Ω/G. It is obvious that the natural conformal structure induced on a hyperbolic n-manifold M by its hyperbolic metric is a uniformized structure. Here the uniformization condition is equivalent to non-surjectivity for the developing map $d: \tilde{M} \to S^n$ of the conformal structure on M (see [KM]); in fact, $d(\tilde{M}) = \Omega$. Another kind of uniformed conformal structures on non-hyperbolic 3-manifolds has been considered in [G-L-T] and [KU]. These 3-manifolds are S^1-bundles over compact surfaces and are related to the 3-manifolds at infinity of infinite volume hyperbolic 4-manifolds which are 2-plane bundles over compact surface (compare similar approach in [AP$_3$]; Ch. 9, §5).

We denote the set of classes of uniformized marked conformal structures (Teichmüller space) on the manifold M by $\mathscr{C}(M)$, where marking is introduced as usual using the fundamental group $\pi_1(M)$. Namely, a marked conformal structure on M is a triple (M, c, f) where c is a conformal structure on M and $f: M \to M$ is a diffeomorphism while two triples (M, c, f) and (M, c', f') are called equivalent if there exists a conformal homeomorphism g: $(M, c) \to (M, c')$ such that $g \circ f$ is isotopic to f'. This space $\mathscr{C}(M)$ is naturally identified with the set of conjugacy classes of faithful representations $\varrho: \pi_1(M) \cong G \to SO(n+1, 1)$ with discrete images that act discontinuously somewhere on the n-sphere S^n (see [LK] and [GM]). From this description, there is a natural topology on $\mathscr{C}(M)$. Let

$$R_n(G) = \text{Hom}(G, SO(n+1, 1))$$

be the space of all representations of $G \cong \pi_1(M)$ into $SO(n+1, 1)$. It has the algebraic topology: representations are close if they are close on a finite generating set. Inside $R_n(G)$ we have the subspace $D_n(G)$ of faithful representations with discrete images that are discontinuous on S^n. $SO(n+1, 1)$ acts on $R_n(G)$ by conjugation leaving $D_n(G)$ invariant. The quotient space

$$T_n(G) = D_n(G)/SO(n+1, 1) \subset \text{Hom}(G, SO(n+1, 1))/SO(n+1, 1)$$

is the space of conjugacy classes of discrete discontinuous faithful representations of $G \cong \pi_1(M)$ into $SO(n+1, 1)$. Since we have a natural identification of $\mathscr{C}(M)$ and $T_n(G)$ (via the holonomy representation $d_*: \pi_1(M) \to SO(n+1, 1)$ induced by developing map $d: \tilde{M} \to S^n$ of the universal covering $\tilde{M}$ of M), this yields a topology on $\mathscr{C}(M)$.

As an example, let M be a closed Riemann surface S_g. In this case investigation of conformal structures on the 2-manifold M is equivalent to studying its geometrical structures. For genus $g > 1$ these are just the hyperbolic structures on M. Here

the Teichmüller space T_g of hyperbolic structures on the surface S_g is homeomorphic (due to the Bers' embedding [BR]) to $\mathbf{R}^{6g-6}$. In particular, $\mathscr{C}(S_g)$ is connected.

The basic object of study in this paper is the space $\mathscr{C}(M)$ for a finite volume hyperbolic n-manifold M, $n \geqq 3$. Here the Mostow rigidity [MW] says that the space of hyperbolic structures on M is a point. The first result showing the non-triviality of $\mathscr{C}(M)$ was [AP$_2$], where a smooth curve in $\mathscr{C}(M)$ was constructed. This curve corresponds to the bending deformation (in the sense of W. Thurston [TH]) of the natural conformal structure on M along a totally geodesic hypersurface in M. (See [AP$_4$], [KO], [J-M]; for further development with stamping deformations and historical remarks see [AP$_7$] and [AP$_8$].)

Also we note that, for finite volume hyperbolic n-manifold M, $n \geqq 3$, the space $\mathscr{C}(M)$ is contained in the space $\mathscr{H}^{n+1}(\pi_1(M))$ of $(n+1)$-dimensional hyperbolic structures on the fundamental group of M. The space $\mathscr{H}^{n+1}(\pi_1(M))$ is compact according to a result of J. Morgan [MG]. Two questions which naturally arise are, first, is the space $\mathscr{C}(M)$ compact for $n \geqq 3$, and, secondly, is it connected?

Dealing here with the second question (for first one, see in [AP$_{10}$]), we construct exotic uniformized structures on a closed hyperbolic 3-manifold M with a sufficiently large set of disjoint totally geodesic surfaces that they cannot be a limit of any sequence of uniformized conformal structures on M obtained from the canonical conformal structure on M, the one induced by the hyperbolic metric, either by bending deformations or by stamping (with torsion or without) deformations along totally geodesic submanifolds of M. (At the present time these are the only known deformations of finite volume hyperbolic manifolds.) Our construction of such exotic structures is based on a block-building method for spatial Kleinian groups with wild limit Julia surfaces developed in [A-T] and [AP$_9$]. Moreover, there is some relation between the set of such exotic uniformized structures and a set of non-trivial hyperbolic $(n+1)$-dimensional homology cobordisms; see [AP$_{10}$].

2 Maximal balls and the Nielsen convex hull

Let us consider the Poincaré model of the hyperbolic space $\mathbf{H}^{n+1}$ in the unit ball $B^{n+1} \subset \mathbf{R}^{n+1}$ (or in the half-space $\mathbf{R}^{n+1}_+$), where the isometry group Isom H^{n+1} acts on the boundary sphere $S^n = \partial H^{n+1}$ as the Möbius group Möb$_n$. Given a discrete group $G \subset$ Möb$_n$ we define the Nielsen convex hull of the limit set $L(G)$ as the minimal set $\mathbf{H}_G \subset \mathbf{H}^{n+1} \cup \partial\mathbf{H}^{n+1}$ such that $L(G) \subset \mathbf{H}_G$ and $\mathbf{H}_G \cap \mathbf{H}^{n+1}$ is convex. There is the G-equivariant retraction $p: \bar{\mathbf{H}}^{n+1} \to \mathbf{H}_G$. If $x \in \mathbf{H}^{n+1} \setminus \mathbf{H}_G$, then $p(x)$ is the point in $\mathbf{H}_G$ with shortest distance to x; if x lies in the discontinuity set $\Omega(G) \subset \partial\mathbf{H}^{n+1}$, then $p(x)$ is the first point of tangency with Nielsen hull $\mathbf{H}_G$ of a horosphere in $\mathbf{H}^{n+1}$ with centre at x.

The geometry of the Nielsen hull $\mathbf{H}_G$ for a geometrically finite Kleinian group $G \subset$ Möb$_n$ was studied in [AP$_6$]. Here the main tool is the notion of maximal balls in the discontinuity set $\Omega(G)$. We say that an open ball $B \subset \Omega(G)$ is maximal if the sphere of lowest dimension containing the set $\partial B \cap L(G)$ is the sphere ∂B itself.

In the planar case $n = 2$, the discontinuity set $\Omega(G)$ of any finitely generated Kleinian group G whose limit set $L(G)$ is not contained in a circle is covered by a family of maximal disks, finite modulo G; see [AP$_6$, Theorem 3.1]. For $n \geqq 3$, it is in general necessary to consider, along with maximal balls, other maximal domains (namely, conical and cylindrical domains), – cf. Theorem 3.3 in [AP$_6$]. In what

appears below, however, we will use only maximal ball coverings for $\Omega(G)$. In this case, the boundary ∂H_G is described in the following way, where $\mathscr{B}(G)$ denotes the family of all maximal balls $B \subset \Omega(G)$. We start with the following statement ([AP_6, Lemma 4.1]):

Lemma 2.1 *Let $G \subset \text{Möb}_n$ be a for $n = 2$ finitely generated and for $n \geqq 3$ geometrically finite Kleinian group whose limit set $L(G)$ is not contained in any $(n-1)$-dimensional sphere. Also, for each maximal ball $B \in \mathscr{B}(G)$ let the subdomain $B' \subset B$ signify the interior of the Nielsen hull of the limit subset $\partial B \cap L(G)$ in the Poincaré hyperbolic metric of the ball B.*

Then the set

$$\Omega' = \cup \{B' : B' \subset B \in \mathscr{B}(G)\} \tag{2.1}$$

is the open set of bijectivity of the retraction p for the set $\Omega_b = \cup \{B : B \in \mathscr{B}(G)\} \subseteq \Omega(G)$. The restriction of the retraction p to Ω' is a conformal homeomorphism: given a connected component $B' \subset \Omega'$, $p|_{B'}$ is the rotation in $\mathbf{R}^{n+1}$ through the angle $\pi/2$ around the fixed sphere ∂B.

We note that every bijectivity domain B' of a maximal ball $B \in \mathscr{B}(G)$ has finite n-dimensional hyperbolic volume modulo the action of the stabilizer $G_B \subset G$ of B, (i.e., vol $B'/G_B < \infty$) due to the geometrical finiteness of the group G and precise invariance of B' with respect to G_B – see [AP_6, Lemma 4.3].

Now we can describe the boundary $\partial \mathbf{H}_G$ of the Nielsen's hull $\mathbf{H}_G$ for the group G as the closure of a union of flat pieces (namely, of non-compact n-dimensional hyperbolic polyhedra $p(B')$, $B' \subset B \in \mathscr{B}(G)$, lying on n-planes $\mathbf{H} \subset \mathbf{H}^{n+1}$, $\partial \mathbf{H} = \partial B$) which are joined to one another by totally geodesic k-dimensional polyhedra, $1 \leqq k \leqq n-1$. We call their edges, if these exist, the (simple) bending geodesics of ∂H_G. In other words, by a bending geodesic we mean a geodesic $l = (x, y)_h$ with end points $x, y \in L(G)$ such that l does not lie inside any flat k-dimensional piece of $\partial \mathbf{H}_G$, $k \geqq 2$. We call such a geodesic $l = (x, y)_h$ a simple bending geodesic if its end points $x, y \in L(G)$ lie on ∂B for some $B \in \mathscr{B}(G)$; otherwise, we call l a limit bending geodesic. (There is a sequence $\{l_i\}$ of simple bending geodesics which converge to l.)

The situation is described more precisely in the following statement ([AP_6, Lemma 4.4]).

Lemma 2.2 *Let $G \subset \text{Möb}_n$ be a Kleinian group whose limit set $L(G)$ is not contained in a $(n-1)$-sphere. Assume that G is finitely generated if $n = 2$ and geometrically finite if $n \geqq 3$. In the latter case assume additionally that the discontinuity set admits a maximal ball covering. Then there exists a finite subfamily $\mathscr{B}_f = \{B_1, \ldots, B_m\} \subset \mathscr{B}(G)$ of maximal balls such that:*

a) $G(B_1 \cup \ldots \cup B_m) = \Omega(G)$ and $G(B_i') \cap G(B_j') = \varnothing$ for $i \neq j$;

b) for each component $\Omega_0 \subset \Omega(G)$ (mod G) there exists a subfamily $\mathscr{B}_0 \subset \mathscr{B}_f$ of joint maximal balls B such that if the hyperbolic polyhedra $p(B') \subset p(\Omega_0) \subset \partial \mathbf{H}_G$ are joined to each other then their common faces are faces of some finite subset of their non-degenerate polyhedral angles $W_1, \ldots, W_k$;

c) the angles W_j have for $n = 2$ the value π, and for $n \geqq 3$ they have the combinatorial type of an angle

$$X_j = \{(tx, x_n) \in \mathbf{R}^{n+1} \times \mathbf{R} : t > 0, x \in Q_j\} \tag{2.2}$$

where Q_j is a convex Euclidean polyhedron lying on an $(n-2)$-plane in $\mathbf{R}^{n-1}$ which does not pass through the origin; furthermore, the edge in W_j corresponding to the

edge $\{(0, x_n): x_n > 0\} \subset X_j$ *(where* Q_j *is bounded) is a simple bending geodesic* $l = (x, y)_h \subset \partial \mathbf{H}_G$;
d) to every simple bending geodesic $l = (x, y)_h \subset \partial \mathbf{H}_G$ *(from the G-orbit of an edge of some angle* W_j*) there corresponds a finite subset* $\mathscr{B}(l)$ *of maximal balls such that an arc* $l' \subset \Omega(G)$ *of the circle meeting the points x and y is the axis of the spherical polyhedral angle*

$$B(l) = \cap \{B: B \in \mathscr{B}(l)\} \tag{2.3}$$

and such that $p(l') = l$.

Let us notice that one can describe the bijectivity domain $B_0' \subset B_0 \in \mathscr{B}(G)$ in another way (see [AP$_6$], p. 220). Namely, let $B_j, j \in J = J(B_0)$, be all the balls from $\mathscr{B}(G)$ that intersect B_0. For every $j \in J$ consider two $(n-1)$-spheres s_{0j} and $c_{0j} = c_{j0}$ in S^n, intersecting the spheres $S_0 = \partial B_0$ and $S_j = \partial B_j$ in the $(n-2)$-sphere $S_0 \cap S_j$. One of them, s_{0j}, is orthogonal to S_0 (s_{j0} is orthogonal to S_j) and the other, $c_{0j} = c_{j0}$, divides the dihedral angle $B_0 \cap B_j$ into halves, i.e., is its bisector. We define $\mathrm{ext}_0(s_{0j})$ and $\mathrm{ext}_0(c_{0j})$ in the ball B_0 as the components of the complements $B_0 \setminus s_{0j}$ and $B_0 \setminus c_{0j}$, respectively, which contain B_0'. Then it is true that

$$B_0' = \cap \{\mathrm{ext}_0(s_{0j}): j \in J(B_0)\} \subset B_0'' = \cap \{\mathrm{ext}_0(c_{0j}): j \in J(B_0)\} \tag{2.4}$$

Remark 2.3 The polyhedral angle $p^{-1}(W_j)$ of a domain $B_i' \subset B_i \in \mathscr{B}(l)$ (from the statement *c*) of Lemma 2.2) that has as its geodesic edge $l_i^0 = \bigcap_{k \in J(i)} s_{ik}$ is contained in a spherical polyhedral angle V_j of the same combinatorial type, having the arc $l' \subset B(l)$ as an edge (cf. (2.3)). This angle V_j is formed by intersection of the balls bounded by the bisector spheres c_{ik} or, more precisely, by the segments of these balls lying in $B_i \cap B_k$, where $B_k \in \mathscr{B}(l) \setminus \{B_i\}$.

It is necessary to point out here that

$$\cup \{V_i: V_i \subset B_i \in \mathscr{B}(l)\} \tag{2.5}$$

is the full neighbourhood of the arc l', $p(l') = l = (x, y)_h$. In other words, the angles V_j form a subdivision of a neighbourhood of the arc l' (for $n = 3$ into non-degenerate dihedral angles).

For the planar case $n = 2$ the boundary of the Nielsen hull $\mathbf{H}_G$ is a 2-dimensional pleated surface (along its bending geodesics) admitting an isometric development in the hyperbolic 2-plane $\mathbf{H}^2$. However, we will show that the n-dimensional pleated surface $\partial \mathbf{H}_G$, $n \geqq 3$, need not admit an isometric development in the n-plane $\mathbf{H}^n \subset \mathbf{H}^{n+1}$: in general, under certain conditions, it has conical singularities; compare the pea-pod example of a quasi-Fuchsian group $G \subset \mathrm{M\ddot{o}b}_3$ in [AP$_6$, §6]. Namely, the following holds:

Lemma 2.4 *Let* $G \subset \mathrm{M\ddot{o}b}_n$, $n \geqq 3$, *be a Kleinian group having at least three maximal balls* $B_i \subset \Omega(G)$ *with two common limit points* $x, y \in L(G)$ *on their boundaries and with* $\mathrm{int}(\cap B_i) \neq \varnothing$. *Then the boundary* $\partial \mathbf{H} \subset \partial \mathbf{H}^{n+1}$ *of the Nielsen hull* $\mathbf{H}_G$ *is not isometrically developable in the n-dimensional hyperbolic space* $\mathbf{H}^n$.

Proof. The statement follows directly from Lemmas 2.1 and 2.3. Indeed, for an isometric development of $\partial \mathbf{H}_G$ in $\mathbf{H}^n$ to exist, it is necessary that the polyhedral angles W_j of all the n-dimensional hyperbolic convex polyhedra $p(B_j')$, $B_j' \subset B_j \in \mathscr{B}(l)$, having as their common edge the simple bending geodesic $l \subset \partial \mathbf{H}_G$ with the ends x and y, should form a full neighbourhood of some geodesic in $\mathbf{H}^n$. However, to these polyhedral angles there correspond the strictly larger polyhedral

angles V_j of the domains $B_j'' \subset B_j$, ($B_j' \subset B_j''$ – see (2.4)) having as their common edge the arc $l' \subset \Omega(G)$(= axis of $B(l)$ from (2.3)) and forming its full neighbourhood. Therefore the above necessary condition for an isometric development is not satisfied if the number card $\mathscr{B}(l)$ of mutually intersecting maximal balls B, $\partial B \supset \{x, y\}$, is greater than 2. The proof is complete.

Note that for $n = 2$ it is always the case that card $\mathscr{B}(l) = 2$, and therefore there exists an isometric development of the 2-surface $\partial \mathbf{H}_G \subset \mathbf{H}^3$ in the plane $\mathbf{H}^2$.

3 Block-building construction for "wild" groups

We will describe here our method for the construction of Kleinian groups in space with wild limit sets. It was first developed (with some applications to topology and analysis) in [A-T] and [AP$_9$]. Under block-building we mean a construction of the desired Kleinian group $G \subset \text{Möb}_3$ from a set of block-groups G_i, $i = 1, \ldots, m$, by sequential amalgamated products (Maskit's first combinations)

$$G = \ldots * G_i \underset{G_{ij}}{*} G_j * \ldots \tag{3.1}$$

and perhaps, finally, by HNN-extension, (Maskit's second combination), where the block-groups G_i are Fuchsian groups acting in corresponding balls $B_i \subset S^3$. (It is possible that $B_1 = B_{m+1}$.) In the first instance the balls are characterized by one of the following two conditions:

For the case of amalgamated $\bar{B}_i \cap \bar{B}_j \neq \varnothing$ if and only if $|i - j| = 1 \pmod m$.

For the HNN-extension case, $\bar{B}_i \cap \bar{B}_j \neq \varnothing$ if and only if $|i - j| = 1$.

Moreover, the intersections $\partial B_i \cap \partial B_j$ (circles or points) are invariant for the amalgam-subgroups $G_{ij} = G_i \cap G_j$.

Then due to additional non-intersection condition concerning a fundamental domain of the group G and $\partial(\bigcup_{i=1}^m B_i)$, the discontinuity set $\Omega(G)$ of the constructed group G has an invariant component $\Omega \subset S^3 \setminus \cup B_i$ with nontrivial (infinitely generated) fundamental group $\pi_1(\Omega)$ and, consequently, the limit set $L(G) = \partial\Omega$ is a wild "Julia surface".

4 Construction of a finite family $\mathscr{C}$ of generator spheres

Let us give a concrete construction from Section 3 which is a modification of [AP$_9$] with some additional geometric conditions on maximal ball coverings of $\Omega(G)$. Here for the construction of the pertinent block-groups, we will use standard groups related to the unit hexagonal prism and the unit truncated icosahedron and generated by reflections.

Namely, let us consider a right prism in $\mathbf{R}^3$ of height 17 whose base is a polygon formed by the union of 55 congruent regular hexagons with unit sides and disjoint interiors. These hexagons cover the regular triangle with side $9\sqrt{3}$, whose vertices are the centres of the extreme hexagons; see Fig. 1. Let us enumerate all the hexagons line by line so that the three extreme ones have numbers 1, 10 and 55, and the central one has the number 31. Divide the prism P into (55×17) small hexagonal prisms $P(k, n)$ of unit height enumerated by pairs (k, n), where $k, 1 \leqq k \leqq 17$, is the floor of $P(k, n)$ and $n, 1 \leqq n \leqq 55$, is the number of the hexagonal projection of $P(k, n)$ in the base.

Now let us form (in this big building P) three apartments P_{gr}, P_{red}, P_{bl}, of green, red and blue colours, respectively. As rooms of these apartments we take the unions of small prisms $P(k, n)$ with the number-pairs (k, n) from the following sets:

(18,14) (18,42)
(17,14) (17,44)(17,48)(17,47)(17,42)
(16,14) (16,44)
(15,14) (15,44)
(14,14) *Blue* (14,44)
(13,14) *apartment,* (13,44)
(12,14) P_{bl}: (12,44)
(11,14) (11,44)
(10,14) (10,44)
(9,14) (9,44)
(8,14) (8,44)
(7,14) (7,44)
(6,14) (6,23) (6,31) (6,38) (6,44)

(18,33)
(17,33)(17,26)(17,17)
(16,17)
(15,17)
Green apartment, P_{gr}: (14,17)
(13,17)
(12,17)
(11,17)
(10,17)
(9,42) (9,37) (9,31) (9,24) (9,16) (9,17)
(8,42)
(7,42)
(6,42)
(5,42)
(4,42)
(3,42)
(2,42)
(1,42)
(0,42)

(12,28)(12,29)(12,30)(12,31)(12,32)(12,33)(12,34)
(11,28) (11,34)
(10,28) (10,34)
(9,28) (9,34)
(8,28) *Red apartment,* P_{red}: (8,34)
(7,28) (7,34)
(6,28) (6,20) (6,12) (6,34)
(5,12) (5,34)
(4,13) (4,12) (4,34)
(3,13) (3,34)
(2,13) (2,34)
(1,14) (1,13) (1,34) (1,33)
(0,14) (0,33)

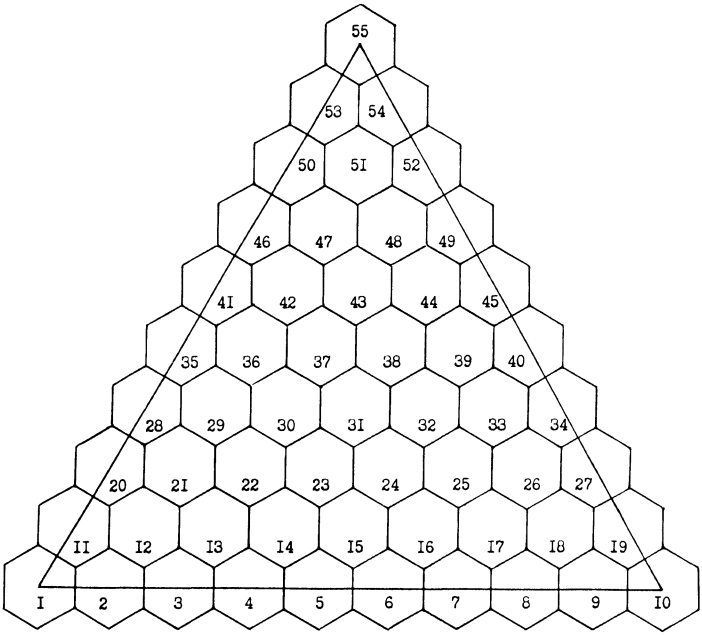


Fig. 1

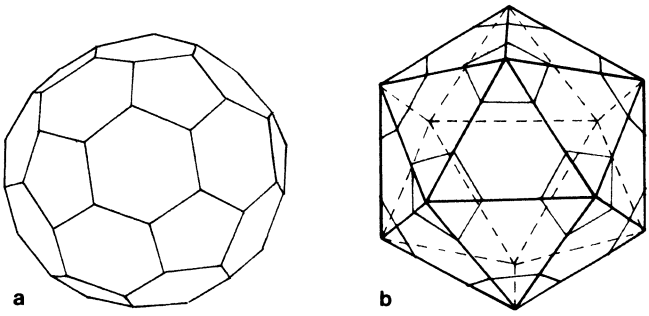


Fig. 2

To the apartments P_{gr}, P_{red}, P_{bl} are adjoined on the 0-th and 18-th floors six extra rooms. These are truncated (semiregular) icosahedra and are labelled $P(0, 14)$, $P(0, 33)$, $P(0, 42)$ and $P(18, 14)$, $P(18, 33)$, $P(18, 42)$, (see Fig. 2a). The new rooms are obtained from the icosahedron in Fig. 2b. They are attached to the existing first and 17-th floors along hexagonal foundation sides.

Such truncated icosahedra with regular pentagonal and hexagonal sides and with unit edge length are obtained from the Euclidean regular icosahedron with edge length 3; see Fig. 2. Since the radii of the circumscribed and inscribed balls for

the icosahedron are

$$\frac{3}{\sqrt{2(\sqrt{5}-1)}} \text{ and } \frac{3}{\sqrt{2(\sqrt{5}-1)}} \cdot \frac{\sqrt{5}+1}{\sqrt{3(\sqrt{10-2\sqrt{5}})}} = \frac{\sqrt{3}(3+\sqrt{5})}{4},$$

respectively, the radii of the inscribed and circumscribed balls for our truncated icosahedron are

$$r = \frac{\sqrt{3}(3+\sqrt{5})}{4} \text{ and } R = \sqrt{r^2+1} = \frac{\sqrt{58+18\sqrt{5}}}{4}, \tag{4.1}$$

respectively.

Now let S_i be the spheres of radius $\sqrt{3}/3$ with centres at vertices of rooms $P(k, n)$, $0 \leqq k \leqq 18$, forming the apartments P_{gr}, P_{red}, P_{bl}. If two such spheres S_i and S_j intersect, their centres are the adjacent vertices of some $P(k, n)$ and their intersection angle is $\pi/3$.

Denote by $B(k, n)$ the ball concentric with the room $P(k, n)$ of radius

$$\sqrt{11/12} \text{ if } 1 \leqq k \leqq 17 \text{ and } \sqrt{\frac{79+27\sqrt{5}}{8}}, \text{ if } k = 0, 18\,. \tag{4.2}$$

Its boundary sphere $S(k, n)$ is orthogonal to each of the spheres S_i whose centres are the vertices of $P(k, n)$. Next, we replace the polyhedral rooms $P(k, n)$ in our apartments by the round rooms $B(k, n)$. In doing so we obtain the linked tubes pictured in Fig. 3. We note that the balls $B(k, n)$ will be balls from the block-building construction, which will use HNN-extension by a hyperbolic transformation $h \in \mathrm{M\ddot{o}b}_3$ with the isometric spheres

$$I = I(h) = \{x \in \mathbf{R}^3 : |Dh(x)| = 1\} = S^2(z, \varrho)$$

and

$$I' = I(h^{-1}) = S^2(z', \varrho)\,,$$

where z, z' and ϱ will be determined later. This hyperbolic transformation h is the composition of reflections with respect to the sphere I and the plane which bisects

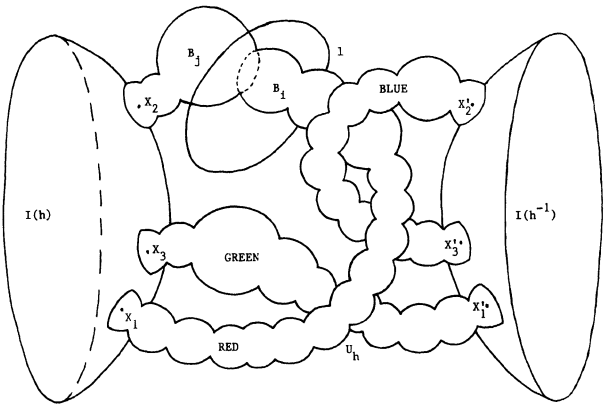


Fig. 3

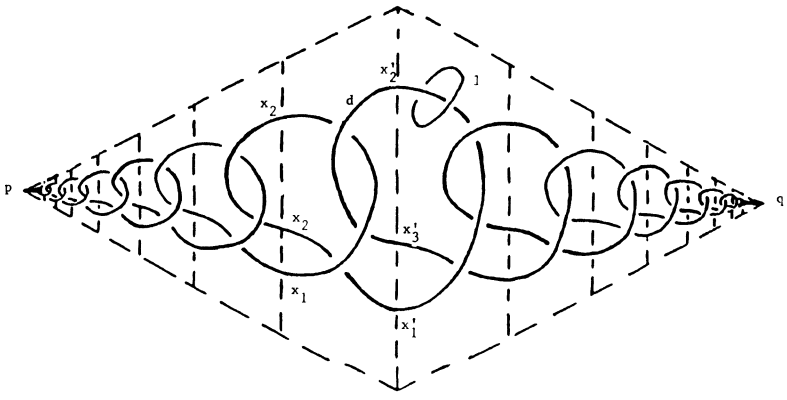


Fig. 4

the segment joining the centres of I and I'; it has two fixed points $p \in \operatorname{int} I$ and $q \in \operatorname{int} I'$. Furthermore, our apartments union is a fattening U_h of the knotted period d_h of a wild Fox-Artin arc d invariant under the cyclic group $\langle h \rangle$ (see Fig. 4).

Now we will finish to form a finite family $\mathscr{C}$ of spheres S_j, with the following properties:

1. The union of the annuli $X_i = \partial B_i \cap \operatorname{ext} I \cap \operatorname{ext} I' \cap (\bigcap_{j \neq i} \operatorname{ext} B_j)$ is covered by the interiors of S_j.
2. For each k, j either $B_k \cap S_j = \varnothing$ or S_j is orthogonal to the sphere ∂B_k; the same is true if $I(h)$ and $I(h^{-1})$ are taken instead of ∂B_k.
3. If $S_i \cap S_j \neq \varnothing$ the exterior dihedral angle (i.e., the one containing ∞) between them equals π/m for some integer m.
4. If $S_i \cap S_j \neq \varnothing$ then there exists a common annulus X_k for which $X_k \cap S_i \neq \varnothing$ and $X_k \cap S_j \neq \varnothing$.
5. There is a one-to-one correspondence between the spheres $S_j \in \mathscr{C}$ which meet $I(h)$ and the spheres $S_{j'} \in \mathscr{C}$ which intersect $I(h^{-1})$ such that $S_{j'} = h(S_j)$.

In other words we form a finite bubble cover of our apartments boundary ∂U_h with good angles between the bubbles and right angles between the bubbles and the spheres ∂B_k, a cover that respects h-periodicity.

Taking into account the Andreev-Thurston rigidity of circular coverings of the sphere S^2 (see Ch. 13 of [TH]), there are only a few such families.

Note that the interiors of the spheres S_i with the centres at the vertices of $P(k, n)$ do not cover the entire boundary ∂U_h, i.e., they don't completely cover the spherical annuli $X_j \subset S(k, n)$. Still uncovered are the hexagonal, quadrangular and pentagonal domains on these annuli corresponding to sides of $P(k, n)$. Let us cover each hexagonal domain on $X_i \subset S(k, n)$ with the union of the interiors of seven spheres orthogonal to the sphere $S(k, n)$. Six of these have equal radii, they are orthogonal to the spheres S_i, and they intersect at the angles $\pi/3$. The seventh sphere intersects the six others at the right angles and does not intersect the spheres S_i with the centres at vertices of $P(k, n)$; see Fig. 5.

We cover each quadrangular domain on $X_i \subset S(k, n)$ by the interiors of five spheres orthogonal to the sphere $S(k, n)$. Four of these spheres have equal radii;

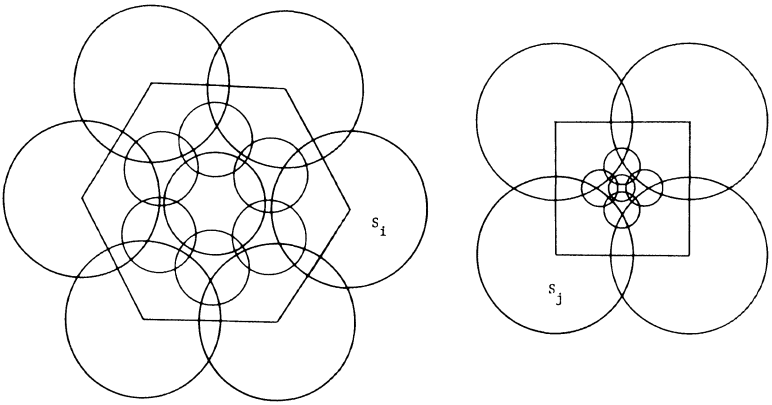


Fig. 5

they are orthogonal to the spheres S_j and thus intersect at the angle $\pi/3$. The fifth sphere intersects the four others at right angles and does not intersect the spheres S_j; see Fig. 5. We do similarly in the case of the pentagonal domains.

We complete the construction of the desired family $\mathscr{C}$ if we construct two isometric spheres $I = I(h)$ and $I' = I(h^{-1})$ with a common radius ϱ and with centres z and z' lying on the axis L of the prism P such that they are orthogonal to the spheres S_i of radius $\sqrt{3}/3$ with centres at the vertices of the truncated icosahedra $P(0, n)$ and $P(18, n)$, $n = 14, 33, 42$, which are nearest to z and z', respectively.

For this purpose, let us consider the three straight lines $L(0, n)$, $n = 14, 33, 42$ (respectively, $L(18, n)$, $n = 14, 33, 42$) which pass through the centres of the truncated icosahedra $P(0, n)$, $n = 14, 33, 42$ (respectively, $P(18, n)$) and also through the centres of its slanting hexagonal sides adjoining to the lower sides of $P(0, n)$, (respectively, the upper sides of $P(18, n)$) and facing the axis L of the building P. The points z and z' are the intersection points of the axis L with the straights $L(0, n)$ and $L(18, n)$, $n = 14, 33, 42$, respectively. In this case the angle φ between $L(0, n)$ (or $L(18, n)$) and axis L is given by

$$\operatorname{tg}(\varphi/2) = \sqrt{3}/2r = 2/(3 + \sqrt{5}), \text{ i.e., } \varphi = \operatorname{arctg}\frac{2}{\sqrt{5}}.$$

Therefore, for the distance between z and z' we get

$$17 + 2(r + 2\sqrt{3}/\operatorname{tg}\varphi) = 17 + 2\sqrt{15} + \sqrt{3}(3 + \sqrt{5})/2 . \tag{4.3}$$

The orthogonality condition for intersection of I and I' with the spheres S_i nearest to z and z', respectively, tells us that

$$\varrho = \sqrt{(493 - 81\sqrt{5})/24} \tag{4.4}$$

is the radius of the isometric spheres $I = I(h)$ and $I' = I(h^{-1})$.

Direct calculation shows that the resulting hyperbolic transformation $h \in \mathrm{M\ddot{o}b}_3$ and the finite covering $\mathscr{C}$ of the boundary ∂U_h by the spheres S_i have all the properties 1–5.

5 The group G construction

Denote by G_{gr}, G_{red}, G_{bl} the Möbius groups generated by reflections with respect to spheres S_j whose interiors intersect our green, red and blue apartments, respectively. These groups are discrete owing to property 3 of the family $\mathscr{C}$ and to the Poincaré-Aleksandrov theorem on fundamental polyhedra (see [AP$_5$]). Maskit's first combination theorem in $\mathbf{R}^3$ (see [AP$_3$, Theorem 4.2]) shows that these groups are amalgamated free products. In fact, they are of the type (3.1), where every amalgam-subgroup $G_{ij} = G_i \cap G_j$ is generated by reflections with respect to spheres from $\mathscr{C}$ whose interiors cover the circle $b_{ij} = \partial B_i \cap \partial B_j$, $G_i(B_i) = B_i$, $G_j(B_j) = B_j$.

Because of the non-intersection of spheres $S_j \in \mathscr{C}$ covering distinct apartments, we see from Klein's combination theorem (see [AP$_3$, Theorem 4.1]) that the discrete group G_0 generated by reflections in the spheres of $\mathscr{C}$ is the free product:

$$G_0 = G_{red} * G_{blue} * G_{green} . \tag{5.1}$$

Now, let us consider the extension G of the group G_0 by the hyperbolic transformation h described in section 4, i.e.,

$$G = \langle G_0, h \rangle . \tag{5.2}$$

We claim that this group has all the properties we are looking for.

Lemma 5.1 *The group $G \subset \text{Möb}_3$ defined in* (5.1) *and* (5.2) *is a discrete group isomorphic to a discrete co-compact group $\Gamma \subset \text{Isom } \mathbf{H}^3$.*

Proof. Let us denote $I(h) \cap d = \{x_1, x_2, x_3\}$ and $I(h^{-1}) \cap d = \{x'_1, x'_2, x'_3\}$, where $h(x_i) = x'_i$ and these points are placed on the Fox-Artin arc d in the following order (see Fig. 4):

$$x_1, x_2, x_3, x'_1, x'_2, x'_3.$$

Denote by H_i (respectively H'_i), $i = 1, 2, 3$, the subgroups of G_0 generated by reflections in the spheres of $\mathscr{C}$, covering the circles $I(h) \cap \partial B_j$ (respectively $I(h^{-1}) \cap \partial B_{j'}$) inside of which the points x_i (respectively x'_i) are located. And let

$$H = H_1 * H_2 * H_3,\ H' = H'_1 * H'_2 * H'_3 \subset G_0 . \tag{5.3}$$

Then, as in the similar proof in [AP$_9$], the statement follows from Maskit's second combination theorem (see [AP$_3$, Theorem 4.5]), Tietze's transformations (see [L-S, Section 2.2]) and Andreev's classification theorem on hyperbolic polyhedra (see [AN]):

$$\begin{aligned} G &= \langle G_{red} * G_{bl} * G_{gr}, h \mid hHh^{-1} = H' \rangle = \\ &= \langle G_{bl} \underset{H'_2}{*} hG_{red}h^{-1} \underset{H'_1}{*} G_{gr}, h \mid hH_3h^{-1} = H'_3 \rangle . \end{aligned} \tag{5.4}$$

Here the following polyhedra are fundamental for this Kleinian group G:

$$F = \cap \{\text{ext}S_i : S_i \in \mathscr{C}\} \cap \text{ext}I \cap \text{ext}I' = F_\infty \cup F_{red} \cup F_{green} \cup F_{blue}, \tag{5.5}$$

and

$$F^* = F_\infty \cup \text{int}\overline{(F_{blue} \cup F_{green} \cup h(F_{red}))} = F_\infty \cup F_0 , \tag{5.6}$$

where F_∞, F_{red}, F_{green}, F_{blue} are connected components of F and F^* has only two connected components: an unbounded component F_∞ and a bounded component F_0, which has the combinatorial type of a compact hyperbolic polyhedron $F'_0 \subset \mathbf{H}^3$ with the same angles.

6 An exotic uniformized structure on a hyperbolic manifold

Let $\Gamma \subset \mathrm{Isom}\,\mathbf{H}^3$ be the discrete co-compact group referred to in Lemma 5.1 which is isomorphic to the constructed Kleinian group $G \subset \mathrm{M\ddot{o}b}_3$, and let $\Gamma_* \subset \Gamma$ be a torsion-free subgroup of finite index. So, $M = \mathbf{H}^3/\Gamma_*$ is a closed hyperbolic 3-manifold containing a large set $\mathcal{S}$ of disjoint totally geodesic surfaces $S \subset M$. Indeed, $\mathrm{card}(\mathcal{S}/\Gamma_* \backslash \Gamma) \geqq 70$. We will show that this manifold M has an exotic conformal structure $c \in \mathcal{C}(M)$ uniformized by the Kleinian torsion-free subgroup $G_* \cong \Gamma_*$ of the Kleinian group G.

In some sense, this structure corresponds to a metric on M which on the one hand coincides with the hyperbolic metric on each piece of M obtained by cutting M along a set of totally geodesic disjoint surfaces $S_{ij} \subset M$, but on the other hand has singularities on the set $\cup S_{ij}$ that are similar to the bending singularities (see [AP_7], [AP_8]): the pieces of M are glued together along S_{ij} up to some angles of rotations around S_{ij}.

Theorem 6.1 *On the closed hyperbolic 3-manifold M there is a uniformized conformal structure $c \in \mathcal{C}(M)$ with the following properties:*
a) c is a quasiconformal image $f(c_0)$ of the canonical conformal structure $c_0 \in \mathcal{C}(M)$ induced by the hyperbolic metric;
b) the holonomy group $d_c^*(\pi_1(M)) = G_* \subset G \subset \mathrm{M\ddot{o}b}_3$ has two invariant components Ω_0 and Ω in its discontinuity set, one of which is the quasiconformal ball $\Omega_0 = \tilde{f}(B^3)$;
c) the holonomy group G_* has no parabolic elements;
d) the structure c is not a limit of any sequence of uniformized conformal structures $\{c_i\} \subset \mathcal{C}(M)$ obtained from the canonical structure c_0 through bending or stamping deformations.

Proof. Using the fact that any component, F_∞ or F_0, of the fundamental domain F^* of the group G (see (5.6)) has side identifications by elements of G, we see ([AP_3, Lemma 3.7]) that the domains

$$\Omega_0 = \cup \{\overline{g(F_0)} : g \in G\} \text{ and } \Omega_1 = \cup \{\overline{g(F_\infty)} : g \in G\} \tag{6.1}$$

are G-invariant components of the discontinuity set $\Omega(G)$ and $\Omega(G) = \Omega_0 \cup \Omega_1$.

Moreover, using the bending construction in [AP_9, (3.2)], we obtain a quasisymmetric homeomorphism

$$\hat{f} : \bar{\Omega}_0 = \Omega_0 \cup L(G) \to \bar{B} \tag{6.2}$$

that maps the closure of the first component Ω_0 to a closed ball $\bar{B} \subset \mathbf{R}^3$, is quasiconformal in Ω_0, and is compatible with the group G. In particular, it follows that the limit set $L(G)$ is a 2-sphere which is a wildly (even local wildly) embedded sphere (due to the Alexander theorem on the non-knottedness of any polyhedral 2-sphere in $\mathbf{R}^3$ and the non-triviality of the loop $l \subset \Omega_1$ (Fig. 3) – see [AP_9, Lemma 2.1]).

Therefore, the group $\hat{f}G\hat{f}^{-1} = \Gamma$ is a discrete co-compact group of hyperbolic isometries in the Poincaré ball model of $\mathbf{H}^3 \cong B$ and $\hat{f}^{-1}$ is the lifting of a quasi-conformal mapping

$$f^{-1}: B/\Gamma \to \Omega_0/G\,, \tag{6.3}$$

which maps the canonical conformal structure c_0 on the closed hyperbolic manifold B/Γ_* to the conformal structure $c \in \mathscr{C}(B/\Gamma_*)$ induced by the natural projection

$$\pi: \Omega_0 \to \Omega_0/G\,.$$

This completes the proof of statements a)–c) of the theorem.

We turn to statement d). For this we will study the geometry of the Nielsen hull $\mathbf{H}_G \subset \mathbf{H}^4$.

Let us consider the sphere $S_0 = S^2((z + z')/2, r_0)$ centered at the middle point $(z + z')/2$ of the segment $[z, z]' \subset L$ and having the radius r_0,

$$r_0 = \sqrt{\left(\frac{17}{2} + r + \frac{2\sqrt{3}}{\operatorname{tg}\varphi}\right)^2 - \varrho^2} = \sqrt{(922 + 153\sqrt{3} + 108\sqrt{5} + 255\sqrt{15})/12}\,. \tag{6.4}$$

This sphere is orthogonal to the isometric spheres $I = I(h)$ and $I' = I(h^{-1})$ with radii ϱ (see (4.4)). Accordingly, the fixed points p and q of the hyperbolic transformation h are the intersection points of S_0 and the axis L of the prism P. The Euclidean distance between them is

$$|p - q| = 2r_0 = 28,37904\ldots. \tag{6.5}$$

We show that the ball $\operatorname{ext} S_0$ lies in the component $\Omega_1 \subset \Omega(G)$. Namely, the construction of our apartments P_{red}, P_{gr} and P_{bl} insures that the vertices of its hexagonal prism rooms $P(k, n)$, $1 \leqq k \leqq 17$, have distances from the centre of the sphere S_0 with the following upper bound:

$$\sqrt{(17/2)^2 + (\sqrt{21} + 1)^2} = 10,169323\ldots. \tag{6.6}$$

On the other hand, among the spheres in the family $\mathscr{C}$ it is the centres of certain spheres $S_i \in \mathscr{C}$ of radius $\sqrt{3}/3$ centered at some vertices of the truncated icosahedra $P(0, n)$ and $P(18, n)$, $n = 14, 33, 42$, that are the most distant from the centre $(z + z')/2$ of the sphere S_0. This maximum distance $\varrho_{\max}$ can be estimated as follows:

$$\begin{aligned} 13,578091\ldots &= \sqrt{(r + 17/2)^2 + (2\sqrt{3})^2} + r < \varrho_{\max} < \sqrt{(17/2 + r)^2 + (2\sqrt{3})^2} + R \\ &= 13,788826\ldots. \end{aligned} \tag{6.7}$$

The estimates (6.6) and (6.7) show that the centres of all spheres in $\mathscr{C}$ lie inside the ball $\operatorname{Int} S_0$, which is invariant under the cyclic group with generator h. Therefore we have the similar property for the set of centres of spheres from the G-orbit of $\mathscr{C}$,

$$G(\mathscr{C}) = \{g(S): S \in \mathscr{C}, g \in G\}\,, \tag{6.8}$$

and hence

$$L(G) \setminus \{p, q\} \subset \operatorname{Int} S_0\,. \tag{6.9}$$

Also, the estimates (6.6) and (6.7) and the symmetry (of order 3) of the set of

extremal points having the distance $\varrho_{\max}$ from the centre of S_0 confirm the existence of at least three distinct maximal balls $B_1, B_2, B_3 \in \mathscr{B}(G)$ in the component Ω_1 such that for $i = 1, 2, 3$:

$$\{p, q\} \subset \partial B_i \cup L(G), \ \infty \in B_i, \ \mathrm{Int}(B_1 \cap B_2 \cap B_3) \neq \varnothing \ , \tag{6.10}$$

Using the construction of the Kleinian group G, especially of its block-subgroups G_i in the balls $B_i (\equiv B(k, n))$, we see that the hyperbolic 3-orbifold B/Γ does not have any family of totally geodesic surfaces intersecting along a closed geodesic γ (it corresponds to the axis of h) and invariant with respect to some automorphism of order 3 fixing γ. The main property of stamping deformations of $\mathbf{H}^3/\Gamma$ along totally geodesic surfaces intersecting on a geodesic (see [AP_7, Section 3]) shows that representations $\varrho : \Gamma \to \mathrm{M\ddot{o}b}_3$ with discontinuous images $\varrho(\Gamma)$ obtained by such stamping deformations cannot approximate the representation $\Gamma \to G \subset \mathrm{M\ddot{o}b}_3$.

Moreover, this representation $\Gamma \to G$ cannot be approximated by any sequence of injective homomorphisms $\varrho : \Gamma \to \mathrm{M\ddot{o}b}_3$ (with discontinuous images) obtained by bending deformations. Indeed, any discontinuous group $\varrho(\Gamma) \subset \mathrm{M\ddot{o}b}_3$ obtained by bending along a hyperbolic plane $\mathbf{H}^2 \subset \mathbf{H}^3 \cong B$ (i.e., along a totally geodesic surface $\mathbf{H}^2/\Gamma \subset B/\Gamma$) has a Nielsen hull $\mathbf{H}_{\varrho(\Gamma)} \subset \mathbf{H}^4$, whose boundary can be isometrically developed in $\mathbf{H}^3$; see [AP_6], [AP_7, Section 2]. On the other hand, the existence of the maximal balls (6.10) and Lemma 2.4 show that the boundary $\partial \mathbf{H}_G \subset \mathbf{H}^4$ in a neighbourhood of the geodesic $(p, q)_h \subset \mathbf{H}^4$ cannot be developed in the 3-plane $\mathbf{H}^3 \subset \mathbf{H}^4$ by means of a quasiisometry in $\mathbf{H}^4$ with any quasiisometry coefficient less than some $K = K(G) > 1$. This completes the proof.

References

[AN] Andreev, E.M.: Convex polyhedra in Lobachevsky spaces. (Russian). Mat. Sb. **81(123)**, 445–478 (1970)

[AP_1] Apanasov, B.N.: On certain analytic method in the theory of Kleinian groups in multidimensional space. (Russian). Dokl. Akad. Nauk SSSR **222**, 11–14 (1975)

[AP_2] Apanasov, B.N.: Nontriviality of Teichmüller space for Kleinian group in space. In: Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference. Ann. Math. Stud. **97**, 21–31 (1981)

[AP_3] Apanasov, B.N.: Discrete groups in space and uniformation problems. Math. Appl., D. Reidel Publ. Co. **40**, (1991) (to appear) (Extended English edition of: Discrete transformation groups and manifold structures (Russian). Novosibirsk: Nauka 1983)

[AP_4] Apanasov, B.N.: Thurston's bends and geometric deformations of conformal structures. In: Complex analysis and applications '85: Proc. Intern. Conf. on Complex Analysis, Varna, May 1985. Sofia: Publ. House Bulgarian Acad. Sci., 14–28 (1986)

[AP_5] Apanasov, B.N.: The effect of dimension 4 in Alexandrov's problem of filling a space by polyhedra. Ann. Global Anal. Geom. **4** (2), 243–261 (1986)

[AP_6] Apanasov, B.N.: The geometry of Nielsen's hull for Kleinian group in space and quasiconformal mappings. Ann. Global Anal. Geom. **6**(3), 207–230 (1988)

[AP_7] Apanasov, B.N.: Bending and stamping deformations of hyperbolic manifolds. Ann. Global Anal. Geom. **8** (1), 3–12 (1990)

[AP_8] Apanasov, B.N.: Deformations of conformal structures on hyperbolic manifolds. (accepted in J. Diff. Geom; see also Preprint 02723-89, Math. Sci. Res. Inst. Berkeley/California, 1989)

[AP_9] Apanasov, B.N.: Quasisymmetric embeddings of a closed ball inextensible in neighbourhoods of any boundary points. Ann. Acad. Sci. Fenn., Ser. A.I. Math. **14**, 243–255 (1989)

[AP_{10}] Apanasov, B.N.: Hyperbolic cobordisms and conformal structures. In: Apanasov, B., Neumann, W., Reid, A., Siebenmann, L. (eds.) Topology **90**, Proceedings, Columbus/Ohio, 1990. Berlin: Walter de Gruyter (to appear)

[A-T] Apanasov, B.N., Tetenov, A.V.: Nontrivial cobordisms with geometrically finite hyperbolic structures. J. Diff. Geom. **28**, 407–422 (1988)

[BR] Bers, L.: Uniformization, Moduli and Kleinian groups, Bull. Lond. Math. Soc. **4**, 257–300 (1972)

[GM] Goldman, W.M.: Geometric structures on manifolds and varieties of representations. In: Goldman, W.M., Magid, A.R. (eds.) Geometry of group representations. Contemp. Math. **74**, Providence 1988

[G-L-T] Gromov, M., Lawson, H.B., Thurston, W.: Hyperbolic 4-manifolds and conformally flat 3-manifolds, Publ. Math. IHES **68**, 27–45 (1988)

[J-M] Johnson, D., Millson, J.J.: Deformation spaces associated to compact hyperbolic manifolds. In: Howe, R. (ed.), Discrete groups in geometry and analysis: Papers in honour of G.D. Mostow on his sixtieth birthday. Boston: Birkhäuser, 48–106 (1987)

[KM] Kamishima, Y.: Conformally flat manifolds whose development maps are not surjective. Trans. Am. Math. Soc. **234**, 607–623 (1986)

[KO] Kourouniotis, C.: Deformations of hyperbolic structures. Math. Proc. Camb. Phil. Soc. **98**, 247–261 (1985)

[KU] Kuiper, N.H.: Hyperbolic 4-manifolds and tesselations. Publ. Math. IHES **68**, 47–76 (1988)

[K-P] Kulkarni, R.S., Pinkall, U.: Uniformization of geometric structures with applications to conformal geometry. In: Differential geometry: Proc. Symp. Peniscola, Spain 1985. Lect. Notes Math., vol. 1209, pp. 190–210. Berlin Heidelberg New York: Springer 1986

[L-S] Lindon, R.C., Schupp, P.E.: Combinatorial group theory. Ergeb. Math. **89**, Berlin Heidelberg New York: Springer 1977

[LK] Lok, W.L.: Deformations of locally homogeneous spaces and Kleinian groups. Thesis Dr. Phil., Columbia Univ. 1984

[MW] Mostow, G.D.: Quasiconformal mappings in n-space and the rigidity of hyperbolic space forms. Publ. Math. Inst. Hautes Études Sci. **34**, 53–104 (1968)

[MG] Morgan, J.: Group actions on trees and the compactification of the space of classes of SO(n, 1)-representations. Topology **25**, 1–33 (1986)

[TH] Thurston, W.: The geometry and topology of 3-manifolds. Mimeographed Lect. Notes, Princeton Univ. 1978/1980

Invent math. 105, 153–156 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Half Moufang implies Moufang for finite generalized quadrangles

J.A. Thas[1], **S.E. Payne**[2], **and Van Maldeghem**[1]

[1] Seminar of Geometry and Combinatorics, State University of Ghent, Krijgslaan 281, B-9000 Gent, Belgium
[2] Department of Mathematics, Box 170, University of Colorado at Denver, Denver, Colorado 80204, USA

Oblatum 1-XI-1989 & 7-XI-1990

Summary. A finite generalized quadrangle has two types of panels. If each panel of one type is Moufang, then every panel is Moufang. Hence by a theorem of Fong and Seitz [1] the quadrangle is classical or dual classical.

1 Introduction

Throughout this note we let $\mathscr{S}=(P, B, I)$ be a finite generalized quadrangle (GQ) of order (s, t), $s>1$, $t>1$. (For definitions and results not given in full here we recommend the monograph [2], hereafter denoted by FGQ.) A panel of $\mathscr{S}$ is an ordered triple (x, L, y), where x and y are distinct points incident with the line L, or an ordered triple (L, p, M), where L and M are distinct lines incident with the point p. For a given panel (x, L, y), let $H(x, L, y)$, be the group of all collineations of $\mathscr{S}$ which fix x and y linewise and L pointwise. For each line M through x, $M \neq L$, $H(x, L, y)$ acts semiregularly on the points of M different from x. Hence $|H(x, L, y)|$ divides s. We say the panel (x, L, y) is Moufang provided $|H(x, L, y)|=s$. For a panel (L, p, M), the group $H(L, p, M)$ is defined in a dual manner. And (L, p, M) is Moufang provided $|H(L, p, M)|=t$. If every panel of $\mathscr{S}$ is Moufang, then by a theorem of Fong and Seitz [1] (cf. J. Tits [3], p. 240), $\mathscr{S}$ must be isomorphic to one of the so-called classical or dual classical examples (cf. FGQ).

The main result of this paper is that if every panel of one type is Moufang, then all panels are Moufang. By point-line duality we may state this as follows:

1.1 Theorem. *If every panel of the form* (x, L, y) *is Moufang, then also every panel of the form* (L, p, M) *is Moufang. Hence* $\mathscr{S}$ *is classical or dual classical.*

The starting point is the following immediate corollary of 9.1.1 and 5.6.2 of FGQ:

1.2 Theorem. *Suppose that every panel of* $\mathscr{S}$ *of the form* (x, L, y) *is Moufang. Then one of the following must hold*:

i. Each point of $\mathscr{S}$ is regular (so $t \leqq s$ by 1.3.6 of FGQ), or
ii. For each pair $\{x, y\}$ of noncollinear points, $|\{x, y\}^{\perp\perp}| = 2$, or
iii. $\mathscr{S} \cong H(4, s)$ (so $\mathscr{S}$ is classical; cf. 3.1.1 of FGQ).

The cases $s \leqq t$ and $s > t$ are somewhat different, but both utilize the theory of elation generalized quadrangles (EGQ). Recall that an elation about a point p is a collineation of $\mathscr{S}$ that fixes p linewise and acts semiregularly on the points of $P \backslash p^{\perp}$. And if G is a group of elations about p acting regularly on $P \backslash p^{\perp}$, we say $(\mathscr{S}^{(p)}, G)$ is an EGQ. See Chapters 8, 9 and 10 of FGQ for definitions and results about EGQ.

Note. For any line N, N^* denotes the set of points incident with N.

2 A local Moufang condition

Main Hypothesis: $\mathscr{S}$ has three distinct points x_1, p, y_1 and two distinct lines L, M with $x_1 I L I p I M I y_1$ for which both (x_1, L, p) and (y_1, M, p) are Moufang.

For $x \in L^* \backslash \{p\}$, put $A(x) = H(x, L, p)$ and $A = \langle A(x) : x \in L^* \backslash \{p\} \rangle$. Analogously, for $y \in M^* \backslash \{p\}$, put $B(y) = H(y, M, p)$ and $B = \langle B(y) : y \in M^* \backslash \{p\} \rangle$. Finally, put $G = \langle A, B \rangle$.

Since $H(x_1, L, p)$ is transitive on $M^* \backslash \{p\}$ and $H(y_1, M, p)$ is transitive on $L^* \backslash \{p\}$, the following result is an immediate consequence of the Main Hypothesis.

2.1 Lemma. *$|A(x)| = |B(y)| = s$ for all $x \in L^* \backslash \{p\}$ and all $y \in M^* \backslash \{p\}$.*

For $a \in A$, $b \in B$, the commutator $[a, b] = a^{-1} b^{-1} a b$ fixes L pointwise, p linewise, and M pointwise. By 2.4.1 of FGQ it follows that $[a, b]$ acts semiregularly on the points of $P \backslash p^{\perp}$. In particular:

2.2 Lemma. *For $x \in L^* \backslash \{p\}$ and $y \in M^* \backslash \{p\}$, and for $a \in A$, $b \in B$, $[a, b]$ acts semiregularly on the points of $\{x, y\}^{\perp} \backslash \{p\}$. Hence $|[a, b]|$ divides t, and $ab = ba$ if $[a, b]$ fixes some point of $P \backslash p^{\perp}$.*

2.3 Lemma. *Let $z \in P \backslash p^{\perp}$ with $L I x I M_1 I z I L_1 I y I M$. Let $\mathrm{id} \neq a \in A(x)$, $\mathrm{id} \neq b \in B(y)$. Then the following are equivalent:*

i. $z^{ab} = z^{ba}$
ii. $[a, b]$ fixes z^{ba}
iii. $[a, b] = \mathrm{id}$
iv. $M_1^{ba} = M_1^b$
v. $L_1^{ab} = L_1^a$
vi. $L_1^a \perp M_1^b$

Proof. Easy to check with the aid of Fig. 1. □

With the setup as in lemma 2.3, suppose $ab = ba$, so $z^{ab} = z^{ba}$, etc. Let N be any line through x, $L \neq N \neq M_1$. Define $z_1 \in P \backslash p^{\perp}$ by $N I z_1 \perp y$. Since $ab = ba$, clearly $z_1^{ab} = z_1^{ba}$, and the above argument shows that N^b is fixed by a. As b is a bijection from the set of lines through x to the set of lines through x^b, each line through x^b must be fixed by a. We have essentially proved the following:

2.4 Lemma. *For $\mathrm{id} \neq a \in A(x)$ and $b \in B(y)$, if a and b commute, then $a \in A(x^b)$. If $ab \neq ba$, then no line through x^b can be fixed by a.*

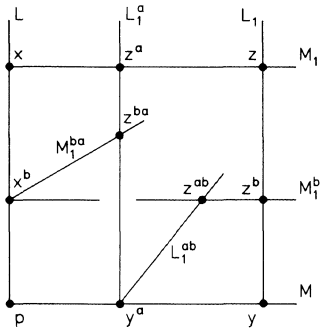


Fig. 1

2.5 Lemma. *If x, w are distinct points of $L^*\setminus\{p\}$ and $a\in A(x)$, then either a fixes all lines through w or only L through w.*

Proof. For any $y\in M^*\setminus\{p\}$, there is a unique $b\in B(y)$ with $w=x^b$. □

2.6 Lemma. *If $\theta\in A$ and θ fixes some line M_1 meeting L at a point x, $x\neq p$, then $\theta\in A(x)$. Hence whenever $\theta\in A$ fixes one line $(\neq L)$ through a point x $(\neq p)$ on L, then θ fixes x linewise.*

Proof. Let $\theta\in A$ satisfy the hypothesis. Let $z\in M_1^*\setminus\{x\}$ and suppose $z^\theta=z_1$. There must be a unique $a\in A(x)$ for which $z_1^a=z$. So $z^{\theta a}=z$. But θa fixes z, fixes L pointwise and p linewise. By 2.4.1 of FGQ, $\theta a=\mathrm{id}$, forcing $\theta\in A(x)$. □

The above proof actually shows the following:

2.7 Lemma. *$A(x)$ is the group of all collineations θ of $\mathscr{S}$ which fix L pointwise, p linewise, and fix some line $(\neq L)$ through x.*

For $c, d\in G$, and for $H\leqq G$, write $c^d=d^{-1}cd$ and $H^c=c^{-1}Hc$. Then the following is immediate.

2.8 Lemma. *$A(x)^b=A(x^b)$ for all $x\in L^*\setminus\{p\}$ and all $b\in B$. Hence $B\leqq N_G(A)$, forcing $A\lhd G$. Similarly, $B\lhd G$, and $G=AB$.*

For fixed $x\in L^*\setminus\{p\}$ and fixed $y\in M^*\setminus\{p\}$, $G=\langle A, B\rangle=\langle A(x), B(y)\rangle$. Hence $G=\langle A(x), B\rangle=\langle A, B(y)\rangle=A(x)\cdot B=A\cdot B(y)$. As $A(x)\cap B=\{\mathrm{id}\}=A\cap B(y)$, also $|G|=|A(x)|\cdot|B|=|A|\cdot|B(y)|$. But $|A(x)|=s=|B(y)|$, so $|A|=|B|$.

2.9 Lemma. *G acts semiregularly on $P\setminus p^\perp$.*

Proof. Write $G=A(x)\cdot B$ for some $x\in L^*\setminus\{p\}$. And suppose $\theta=ab$ fixes $z\in P\setminus p^\perp$, for some $a\in A(x)$, $b\in B$. Define y by $MIy\perp z$. Then $z=z^{ab}\perp y$ and $y^{ab}=y^a$ imply $y^a=y$. Hence $a=\mathrm{id}$. So $z^b=z$, forcing $b\in B(y)$ (by lemma 2.6). But then $b=\mathrm{id}$. □

3 EGQ and the local Moufang condition

Suppose that for some point p there is a group G for which $(\mathscr{S}^{(p)}, G)$ is an EGQ. Let $L_0, \ldots, L_t$ be the lines through p. Let e be a fixed but arbitrary point of $P\setminus p^\perp$. For $g\in G$ we can identify g with the point e^g. For $0\leqq i\leqq t$, define x_i, M_i by $eIM_iIx_iIL_i$. Put $S_i=\{\theta\in G: M_i^\theta=M_i\}$, $S_i^*=\{\theta\in G: x_i^\theta=x_i\}$. Then $|S_i|=s$; $|S_i^*|=st$; $|G|=s^2t$; $|S_iS_j\cap S_k|=1=|S_i^*\cap S_j|$ if i, j, k are distinct. And $\{S_j^*,$

$S_jS_0\backslash S_j$, $S_jS_1\backslash S_j$, ..., $S_jS_t\backslash S_j\}$ (omitting $S_jS_j\backslash S_j$) is a partition of G. (See Chapters 8, 9 and 10 of FGQ for these and many other results about EGQ).

We want to develop conditions that guarantee that panels of the form (L_j, p, L_k) are Moufang.

3.1 Lemma. *With the same notation as above, suppose* $0<j<k\leqq t$ *and suppose that both panels* (x_i, L_i, p), $i\in\{j, k\}$, *are Moufang, with* $H(x_i, L_i, p)=S_i\leqq G$ *and* $(\mathscr{S}^{(p)}, G)$ *an EGQ. Then the panel* (L_j, p, L_k) *is Moufang.*

Proof. Since S_j fixes all points of L_j and S_k acts regularly on the points of L_j different from p, it must be that no element of $S_jS_k\backslash S_j$ can fix any point of L_j different from p. Hence from the partition of G given above, S_j^* must contain all the elements of G that fix some point of L_j different from p. But each such point has a stabilizer of order st in G. Hence $\theta\in G$ fixes some point of L_j different from p if and only if $\theta\in S_j^*$ if and only if θ fixes all points of L_j. (Note: A corollary is that $S_j^*\lhd G$). Similarly, $\theta\in G$ fixes some point of L_k different from p if and only if $\theta\in S_k^*$ if and only if θ fixes all points of L_k. Then from $S_j^*\cap S_k^*\leqq H(L_j, p, L_k)$ and $t=|S_j^*\cap S_k^*|\leqq|H(L_j, p, L_k)|\leqq t$, it follows that (L_j, p, L_k) is Moufang. □

In the next lemma, the x_i, L_i, p, ... are as defined in the beginning of this section, but we do not require $\mathscr{S}$ to be an EGQ for some group.

3.2 Lemma. *Suppose* $s\leqq t$ *and* $|\{p, y\}^{\perp\perp}|=2$ *whenever* $y\in P\backslash p^\perp$. *Suppose* (x_i, L_i, p) *is Moufang for each* i, $0\leqq i\leqq t$. *Then each panel of the form* (L_i, p, L_j), $0\leqq i<j\leqq t$, *is Moufang.*

Proof. Put $G_p=\langle H(p, py, y)\colon y\in p^\perp\backslash\{p\}\rangle$. Then by 9.4.3 of FGQ, $(\mathscr{S}^{(p)}, G_p)$ is an EGQ. Now by lemma 3.1, the panel (L_i, p, L_j) is Moufang for $0\leqq i<j\leqq t$. □

4 Proof of Theorem 1.1

Suppose that every panel of $\mathscr{S}$ of the type (x, L, y) is Moufang, so the conclusion of theorem 1.2 holds.

If all points of $\mathscr{S}$ are regular and $s=t$, then $\mathscr{S}\cong W(s)$ (cf. 5.2.1 of FGQ).

Next assume that $s>t$. For some point p, let $L_0, \ldots, L_t$ be the lines through p and consider a panel of the form (L_j, p, L_k), $0\leqq j<k\leqq t$. For $i=j, k$ put $\hat{S}_i=\langle H(x, L_i, p)\colon L_i I x\neq p\rangle$, and let $G_{jk}=\langle\hat{S}_j, \hat{S}_k\rangle$. It is clear that G_{jk} fixes p linewise. By lemma 2.9 G_{jk} acts semiregularly on $P\backslash p^\perp$. Then by 9.4.2 of FGQ (with $r=1$), G_{jk} is transitive on $P\backslash p^\perp$, so $(\mathscr{S}^{(p)}, G_{jk})$ is an EGQ. Now by lemma 3.1, it follows that (L_j, p, L_k) is Moufang.

If $s\leqq t$ and $|\{x, y\}^{\perp\perp}|=2$ for each pair (x, y) of noncollinear points, then the desired result follows from lemma 3.2.

Finally, if $\mathscr{S}\cong H(4, s)$, then $\mathscr{S}$ is classical (and hence Moufang). □

Acknowledgement. The third author is supported by the National Fund for scientific research (Belgium).

References

1. Fong, P., Seitz, G.M.: Groups with a (B, N)-pair of rank 2, I and II. Invent. Math. **21**, 1–57 (1973) and **24**, 191–239 (1974)
2. Payne, S.E., Thas, J.A.: Finite generalized quadrangles. Pitman 1984
3. Tits, J.: Classification of buildings of spherical type and Moufang polygons: a survey. In: Segre, B. (ed.) Teorie Combinatorie. Roma Accad. Naz. Lincei 229–246 (1976)

Invent. math. 105, 157–175 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Pointwise analysis of Riemann's "nondifferentiable" function

M. Holschneider[1] **and Ph. Tchamitchian**[2]

Centre de Physique Théorique*, C.N.R.S.-Luminy-Case 907, F-13288 Marseille Cedex 9, France

Oblatum 27-III-1990 & 20-XI-1990

Summary. We will show how to analyse the local regularity of functions with the help of the wavelet transform. These results will be applied to the function of Riemann, where we show the existence of a dense set of points where this function is differentiable. On another dense set we show the existence of local singularities of cusp type. On a third set we show differentiability to the right (left). On the remaining set the functions will be shown to be not differentiable.

1 Introduction and historical remarks

In the year 1872 Weierstrass introduced his famous function [1]

$$\sigma(t)=\sum_{n=1}^{\infty} b^n \cos(a^n t), \qquad 0<b<1$$

for which he could show that it is continuous but nowhere differentiable whenever a is an odd integral number and $a\cdot b$ exceeds a certain value. Hardy later showed [2] that $a\cdot b>1$ is all that is needed to imply nondifferentiability everywhere as it can now be shown very easily. The history of nondifferentiable functions represented by lacunary Fourier series goes even back to Riemann, who is reported to have proposed before the function

$$W(t)=\sum_{n=1}^{\infty} \frac{1}{n^2} \sin(\pi n^2 t), \tag{1.1}$$

– which is less lacunary than Weierstrass's function σ – as a continuous, nowhere differentiable function, but neither he himself nor Weierstrass who took his inspiration from this example did succeed in proving this. However Riemann's assertion was partially confirmed by Hardy who proved that W is not differentia-

[1] and Université de Paris-Dauphine, CEREMADE
[2] and Faculté des Sciences et Techniques de Saint-Jerôme, Marseille
* Laboratoire Propre LP-7061, Centre National de la Recherche Scientifique Juin 1989

ble at any irrational point and at any rational point of the form P/Q with either $P=0 \bmod 2$ and $Q=1 \bmod 4$ or $P=1 \bmod 2$ and $Q=2 \bmod 4$. In his demonstration he used the analytic continuation of Riemann's function into the complex upper half-plane, which is a Jacobi theta function. Inparticular he used the analysis of the behaviour of this Jacobi theta function near the real axis that he and Littlewood had obtained [3] before.

So everything seemed to indicate that Riemann's conjecture was right until in 1970 Gerver proved that W actually is differentiable at any rational point of the form P/Q with $P=Q=1 \bmod 2$ [4] with derivative $-1/2$. He later proved that Riemann's function was not differentiable at any other rational point, and he showed [5] that his technique of analysis applied to other trigonometric sums of the type $\sum_1^\infty \sin(\pi r(n))/r(n)$. So the question of differentiability of Riemann's function was completely solved. Later Queffelec [6, 7] simplified the proof of Gerver that was based on elementary, but complicated estimates on trigonometric sums.

In 1988 one of us showed that also the wavelet analysis of W may lead to a Jacobi theta function [8]. The wavelet transform is a sort of mathematical microscope analysing arbitrary functions on various length scales around any point of the real axis. However all results in [8] where based on strong "a priori" hypothesis' on the function to be analyzed, which could not be verified for W.

The aim of the following paper is to prove rigorous results for the analysis of the pointwise regularity of functions through wavelet transforms in general. Such a kind of study of local regularity is not new. In their paper "Local properties of solutions of elliptic partial differential equations" published in 1961 [9], Calderon and Zygmund established pointwise estimates for solutions of elliptic partial differential equations. In the famous book by Stein, "Singular integrals and differentiability properties of functions" [10], explicit characterisations of pointwise differentiability, in the usual sense or in some weaker form (defined by means of suitable integral oscillations), are given. Moreover, their results are obtained through the use of one of the ancestors of the wavelet transform: the gradient of the harmonic extension of functions. So the reader may ask about what is new in this paper. The point is that our results are simultaneously on pointwise differentiability and on usual sense. We mean by this that they are not of the "almost everywhere" type and that regularity will not be expressed in any integral sense. The price we had to pay is that our theorems are not characterising the desired regularity properties, but are of Tauberian nature. They are however optimal. The analysis of local regularity using the wavelet transform had been established before in the case of local Hölder regularity by Jaffard [11]. The theorem given in this paper prooving pointwise differentiability through wavelet coefficients however is new, as far as we know.

Applying these results to the Riemann function W we shall be able to reprove the results of Hardy, Gerver and Queffelec concerning the question of differentiability of Riemann's function. In addition we prove that for any rational P/Q with $P=1 \bmod 2$ and $Q=0 \bmod 2$ or its inverse the function of Riemann has a cusp, which we shall be able to characterize completely. In particular we will show that there are infinite many points where W is differentiable to the

right (left) but not to the left (right). The key element in this analysis is again the Jacobi theta function, that was already used by Hardy. But this time we shall interpret this theta function as a wavelet analysis of W.

The paper is organized as follows. In Sect. 2 we give a brief introduction to the wavelet transform. In particular we give a sufficient condition, valid for a large class of functions, under which the inverse wavelet transform converges pointwise. In Sect. 3 we recall how to prove global Hölder regularity through wavelet transforms. The analysis of local Hölder regularity will make the subject of Sect. 4. In Sect. 5 we will show how to prove local differentiability with the help of the wavelet transform. All these results shall be applied to the Riemann-Weierstrass function W in Sects. 6, 7, 8 and 9.

2 Introduction to the wavelet transform

We will not give an exhausting definition of what we will call a *wavelet*, since in the literature the term wavelet is used for various kinds of functions depending on the application. However to fix the ideas consider a locally integrable complex-valued function φ, which in general is well localized and regular as for instance expressed by the following condition:

$$|\varphi(t)|+|\varphi'(t)|=O\left(\frac{1}{1+|t|^{2+\varepsilon}}\right).$$

In addition suppose that the first two moments of φ vanish:

$$\int_{-\infty}^{+\infty} dt\,\varphi(t)=\int_{-\infty}^{+\infty} dt\,t\,\varphi(t)=0.$$

These two condition are maximal in the sense that all theorems stated in this paper will hold for those functions taken as wavelets. Depending on the problem we shall be able to relax these conditions considerably. Since some moments of φ vanish it necessarily has some oscillations, which justifies the term wavelet. Dilating and translating[3] the wavelet φ we obtain a two parameter family of functions $\frac{1}{a}\varphi\left(\frac{t-b}{a}\right)$. The parameter $b\in\mathbb{R}$ is a position parameter, whereas $a>0$ may be interpreted as a scale parameter. The wavelet transform of an arbitrary function s with respect to a wavelet φ is given [12, 13] by

$$T_\varphi s(b,a)=T(b,a)=\int_{-\infty}^{+\infty} dt\,\frac{1}{a}\bar{\varphi}\left(\frac{t-b}{a}\right)s(t)$$

whenever these Lebesgue integrals are well defined for all b and a. In general T is a smooth function over the position-scale half-plane. Analysing a function with the help of the wavelet transform amounts to analyse it on different length scales around arbitrary positions. This transform is a sort of mathematical micro-

[3] In general a continuous wavelet analysis is constructed with the help of an irreducible, square integrable representation of a group [12]. Here it is the affine group.

scope, where $1/a$ is the enlargement and b is its position over the function to be analyzed. The specific optic is determined by the wavelet itself. This interpretation is useful in the analysis of fractal objects with the help of this transform [8, 14].

The wavelet transform is invertible. An explicit inversion formula is given by the following double integral over the position-scale half-plane

$$s(t)=\int_0^\infty \frac{da}{a}\int_{-\infty}^{+\infty} db\,\frac{1}{a}\,g\left(\frac{t-b}{a}\right)T(b,a),$$

where g is a fairly arbitrary, integrable function subject only to the following restrictions

(2.1) (i) $a^{-1}\hat{\varphi}(a)\hat{g}(a)\in L^1(\mathbb{R})$,

(ii) $\int_0^\infty \frac{da}{a}\,\bar{\hat{\varphi}}(a)\hat{g}(a)=1$,

(iii) $\int_0^\infty \frac{da}{a}\,\bar{\hat{\varphi}}(-a)\hat{g}(-a)=1$.

Such a function is called a *reconstruction wavelet* for the wavelet φ. In particular for an integrable wavelet its reconstruction wavelet may be chosen compactly supported and two times continuously differentiable. In this case for any t the reconstruction integral actually runs over a cone with top in t and with its opening angle determined by the support of g. This cone is sometimes called the *influence cone* of g at t. The inversion formula holds pointwise as is shown by the following theorem.

(2.2) Theorem. *Let s be a bounded*[4] *function that is weakly oscillating around* 0 *as expressed by the following condition:*

$$\lim_{u\to\infty}\frac{1}{2u}\int_{t-u}^{t+u} dy\,s(y)=0 \qquad \textit{uniformly in } t. \tag{2.3}$$

Let T be its wavelet transform with respect to an integrable wavelet φ and let g be an integrable reconstruction wavelet. Assume that φ and g decrease fast enough so that

$$\log(2+|t|)\,\varphi(t)\in L^1(\mathbb{R}) \quad \textit{and} \quad \log(2+|t|)\,g(t)\in L^1(\mathbb{R}). \tag{2.4}$$

Then in any point t where s is continuous we have

$$s(t)=\lim_{\substack{\varepsilon\to 0\\ \rho\to\infty}}\int_\varepsilon^\rho \frac{da}{a}\int_{-\infty}^{+\infty} db\,\frac{1}{a}\,g\left(\frac{t-b}{a}\right)T(b,a). \tag{2.5}$$

[4] Actually s may be growing polynomially if in turn the wavelet decreases fast enough at infinity.

Proof. Replacing the definition of T under the integral and exchanging the order of integration yields

$$\begin{aligned} s_{\varepsilon,\rho}(t) &= \int_{\varepsilon}^{\rho} \frac{da}{a} \int_{-\infty}^{+\infty} db\, T(b,a) \frac{1}{a} g\left(\frac{t-b}{a}\right) \\ &= \int_{\varepsilon}^{\rho} \frac{da}{a} \int_{-\infty}^{+\infty} db \int_{-\infty}^{+\infty} dy \frac{1}{a} \bar{\varphi}\left(\frac{y-b}{a}\right) \frac{1}{a} g\left(\frac{t-b}{a}\right) s(y) \\ &= \int_{-\infty}^{+\infty} dy\, M_{\varepsilon,\rho}(y-t)\, s(y) \end{aligned}$$

where we have written

$$M_{\varepsilon,\rho}(t) = \int_{\varepsilon}^{\rho} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{a} \bar{\varphi}\left(\frac{y-b}{a}\right) \frac{1}{a} g\left(\frac{t-b}{a}\right).$$

We have

$$M_{\varepsilon,\rho}(t) = \frac{1}{\varepsilon} M\left(\frac{t}{\varepsilon}\right) - \frac{1}{\rho} M\left(\frac{t}{\rho}\right),$$

where (almost everywhere)

$$M(t) = \frac{1}{t} \int_{0}^{t} d\tau\, h(\tau), \quad \text{and} \quad h(t) = \int_{-\infty}^{+\infty} db\, \bar{\varphi}(t+b)\, g(b). \tag{2.6}$$

The hypothesis on φ and g implies that

(i) $\int_{-\infty}^{0} h = \int_{0}^{\infty} h = 0$, thanks to (2.1)

(ii) $\log(2+|t|)\, h(t) \in L^1(\mathbb{R})$, thanks to (i) and (2.4),

(iii) $M \in L^1(\mathbb{R})$, and $\int M = 1$ consequence of (i), (ii) and (2.1).

It follows that

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{+\infty} dy \frac{1}{\varepsilon} M\left(\frac{y-t}{\varepsilon}\right) s(y) = s(t).$$

On the other hand the hypothesis (2.3) on s implies that

$$\lim_{\rho \to 0} \int_{-\infty}^{+\infty} dy \frac{1}{\rho} M\left(\frac{y-t}{\rho}\right) s(y) = 0,$$

and this concludes the proof. □

Remark. It is clear that the same theorem holds for functions s and wavelets φ the Fourier transform of which is supported by the positive (negative) frequencies only. In this case a reconstruction wavelet needs only satisfy the second (third) equation of condition (2.1).

We shall call a representation of the form (2.5) of an arbitrary function as superposition of functions that are localized around a certain length scale

and position a *scale-space-representation* of the function s. The function T is called the *scale-space-coefficients* with respect to the reconstruction wavelet g. Note that we suppose that the limit $\varepsilon \to 0$ and $\rho \to \infty$ are independent. As we have seen those representations exist, but they are far from being unique – there are at least as many as there are reconstruction wavelets. Furthermore not all scale-space-representations arise from wavelet transforms.

3 The wavelet transform of Hölder continuous functions

In this section we recall how to analyze uniform Hölder regularity through wavelet transforms [13, 15]. Generally speaking the amount of uniform regularity of a function is reflected in its wavelet transform by the decrease of the wavelet coefficients at small scales as shown by the following well known theorem (for a proof see [10] or [15]).

(3.1) Theorem. *If a bounded function s satisfies with some constant $c>0$*

$$|s(t)-s(y)| \leqq c|t-y|^{\alpha}, \qquad \alpha \in (0, 1]$$

then its wavelet transform with respect to a wavelet φ satisfies

$$T(b, a)=o(a^{\alpha})$$

uniformly in b. The analysing wavelet φ should satisfy $\varphi \in L^1(\mathbb{R})$, $t^{\alpha}\varphi \in L^1(\mathbb{R})$ and $\int \varphi = 0$.

In turn a uniform decrease of the scale-space-coefficients in $O(a^{\alpha})$ implies Hölder regularity with exponent α, provided that $\alpha < 1$.

(3.2) Theorem. *If the scale-space-coefficients T with respect to a one time continuously differentiable, compactly supported reconstruction wavelet g of a function s satisfy at*

$$T(b, a)=o(a^{\alpha})$$

uniformly in b with some $\alpha \in (0, 1)$ then s is Hölder continuous with regularity exponent α.

Although this is a classical result (e.g. [10]) we recall for the convenience of the reader the main elements of the proof. By hypothesis we may write

$$\begin{aligned} s(t) &= \int_0^{\infty} \frac{da}{a} \int_{-\infty}^{+\infty} db \, \frac{1}{a} g\left(\frac{t-b}{a}\right) T(b, a) \\ &= \left\{ \int_0^1 \frac{da}{a} + \int_1^{\infty} \frac{da}{a} \right\} \int_{-\infty}^{+\infty} db \, \frac{1}{a} g\left(\frac{t-b}{a}\right) T(b, a) \\ &= s_{\text{small}}(t) + s_{\text{large}}(t). \end{aligned}$$

where from now on – unless otherwise stated – we mean by $\int_0^\infty \frac{da}{a}$ the generalized integral $\lim_{\varepsilon\to 0} \int_\varepsilon^{1/\varepsilon} \frac{da}{a}$. The integral over the large scales, $a>1$, yields a smooth function s_{large}. Therefore the local behaviour of s is determined by the small scale integral s_{small}. This integral is absolutely convergent, and we decompose it in the following way:

$$s_{\text{small}}(t+h)-s_{\text{small}}(t) = \int_0^h \frac{da}{a} \int_{-\infty}^{+\infty} db\, \frac{1}{a} g\left(\frac{t+h-b}{a}\right) T(b,a) \qquad (T_1)$$

$$- \int_0^h \frac{da}{a} \int_{-\infty}^{+\infty} db\, \frac{1}{a} g\left(\frac{t-b}{a}\right) T(b,a) \qquad (T_2)$$

$$+ \int_h^1 \frac{da}{a} \left\{\frac{1}{a} g\left(\frac{t+h-b}{a}\right) - \frac{1}{a} g\left(\frac{t-b}{a}\right)\right\} T(b,a). \qquad (T_3)$$

The first two terms are treated in a straightforward way by using the hypothesis on $T(b, a)$. For the last one we have to use the interaction between the wavelets:

$$g\left(\frac{t+h-b}{a}\right) - g\left(\frac{t-b}{a}\right) = \frac{h}{a} g'\left(\frac{t+\tau-b}{a}\right),$$

and the end of the proof follows. □

It already contains, in a very simple form, the main ingredients of the proofs in the subsequent theorems.

4 The wavelet analysis of local Hölder regularity

The local regularity of a function implies an equivalent local decrease of its wavelet coefficients at small scale as shown by the following theorem.

(4.1) Theorem. *Let s be a bounded, locally integrable function that satisfies*

$$s(t_0+h)-s(t_0)=O(h^\alpha), \qquad \alpha\in(0,1]$$

at some point $t_0\in\mathbb{R}$. Then its wavelet transform with respect to a wavelet φ satisfies

$$T(t_0+b,a)=O(a^\alpha+|b|^\alpha).$$

The wavelet is supposed to satisfy $\varphi\in L^1(\mathbb{R})$, $t^\alpha\varphi\in L^1(\mathbb{R})$ and $\int\varphi=0$.

Proof. The argument is well known: because of the fact that the wavelet is of zero mean, it is possible to make use of the local behaviour of s. By an overall translation we suppose that $t_0=0$ and write

$$T(b,a)=\int_{-\infty}^{+\infty} dt\, \frac{1}{a}\bar{\varphi}\left(\frac{t-b}{a}\right)(s(t)-s(0)).$$

The end of the proof is straightforward. □

We now turn our attention to the reciprocal problem. The following theorem is similar to a theorem proved before by S. Jaffard [11]. Both are of Tauberian type.

(4.2) Theorem. *Suppose that a bounded function s has a scale-space-representation with a one time continuously differentiable, compactly supported reconstruction wavelet g and coefficients T satisfying for some $\gamma>0$ and $\alpha\in(0,1)$*

(i) $T(b,a)=O(a^\gamma)$ *uniformly in b,*

(ii) $T(t_0+b,a)=O\left(a^\alpha+\frac{|b|^\alpha}{|\log b|}\right)(b,a\mapsto O)$.

Then s has at t_0 a local regularity exponent α in the sense that $s(t_0+h)-s(t_0)=O(h^\alpha)$.

Proof. By an overall translation and dilation which does not change the local regularity of s we may suppose that the support of g is contained in $[-1/2, +1/2]$ and that $t_0=0$.

As in the proof of Theorem (3.2) we only need to estimate the small scale integral

$$\Delta(h)=s_{\text{small}}(h)-s_{\text{small}}(0)$$
$$=\int_0^1 \frac{da}{a}\int_{-\infty}^{+\infty} db\left\{\frac{1}{a}g\left(\frac{h-b}{a}\right)-\frac{1}{a}g\left(-\frac{b}{a}\right)\right\}T(b,a).$$

We only treat the case $h>0$, the case $h<0$ being analogous. Then with the help of the function $\eta(h)=|h|^{\alpha/\gamma}$, $\alpha>\gamma$ we may split the integral into several parts.

$$\Delta(h)=\int_0^{\eta(h)}\frac{da}{a}\int_{-\infty}^{+\infty}db\,\frac{1}{a}g\left(\frac{h-b}{a}\right)T(b,a) \tag{T_1}$$

$$+\int_{\eta(h)}^{h}\frac{da}{a}\int_{-\infty}^{+\infty}db\,\frac{1}{a}g\left(\frac{h-b}{a}\right)T(b,a) \tag{T_2}$$

$$-\int_0^{h}\frac{da}{a}\int_{-\infty}^{+\infty}db\,\frac{1}{a}g\left(-\frac{b}{a}\right)T(b,a) \tag{T_3}$$

$$+\int_h^1\frac{da}{a}\int_{-\infty}^{+\infty}db\left\{\frac{1}{a}g\left(\frac{h-b}{a}\right)-\frac{1}{a}g\left(-\frac{b}{a}\right)\right\}T(b,a). \tag{T_4}$$

These terms will be estimated separately.

T_1. Using the global Hölder regularity of s we have $T=O(a^\gamma)$ which leads us to

$$|T_1|\leqq\int_0^{\eta(h)}\frac{da}{a}\int_{-\infty}^{+\infty}db\left|\frac{1}{a}g\left(\frac{h-b}{a}\right)\right||T(b,a)|$$
$$=O(1)\,\|g\|_{L^1(\mathbb{R})}\int_0^{\eta(h)}\frac{da}{a}a^\gamma$$
$$=O(h^\alpha).$$

T_2. Since the support of g is in $[-1/2, +1/2]$ we have under the integral $a\leqq h$ and $b\leqq 2h$. Therefore we may estimate $T=O(a^\alpha+h^\alpha/|\log h|)$. The term in

a^α may be estimated as in T_3 below to be of order $O(h^\alpha)$, and thus we end up with

$$\begin{aligned}|T_2| &\leqq \int_{\eta(h)}^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \left|\frac{1}{a} g\left(\frac{h-b}{a}\right)\right| |T(b,a)| \\ &= O(1)\frac{h^\alpha}{|\log h|} \int_{\eta(h)}^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \left|\frac{1}{a} g\left(\frac{h-b}{a}\right)\right| + O(h^\alpha) \\ &= \|g\|_{L^1(\mathbb{R})} O(1)\frac{h^\alpha}{|\log h|} \log\left(\frac{h}{\eta(h)}\right) + O(h^\alpha) \\ &= O(h^\alpha).\end{aligned}$$

T_3. This integral runs over the influence cone of g in 0. Inside this cone we have $T = O(a^\alpha)$. As term T_1 in the proof of Theorem (3.2) it may be estimated to be $T_3 = O(h^\alpha)$.

T_4. Using as in the proof of Theorem (3.2) the regularity of the reconstruction wavelet we may write

$$g\left(\frac{h-b}{a}\right) - g\left(-\frac{b}{a}\right) = \frac{h}{a} g'\left(\frac{\tau - b}{a}\right)$$

with some τ between 0 and h. Then,

$$\begin{aligned}|T_4| &\leqq \int_{h}^{1} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{a} \left| g\left(\frac{h-b}{a}\right) - g\left(-\frac{b}{a}\right)\right| |T(b,a)| \\ &= h \int_{h}^{1} \frac{da}{a} \int_{-\infty}^{+\infty} db \left|\frac{1}{a^2} g'\left(\frac{\tau - b}{a}\right)\right| |T(b,a)|.\end{aligned}$$

Since g' is together with g supported by $[-1/2, +1/2]$ the integral runs over the cone $-|b| \leqq 2a$ – so that $T = o(a^\alpha)$. This implies

$$|T_4| = O(h)\|g'\|_{L^\infty(\mathbb{R})} \int_{h}^{1} \frac{da}{a} a^{\alpha - 1} = O(h^\alpha).$$

This shows that s_{small} and therefore s is Hölder continuous with exponent α, which proves the theorem. □

5 Pointwise differentiability and wavelet analysis

In this section we shall study pointwise differentiability through wavelet transforms. As is shown in the following theorem the local smoothness is mirrored on the wavelet side by a rapid decrease of the wavelet coefficients.

(5.1) Theorem. *Let s be a bounded function that is differentiable in t_0. Then its wavelet transform satisfies:*

$$T(t_0 + b, a) = o(|b| + a).$$

The analysing wavelet should satisfy $\varphi \in L^1(\mathbb{R})$, $t\varphi \in L^1(\mathbb{R})$ and $\int \varphi = \int t\varphi = 0$.

The proof is similar to that of Theorem (4.1), and is left to the reader. Now we want to relate the local decrease of the scale-space-coefficients of the function to the differentiability of the function itself. As in the previous section, we will see that it is not possible to prove the exact converse of the preceding theorem: a slightly stronger hypothesis on the wavelet side is needed to prove the differentiability of a function through its scale-space-coefficients.

(5.2) Theorem. *If the scale-space-coefficients T of s with respect to a compactly supported, two times continuously differentiable reconstruction wavelet g satisfy*

(i) $T(b, a) = O(a^{\alpha})$ *uniformly in* b

(ii) $T(t_0 + b, a) = O(a\rho(a) + |b|\rho(b))$

with an arbitrary $\alpha > 0$ *and with a monotone function* $\rho = \rho(|a|)$ *satisfying the condition of Dini:*

$$\int_0^1 \frac{da}{a} \rho(a) < \infty,$$

then s is differentiable in t_0. Furthermore the condition on ρ is optimal.

Proof. By an overall dilation and translation, which does not change the local regularity we may suppose that the support of g is contained in $[-1/2, +1/2]$ and that $t_0 = 0$. Let $\Delta(h) = h^{-1}(s_{\text{small}}(h) - s_{\text{small}}(0))$. We want to show that $\lim_{h \to 0} \Delta(h)$ exists. This is equivalent to say that $\Delta(h, h') = \Delta(h) - \Delta(h')$ goes to 0 together with h and h'. By hypothesis we may write (in the following g_h will stand for $\frac{1}{a} g\left(\frac{h-b}{a}\right)$)

$$\Delta(h, h') = \int_0^1 \frac{da}{a} \int_{-\infty}^{+\infty} db \left\{ \frac{1}{h}(g_h - g_0) - \frac{1}{h'}(g_{h'} - g_0) \right\} T(b, a).$$

Let $\eta(h) = |h|^{\frac{2}{\alpha}}$. Again we only consider the case $h > 0$ since the case $h < 0$ is analogous. We may suppose that $0 < h' < h$. We now cut the integral over the half plane into several parts.

$$\Delta(h, h') = \int_0^{\eta(h)} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{h} g_h T(b, a) \tag{T$_1$}$$

$$- \int_0^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{h} g_0 T(b, a) \tag{T$_2$}$$

$$+ \int_{\eta(h)}^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{h} g_h T(b, a) \tag{T$_3$}$$

$$- \int_0^{\eta(h')} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{h'} g_{h'} T(b, a) \tag{T$_4$}$$

$$+ \int_0^{h'} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{h'} g_0 T(b, a) \tag{T$_5$}$$

$$- \int_{\eta(h')}^{h'} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{h'} g_{h'} T(b, a) \tag{T$_6$}$$

$$- \int_{h'}^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{h'} (g_{h'} - g_0) T(b, a) \tag{T$_7$}$$

$$+ \int_{h}^{1} \frac{da}{a} \int_{-\infty}^{+\infty} db \left\{ \frac{1}{h}(g_h - g_0) - \frac{1}{h'}(g_{h'} - g_0) \right\} T(b, a). \tag{T$_8$}$$

These terms will be treated separately.

T_1, T_4. Going to the modulus yields

$$h|T_1| \leqq \int_0^{\eta(h)} \frac{da}{a} \int_{-\infty}^{+\infty} db \left| \frac{1}{a} g\left(\frac{h-b}{a}\right) \right| |T(b, a)|.$$

Majorizing $T = O(a^\alpha)$ we obtain

$$h|T_1| = O(1) \|g\|_{L^1(\mathbb{R})} \int_0^{\eta(h)} \frac{da}{a} a^\alpha = o(h)$$

by hypothesis on η, and thus $T_1 = o(1)$. Replacing h by h' we find $T_4 = o(1)$ too.

T_2, T_5. Estimating the modulus yields

$$h|T_2| \leqq \int_0^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \left| \frac{1}{a} g\left(-\frac{b}{a}\right) \right| |T(b, a)|.$$

The integral runs over the influence cone of g in 0; that is we have $|b| \leqq a/2$. Therefore we may estimate $T = O(a\rho(a))$ yielding

$$h|T_2| = O(1) \|g\|_{L^1(\mathbb{R})} \int_0^{h} \frac{da}{a} a\rho(a) = o(h)$$

since ρ satisfies the condition of Dini. This proves $T_2 = o(1)$ and replacing h by h' the same holds for T_5.

T_3, T_6. Again we estimate the modulus

$$h|T_3| \leqq \int_{\eta(h)}^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \left| \frac{1}{a} g\left(\frac{h-b}{a}\right) \right| |T(b, a)|.$$

Because the support of g is contained in $[-1/2, +1/2]$ we can verify that under the integral $|b|$ never exceeds $2h$. Since ρ was chosen monotone we may estimate $T = O(h\rho(2h))$. All this together yields

$$|T_3| = O(1)\rho(2h) \|g\|_{L^1(\mathbb{R})} \int_{\eta(h)}^{h} \frac{da}{a} = O(1)\rho(2h) \log(h)$$

by hypothesis on η. Since ρ satisfies the condition of Dini and is monotone it follows that $\log(h)\rho(h) \to 0$ as $h \to 0$. This shows that $T_3 = o(1)$. The same holds for T_6 as is seen when replacing h by h'.

T_7. The usual estimate yields

$$|T_7| \leqq \int_{h'}^{h} \frac{da}{a} \int_{-\infty}^{+\infty} db \frac{1}{a} \frac{1}{h'} \left| g\left(\frac{h'-b}{a}\right) - g\left(-\frac{b}{a}\right) \right| |T(b, a)|.$$

Since g is one time differentiable with derivative g' we may write:

$$g\left(\frac{h'-b}{a}\right) - g\left(-\frac{b}{a}\right) = \frac{h'}{a} g'\left(\frac{\tau-b}{a}\right)$$

with $\tau \in [0, h']$. Together with g its derivative is supported by $[-1/2, +1/2]$ and again the integral runs over a cone $-|b| \leqq 2a$ – which implies $T = O(a\rho(2a))$. All together yields

$$|T_7| = O(1) \|g\|_{L^\infty(\mathbb{R})} \int_{h'}^{h} \frac{da}{a} \rho(2a) = o(1),$$

since ρ satisfies the condition of Dini.

T_8. The last term we must estimate reads

$$|T_8| \leqq \int_h^1 \frac{da}{a} \int_{-\infty}^{+\infty} db \left| \frac{1}{h}(g_h - g_0) - \frac{1}{h'}(g_{h'} - g_0) \right| |T(b, a)|.$$

Using the fact that g is two times continuously differentiable with second derivative g'' we may write

$$\frac{1}{h}(g_h - g_0) - \frac{1}{h'}(g_{h'} - g_0) = \frac{\sigma}{a^2} g''\left(\frac{\tau - b}{a}\right)$$

with $\sigma, \tau \in [0, h]$. Again the integral runs over a cone $-|b| \leqq 2a$ – because g'' is supported by $[-1/2, +1/2]$. Estimating again $T = O(a\rho(2a))$ we obtain

$$|T_8| = O(1) \|g''\|_{L^\infty(\mathbb{R})} h \int_h^1 \frac{da}{a} \rho(2a).$$

Clearly $\lim_{a \mapsto 0} \rho(a) = 0$ and therefore

$$\limsup_{h \mapsto 0} h \int_h^1 \frac{da}{a} \rho(2a) = 0.$$

This shows that $T_8 = o(1)$ and thus the first part of the theorem is proved.

That the condition on ρ is actually the weakest possible can be seen as follows. Consider a function g in the Schwartz-class whose Fourier transform is supported by $[1, 2]$ such that the functions $2^{j/2} g(2^j t)$, $j = 1, 2 \ldots$ are an orthonormal set. In addition we suppose that $g'(0) = 1$. Then let

$$s(t) = \sum_{j=1}^{\infty} 2^{-j} \rho(2^{-j}) g(2^j t)$$

with a continuous, positive, monotone, bounded function ρ. The wavelet transform of s with respect to g satisfies

$$T(0, 2^{-j}) = 2^{-j} \rho(2^{-j}), \quad j = 1, 2 \ldots$$

Now s is differentiable for $t \neq 0$ and therefore $\Delta(h) = h^{-1}(s(h) - s(0))$ can be written as

$$\Delta(h) = \sum_{j=1}^{\infty} \rho(2^{-j}) g'(2^j \tau_j),$$

with some $\tau_j \in [0, h]$. Now ρ satisfies a condition of Dini if and only if

$$\sum_{j=1}^{\infty} \rho(2^{-j}) < \infty$$

and therefore this same condition implies the differentiability of s at $t=0$.

On the other hand suppose that s is differentiable at $t=0$. By an overall dilation we may suppose that $g'(t) \sim 1$ for $t \in [-1, 1]$. Then

$$\Delta(h) \sim \sum_{2^{-j} \geqq \tau_j} \rho(2^{-j}) + \sum_{2^{-j} < \tau_j} \rho(2^{-j}) g'(2^j \tau_j) = T_1 + T_2.$$

Since ρ is uniformly bounded, T_2 stays bounded as $h \to 0$. The same then must be true for T_1. This shows that ρ satisfies a condition of Dini and so the wavelet transform of s satisfies the condition of Theorem (5.2) if and only if s is differentiable in 0. □

6 The function of Riemann-Weierstrass

In this section we will study the local behaviour of W with the help of a suitable chosen wavelet transform. Instead of W we will consider the following family of complex-valued functions over the real line:

$$W_\beta(t) = \frac{2}{\pi^\beta} \sum_{n=1}^{\infty} n^{-2\beta} e^{i\pi n^2 t}, \quad \text{for } \beta > 1/2. \tag{6.1}$$

The classical function of Riemann and Weierstrass (1.1) is the imaginary part of W_1. By direct computation we can verify that the following scale-space-representation for W_β holds

$$W_\beta(t) = \int_0^\infty \frac{da}{a} \int_{-\infty}^{+\infty} db \, \frac{1}{a} g\left(\frac{t-b}{a}\right) T(b+ia) \tag{6.2}$$

with $T(\tau) = (\operatorname{Im}\tau)^\beta \vartheta(\tau)$, $\tau = b + ia$, $\operatorname{Im}\tau = a$ and

$$\vartheta(\tau) = 1 + 2 \sum_{n=1}^{\infty} e^{i\pi n^2 \tau}, \quad \text{for } \operatorname{Im}\tau > 0. \tag{6.3}$$

$\vartheta(\tau)$ is a Jacobi theta function and g any integrable function that satisfies

$$\int_0^\infty \frac{da}{a} \hat{g}(a) a^\beta e^{-a} = 1 \quad \text{and} \quad \hat{g}(0) = 0. \tag{6.4}$$

Thus the analysis of the local behaviour of W_β is transformed into the investigation of the behaviour of an analytic function near the boundary of its domain of analyticity.

The fact that analytic functions appear as scale space representations of functions is already implicit in [10]. In the context of Quantum Mechanics, T. Paul has systematically investigated the representations by the functions [16]

$$\varphi_m(t)=\frac{1}{2\pi}\frac{\Gamma(m+1)}{(1-it)^{m+1}}, \quad \text{for fixed } m>0. \tag{6.5}$$

Looking at the family of their dilates and translates we see that we can write[5]

$$\frac{1}{a}\bar{\varphi}_m\left(\frac{t-b}{a}\right)=\frac{1}{2\pi}(\operatorname{Im}\tau)^m\Gamma(m+1)\left(\frac{i}{\tau-t}\right)^{m+1}.$$

Thus a wavelet analysis of a bounded function with respect to these wavelets yields – up to the prefactor $(\operatorname{Im}\tau)^m$ – to an analytic function over the complex upper half plane. The first condition in (6.4) says that g is a reconstruction wavelet for φ_β.

The wavelet transform of W_β with respect to the wavelet φ_β is given by

$$T(b,a)=T(\tau)=(\operatorname{Im}\tau)^\beta(\vartheta(\tau)-1). \tag{6.6}$$

Note however that the analysing wavelet does not satisfy the hypothesis of Theorem (5.1). Therefore the apparent decrease of at most $O(a)$ does not tell us anything about the local differentiability properties of W. However since it is known [19] that

$$\vartheta(\tau)=O((\operatorname{Im}\tau)^{-1/2}) \tag{6.7}$$

it follows from Theorem (3.2) as a first result that W_β satisfies uniformly a Hölder condition with regularity exponent $\beta-1/2$ for $\beta\in(1/2, 3/2)$.

We now come to the analysis of ϑ near the real axis. The following transformation formulas for ϑ are known

$$\vartheta(K^2\tau)=\vartheta(\tau), \qquad \vartheta(U\tau)=\sqrt{-i\tau}\,\vartheta(\tau) \tag{6.8}$$

where $K:\tau\to\tau+1$ and $U:\tau\to-\frac{1}{\tau}$ are the generators of the modular group. The subgroup generated by K^2 and U is called the *theta* group G_ϑ. It leaves invariant the real axis and the rationals, which split into two orbits [18]: the orbit of 1 consisting of all rationals of the form $(2P+1)/(2Q+1)$ and the orbit of 0 consisting of the rationals of the form $(2P+1)/(2Q)$ and their inverses. These two sets will be treated separately.

[5] In all formulas involving powers of complex numbers we use the convention that the argument of a complex number is continuous with the exception of the negative real axis.

7 The orbit of 0

The local behaviour of the scale-space-representation (6.2) is determined by the following theorem

(7.1) Theorem. *At any point t in the orbit of 0 the scale-space-coefficients look like*

$$(\mathrm{Im}\,\tau)^{\beta}\,\vartheta(t+\tau)=C_t(\mathrm{Im}\,\tau)^{\beta}\sqrt{\frac{i}{\tau}}+O(\tau^{2\beta-1/2}) \tag{7.2}$$

for all $\beta>1/2$. The constant C_t satisfies the following transformation formulas

$$C_{K^2t}=C_t, \qquad C_{Ut}=\sqrt{-\frac{i}{t}}\,C_t$$

which together with $C_0=1$ determine C_t along the orbit of 0 under G_ϑ.

The proof of this theorem is an immediate consequence of the following two Lemmas.

(7.3) Lemma. *Theorem* (7.1) *holds at $t=0$.*

Proof. From the series expansion of ϑ (6.3) we can easily read the following estimation

$$\vartheta(\tau)=1+O(e^{-Im\tau}) \qquad (\mathrm{Im}\,\tau\to\infty). \tag{7.4}$$

From the transformation behaviour (6.8) of ϑ we can write

$$\vartheta(\tau)=\sqrt{\frac{i}{\tau}}\,\vartheta\left(-\frac{1}{\tau}\right).$$

Using now the estimate (7.4) and the global estimate (6.7) with τ replaced by $-1/\tau$ we can read the following behaviour at 0

$$\vartheta(\tau)=\sqrt{\frac{i}{\tau}}+\rho(\tau)$$

and the remainder can be estimated by

$$\rho(b+ia)=\begin{cases} O((b^2+a^2)^{-1/4}e^{-a/(b^2+a^2)}), & \text{for } \dfrac{a}{b^2+a^2}>1 \\ O(a^{-1/2}(b^2+a^2)^{1/4}), & \text{uniformly.}\end{cases}$$

It is not hard to see that

$$a^{\beta}\rho(b+ia)=O((b^2+a^2)^{\beta-1/4}).$$

All together actually shows that (7.2) holds at $t=0$. □

It remains to show that everything is well behaved under the action of the theta group. Therefore it is enough to check on the generators of G_ϑ. Since this is of course true for K^2 the theorem will be proved by the following lemma.

(7.5) Lemma. *Suppose that at some* $t \in \mathbb{R}$, $t \neq 0$

$$\vartheta(t+\tau) = C_t \sqrt{\frac{i}{\tau}} + \rho(\tau)$$

holds with ρ *satisfying the estimate* (7.2). *Then at* $Ut = -\frac{1}{t}$ *we have*

$$\vartheta\left(-\frac{1}{t}+\tau\right) = \sqrt{-\frac{i}{t}}\, C_t \sqrt{\frac{i}{\tau}} + \tilde{\rho}(\tau)$$

where $\tilde{\rho}$ *satisfies again the estimate* (7.2).

Proof. By direct computation using the transformation behaviour (6.8) of ϑ we can write

$$\vartheta\left(-\frac{1}{t}+\tau\right) = \sqrt{\frac{it}{t\tau-1}}\, \vartheta\left(t+\frac{t^2\tau}{1-t\tau}\right)$$

and by hypothesis on the behaviour of ϑ in t

$$\vartheta\left(-\frac{1}{t}+\tau\right) = C_t \sqrt{\frac{1}{t\tau}} + \rho\left(\frac{t^2\tau}{1-t\tau}\right).$$

Since $1/(1-t\tau) \simeq 1$ for $\tau \to 0$ the remainder satisfies again the estimate (7.2). Therefore the lemma is proved and so is the Theorem (7.1). □

We now interprete these results in terms of W_β by means of the following theorem. It turns out that W_β is in the exact scaling class as was conjectured in [8].

(7.6) Theorem. *At any point* t *in the orbit of* 0 *the function of Riemann and Weierstrass has local cusps of the following explicit form* ($\gamma = \beta - 1/2$)

$$W_\beta(t+h) - W_\beta(t) = C_t^- |h|_-^\gamma + C_t^+ |h|_+^\gamma + \rho(h)$$

with $|h|_\pm = |h \pm |h||/2$. *The remainder* ρ *is differentiable in* 0 *for* $\beta > 3/4$ *whereas for* $1/2 < \beta \leqq 3/4$ *it is of local Hölder regularity with exponent* $2\beta - 1/2 - \varepsilon$ *for all* $\varepsilon > 0$. *The two complex constants* $C_t^\pm$ *determining the local behaviour of* W_β *to the right (to left) of* t *obey the transformation equations*

$$C^\pm_{K^2 t} = C_t^\pm, \qquad C^\pm_{Ut} = \sqrt{-\frac{i}{t}}\, C_t^\pm$$

which together with

$$C_0^\pm = -\frac{\sqrt{\pi}}{\Gamma(\gamma+1)\sin(\pi\gamma)}\, e^{\mp \gamma \pi i/2}$$

determine $C_t^\pm$ *along the orbit of* 0 *under* G_ϑ.

Proof. Consider the function

$$s_{\text{loc}}(t) = c_- |t|_-^\alpha + c_+ |t|_+^\alpha$$

with some $\alpha \in (0, 1]$ and two complex constants $c_\pm$ satisfying

$$c_- = e^{\alpha \pi i} c_+$$

in order to make sure that the Fourier transform of s_{loc} – in the sense of distribution – is supported by the positive frequencies only. The transformation of s_{loc} with respect to the wavelet φ_m (6.5) is well defined for $m > \alpha$, and yields

$$\tilde{T}(\tau) = -\frac{1}{\pi} \Gamma(\alpha+1)\, \Gamma(m-\alpha) \sin(\alpha\pi)\, e^{\alpha \pi i/2} c_+ (\operatorname{Im} \tau)^m \left(\frac{i}{\tau}\right)^{m-\alpha},$$

as is shown by direct computation. For $\alpha = \beta - 1/2$, $m = \beta$ and $c_\pm$ as stated in the theorem, the local behaviour of the scale-space-coefficients of W_β at any point in the orbit of 0 is – up to a remainder (7.2) – of the same type as $\tilde{T}$. Since we may estimate the small scale contribution of the remainder with the help of Theorems (5.2) and (4.2) respectively it is enough to show that the small scale reconstruction integral over $\tilde{T}$ gives back – up to an additional smooth function – the local behaviour of s_{loc}. Therefore let g be a real valued, symmetric, compactly supported reconstruction wavelet for φ_m, that is in the class of Schwartz, to be chosen later. By the proof of Theorem (2.2) we have

$$\int_0^1 \frac{da}{a} \int_{-\infty}^{+\infty} db\, \frac{1}{a} g\left(\frac{t-b}{a}\right) \tilde{T}(b + ia) = s_{\text{loc}}(t) - (M * s_{\text{loc}})(t),$$

where M is essentially given by the formula (2.6). But since φ_m is supported by the positive frequencies only we rather have

$$\hat{M}(\omega) = \int_{|\omega|}^{\infty} \frac{da}{a} a^m e^{-a} \hat{g}(a).$$

If we choose g to have sufficiently many vanishing moments then M is smooth and localized enough so that $M * s_{\text{loc}}$ is smooth too, which proves the theorem. □

(7.7) Corollary. *The classical function of Riemann has a dense set of points where it is differentiable to the right (left) but not to the left (right).*

Proof. Since $\beta = 1$ we have $\arg C_0^{+(-)} = \pi/4$. The action of the translation K^2 does not change the argument, whereas U turns the phase by $\pm \pi/4$ depending on whether $t > 0$ or $t < 0$. Therefore we can by successive application of K^{2n} and U construct an infinite and dense set of points $\{t\}$ for which the argument of the local constant $C_t^{+(-)}$ is equal to 0 or π, and therefore the imaginary part of $C_t^{+(-)}$ vanishes. Since the function of Riemann is the imaginary part of W_1, the differentiability to the right (to the left) follows from the differentiability of the remainder in the previous theorem. □

8 The orbit of 1

Along the orbit of 1 under G_ϑ the local regularity of W_β is characterized by the following theorem.

(8.1) Theorem. *At any point of the orbit of* 1 *under the group* G_ϑ *the function of Riemann-Weierstrass is differentiable for* $\beta > 3/4$. *For* $1/2 < \beta \leqq 3/4$ *it is locally Hölder continuous with exponent* $2\beta - 1/2 - \varepsilon$ *for all* $\varepsilon > 0$.

Proof. Denote $\Theta(\tau) = \vartheta(1+\tau)$ the translate of ϑ. Then we may write

$$\begin{aligned}\Theta(\tau) &= \sum_{j=-\infty}^{\infty} (-1)^j e^{i\pi j^2 \tau} \\ &= \sum_{j\,\text{even}} e^{i\pi j^2\tau} - \sum_{j\,\text{odd}} e^{i\pi j^2\tau} \\ &= 2 \sum_{j\,\text{even}} e^{i\pi j^2\tau} - \vartheta(\tau) \\ &= 2\vartheta(4\tau) - \vartheta(\tau).\end{aligned}$$

Therefore the local behaviour of ϑ in the points of the orbit of 0 implies that Θ satisfies the same estimate as the remainder in (7.2) in every point of the orbit of 1. The theorem then follows then follows from Theorem (5.2) and Theorem (4.2). □

9 The irrational points

For any irrational point t, Hardy and Littlewood have shown [3] that there is a positive constant c such that $a_n^{1/4} |Im\,\vartheta(t + i a_n)| > c$ and $a_n^{1/4} |\mathrm{Re}\,\vartheta(t + i a_n)| > c$ for a sequence of positive real numbers $a_n \to 0$. This implies via Theorem (4.1) that neither W_β nor its real or imaginary part can be Lipschitz continuous and hence they are not differentiable at any irrational point for any $\beta \in (1/2, 5/4)$ since the wavelet transform (6.6) is only decreasing as $O(a^{1-\varepsilon})$, $\varepsilon > 0$ at these points.

We remark here that this proof follows essentially the reasoning of Hardy [2] who has proved the nondifferentiability of W_β at these points before. He used the angular derivative of the Poisson integral for functions on the circle, which in this case was the theta function ϑ.

10 Conclusion and outlook

We have shown that the wavelet transform is a powerful tool to analyze the local regularity of functions. In the case of the function of Riemann we could connect its local analysis to the investigation of a Jacobi theta function near its domain of analyticity. It is clear that the same kind of analysis applies to other functions associated to modular forms through wavelet transforms.

Acknowledgements. We want to thank Professor Y. Meyer for interesting discussions and for carefully reading the manuscript. One of us (M.H.) wants to express his thanks for warm

hospitality at the CEREMADE at the university of Paris Dauphine where part of the work was carried out. For exactly the same reason, the second author wants to gratefully thank M. and M. Benedetti.

References

1. Weierstrass, K.: Über continuirliche Functionen eines reellen Arguments, die für keinen Werth des letzteren einen bestimmten differentialquotienten besitzen. Mathematische Werke II, 71–74. Königl. Akad. Wiss.: 1872
2. Hardy, G.H.: Weierstrass's nondifferentiable function, Trans. Am. Math. Soc. **17**, 301–325 (1916)
3. Hardy, G.H., Littlewood, E.: Some problems of diophantine approximation II. Acta Math. **37**, 194–238 (1914)
4. Gerver, J.: The Differentiability of the Riemann function at certain rational multiples of π. Am. J. Math. XCII., Nr. 1 (1970)
5. Gerver, J.: More on the differentiability of the Riemann function. Am. J. Math. **93**, 33–41 (1970)
6. Queffelec, M.: Dérivabilité de certains sommes de séries de Fourier lacunnaires. Note au C.R. Séan. Acad. Sci. **273**, Sér. A, 291–293 (2.8.1971)
7. Queffelec, M.: Dérivabilité de certains sommes de séries de Fourier lacunnaires. Centre de Orssay, Université Paries-Sud (Thèse 3^e cycle) juin 1971
8. Holschneider, M.: On the Wavelet transform of Fractal Objects. J. Stat. Phys. **5/6**, 963–993 (1988)
9. Calderon, A.P., Zygmund, A.: Local properties of solutions of elliptic partial differential equations. Stud. Math. **20**, 171–225 (1961)
10. Stein, E.: Singular Integrals and Differentiability Properties of Functions. Princeton: Princeton University Press 1979
11. Jaffard, S.: Exposants de Holder en des points donnes et coefficients d'ondelettes. C.R. Acad. Sci., Paris, Sér. I. **308**, 7 (1989)
12. Grossmann, A., Morlet, J.: In: Streit, L. (ed.) Mathematics and Physics, Lectures on recent results. Singapore: World Scientific Publishing 1987
13. Lemarie, P.G., Meyer, Y.: Ondelettes et bases Hilbertiennes. Rev. Mat. Iberoam. **1**, 1286 (1987)
14. Arneodo, A., Grasseau, G., Holschneider, M.: On the Wavelet transform of Multifractals. Phys. Rev. Lett. **61**, Nr. 20 (1988)
15. Coifman, R., Meyer, Y.: Au délà des opérateurs pseudo-différentiels. Astérisque **57** (1978)
16. Paul, T.: Ondelettes et mécanique quantique. Thesis. Marseile (1985)
17. Apostol, T.M.: Modular Functions and Dirichlet Series in Number Theory. Berlin Heidelberg New York: Springer (1976)
18. Gunning, R.C.: Lectures on Modular Forms. Princeton: Princeton University Press 1962
19. Mumford, D.: Tata Lectures on Theta 1. Boston: Birkhäuser (1983)

Invent math. N 105, 177–183 (191) (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Finite presentations of pro-p groups and discrete groups

John S. Wilson

Christ's College, Cambridge CB2 3BU, UK

Oblatum 16-X-1990

1 Introduction

Our object in this paper is to extend the Golod-Šafarevič theorem for finite groups to a large class of infinite pro-p groups and discrete groups, including all soluble groups. We write $d(G)$ for the minimal number of generators of a group G: of course, in the case of pro-p groups this notation refers to topological generators.

Theorem A. *Let G be a pro-p group which has a finite presentation with n generators and r relations, and suppose $d = d(G) > 1$. Then either*

$$r \geqq n + \tfrac{1}{4}d^2 - d, \tag{1}$$

or for each finitely generated dense discrete subgroup X of G there is a closed normal subgroup K of G such that XK/K is an infinite (finitely generated) torsion group.

Theorem B. *Let G be a discrete group which has a finite presentation with n generators and r relations, and let $d = d(G^{\mathrm{ab}})$ (where G^{ab} is the abelianization of G). Then either*

$$r \geqq n + \tfrac{1}{4}(d^2 - 1) - d, \tag{2}$$

or for some prime p, G has a normal subgroup K such that G/K is an infinite residually finite p-torsion group.

The inequality (1) for finite p-groups is essentially the Golod-Šafarevič theorem [3], as improved by Vinberg [12] and Gaschütz. It was shown by Koch [5] that (1) holds for p-adic analytic groups and by Lubotsky [8] that (2) holds for groups whose pro-p completion is p-adic analytic for each prime p. Since finitely generated linear torsion groups are finite (cf. Wehrfritz [13, Corollary 4.9]) and since quotients of p-adic analytic groups by closed subgroups are again

p-adic analytic (see Lazard [7, 3.2.3.5]), Theorems A and B include the earlier results.

The classes of groups for which one can conclude that the inequalities (1), (2) hold are rather extensive. Since finitely generated discrete soluble torsion groups are finite, these inequalities hold for finite presentations of soluble pro-p and discrete groups. Indeed Zel'manov [14] has shown that every finitely generated discrete residually finite p-torsion group which satisfies a law (in the sense of Neumann [10]) is finite, and so (1), (2) hold for presentations of groups which satisfy a law. Finally, taking the case of discrete groups, we note that the class of groups whose finitely generated subgroups have no infinite residually finite p-torsion images is closed with respect to extensions, direct limits and quotients, and so contains, for example, all groups obtained by iterating transfinitely the operations of taking extensions and direct limits, starting from finite groups and abelian groups.

The link between Theorem A and Theorem B comes from taking pro-p completions, an idea pioneered by Lubotsky [8]. Theorems A and B assert that either a finite presentation contains many relations or the group determined by the presentation is 'large'. The sense in which the group is 'large' may at first sight seem surprising; it is however a direct consequence of the method of proof, which resembles the procedure first used by Golod [2] to construct infinite finitely generated torsion groups. It is natural to speculate that the groups for which the inequalities (1), (2) fail do, in fact, have free subgroups of rank 2: proving this is likely to be difficult.

We shall show that Theorem A and Theorem B give some information about the growth in the number of generators of subgroups of finite index in finitely presented groups and about the kernels of maps from finitely presented pro-p groups to $\mathbb{Z}_p$, the infinite procyclic pro-p group.

Corollary A. *Let G be a finitely presented pro-p group and assume that there do not exist a closed subgroup K of G and a finitely generated discrete subgroup X of G normalizing K such that XK/K is an infinite torsion group. Then*

(i) *there is a constant $k > 0$ such that $d(H) \leqq k|G : H|^{\frac{1}{2}}$ for each open subgroup H of G, and*
(ii) *if N is any (closed) normal subgroup of G such that $G/N \cong \mathbb{Z}_p$, then N is finitely generated.*

Corollary B. *Let G be a finitely presented discrete group and assume that there do not exist a prime p and subgroups X, K such that $K \triangleleft X$ and X/K is an infinite finitely generated residually finite p-torsion group. Then there is a constant $k > 0$ such that $d(H^{\mathrm{ab}}) \leqq k|G : H|^{\frac{1}{2}}$ for each subgroup H of finite index in G.*

Of course the hypotheses of Corollary A and Corollary B are satisfied by soluble groups. Corollary A(ii) invites comparison with the case of discrete soluble groups. (For an account of the extensive theory of finitely presented discrete soluble groups, see Strebel [11].) It is by no means true that the kernel K of every map from a finitely presented discrete soluble group G to $\mathbb{Z}$ is finitely generated: the metabelian group $\langle x, y; y^{-1}xy = x^2\rangle$ illustrates this, and indeed, if G is a finitely presented metabelian group which has the above property, then G must be polycyclic (see Bieri and Strebel [1, Theorem A]).

We refer the reader to Koch [6] for elementary material concerning pro-p groups and their presentations, and to Lyndon and Schupp [9] for presentations of discrete groups.

2 Proof of Theorem A

We shall use a version of the main lemma in the proof of the Golod-Šafarevič theorem for finite groups which incorporates improvements due to Vinberg [12], Gaschütz, and Koch [5]. A convenient reference is Koch [6, Chapter 7]. First we recall briefly the definition of the Zassenhaus filtration of a free pro-p group. Let A be the formal power series algebra over $\mathbb{F}_p$, the field of p elements, in non-commuting indeterminates $x_1, \ldots, x_d$. Then the elements $1+x_1, \ldots, 1+x_d$ are units in A, and the (closed) multiplicative group F generated by these elements is the free pro-p group with free generators $1+x_1, \ldots, 1+x_d$ (see [6, Satz 7.16]). Let I be the ideal of A consisting of all power series with zero constant term. The Zassenhaus filtration of F is the descending chain

$$F = F_1 \geqq F_2 \geqq F_3 \geqq \cdots$$

of (closed) normal subgroups of F, where

$$F_n = \{ g \in F; g - 1 \in I^n \}$$

for each n. An element $g \in F \setminus \{1\}$ has *degree* n with respect to this filtration if $f \in F_n \setminus F_{n+1}$. The result that we need is the following lemma (see [6, Satz 7.20]):

Lemma 1. *Let F be the free pro-p group on d generators and let $\{S_n; n \geqq 2\}$ be a family of disjoint finite subsets such that S_n consists of elements of degree at least n for each n. Let R be the (closed) normal subgroup of F generated by $\bigcup\{S_n; n \geqq 2\}$. Define the real power series $\psi(t)$ by*

$$\psi(t) = 1 - dt + r_2 t^2 + r_3 t^3 + \cdots,$$

where $r_n = |S_n|$ for each n, and suppose that $\psi(t)$ is convergent for $0 < t < 1$. If the group $G = F/R$ is finite, then $\psi(t) > 0$ for $0 < t < 1$.

Now if a pro-p group G has a presentation with n generators and r relations then it also has a presentation with $d(G)$ generators and $r - (d(G) - n)$ relations (see for example Lubotsky [8, Lemma 1.1]). Moreover, if $G \cong F/R$, where F is free of rank d and $d(G) = d$, then R must be contained in F_2, the second term of the Zassenhaus filtration. Therefore Theorem A is essentially the case $m = 2$ of the following result:

Theorem A′. *Suppose that $G = F/R$, where F is the free pro-p group of rank $d > 1$ and where R is the (closed) normal subgroup generated by $U = \{u_1, \ldots, u_r\}$. If U is contained in F_m, where $m \geqq 2$ (and where (F_i) is the Zassenhaus filtration of F), and if*

$$r < \frac{(m-1)^{m-1}}{m^m} d^m, \tag{3}$$

then for each finitely generated dense discrete subgroup X of G there is a closed normal subgroup K of G such that XK/K is an infinite torsion group.

Proof. Let $\varepsilon \geqq 0$ and consider the function

$$\psi_\varepsilon(t) = 1 - dt + rt^m + \varepsilon t^m/(1-t).$$

We claim that $\psi_0(t_1) < 0$ for some $t_1 \in (0,1)$. We note that

$$\psi_0(t_0) < 0 \quad \text{where} \quad t_0 = \left(\frac{d}{mr}\right)^{\frac{1}{m-1}}.$$

Therefore if the claim is false then $t_0 > 1$, so that $d > mr$, and also

$$0 \leqq \psi_0(1) = 1 - d + r < 1 - \left(\frac{m-1}{m}\right)d.$$

Since $d \geqq 2$ this is impossible, and the claim follows. By continuity there is a positive ε with $\psi_\varepsilon(t_1) < 0$. Choose an integer q with $q\varepsilon \geqq 1$.

We list the elements $\bar{s}_1, \bar{s}_2, \ldots$ of X and choose for each $\bar{s}_i$ a preimage $s_i \in F$. Let

$$v_i = s_i^{p^{iq+m}}$$

for each i. Then $v_i \in F_{iq+m}$ for each m; indeed, calculating in the formal power series ring A introduced above, we have

$$v_i - 1 = (s_i - 1)^{p^{iq+m}} \in I^{p^{iq+m}} \leqq I^{iq+m}.$$

Let $S = U \cup \{v_i\,; i \geqq 1\}$, and suppose that the quotient of F by the (closed) normal subgroup R_+ generated by S is finite. Then $\phi(t_1) > 0$ by Lemma 1, where

$$\phi(t) = 1 - dt + rt^m + \sum_{i=1}^{\infty} t^{qi}.$$

But for $t \in (0,1)$ we have

$$qt^q \leqq 1 + t + \cdots + t^{q-1}, \quad \text{and so} \quad t^q \leqq \varepsilon(1 + t + \cdots + t^{q-1}).$$

Thus

$$\phi(t) \leqq 1 - dt + rt^m + \varepsilon t^m \sum_{j=0}^{\infty} t^j = \psi_\varepsilon(t).$$

This is a contradiction. It follows that F/R_+ is infinite; and by construction the image in F/R_+ of the subgroup X of G is periodic, and, being dense, infinite. Thus we may take for K the kernel of the map from G to F/R_+, and Theorem A is proved.

3 Proof of Theorem B

If G is a discrete group which is isomorphic to the quotient group of the free (discrete) group on $\{x_1, \ldots, x_n\}$ modulo the normal subgroup generated by $\{u_1, \ldots, u_r\}$, then the quotient group of the free pro-p group on $\{x_1, \ldots, x_n\}$

modulo the closed normal subgroup generated by $\{u_1, \ldots, u_r\}$ satisfies the universal property of the pro-p completion $\widehat{G}$ of G, and so is isomorphic to $\widehat{G}$. This observation, made by Lubotsky [8], is the key to Theorem B.

Let G be as in the statement of Theorem B: thus G has a presentation with n generators and r relations; and d denotes $d(G/G')$, where G' is the commutator subgroup of G. Because G/G' is finitely generated, there is a prime p such that $d(G/G') = d(G/G'G^p)$. Let $\widehat{G}$ be the pro-p completion of G. We have $d(\widehat{G}) = d(G/G'G^p)$, for example by Lemma 2.2 of Lubotsky [8]. Applying Theorem A (with an extra summand $-\frac{1}{4}$ inserted in (1) to include presentations of $\mathbb{Z}_p$ with one more generator than relations), either we have

$$r \geqq n + \tfrac{1}{4}(d^2 - 1) - d,$$

or there is a closed normal subgroup $\overline{K}$ of $\widehat{G}$ such that the image of G in $\widehat{G}/\overline{K}$ is an infinite torsion group, and since $\widehat{G}/\overline{K}$ is a pro-p group this image is clearly a residually finite p-torsion group. This concludes the proof of Theorem B.

4 Subgroups of finite index

We need the following lemma to relate presentations of a group to presentations of subgroups of finite index.

Lemma 2. *Let G be a group (pro-p or discrete) having a presentation with n generators and r relations, and let H be a subgroup of finite index h. Then H has a presentation with $hn - (h-1)$ generators and hr relations.*

Proof. The lemma is essentially well known but we know of no reference which is entirely satisfactory. Let $G \cong F/R$, where F is free of rank n and R is generated as normal subgroup by $\{u_1, \ldots, u_r\}$, and let H correspond to E/R. Then $|F : E| = h$, and E is free of rank $hn - (h-1)$ (see Koch [6, Beispiel 6.3 on p. 48] for the pro-p case and Lyndon and Schupp [9, Proposition 3.9] for the discrete case). Let $\{t_1, \ldots, t_h\}$ be a left transversal to E in F, so that $F = \bigcup_{i=1}^{h} t_i E$. Clearly R is generated as a normal subgroup of E by the set $\{t_j^{-1} u_i t_j \,;\, 1 \leqq i \leqq r,\ 1 \leqq j \leqq h\}$.

Corollary A(i) and Corollary B are now simple consequences of Theorem A and Theorem B. Assume that G satisfies the hypotheses of Corollary A and let G have a presentation with n generators and r relations; clearly we may also assume that no open subgroup is isomorphic to $\mathbb{Z}_p$. Let H be an open subgroup with index h; thus H has a presentation with $hn - (h-1)$ generators and hr relations. Applying Theorem A for the subgroup H we have

$$hr \geqq hn - (h-1) + \tfrac{1}{4} d(H)^2 - d(H).$$

Therefore

$$(d(H) - 2)^2 \leqq 4(h(r-n) + h)$$

and

$$d(H) \leqq 2h^{\frac{1}{2}}(r - n + 1)^{\frac{1}{2}} + 2 \leqq (2 + 2(r-n+1)^{\frac{1}{2}})h^{\frac{1}{2}}.$$

This concludes the proof of Corollary A(i). The proof of Corollary B is very similar and it may be omitted.

5 Kernels of maps to $\mathbb{Z}_p$

For the proof of Corollary A(ii) it will be helpful to use the properties of the Frattini subgroup of a pro-p group and to show that one particular group cannot be an image of a 'small' finitely presented pro-p group. The Frattini subgroup $\Phi(G)$ of a profinite group G is the intersection of all maximal open subgroups, and it has the property that if H is a closed subgroup satisfying $G = H\Phi(G)$ then $H = G$ (see Gruenberg [4, Lemma 3]); thus $d(G) = d(G/\Phi(G))$. If G is a pro-p group then $\Phi(G)$ is the closed subgroup generated by all commutators and pth powers of elements of G (see [4]).

Let T be a infinite procyclic pro-p group and let A be the completed group algebra of T over $\mathbb{F}_p$ (see e.g. Koch [6, Chapter 7]). Then A is isomorphic to the ring of formal power series in one indeterminate over $\mathbb{F}_p$ (see [6, Satz 7.16]) and in particular A is a principal ideal domain all of whose proper quotients are finite. Let G be the split extension of a free A-module M of rank 1 by T. Thus G is a metabelian pro-p group generated by two elements. Consider the subgroup $G_m = MT^{p^m}$ for each integer m. As a module for the completed group algebra of T^{p^m}, M is free of rank p^m. Thus $G_m/\Phi(G_m)$ is an extension of an $\mathbb{F}_p$-module of rank p^m by a cyclic group of order p, and we have

$$d(G_m) = d(G_m/\Phi(G_m)) = p^m + 1.$$

On the other hand G_m has index p^m in G. It follows by Corollary A(i) that G cannot be an image of a finitely presented pro-p group satisfying the hypotheses of Corollary A.

We are now ready to prove Corollary A(ii). Let G satisfy the hypotheses, and suppose that $G/N \cong \mathbb{Z}_p$ and that N is not finitely generated, so that $N/\Phi(N)$ is infinite. However N is finitely generated as a normal subgroup of G, since $\mathbb{Z}_p$ is a finitely presented pro-p group. Therefore $N/\Phi(N)$ is a finitely generated module for the completed group ring of T, where T is a complement to N in G. Using the properties of A mentioned above, we may write $N/\Phi(N) \cong U \oplus V$, where U, V are A-submodules and V is free of rank 1. Therefore if $U = L/\Phi(N)$ we find that G/L is an extension of a free A-module by T, and we have a contradiction from the paragraph above. This concludes the proof of Corollary A(ii).

References

1. Bieri, R., Strebel, R.: Valuations and finitely presented metabelian groups. Proc. Lond. Math. Soc. **41**, 439–464 (1980)
2. Golod, E.S.: On nil-algebras and finitely approximable p-groups. Izv. Akad. Nauk SSSR, Ser. Mat. **28**, 273–276 (1964)
3. Golod, E.S., Šafarevič, I.R.: On the class field tower. Izv. Akad. Nauk SSSR, Ser. Mat. **28**, 261–272 (1964)
4. Gruenberg, K.W.: Projective profinite groups. J. Lond. Math. Soc. **42**, 155–165 (1967)

5. Koch, H.: Zum Satz von Golod-Šafarevič. Math. Nachr. **42**, 321–333 (1969)
6. Koch, H.: Galoissche Theorie der p-Erweiterungen. Berlin: VEB Deutscher Verlag der Wissenschaften 1970
7. Lazard, M.: Groupes analytiques p-adiques. Publ. Math., Inst. Hautes Étud. **26**, 389–603 (1965)
8. Lubotsky, A.: Group presentation, p-adic analytic groups and lattices in $SL_n(\mathbf{C})$. Ann. Math. (2) **118**, 115–130 (1983)
9. Lyndon, R.C., Schupp, P.E.: Combinatorial group theory. (Ergeb. Math. Grenzgeb., Bd. 89) Berlin Heidelberg New York: Springer 1977
10. Neumann, H.: Varieties of groups. (Ergeb. Math. Grenzgeb., Bd. 37) Berlin Heidelberg New York: Springer 1967
11. Strebel, R.: Finitely presented soluble groups. In: Gruenberg, K.W., Roseblade, J.E. (eds.) Group Theory: essays for Philip Hall, pp. 257–314. London Orlando: Academic Press 1984
12. Vinberg, E.B.: On the theorem concerning the infinite dimension of an associative algebra. Izv. Akad. Nauk SSSR, Ser. Mat. **29**, 209–214 (1965)
13. Wehrfritz, B.A.F.: Infinite linear groups. (Ergeb. Math. Grenzgeb., Bd 76) Berlin Heidelberg New York: Springer 1973
14. Zel'manov, E.I.: Personal communication

Inv. Math. N 105, 185–216 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

R-torsion and zeta functions for locally symmetric manifolds

Henri Moscovici* and Robert J. Stanton*,**

Department of Mathematics, Ohio State University, Columbus, Ohio 43210–1174, USA

Oblatum 27–XII–1989 & 18–XII–1990

Introduction

In 1968 Milnor [M] pointed out the remarkable similarity between the algebraic formalism of R-torsion in topology and zeta functions à la Weil in dynamical systems. From a different and purely analytical perspective, Ray and Singer [R-S2] exhibited an identity expressing the holomorphic analogue of R-torsion for surfaces (of genus > 1) in terms of classical Selberg zeta functions. They also conjectured the equality of R-torsion and analytic torsion, subsequently proved independently by Cheeger [C] and Müller [Mü]. These developments pointed towards an intimate relationship between torsion invariants and dynamical zeta functions. This theme has been thoroughly investigated by Fried who devised for any smooth flow and any flat bundle over the underlying manifold a certain formal zeta function counting the periodic orbits of a flow with appropriate multiplicities. He was able to show for a variety of flows ([F1, F2]) that the zeta function associated to any *acyclic* flat bundle is actually meromorphic on a neighborhood of $[0, \infty)$, regular at 0, and that its value at 0 coincides with the R-torsion with coefficients in the given flat bundle and thus is a topological invariant. Because of the analogy with the Lefschetz fixed point formula, Fried coined the term "flow with the Lefschetz property" in reference to such a flow. In particular, he proved that the geodesic flow of a closed manifold of constant negative curvature has the Lefschetz property [F3] and conjectured that this remains true for any closed locally homogeneous Riemannian manifold.

In this paper we shall prove Fried's conjecture in what is arguably the most interesting case, namely for closed locally symmetric manifolds of nonpositive sectional curvature. The main novelty of our approach consists in the introduction of certain "supersymmetrized" zeta functions of Selberg type for higher rank locally symmetric spaces of non-compact type, which admit meromorphic

* Supported in part by NSF Grant DMS-8803072

** Part of this work was done while visiting the University of Chicago and Université Louis Pasteur

continuation and satisfy a functional equation. They play the same key role in the proof as the usual Selberg zeta functions do in the rank 1 case.

We now proceed to state our result and outline the main steps of the proof. First, let us recall Fried's general definition of the formal zeta function associated to a flow [F2]. Let M be a compact, connected, smooth manifold, equipped with a smooth flow Φ which admits an invariant complementary distribution. The set of periods of closed orbits of Φ is discrete and the periodic set breaks up into connected components. Each such component C is compact, hence has a Fuller index, $\text{ind}_F(C) \in Q$. All orbits in C have the same period, ℓ_c, and are freely homotopic, thus determining a conjugacy class $[\gamma_c]$ in $\pi_1(M)$. Given a finite-dimensional representation φ of $\pi_1(M)$, define a formal Dirichlet-type function by the formula

$$Z_\varphi(s) = \exp - \sum_C \operatorname{Tr} \varphi(\gamma_c) \operatorname{ind}_F(C) e^{-s\ell_c}.$$

If M is the unit tangent bundle SX to a compact locally symmetric space X and Φ is the geodesic flow, then the connected components of the periodic set are parametrized by the non-trivial conjugacy classes $[\gamma]$ in $\Gamma = \pi_1(X)$. Each connected component X_γ is itself a compact locally symmetric manifold of non-positive sectional curvature. Φ restricts to a periodic flow on X_γ and the quotient $\widehat{X}_\gamma = X_\gamma/\Phi$ is a smooth orbifold whose Fuller index equals $\chi(\widehat{X}_\gamma)/\mu_\gamma$, where $\chi(\widehat{X}_\gamma) \in Q$ is the orbifold Euler characteristic of $\widehat{X}_\gamma$ and μ_γ is the multiplicity of a generic orbit of $\Phi|X_\gamma$. Given an acyclic unitary representation φ of Γ, the above formal zeta function takes the form

$$Z_\varphi(s) = \exp - \sum_{[\gamma]\neq 1} \operatorname{Tr} \varphi(\gamma) \chi(\widehat{X}_\gamma) \frac{e^{-s\ell_\gamma}}{\mu_\gamma}.$$

Our main result can be stated as follows:

Theorem. *$Z_\varphi(s)$ is analytic for $\operatorname{Re} s \gg 0$ and can be meromorphically continued to the entire complex plane with $s = 0$ a regular point. Moreover,*

$$Z_\varphi(0) = \tau_\varphi(SX),$$

where $\tau_\varphi(SX) = \tau_\varphi(X)^2$ is the R-torsion of the acyclic flat bundle defined by the unitary representation φ of $\pi_1(X) \cong \pi_1(SX)$.

A brief synopsis of the proof is now in order. First, one uses the Cheeger-Müller theorem to replace the R-torsion by analytic torsion. Next, one passes from zeta-function determinants to heat kernels, i.e., in the formalism we expound in Sect. 1, to theta-function determinants. The trace of the heat flow is then decomposed, in Sec. 2, according to free homotopy classes, by means of the well-known recipe of Selberg [Se]. A certain amount of work is required at this point to recognize the coefficients of this expansion as Euler characteristics of orbifolds (cf. Sec. 3). This is one of the new features distinguishing the higher rank case from the rank one and the flat case, in which the periodic geodesics are either all isolated or only the isolated ones contribute to the expansion. The main difficulty however, comes from the lack of an adequate theory of Selberg-

type zeta functions for locally symmetric spaces of higher rank. We finesse this impediment by constructing, in Sec. 4, certain "super" Selberg zeta functions, $Z_\varphi^\ell(s)$, $0 \leqq \ell \leqq 2m < \dim X$, as alternating products of formal Selberg-like functions, which reduce to genuine Selberg zeta functions only in the rank 1 case. Each function $Z_\varphi^\ell(s)$ is related to a theta-regularized determinant hence is meromorphic on $\mathbb{C}$ and moreover satisfies a functional equation. Finally, Z_φ itself can be expressed as an alternating product of these,

$$Z_\varphi(s) = \prod_{\ell=0}^{2m} Z_\varphi^\ell(s-m+\ell)^{(-1)^\ell}.$$

Not surprisingly, the functional equations play a crucial role in identifying the special value $Z_\varphi(0)$ with the R-torsion.

1 Operator determinants

In this preliminary section we introduce a regularization procedure for determinants of Laplacians which is especially well suited for our purposes. For the sake of clarity, it will be convenient to axiomatize the spectral properties required by this procedure.

So, throughout this section, A will denote a positive, possibly unbounded, self-adjoint operator on a separable Hilbert space, which is invertible with p-summable inverse, i.e.

$$\operatorname{Tr} A^{-p} < \infty \quad \text{for some} \quad p \in [1, \infty). \tag{1.1}$$

This implies that the associated theta-function

$$\theta_A(t) = \operatorname{Tr} e^{-tA}, \quad t > 0$$

is defined and moreover satisfies the estimate:

$$\theta_A(t) = O(t^{-p}) \quad \text{as} \quad t \to 0^+.$$

We impose the stronger requirement that it admits an asymptotic expansion of the form

$$\theta_A(t) \sim \sum_{k=0}^{\infty} c_{\alpha_k} t^{\alpha_k} \quad \text{as} \quad t \to 0^+, \tag{1.2}$$

where $\alpha_k \in \mathbb{R}$, $-p = \alpha_0 < \alpha_1 < \cdots < \alpha_k < \cdots$ and $\alpha_k \to \infty$.

On the other hand, one automatically has the "large time" estimate

$$\theta_A(t) = O(e^{-t\lambda_1}) \quad \text{as} \quad t \to \infty. \tag{1.3}$$

where λ_1 is the first eigenvalue of A.

From (1.2) and (1.3) it follows that θ_A is a pseudofunction [S, p. 38] and hence the *finite part* (in the sense of Hadamard) of its Laplace transform

$$Pf \int_0^\infty e^{-t\lambda} \theta_A(t)\, dt \tag{1.4}$$

is defined and meromorphic on $\mathbb{C}$. Indeed, (1.4) can be written as a sum

$$Pf \int_0^1 e^{-t\lambda}\theta_A(t)\,dt + \mathrm{Tr}\,[(A+\lambda)^{-1}e^{-(A+\lambda)}]$$

with the first term an entire function and the second meromorphic, with simple poles at $\lambda = -\lambda_k$, $k \geqq 1$. Here

$$0 < \lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_k \leqq \cdots, \quad \lambda_k \to \infty$$

are the eigenvalues of A, repeated according to their multiplicities. Note also that the residues at these poles are integers.

Since the expression (1.4) can be viewed as a regularization of the divergent sum

$$\sum_{k=0}^{\infty} \int_0^{\infty} e^{-t(\lambda+\lambda_k)}\,dt = \sum_{k=0}^{\infty} (\lambda+\lambda_k)^{-1},$$

we shall refer to it as the *θ-regularized trace* of $(A+\lambda)^{-1}$, whence the notation

$$\mathrm{Tr}_\theta (A+\lambda)^{-1} = Pf \int_0^{\infty} e^{-t\lambda}\theta_A(t)\,dt. \tag{1.4'}$$

Furthermore, we define the *θ-regularized determinant* of $I + \lambda A^{-1}$ as the exponential of the meromorphic function $\log\det_\theta(I+\lambda A^{-1})$, uniquely characterized by the requirements:

$$\begin{cases} \frac{d}{d\lambda}\log\det_\theta(I+\lambda A^{-1}) = \mathrm{Tr}_\theta(A+\lambda)^{-1}; \\ \log\det_\theta(I+\lambda A^{-1})|_{\lambda=0} = 0. \end{cases} \tag{1.5}$$

We shall need to relate it to the more familiar *ζ-regularized determinant* $\det_\zeta$ of Ray and Singer [R-S]. Let us recall that, by definition,

$$\log\det_\zeta A = -\zeta'_A(0) \tag{1.6}$$

where ζ_A is the zeta-function of A, i.e.

$$\zeta_A(s) = \mathrm{Tr}\,A^{-s} \quad \text{for} \quad \mathrm{Re}\,s > p.$$

Because of (1.2) and (1.3), ζ_A can be meromorphically continued to $\mathbb{C}$, with the (simple) poles:

$$s = -\alpha_k, \quad \mathrm{Res}_{s=-\alpha_k}\zeta_A = \frac{1}{\Gamma(-\alpha_k)}c_{\alpha_k}, \quad \forall k \geqq 0.$$

This follows (see e.g. [V]) from the fact that

$$\zeta_A(s) = \frac{1}{\Gamma(s)}\mu_A(s)$$

where μ_A is the Mellin transform of the theta-function, i.e.

$$\mu_A(s) = \int_0^\infty t^{s-1}\theta_A(t)\,dt, \quad \operatorname{Re} s > p.$$

The function μ_A itself is meromorphic on $\mathbb{C}$, with simple poles at $s = -\alpha_k$ and residues $\operatorname{Res}_{s=-\alpha_k}\mu_A = c_{\alpha_k}$, $\forall k \geqq 0$. Since $\frac{1}{\Gamma(s)}$ vanishes at zero and all negative integers, these are regular values for ζ_A; more precisely, one has

$$\zeta_A(-k) = (-1)^k k!\, c_{-k}, \quad \forall k \geqq 0, \tag{1.7}$$

with the usual convention: $c_\alpha = 0$ if $\alpha \notin \{\alpha_k\}_{k\geqq 0}$.

Similar statements apply to the one-parameter family of functions, defined for $\operatorname{Re}\lambda > -\lambda_1$ as follows:

$$\mu_A(s,\lambda) = \int_0^\infty t^{s-1}e^{-t\lambda}\theta_A(t)\,dt, \quad \operatorname{Re} s > p$$

and

$$\zeta_A(s,\lambda) = \frac{1}{\Gamma(s)}\mu_A(s,\lambda).$$

In particular,

$$\zeta_A(0,\lambda) = \sum_{k=0}^{[p]} \frac{(-1)^k}{k!} c_{-k}\lambda^k \tag{1.8}$$

and

$$\log\det{}_\zeta(A+\lambda) = -\zeta'_A(0,\lambda). \tag{1.9}$$

The theta and zeta regularizations are related by the following simple formula.

Proposition 1.1. *If A is a positive definite operator satisfying (1.1) and (1.2), then*

$$\det{}_\theta(I+\lambda A^{-1}) = \frac{\det_\zeta(A+\lambda)}{\det_\zeta A}\exp C\sum_{k=1}^{[p]} c_{-k}\frac{(-\lambda)^k}{k!}, \tag{1.10}$$

where C is Euler's constant.

Proof. We first recall the *finite part* prescription for a meromorphic function F, at a simple pole α:

$$PfF(\alpha) = \lim_{\varepsilon\to 0}\left[F(\alpha+\varepsilon) - \frac{1}{\varepsilon}\operatorname{Res}_{s=\alpha}F\right].$$

With this notation, it follows from [S, p. 41] and the preceding discussion that

$$Pf\int_0^\infty t^{\alpha-1}e^{-t\lambda}\theta_A(t)\,dt = Pf\,\mu_A(\alpha,\lambda), \quad \operatorname{Re}\lambda > -\lambda_1$$

at any pole $s = \alpha$. In particular, (1.5) becomes

$$\left\{ \begin{array}{l} \frac{d}{d\lambda} \log \det_\theta (I + \lambda A^{-1}) = Pf\, \mu_A(1, \lambda), \\ \log \det_\theta (I + \lambda A^{-1})|_{\lambda=0} = 0, \end{array} \right. \tag{1.5'}$$

i.e.

$$\log \det_\theta (I + \lambda A^{-1}) = \int_0^\lambda Pf\, \mu_A(1, v)\, dv\,. \tag{1.5''}$$

Using the relationship

$$\frac{d}{d\lambda} \mu_A(s, \lambda) = -\mu_A(s+1, \lambda)$$

formula (1.5″) gives:

$$\log \det_\theta (I + \lambda A^{-1}) = Pf\, \mu_A(0) - Pf\, \mu_A(0, \lambda) \tag{1.5'''}$$

By [V, (4.10)], (1.8) and (1.9),

$$\begin{aligned} Pf\, \mu_A(0, \lambda) &= \zeta'_A(0, \lambda) - C\zeta_A(0, \lambda) \\ &= -\log \det_\zeta (A + \lambda) - C \sum_{k=0}^{[p]} c_{-k} \frac{(-\lambda)^k}{k!}\,. \end{aligned}$$

Together with (1.5‴), this gives the formula (1.10). □

As remarked by Voros, $\det_\zeta (A + \lambda)$ has the following asymptotic expansion at ∞ [V, 5.1)]:

$$\begin{aligned} -\log \det_\zeta (A + \lambda) \underset{\lambda \to \infty}{\sim} & \sum_{\alpha_n \neq 0, -1, -2, \ldots} c_{\alpha_n} \Gamma(\alpha_n) \lambda^{-\alpha_n} \\ & - \sum_{k=0}^{[p]} c_{-k} \left(\log \lambda - \sum_{j=1}^{k} \frac{1}{j} \right) \frac{(-\lambda)^k}{k!}\,. \end{aligned} \tag{1.11}$$

Combined with (1.10), this implies a Stirling-type formula for the θ-regularized determinant:

$$\begin{aligned} \log \det_\theta (I + \lambda A^{-1}) \underset{\lambda \to \infty}{\sim} & - \sum_{\alpha_n \neq 0, -1, -2, \ldots} c_{\alpha_n} \Gamma(\alpha_n) \lambda^{-\alpha_n} \\ & + \sum_{k=1}^{[p]} c_{-k} \left(C + \log \lambda - \sum_{j=1}^{k} \frac{1}{j} \right) \frac{(-\lambda)^k}{k!} \\ & + c_0 \log \lambda - \log \det_\zeta A\,. \end{aligned} \tag{1.12}$$

Notice that if none of the α_k are non-positive integers then (1.10) becomes

$$\det_\zeta (A + \lambda) = \det_\zeta A \det_\theta (I + \lambda A^{-1})$$

and (1.12) gives

$$\log \det{}_\theta(I + \lambda A^{-1}) \underset{\lambda \to \infty}{\sim} -\sum_{\alpha_n} c_{\alpha_n} \Gamma(\alpha_n) \lambda^{-\alpha_n} - \log \det{}_\zeta A. \tag{1.12$'$}$$

For such operators A there is a common modification of the regularized determinants that can be normalized to be 1 at $+\infty$ given by

$$e^{\sum_{\alpha_k<0} c_{\alpha_k}\Gamma(\alpha_k)\lambda^{-\alpha_k}} \det{}_\zeta(A+\lambda) = \det{}_\zeta A \, e^{\sum_{\alpha_k<0} c_{\alpha_k}\Gamma(\alpha_k)\lambda^{-\alpha_k}} \det{}_\theta(I + \lambda A^{-1}), \tag{1.13}$$

holomorphic for $\operatorname{Re}\lambda > 0$. If the negative α_k have some uniformity one can obtain a meromorphic function on $\mathbb{C}$. For example if all the negative α_k are half-integral then

$$\det{}_\zeta A \, e^{\sum_{\alpha_k<0} c_{\alpha_k}\Gamma(\alpha_k)s^{-2\alpha_k}} \det{}_\theta(I + s^2 A^{-1})$$

is a meromorphic function of s.

A similar statement applies to the one parameter family of functions obtained by replacing A with $A + \lambda$ where $\operatorname{Re}\lambda > -\lambda_1$:

Corollary 1.2. *Let A be a positive definite operator satisfying (1.1) and (1.2) and suppose for all $\alpha_k < 0, 2\alpha_k$ is an integer. Then*

$$\det{}_\zeta A \, e^{\sum_{\alpha_k<0} c_{\alpha_k}\Gamma(\alpha_k)s^{-2\alpha_k}} \det{}_\theta(I + (s^2 - \lambda)A^{-1}) \tag{1.14}$$

is a meromorphic function of s, and

$$\lim_{s\to+\infty} \det{}_\zeta A \, e^{\sum_{\alpha_k<0} c_{\alpha_k}\Gamma(\alpha_k)s^{-2\alpha_k}} \det{}_\theta(I + (s^2 - \lambda)A^{-1}) = 1. \tag{1.15}$$

Proof. If suffices to prove (1.15). But this follows directly from (1.12)$'$ and $(s^2 - \lambda)^{-\alpha_k} = s^{-2\alpha_k}(1 + O(s^{-2}))$ for $\alpha_k < 0$. □

Corollary 1.3. *If $2\alpha_k$ is an odd integer $\alpha_k < 0$, denote for any fixed λ the function in (1.14) by $D(s, A)$. Then $D(s, A)$ has the following properties:*

(i) *$D(\cdot, A)$ is meromorphic in $\mathbb{C}$;*

(ii) $\lim_{s\to+\infty} D(s, A) = 1$;

(iii) $D(-s, A) = e^{2\sum_{\alpha_k<0} c_{\alpha_k}\Gamma(\alpha_k)s^{-2\alpha_k}} D(s, A)$;

(iv) $D(-s, A)D(s, A) = [\det{}_\zeta A \det{}_\theta(I + (s^2 - \lambda)A^{-1})]^2$.

Given a finite collection of operators $A_1, \ldots, A_m$ each satisfying (1.1) and (1.2), define the *torsion zeta function* by the formula

$$\log \tau_\zeta(\lambda) = \frac{1}{2}\sum_{q=0}^{m} (-1)^{q+1} q \log \det{}_\zeta(A_q + \lambda).$$

The analytic torsion of the family is by definition the number $\tau_\zeta(0)$. The actual analytic torsion of a Riemannian manifold [R-S1], in which we are interested, is associated to the collection of Laplacians on forms.

2 The torsion theta function

This section is devoted to the "torsion θ-function" associated to the Laplace operators on a compact locally symmetric manifold. The main result is a formula expressing it as an explicit sum over conjugacy classes in the fundamental group of the manifold. We begin with a description of the operators involved and the Schwartz kernels for the corresponding heat operators. Then we express the θ-function as a series of orbital integrals. The evaluation of these orbital integrals is accomplished by computing their group theoretical Fourier transform and then using the inversion formula. However, it is made explicit through some fortuitous cancellations that lead to a simple closed form. The identity element's contribution to θ also takes an unexpectedly simple form which will be used in a crucial way in a later section.

Let $\widetilde{X}$ be a globally symmetric space of noncompact type and dimension $2n+1$, and let G denote the identity component of the group of orientation preserving isometries of $\widetilde{X}$. Then G is a connected semisimple Lie group, and if K is a fixed maximal compact subgroup, then $\widetilde{X}$ is naturally isometric to G/K. Let $\mathfrak{p}$ denote the tangent space of $\widetilde{X}$ at eK. K acts on $\mathfrak{p}$ and hence on the exterior powers $\Lambda^k\mathfrak{p}$.

Let $\Gamma \subseteq G$ be a discrete, torsion free, co-compact subgroup of G. Then $X = \Gamma \backslash \widetilde{X}$ is a compact locally symmetric manifold covered by $\widetilde{X}$ and with $\pi_1(X) = \Gamma$. We are interested in differential forms on X with coefficients in a flat bundle. So let $\varphi : \Gamma \to U(F)$ be a unitary representation of Γ on the finite dimensional complex vector space F and form the associated flat bundle $\mathbb{F}$. Form the induced representation $I_\Gamma^G\varphi$ which we denote by $(R_{\Gamma,\varphi}, L^2(\Gamma \backslash G;\varphi))$. The smooth vectors in this representation are invariant by G and known to be given by

$$C^\infty(\Gamma \backslash G;\varphi) = \{f : G \to F | f \quad \text{is smooth and} \quad f(\gamma g) = \varphi(\gamma) f(g)\}.$$

The group K then acts on $C^\infty(\Gamma \backslash G;\varphi) \underset{\mathbb{R}}{\otimes} \Lambda^k\mathfrak{p}$ and the invariant vectors $(C^\infty(\Gamma \backslash G;\varphi) \underset{\mathbb{R}}{\otimes} \Lambda^k\mathfrak{p})^K$ are identified with the space of smooth $\mathbb{F}$-valued differential k-forms on X (cf. [M-S, § 7]). In the same way $(L^2(\Gamma \backslash G;\varphi) \underset{\mathbb{R}}{\otimes} \Lambda^k\mathfrak{p})^K$ can be identified with the L^2-forms. Furthermore, the differential operators d_φ, d_φ^* are given by the formulas

$$d_\varphi = \sum_i R_{\Gamma,\varphi}(X_i) \otimes \varepsilon(X_i)$$

$$d_\varphi^* = \sum_i R_{\Gamma,\varphi}(X_i) \otimes i(X_i),$$

where $\{X_i\}$ is an orthonormal basis of $\mathfrak{p}$, and ε (resp. i) is exterior (resp. interior) multiplication of $\Lambda^*\mathfrak{p}$. We denote the Laplacian on smooth $\mathbb{F}$-valued k-forms as well as the closed extension to L^2-forms by

$$\triangle_{\varphi,k} = d_\varphi d_\varphi^* + d_\varphi^* d_\varphi .$$

Denote by $\tilde{d}, \tilde{d}^*, \tilde{\triangle}_k$ the corresponding operators for $\tilde{X}$ viewed as acting on $(C^\infty(G) \otimes \Lambda^k\mathfrak{p})^K$. The heat kernel $\tilde{h}_t^{(k)} = e^{-t\tilde{\triangle}_k}$ is in $[\mathscr{S}(G) \otimes \mathrm{End}\,(\Lambda^k\mathfrak{p})]^{K\times K}$ ([B-M]), where $\mathscr{S}(G) = \cap_{p>0}\mathscr{C}^p(G)$. Thus the Schwartz kernel for $e^{-t\triangle_{\varphi,k}}, h_{\varphi,t}^{(k)}$, is given by

$$h_{\varphi,t}^{(k)}(p,q) = \sum_\Gamma \tilde{h}_t^{(k)}(y^{-1}\gamma x) \otimes \varphi(\gamma)$$

where $p = \Gamma x$, $q = \Gamma y$ and x, y are in G. We set

$$\tilde{k}_t = \sum (-1)^k k \,\mathrm{tr}\, \tilde{h}_t^{(k)} ,$$

tr being the pointwise trace. Then following the familiar Selberg approach (for a succinct and lucid presentation see [W]) one gets the orbital integral expansion

$$\sum (-1)^k k \,\mathrm{Tr}\, e^{-t\triangle_{\varphi,k}} = \sum_{[\gamma]} \mathrm{Tr}\,\varphi(\gamma)\mathrm{vol}\,(\Gamma_\gamma \backslash G_\gamma)\mathscr{O}_{\tilde{k}_t}(\gamma) , \tag{2.1}$$

where

$$\mathscr{O}_{\tilde{k}_t}(\gamma) = \int_{G/G_\gamma} \tilde{k}_t(g\gamma g^{-1})\, d\bar{g} .$$

Since the distribution $f \mapsto \mathscr{O}_f(\gamma)$ is tempered and $\tilde{X}$ is odd dimensional, it suffices to evaluate the Fourier transform of $\mathscr{O}_{\tilde{k}_t}(\gamma)$ on principal series representations. So we now choose $\mathfrak{a} \subseteq \mathfrak{p}$ a maximal abelian subalgebra and recall the construction of these representations.

Let $\mathfrak{q}$ be a standard cuspidal parabolic subalgebra which we may assume is expressed as $\mathfrak{q} = \mathfrak{m}_\mathfrak{q} \oplus \mathfrak{a}_\mathfrak{q} \oplus \mathfrak{n}_\mathfrak{q}$ with $\mathfrak{a}_\mathfrak{q} \subseteq \mathfrak{a}$, and $\mathfrak{m}_\mathfrak{q} = \mathfrak{t}_\mathfrak{m} \oplus \mathfrak{p}_\mathfrak{m} = \mathfrak{m}_\mathfrak{q} \cap \mathfrak{t} \oplus \mathfrak{m}_\mathfrak{q} \cap \mathfrak{p}$. Let Q be the normalizer of $\mathfrak{q}$ in G; Q has a Langlands decomposition $Q = M_Q A_Q N_Q$. Let (ξ, W_ξ) be an irreducible unitary representation of M_Q and e^ν a quasi-character of A_Q. Let $\pi_{\xi,\nu} = \mathrm{Ind}_Q^G \xi \otimes e^\nu \otimes I$, acting by the left regular representation on

$$H_{\xi,\nu} = \{f : G \to W_\xi | f(gman) = e^{-(\nu+\varrho_Q)\log a}\xi(m)^{-1} f(g)\} ,$$

and with norm squared $\int_K |f(k)|^2_{W_\xi}\, dk$. Let us note that unitary induction corresponds to ν imaginary valued. In order to handle characters of discrete series representations it is more convenient to use $M_Q^+ = M_Q^0 C$ where $C = \ker \mathrm{Ad}\,|_{M\widehat{Q}}$.

For each k there is an operator $\pi_{\xi,\nu}(\tilde{h}_t^{(k)})$ on $[H_{\xi,\nu} \otimes \Lambda^k\mathfrak{p}]^K$. By abuse of notation we shall denote by $\pi_{\xi,\nu}(\tilde{k}_t)$ the expression $\sum(-1)^k k \pi_{\xi,\nu}(\tilde{h}_t^{(k)})$.

Proposition 2.1. *Let $Y \in \mathfrak{a}_\mathfrak{q}$ be a unit vector and let $\mathfrak{p}_Y$ be the orthogonal complement to Y in* $\mathfrak{p}$. *Then*

(i) $\quad \mathrm{Tr}\,\pi_{\xi,\nu}(\widetilde{k}_t) = e^{t\pi_{\xi,\nu}(\Omega)}\dim\,[W_\xi \otimes (\Lambda^{\mathrm{odd}}\,\mathfrak{p}_Y - \Lambda^{\mathrm{even}}\,\mathfrak{p}_Y)]^{K\cap M_Q^+}$,

(2.2) *where Ω is the Casimir element.*

(ii) $\quad \mathrm{Tr}\,\pi_{\xi,\nu}(\widetilde{k}_t) = 0$ *if* $\dim\,\mathfrak{a}_\mathfrak{g} \geqq 2$.

Proof. First were call Kuga's result ([B-W, p. 49]):

$$\pi_{\xi,\nu}(\widetilde{\Delta}_k) = -\pi_{\xi,\nu}(\Omega)$$

independently of k. Hence

$$\begin{aligned}\mathrm{Tr}\,\pi_{\xi,\nu}(\widetilde{k}_t) &= \sum(-1)^k k\,\mathrm{Tr}\,\pi_{\xi,\nu}(\widetilde{h}_t^{(k)}) \\ &= e^{t\pi_{\xi,\nu}(\Omega)}\sum(-1)^k k\,\dim\,[H_{\xi,\nu}\otimes\Lambda^k\mathfrak{p}]^K \\ &= e^{t\pi_{\xi,\nu}(\Omega)}\sum(-1)^k k\,\dim\,[W_\xi\otimes\Lambda^k\mathfrak{p}]^{K\cap M_Q^+},\end{aligned}$$

here we have used Frobenius reciprocity. Now $\mathfrak{p} = \mathbb{R}Y \oplus \mathfrak{p}_Y$ as $K\cap M_Q^+$ module, so in the Grothendieck ring of $K\cap M_Q^+$ we have

$$\begin{aligned}\sum_{k=0}^{2n+1}(-1)^k k\Lambda^k\mathfrak{p} &= \sum_{k=1}^{2n+1}(-1)^k k[\Lambda^k\mathfrak{p}_Y \oplus \Lambda^{k-1}\mathfrak{p}_Y] \\ &= \sum_{k=1}^{2n+1}(-1)^k k\Lambda^k\mathfrak{p}_Y + \sum_{k=0}^{2n}(-1)^{k+1}(k+1)\Lambda^k\mathfrak{p}_Y \\ &= \sum_{\ell=0}^{2n}(-1)^{\ell+1}\Lambda^\ell\mathfrak{p}_Y\,.\end{aligned}$$

Tensoring with W_ξ and taking $K\cap M_Q^+$ invariants one obtains (*i*).

For (ii) suppose there is a nonzero $H \in \mathfrak{a}_\mathfrak{q}\cap\mathfrak{p}_Y$. Since M_Q^+ centralizes H, $\varepsilon(H) + i(H)$ is a $K\cap M_Q^+$ intertwining operator between $\Lambda^{\mathrm{ev}}\mathfrak{p}_Y$ and $\Lambda^{\mathrm{odd}}\mathfrak{p}_Y$, and non-trivial since H is. Hence, $\Lambda^{\mathrm{ev}}\mathfrak{p}_Y$ and $\Lambda^{\mathrm{odd}}\mathfrak{p}_Y$ are equivalent $K\cap M_Q^+$ modules and (ii) follows. □

Corollary 2.2. *If G has no simple factor locally isomorphic to* $\mathrm{SO}_e(p,q)$, *pq* odd *or* $\mathrm{SL}(3,\mathbb{R})$ *then*

$$\sum(-1)^k k\,\mathrm{Tr}\,e^{-t\,\Delta_{\varphi,k}} = 0\,.$$

Proof. Proposition 2.1 implies that $\widetilde{k}_t$ is a pseudocusp form (see [M-S, Proof of Lemma 4.4]). Therefore $\mathcal{O}_{\widetilde{k}_t}(\gamma) = 0$ for regular semisimple γ unless G has a fundamental split rank 1 Cartan subgroup, hence as stated. As Γ is cocompact all elements are semisimple and the orbital integrals for singular elements are obtained from those of regular elements in a well known way (see Proof of Corollary 4.8 in [M-S]). The result then follows from (2.1). □

We now fix B a fundamental Cartan subgroup with $B = A_I A_{\mathbb{R}}$ and we may suppose $\dim A_{\mathbb{R}} = 1$. Let $M_Q A_{\mathbb{R}} N_Q$ be a cuspidal parabolic subgroup associated to B. Since B is fundamental, A_I is connected and $M_Q^+ = M_Q^0$, but of course M_Q is not necessarily connected. Under these circumstances certain useful simplifications arise in the Fourier transform of $\tilde{k}_t$.

Lemma 2.3.

$$\Lambda^{\mathrm{ev}} \mathfrak{p}_Y - \Lambda^{\mathrm{odd}} \mathfrak{p}_Y \cong (\Lambda^{\mathrm{ev}} \mathfrak{p}_{\mathfrak{m}} - \Lambda^{\mathrm{odd}} \mathfrak{p}_{\mathfrak{m}}) \otimes \sum (-1)^{\ell} \Lambda^{\ell} \mathfrak{n}_Q$$

as $K \cap M_Q^0$ modules.

Proof. As $K \cap M_Q^0$ module $\mathfrak{p}_Y \cong \mathfrak{p}_{\mathfrak{m}} \oplus \mathfrak{n}_Q$. The claim follows from elementary linear algebra. □

Lemma 2.4. *Let V_μ be a finite dimensional M_Q^0 module with infinitesimal character μ. If (ξ, W_ξ) is a unitary irreducible representation of M_Q^0 with infinitesimal character Λ_ξ and*

$$\dim [W_\xi \otimes (\Lambda^{\mathrm{ev}} \mathfrak{p}_{\mathfrak{m}} - \Lambda^{\mathrm{odd}} \mathfrak{p}_{\mathfrak{m}}) \otimes V_\mu]^{K \cap M_Q^0} \neq 0 \quad \text{then} \tag{2.3}$$
$$\|\Lambda_\xi\|^2 = \|\mu\|^2 .$$

Proof. Consider the De Rham operator for the equirank group M_Q^0:

$$d_{M^0,\mu} + d_{M^0,\mu}^* : [W_\xi \otimes \Lambda^{\mathrm{ev}} \mathfrak{p}_{\mathfrak{m}} \otimes V_\mu]^{K \cap M_Q^0} \to [W_\xi \otimes \Lambda^{\mathrm{odd}} \mathfrak{p}_{\mathfrak{m}} \otimes V_\mu]^{K \cap M_Q^0} .$$

One knows from Kuga's formula that

$$(d_{M^0,\mu} + d_{M^0,\mu}^*)^2 = -c_\xi + c_\mu$$

where $c_\xi (\mathrm{resp.} c_\mu)$ is the respective eigenvalue of the M_Q^0 Casimir element. In particular $c_\xi = \|\Lambda_\xi\|^2 - \|\varrho_M\|^2$ and $c_\mu = \|\mu\|^2 - \|\varrho_M\|^2$. As $K \cap M_Q^0$ modules

$$W_\xi \otimes \Lambda^{\mathrm{ev}} \mathfrak{p}_{\mathfrak{m}} \otimes V_\mu - W_\xi \otimes \Lambda^{\mathrm{odd}} \mathfrak{p}_{\mathfrak{m}} \otimes V_\mu = \ker (d_{M^0,\mu} + d_{M^0,\mu}^*)|_{\mathrm{even}} - \ker (d_{M^0,\mu} + d_{M^0,\mu}^*)|_{\mathrm{odd}}$$

where the right side refers to the restriction to even and odd powers of $\mathfrak{p}_{\mathfrak{m}}$. Taking $K \cap M_Q^0$ invariants and using the assumption they are nonzero, it follows $(d_{M^0,\mu} + d_{M^0,\mu}^*)^2$ must be zero. Hence $c_\xi = c_\mu$. □

Lemma 2.5. *Assume G is simple and has a split rank 1 fundamental Cartan subgroup B. Let Q be a parabolic subgroup associated to B. Let V_{μ_ℓ} be an irreducible M_Q^0 module occuring in $\Lambda^\ell \mathfrak{n}_Q$ and with infinitesimal character μ_ℓ. Then*

$$\|\mu_\ell\|^2 - \|\varrho\|^2 = -\left(\frac{\dim_{\mathbb{R}} \mathfrak{n}_Q}{2} - \ell\right)^2 \tag{2.4}$$

Proof. Examining the two possible cases gives the shortest argument.

Consider $SO_e(p,q)$, pq odd then $M_Q^0 = SO_e(p-1,q-1)$; $\mathfrak{n}_Q \cong \mathbb{R}^{p-1} \oplus \mathbb{R}^{q-1}$ with the standard $SO_e(p-1,q-1)$ action. Clearly the result depends only on the complexified groups. Set $m = (p+q)/2$. Then D_m has $\varrho = (m-1, \ldots, 1, 0)$, while D_{n-1} has $\mu_\ell = (m-1, \ldots, 1) - (0, \ldots, 0, 1, \ldots, 1)$ with $m-1-\ell$ ones, or $\mu_{m-1} = (m-1, \ldots, \pm 1)$. Then

$$\begin{aligned}\|\mu_\ell\|^2 - \|\varrho\|^2 &= m-1-\ell - 2\frac{(m-1-\ell)(m-1-\ell+1)}{2} \\ &= -(m-1-\ell)^2 \\ &= -\left(\frac{\dim_{\mathbb{R}} \mathfrak{n}_Q}{2} - \ell\right)^2.\end{aligned}$$

Consider now $SL(3,\mathbb{R})$. Then $M_Q^0 \cong SL(2,\mathbb{R})$; $\mathfrak{n}_Q \cong \mathbb{R}^2$ with the standard action. It is enough to check the result for $\ell = 1$. For this let (α,β), $(-\alpha,\beta)$ and $(2\alpha,0)$ denote the positive roots. Notice that $\|2\alpha\|^2 = \|(\pm\alpha,\beta)\|^2$ in A_2, while $\varrho = (\alpha,\beta)$ and $\mu_1 = 2\varrho_M = 2\alpha$. Hence $\|\mu_1\|^2 - \|\varrho\|^2 = 0$ in agreement with the Lemma. □

Remark. We can prove Lemma 2.5 without case-by-case analysis by using $\mathfrak{n}$-cohomology. Here this would be an unnecessary complication but in [M-S,2] it is critical.

The next statement uses the notation of page 647 in [M-S] with one minor modification.

Lemma 2.6. *Let V be a finite dimensional representation of $K \cap M_Q^0$. Let $\sigma_\pm$ be the Spin representations of $K \cap M_Q^0$ in Spin $(\mathfrak{p}_\mathfrak{m})$. Then on A_I*

(2.5)

$$\begin{aligned}&(-1)^{d_M+1}(\chi_{\sigma_+} - \chi_{\sigma_-})^2 \chi_V \frac{\overline{{}'\triangle}_{M,c}}{{}'\triangle_{M,n}} \\ &= \sum_{\omega \in \mathscr{E}_2(M_Q^0)} \dim\,[H(\omega) \otimes (\Lambda^{\mathrm{odd}}\mathfrak{p}_\mathfrak{m} - \Lambda^{\mathrm{ev}}\mathfrak{p}_\mathfrak{m}) \otimes V]^{K\cap M_Q^0}\,\overline{\Phi}_\omega \\ &+ \sum_{\omega \in \mathscr{E}_2^s(M_Q^0)} \frac{1}{|W(\omega)|} \sum_{W(\omega)} \varepsilon(\omega) \dim\,[H^w(\omega) \otimes (\Lambda^{\mathrm{odd}}\mathfrak{p}_\mathfrak{m} - \Lambda^{\mathrm{ev}}\mathfrak{p}_\mathfrak{m}) \otimes V]^{K\cap M_Q^0}\,\overline{{}'\triangle}_M\,\overline{\Theta}'_{w\omega}.\end{aligned}$$

Proof. This is the statement of Lemma 4.2 in [M-S] with

$$W = (\Lambda^{\mathrm{odd}}\mathfrak{p}_\mathfrak{m} - \Lambda^{\mathrm{ev}}\mathfrak{p}_\mathfrak{m}) \otimes V$$

and

$$\Lambda^{\mathrm{odd}}\mathfrak{p}_\mathfrak{m} - \Lambda^{\mathrm{ev}}\mathfrak{p}_\mathfrak{m} = (-1)^{d_M+1}(\sigma_+ - \sigma_-)^2 (d_M = \dim M_Q^0/K \cap M_Q^0).\quad \square$$

We are now able to obtain a simple closed expression for orbital integrals of $\tilde{k}_t$ for regular elements. Let us recall that we have fixed $B = A_I A_\mathbb{R}$ a fundamental split rank 1 Cartan subgroup, an associated cuspidal parabolic subgroup $M_Q A_\mathbb{R} N_Q$,

and $m = \frac{1}{2}\dim N_Q$. Set $\mathfrak{m}^1 = \mathfrak{m}_\mathfrak{q} \oplus \mathfrak{a}_\mathbb{R}$ and for compatible orders on $\triangle_\mathfrak{m}$ and $\triangle_\mathfrak{g}$ define the functions

$$\triangle_+(a) = \left|\det\left(I - \mathrm{Ad}\, a^{-1}\right)|_{\mathfrak{g}/\mathfrak{m}^1}\right|^{1/2}$$

$$'\triangle_I(a) = \prod_{\alpha \in \triangle_\mathfrak{m}^+} [1 - \xi_{-\alpha}(a)],$$

and recall that Y is a basis vector for $\mathfrak{a}_\mathbb{R}$. For normalization of measures we follow Harish-Chandra ([H-C, I]). For a brief summary of this see [M-S, §4], with the minor correction that $d\bar{x}$ is normalized so that $dx = d\bar{x}\, dg_h$.

Lemma 2.7. *Let $h = h_I h_R = h_I \exp rY$ be in B'. Then*

$$\mathcal{O}_{\widetilde{k}_t}(h_I h_R) = \frac{e^{-r^2/4t}}{(4\pi t)^{1/2}} \sum_{\ell=0}^{2m} (-1)^{\ell+1} \frac{e^{-t(m-\ell)^2}}{\triangle_+(h)} \chi_{\Lambda^\ell \mathfrak{n}}(h_I). \tag{2.6}$$

Proof. Using Lemma 2.3 and (2.2) we get for the Fourier transform of $\widetilde{k}_t$

$$\mathrm{Tr}\, \pi_{\xi,\nu}(\widetilde{k}_t) = e^{t\pi_{\xi,\nu}(\Omega)} \sum_{\ell=0}^{2m} (-1)^\ell \dim\left[W_\xi \otimes (\Lambda^{\mathrm{odd}}\, \mathfrak{p}_\mathfrak{m} - \Lambda^{\mathrm{ev}}\, \mathfrak{p}_\mathfrak{m}) \otimes \Lambda^\ell \mathfrak{n}\right]^{K \cap M_Q^0}.$$

Recall that $\widetilde{k}_t$ is a pseudocusp form. We are then in the position to follow the proof of Lemma 4.3 [M-S] but not the result itself due to the presence of $e^{t(\|\Lambda_\xi\|^2 - \|\varrho\|^2)}$. In this way and referring to loc. cit. for unexplained notation we obtain

$$'F_{\widetilde{k}_t}(h_I h_R)$$

$$= \frac{e^{-r^2/4t}}{(4\pi t)^{1/2}} \sum_{\ell=0}^{2m} (-1)^\ell \sum_{\omega \in \mathscr{E}_2(M^+)} e^{t(\|\Lambda_\omega\|^2 - \|\varrho\|^2)} [W(G/B)]^{-1}$$

$$\sum_{W(G/B)} \varepsilon_I(s) \xi_{s\cdot\varrho_M - \varrho_M}(h_I)\, '\Phi_\omega(s^{-1} h_I) \dim\left[H(\omega) \otimes (\Lambda^{\mathrm{odd}}\, \mathfrak{p}_\mathfrak{m} - \Lambda^{\mathrm{ev}}(\mathfrak{p}_\mathfrak{m})) \otimes \Lambda^\ell \mathfrak{n}\right]^{K \cap M_Q^0}$$

$+$ similar term form $\mathscr{E}_2^s(M_Q^0)$.

Next using Lemma 2.4 we replace $\|\Lambda_\omega\|^2$ with $\|\mu_\ell\|^2$, and using Lemma 2.5 we replace $\|\mu_\ell\|^2 - \|\varrho\|^2$ with $-(m-\ell)^2$. This gives

$$'F_{\widetilde{k}_t}(h_I h_R)$$

$$= \frac{e^{-r^2/4t}}{(4\pi t)^{1/2}} \sum_{\ell=0}^{2m} (-1)^\ell e^{-t(m-\ell)^2} [W(G/B)]^{-1} \sum_{W(G/B)} \varepsilon_I(s) \xi_{s\cdot\varrho_M - \varrho_M}(h_I)$$

$$\sum_{\omega \in \mathscr{E}_2(M_Q^0)} {}'\Phi_\omega(s^{-1} h_I) \dim\left[H(\omega) \otimes (\Lambda^{\mathrm{odd}}\, \mathfrak{p}_\mathfrak{m} - \Lambda^{\mathrm{ev}}(\mathfrak{p}_\mathfrak{m})) \otimes \Lambda^\ell \mathfrak{n}\right]^{K \cap M_Q^0}$$

$+$ similar term form $\mathscr{E}_2^s(M_Q^0)$.

But these series are summed in Lemma 2.6 giving

$$'F_{\widetilde{k_t}}(h_I h_R) = \frac{e^{-r^2/4t}(-1)^{d_M+1}}{[W(G/B)](4\pi t)^{1/2}} \sum_{W(G/B)} \varepsilon_I(s)\xi_{s\cdot\varrho_M-\varrho_M}(h_I)$$
$$\sum_{\ell=0}^{2m}(-1)^\ell e^{-t(m-\ell)^2} \frac{'\triangle_{M,c}(s^{-1}h_I)}{\overline{'\triangle_{M,n}}(s^{-1}h_I)} (\overline{\chi}_{\sigma_+} - \overline{\chi}_{\sigma_-})^2 \overline{\chi}_{\Lambda^\ell \mathfrak{n}}(s^{-1}\cdot h_I).$$

Write $\frac{'\triangle_{M,c}}{\overline{'\triangle_{M,n}}} = \frac{'\triangle_M}{|'\triangle_{M,n}|^2}$; use $(\overline{\chi}_{\sigma_+} - \overline{\chi}_{\sigma_-})^2 = (-1)^{|\triangle^+_{M,n}|}|'\triangle_{M,n}|^2$, Lemma 27.1 from [H-C,I], to get the expression

$$'F_{\widetilde{k_t}}(h_I h_R)$$
$$= \frac{e^{-r^2/4t}}{(4\pi t)^{1/2}}(-1)^{d_M+|\triangle^+_{M,n}|+1} \sum_{\ell=0}^{2m}(-1)^\ell e^{-t(m-\ell)^2} {}'\triangle_M(h_I) \frac{1}{|W(G/B)|} \sum \overline{\chi}_{\Lambda^\ell \mathfrak{n}}(s^{-1}h_I).$$

Since B is fundamental a representative s may be taken from $K \cap M_Q$. But $\Lambda^\ell \mathfrak{n}$ is a representation of M_Q not just $K \cap M_Q^0$ and so the character is a class function. Hence we obtain

$$(2.7) \quad 'F_{\widetilde{k_t}}(h_I h_R) = \frac{e^{-r^2/4t}(-1)^{d_M+|\triangle^+_{M,n}|+1}}{(4\pi t)^{1/2}} \sum_{\ell=0}^{2m}(-1)^\ell e^{-t(m-\ell)^2} {}'\triangle_M(h_I)\overline{\chi}_{\Lambda^\ell \mathfrak{n}}(h_I).$$

But $|\triangle^+_{M,n}| = \dim \mathfrak{p}_\mathfrak{m} = d_M$, also $\mathfrak{n}$ is an $\mathbb{R}$-space so $\chi_{\Lambda^\ell \mathfrak{n}}$ is $\mathbb{R}$-valued.
The proof will be complete once we recall the connection between $'F_{\widetilde{k_t}}$ and $\mathcal{O}_{\widetilde{k_t}}$,

$$\mathcal{O}_{\widetilde{k_t}}(h_I h_R) = \frac{1}{\triangle_+(h)} \frac{1}{'\triangle_M(h_I)} {}'F_{\widetilde{k_t}}(h). \quad \square$$

The evaluation of the orbital integrals for h singular but $h \neq e$ introduces nothing different from [M-S]. Multiply (2.7) by $\xi_{\varrho\widehat{M}}$ to get $F_{\widetilde{k_t}}$, then apply the usual differential operator. Following Harish-Chandra [DSII, p. 32–37] one has

$$(2.8) \qquad \mathcal{O}_{\widetilde{k_t}}(h) = \frac{N^{-1}c_h^{-1}}{\xi_\varrho(h)\prod_{\alpha\in P_h^c}[1-\xi_{-\alpha}(h)]} \widetilde{\omega}_h F_{\widetilde{k_t}}(h),$$

where $N = [G_h : G_h^0]$ and

$$(2.9) \qquad c_h = (-1)^{q_h}(2\pi)^{q_h} 2^{\nu_h/2} \frac{v(A_I)}{v(K\cap G_h)} |W(G_h/A_I)|$$

with $q_h = \#P_h$, $\nu_h = \dim G_h/K_h - \operatorname{rank} G_h/K_h$. We write

$$\Lambda^\ell \mathfrak{n} = \sum a_{i,\ell} W_{\lambda_i}$$

with $a_{i,\ell}$ nonnegative integers and λ_i highest weights relative to $\triangle^+(\mathfrak{m}_{\mathfrak{q},\mathbb{C}}, \mathfrak{a}_{I,\mathbb{C}})$.

Lemma 2.8. *Let $h = h_l h_R$ be in B but not the identity element, then*

$$\int_{G/G_h^0} \tilde{k}_t(h^{\bar{x}})\, d\bar{x} = (c_h^{-1}) \frac{1}{\xi_\varrho(h) \prod_{\alpha \in P_h^c} [1 - \xi_\alpha(h^{-1})]} \frac{e^{-r^2/4t}}{(4\pi t)^{1/2}}$$
$$\sum_\ell (-1)^{\ell+1} e^{-t(m-\ell)^2} \sum a_{i,\ell} \sum_{W_{\mathbb{C}}(M_Q^0)} \varepsilon(w) \tilde{\omega}_h(w(\lambda_i + \varrho_M)) \xi_{w\lambda_i}(h_I).$$

From the usual asymptotics of heat kernels the dependence on t of $\tilde{k}_t(e)$, e the identity element, would be expected to be $O(t^{-\frac{2n+1}{2}})$. Instead, here a better estimate helds. (cf. [G-S] for the general case.)

Proposition 2.9.

$$\tilde{k}_t(e) = O(t^{-1/2}), \quad t \to 0^+ \tag{2.10}$$

Proof. One could, as B is fundamental, start with (2.8) to obtain $\tilde{k}_t(e)$. This leads to some intricate bookkeeping involving differentiation with H_γ, γ a complex root, that we choose to avoid. Instead, we use the inversion formula of Harish-Chandra

$$\tilde{k}_t(e) = \sum_{\{A_i\}} \int_{A_i^*} \operatorname{Tr} \pi_{\xi,\nu}(\tilde{k}_t) d\mu_i(\xi, \nu).$$

From the fact that $\tilde{k}_t$ is a pseudocusp form and Lemma 2.3 one obtains

$$\begin{aligned} \tilde{k}_t(e) &= \dim F \int_{-\infty}^{\infty} e^{-t\nu^2} \sum_{\xi \in \mathscr{E}_2(M^+)} e^{t(\|\Lambda_\xi\|^2 - \|\varrho\|^2)} \sum (-1)^\ell \\ &\quad \dim [W_\xi \otimes (\Lambda^{\text{odd}} \mathfrak{p} - \Lambda^{\text{ev}} \mathfrak{p})]^{K \cap M_Q^0} d\mu_B(\xi, \nu) \\ &= \dim F \int_{-\infty}^{\infty} e^{-t\nu^2} \sum_{\xi \in \mathscr{E}_2(M_Q^0)} e^{t(\|\Lambda_\xi\|^2 - \|\varrho\|^2)} \\ &\quad \dim [W_\xi \otimes (\Lambda^{\text{odd}} \mathfrak{p}_{\mathfrak{n}} - \Lambda^{\text{ev}} \mathfrak{p}_{\mathfrak{n}}) \otimes \sum (-1)^\ell \Lambda^\ell \mathfrak{n}]^{K \cap M_Q^0} d\mu_B(\xi, \nu). \end{aligned} \tag{2.11}$$

From [H-C; III] one knows that the Plancherel density $\mu_B(\xi, \nu) = d_\xi \prod_{\gamma \in \Delta_{\mathbb{C}+}} \langle (\Lambda_\xi, \nu), \gamma \rangle$, where d_ξ, the formal degree, is given by $c \prod_{\Delta_M^+} \langle \Lambda_\xi, \alpha \rangle$. Clearly, $\mu_B(\xi, \nu)$ is a polynomial in ν, even ν^2 since roots occur in pairs $(\alpha, \beta), (-\alpha, \beta)$, of degree $|\Delta_{\mathbb{C}}^+| = 2m$. Set $\mu_B(\xi, \nu) = d_\xi \sum_{j=0}^m \nu^{2m-2j} p_{2j}(\Lambda_\xi)$. Now M acts on $\mathfrak{n}$, so if $w \in W(M, A_I)$ w must preserve $\Delta_{\mathbb{C}}^+$. Hence p_{2j} is $W(M, A_I)$ invariant.

Let us examine (2.5). Let $\tilde{\omega}_M$ be the differential operator $\prod_{\Delta_M^+} H_\alpha$, apply it to (2.5) and evaluate at e. Since $(\chi_{\sigma_+} - \chi_{\sigma_-})' \overline{\Delta}_{M,C}$, vanishes at e to the same order

as $\widetilde{\omega}_M$ the only possible non-zero contribution from the left side is $c \cdot \chi_V(e)$. On the right side the singular series will vanish when we apply $\widetilde{\omega}_M$ (by definition of singular), while the $\mathscr{E}_2(M_Q^0)$ series will give

$$c \sum \dim [H(\xi) \otimes (\Lambda^{\text{odd}} \mathfrak{p}_{\mathfrak{m}} - \Lambda^{\text{ev}} \mathfrak{p}_{\mathfrak{m}}) \otimes V]^{K \cap M_Q^0} d_\xi .$$

Also we can obtain $\sum v^{2m-2j} p_{2j}(\Lambda_\xi)$ by applying a differential operator to 2.5. On A_I let P_{2j} be the $W(M, A_I)$ invariant homogeneous differential operator with symbol p_{2j}. Then clearly $\widetilde{\omega}_M \sum v^{2m-2j} P_{2j}$ applied to the right side of (2.5) and evaluated at e gives (up to a constant)

$$\sum_{\xi \in \mathscr{E}_2(M^+)} \dim [H(\xi) \otimes (\Lambda^{\text{odd}} \mathfrak{p}_{\mathfrak{m}} - \Lambda^{\text{ev}} \mathfrak{p}_{\mathfrak{m}}) \otimes V]^{K \cap M_Q^0} \mu_B(\xi, v) .$$

The same reasoning applies to $e^{t\|\Lambda_\xi\|^2}$; we denote the corresponding operator by $e^{t\Delta}$. Thus, we get

$$\widetilde{k}_t(e) = c e^{-t\|\varrho\|^2} \int_{-\infty}^{\infty} e^{-t v^2} \sum_{j=0}^{m} v^{2m-2j} P_{2j} \widetilde{\omega}_M e^{t\Delta} \{(\chi_{\sigma_+} - \chi_{\sigma_-})' \overline{\triangle}_{M,C} \chi_{\Sigma(-1)^\ell \Lambda^\ell \mathfrak{n}}\}_e \, dv .$$

Now

$$\begin{aligned} \chi_{\Sigma(-1)^\ell \Lambda^\ell \mathfrak{n}}(h_I) &= \det (I - \mathrm{Ad}\,|_{\mathfrak{n}}(h_I)) \\ &= \prod_{\gamma \in \triangle_{\mathbb{C}}^+} [1 - \xi_\gamma(h_I)] , \end{aligned}$$

and so $(\chi_{\sigma_+} - \chi_{\sigma_-})' \overline{\triangle}_{M,C} \chi_{\Sigma(-1)^\ell \Lambda^\ell \mathfrak{n}}$ vanishes at e to order $2m + |\triangle_M^+|$. Thus we need only consider differentiations of order $\geqq 2m + |\triangle_M^+|$. If we expand $e^{t\Delta} = \sum \frac{t^k}{k!} \triangle^k$, we get

$$\sum_{j=0}^{m} v^{2m-2j} P_{2j} \widetilde{\omega}_M e^{t\Delta} = \sum_{j=0}^{m} (t^{1/2} v)^{2(m-j)} P_{2j} \frac{\triangle^{m-j}}{(m-j)!} \widetilde{\omega}_M + O(t^m v^{2\ell}),$$

$$\text{with} \quad 2\ell < m .$$

Substituting this into the integral and letting $t \to 0^+$, one gets

$$\widetilde{k}_t(e) \sim \frac{c}{t^{1/2}} \left[\sum_{j=0}^{m} \triangle^{m-j} P_{2j} \widetilde{\omega}_M \{(\chi_{\sigma_+} - \chi_{\sigma_-})' \overline{\triangle}_{M,C} \chi_{\Sigma(-1)^\ell \Lambda^\ell \mathfrak{n}}\}_e \int_{-\infty}^{\infty} v^{2m-2j} e^{-v^2} \, dv \right] . \quad \square$$

Corollary 2.10. *One has*

$$\mathcal{O}_{\widetilde{k_t}}(e) = \sum(-1)^{\ell} \int_{-\infty}^{\infty} e^{-t(\nu^2+(m-\ell)^2)} \tag{i}$$

$$\left(\sum_{\xi \in \mathscr{E}_2(M^+)} \dim [W_\xi \otimes (\Lambda^{\text{odd}} \mathfrak{p}_{\mathfrak{m}} - \Lambda^{\text{ev}} \mathfrak{p}_{\mathfrak{m}}) \otimes \Lambda^{\ell} \mathfrak{n}] \right)^{K \cap M_Q^0} d\mu_B(\xi, \nu) .$$

Moreover

$$\sum(-1)^k k \operatorname{Tr} e^{-t\Delta_{\varphi,k}} \sim \frac{c}{t^{1/2}}, \quad t \to 0^+ . \tag{ii}$$

Proof. It suffices to take (2.11) and use Lemma 2.5 to obtain (i). The assertion (ii) is immediate from Proposition 2.9 and standard estimates for the contributions from Lemma 2.7 from other Γ conjugacy classes to (2.1). □

We summarize the results of this section with the expression of $\sum(-1)^k k \operatorname{Tr} e^{-t\Delta_{\varphi,k}}$ in group theoretical terms. As in [M-S] let $\mathscr{E}_1(\Gamma)$ denote the Γ-conjugacy classes of non-trivial elements in Γ contained in a fundamental, $\mathbb{R}$-rank one Cartan subgroup of G.

Theorem 2.11. *Let X be a compact locally symmetric manifold of odd dimension and $\mathbb{F}$ a flat vector bundle over X associated to a unitary representation (φ, F) of $\Gamma = \pi_1(X)$. Let $\Delta_{\varphi,k}$ be the Laplacian on k-forms with values in $\mathbb{F}$. Then*

(2.12)

$$\theta(t) = \sum(-1)^k k \operatorname{Tr} e^{-t\Delta_{\varphi,k}}$$

$$= \sum_{[\gamma] \in \mathscr{E}_1(\Gamma)} \operatorname{Tr} \varphi(\gamma) \frac{[G_\gamma : G_\gamma^0]^{-1} \operatorname{vol}(G_\gamma/\Gamma_\gamma)}{\xi_\varrho(\gamma) \prod_{\alpha \in P_\gamma^c} [1 - \xi_\alpha(\gamma^{-1})]} c_\gamma^{-1}$$

$$\frac{e^{-\ell_\gamma^2/4t}}{(4\pi t)^{1/2}} \sum(-1)^{\ell+1} e^{-t(m-\ell)^2} \sum a_{i,\ell} \sum_{W_{\mathbb{C}}(M_Q^0)} \varepsilon(w) \widetilde{\omega}_\gamma(w(\lambda_i + \varrho_M)) \xi_{w\lambda_i}(\gamma_I)$$

$$+ \dim F \operatorname{vol} X \sum(-1)^{\ell} \int_{-\infty}^{\infty} e^{-t(\nu^2+(m-\ell)^2)}$$

$$\left(\sum_{\xi \in \mathscr{E}_2(M_Q^0)} \dim [W_\xi \otimes (\Lambda^{\text{odd}} \mathfrak{p}_{\mathfrak{m}} - \Lambda^{\text{ev}} \mathfrak{p}_{\mathfrak{m}}) \otimes \Lambda^{\ell} \mathfrak{n}]^{K \cap M_Q^0} \right) d\mu(\xi, \nu) .$$

3 Euler characteristics

The goal of this section is to evaluate in group theoretical terms the Euler characteristics of the orbifolds arising from the periodic geodesics. This will enable us to express the Laplace transform of the torsion theta function $\theta(t)$, discussed in the preceding section, in purely dynamical terms.

We begin by reviewing briefly some basic facts about the geodesic flow of a locally symmetric manifold. A thorough discussion of this subject can be found in [D-K-V, § 5]. Then, we outline some results from [M-S, § 5] related to the disconnectedness and reductive nature of G_γ. These allow us to construct an algebraic descent to an equirank group that facilitates the evaluation of the Euler characteristics.

The set of periodic orbits of the geodesic flow Φ on SX consists of a disjoint union of connected smooth submanifolds parametrized by conjugacy classes $[\gamma] \neq [e]$ in $\Gamma = \pi_1(X)$. For each $[\gamma]$ the corresponding submanifold X_γ is diffeomorphic to $\Gamma_\gamma \backslash G_\gamma / U_\gamma$, U_γ a maximal compact subgroup of G_γ. As Γ is co-compact each γ is semisimple and so we may assume it has been put in standard position relative to the original choice of K. Then $X_\gamma \cong \Gamma_\gamma \backslash G_\gamma / K_\gamma$ and we write $\gamma = \gamma_I \gamma_{\mathbb{R}}$ with $\gamma_{\mathbb{R}} = \exp \ell_\gamma Y \in \exp \mathfrak{p}$ where $Y \in \mathfrak{p}$ is a unit vector. The restriction of the geodesic flow to each X_γ can be described, in this picture, by

$$\Phi_s(\Gamma_\gamma g K_\gamma) = \Gamma_\gamma g \exp sY K_\gamma, \quad g \in G_\gamma.$$

The orbits of $\Phi|_{X_\gamma}$ are periodic geodesics in the free homotopy class corresponding to $[\gamma]$. They all have the same length ℓ_γ and one has $\gamma_{\mathbb{R}} = \exp \ell_\gamma Y$.

For most of the ensuing discussion we shall assume that $[\gamma] \in \mathscr{E}_1(\Gamma)$ and that $\gamma = \gamma_I \exp \ell_\gamma Y \in A_I A_{\mathbb{R}} = B$.

The group G_γ is reductive, possibly disconnected. Both these features introduce technical complications. Let G_γ^0 be the component of the identity of G_γ and let C_γ be the center of G_γ^0. Notice that since B is connected, $B \subseteq G_\gamma^0$, in particular $\gamma \in C_\gamma$. Also it is easy to see that C_γ is a normal subgroup of G_γ and hence $\Gamma_\gamma \cap C_\gamma \triangleleft \Gamma_\gamma$.

Lemma 3.1. *Let* $[\gamma] \in \mathscr{E}_1(\Gamma)$. *Then,* $\Gamma_\gamma \cap C_\gamma$ *is free abelian of rank 1.*

Proof. This can be proved as in Lemma 5.3 (ii) of [M- S]. □

Lemma 3.2. *Let* $[\gamma] \in \mathscr{E}_1(\Gamma)$ *and let* γ_1 *be a generator of* $\Gamma_\gamma \cap C_\gamma$. *Then* $[\gamma_1] \in \mathscr{E}_1(\Gamma)$ *and we may take* γ_1 *in* B.

Proof. $\gamma_1 \in C_\gamma$ = center of G_γ^0 which is contained in every Cartan subgroup of G_γ^0. It follows that $\gamma_1 \in B$. As γ_1 is in $\Gamma_\gamma \subset \Gamma$ and γ_1 is in B, one has $[\gamma_1] \in \mathscr{E}_1(\Gamma)$. □

Definition 3.3. Let $[\gamma] \in \mathscr{E}_1(\Gamma)$. One may pick a generator γ^* of $\Gamma_\gamma \cap C_\gamma$ such that $\gamma = (\gamma^*)^{m_\gamma}$ with $m_\gamma \geqq 1$.

It follows from Lemma 3.2 that $\ell_\gamma = m_\gamma \ell_{\gamma^*}$.

Returning to G_γ^0, we note that $A_{\mathbb{R}}$ is the split component of G_γ^0. Also $C_\gamma = C_{\gamma\hat{I}} A_{\mathbb{R}}$ with $C_{\gamma\hat{I}}$ a compact abelian, possibly disconnected, group. Let G'_γ denote the semisimple group (with finite center) such that $G_\gamma = G'_\gamma C_\gamma^0$. Set $\Gamma'_\gamma = G'_\gamma \cap \Gamma_\gamma C_\gamma$. Then Γ'_γ is a discrete, co-compact subgroup of G'_γ and, if necessary, we may take ${}^0\Gamma'_\gamma \triangleleft \Gamma'_\gamma$ and ${}^0\Gamma'_\gamma$ torsion-free and of finite index in Γ'_γ. Set ${}^0\Gamma_\gamma = \Gamma_\gamma \cap {}^0\Gamma'_\gamma C_\gamma$. Then ${}^0\Gamma_\gamma$ is normal in Γ_γ, torsion-free and of finite index in Γ_γ.
Set ${}^0X_\gamma = {}^0\Gamma_\gamma \backslash G_\gamma / K_\gamma$ and ${}^0X'_\gamma = {}^0\Gamma'_\gamma \backslash G'_\gamma / K'_\gamma$. Then S^1 acts freely on ${}^0X_\gamma$ ([M-S, Lemma 5.4]) with quotient space ${}^0X'_\gamma$. Let $\chi({}^0X'_\gamma)$ denote the Euler characteristic of ${}^0X'_\gamma$.

Lemma 3.4.

$$\chi({}^0X'_\gamma)\ell_{\gamma^0}[\Gamma_\gamma : {}^0\Gamma_\gamma]^{-1} = c_\gamma^{-1}|W^{\mathbb{C}}(G_\gamma)| \prod_{\alpha \in P_\gamma} \langle \varrho_\gamma, \alpha \rangle [G_\gamma : G_\gamma^0]^{-1} \mathrm{vol}\,(G_\gamma/\Gamma_\gamma).$$

Proof. We shall reduce the computation of $\chi({}^0X'_\gamma)$ to an equirank group situation and then use results from [H-P] and [Sc]. As usual $MA_{\mathbb{R}}N$ is the cuspidal parabolic associated to B. Since $\gamma = \gamma_I \exp \ell_\gamma Y \in B$, γ_I is an elliptic element in M^0. Notice that for such B we have $M_{\gamma_I}^0 \equiv (M_{\gamma_I})^0 = (G'_\gamma)^0 C_{\gamma_I}^0$ and $M_{\gamma_I} = G'_\gamma C_{\gamma_I}$. Let $K_{\gamma I}^M$ (resp. $K_{\gamma I}^{M^0}$) denote the maximal compact subgroup of M_{γ_I} (resp. $M_{\gamma_I}^0$) relative to the restriction to M of the Cartan involution of G. Then $M_{\gamma_I}/K_{\gamma_I}^M \cong M_{\gamma_I}^0/K_{\gamma_I}^{M^0}$ and ${}^0\Gamma'_\gamma \backslash M_{\gamma_I}/K_{\gamma_I}^M \cong^0 X'_\gamma$ is a finite cover (of order $[{}^0\Gamma'_\gamma : {}^0\Gamma'_\gamma \cap G'^0_\gamma]$) of ${}^0\Gamma'_\gamma \cap M_{\gamma_I}^0 \backslash M_{\gamma_I}^0/K_{\gamma_I}^{M^0}$. Although we have not produced a discrete, co-compact subgroup of M^0, nevertheless the elliptic element γ_I and the discrete co-compact group ${}^0\Gamma'_\gamma \cap M_{\gamma_I}^0$ places ${}^0\Gamma'_\gamma \cap M_{\gamma_I}^0 \backslash M_{\gamma_I}^0/K_{\gamma_I}^{M^0}$ in the local setting of [H-P]. So following their argument (p. 220–225) we take the orientable cover ${}^0\overline{X}'_\gamma = {}^0\Gamma'_\gamma \backslash G'_\gamma / K'^0_\gamma$ of ${}^0X'_\gamma$. Then $\chi({}^0X'_\gamma) = \chi({}^0\overline{X}'_\gamma)/[K'_\gamma : K'^0_\gamma]$ and $\chi({}^0\overline{X}'_\gamma)$ can be obtained by orienting ${}^0\overline{X}'_\gamma$ then evaluating the Euler class on the fundamental cycle.
Let A_{γ_I} be a maximal torus in $K'^0_{\gamma_I}$ and set $Y'_\gamma = {}^0\Gamma'_\gamma \backslash G'_\gamma / A_{\gamma_I}$. Then for a choice of orientation $c(\prod_{\alpha \in P_\gamma} \alpha)$ is the Euler class of Y'_γ and

$$\chi({}^0\overline{X}'_\gamma) = \frac{c\Big(\prod\limits_{\alpha \in P_\gamma} \alpha\Big)[Y'_\gamma]}{|W(K'^0_\gamma/A_{\gamma_I})|}.$$

Schmid ([Sc, p. 38]) evaluated $c(\prod_{\alpha \in P_\gamma} \alpha)[Y'_\gamma]$ obtaining $(2\pi)^{-\#P_\gamma}(-1)^{\#P_{\gamma,n}}$ $v({}^0\Gamma'_\gamma \backslash G'_\gamma)\,|W^{\mathbb{C}}(G'_\gamma)| \prod_{\alpha \in P_\gamma} \langle \varrho_\gamma, \alpha \rangle$. The choice of Haar measure in [Sc] is different from the one here with the relationship being

$$v({}^0\Gamma'_\gamma \backslash G'_\gamma) = \mathrm{vol}\,({}^0\Gamma'_\gamma \backslash G'_\gamma) 2^{-\nu_\gamma/2} v(K'_\gamma) v(A_{\gamma_I})^{-1}.$$

After these substitutions we obtain

$$\chi({}^0X'_\gamma) = [K'_\gamma : K'^0_\gamma]^{-1}|W(K'_\gamma/A_{\gamma_I})|^{-1}$$
$$\times (-1)^{\#P_{\gamma,n}}(2\pi)^{-\#P_\gamma}2^{-\nu_\gamma/2}v(K'_\gamma)v(A_{\gamma_I})^{-1}\mathrm{vol}\,({}^0\Gamma'_\gamma \setminus G'_\gamma)$$
$$\times |W^{\mathbb{C}}(G'_\gamma)| \prod_{\alpha\in P_\gamma}\langle \varrho_\gamma, \alpha\rangle .$$

A comparison with (2.9), together with the standard facts $[K'_\gamma : K'^0_\gamma] = [G'_\gamma : G'^0_\gamma] = [G_\gamma : G^0_\gamma]$, $|W(K'^0_\gamma/A_{\gamma_I})| = |W(G'^0_\gamma/A_{\gamma_I}| = |W(G_\gamma/A_I)|$ give

$$\chi({}^0X'_\gamma) = c_\gamma^{-1}\mathrm{vol}\,({}^0\Gamma'_\gamma \setminus G'_\gamma)|W^{\mathbb{C}}(G_\gamma)| \prod_{\alpha\in P_\gamma}\langle \varrho_\gamma, \alpha\rangle .$$

Now using the S^1 fibration ${}^0X_\gamma \to {}^0X'_\gamma$, and integration along the fiber (Lemma 5.8 in [M-S]) we get

$$\mathrm{vol}\,({}^0\Gamma'_\gamma \setminus G'_\gamma) = \mathrm{vol}\,({}^0\Gamma_\gamma \setminus G_\gamma)/\ell_{\gamma^*} .$$

It remains to show that $[\Gamma_\gamma : {}^0\Gamma_\gamma] = [\Gamma'_\gamma : {}^0\Gamma'_\gamma]$. Indeed, $\Gamma'_\gamma/{}^0\Gamma'_\gamma$ acts freely on $G'_\gamma/{}^0\Gamma'_\gamma$. Similarly, $\Gamma_\gamma/{}^0\Gamma_\gamma \cong \Gamma_\gamma C_\gamma/{}^0\Gamma'_\gamma C_\gamma$ acts freely on $G_\gamma/{}^0\Gamma'_\gamma C_\gamma$ with quotient $G_\gamma/\Gamma_\gamma C_\gamma$. But $G_\gamma/{}^0\Gamma'_\gamma C_\gamma \cong G'_\gamma/{}^0\Gamma'_\gamma$ and $G_\gamma/\Gamma_\gamma C_\gamma \cong G'_\gamma/\Gamma'_\gamma$. □

As in [M-S], we let θ_γ be the closed 1-form on X_γ induced by the invariant form on G_γ/K_γ determined by the covector Y^*, where Y is as in Prop. 2.1. We then let NX_γ denote the orthogonal complement in TX_γ of the tangent bundle to the and more foliation to the left foliation by the orbits of Θ_t on X_γ (formerly denoted ${}^0TX_\gamma$ in [M-S]). The lifts to the orientable cover $\overline{X}_\gamma$ of these (and all other) constructs will be denoted by the same symbol superscribed by a bar.

Lemma 3.5. *Let $\bar{e}_N$ denote the Euler class of $N\overline{X}_\gamma$. Then*

$$[\Gamma_\gamma :^0 \Gamma_\gamma]^{-1}\chi({}^0\overline{X}'_\gamma)\ell_{\gamma^*} = \bar{e}_N([\overline{X}_\gamma] \cap \overline{\theta}_\gamma) .$$

Proof. Let $\overline{\pi} :^0 \overline{X}_\gamma \to {}^0\overline{X}'_\gamma$ be the natural projection. With the obvious notation, one has $N^0\overline{X}_\gamma = \overline{\pi}^*(T^0\overline{X}'_\gamma)$, therefore the Euler class ${}^0\bar{e}_N$ of $N^0\overline{X}_\gamma$ is the pull-back of the Euler class of ${}^0\overline{X}'_\gamma$. By Lemmas 5.8 and 5.9 in [M-S],

$${}^0\bar{e}_N([{}^0\overline{X}_\gamma] \cap^0 \overline{\theta}_\gamma) = \chi({}^0\overline{X}'_\gamma)\ell_{\gamma^*} .$$

Finally, since ${}^0\Gamma_\gamma \setminus \Gamma_\gamma$ acts freely on ${}^0\overline{X}_\gamma$, one has:

$${}^0\bar{e}_N([\overline{X}_\gamma] \cap^0 \overline{\theta}_\gamma) = [\Gamma_\gamma :^0 \Gamma_\gamma]\bar{e}_N([\overline{X}_\gamma] \cap \overline{\theta}_\gamma) . \quad \square$$

We now come to the main result of this section. Let $\hat{X}_\gamma$ (resp. $\hat{\overline{X}}_\gamma$) be the quotient space of X_γ (resp. $\overline{X}_\gamma$) modulo the (circle) action of the flow Φ (resp. $\overline{\Phi}$). It has

a natural (smooth) orbifold structure. We denote by $\chi(\widehat{X}_\gamma)$ (resp. $\chi(\widehat{\overline{X}}_\gamma)$) its Euler characteristic, *as an orbifold* (cf. e.g. [H]).

Proposition 3.6. *With λ_γ denoting the length in X_γ of a generic orbit of $\Phi|X_\gamma$, one has:*

$$\chi(\widehat{X}_\gamma) = \frac{1}{\lambda_\gamma}[G_\gamma : G_\gamma^0]^{-1}\mathrm{vol}(\Gamma_\gamma \setminus G_\gamma)c_\gamma^{-1}|W^{\mathfrak{C}}(G_\gamma)| \prod_{\alpha \in P_\gamma} \langle \varrho_\gamma, \alpha \rangle .$$

Proof. Assume, for the moment that $\widehat{X}_\gamma$ is orientable. Since NX_γ is isomorphic to the pull-back of $T\widehat{X}_\gamma$, e_N is the pull-back of the Euler class $\hat{e}$ of $\widehat{X}_\gamma$. Arguing as in the proof of the preceding lemma, but using integration along the fiber only over the generic stratum, one sees that

$$\chi(\widehat{X}_\gamma) = \frac{1}{\lambda_\gamma} e_N([X_\gamma] \cap \theta_\gamma) .$$

But

$$e_N([X_\gamma] \cap \theta_\gamma) = [G_\gamma : G_\gamma^0]^{-1} \bar{e}_N([\overline{X}_\gamma] \cap \bar{\theta}_\gamma)$$

and thus the stated formula follows from Lemmas 3.4 and 3.5
The non-orientable case can be treated in a similar manner, with minor and routine changes. □

The quotient $\ell_\gamma/\lambda_\gamma$ of the geometric length by the length of a generic orbit shall be called the *generic multiplicity* [1] and denoted μ_γ.

We conclude this section with the following comment on the case when $[\gamma] \notin \mathscr{E}_1(\Gamma)$.

Remark 3.7. Assume that the local Euclidean factor of X_γ has dimension $r \geqq 2$. Then $\chi(\widehat{X}_\gamma) = 0$.

Proof. It is an easy exercise, which we leave to the reader, to show that the dimension of the local Euclidean factor coincides with the minimal dimension of the split part of a Cartan subgroup containing γ. Now according to [E], there exists a finite cover Y_γ of X_γ which is diffeomorphic to a product of a torus T^r by a compact manifold $\overline{Y}_\gamma$. Moreover the flow Φ lifts to the standard action of S^1 on $S^1 \times T^{r-1}$. It follows that $\widehat{X}_\gamma$ is covered by a manifold diffeomorphic to $T^{r-1} \times \overline{Y}_\gamma$, hence our claim. [2] □

[1] It is the same as the algebraic multiplicity defined in [M-S, § 5].
[2] This argument also applies to the Lefschetz numbers $L(\gamma, D)$ in [M-S].

4 Zeta functions

The goal of this section is to show that the Weil-type zeta function

$$Z_\varphi(s) = \exp - \sum_{[\gamma]\neq[1]} \operatorname{Tr} \varphi(\gamma)\chi(\widehat{X}_\gamma)\frac{e^{-s\ell_\gamma}}{\mu_\gamma}$$

exists as a meromorphic function on $\mathbb{C}$ and to relate $Z_\varphi(0)$ to the analytic torsion of (X, φ). In the process, using a natural bi-grading of $\Lambda^*\mathfrak{p}$, we shall construct some operators whose θ-regularized determinants will serve as super Selberg-type zeta functions in the higher rank situation. When the rank is one, these functions reduce to Selberg zeta functions but in the higher rank case there is an additional $\mathbb{Z}_2$ grading (whence the terminology "super") which is essential to obtain a tractable function. We will then be able to express $Z_\varphi(s)$ in terms of these super Selberg zeta functions and thus obtain the expected function theoretic properties by appealing to the techniques from § 1. This requires that we assume now that φ defines an acyclic complex (there are no harmonic $\mathbb{F}$-valued forms). Let us recall that $B = A_I A_{\mathbb{R}}$ denotes a fundamental split rank one Cartan subgroup, and that $Q = M_Q A_{\mathbb{R}} N_Q$ is a corresponding parabolic subgroup. We make some simplifications in the notation for this section. We set $M = M_Q^0$ $(= M_Q^+)$, $K^M = K \cap M_Q^+$, and usually omit the subscript Q. Let $\Omega^k = [L^2(\Gamma \setminus G; \varphi) \oplus \Lambda^k\mathfrak{p}]^K$, $k \geqq 0$ and $\Omega^{-1} \equiv 0$. Set $\mathfrak{c} = \mathfrak{a} \oplus \mathfrak{p}_{\mathfrak{m}}$. It is well known $\mathfrak{p}$ is equivalent to $\mathfrak{c} \oplus \mathfrak{n}$ as K^M module, and consequently

$$\Lambda^k\mathfrak{p} \simeq \sum_{j+\ell=k} \oplus \Lambda^j\mathfrak{c} \otimes \Lambda^\ell\mathfrak{n}.$$

We fix this latter orthogonal sum and shall use this isomorphism to identify $\Lambda^j\mathfrak{c}\otimes\Lambda^\ell\mathfrak{n}$ as a subspace of $\Lambda^k\mathfrak{p}$. Then let $Q^{j,\ell} : \Lambda^k\mathfrak{p} \to \Lambda^j\mathfrak{c}\otimes\Lambda^\ell\mathfrak{n}$ be the orthogonal projection operator and ${}^tQ^{j,\ell}$ the inclusion. We set $\Omega^{j,\ell} = I \otimes Q^{j,\ell}(\Omega^{j+\ell})$; $\Omega^{j,\ell}$ is a closed subspace of $[L^2(\Gamma \setminus G; \varphi) \otimes \Lambda^j\mathfrak{c} \otimes \Lambda^\ell\mathfrak{n}]^{K^M}$.
Recall that the heat operator $\widetilde{h}_t^{(k)}$ is in $[\mathscr{S}(G) \otimes \operatorname{End}\Lambda^k\mathfrak{p}]^{K\times K}$. We set

$$\widetilde{h}_t^{(j,\ell)} = I \otimes Q^{j,\ell} \circ e^{-t\widetilde{\Delta}_{\varphi,k}} \circ I \otimes {}^tQ^{j,\ell}.$$

Clearly $\widetilde{h}_t^{(j,\ell)}$ is in $[\mathscr{S}(G) \otimes \operatorname{End}(\Lambda^j\mathfrak{c} \otimes \Lambda^\ell\mathfrak{n})]^{K^M\times K^M}$, and so we can define an operator $e^{-t\Delta_\varphi^{j,\ell}}$ on $\Omega^{j,\ell}$ by the kernel

$$h_{\varphi,t}^{(j,\ell)}(p,q) = \sum_\Gamma \widetilde{h}_t^{(j,\ell)}(x^{-1}\gamma y) \otimes \varphi(\gamma).$$

The non-zero spectrum of $\Delta_\varphi^{j,\ell}$ is contained in that of $\Delta_{\varphi,\, j+\ell}$ thus $e^{-t\Delta_\varphi^{j,\ell}}$ is a trace class operator on $\Omega^{j,\ell}$.
Recall that the dimension $X = 2n+1$ and the dimension $n = 2m$, hence dimension $\mathfrak{p}_{\mathfrak{m}} \equiv 2k = 2(n-m)$.

Lemma 4.1. *Fix ℓ, $0 \leqq \ell \leqq 2m$. As $K \cap M^+$ module one has*

$$\sum_{j=0}^{2k+1} \oplus(-1)^{j+\ell}(j+\ell)\Lambda^j \mathfrak{c} \otimes \Lambda^\ell \mathfrak{n} \simeq (-1)^\ell(\Lambda^{\text{odd}}\mathfrak{p}_m - \Lambda^{\text{ev}}\mathfrak{p}_m) \otimes \Lambda^\ell \mathfrak{n}. \tag{4.1}$$

Proof. Recall that $\dim \mathfrak{a} = 1$ and that $K \cap M$ centralizes $\mathfrak{a}$. Then $\Lambda^j \mathfrak{c} \simeq \Lambda^j \mathfrak{p}_m \oplus \Lambda^{j-1}\mathfrak{p}_m$ $(0 < j < 2k)$, while $\Lambda^0 \mathfrak{c} \simeq \Lambda^0 \mathfrak{p}_m$ and $\Lambda^{2k+1}\mathfrak{c} \simeq \Lambda^{2k}\mathfrak{p}_m$ trivially. Hence as in the proof of (2.2)

$$\begin{aligned}
&\sum_{j=0}^{2k+1} \oplus(-1)^{j+\ell}(j+\ell)\Lambda^j \mathfrak{c} \otimes \Lambda^\ell \mathfrak{n} \\
&= \ell \sum_{j=0}^{2k+1} \oplus(-1)^{j+\ell}\Lambda^j \mathfrak{c} \otimes \Lambda^\ell \mathfrak{n} \oplus \sum_{j=0}^{2k+1} \oplus(-1)^{j+\ell} j\Lambda^j \mathfrak{c} \otimes \Lambda^\ell \mathfrak{n} \\
&= (-1)^\ell \ell(\Lambda^{\text{ev}}\mathfrak{c} - \Lambda^{\text{odd}}\mathfrak{c}) \otimes \Lambda^\ell \mathfrak{n} \oplus (-1)^\ell(\Lambda^{\text{odd}}\mathfrak{p}_m - \Lambda^{\text{ev}}\mathfrak{p}_m) \otimes \Lambda^\ell \mathfrak{n}.
\end{aligned}$$

If $Y \in \mathfrak{a}$ has norm 1, then $e(Y) + i(Y)$ is a $K \cap M^+$ intertwining operator between $\Lambda^{\text{ev}}\mathfrak{c}$ and $\Lambda^{\text{odd}}\mathfrak{c}$, and moreover is invertible. So as $K \cap M^+$ modules $\Lambda^{\text{ev}}\mathfrak{c} \simeq \Lambda^{\text{odd}}\mathfrak{c}$ and the Lemma follows. □

The complex Ω^* has a natural $\mathbb{Z}_2$ grading $(\varepsilon|\Omega^k = (-1)^k \text{Id})$ and a $\mathbb{Z}$ grading implemented by the number operator $N, N|\Omega^k = k\text{Id}$. In terms of the datum $(\Omega^*, \varepsilon, N, \triangle_{\varphi,k})$, the θ-function is given by

$$\theta(t) = \text{Tr}_s N \exp -t\triangle_\varphi,$$

where Tr_s denotes the supertrace $\text{Tr} \circ \varepsilon$. On the other hand we have seen there is a natural bigrading, $\Omega^{*,**}$, with each row and column having a natural $\mathbb{Z}_2$ grading. If we denote by $\triangle_\varphi^\ell$ the operators $\triangle_\varphi^{*,\ell}$ on $\Omega^{*,\ell}$, then Lemma 4.1 suggests we compute

$$\theta_\ell(t) = \text{Tr}_s \exp -t\triangle_\varphi^\ell.$$

So we set $f_t^{j,\ell} = \text{tr}\,\tilde{h}_t^{(j,\ell)}$ (tr as before) and we set $f_t^\ell = \sum_{j=0}^{2k}(-1)^j f_t^{j,\ell}$.

Lemma 4.2. *f_t^ℓ is a pseudocusp form on G, moreover,*

$$\hat{f}_{tB}^\ell(\xi, \mu) = e^{t\pi_{\xi,\mu}(\Omega)} \dim [W_\xi \otimes (\Lambda^{\text{odd}}\mathfrak{p}_m - \Lambda^{\text{ev}}\mathfrak{p}_m) \otimes \Lambda^\ell \mathfrak{n}]^{K^M}. \tag{4.2}$$

Proof. We need to show that $\hat{f}_{tA}^\ell = 0$ for $A \succ B$, $A \neq B$. So fix such an A. We may assume that $A \subseteq M_1 = MA_{\mathbb{R}}$; $A = H_I H_{\mathbb{R}}$ with $H_{\mathbb{R}} \supseteq A_{\mathbb{R}}$; $Q_A = M_A H_{\mathbb{R}} N_A$ with $M_A^+ \subseteq M^+$. Let $Y \in H_{\mathbb{R}} \setminus A_{\mathbb{R}}$ with $\|Y\| = 1$. Let $a^* = (\xi, \mu, \nu)\varepsilon A^* = H_I^* \times H_{\mathbb{R}}^*$. By double induction $\pi_{a^*} = \text{Ind}_{M_A^+ H_{\mathbb{R}} N_A}^G(\xi, \mu, \nu) \otimes I = \text{Ind}_{M A_{\mathbb{R}} N}^G(\text{Ind}\,(\xi, \mu) \otimes \nu \otimes I)$. As in the proof of Proposition 2.1 we have

$$\hat{f}_{tA}^\ell(a^*) = e^{t\pi_{a^*}(\Omega)} \dim [W_{\xi,\nu} \otimes (\Lambda^{\text{odd}}\mathfrak{p}_m - \Lambda^{\text{ev}}\mathfrak{p}_m) \otimes \Lambda^\ell \mathfrak{n}]^{K \cap M}$$

and by Frobenius reciprocity

$$\dim [W_{\xi,\mu} \otimes (\Lambda^{\text{odd}}\mathfrak{p}_m - \Lambda^{\text{ev}}\mathfrak{p}_m) \otimes \Lambda^\ell \mathfrak{n}]^{K \cap M} = \dim [W_\xi \otimes (\Lambda^{\text{odd}}\mathfrak{p}_m - \Lambda^{\text{ev}}\mathfrak{p}_m) \otimes \Lambda^\ell \mathfrak{n}]^{K \cap M_A}.$$

But $\Lambda^{\text{odd}}\mathfrak{p}_{\mathfrak{m}} - \Lambda^{\text{ev}}\mathfrak{p}_{\mathfrak{m}} = \pm(\sigma_+ - \sigma_-)^2$, with $\sigma_\pm$ the spin representations of $K \cap M^+$ and $c(Y)$ will be a nontrivial $K \cap M_A$ intertwining operator between σ_+ and σ_-. Hence this multiplicity is zero and so $\hat{f}^\ell_{tA}(a^*) = 0$. □

Lemma 4.3.

$$\theta(t) = \sum_{\ell=0}^{2m} (-1)^\ell \theta_\ell(t).$$

Proof. Since $\operatorname{tr}\tilde{h}_t^{(k)}$ (resp. f_t^ℓ) is admissible, $\theta(t) = \operatorname{Tr}_s N\exp - t\triangle_\varphi$ (resp. $\theta_\ell(t) = \operatorname{Tr}_s\exp - t\triangle_\varphi^\ell$) is given by the orbital integral expansion (2.1). From Proposition 2.1, Lemma 2.3, and Lemma 4.2 we see that

$$\operatorname{Tr}\pi_{\xi,\nu}(\tilde{k}_t) = \sum (-1)^\ell \operatorname{Tr}\pi_{\xi,\nu}(f_t^\ell).$$

Since $\tilde{k}_t$ and the f_t^ℓ's are pseudocusp forms, it follows from Proposition 4.1 in [M-S] that for any regular semisimple element (and hence any semisimple element) the corresponding orbital integrals for $\tilde{k}_t$ and $\sum_{\ell=0}^{2m}(-1)^\ell f_t^\ell$ are equal. In view of (2.1) this completes the proof. □

Remark 4.4. Actually the supports of the Fourier transform of the θ_ℓ's are mutually disjoint for $0 \leqq \ell \leqq m$. Indeed, $\hat{\theta}_\ell(\xi, \mu)$ is supported in ξ in the set $\{\xi \mid \|\Lambda_\xi\|^2 = \|\varrho\|^2 - (m-\ell)^2\}$ as follows from (2.3) and (2.4).

Lemma 4.5. *Let* $h = h_I h_{\mathbb{R}} = h_I \exp rY \in B$.

(i) *If* $h \in B'$ *then* $\mathcal{O}_{f_t^\ell}(h) = \dfrac{e^{-r^2/4t}}{(4\pi t)^{1/2}} \dfrac{e^{-t(m-\ell)^2}}{\triangle_+(h)} \chi_{\Lambda^\ell \mathfrak{n}}(h_I)$.

(ii) *If* $h \neq e$ *then*

$$\mathcal{O}_{f_t^\ell}(h) = c_h^{-1} \frac{e^{-r^2/4t}}{(4\pi t)^{1/2}} \frac{e^{-t(m-\ell)^2} \sum a_{i,\ell} \sum \varepsilon(w)\tilde{\omega}_h(w(\lambda_i + \varrho_M))\xi_{w\cdot\lambda_i}(h_I)}{\xi_\varrho(h) \prod_{\alpha \in P_h^c} [1 - \xi_{-\alpha}(h)]}$$

(iii) $f_t^\ell(e) = \dim F \dim X \int_{-\infty}^{\infty} e^{-t(\nu^2 + (m-\ell)^2)}$
$\times \sum \dim [W_\xi \otimes (\Lambda^{\text{odd}}\mathfrak{p}_{\mathfrak{m}} - \Lambda^{\text{ev}}\mathfrak{p}_{\mathfrak{m}}) \otimes \Lambda^\ell \mathfrak{n}]^{K^M} d\mu(\xi, \nu)$.

In particular, with

$$\sum \dim [W_\xi \otimes (\Lambda^{\text{odd}}\mathfrak{p}_{\mathfrak{m}} - \Lambda^{\text{ev}}\mathfrak{p}_{\mathfrak{m}}) \otimes \Lambda^\ell \mathfrak{n}]^{K^M} \frac{d\mu(\xi, \nu)}{d\nu} = \sum_{k=0}^{m} a_k^\ell \nu^{2k}$$

one has

$$f_t^\ell(e) = \dim F \dim X e^{-t(m-\ell)^2} \sum a_k^\ell \frac{\Gamma(k + 1/2)}{t^{k+1/2}}.$$

Proof. These follow directly from §2, with the last statement a consequence of the fact that only finitely many ξ contribute a non-zero multiplicity and the density $\frac{d\mu(\xi,\nu)}{d\nu}$ is a polynomial in ν^2. □

From this Lemma one immediately gets an expression for $\theta_\ell(t)$. We include it for future reference.

Proposition 4.6.

$$\begin{aligned}\theta_\ell(t) = (-1)\sum_{[\gamma]\in\mathscr{E}_1(\Gamma)} \operatorname{Tr}\varphi(\gamma)\frac{[G_\gamma : G_\gamma^0]^{-1}\operatorname{vol}(G_\gamma/\Gamma_\gamma)c_\gamma^{-1}}{\xi_\varrho(\gamma)\prod_{\alpha\in P_\gamma^c}[1-\xi_{-\alpha}(\gamma)]}\frac{e^{-\ell_\gamma^2/4t}}{(4\pi t)^{1/2}}e^{-t(m-\ell)^2}\\ \sum a_{i,\ell}\sum \varepsilon(w)\tilde{\omega}_\gamma(w(\lambda_i+\varrho_\mu))\xi_{w\cdot\lambda_i}(\gamma_I)\\ +(-1)^\ell \dim F \operatorname{vol} X\int_{-\infty}^{\infty} e^{-t(\nu^2+(m-\ell)^2)}\\ \left(\sum_{\xi\in\mathscr{E}_2(M)}\dim[W_\xi\otimes(\Lambda^{\mathrm{odd}}\mathfrak{p}_{\mathfrak{m}}-\Lambda^{\mathrm{ev}}\mathfrak{p}_{\mathfrak{m}})\otimes\Lambda^\ell\mathfrak{n}]^{K^M}\right)d\mu(\xi,\nu).\end{aligned} \tag{4.3}$$

The results on θ-regularized traces and determinants shall now be brought to bear on the problem of constructing Selberg-type zeta functions in higher rank spaces.

Proposition 4.7.

$$\operatorname{Tr}_\theta(\triangle_\varphi^\ell+\lambda)^{-1} = Pf\int_0^\infty e^{-t\lambda}\theta_\ell(t)\,dt$$

is an analytic function when $\operatorname{Re}\lambda > -\varepsilon_\ell$. *Moreover, with* $\lambda = s^2+2s|m-\ell|$ *and in addition* $\operatorname{Re}(s+|m-\ell|)\notin(-\infty,0]$,

$$\begin{aligned}&2(s+|m-\ell|)\operatorname{Tr}_\theta(\triangle_\varphi^\ell\oplus s^2+2s|m-\ell)^{-1}\\ &\quad=\sum_{[\gamma]\in\mathscr{E}_1(\Gamma)}\operatorname{Tr}\varphi(\gamma)\chi(\widehat{X}_\gamma)\lambda_\gamma e^{-s\ell_\gamma}\cdot\frac{e^{(m-|m-\ell|)\ell_\gamma}}{\det(I-\operatorname{Ad}\gamma|_{\mathfrak{n}})}D_\ell(\gamma)\\ &\quad+Pf\,2(s+|m-\ell|)\int_0^\infty e^{-t(s^2+2s|m-\ell|)}f_t^\ell(e)\,dt\end{aligned} \tag{4.4}$$

where

$$D_\ell(\gamma)=\frac{|W^{\mathbb{C}}(G_\gamma)|^{-1}\prod_{\alpha\in P_\gamma}\langle\varrho_\gamma,\alpha\rangle^{-1}}{\xi_{\varrho_M}(\gamma)\prod_{\alpha\in P_{\gamma,I}^c}[1-\xi_{-\alpha}(\gamma)]}\sum_i a_{i,\ell}\sum\varepsilon(w)\tilde{\omega}_\gamma(w(\lambda_i+\varrho_\mu)\xi_{w\cdot\lambda_i}(\gamma_I). \tag{4.5}$$

Proof. As θ_ℓ is a pseudofunction, the analyticity follows from § 1. Also, given the assumption $\mathbb{F}$ is acyclic the domain $\operatorname{Re}\lambda > -\varepsilon_\ell$ results, ε_ℓ being a lower bound for the positive eigenvalues of the $\triangle_\varphi^{j,\ell}$.

To obtain (4.4) we begin with (4.3) and make various identifications. From

Proposition (3.4)

$$\chi(\widehat{X}_\gamma)\lambda_\gamma = [G_\gamma : G_\gamma^0]^{-1}\mathrm{vol}\,(\Gamma_\gamma \setminus G_\gamma)c_\gamma^{-1}|W^{\mathbb{C}}(G_\gamma)| \prod_{\alpha\in P_\gamma}\langle \varrho_\gamma, \alpha\rangle\,.$$

Also $P_\gamma^c = P_{\gamma,I}^c \cup P_{\mathbb{C}}$, $\xi_\varrho(\gamma) = \xi_{\varrho_M}(\gamma)\xi_{\varrho_{\mathbb{C}}}(\gamma)$, and $\xi_{\varrho_{\mathbb{C}}}(\gamma)\prod_{\alpha\in P_{\mathbb{C}}}[1-\xi_{-\alpha}(\gamma)] = \xi_{\varrho_{\mathbb{C}}}(\gamma^{-1})\det(I-\mathrm{Ad}\,\gamma|_{\mathfrak{n}})$. And as B is fundamental and split rank one, with $\gamma = \gamma_I \exp \ell_\gamma Y$, one has $\xi_{\varrho_{\mathbb{C}}}(\gamma) = e^{m\ell_\gamma}$. Then the result follows from the identity

$$\frac{e^{-z\ell_\gamma}}{2z} = \int_0^\infty e^{-z^2 t}\frac{e^{-\ell_\gamma^2/4t}}{(4\pi t)^{1/2}}\,dt, \quad \mathrm{Re}\,z \notin (-\infty, 0] \tag{4.6}$$

and an interchange of integrals. The justification for the interchange proceeds exactly as in the proof of Proposition 6.1 [M-S] and consequently we only sketch it.

Fix T, $0 < T < \infty$, and set

$$I_T = Pf\int_0^T 2(s+|m-\ell|)\theta_\ell(t)e^{-t(s^2+2s|m-\ell|)}\,dt$$

and

$$I_\infty = \int_T^\infty 2(s+|m-\ell|)\theta_\ell(t)e^{-t(s^2+2s|m-\ell|)}\,dt.$$

Consider first $I_\infty = \mathrm{Tr}_s(2(s+|m-\ell|)\ (\triangle_\varphi^\ell + s^2 + 2s|m-\ell|)^{-1})\ e^{-T(\triangle_\varphi^\ell + s^2 + 2s|m-\ell|)}$. Since $\mathrm{Re}\,(s^2+2s|m-\ell|) > 0$ the kernel of this operator is in $\mathscr{C}^1(G)$ and the trace can be computed by orbital integrals. From Kuga's formula and the argument in [M-S, p. 660], one sees that this kernel is also a pseudocusp form as in Lemma 4.2. Thus, we get the absolutely convergent series

$$\begin{aligned}&\sum_{[\gamma]\in\mathscr{E}_1(\Gamma)}\mathrm{Tr}\,\varphi(\gamma)\chi(\widehat{X}_\gamma)\lambda_\gamma\frac{e^{m\ell_\gamma}}{\det(I-\mathrm{Ad}\,\gamma|_{\mathfrak{n}})}D_\ell(\gamma)\\ &\int_{-\infty}^\infty \frac{2(s+|m-\ell|)}{v^2+(s+|m-\ell|)^2}e^{iv\ell_\gamma}e^{-T(v^2+(s+|m-\ell|)^2)}\,dv\,.\end{aligned} \tag{4.7}$$

For I_T, if we ignore the contribution from the identity element, as we may, we have a convergent integral. We can then proceed as on *p.* 661 getting the absolutely convergent

$$\sum_{[\gamma]\in\mathscr{E}_1(\Gamma)}\mathrm{Tr}\,\varphi(\gamma)\chi(\widehat{X}_\gamma)\frac{\lambda_\gamma e^{m\ell_\gamma}D_\ell(\gamma)}{\det(I-\mathrm{Ad}\,\gamma|_{\mathfrak{n}})}\int_0^T 2(s+|m-\ell|)e^{-t(s+|m-\ell|)^2}\frac{e^{-\ell_\gamma^2/4t}}{(4\pi t)^{1/2}}\,dt\,. \tag{4.8}$$

Adding (4.7) and (4.8) and undoing the Fourier transform in (4.7) yields the expression

$$\sum_{[\gamma]\in\mathscr{E}_1(\Gamma)} \operatorname{Tr}\varphi(\gamma)\chi(\widehat{X}_\gamma)\frac{\lambda_\gamma e^{m\ell_\gamma}D_\ell(\gamma)}{\det(I-\operatorname{Ad}\gamma|_{\mathfrak{n}})}\int_0^\infty 2(s+|m+\ell|)e^{-t(s+|m-\ell|)^2}\frac{e^{-\ell_\gamma^2/4t}}{(4\pi t)^{1/2}}\,dt$$

$$+ Pf\, 2(s+|m-\ell|)\int_0^\infty e^{-t(s^2+2s|m-\ell|)}f_t^\ell(e)\,dt.$$

The integral is evaluated using the identity (4.6), completing the proof. □

We are now in a position to introduce the super Selberg zeta functions.

Definition. For $z\in\mathbb{C}$ satisfying $\operatorname{Re}z^2 > (m-\ell)^2-\varepsilon_\ell$ and $\operatorname{Re}z\notin(-\infty,0]$ define

$$\log Z_\varphi^\ell(z) = -\sum_{[\gamma]\neq 1}\operatorname{Tr}\varphi(\gamma)\chi(\widehat{X}_\gamma)\frac{D_\ell(\gamma)}{\det(I-\operatorname{Ad}\gamma|_{\mathfrak{n}})}\frac{e^{-(z-m)\ell_\gamma}}{\mu_\gamma};$$

recall that μ_γ is the generic multiplicity.

Note that by Remark 3.7 only classes $[\gamma]\in\mathscr{E}_1(\Gamma)$ contribute to the sum. Moreover, this series converges absolutely and uniformly on compact subsets, as follows from the previous proposition and the fact that the length spectrum is bounded away from zero. For the same reason, one can see that $\lim_{z\to+\infty}\log Z_\varphi^\ell(z)=0$.

Remark. The function $Z_\varphi^\ell(s)$ reduces to an ordinary Selberg zeta function only when rank $G/K=1$. For higher rank spaces, $Z_\varphi^\ell(s)$ can be viewed as an alternating product

$$Z_\varphi^\ell(s)=\prod_{j=0}^{2k} Z_\varphi^{\ell,j}(s)^{(-1)^j},\quad 2k=\dim\mathfrak{p}_{\mathfrak{m}}$$

where $\log Z_\varphi^{\ell,j}(s)$ can be defined, at least at a formal level, by a sum reminiscent of the expression for a rank 1 Selberg zeta function, but involving *all* conjugacy classes $[\gamma]$ in Γ, $[\gamma]\neq\{1\}$. Thus, the formula for $Z_\varphi^\ell(s)$ incorporates hidden cancellations which remove the contribution of the conjugacy classes outside $\mathscr{E}_1(\Gamma)$.

Theorem 4.8.

$$Z_\varphi^\ell(z)=\exp-\sum_{[\gamma]\neq 1}\operatorname{Tr}\varphi(\gamma)\chi(\widehat{X}_\gamma)\frac{D_\ell(\gamma)}{\det(I-\operatorname{Ad}\gamma|_{\mathfrak{n}})}\frac{e^{-(z-m)\ell_\gamma}}{\mu_\gamma}$$

is holomorphic if $\operatorname{Re}z^2>(m-\ell)^2-\varepsilon_\ell$ *and* $\operatorname{Re}z\notin(-\infty,0]$. *Moreover* Z_φ^ℓ *has a meromorphic continuation to* $\mathbb{C}$ *given by*

(i) $Z_\varphi^\ell(z)=\det{}_\zeta(\Delta_\varphi^\ell)e^{-2\pi c\sum_{k=0}^{m}(-1)^k a_k^\ell\frac{z^{2k+1}}{2k+1}}\det{}_\theta(I+(z^2-(m-\ell)^2)\Delta_\varphi^{\ell-1})$,
$(c=\dim F\dim X)$;

satisfies a functional equation

(ii) $Z_\varphi^\ell(-z) = e^{4\pi c \sum_{k=0}^{m} (-1)^k a_k^\ell \frac{z^{2k+1}}{2k+1}} Z_\varphi^\ell(z)$;

has the product formula

(iii) $Z_\varphi^\ell(z) Z_\varphi^\ell(-z) = [\det_\zeta(\triangle_\varphi^\ell) \det_\theta(I + (z^2 - (m-\ell)^2)\triangle_\varphi^{\ell-1})]^2$;

and satisfies the duality relation

(iv) $Z_\varphi^\ell(z) = Z_\varphi^{2m-\ell}(z)$.

Proof. From (4.4) we have

$$\frac{d}{dz} \log Z_\varphi^\ell(z) = 2z \operatorname{Tr}_\theta(\triangle_\varphi^\ell + z^2 - (m-\ell)^2)^{-1} - Pf\, 2z \int_0^\infty e^{-t(z^2-(m-\ell)^2)} f_t^{\ell}(e)\, dt$$

while (1.5) gives

$$\frac{d}{dz} \log \det_\theta(I + (z^2 - (m-\ell)^2)\triangle_\varphi^{\ell-1}) = 2z \operatorname{Tr}_\theta(\triangle_\varphi^\ell + z^2 - (m-\ell)^2)^{-1}.$$

Thus we get

$$\log Z_\varphi^\ell(z) = c + \log \det_\theta(I + (z^2 - (m-\ell)^2)\triangle_\varphi^{\ell-1}) + Pf \int_0^\infty e^{-t(z^2-(m-\ell)^2)} f_t^{\ell}(e) \frac{dt}{t}.$$

To prove (i) we evaluate the Pf integral and determine c using $\lim_{z\to+\infty}$. Using (iii) of Lemma 4.5 we compute

$$\begin{aligned} Pf \int_0^\infty e^{-t(z^2-(m-\ell)^2)} f_t^{\ell}(e) \frac{dt}{t} &= Pf \int_0^\infty e^{-tz^2} c \sum_{k=0}^{m} a_k^\ell \Gamma\left(k + \frac{1}{2}\right) \frac{dt}{t^{k+3/2}} \\ &= c \sum_{k=0}^{m} a_k^\ell \Gamma\left(k + \frac{1}{2}\right) \Gamma\left(-k - \frac{1}{2}\right) z^{2k+1} \\ &= -2\pi c \sum_{k=0}^{m} (-1)^k a_k^\ell \frac{z^{2k+1}}{2k+1}, \end{aligned}$$

and we note the small time asymptotics of $\theta_\ell(t)$

$$\theta_\ell(t) \sim c e^{-t(m-\ell)^2} \sum_{k=0}^{m} a_k^\ell \frac{\Gamma(k+\frac{1}{2})}{t^{k+1/2}} = \sum t^{\alpha_n} c_n,$$

where $\alpha_n = n + \frac{1}{2}$, $n \geqq -m-1$, and

$$c_n = \sum_{j-k=n+1} c a_k^\ell \Gamma\left(k + \frac{1}{2}\right) (-1)^j \frac{(m-\ell)^{2j}}{j!},$$

of course $j \geqq 0$ and $m \geqq k \geqq 0$. In the Stirling type estimate (1.12)′ we set $\lambda = z^2 - (m-\ell)^2$ and expand $[1 - (\frac{m-\ell}{z})^2]^{\alpha_n}$ obtaining

$$\begin{aligned}
&\log\det{}_\theta(I + (z^2 - (m-\ell)^2)\Delta_\varphi^{\ell-1}) \\
&\sim -\sum (z^2 - (m-\ell)^2)^{-\alpha_n}\Gamma(\alpha_n)c_n - \log\det{}_\zeta(\Delta_\varphi^\ell) \\
&= -\sum_{n \geqq -m-1} c_n \sum_{r \geqq 0} z^{-(2n+1+2r)}\Gamma\left(n + \frac{1}{2} + r\right)\frac{(m-\ell)^{2r}}{r!} - \log\det{}_\zeta(\Delta_\varphi^\ell).
\end{aligned}$$

Substituting for c_n and summing over j, k we get

$$\begin{aligned}
&\log\det{}_\theta(I + (z^2 - (m-\ell)^2)\Delta_\varphi^{\ell-1}) \\
&\sim -\log\det{}_\zeta(\Delta_\varphi^\ell) \\
&\quad -\sum_{k=0}^{m} z^{+2k+1} c a_k^\ell \Gamma\left(k + \frac{1}{2}\right) \sum_{j,r \geqq 0} (-1)^j \frac{(m-\ell)^{2j+2r}}{j!\,r!} z^{-2j-2r}\Gamma\left(j + r - k - \frac{1}{2}\right) \\
&= -\log\det{}_\zeta(\Delta_\varphi^\ell) \\
&\quad -\sum_{k=0}^{m} z^{+2k+1} c a_k^\ell \Gamma\left(k + \frac{1}{2}\right) \sum_{u=0}^{\infty} z^{-2u}(m-\ell)^{2u}\Gamma\left(u - k - \frac{1}{2}\right) \sum_{v=0}^{u} (-1)^v \frac{1}{v!\,(u-v)!}.
\end{aligned}$$

But the sum over v is zero unless $u = 0$, hence

$$\begin{aligned}
&\log\det{}_\theta(I + (z^2 - (m-\ell)^2)\Delta_\varphi^{\ell-1}) \\
&\quad \sim -\log\det{}_\zeta(\Delta_\varphi^\ell) - \sum_{k=0}^{m} z^{+2k+1} c a_k^\ell \Gamma\left(k + \frac{1}{2}\right)\Gamma\left(-k - \frac{1}{2}\right).
\end{aligned}$$

As $z \to +\infty$, $\log Z_\varphi^\ell(z) \to 0$ and hence $c = \log\det{}_\zeta \Delta_\varphi^\ell$ and the proof of (i) is complete. There is a shorter proof of (i) using the approximation $(1-x)^{-\alpha_n} \sim 1 + O(x)$, $x \to 0^+$ but the one given shows that $\log Z_\varphi^\ell$ vanishes to ∞-order at ∞.

The functional equation (ii) follows immediately from the form of the exponential factor and the z^2 in the θ-regularized determinant. Also (iii) is obvious from (ii). For (iv) we note that as representations of $M, \Lambda^\ell \mathfrak{n}$ and $\Lambda^{2m-\ell}\mathfrak{n}$ are equivalent hence $D_\ell(\gamma) = D_{2m-\ell}(\gamma)$. □

Lemma 4.9. *Let $h = h_I h_{\mathbb{R}} \in B$, $h \neq e$. Then*

$$\det(I - \mathrm{Ad}\, h|_{\mathfrak{n}}) = \frac{\widetilde{\omega}_h\{\xi_{\varrho_M}{}' \Delta_M \sum (-1)^\ell \chi_{\Lambda^\ell \mathfrak{n}}\}(h)}{|W^{\mathbb{C}}(G_h)| \prod_{\alpha \in P_h} \langle \varrho_h, \alpha\rangle \xi_{\varrho_M}(h) \prod_{\alpha \in P_{h,I}^c} [1 - \xi_{-\alpha}(h)]}. \tag{4.9}$$

Proof. When h is regular, $\widetilde{\omega}_h = 1$. Then (4.9) is the well-known

$$\det(I - \mathrm{Ad}\, h|_{\mathfrak{n}}) = \sum (-1)^\ell \mathrm{Tr}\,(\Lambda^\ell \mathrm{Ad}\, h|_{\mathfrak{n}}). \tag{4.10}$$

If we multiply (4.10) by $\xi_{\varrho_M}{}'\triangle_M$, then (4.9) results essentially from L'Hôpital's rule. Indeed since B is fundamental, split-rank one and $h \neq e$, we have $\triangle(\mathfrak{g}_h)$ consists of imaginary roots. Then the only non-zero contribution to $\widetilde{\omega}_h\{\xi_{\varrho_M}{}'\triangle_M \det(I - \mathrm{Ad}(\cdot)|_{\mathfrak{n}})\}(h)$ is $\det(I - \mathrm{Ad}\, h|_{\mathfrak{n}})\xi_{\varrho_M}(h)\widetilde{\omega}_h({}'\triangle_M)(h)$. But

$$\widetilde{\omega}_h\Big\{\prod_{\alpha\in\triangle_M^+}[1-\xi_{-\alpha}(\cdot)]\Big\}(h) = \sum_{\sigma}\prod_{\alpha\in P_{h,I}}\langle\alpha,\sigma\alpha\rangle\prod_{\alpha\in P_{h,I}^c}[1-\xi_{-\alpha}(h)]$$

where σ runs over the permutation group on $P_{h,I}$. In [Sc, (5.14)], we find

$$\sum_{\sigma}\prod_{\alpha\in P_{h,I}}\langle\alpha,\sigma\alpha\rangle = |W^{\mathbb{C}}(G_h)|\prod_{\alpha\in P_{h,I}}\langle\varrho_h,\alpha\rangle\,. \quad \square$$

Lemma 4.10. *Let $h = h_I \exp rY \in B$, $h \neq e$. Then*

$$\det(I - \mathrm{Ad}\, h|_{\mathfrak{n}}) = \frac{\sum_{\ell}(-1)^{\ell}e^{r\ell}\sum_i a_{i,\ell}\sum\varepsilon(w)\widetilde{\omega}_h(w(\lambda_i+\varrho_M))\xi_{w\cdot\lambda_i}(h_I)}{|W^{\mathbb{C}}(G_h)|\prod_{\alpha\in P_{h,I}}\langle\varrho_h,\alpha\rangle\xi_{\varrho_M}(h)\prod_{\alpha\in P_{h,I}^c}[1-\xi_{-\alpha}(h)]}\,. \tag{4.11}$$

Proof. Recall the notation from Lemma 2.8,

$$\chi_{\Lambda^{\ell}\mathfrak{n}} = \sum_i a_{i,\ell}\chi_{\lambda_i}$$

as representations of M^+. As observed before $\mathrm{Ad}(\exp rY)|_{\mathfrak{n}} = e^r\mathrm{Id}$. Then (4.11) follows from (4.9) and the Weyl character formula. $\square$

We are now able to prove the main result that relates R-torsion to a special value of a zeta function.

Theorem 4.11.

$$Z_{\varphi}(s) = \exp -\sum_{[\gamma]\neq 1}\mathrm{Tr}\,\varphi(\gamma)\chi(\widehat{X}_{\gamma})\frac{e^{-s\ell_{\gamma}}}{\mu_{\gamma}}$$

is an analytic function in the region $\mathrm{Re}\, s > m$ *and* $\mathrm{Re}\,(s-m)^2 \geqq m^2$. *Moreover*

$$Z_{\varphi}(s) = \prod_{\ell=0}^{2m} Z_{\varphi}^{\ell}(s-m+\ell)^{(-1)^{\ell}}\,;$$

in particular Z_{φ} admits a meromorphic continuation to $\mathbb{C}$ with $s = 0$ a regular point. Also

$$Z_{\varphi}(0) = \tau_{\zeta}(0)^2\,.$$

Proof. For $\mathrm{Re}\, s$ sufficiently large we have

$$\sum_{\ell=0}^{2m}(-1)^{\ell}\log Z_{\varphi}^{\ell}(s-m+\ell)$$

$$= -\sum_{[\gamma]\in\mathscr{P}_1(\Gamma)}\mathrm{Tr}\,\varphi(\gamma)\frac{\chi(\widehat{X}_{\gamma})}{\det(I-\mathrm{Ad}\,\gamma|_{\mathfrak{n}})}\frac{e^{-s\ell_{\gamma}}}{\mu_{\gamma}}\cdot\Big[\sum_{\ell=0}^{2m}(-1)^{\ell}e^{(2m-\ell)\ell_{\gamma}}D_{\ell}(\gamma)\Big]\,.$$

But as noted before $D_\ell(\gamma) = D_{2m-\ell}(\gamma)$, so the expression in brackets is $\sum_{\ell=0}^{2m}(-1)^\ell e^{\ell\ell_\gamma} D_\ell(\gamma)$. However, from (4.11) we see that this is just $\det(I - \mathrm{Ad}\,\gamma|_{\mathfrak{n}})$, hence we obtain the series for $\log Z_\varphi(s)$. An examination of the domain of convergence for each super Selberg zeta function shows that each is contained in the stated domain. The meromorphic continuation follows from that of each Z_φ^ℓ and the regularity of $s = 0$ from the acyclicity assumption.
Finally, using (iv) and (iii) of Theorem 4.8 we obtain

$$\begin{aligned}
Z_\varphi(0) &= \prod_{\ell=0}^{2m} Z_\varphi^\ell(\ell - m)^{(-1)^\ell} \\
&= \prod_{\ell=0}^{m} Z_\varphi^\ell(-|m-\ell|)^{(-1)^\ell} \prod_{\ell=m+1}^{2m} Z_\varphi^\ell(|m-\ell|)^{(-1)^\ell} \\
&= \prod_{\ell=0}^{m} Z_\varphi^\ell(-|m-\ell|)^{(-1)^\ell} \prod_{\ell=0}^{m-1} Z_\varphi^\ell(|m-\ell|)^{(-1)^\ell} \\
&= \prod_{\ell=0}^{m-1} [\det{}_\zeta^2(\triangle_\varphi^\ell)]^{(-1)^\ell} \det{}_\zeta(\triangle_\varphi^m)^{(-1)^m} \\
&= \prod_{\ell=0}^{2m} [\det{}_\zeta(\triangle_\varphi^\ell)]^{(-1)^\ell} \\
&= \tau_\zeta(0)^2.
\end{aligned}$$

Here, the last equality follows from Lemma 4.3. □

References

[B-M] Barbasch, D., Moscovici, H.: L^2-index and the Selberg trace formula. J. Funct. Anal. **53**, 151–201 (1983)

[B-W] Borel, A., Wallach, N.: Continuous cohomology, discrete subgroups, and representations of reductive groups. Ann. Math. Stud. **94**, 1–387 (1980)

[C] Cheeger, J.: Analytic torsion and the heat equation. Ann. Math. **109**, 259–322 (1979)

[D-K-V] Duistermaat, J.J., Kolk, J.A.C., Varadarajan, V.S.: Spectra of compact locally symmetric manifolds of negative curvature. Invent. Math. **52**, 27–93 (1979)

[E] Eberlein, P.: A canonical form for compact nonpositively curved manifolds whose fundamental groups have nontrivial center. Math. Ann. **260**, 23–29 (1982)

[F1] Fried, D.: Lefschetz formulas for flows. In: Verjovsky A. (ed.) The Lefschetz Centennial Conference, Part III. Contemp. Mathematics, vol. **58** Providence: Am. Math. Soc. 1987

[F2] Fried, D.: Counting circles. *Dynamical Systems.* (Lect. Notes Math., vol. **1342**, pp. 196–215 Berlin Heidelberg New York, Springer 1988

[F3] Fried, D.: Analytic torsion and closed geodesics on hyperbolic manifolds. Invent. Math. **84**, 523–540 (1986)

[G-S] Günther, P., Schimming, R.: Curvature and Spectrum of Compact Riemannian manifolds. J. Differ. Geometry **12**, 599–618 (1977)

[H] Haefliger, A.: Groupoids d'holonomie et classifiants, *Structure transverse des feuilletages.* Astérisque **116**, 70–97 (1984)

[H-C;I] Harish-Chandra: Harmonic analysis on real reductive groups, I. J. Funct. Anal. **19**, 104–204 (1975)

[H-C;III] Harish-Chandra: Harmonic analysis on real reductive groups, III. The Maass-Selberg relations and the Plancherel formula. Ann. Math. **104**, 117–201 (1976)

[H-C; DSII] Harish-Chandra: Discrete series for semisimple Lie groups, II. Acta Math. **116**, 1–111 (1966)

[H-C; S] Harish-Chandra: Supertempered distributions on real reductive groups. Adv. Math., Suppl. Stud. **8**, 139–153 (1983)

[H-P] Hotta, R., Parthasarathy, R.: A geometric meaning of the multiplicity of integrable discrete classes in $L^2(\Gamma \setminus G)$. Osaka J. Math. **10**, 211–234 (1973)

[M] Milnor, J.: Infinite cyclic coverings. In: Conf. on the Topology of Manifolds, pp. 115–133, Boston: Prindle, Weber & Schmidt 1968

[M-S] Moscovici, H., Stanton, R.J.: Eta invariants of Dirac operators on locally symmetric manifolds. Invent. Math. **95**, 629–666 (1989)

[M-S,2] Moscovici, H., Stanton, R.J.: Holomorphic torsion for Hermitian locally symmetric manifolds. (in preparation)

[Mu] Müller, W.: Analytic torsion and R-torsion of Riemannian manifolds, Adv. Math. Suppl. Stud. **28**, 233–305 (1978)

[R-S1] Ray, D.B., Singer, I.: R-torsion and the Laplacian on Riemannian manifolds. Adv. Math. Suppl. Stud. **7**, 145–210 (1971)

[R-S2] Ray, D.B., Singer, I.: Analytic torsion for complex manifolds. Ann. Math. **98**, 154–177 (1973)

[Sc] Schmid, W.: On a conjecture of Langlands, Ann. Math. **93**, 1–42 (1971)

[S] Schwartz, L.: Théorie des distributions. Paris: Hermann 1966

[Se] Selberg, A.: Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. J. Indian Math. Soc. **20**, 47–87 (1956)

[V] Voros, A.: Spectral functions, special functions and the Selberg zeta function. Commun. Math. Phys. **110**, 439–465 (1987)

[W] Wallach, N.R.: On the Selberg trace formula in the case of compact quotient. Bull. Am. Math. Soc. **82**, 171–195 (1976)

Invent. math. 105, 217–220 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Nichtlinearisierbare Operationen halbeinfacher Gruppen auf affinen Räumen

Friedrich Knop*

Mathematisches Institut, Rheinsprung 21, CH-4051 Basel

Oblatum 26-X-1990 & 8-XII-1990

Es bestand längere Zeit das Problem, ob jede algebraische Operation einer komplexen reduktiven Gruppe G auf einem affinen Raum $\mathbb{C}^N$ linearisierbar ist (siehe etwa [K] und die dort angegebene Literatur). In [S] gab Schwarz erste Gegenbeispiele für die orthogonale Gruppe $O_2(\mathbb{C})$ sowie für $\mathrm{SL}_2(\mathbb{C})$ an. Die Konstruktion beruht auf den von Kraft und Schwarz in [KS] entwickelten Methoden, und wird dort auch angewendet, um Gegenbeispiele für alle einfachen Gruppen außer F_4 und E_6 zu konstruieren.

In der vorliegenden Arbeit konstruieren wir für jede nichtkommutative, zusammenhängende, reduktive Gruppe, insbesondere also für jede halbeinfache Gruppe, eine effektive nichtlinearisierbare Operation auf einem affinen Raum. Dazu führen wir den allgemeinen Fall auf den Fall $G=\mathrm{SL}_2(\mathbb{C})$ aus [S] zurück. Die Klassifikation der halbeinfachen Gruppen wird nicht benutzt.

Weitere Konstruktionen für gewisse einfache oder endliche Gruppen werden in [MP1], [MP2] und [MMP] angegeben.

Sei G im folgenden eine nichtkommutative, zusammenhängende, reduktive Gruppe. Wähle einen maximalen Torus $T\subset G$ mit Liealgebra $\mathfrak{t}\subset\mathfrak{g}$. Sei $\Sigma\subseteq\mathfrak{t}^*$ ein System einfacher Wurzeln, und wähle $\alpha\in\Sigma$ beliebig, aber fest. Sei dann L die von T und den Wurzeluntergruppen U_α, $U_{-\alpha}$ erzeugte Leviuntergruppe. Mit L' bezeichnen wir ihre Kommutatoruntergruppe und mit Z ihr Zentrum. Dann ist $L=ZL'$ und $L'\cong\mathrm{SL}_2(\mathbb{C})$ oder $\cong\mathrm{SO}_3(\mathbb{C})$. Wähle ein $\xi_0\in\mathfrak{t}$, dessen Zentralisator genau L ist. Mit $\mathfrak{a}$ bezeichnen wir den dreidimensionalen affinen Teilraum $\xi_0+\mathrm{Lie}\,L'\subseteq\mathfrak{g}$. Als L'-Varietät ist er isomorph zur adjungierten Darstellung.

Sei $\chi\in\mathfrak{X}(T)$ ein dominantes Gewicht und M_χ (bzw. M'_χ) der zugehörige irreduzible G-Modul (bzw. L-Modul). Die Einschränkung von χ auf Z sei χ'.

Für eine G-Varietät X, ein $x\in X^G$ und eine G-Darstellung M sei $\mathrm{VB}_G(X,x,M)$ die Menge der Isomorphieklassen von G-Vektorbündeln auf X, deren Faser über x isomorph zu M ist. Das triviale Vektorbündel $X\times M$ werde mit $\Theta_X(M)$ bezeichnet.

Sei $\mathscr{V}'$ ein L-Vektorbündel auf $\mathfrak{a}$. Dann operiert Z trivial auf der Basis und linear auf jeder Faser. Also bildet die Vereinigung der χ'-Eigenräume der Fasern ein Unterbündel $\mathscr{V}'_{\chi'}$. Betrachte M_χ als L-Modul. Dann sieht man leicht,

* Unterstützt durch den Schweizerischen Nationalfonds zur Förderung der wissenschaftlichen Forschung.

daß der χ'-Eigenraum von Z isomorph zu $M_{\chi'}$ ist. Dies liefert damit folgende Abbildung:

$$\Phi_\chi\colon \mathrm{VB}_G(\mathfrak{g}, 0, M_\chi) \to \mathrm{VB}_L(\mathfrak{a}, \xi_0, M'_\chi)\colon \mathscr{V} \mapsto (\mathscr{V}|_{\mathfrak{a}})_{\chi'}$$

Das Hauptresultat ist dann:

Satz. *Für ein reguläres, dominantes Gewicht χ ist Φ_χ surjektiv.*

Daraus folgt:

Korollar. *Es gibt nichttriviale G-Vektorbündel über* $\mathfrak{g}$.

Beweis. Nach [S] ist $|\mathrm{VB}_L(\mathfrak{a}, \xi_0, M'_\chi)| > 1$ für $\langle \chi, \alpha \rangle \geqq 3$. □

Korollar. *Es gibt eine effektive, nichtlinearisierbare Operation von G auf einem affinen Raum.*

Beweis. Nach dem Bass-Haboush-Trick [BH] genügt es, ein Vektorbündel $\mathscr{V}$ über $\mathfrak{g}$ zu finden, für das $\mathscr{V} \oplus \Theta_{\mathfrak{g}}(\mathfrak{g})$ nichttrivial ist. Wegen [S] können wir uns auf den Fall $\mathrm{rg}\, G \geqq 2$ beschränken. Wähle zunächst irgendeine treue Darstellung E von G und einen regulären dominanten Charakter χ mit:
(i) Es gibt $\mathscr{V}^0 \in \mathrm{VB}_G(\mathfrak{g}, 0, M_\chi)$, so daß $\Phi_\chi(\mathscr{V}^0)$ nichttrivial ist.
(ii) χ' ist kein Gewicht von Z in $E \oplus \mathfrak{g}$.
Setze dann $\mathscr{V} := \mathscr{V}^0 \oplus \Theta_{\mathfrak{g}}(E)$. Wegen

$$((\mathscr{V} \oplus \Theta_{\mathfrak{g}}(\mathfrak{g}))|_{\mathfrak{a}})_{\chi'} = (\mathscr{V}^0|_{\mathfrak{a}} \oplus \Theta_{\mathfrak{a}}(E \oplus \mathfrak{g}))_{\chi'} = \Phi_\chi(\mathscr{V}^0)$$

leistet $\mathscr{V}$ das Gewünschte. □

Beweis des Satzes. Sei $\mathfrak{g}_0 \subset \mathfrak{g}$ die Menge der regulär halbeinfachen Elemente. Ihr Komplement besteht genau aus den Punkten, in denen die Diskriminante $\Delta \in \mathbb{C}[\mathfrak{g}]$ verschwindet. Für eine endliche Teilmenge $S \subset \mathbb{C}$ sei $\mathfrak{g}_S := \{\xi \in \mathfrak{g} \mid \Delta(\xi) \notin S\}$. Analog setzen wir $\mathfrak{t}_0 := \mathfrak{t} \cap \mathfrak{g}_0$ und $\mathfrak{t}_S := \mathfrak{t} \cap \mathfrak{g}_S$. Sei W die Weylgruppe von G. Weiter sei $\delta \in \mathbb{C}[\mathfrak{t}]$ das Produkt der positiven Wurzeln. Es gilt $\Delta|_{\mathfrak{t}} = \delta^2$.

Wir konstruieren nun ein Vektorbündel $\mathscr{V}$, indem wir triviale Vektorbündel über $\mathfrak{g}_0$ und $\mathfrak{g}_S$ für geeignetes $S \subset \mathbb{C}^*$ zusammenkleben. Wir haben $\mathfrak{g}_0 \cap \mathfrak{g}_S = \mathfrak{g}_{S_0}$ mit $S_0 = S \cup \{0\}$. Die Verklebungsabbildung ist ein G-äquivarianter Morphismus

$$\varphi\colon \mathfrak{g}_{S_0} \to \mathrm{GL}(M_\chi),$$

wobei G auf der rechten Seite durch Konjugation operiert. Wegen $\mathfrak{g}_{S_0} \subseteq \mathfrak{g}_0$ gilt $\mathfrak{g}_{S_0} = G \times^{N(T)} \mathfrak{t}_{S_0}$. Also ist φ eindeutig durch die $N(T)$−äquivariante Einschränkung

$$\varphi'\colon \mathfrak{t}_{S_0} \to \mathrm{GL}(M_\chi)$$

bestimmt. Sei

$$M_\chi = \bigoplus_{\eta \in \mathscr{X}(T)} M(\eta)$$

die Zerlegung von M_χ in Gewichträume. Das Bild von φ' liegt im Zentralisator von T in $\mathrm{GL}(M_\chi)$, und dieser ist gleich $\prod_{\eta \in \mathfrak{X}(T)} \mathrm{GL}(M(\eta))$. Also ist φ' gegeben durch eine Familie von Morphismen

$$\varphi_\eta \colon \mathfrak{t}_{S_0} \to \mathrm{GL}(M(\eta))$$

mit

$$\varphi_{w\eta}(\xi) = \bar{w} \circ \varphi_\eta(w^{-1}\xi) \circ \bar{w}^{-1} \qquad \text{für alle } w \in W.$$

Dabei ist $\bar{w} \in N(T)$ ein beliebiger Repräsentant von w und bezeichnet gleichzeitig den induzierten Homomorphismus $M(\eta) \to M(w\eta)$. Wir interessieren uns im folgenden nur für solche φ, für die $\varphi_\eta \equiv 1$ ist, außer, wenn $\eta = w(\chi - n\alpha)$ mit $w \in W$ und $n \in \mathbb{N}$ ist. Für solche η ist $M(\eta)$ eindimensional und daher $\varphi_\eta \colon \mathfrak{t}_{S_0} \to \mathbb{C}^*$. Da χ nach Voraussetzung regulär ist, ist die Standgruppe von $\chi - n\alpha$ in W eine Untergruppe von $W_\alpha := \{1, s_\alpha\}$, wobei s_α die Spiegelung an α ist. Also wird φ gegeben durch eine Familie von Einheiten

$$\varphi_n := \chi_{\chi - n\alpha} \colon \mathfrak{t}_{S_0} \to \mathbb{C}^* \qquad (n = 0, \ldots, n_0 := \langle \chi, \alpha \rangle),$$

die nur der Einschränkung

$$\varphi_n(s_\alpha \xi) = \varphi_{n_0 - n}(\xi) \qquad \text{für alle } n \in \mathbb{N}$$

unterliegen. Wir können also φ_n für $n < n_0/2$ frei wählen, während $\varphi_n \circ s_\alpha = \varphi_n$ für $n = n_0/2$ gelten muß. Weiterhin beschränken wir uns auf solche Funktionen, die sich als rationale Funktion in δ ausdrücken lassen: $\varphi_n(\xi) = p_n(\delta(\xi))$. Die p_n sind für $0 \leqq n < n_0/2$ beliebig. Für $n = n_0/2$ gilt $p_n(-t) = p_n(t)$. Sind umgekehrt solche Funktionen gegeben, so definieren sie einen Morphismus $\varphi_p \colon \mathfrak{g}_{S_0} \to \mathrm{GL}(M_\chi)$.

Seien

$$\varphi_0 \colon \mathfrak{g}_0 \to \mathrm{GL}(M_\chi), \qquad \varphi_S \colon \mathfrak{g}_S \to \mathrm{GL}(M_\chi)$$

G-Morphismen. Dann liefern φ und $\varphi_0 \cdot \varphi \cdot \varphi_S$ offenbar isomorphe Vektorbündel. Für $0 \leqq n \leqq n_0/2$ sei $a_n \in \mathbb{C}^*$ und $\nu_n \in \mathbb{Z}$ (mit ν_n gerade für $n = n_0/2$). Dann definieren die Funktionen $a_n \delta^{\nu_n}$ einen G-Morphismus $\varphi_0 \colon \mathfrak{g}_0 \to \mathrm{GL}(M_\chi)$. Also führen die Funktionen p_n und $(a_n \delta^{\nu_n} p_n)$ zu isomorphen Vektorbündeln, d.h. man kann die p_n so normieren, daß sie bei Null definiert und dort gleich eins sind. Wir betrachten ab jetzt nur noch solcherart normierte Funktionen.

Im folgenden Lemma nützen wir φ_S aus. Sei v_0 die Bewertung des Körpers $\mathbb{C}(\delta)$ in $\delta = 0$.

Lemma. *Es gibt ein $N \in \mathbb{N}$ mit folgender Eigenschaft: Zwei Familien $p_n(\delta)$, $\bar{p}_n(\delta)$ normierter, rationaler Funktionen mit*

$$v_0(p_n - \bar{p}_n) \geqq N \qquad \textit{für alle } n$$

definieren isomorphe Vektorbündel.

Beweis. Sei $q_n := p_n/\bar{p}_n$. Das Problem ist, ob sich $\varphi_{q_\cdot}$ zu einem Morphismus $\varphi_S \colon \mathfrak{g}_S \to \mathrm{GL}(M_\chi)$ fortsetzen läßt. Entwickle dazu $q_n(\delta) = 1 + a_{n1}\,\delta + a_{n2}\,\delta^2 + \ldots$ in eine

Potenzreihe. Dann hat die induzierte Abbildung $\varphi_{q.}\colon \mathfrak{g}_{S_0} \to \mathrm{GL}(M_\chi) \subset \mathrm{End}(M_\chi)$ eine Reihenentwicklung der Form

$$\sum_{\nu \gg -\infty} f_\nu(\xi, a_{ni})\, \Delta^\nu,$$

wobei jeder Koeffizient f_ν polynomial von $\xi \in \mathfrak{g}$ und endlich vielen a_{ni} abhängt. Wenn alle $a_{ni}=0$ sind, ist $\varphi_{q.} \equiv \mathrm{id}_{M_\chi}$ und damit fortsetzbar. Also ist $\varphi_{q.}$ auch fortsetzbar, wenn nur die endlich vielen a_{ni}, die in einem f_ν mit $\nu \leqq 0$ vorkommen, gleich null sind. □

Im Fall $\mathrm{rg}\, G = 1$ liefert unsere Konstruktion *alle* G-Vektorbündel $\mathscr{V}$ über $\mathfrak{g}$ ([S]). Dies folgt aus der Tatsache, daß es dann immer eine endliche Menge $S \subset \mathbb{C}^*$ gibt, so daß $\mathscr{V}$ über $\mathfrak{g}_0$ und $\mathfrak{g}_S$ trivial ist ([KS, Chap. VII: Prop. 1.1 (3) & Thm. 2.6 (1)]). Insbesondere wird jedes $\mathscr{V}' \in \mathrm{VB}_L(\mathfrak{a}, \xi_0, M'_\chi)$ definiert durch geeignete normierte rationale Funktionen $q_n(\lambda)$ auf $\mathfrak{s} := \mathfrak{t} \cap \mathfrak{a} = \{\xi_0 + \lambda h_\alpha \mid \lambda \in \mathbb{C}\}$. Dabei ist $h_\alpha \in \mathfrak{t} \cap \mathrm{Lie}\, L'$ mit $\alpha(h_\alpha) = 1$.

Das Polynom δ verschwindet genau in den Spiegelungsebenen von $\mathfrak{t}$, und zwar mit Multiplizität eins. Nach Wahl liegt ξ_0 auf genau einer solchen Spiegelungsebene, nämlich der zu s_α, und $\mathfrak{s}$ schneidet diese dort transversal. Es folgt, daß $\delta_0 := \delta|_{\mathfrak{s}}$ eine einfache Nullstelle in ξ_0 hat, d.h. v_0-adisch liegt $\mathbb{C}(\delta_0)$ dicht in $\mathbb{C}(\mathfrak{s}) = \mathbb{C}(\lambda)$. Daher gibt es zu den Funktionen q_n, die $\mathscr{V}'$ definieren, normierte rationale Funktionen p_n mit $v_0(p_n(\delta_0) - q_n) \geqq N$. Dabei ist $N \in \mathbb{N}$ wie im Lemma, aber auf L' und nicht auf G angewendet. Durch $p_n(\delta)$ wird dann ein Vektorbündel $\mathscr{V}$ auf $\mathfrak{g}$ definiert. Das Bündel $\Phi_\chi(\mathscr{V})$ wird wiederum geliefert durch $p_n(\delta)|_{\mathfrak{s}} = p_n(\delta_0)$. Also ist $\Phi_\chi(\mathscr{V}) \cong \mathscr{V}'$. □

Diese Arbeit entstand im Frühjahr 1989 während eines Aufenthaltes am Max-Planck-Institut für Mathematik in Bonn. Ich danke dem Institut für die Unterstützung und Gastfreundschaft.

Literatur

[BH] Bass, H., Haboush, W.: Some equivariant K-theory of affine algebraic group actions. Commun. Alg. **15**, 181–217 (1987)

[K] Kraft, H.: Algebraic automorphisms of affine space. In: H. Kraft, T. Petrie, G. Schwarz (eds.) Topological methods in algebraic transformation groups. (Prog. Math., vol. 80, pp. 81–105) Boston Basel Berlin: Birkhäuser 1989

[KS] Kraft, H., Schwarz, G.: Reductive group actions with one dimensional quotient. (Preprint 1990)

[MP1] Masuda, M., Petrie, T.: Equivariant Serre conjecture and algebraic actions on $\mathbb{C}^n$. (Preprint 1990)

[MP2] Masuda, M., Petrie, T.: Equivariant algebraic vector bundles over representations of reductive groups, I. (Preprint 1990)

[MMP] Masuda, M., Moser-Jauslin, L., Petrie, T.: Equivariant algebraic vector bundles over representations of reductive groups, II. (Preprint 1990)

[S] Schwarz, G.: Exotic algebraic group actions. C.R. Acad. Sci., Paris, Sér. I. **309**, 89–94 (1989)

Nachtrag bei der Korrektur

Durch eine genauere Betrachtung des $\mathrm{SL}_2(\mathbb{C})$-Falles kann man zeigen, daß die beiden Korollare auch für nicht zusammenhängende Gruppen gültig sind, wenn nur die Zusammenhangskomponente der Eins kein Torus ist.

Invent. math. 105, 221–245 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Groupe de Chow de codimension deux des variétés définies sur un corps de nombres: un théorème de finitude pour la torsion

Jean-Louis Colliot-Thélène[1] **et Wayne Raskind**[2]

[1] CNRS, URA 752, Mathématiques, Bâtiment 425, Université de Paris-Sud, F-91405 Orsay, France
[2] Faculty of Science, Department of Mathematics, University of Tokyo, Hongo, Tokyo, 113, Japan and Department of Mathematics, University of Arizona, Tucson AZ 85721, USA

Oblatum 17-IX-1990

Summary. Let X be a smooth, projective variety defined over a number field and let $\mathrm{CH}^2(X)$ denote the Chow group of codimension two cycles modulo rational equivalence. We show that if the cohomology group $H^2(X, \mathcal{O}_X)$ vanishes then the torsion subgroup of $\mathrm{CH}^2(X)$ is a finite group. This result covers all previous results in this direction. The hypothesis $H^2(X, \mathcal{O}_X)=0$ is used to lift line bundles.

Sur la base de résultats classiques, tels que les théorèmes de Mordell-Weil et de Néron-Severi, H. Bass a conjecturé que le groupe de Grothendieck $K_0(X)$ d'une variété lisse X définie sur un corps de nombres est un groupe de type fini. Une variante légèrement plus forte de cette conjecture postule que les groupes de Chow $\mathrm{CH}^i(X)$ des cycles de codimension i modulo l'équivalence rationnelle sont des groupes de type fini; lorsque X est une surface, la formule de Riemann-Roch sans dénominateurs de Grothendieck-Jouanolou ([F, 15.3.6]) montre que ces conjectures sont équivalentes.

Dans cet article, nous nous intéressons au sous-groupe de torsion $\mathrm{CH}^2(X)_{\mathrm{tors}}$ du groupe de Chow $\mathrm{CH}^2(X)$ des cycles de codimension 2. Notre premier résultat est le théorème 3.4:

Théorème A *Soit X une variété projective, lisse et géométriquement connexe définie sur un corps de nombres k. Supposons le groupe $H^2(X, \mathcal{O}_X)$ nul. Alors pour tout entier $n>0$, le sous-groupe de n-torsion du groupe de Chow $\mathrm{CH}^2(X)$ des cycles de codimension deux est fini.*

Ce résultat permet déjà d'établir la finitude de tout le groupe $\mathrm{CH}^2(X)_{\mathrm{tors}}$ sous l'une des hypothèses supplémentaires suivantes:

(i) Le groupe $H^1(X, \mathcal{O}_X)$ est aussi nul (Théorème 4.1);
(ii) X est une surface (§ 5).

Lorsque ces deux hypothèses sont simultanément satisfaites, Coombes [Cb] a une autre démonstration de ce résultat.

Cependant Salberger vient d'établir le résultat suivant:

Théorème B (Salberger) *Soit X une variété projective, lisse et géométriquement connexe définie sur un corps de nombres k. Supposons le groupe $H^2(X, \mathcal{O}_X)$ nul. Alors le groupe $\mathrm{CH}^2(X)_{\mathrm{tors}}$ est d'exposant fini.*

La démonstration de ce théorème (Théorème 4.3 ci-dessous) repose sur plusieurs de nos résultats antérieurs [CT/R1, R3]. L'innovation est la Proposition 4.2 (nous offrons une variante de la démonstration de Salberger).

La combinaison des théorèmes A et B donne immédiatement le résultat général, qui les englobe tous deux:

Théorème C *Soit X une variété projective, lisse et géométriquement connexe définie sur un corps de nombres k. Supposons le groupe $H^2(X, \mathcal{O}_X)$ nul. Alors le groupe $\mathrm{CH}^2(X)_{\mathrm{tors}}$ est fini.*

Soit X une surface projective et lisse sur un corps k (car.$(k)=0$), et soit $A_0(X) \subset \mathrm{CH}^2(X)$ le sous-groupe des classes de cycles de degré zéro. Soit $\underline{\mathrm{Alb}}_X$ la variété d'Albanese de X (c'est la variété abélienne duale de la variété de Picard). Si X satisfait $H^2(X, \mathcal{O}_X)=0$, une conjecture de Bloch (voir [Bl/K/L]) implique que le noyau de l'application d'Albanese $A_0(X) \to \underline{\mathrm{Alb}}_X(k)$ est contenu dans le sous-groupe de torsion de $\mathrm{CH}^2(X)$. Cette conjecture de Bloch a été vérifiée (en caractéristique zéro) pour toutes les surfaces qui ne sont pas de type général [Bl/K/L], par exemple pour les surfaces d'Enriques, mais elle l'a aussi été pour certaines surfaces de type général, comme les surfaces de Godeaux (voir [R1, Remark 2.13]). Dans tous ces cas, lorsque k est un corps de nombres, le théorème A et ses conséquences, ou le théorème C, impliquent la validité de la conjecture de Bass: le groupe $K_0(X)$ est de type fini.

Nous obtenons aussi des résultats de finitude sur un corps de base local (Théorèmes 6.1 et 6.3):

Théorème D *Soit k un corps p-adique et X une surface projective, lisse et géométriquement connexe sur le corps k, satisfaisant $H^2(X, \mathcal{O}_X)=0$. Alors:*
(i) *Le groupe $\mathrm{CH}^2(X)_{\mathrm{tors}}$ est fini.*
(ii) *Si la surface X a bonne réduction Y, avec $H^2(Y, \mathcal{O}_Y)=0$, la torsion première à p de $\mathrm{CH}^2(X)$ s'identifie à la torsion première à p de $\mathrm{CH}^2(Y)$.*

(Rappelons que la théorie du corps de classes non ramifié supérieur [K/S] décrit le groupe fini $\mathrm{CH}^2(Y)_{\mathrm{tors}}$.)

Des cas particuliers des résultats ci-dessus, tant dans le cas global que dans le cas local, ont déjà été établis par divers auteurs: Bloch [Bl4], les auteurs du présent article [CT, CT/R1, R2], Bloch et Srinivas [Bl/Sr], Gros [Gs], Ōkōchi [O], Coombes [Cb].

Les démonstrations de nos résultats combinent deux approches assez dissemblables, mais reposant toutes deux sur les résultats fondamentaux de Merkur'ev et Suslin et sur leurs développements subséquents (voir [R1]).

La première méthode utilisée est la méthode de localisation. On prend un modèle propre et lisse $\mathcal{X}$ de X au-dessus d'un ouvert du spectre de l'anneau des entiers du corps de nombres, et l'on compare cycles de codimension deux et de torsion sur $\mathcal{X}$ et sur X via la suite de localisation en K-théorie. Par cette méthode, on établit le théorème A (finitude de la n-torsion pour tout $n>0$, sous l'hypothèse $H^2(X, \mathcal{O}_X)=0$). Outre la suite de localisation, la démonstration utilise l'exactitude du complexe de Gersten pour les schémas lisses au-dessus d'un anneau de Dedekind (Bloch et al.) (§ 1) et l'analogue de la suite exacte de Moore en théorie du corps de classes supérieur (Bloch et al. [So])

(§ 2). L'hypothèse $H^2(X, \mathcal{O}_X)=0$, ou plutôt l'hypothèse analogue pour les spécialisations lisses de X, permet de relever les classes de faisceaux inversibles (théorème de Grothendieck [Gk1]).

La seconde méthode, que nous appelons méthode galoisienne, consiste à considérer tous les objets (cycles, diviseurs, groupes de K-cohomologie) au niveau de la clôture algébrique de k et à faire agir le groupe de Galois sur toute la situation. Si on veut la pousser à fond, elle implique un travail délicat avec la torsion du groupe de Néron-Severi. Mais, ici encore sous l'hypothèse $H^2(X, \mathcal{O}_X)=0$ (voir § 4, Rappel **R3**), cette méthode permet à moindres frais d'obtenir le Théorème B (la torsion du groupe de Chow $\mathrm{CH}^2(X)$ est d'exposant fini). Parmi les ingrédients ici utilisés, mentionnons le théorème de Deligne (conjectures de Weil) dans [CT/R1] et, de nouveau, la théorie du corps de classes supérieur par le biais de l'utilisation dans la Proposition 4.2 du résultat principal de [R3], résultat qui repose lui-même sur des travaux de Saito et de Jannsen.

Disons quelques mots sur l'origine des deux méthodes et leurs mérites respectifs.

Avec du recul, on voit que, lorsque k est un corps de fonctions d'une variable sur un corps fini et X/k une surface rationnelle ou une surface fibrée en coniques sur une courbe de genre quelconque, ce sont les derniers termes de la suite de localisation qui mènent, dans [CT/S/S, § 3], à un résultat de finitude. Récemment, l'un des auteurs [R2] a utilisé les termes supérieurs de cette suite pour étudier la torsion du groupe de Chow dans le cas de surfaces ayant bonne réduction sur un corps k p-adique. La remarque que cette approche vaut encore pour un corps de base global a été le point de départ du présent article.

La méthode galoisienne a été inaugurée par Bloch [Bl4] en 1981, et elle a été poursuivie par divers auteurs [CT, Bl/Sr, CT/S, CT/R1, Gs, O, R1]. C'est cette méthode, systématisée dans [CT/R1], qu'utilise Coombes [Cb], qui doit faire face aux difficultés présentées par la torsion dans le groupe de Néron-Severi; Salberger a une approche légèrement différente du travail de Coombes.

La méthode de localisation est plus qualitative que la méthode galoisienne. Elle a ainsi l'avantage de la légèreté. Mais la méthode purement galoisienne de Coombes pourrait permettre d'obtenir, comme c'est déjà le cas pour les surfaces rationnelles [CT/S, Sa], des résultats quantitatifs plus fins.

Lorsque le corps de base k est non un corps de nombres, mais un corps de fonctions d'une variable sur un corps fini de caractéristique $p>0$, on peut établir le théorème A, *à la p-torsion près*. Si l'on disposait de la résolution des singularités pour les variétés lisses sur un corps fini et si l'on contrôlait la p-torsion dans l'analogue du résultat de Somekawa (§ 2)), on pourrait utiliser le théorème 1 de [CT/S/S] pour établir tout le théorème A, i.e. en contrôlant aussi la p-torsion. En supposant de plus que X satisfait la condition «$B_2-\rho=0$», on peut établir le théorème B, *à la p-torsion près*.

Remerciements. K. Coombes nous a tenus informés de l'avancement de ses recherches [Cb]. C. Soulé a vu que le travail de Somekawa [So] pouvait nous être utile. Nous avons eu plusieurs échanges avec P. Salberger, qui nous a communiqué la Proposition 4.2. Le théorème 1.1 est dû à H. Gillet (non publié) et le lemme 6.4 à S. Bloch. Ce travail a été commencé lors d'un séjour au PN 48 (Sainte-Gauburge, Orne). Nous avons pu le continuer à la rencontre de K-théorie algébrique de Santa Margherita (Italie). Le premier auteur a bénéficié d'une invitation du Centre Culturel Européen de Delphes et de la splendeur de l'île de Santorin (Grèce). Pendant la mise au point de l'article, le second auteur séjournait à l'Université de Tokyo, grâce à une bourse de la Société Japonaise pour la Promotion de la Science (JSPS). Il remercie K. Kato pour son hospitalité et l'intérêt qu'il a pris à ce travail.

0 Notations et rappels

Ce sont en grande partie les mêmes que dans l'article [CT/R1].

Soit n un entier et M un groupe abélien. On note ${}_nM$ le sous-groupe de M formé des éléments annulés par n, et on note M_{tors} le sous-groupe de torsion de M. On note $M\{l\}$ le sous-groupe de torsion l-primaire. Lorsque M est un groupe de torsion, on note $M^0 \subset M$ le sous-groupe divisible maximal de M.

Etant donné un corps k, on note $\bar{k}$ une clôture séparable de k et $\mathfrak{g} = \mathrm{Gal}(\bar{k}/k)$ le groupe de Galois absolu. Si X/k est une variété algébrique définie sur k, et L/k une extension de corps, on note $X_L = X \times_k L$, et l'on note $\bar{X} = X \times_k \bar{k}$. Lorsque X est intègre, on note $k(X)$ son corps des fonctions, et lorsque X est géométriquement intègre, on note $\bar{k}(X)$ le corps des fonctions de $\bar{X}$. On note $X(k)$ l'ensemble des k-points rationnels d'une k-variété X.

On note $H^i(X, \mathcal{O}_X)$ les groupes de cohomologie, pour la topologie de Zariski, du faisceau structural d'un schéma X, et on note $\mathrm{Pic}(X) = H^1_{\mathrm{Zar}}(X, \mathcal{O}_X^*)$ le groupe de Picard de X.

Etant donné un schéma noethérien Z, on note $Z^{(i)}$ (resp. $Z_{(i)}$) l'ensemble des points de codimension i (resp. de dimension i) du schéma Z, et on note $\mathrm{CH}^i(Z)$ (resp. $\mathrm{CH}_i(Z)$) le groupe de Chow des cycles de codimension i (resp. de dimension i) sur Z modulo l'équivalence rationnelle [F]. Lorsque Z est le spectre d'un anneau noethérien A, on note $A^{(i)}$ l'ensemble des idéaux premiers de A de hauteur i.

Lorsque Z est une variété algébrique propre sur un corps k, l'application degré envoie $\mathrm{CH}_0(Z)$ dans le groupe des entiers naturels $\mathbb{Z}$; on note $A_0(Z)$ son noyau.

On désigne par $K_j(X)$, resp. $K'_j(X)$, le j-ième groupe de K-théorie de Quillen, pour la catégorie des $\mathcal{O}_X$-modules cohérents localement libres, resp. pour la catégorie des modules cohérents sur un schéma noethérien X. Les groupes $K_j(X)$ et $K'_j(X)$ coïncident lorsque X est un schéma régulier. Lorsque X est le spectre d'un anneau A, on note $K_j(A) = K_j(X)$. On désigne par $\mathcal{K}_j$ le faisceau pour la topologie de Zariski sur X associé au préfaisceau $U \to K_j(\Gamma(U, \mathcal{O}_U))$. Les groupes de cohomologie $H^i(X, \mathcal{K}_j)$ sont pris pour la topologie de Zariski sur X.

Pour tout schéma noethérien séparé X, on dispose de la suite spectrale BGQ (Brown, Gersten, Quillen):

$$E_1^{pq}(X) = \sum_{x \in X^p} K_{-p-q}(k(x)) \Rightarrow K'_{-p-q}(X)$$

($k(x)$ désigne le corps résiduel en x), suite spectrale qui est covariante fonctorielle par morphismes finis (Gillet [Gi1, 7.22]). Pour un schéma noethérien séparé quelconque X, on a un isomorphisme:

$$E_2^{p,-p}(X) \simeq \mathrm{CH}^p(X)$$

(Ceci s'établit en combinant un calcul de Sherman ([Sh1, Prop. 1.4]) et une proposition de Grayson ([Gn1, p. 73])).

Lorsque le complexe (de faisceaux) de Gersten-Quillen associé au faisceau $\mathcal{K}_q$ sur un schéma X est exact, on a en outre des isomorphismes:

$$E_2^{p,-q}(X) \simeq H^p(X, \mathcal{K}_q).$$

Il en est toujours ainsi lorsque X est une variété lisse sur un corps (Quillen [Q]).

Etant donné un entier $n>0$ inversible sur un schéma X, on note μ_n le faisceau pour la topologie étale sur X défini par le groupe (fini étale) des racines n-ièmes de l'unité, et pour j entier naturel, on définit le faisceau (étale) $\mu_n^{\otimes j}=\mu_n\otimes\ldots\otimes\mu_n$ (j fois). On note $H^i_{\text{ét}}(X,\mu_n^{\otimes j})$, et parfois simplement $H^i(X,\mu_n^{\otimes j})$ les groupes de cohomologie étale de X à valeurs dans $\mu_n^{\otimes j}$. On note $\mathscr{H}^i(\mu_n^{\otimes j})$ le faisceau pour la topologie de Zariski sur X associé au préfaisceau qui à un ouvert U associe le groupe $H^i_{\text{ét}}(U,\mu_n^{\otimes j})$.

On note $H^i_{\text{ét}}(X,\mathbb{Q}/\mathbb{Z}(j))$, resp. $H^i_{\text{ét}}(X,\mathbb{Q}_l/\mathbb{Z}_l(j))$, la limite inductive des groupes $H^i_{\text{ét}}(X,\mu_n^{\otimes j})$ lorsque n varie parmi les entiers, resp. parmi les puissances d'un nombre premier l.

On note $\mathbb{G}_m$ le groupe multiplicatif sur un schéma X, et le faisceau pour la topologie étale sur X qu'il définit. On a $\mathrm{Pic}(X)\simeq H^1_{\text{ét}}(X,\mathbb{G}_m)$. Le groupe $H^2_{\text{ét}}(X,\mathbb{G}_m)$ est le groupe de Brauer $\mathrm{Br}(X)$ du schéma X.

1 Un théorème de finitude pour les modèles entiers

Le théorème suivant a été annoncé par H. Gillet en 1985.

Théorème 1.1 *Soit* $\mathrm{Spec}(A)$ *un ouvert non vide du spectre de l'anneau des entiers d'un corps de nombres, et soit* $\mathscr{X}$ *un* A*-schéma projectif et lisse. Soit* n *un entier inversible dans* A*. Alors le sous-groupe de* n*-torsion* ${}_n\mathrm{CH}^2(\mathscr{X})$ *est fini.*

Démonstration.

(a) On dispose du complexe de Bloch-Ogus/Gersten ([K, § 1])

$$\bigoplus_{x\in\mathscr{X}^{(0)}} i_{x*}H^2(\kappa(x),\mu_n^{\otimes 2})\to\bigoplus_{x\in\mathscr{X}^{(1)}} i_{x*}H^1(\kappa(x),\mu_n)\to\bigoplus_{x\in\mathscr{X}^{(2)}} i_{x*}H^0(\kappa(x),\mathbb{Z}/n)$$

complexe que nous noterons $\mathbf{H}^2_{\mathbf{n}}$, les degrés étant donnés par la codimension.

On dispose par ailleurs d'un complexe analogue (Gersten), soit $\mathbf{K_2}$, pour la K-théorie elle-même:

$$\bigoplus_{x\in\mathscr{X}^{(0)}} i_{x*}K_2(\kappa(x))\to\bigoplus_{x\in\mathscr{X}^{(1)}} i_{x*}K_1(\kappa(x))\to\bigoplus_{x\in\mathscr{X}^{(2)}} i_{x*}K_0(\kappa(x))$$

ainsi que du complexe qu'on en déduit pour la K-théorie modulo n, complexe que nous noterons $\mathbf{K_2/n}$:

$$\bigoplus_{x\in\mathscr{X}^{(0)}} i_{x*}K_2(\kappa(x))/n\to\bigoplus_{x\in\mathscr{X}^{(1)}} i_{x*}K_1(\kappa(x))/n\to\bigoplus_{x\in\mathscr{X}^{(2)}} i_{x*}K_0(\kappa(x))/n.$$

Enfin, il y a un morphisme naturel de complexes $\mathbf{K_2/n}\to\mathbf{H^2_n}$ ([Bl3, Theorem 2.3]; [K, Lemma 1.4]).

Un argument formel, dû à Bloch ([Bl3, Cor. 1.11]; [CT/S/S, § 1, p. 767]), permet de déduire du théorème de Merkur'ev-Suslin une surjection:

$$H^1_{\text{Zar}}(\mathscr{X},\mathbf{H^2_n})\twoheadrightarrow{}_n\mathrm{CH}^2(\mathscr{X}).$$

(b) Soit $\mathscr{H}^2(\mu_n^{\otimes 2})$ le faisceau (Zariski) sur $\mathscr{X}$ associé au préfaisceau $U \to H^2_{\text{ét}}(U, \mu_n^{\otimes 2})$. On dispose d'un carré commutatif de morphismes de complexes de faisceaux (les faisceaux $\mathscr{K}_2/n$ et $\mathscr{H}^2(\mu_n^{\otimes 2})$ étant identifiés à des complexes concentrés en degré zéro):

$$\begin{array}{ccc} \mathscr{K}_2/n & \longrightarrow & \mathbf{K_2/n} \\ \downarrow & & \downarrow \\ \mathscr{H}^2(\mu_n^{\otimes 2}) & \longrightarrow & \mathbf{H^2_n} \end{array}$$

([Bl 3, Theorem 2.3]; [K, Lemma 1.4]).

Nous nous proposons de montrer que la flèche

$$\mathscr{H}^2(\mu_n^{\otimes 2}) \to \mathbf{H^2_n}$$

définit un quasi-isomorphisme de complexes de faisceaux.

Il résulte de travaux de Bloch [Bl6], Grayson [Gn2], Gillet et Levine [Gi/Le] et Gillet [Gi2] (voir aussi [Sh3]) sur les schémas lisses sur un anneau de valuation discrète que la flèche $\mathscr{K}_2 \to \mathbf{K_2}$ est un quasi-isomorphisme de complexes de faisceaux, et qu'il en est de même de la flèche $\mathscr{K}_2/n \to \mathbf{K_2/n}$ (voir [R2, § 1 et Prop. 1.1]).

Soit R un anneau local de $\mathscr{X}$, et K son corps des fractions. On a donc un diagramme commutatif de complexes:

$$\begin{array}{ccccccccccc} 0 \to & K_2(R)/n & \to & K_2(K)/n & \to & \bigoplus_{x \in R^{(1)}} K_1(\kappa(x))/n & \to & \bigoplus_{x \in R^{(2)}} K_0(\kappa(x))/n & \to 0 \\ & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\ 0 \to & H^2_{\text{ét}}(R, \mu_n^{\otimes 2}) & \to & H^2_{\text{ét}}(K, \mu_n^{\otimes 2}) & \to & \bigoplus_{x \in R^{(1)}} H^1(\kappa(x), \mu_n) & \to & \bigoplus_{x \in R^{(2)}} H^0(\kappa(x), \mathbb{Z}/n) & \to 0. \end{array}$$

Le complexe supérieur est exact. Pour établir que la flèche $\mathscr{H}^2(\mu_n^{\otimes 2}) \to \mathbf{H^2_n}$ est un quasi-isomorphisme, il suffit d'établir (pour tout R) que les flèches verticales sont des isomorphismes. Sauf pour la flèche $K_2(R)/n \to H^2_{\text{ét}}(R, \mu_n^{\otimes 2})$, cela résulte du théorème 90 de Hilbert et du théorème de Merkur'ev et Suslin. Comme le premier complexe est exact, il en résulte que la flèche $K_2(R)/n \to H^2_{\text{ét}}(R, \mu_n^{\otimes 2})$ est injective, et que le deuxième complexe est exact sauf peut-être au terme $H^2_{\text{ét}}(R, \mu_n^{\otimes 2})$. Il est exact si et seulement si la flèche $K_2(R)/n \to H^2_{\text{ét}}(R, \mu_n^{\otimes 2})$ est surjective, ou encore si et seulement si la flèche $H^2_{\text{ét}}(R, \mu_n^{\otimes 2}) \to H^2_{\text{ét}}(K, \mu_n^{\otimes 2})$ est injective.

Supposons d'abord n premier, et soit $S = R[T]/\Phi_n(T)$, avec $\Phi_n(T) = (T^n - 1)/(T - 1)$. C'est là un anneau semi-local régulier, et l'on sait ([Gk3, Corollaire 1.8]) qu'on a l'injection de groupes de Brauer $H^2_{\text{ét}}(S, \mathbb{G}_m) \hookrightarrow H^2_{\text{ét}}(L, \mathbb{G}_m)$, où $L = S \otimes_R K$. La suite de Kummer et la nullité de $H^1_{\text{ét}}(S, \mathbb{G}_m) = \text{Pic}(S)$ donnent alors l'injection $H^2_{\text{ét}}(S, \mu_n) \hookrightarrow H^2_{\text{ét}}(L, \mu_n)$. Comme n est premier et donc μ_n déployé sur S (i.e. isomorphe à $\mathbb{Z}/n$), on en déduit aussi l'injection $H^2_{\text{ét}}(S, \mu_n^{\otimes 2}) \hookrightarrow H^2_{\text{ét}}(L, \mu_n^{\otimes 2})$. Le groupe $H^2_{\text{ét}}(R, \mu_n^{\otimes 2})$ étant annulé par l'entier n, premier au degré $(n-1)$ du revêtement fini S/R, un argument de transfert permet alors de conclure à l'injectivité de $H^2_{\text{ét}}(R, \mu_n^{\otimes 2}) \to H^2_{\text{ét}}(K, \mu_n^{\otimes 2})$. Pour n premier,

l'application $K_2(R)/n \to H^2_{\text{ét}}(R, \mu_n^{\otimes 2})$ est donc surjective. Supposons alors $n = rs$ avec $r > 1$ et $s > 1$. On dispose d'une suite exacte

$$0 \to \mu_r^{\otimes 2} \to \mu_n^{\otimes 2} \to \mu_s^{\otimes 2} \to 0$$

(la deuxième flèche étant l'élévation à la puissance r) qui induit la seconde ligne du diagramme commutatif de suites exactes:

$$\begin{array}{ccccc} K_2(R)/r & \longrightarrow & K_2(R)/n & \longrightarrow & K_2(R)/s \\ \downarrow & & \downarrow & & \downarrow \\ H^2_{\text{ét}}(R, \mu_r^{\otimes 2}) & \longrightarrow & H^2_{\text{ét}}(R, \mu_n^{\otimes 2}) & \longrightarrow & H^2_{\text{ét}}(R, \mu_s^{\otimes 2}). \end{array}$$

Dans la première ligne, la première flèche est la multiplication par s et la seconde la projection naturelle, qui est évidemment surjective. Si donc les flèches verticales extrêmes sont surjectives, il en est de même de la flèche centrale. Une récurrence évidente achève d'établir l'isomorphisme $K_2(R)/n \xrightarrow{\cong} H^2_{\text{ét}}(R, \mu_n^{\otimes 2})$ pour n quelconque, et donc aussi le quasi-isomorphisme de complexes de faisceaux (Zariski)

$$\mathscr{H}^2(\mu_n^{\otimes 2}) \to \mathbf{H}^2_{\mathbf{n}}.$$

(c) Comme sur tout schéma où n est inversible, on a la suite spectrale associée au morphisme de sites $\mathscr{X}_{\text{ét}} \to \mathscr{X}_{\text{Zar}}$:

$$E_2^{pq} = H^p_{\text{Zar}}(\mathscr{X}, \mathscr{H}^q(\mu_n^{\otimes 2})) \Rightarrow H^n_{\text{ét}}(\mathscr{X}, \mu_n^{\otimes 2}).$$

Un argument plus simple que celui donné ci-dessus, et valable pour tout schéma régulier, assure que les complexes de Gersten de $\mathscr{H}^q(\mu_n^{\otimes 2})$ pour $q = 0$ et 1 sont exacts, et donc que ces faisceaux admettent des résolutions flasques de longueur 1, resp. 2. Il en résulte que les groupes E_2^{pq} sont nuls pour $q = 0$ et $p > 0$ et pour $q = 1$ et $p > 1$. En particulier $H^1_{\text{Zar}}(\mathscr{X}, \mathscr{H}^2(\mu_n^{\otimes 2}))$ est un sous-groupe du groupe $H^3_{\text{ét}}(\mathscr{X}, \mu_n^{\otimes 2})$.

(d) Soit $\pi: \mathscr{X} \to S = \text{Spec}(A)$ le morphisme structural. On dispose de la suite spectrale $E_2^{pq} = H^p_{\text{ét}}(S, R^q\pi_*(\mu_n^{\otimes 2})) \Rightarrow H^m_{\text{ét}}(\mathscr{X}, \mu_n^{\otimes 2})$. Comme π est propre et lisse, et n inversible, les faisceaux $R^q\pi_*(\mu_n^{\otimes 2})$ sont constructibles ([Mi1, VI.4.2]) et annulés par n, et donc tous les groupes E_2^{pq} sont finis ([Mi2, Chapter II, Theorem 2.13, p. 213]). Il en est donc ainsi aussi des groupes $H^m_{\text{ét}}(\mathscr{X}, \mu_n^{\otimes 2})$.

(e) Conclusion. En résumé, d'après *(a)*, le groupe ${}_n\text{CH}^2(\mathscr{X})$ est un quotient de $H^1_{\text{Zar}}(\mathscr{X}, \mathbf{H}^2_{\mathbf{n}})$, groupe lui-même isomorphe à $H^1_{\text{Zar}}(\mathscr{X}, \mathscr{H}^2(\mu_n^{\otimes 2}))$ d'après *(b)*. D'après *(c)*, ce dernier groupe est un sous-groupe du groupe fini $H^3_{\text{ét}}(\mathscr{X}, \mu_n^{\otimes 2})$, groupe fini d'après *(d)*. Le groupe ${}_n\text{CH}^2(\mathscr{X})$ est donc fini. □

Remarque 1.2 Gillet sait en fait démontrer (directement, i.e. sans passer par la K-théorie) l'exactitude des complexes de Gersten pour tous les faisceaux $\mathscr{H}^i(\mu_n^{\otimes j})$ (i, j entiers arbitraires) sur les schémas lisses sur un anneau de valuation discrète.

Remarque 1.3 En utilisant la suite de localisation et les théorèmes de finitude classique pour le groupe de Chow de codimension 1 pour une variété lisse

sur un corps fini, on voit que le théorème 1.1 vaut encore même lorsque $n \neq 0$ n'est pas inversible dans A. Grâce à cette remarque, on pourrait éviter la réduction au cas «r inversible» dans la démonstration du théorème 3.4 ci-dessous.

2 Une conséquence de la suite exacte de Bloch-Kato-Saito-Somekawa

Soit A l'anneau des entiers d'un corps p-adique k (extension finie de $\mathbb{Q}_p$), et soit J un A-schéma abélien. Soit $\mathbb{F}$ le corps résiduel de A. On définit alors un homomorphisme

$$\rho_k : J(k) \underset{\mathbb{Z}}{\otimes} k^* \to J(\mathbb{F})$$

de la façon suivante. Comme on a l'identification $J(k) = J(A)$, on dispose de la flèche de réduction $J(A) \to J(\mathbb{F}), x \to \bar{x}$. On envoie alors $x \otimes \lambda \in J(k) \underset{\mathbb{Z}}{\otimes} k^* = J(A) \underset{\mathbb{Z}}{\otimes} k^*$ sur $v(\lambda) \cdot \bar{x}$, où v désigne la valuation normalisée de k. Plus généralement, si A, k et J sont comme ci-dessus, et si L est un corps extension finie de k, de corps résiduel $\mathbb{F}_L$, on définit l'application

$$\rho_{L/k} : J(L) \underset{\mathbb{Z}}{\otimes} L^* \to J(\mathbb{F})$$

comme la composée de l'application $\rho_L : J(L) \underset{\mathbb{Z}}{\otimes} L^* \to J(\mathbb{F}_L)$ et de la norme $N_{\mathbb{F}_L/\mathbb{F}} : J(\mathbb{F}_L) \to J(\mathbb{F})$.

Soit B l'anneau des entiers d'un corps de nombres k, soit $f \in B$, $f \neq 0$, et soit $A = B_f$. Soit J un A-schéma abélien. Pour tout idéal premier m non nul de A, de corps résiduel $\mathbb{F}_m$ et toute extension finie L/k de corps, on dispose alors d'une application

$$\rho_{L/k,m} : J(L) \underset{\mathbb{Z}}{\otimes} L^* \to J(\mathbb{F}_m)$$

obtenue par composition:

$$J(L) \underset{\mathbb{Z}}{\otimes} L^* \to \bigoplus_{M|m} (J(L_M) \underset{\mathbb{Z}}{\otimes} L_M^*) \xrightarrow{\oplus\{\rho_{L_M/k_m}\}} J(\mathbb{F}_m)$$

(ici k_m est le complété de k en m, et les L_M sont les diverses complétions de L en les places M au-dessus de m).

Le théorème suivant est une conséquence immédiate du Théorème 4.1 de Somekawa [So]. Il s'agit là d'une suite de réciprocité à la Moore, que Somekawa établit en suivant les travaux fondamentaux de Bloch [Bl5] et de Kato et Saito [K/S].

Théorème 2.1. *Soit B l'anneau des entiers d'un corps de nombres k, soit $f \in B$, $f \neq 0$, et soit $A = B_f$. Soit J un A-schéma abélien. L'application*

$$\bigoplus_{L/k} \bigoplus_{m} \rho_{L/k,m} : \bigoplus_{L/k} (J(L) \underset{\mathbb{Z}}{\otimes} L^*) \to \bigoplus_{m} J(\mathbb{F}_m),$$

où L/k parcourt les extensions finies de k – une par classe d'isomorphie – et m parcourt les idéaux premiers non nuls de A, a un conoyau fini.

(Bloch [Bl5] et Kato et Saito [K/S] traitent le cas où J_k est la jacobienne d'une courbe projective et lisse sur k possédant un point rationnel.)

3 Démonstration du théorème A

Commençons par quelques rappels de résultats de Grothendieck [Gk2]. Soit S un schéma noethérien et $\mathscr{X}/S$ un S-schéma projectif et lisse à fibres géométriques intègres. Notons $\mathscr{X}_m/\kappa(m)$ la fibre de $\mathscr{X}/S$ en un point $m \in S$, de corps résiduel $\kappa(m)$. Faisons les deux hypothèses suivantes:

(i) La projection $\mathscr{X} \to S$ admet une section, i.e. $\mathscr{X}(S) \neq \emptyset$.

(ii) $H^2(\mathscr{X}_m, \mathscr{O}_{\mathscr{X}_m}) = 0$ pour tout $m \in S$.

Alors le foncteur de Picard $\underline{\mathrm{Pic}}_{\mathscr{X}/S}$ qui à tout S-schéma T associe le groupe $\mathrm{Pic}(\mathscr{X} \times_S T)/\mathrm{Pic}(T)$ est représentable par un S-schéma en groupes localement lisse. La composante de la section unité, soit $\underline{\mathrm{Pic}}^0_{\mathscr{X}/S}$, est un S-schéma abélien. Pour tout S-schéma T, on a:

$$\underline{\mathrm{Pic}}_{\mathscr{X}/S} \times_S T \simeq \underline{\mathrm{Pic}}_{\mathscr{X}_T/T}$$

et

$$\underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/S} \times_S T \simeq \underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}_T/T}.$$

Pour tout S-schéma T, l'image inverse définit alors un diagramme commutatif:

$$\begin{array}{ccccc} \underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/S}(S) & \longrightarrow & \underline{\mathrm{Pic}}_{\mathscr{X}/S}(S) & = & \mathrm{Pic}(\mathscr{X})/\mathrm{Pic}(S) \\ \downarrow & & \downarrow & & \downarrow \\ \underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}_T/T}(T) & \longrightarrow & \underline{\mathrm{Pic}}_{\mathscr{X}_T/T}(T) & = & \mathrm{Pic}(\mathscr{X}_T)/\mathrm{Pic}(T). \end{array}$$

Pour $\mathscr{X}/S$ comme ci-dessus, nous *désignerons* dans cet article par $\mathrm{NS}(\mathscr{X})$ le conoyau de l'application $\underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/S}(S) \to \underline{\mathrm{Pic}}_{\mathscr{X}/S}(S) = \mathrm{Pic}(\mathscr{X})/\mathrm{Pic}(S)$. Le diagramme ci-dessus induit un homomorphisme $\mathrm{NS}(\mathscr{X}) \to \mathrm{NS}(\mathscr{X}_T)$.

Lorsque $S = \mathrm{Spec}(A)$ est le spectre d'un anneau de valuation discrète de corps des fractions k, dans le diagramme de suites exactes:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/S}(A) & \longrightarrow & \mathrm{Pic}(\mathscr{X}) & \longrightarrow & \mathrm{NS}(\mathscr{X}) & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}_k/k}(k) & \longrightarrow & \mathrm{Pic}(\mathscr{X}_k) & \longrightarrow & \mathrm{NS}(\mathscr{X}_k) & \longrightarrow & 0 \end{array}$$

(où l'on utilise les identifications $\underline{\mathrm{Pic}}_{\mathscr{X}/S}(A) = \mathrm{Pic}(\mathscr{X})/\mathrm{Pic}(A) = \mathrm{Pic}(\mathscr{X})$ et $\underline{\mathrm{Pic}}_{\mathscr{X}_k/k}(k) = \underline{\mathrm{Pic}}_{\mathscr{X}_k/k}(k)/\mathrm{Pic}(k) = \mathrm{Pic}(\mathscr{X}_k)$, qui reposent sur l'hypothèse $\mathscr{X}(S) \neq \emptyset$), la flèche verticale de gauche est un isomorphisme car $\underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/S}$ est un S-schéma propre, et la flèche verticale médiane est un isomorphisme ($\mathscr{X}$ est régulier, et le diviseur $\mathscr{X}_m$ est clairement défini par une seule équation). Ainsi l'application $\mathrm{NS}(\mathscr{X}) \to \mathrm{NS}(\mathscr{X}_k)$ est un isomorphisme.

Lorsque $S=\mathrm{Spec}(k)$ est le spectre d'un corps k, de clôture séparable $\bar{k}$, notant $\mathfrak{g}=\mathrm{Gal}(\bar{k}/k)$ et $\bar{\mathscr{X}}=\mathscr{X}\times_k \bar{k}$, les identifications $\underline{\mathrm{Pic}}_{\mathscr{X}/S}(k)=\mathrm{Pic}(\mathscr{X})$ (qui provient de l'hypothèse $\mathscr{X}(S)\neq\emptyset$) et $\underline{\mathrm{Pic}}_{\mathscr{X}/S}(\bar{k})=\mathrm{Pic}(\bar{\mathscr{X}})$ permettent d'écrire le diagramme commutatif de suites exactes:

$$\begin{array}{ccccccc} 0 & \longrightarrow & \underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/S}(k) & \longrightarrow & \mathrm{Pic}(\mathscr{X}) & \longrightarrow & \mathrm{NS}(\mathscr{X}) \longrightarrow 0 \\ & & \downarrow & & \downarrow & & \downarrow \\ 0 & \longrightarrow & \underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/S}(\bar{k})^{\mathfrak{g}} & \longrightarrow & \mathrm{Pic}(\bar{\mathscr{X}})^{\mathfrak{g}} & \longrightarrow & \mathrm{NS}(\bar{\mathscr{X}})^{\mathfrak{g}} \end{array} .$$

Les deux flèches verticales de gauche sont des isomorphismes. Ainsi l'application naturelle $\mathrm{NS}(\mathscr{X})\to\mathrm{NS}(\bar{\mathscr{X}})$ est alors une injection.

Lemme 3.1 *Soit A un anneau de valuation discrète, de corps des fractions k de caractéristique zéro, d'idéal maximal m et de corps résiduel κ, et soit $\bar{k}$ une clôture algébrique de k. Soit $\mathscr{X}/A$ un schéma projectif et lisse à fibres géométriques intègres. Notons $X=\mathscr{X}\times_A k$, puis $\bar{X}=X\times_k \bar{k}$, enfin $\mathscr{X}_m/\kappa$ la fibre au-dessus de m. Supposons*:

(i) $H^2(\mathscr{X}_m, \mathcal{O}_{\mathscr{X}_m})=0$.
(ii) *L'application composée* $\mathrm{Pic}(X)\to\mathrm{Pic}(\bar{X})\to\mathrm{NS}(\bar{X})$ *est surjective.*
(iii) $X(A)=X(k)\neq\emptyset$.

Alors l'application composée $\mathrm{Pic}(\mathscr{X})\to\mathrm{Pic}(\mathscr{X}_m)\to\mathrm{NS}(\mathscr{X}_m)$ *est surjective.*

Démonstration. Soit $\hat{A}$ le complété de A, et $\hat{k}$ son corps des fractions. Soit $\bar{\hat{k}}$ une clôture algébrique de $\hat{k}$ contenant $\bar{k}$. Compte tenu des rappels ci-dessus, on dispose du diagramme commutatif suivant:

$$\begin{array}{ccccc} \mathrm{Pic}(\mathscr{X}) & \longrightarrow & \mathrm{Pic}(\mathscr{X}_{\hat{A}}) & \longrightarrow & \mathrm{Pic}(\mathscr{X}_m) \\ \downarrow & & \downarrow & & \downarrow \\ & & \mathrm{NS}(\mathscr{X}_{\hat{A}}) & \longrightarrow & \mathrm{NS}(\mathscr{X}_m) \\ & & \downarrow\simeq & & \downarrow= \\ \mathrm{Pic}(X) & \longrightarrow & \mathrm{NS}(\mathscr{X}_{\bar{\hat{k}}}) & \longrightarrow & \mathrm{NS}(\mathscr{X}_m) \\ \downarrow & & \downarrow & & \\ \mathrm{NS}(\bar{X}) & \longrightarrow & \mathrm{NS}(\mathscr{X}_{\bar{\hat{k}}}). & & \end{array}$$

Par hypothèse, la flèche $\mathrm{Pic}(\mathscr{X})\to\mathrm{NS}(\bar{X})$ est surjective. La flèche $\mathrm{NS}(\bar{X})\to\mathrm{NS}(\mathscr{X}_{\bar{\hat{k}}})$ est un isomorphisme (le groupe de Néron-Severi ne change pas par extensions de corps de base algébriquement clos). Par ailleurs, la flèche $\mathrm{NS}(\mathscr{X}_{\hat{k}})\to\mathrm{NS}(\mathscr{X}_{\bar{\hat{k}}})$ est injective. L'hypothèse (ii) assure alors que la flèche $\mathrm{Pic}(X)\to\mathrm{NS}(X_{\hat{k}})$ est surjective, donc aussi la flèche $\mathrm{Pic}(\mathscr{X})\to\mathrm{NS}(\mathscr{X}_{\hat{A}})$. Mais d'après un théorème de Grothendieck ([Gk1, Corollaire 1 de la Proposition 3, p. 11]), l'hypothèse (i) assure que l'application $\mathrm{Pic}(\mathscr{X}_{\hat{A}})\to\mathrm{Pic}(\mathscr{X}_m)$ est surjective, ce qui achève d'établir le lemme. □

Soit A un anneau de Dedekind, et soit k son corps des fractions. Soit $\mathscr{X}/A$ un schéma projectif et lisse à fibres géométriques intègres, et soit $X=\mathscr{X}\times_A k$. Soit $\mathscr{X}_m/\kappa(m)$ la fibre au-dessus de $m\in\mathrm{Spec}(A)$. En combinant le théorème 2.1

de [Sh2], la proposition 1.4 de [Sh1] (Sherman) et le résultat de Grayson [Gn1] rappelé au §0, on obtient la suite exacte de localisation (cf. [R2, Prop. 1.2]):

$$(\mathscr{L}) \qquad H^1(X, \mathscr{K}_2) \xrightarrow{\{\partial_m\}} \bigoplus_{m \in A^{(1)}} \mathrm{Pic}(\mathscr{X}_m) \to \mathrm{CH}^2(\mathscr{X}) \to \mathrm{CH}^2(X) \to 0.$$

On dispose par ailleurs du cup-produit $\mathrm{Pic}(X) \underset{\mathbb{Z}}{\otimes} k^* \to H^1(X, \mathscr{K}_2)$, qu'on peut définir de façon simple à partir de la présentation de $H^1(X, \mathscr{K}_2)$ comme homologie du complexe de Gersten sur X.

Lemme 3.2 *Soit A un anneau principal, k son corps des fractions supposé de caractéristique nulle, $\bar{k}$ une clôture algébrique de k. Soit $\mathscr{X}/A$ un schéma projectif et lisse à fibres géométriques intègres. Notons $X = \mathscr{X} \times_A k$ et $\bar{X} = X \times_k \bar{k}$, puis $\mathscr{X}_m/\kappa(m)$ la fibre au-dessus de $m \in \mathrm{Spec}(A)$. Supposons*:

(i) $H^2(\mathscr{X}_m, \mathscr{O}_{\mathscr{X}_m}) = 0$ *pour tout* $m \in \mathrm{Spec}(A)$.

(ii) *L'application composée* $\mathrm{Pic}(X) \to \mathrm{Pic}(\bar{X}) \to \mathrm{NS}(\bar{X})$ *est surjective.*

(iii) $\mathscr{X}(A) = X(k) \neq \emptyset$.

Alors l'application composée

$$\mathrm{Pic}(\mathscr{X}) \underset{\mathbb{Z}}{\otimes} k^* \to \mathrm{Pic}(X) \underset{\mathbb{Z}}{\otimes} k^* \to H^1(X, \mathscr{K}_2) \xrightarrow{\{\partial_m\}} \bigoplus_{m \in A^{(1)}} \mathrm{Pic}(\mathscr{X}_m) \to \bigoplus_{m \in A^{(1)}} \mathrm{NS}(\mathscr{X}_m)$$

est surjective.

Démonstration. Soit m un idéal premier non nul de A, et soit A_m le localisé de A en m. Comme $\mathscr{X}$ est régulier, l'application $\mathrm{Pic}(\mathscr{X}) \to \mathrm{Pic}(\mathscr{X}_{A_m})$ est surjective. Il résulte alors du lemme 3.1 et des hypothèses que pour tout $m \in A^{(1)}$, l'application composée $\mathrm{Pic}(\mathscr{X}) \to \mathrm{Pic}(\mathscr{X}_m) \to \mathrm{NS}(\mathscr{X}_m)$ est surjective.

Pour tout élément $\{\lambda_m\} \in \bigoplus_{m \in A^{(1)}} \mathrm{NS}(\mathscr{X}_m)$, on peut donc trouver une famille $\{L^m\}_{m \in A^{(1)}}$ d'éléments de $\mathrm{Pic}(X)$, avec $L^m = 0$ sauf pour m dans un ensemble fini S d'idéaux premiers de hauteur 1 de A, et telle que, pour tout m, L^m induise λ_m dans $\mathrm{NS}(\mathscr{X}_m)$.

Soit v_m la valuation $k^* \to \mathbb{Z}$ associée à m. Soit $L \in \mathrm{Pic}(\mathscr{X})$ et L_m l'image de L par la flèche de restriction $\mathrm{Pic}(\mathscr{X}) \to \mathrm{Pic}(\mathscr{X}_m)$. En suivant la construction de la suite exacte de localisation $(\mathscr{L})$, on vérifie que l'application composée

$$\mathrm{Pic}(\mathscr{X}) \underset{\mathbb{Z}}{\otimes} k^* \to \mathrm{Pic}(X) \underset{\mathbb{Z}}{\otimes} k^* \to H^1(X, \mathscr{K}_2) \xrightarrow{\partial_m} \mathrm{Pic}(\mathscr{X}_m)$$

envoie l'élément $L \otimes \alpha$ sur l'élément $L_m^{\otimes v_m(\alpha)}$ (cf. [R2, Prop. 1.3]).

Comme A est principal, on peut, pour chaque $m \in S$, trouver un élément π_m qui engendre l'idéal m. L'élément $\sum_{m \in S} L^m \otimes \pi_m$ a alors pour image $\{\lambda_m\} \in \bigoplus_{m \in A^{(1)}} \mathrm{NS}(\mathscr{X}_m)$ par l'application composée apparaissant dans l'énoncé. □

Lemme 3.3 *Soit k un corps de nombres, et $A=B_f$ un localisé de l'anneau B des entiers de k. Supposons A principal. Soit $\mathscr{X}/A$ un schéma projectif et lisse à fibres géométriques intègres. Notons $X=\mathscr{X}\times_A k$ et $\bar{X}=X\times_k \bar{k}$, puis $\mathscr{X}_m/\kappa(m)$ la fibre au-dessus de $m\in\mathrm{Spec}(A)$. Supposons*:

(i) $H^2(\mathscr{X}_m, \mathscr{O}_{\mathscr{X}_m})=0$ *pour tout* $m\in\mathrm{Spec}(A)$.

(ii) *L'application composée* $\mathrm{Pic}(X)\to\mathrm{Pic}(\bar{X})\to\mathrm{NS}(\bar{X})$ *est surjective.*

(iii) $\mathscr{X}(A)=X(k)\neq\emptyset$.

Alors l'application

$$H^1(X,\mathscr{K}_2)\xrightarrow{\{\partial_m\}}\bigoplus_{m\in A^{(1)}}\mathrm{Pic}(\mathscr{X}_m)$$

dans la suite de localisation ($\mathscr{L}$) précédant le Lemme 3.2 *a un conoyau fini; ainsi l'application de restriction* $\mathrm{CH}^2(\mathscr{X})\to\mathrm{CH}^2(X)$ *a un noyau fini.*

Démonstration. Soit J le A-schéma abélien $\underline{\mathrm{Pic}}^{\circ}_{\mathscr{X}/A}$. On dispose du diagramme:

$$\begin{array}{ccc}
 & & \bigoplus_{m\in A^{(1)}}\mathrm{NS}(\mathscr{X}_m)\\
 & \nearrow & \uparrow\\
H^1(X,\mathscr{K}_2) & \xrightarrow{\{\partial_m\}} & \bigoplus_{m\in A^{(1)}}\mathrm{Pic}(\mathscr{X}_m)\\
\uparrow & & \uparrow\\
\bigoplus_{L/k}(J(L)\otimes_{\mathbb{Z}} L^*) & \xrightarrow{\oplus_{L/k}\oplus_m\rho_{L/k,m}} & \bigoplus_{m\in A^{(1)}}J(\mathbb{F}_m),
\end{array}$$

où L parcourt les extensions finies de k (une par classe d'isomorphie), les flèches $\rho_{L/k,m}$ ont été définies au § 2 et les applications $J(L)\otimes_{\mathbb{Z}} L^*\to H^1(X,\mathscr{K}_2)$ sont définies comme les applications composées:

$$J(L)\otimes_{\mathbb{Z}} L^*\to\mathrm{Pic}(X_L)\otimes_{\mathbb{Z}} L^*\to H^1(X_L,\mathscr{K}_2)\xrightarrow{N_{L/k}}H^1(X,\mathscr{K}_2).$$

Dans ce diagramme, le triangle supérieur est commutatif par définition. En utilisant la fonctorialité covariante de la suite spectrale de Brown-Gersten-Quillen par morphismes finis ([Gi1, Thm. 7.22]), qui donne celle de la suite de localisation ($\mathscr{L}$) (cf. [Sh2, Thm. 2.1 et Cor. 2.3]), on vérifie que le carré dans le diagramme est commutatif. La suite verticale de droite est exacte (par définition).

La flèche oblique est surjective (Lemme 3.2) et la flèche inférieure a un conoyau fini d'après Somekawa (Théorème 2.1). Ceci suffit à établir le lemme 3.3. □

Nous pouvons maintenant démontrer le théorème A de l'introduction:

Théorème 3.4 *Soit X une variété projective, lisse et géométriquement connexe définie sur un corps de nombres k. Supposons le groupe $H^2(X, \mathcal{O}_X)$ nul. Alors, pour tout entier $n>0$, le sous-groupe de n-torsion du groupe de Chow $\mathrm{CH}^2(X)$ des cycles de codimension deux est fini.*

Démonstration. Choisissons un ouvert non vide $\mathrm{Spec}(A)$ de l'anneau des entiers de k sur lequel la k-variété projective, lisse et géométriquement intègre X s'étend en un A-schéma projectif et lisse $\mathscr{X}$ à fibres géométriques intègres.

Comme la fibre générique $X=\mathscr{X}\times_A k$ du A-schéma propre et plat $\mathscr{X}$ satisfait $H^2(X, \mathcal{O}_X)=0$, un théorème de semi-continuité dû à Grothendieck (voir [Mu, § 5, p. 50] ou [H, Chapter III, Thm. 12.8]) garantit l'existence d'un ouvert non vide de $\mathrm{Spec}(A)$ tel que $H^2(\mathscr{X}_m, \mathcal{O}_{\mathscr{X}_m})=0$ pour tout m dans cet ouvert. Quitte à remplacer A par A_f avec $f\in A$ non nul, on peut donc supposer qu'il en est ainsi pour tout $m\in\mathrm{Spec}(A)$. On a alors $H^2(\mathscr{X}\times_A \kappa, \mathcal{O}_{\mathscr{X}\times_A\kappa})=0$ pour tout homomorphisme de A dans un corps κ.

On peut trouver une sous-extension finie L de $\bar{k}/k$ telle que $\mathrm{Gal}(\bar{k}/L)$ agisse trivialement sur le groupe de type fini $\mathrm{NS}(\bar{X})$ et que X_L possède un point L-rationnel. Pour une telle extension, que nous fixons désormais, l'application composée $\mathrm{Pic}(X_L)\to\mathrm{Pic}(\bar{X})\to\mathrm{NS}(\bar{X})$ est surjective.

Soit B la clôture intégrale de A dans L. Comme B est fini sur A, quitte à restreindre A, on peut supposer que B est principal, et que n et $r=[L:k]=[B:A]$ sont inversibles dans A.

On voit ainsi que les hypothèses du lemme 3.3 sont satisfaites par le B-schéma $\mathscr{X}_B$. La restriction $\mathrm{CH}^2(\mathscr{X}_B)\to\mathrm{CH}^2(X_L)$, qui est surjective, a donc un noyau fini, soit N.

L'image réciproque donne lieu à un diagramme commutatif de suites exactes:

$$\begin{array}{ccccccc} 0 & \longrightarrow & N & \longrightarrow & \mathrm{CH}^2(\mathscr{X}_B) & \longrightarrow & \mathrm{CH}^2(X_L) \longrightarrow 0 \\ & & & & \uparrow{\scriptstyle i} & & \uparrow{\scriptstyle j} \\ & & & & \mathrm{CH}^2(\mathscr{X}) & \longrightarrow & \mathrm{CH}^2(X) \longrightarrow 0. \end{array}$$

Par un argument de transfert ([Gi1, Thm. 7.22]), le groupe $\mathrm{Ker}(i)$ est contenu dans le groupe ${}_r\mathrm{CH}^2(\mathscr{X})$. Comme r est inversible, ce dernier groupe est fini d'après le Théorème 1.1. Le groupe N est aussi fini.

Plaçons-nous un instant dans la catégorie de Serre des groupes abéliens modulo les groupes abéliens finis. On a alors un carré commutatif

$$\begin{array}{ccc} \mathrm{CH}^2(\mathscr{X}_B) & \simeq & \mathrm{CH}^2(X_L) \\ \uparrow{\scriptstyle i} & & \uparrow{\scriptstyle j} \\ \mathrm{CH}^2(\mathscr{X}) & \longrightarrow & \mathrm{CH}^2(X) \end{array}$$

où i est injective et la flèche inférieure est surjective. Ainsi, $\mathrm{CH}^2(X)$ s'identifie à un sous-groupe de $\mathrm{CH}^2(\mathscr{X}_B)$, et ${}_n\mathrm{CH}^2(X)$ s'identifie à un sous-groupe de ${}_n\mathrm{CH}^2(\mathscr{X}_B)$, lequel est nul d'après le théorème 1.1.

Revenant dans la catégorie des groupes abéliens, on voit que ${}_n\mathrm{CH}^2(X)$ est fini, ce qui achève la démonstration. □

Remarque 3.5 De la suite de localisation et des théorèmes de finitude classique pour le groupe de Chow de codimension 1 (théorèmes de Mordell-Weil et de Néron-Severi) on déduit du théorème 3.4 que pour tout ouvert U d'une variété X comme dans le théorème, le groupe ${}_n\mathrm{CH}^2(U)$ est fini.

4 Rappels sur la méthode galoisienne; contribution de Salberger; démonstration des théorèmes B et C

Soit k un corps de caractéristique nulle, de clôture algébrique $\bar{k}$. Soit $\mathfrak{g} = \mathrm{Gal}(\bar{k}/k)$. Soit X une k-variété géométriquement intègre, projective et lisse, et $\bar{X} = X \times_k \bar{k}$. Tout élément z de $\mathrm{CH}^2(X)$ d'image nulle dans $\mathrm{CH}^2(\bar{X})$ a une image nulle dans un groupe $\mathrm{CH}^2(X_L)$ pour L/k extension finie convenable. Un argument de transfert montre alors que z est un élément de torsion dans $\mathrm{CH}^2(X)$. On a ainsi la suite exacte:

$$(*) \qquad 0 \to \mathrm{Ker}[\mathrm{CH}^2(X) \to \mathrm{CH}^2(\bar{X})] \to \mathrm{CH}^2(X)_{\mathrm{tors}} \to (\mathrm{CH}^2(\bar{X})_{\mathrm{tors}})^{\mathfrak{g}}.$$

Dans [CT/R1], nous avons montré:

R1 *Pour toute variété X projective, lisse et géométriquement connexe sur un corps de nombres, ou plus généralement sur un corps k de type fini sur le corps $\mathbb{Q}$ des rationnels, le groupe $(\mathrm{CH}^2(\bar{X})_{\mathrm{tors}})^{\mathfrak{g}}$ est fini* ([CT/R1, Theorem 3.3]; combinaison des théorèmes de Merkur'ev/Suslin et du théorème de Deligne).

R2 *Pour toute variété X lisse géométriquement connexe sur un corps k de clôture séparable $\bar{k}$, avec $\mathfrak{g} = \mathrm{Gal}(\bar{k}/k)$, on a une suite exacte:*

$$(**) \quad H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2)) \to \mathrm{Ker}[\mathrm{CH}^2(X) \to \mathrm{CH}^2(\bar{X})] \to H^1(\mathfrak{g}, H^1(\bar{X}, \mathscr{K}_2))$$

([CT/R1, Prop. 3.6]; c'est l'analyse galoisienne à la Bloch).

R3 *Soit X une variété propre, lisse, géométriquement connexe sur un corps k de caractéristique zéro. Supposons $H^2(X, \mathcal{O}_X) = 0$. Alors il existe un entier $n > 0$ qui annule le groupe $H^1(\mathfrak{g}, H^1(\bar{X}, \mathscr{K}_2))$* ([CT/R1, Prop. 3.9.b]).

R4 *Soit X une variété propre, lisse, géométriquement connexe sur un corps k de caractéristique zéro. Supposons $H^1(X, \mathcal{O}_X) = 0$. Alors il existe un entier $m > 0$ qui annule le groupe $H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2))$* ([CT/R1 Prop. 3.7]).

Avant d'établir le théorème principal dans sa généralité, observons que la suite exacte (*), les rappels **R1**, **R2**, **R3** et **R4** et le théorème 3.4 impliquent immédiatement le:

Théorème 4.1 *Soit k un corps de nombres et X une k-variété projective, lisse, géométriquement intègre, telle que $H^2(X, \mathcal{O}_X) = 0 = H^1(X, \mathcal{O}_X)$. Alors, le sous-groupe de torsion du groupe $\mathrm{CH}^2(X)$ est fini.*

Remarque 4.1.1 La démonstration de ce dernier théorème ne fait pas appel à la théorie du corps de classes supérieur (§ 2). Lorsque $H^1(X, \mathcal{O}_X) = 0$, il suffit en effet, pour établir le théorème 3.4, d'utiliser le Théorème 1.1 et les lemmes 3.1 et 3.2.

Comme l'a observé Salberger, on peut s'affranchir de l'hypothèse $H^1(X, \mathcal{O}_X) = 0$ en utilisant la proposition suivante, qui s'appuie sur, et généralise, un résultat dû à l'un d'entre nous [R3] dans le cas des courbes:

Proposition 4.2 (P. Salberger) *Soit k un corps de nombres, ou un corps p-adique, ou le corps des réels. Pour toute k-variété projective lisse et géométriquement intègre X sur k, le groupe $H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2))$ est d'exposant fini.*

(Salberger nous a communiqué cette proposition pour k un corps de nombres, mais les résultats de [R3] couvrent aussi le cas local.)

Démonstration (variante de celle de Salberger). Quitte à faire une extension finie de k, on peut supposer $X(k) \neq \emptyset$. Alors $H^1(\mathfrak{g}, K_2\,\bar{k}(X)) = 0$ ([CT, Theorem B]), et on a la suite exacte

$$0 \to H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2)) \longrightarrow H^2(\mathfrak{g}, H^0(\bar{X}, \mathscr{K}_2)) \xrightarrow{\rho_X} H^2(\mathfrak{g}, K_2\,\bar{k}(X)).$$

Soit $Y \subset X$ une section hyperplane lisse, géométriquement intègre.

Montrons que l'application naturelle $H^2(\mathfrak{g}, H^0(\bar{X}, \mathscr{K}_2)) \to H^2(\mathfrak{g}, H^0(\bar{Y}, \mathscr{K}_2))$ induit une application $\mathrm{Res}_1 : \mathrm{Ker}(\rho_X) \to \mathrm{Ker}(\rho_Y)$.

Soit A l'anneau local sur X du point générique de Y, et F le corps résiduel, qui n'est autre que $k(Y)$. L'anneau $\bar{A} = A \otimes_k \bar{k}$ est un anneau de valuation discrète de corps des fractions $\bar{k}(X)$ et de corps résiduel $\bar{k}(Y)$. On a donc la suite exacte de $\mathfrak{g}$-modules

$$0 \to K_2\,\bar{A} \to K_2\,\bar{k}(X) \to (F \otimes_k \bar{k})^* \to 1$$

(Dennis/Stein; voir [Sh3, p. 81]) et donc, grâce au théorème 90 de Hilbert, une injection

$$H^2(\mathfrak{g}, K_2\,\bar{A}) \to H^2(\mathfrak{g}, K_2\,\bar{k}(X)).$$

Ainsi le noyau de l'application ρ_X coïncide avec le noyau de l'application naturelle $H^2(\mathfrak{g}, H^0(\bar{X}, \mathscr{K}_2)) \to H^2(\mathfrak{g}, K_2\,\bar{A})$. Du diagramme commutatif

$$\begin{array}{ccc} H^2(\mathfrak{g}, H^0(\bar{X}, \mathscr{K}_2)) & \to & H^2(\mathfrak{g}, K_2\,\bar{A}) \\ \downarrow & & \downarrow \\ H^2(\mathfrak{g}, H^0(\bar{Y}, \mathscr{K}_2)) & \to & H^2(\mathfrak{g}, K_2\,\bar{k}(Y)) \end{array}$$

on déduit alors l'application $\mathrm{Res}_1 : \mathrm{Ker}(\rho_X) \to \mathrm{Ker}(\rho_Y)$.

On dispose par ailleurs du diagramme commutatif de suites exactes ([CT/R1], démonstration de la proposition 3.8; on utilise ici le calcul dû à Suslin ([Su, Prop. 5.2]) de la torsion de $H^0(\bar{X}, \mathscr{K}_2)$, calcul qui donne la commutativité du carré gauche du diagramme):

$$\begin{array}{ccccccccc} 0 & \longrightarrow & H^1_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2)) & \longrightarrow & H^0(\bar{X}, \mathscr{K}_2) & \longrightarrow & H^0(\bar{X}, \mathscr{K}_2) \otimes_{\mathbf{Z}} \mathbb{Q} & \longrightarrow & 0. \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & H^1_{\text{ét}}(\bar{Y}, \mathbb{Q}/\mathbb{Z}(2)) & \longrightarrow & H^0(\bar{Y}, \mathscr{K}_2) & \longrightarrow & H^0(\bar{Y}, \mathscr{K}_2) \otimes_{\mathbf{Z}} \mathbb{Q} & \longrightarrow & 0. \end{array}$$

L'application Res_1 est donc induite par une application

$$\mathrm{Res}_2 : H^2(\mathfrak{g}, H^1_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2))) \to H^2(\mathfrak{g}, H^1_{\text{ét}}(\bar{Y}, \mathbb{Q}/\mathbb{Z}(2))),$$

elle-même induite par l'application image réciproque:

$$H^1_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2)) \to H^1_{\text{ét}}(\bar{Y}, \mathbb{Q}/\mathbb{Z}(2)).$$

Au niveau des groupes fondamentaux, l'application $\pi_1(\bar{Y}) \to \pi_1(\bar{X})$ est surjective ([SGA1, X, Lemme 2.10, p. 274]). Ceci entraîne que le morphisme de k-variétés abéliennes $J_X = \underline{\text{Pic}}^{\circ}_{X/k} \to J_Y = \underline{\text{Pic}}^{\circ}_{Y/k}$ est injectif. Le théorème de complète réductibilité de Poincaré ([Mu, Chap. IV, Thm. 1 p. 173]) assure alors l'existence d'un k-morphisme $J_Y \to J_X$ tel que le composé $J_X \to J_Y \to J_X$ soit une isogénie (surjection à noyau fini).

Par restriction aux points de torsion puis par torsion par les racines de l'unité, on obtient des homomorphismes de modules galoisiens divisibles

$$i\colon \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{X/k}(\bar{k}) \otimes \mu_n(\bar{k})) \to \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{Y/k}(\bar{k}) \otimes \mu_n(\bar{k}))$$

$$j\colon \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{Y/k}(\bar{k}) \otimes \mu_n(\bar{k})) \to \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{X/k}(\bar{k}) \otimes \mu_n(\bar{k}))$$

avec $j \circ i$ surjectif et $\text{Ker}(j \circ i)$ fini. Comme les groupes de cohomologie galoisienne d'un $\mathfrak{g}$-module d'exposant fini sont aussi d'exposant fini, on en déduit aisément que le noyau de l'application

$$H^2(\mathfrak{g}, i)\colon\ H^2(\mathfrak{g}, \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{X/k}(\bar{k}) \otimes \mu_n(\bar{k}))) \to H^2(\mathfrak{g}, \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{Y/k}(\bar{k}) \otimes \mu_n(\bar{k})))$$

est d'exposant fini.

On dispose par ailleurs d'un diagramme commutatif de suites exactes de modules galoisiens:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{X/k}(\bar{k}) \otimes \mu_n(\bar{k})) & \longrightarrow & H^1_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2)) & \longrightarrow & F_1 & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \varinjlim_n ({}_n\underline{\text{Pic}}^{\circ}_{Y/k}(\bar{k}) \otimes \mu_n(\bar{k})) & \longrightarrow & H^1_{\text{ét}}(\bar{Y}, \mathbb{Q}/\mathbb{Z}(2)) & \longrightarrow & F_2 & \longrightarrow & 0
\end{array}$$

où les modules F_1 et F_2 sont finis. On voit alors que le noyau de l'application

$$\text{Res}_2\colon\ H^2(\mathfrak{g}, H^1_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2))) \to H^2(\mathfrak{g}, H^1_{\text{ét}}(\bar{Y}, \mathbb{Q}/\mathbb{Z}(2)))$$

est d'exposant fini (il est ici commode de raisonner dans la catégorie de Serre des groupes abéliens modulo les groupes d'exposant fini).

Ainsi, par sections hyperplanes lisses géométriquement intègres successives (Bertini, [A/K]), on trouve une courbe projective et lisse géométriquement intègre C et une application de $\text{Ker}(\rho_X)$ dans $\text{Ker}(\rho_C)$ à noyau d'exposant fini.

(Au lieu d'utiliser les propriétés des variétés abéliennes, Salberger utilise le théorème de Lefschetz difficile).

Jusqu'ici on n'a pas utilisé d'autre hypothèse sur k que $\text{car}(k) = 0$. Si k est un corps de nombres, ou un corps p-adique, ou le corps des réels, le Théorème

3.7 de [R3] assure que lorsque C est une courbe projective et lisse, $\mathrm{Ker}(\rho_C)$ est d'exposant fini. Il en est donc de même de $\mathrm{Ker}(\rho_X)$. □

Les rappels et la proposition 4.2 donnent immédiatement le théorème B de l'introduction:

Théorème 4.3 *Soit X une variété projective, lisse et géométriquement connexe définie sur un corps de nombres k. Supposons le groupe $H^2(X, \mathcal{O}_X)$ nul. Alors le sous-groupe de torsion $\mathrm{CH}^2(X)_{\mathrm{tors}}$ du groupe de Chow $\mathrm{CH}^2(X)$ des cycles de codimension deux est d'exposant fini.*

Démonstration. Il suffit de combiner la suite (*), les rappels **R1**, **R2**, **R3** et la proposition 4.2. □

La combinaison de notre théorème 3.4 et du théorème 4.3 donne alors le théorème C de l'introduction, c'est-à-dire la finitude de $\mathrm{CH}^2(X)_{\mathrm{tors}}$ sous les hypothèses de ces deux théorèmes.

Remarque 4.4 Le théorème C (ou encore, puisqu'il s'agit de surfaces, la combinaison des théorèmes 3.4 et 5.1 ci-après) recouvre en particulier le cas étudié par Gros [Gs] et Ōkōchi [O]: celui d'une surface fibrée en courbes de genre zéro au-dessus d'une courbe. Un point-clé dans leur démonstration était l'utilisation d'un théorème de Kato [K] sur la cohomologie étale des courbes, dont la démonstration repose précisément sur la théorie du corps de classes supérieur de Bloch et Kato-Saito (voir [K, Prop. 2.1]).

5 Surfaces sur un corps quelconque

Lorsque X est une surface avec $H^2(X, \mathcal{O}_X)=0$, on peut, en utilisant la classification des surfaces, établir le théorème C sans utiliser la Proposition 4.2 ni le résultat principal de [R3]. On a en effet le:

Théorème 5.1 *Soit k un corps, $\mathrm{car.}(k)=0$, et soit X une k-surface projective, lisse, géométriquement intègre, telle que $H^2(X, \mathcal{O}_X)=0$. Alors le groupe $\mathrm{Ker}[\mathrm{CH}^2(X)\to \mathrm{CH}^2(\bar{X})]$ est d'exposant fini.*

Dans le cas particulier où k est un corps de nombres, il résulte alors du théorème 3.4 que $\mathrm{Ker}[\mathrm{CH}^2(X)\to\mathrm{CH}^2(\bar{X})]$ est fini. Comme par ailleurs le groupe $(\mathrm{CH}^2(\bar{X})_{\mathrm{tors}})^{\mathfrak{g}}$ est alors fini (§ 4, Rappel **R1**), on obtient la finitude du groupe $\mathrm{CH}^2(X)_{\mathrm{tors}}$ tout entier, c'est-à-dire le théorème C pour les surfaces.

Démonstration du théorème 5.1 Soit $\mathcal{B}$ la sous-catégorie de Serre de la catégorie des groupes abéliens, formée des groupes d'exposant fini. Nous devons montrer que le groupe $\mathrm{Ker}[\mathrm{CH}^2(X)\to\mathrm{CH}^2(\bar{X})]$ est dans $\mathcal{B}$. Par un argument de transfert, on voit qu'il suffit d'établir cet énoncé pour $\mathrm{Ker}[\mathrm{CH}^2(X_K)\to\mathrm{CH}^2(\bar{X})]$, pour K/k une extension finie de k. On s'autorisera donc de telles extensions finies du corps de base.

On dispose de la suite exacte (§ 4, **R2**):

$$(**)\quad H^1(\mathfrak{g}, K_2\bar{k}(X)/H^0(\bar{X}, \mathcal{K}_2))\to \mathrm{Ker}[\mathrm{CH}^2(X)\to\mathrm{CH}^2(\bar{X})]\to H^1(\mathfrak{g}, H^1(\bar{X}, \mathcal{K}_2)).$$

L'hypothèse $H^2(X, \mathcal{O}_X)=0$ implique que le groupe $H^1(\mathfrak{g}, H^1(\bar{X}, \mathcal{K}_2))$ est dans $\mathcal{B}$ ([CT/R1, 3.9 (b)]).

Il suffit donc de montrer que l'image du groupe $H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2))$ dans $\mathrm{Ker}[\mathrm{CH}^2(X) \to \mathrm{CH}^2(\bar{X})]$ est dans $\mathscr{B}$.

Comme le groupe $H^0(\bar{X}, \mathscr{K}_2)$ est un invariant birationnel (cf. [Su, démonstration du Cor. 5.6]), il en est de même du groupe $H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2))$. On peut donc se ramener, après extension finie du corps de base, au cas où la surface X est minimale, et le reste après extension à $\bar{k}$.

Utilisons alors la classification algébrique des surfaces projectives et lisses sur un corps algébriquement clos de caractéristique zéro ([Be, B/M]). Soit X une surface minimale sur un tel corps, vérifiant $H^2(X, \mathcal{O}_X) = 0$. Alors ([Be, Lemme V. 18]),

– soit $H^1(X, \mathcal{O}_X) = 0$, et dans ce cas, le groupe $H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2))$ lui-même est dans $\mathscr{B}$ ([CT/R1, 3.7]);

– soit X admet une fibration $p: X \to C$ au-dessus d'une courbe projective et lisse C, à fibre générique lisse et géométriquement intègre, et telle que la flèche induite $\underline{\mathrm{Pic}}^{\circ}_{C/k} \to \underline{\mathrm{Pic}}^{\circ}_{X/k}$ soit un isomorphisme (la flèche $p: X \to C$ provenant de l'application d'Albanese).

Il suffit donc d'examiner le dernier cas. Par un argument de transfert, quitte à faire une extension finie du corps de base, on peut supposer C et le morphisme $X \to C$ définis sur k, et $X(k) \neq \emptyset$, donc aussi $C(k) \neq \emptyset$. D'après la variante cohomologique [CT] du théorème 90 pour K_2, on sait alors que $H^1(\mathfrak{g}, K_2\,\bar{k}(X)) = 0$ et $H^1(\mathfrak{g}, K_2\,\bar{k}(C)) = 0$. Par image réciproque via p, on a donc un diagramme commutatif de suites exactes:

$$
\begin{array}{ccccccc}
0 \longrightarrow & H^1(\mathfrak{g}, K_2\,\bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2)) & \longrightarrow & H^2(\mathfrak{g}, H^0(\bar{X}, \mathscr{K}_2)) & \longrightarrow & H^2(\mathfrak{g}, K_2\,\bar{k}(X)) \\
(***) & \uparrow p_1^* & & \uparrow p_2^* & & \uparrow p_3^* \\
0 \longrightarrow & H^1(\mathfrak{g}, K_2\,\bar{k}(C)/H^0(\bar{C}, \mathscr{K}_2)) & \longrightarrow & H^2(\mathfrak{g}, H^0(\bar{C}, \mathscr{K}_2)) & \longrightarrow & H^2(\mathfrak{g}, K_2\,\bar{k}(C)).
\end{array}
$$

Montrons tout d'abord que l'application p_3^* a un noyau dans $\mathscr{B}$. Soit M un point fermé de la fibre générique de p, soit A l'anneau local de M sur cette fibre, et F le corps résiduel, qui est une extension finie du corps $k(C)$.

L'anneau $\bar{A} = A \otimes_k \bar{k}$ est un anneau de Dedekind semi-local de corps des fractions $\bar{k}(X)$. On a donc la suite exacte de $\mathfrak{g}$-modules (voir [Sh3, p. 81]):

$$0 \to K_2\,\bar{A} \to K_2\,\bar{k}(X) \to (F \otimes_k \bar{k})^* \to 1,$$

et donc grâce au théorème 90 de Hilbert, une injection

$$H^2(\mathfrak{g}, K_2\,\bar{A}) \to H^2(\mathfrak{g}, K_2\,\bar{k}(X)).$$

Ainsi le noyau de l'application p_3^* coïncide avec le noyau de l'application naturelle $H^2(\mathfrak{g}, K_2\,\bar{k}(C)) \to H^2(\mathfrak{g}, K_2\,\bar{A})$. Par la composition $k(C) \to A \to F$, ce dernier noyau est contenu dans celui de l'application naturelle $H^2(\mathfrak{g}, K_2\,\bar{k}(C)) \to H^2(\mathfrak{g}, K_2(F \otimes_k \bar{k}))$, lequel est clairement dans $\mathscr{B}$.

Montrons maintenant que noyau et conoyau de l'application p_2^* sont dans $\mathscr{B}$. Comme il a été rappelé dans la démonstration de la Proposition 4.2, grâce à Suslin on dispose d'un diagramme commutatif de suites exactes naturelles ([CT/R1, p. 190]):

$$
\begin{array}{ccccccc}
0 \longrightarrow & H^1_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2)) & \longrightarrow & H^0(\bar{X}, \mathscr{K}_2) & \longrightarrow & H^0(\bar{X}, \mathscr{K}_2) \otimes_{\mathbb{Z}} \mathbb{Q} \longrightarrow 0. \\
 & \uparrow & & \uparrow & & \uparrow \\
0 \longrightarrow & H^1_{\text{ét}}(\bar{C}, \mathbb{Q}/\mathbb{Z}(2)) & \longrightarrow & H^0(\bar{C}, \mathscr{K}_2) & \longrightarrow & H^0(\bar{C}, \mathscr{K}_2) \otimes_{\mathbb{Z}} \mathbb{Q} \longrightarrow 0.
\end{array}
$$

Les hypothèses géométriques sur l'application p impliquent que la flèche verticale de gauche identifie le groupe divisible $H^1_{\text{ét}}(\bar{C}, \mathbb{Q}/\mathbb{Z}(2))$ avec le sous-groupe divisible maximal de $H^1_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2))$. Le conoyau de cette flèche peut être identifié avec le groupe *fini*

$$\bigoplus_{l \text{ premier}} H^2_{\text{ét}}(\bar{X}, \mathbb{Z}_l(2))_{\text{tors}},$$

lequel est isomorphe (non canoniquement) au sous-groupe de torsion du groupe de Néron-Severi de $\bar{X}$. Des propriétés de base de la cohomologie galoisienne résulte alors que noyau et conoyau de

$$p_2^*: H^2(\mathfrak{g}, H^0(\bar{C}, \mathscr{K}_2)) \to H^2(\mathfrak{g}, H^0(\bar{X}, \mathscr{K}_2))$$

sont dans $\mathscr{B}$.

De $\mathrm{Ker}(p_3^*) \in \mathscr{B}$, $\mathrm{Ker}(p_2^*) \in \mathscr{B}$, $\mathrm{Coker}(p_2^*) \in \mathscr{B}$, et du diagramme (***), il résulte formellement $\mathrm{Coker}(p_1^*) \in \mathscr{B}$.

La suite exacte (**) est fonctorielle contravariante pour les morphismes dominants de variétés lisses, et l'on a $\mathrm{CH}^2(C) = 0$ puisque C est une courbe. Ainsi l'application qui envoie $H^1(\mathfrak{g}, K_2 \bar{k}(X)/H^0(\bar{X}, \mathscr{K}_2))$ dans le groupe $\mathrm{Ker}[\mathrm{CH}^2(X) \to \mathrm{CH}^2(\bar{X})]$ se factorise par le conoyau de p_1^*. Comme ce dernier est dans $\mathscr{B}$, ceci suffit à conclure. □

Remarque 5.2 Il serait intéressant de reprendre la démonstration de Bloch et al. [Bl/K/L] et de voir si, sur un corps non algébriquement clos, elle permet d'établir le théorème 5.1 directement, i.e. sans passer par la K-théorie.

6 Surfaces sur un corps local

Dans [CT/R1] et [R2], nous avons obtenu des résultats de finitude pour la torsion de $\mathrm{CH}^2(X)$ pour X propre et lisse sur un corps p-adique. (Certains des résultats de [R2] valent sans supposer $H^2(X, \mathscr{O}_X)$ nul.)

L'objet du présent paragraphe est d'affiner les résultats de finitude sous l'hypothèse $H^2(X, \mathscr{O}_X) = 0$, particulièrement lorsque X est une surface.

Théorème 6.1 *Soit X une variété projective, lisse et géométriquement connexe, définie sur un corps k p-adique. Supposons le groupe $H^2(X, \mathscr{O}_X)$ nul. Alors:*
(i) *Le groupe* $\mathrm{Ker}[\mathrm{CH}^2(X) \to \mathrm{CH}^2(\bar{X})]$ *est fini.*
(ii) *Si X est une surface, le groupe* $\mathrm{CH}^2(X)_{\text{tors}}$ *est fini.*

Démonstration. Pour établir (i), il suffit de combiner les rappels **R2** et **R3** (§ 4), la Proposition 4.2 et le fait suivant: pour une variété lisse X sur un corps p-adique, et n un entier positif, le groupe ${}_n\mathrm{CH}^2(X)$ est fini (cf. [CT/S/S, Corollaire 2, p. 773]).

Pour établir (ii), on commence par observer que l'approche à la Merkur'ev/Suslin du théorème de Roitman (Bloch [Bl1], voir aussi [CT/R1, Prop. 3.1]) donne, lorque X est une surface, un isomorphisme $\mathfrak{g}$-équivariant:

$$\mathrm{CH}^2(\bar{X})_{\text{tors}} \simeq H^3_{\text{ét}}(\bar{X}, \mathbb{Q}/\mathbb{Z}(2)),$$

qui, par les identifications usuelles, donne un isomorphisme $\mathfrak{g}$-équivariant, où $\underline{\mathrm{Alb}}_X$ désigne la variété d'Albanese de X:

$$A_0(\bar{X})_{\mathrm{tors}} = \underline{\mathrm{Alb}}_X(\bar{k})_{\mathrm{tors}}.$$

Ainsi le groupe $(A_0(\bar{X})_{\mathrm{tors}})^{\mathfrak{g}}$ s'identifie au groupe $\underline{\mathrm{Alb}}_X(k)_{\mathrm{tors}}$, groupe qui est fini si k est un corps p-adique. D'après la suite exacte (*) du § 4, il suffit donc d'établir la finitude de $\mathrm{Ker}[\mathrm{CH}^2(X) \to \mathrm{CH}^2(\bar{X})]$. Celle-ci a fait l'objet de (i); pour une surface on peut aussi l'établir en utilisant le théorème 5.1 (et la finitude de la n-torsion de $\mathrm{CH}^2(X)$ pour $n>0$ et X lisse sur un corps p-adique). □

Remarque 6.1.2 Dans [CT/R1], nous n'avions obtenu les résultats de finitude du théorème 6.1 ci-dessus que sous l'une des hypothèses supplémentaires suivantes: $H^1(X, \mathcal{O}_X)=0$ ([CT/R1, 3.14]), ou bien X a potentiellement bonne réduction ([CT/R1, 3.16]). Pour une surface X avec (très) bonne réduction sur un corps p-adique, le théorème 6.3 ci-dessous permet de calculer la partie première à p du groupe fini $\mathrm{CH}^2(X)_{\mathrm{tors}}$.

Proposition 6.2 *Soit A un anneau de valuation discrète, de corps résiduel κ, et de corps des fractions k supposé de caractéristique nulle. Soit $\mathcal{X}/A$ un schéma projectif et lisse à fibres géométriques intègres, soit $X=\mathcal{X}_k$ la fibre générique et soit $Y=\mathcal{X}_\kappa$ la fibre spéciale. Supposons $H^2(Y, \mathcal{O}_Y)=0$. Alors, dans la suite de localisation ($\mathcal{L}$)* (cf. § 3), *l'application*

$$H^1(X, \mathcal{K}_2) \xrightarrow{\partial} \mathrm{Pic}(Y)$$

est surjective, et l'application $\mathrm{CH}^2(\mathcal{X}) \to \mathrm{CH}^2(X)$ est un isomorphisme.

Démonstration. Soit A_h le hensélisé de A en son idéal maximal m. L'hypothèse $H^2(Y, \mathcal{O}_Y)=0$, le théorème de Grothendieck cité plus haut ([Gk1, p. 11]) et le théorème d'approximation d'Artin ([Ar, Cor. 2.2, p. 27]) assurent que l'application

$$\mathrm{Pic}(\mathcal{X}_{A_h}) \to \mathrm{Pic}(Y)$$

est surjective. Etant donné une classe $\mathcal{L}_0 \in \mathrm{Pic}(Y)$, il existe donc une extension finie L de k, un idéal maximal M de la clôture intégrale B de A dans L, de corps résiduel $\kappa(M)$ isomorphe à κ via l'application naturelle $\kappa=\kappa(m) \to \kappa(M)$, et une classe $\mathcal{L} \in \mathrm{Pic}(\mathcal{X}_B)$ qui induise $\mathcal{L}_0$ via l'application composée

$$\mathrm{Pic}(\mathcal{X}_B) \to \mathrm{Pic}(\mathcal{X}_{B_M}) \to \mathrm{Pic}(\mathcal{X}_{\kappa(M)}) \simeq \mathrm{Pic}(\mathcal{X}_{\kappa(m)}) = \mathrm{Pic}(Y).$$

Soient M_i $(i=1, \ldots, n)$ les idéaux premiers de B non nuls, et $M_1=M$. Soit $\pi \in L^*$ un générateur de l'idéal principal M. Par fonctorialité covariante de la suite de localisation (fonctorialité qui résulte de celle de la suite de Brown-Gersten-Quillen, voir Gillet [Gi1, Thm. 7.22]), on a un diagramme commutatif:

$$\begin{array}{ccc} H^1(X_L, \mathcal{K}_2) & \xrightarrow{\{\partial_{M_i}\}} & \bigoplus_{i=1,\ldots,n} \mathrm{Pic}(\mathcal{X}_{\kappa(M_i)}) \\ \downarrow & & \downarrow \\ H^1(X, \mathcal{K}_2) & \xrightarrow{\partial} & \mathrm{Pic}(Y), \end{array}$$

où les flèches verticales sont les normes, la flèche $\mathrm{Pic}(\mathscr{X}_{\kappa(M_1)}) \to \mathrm{Pic}(Y)$ étant un isomorphisme. Le calcul indiqué dans la démonstration du lemme 3.2 montre que l'image de $\mathscr{L} \otimes \pi$ par la flèche composée

$$\mathrm{Pic}(\mathscr{X}_B) \otimes L^* \to \mathrm{Pic}(X_L) \otimes L^* \to H^1(X_L, \mathscr{K}_2) \to \bigoplus_{i=1,\ldots,n} \mathrm{Pic}(\mathscr{X}_{\kappa(M_i)})$$

est une famille $\{\mathscr{L}_i\}$ avec $\mathscr{L}_i = 0$ pour $i \neq 1$, et $\mathscr{L}_1$ d'image $\mathscr{L}_0$ via l'isomorphisme $\mathrm{Pic}(\mathscr{X}_{\kappa(M)}) \to \mathrm{Pic}(Y)$.

Par commutativité du diagramme, on voit que $\mathscr{L}_0$ est dans l'image de $H^1(X, \mathscr{K}_2)$, et la proposition est démontrée. □

Remarque 6.2.1 Lorsque A est supposé complet, la démonstration de la proposition se simplifie considérablement; on n'a pas à utiliser le théorème d'approximation d'Artin, ni à passer à une extension finie de k. La proposition dans ce cas est un cas particulier de [R2, Corollary 1.4].

Remarque 6.2.2 La démonstration donnée ci-dessus présente une certaine analogie avec un argument de Gillet [Gi2].

Le théorème suivant étend la proposition 1.11 de [R2] ainsi qu'un résultat de Coombes [Cb]:

Théorème 6.3 *Soit k un corps p-adique* (car.$(k) = 0$), *A l'anneau des entiers de k, m l'idéal maximal, κ le corps résiduel. Soit $\mathscr{X}$ un A-schéma projectif et lisse, à fibres géométriques intègres, de dimension relative 2. Soit $X = \mathscr{X} \times_A k$ et $Y = \mathscr{X} \times_A \kappa$. Supposons $H^2(Y, \mathscr{O}_Y) = 0$. Pour tout l premier différent de p, les sous-groupes de torsion l-primaire $\mathrm{CH}^2(X)\{l\}$ et $\mathrm{CH}^2(Y)\{l\}$ sont des groupes finis naturellement isomorphes.*

(On trouvera plus bas (6.5) des rappels sur le groupe *fini* $\mathrm{CH}^2(Y)_{\mathrm{tors}}$.)

Démonstration. D'après la proposition 6.2 (ou [R2] pour la partie première à p), l'hypothèse $H^2(Y, \mathscr{O}_Y) = 0$ assure que la flèche de restriction $H^2(\mathscr{X}, \mathscr{K}_2) \to H^2(X, \mathscr{K}_2)$ est un isomorphisme. Ainsi, pour *tout* premier l, la restriction $H^2(\mathscr{X}, \mathscr{K}_2)\{l\} \to H^2(X, \mathscr{K}_2)\{l\}$ est un isomorphisme.

Par ailleurs, l'un des principaux résultats de [R2, Theorem 1.9] assure que pour l premier différent de p, la réduction $H^2(\mathscr{X}, \mathscr{K}_2) \to H^2(Y, \mathscr{K}_2)$ (obtenue par fonctorialité contravariante) induit une *injection* $H^2(\mathscr{X}, \mathscr{K}_2)\{l\} \hookrightarrow H^2(Y, \mathscr{K}_2)\{l\}$ (ce résultat n'utilise pas l'hypothèse $H^2(Y, \mathscr{O}_Y) = 0$.)

On voit donc que l'application composée

$$\rho\colon\ \mathrm{CH}^2(X) \simeq H^2(X, \mathscr{K}_2) \simeq H^2(\mathscr{X}, \mathscr{K}_2) \to H^2(Y, \mathscr{K}_2) \simeq \mathrm{CH}^2(Y)$$

induit une application injective:

$$\rho_l\colon\ \mathrm{CH}^2(X)\{l\} \to \mathrm{CH}^2(Y)\{l\}.$$

Notons qu'il existe une flèche de spécialisation naturelle ([F, 20.3]) $\sigma\colon \mathrm{CH}^2(X) \to \mathrm{CH}^2(Y)$, qui est induite à partir de la flèche de Gysin $\mathrm{CH}^2(\mathscr{X}) \to \mathrm{CH}^2(Y)$ (obtenue par spécialisation directe des cycles): on dispose en effet de la suite exacte

$$\mathrm{Pic}(Y) \to \mathrm{CH}^2(\mathscr{X}) \to \mathrm{CH}^2(X) \to 0,$$

et l'application composée

$$\mathrm{Pic}(Y) \to \mathrm{CH}^2(\mathscr{X}) \to \mathrm{CH}^2(Y)$$

est nulle (loc. cit.). Nous allons montrer que l'application induite

$$\sigma_l : \mathrm{CH}^2(X)\{l\} \to \mathrm{CH}^2(Y)\{l\}$$

est surjective.

Nous utiliserons le lemme suivant:

Lemme 6.4 ([Bl2]) *Soit A un anneau de valuation discrète complet, de corps résiduel κ fini. Soit $\mathscr{X}$ un A-schéma projectif et lisse, à fibres géométriques intègres, de dimension relative 2. Soit $X = \mathscr{X} \times_A k$ et $Y = \mathscr{X} \times_A \kappa$. Supposons $H^2(Y, \mathscr{O}_Y) = 0$. On peut alors trouver deux extensions finies κ_1 et κ_2 de κ, avec $[\kappa_1 : \kappa]$ et $[\kappa_2 : \kappa]$ premiers entre eux, et pour chaque $i = 1, 2$, une courbe relative $\mathscr{C}_i \subset \mathscr{X}_i = \mathscr{X} \times_A A_i$ projective, lisse et à fibres géométriquement intègres sur l'anneau des entiers A_i de l'extension non ramifiée k_i de k correspondant à κ_i, et de fibre spéciale C_i/κ_i contenant le support du 0-cycle $z_i = z \times_\kappa \kappa_i$.*

(La démonstration sera donnée à la fin du paragraphe.)

Soit alors $\tilde{z} \in \mathrm{CH}^2(Y)\{l\}$, et soit z un 0-cycle (de degré zéro) sur Y, de classe $\tilde{z}$. Soient $A_i, \mathscr{C}_i/A_i$ et C_i/κ_i comme dans le lemme. Soit J_i le A_i-schéma abélien $\underline{\mathrm{Pic}}^\circ_{\mathscr{C}_i/A_i}$. Notons $Y_i = Y \times_\kappa \kappa_i$, puis $\mathscr{X}_i = \mathscr{X} \times_A A_i$ et $X_i = X \times_k k_i$.

La classe $\tilde{z}_i \in \mathrm{CH}^2(Y_i)\{l\}$ provient d'une classe $\tilde{t}_i$ dans $J_i(\kappa_i)$ par l'application naturelle $J_i(\kappa_i) \subset \mathrm{Pic}(C_i) \to \mathrm{CH}^2(Y_i)$. Comme $J_i(\kappa_i)$ est fini, et $\tilde{z}_i$ de torsion l-primaire, on peut choisir $\tilde{t}_i \in J_i(\kappa_i)\{l\}$. Soit $n \geqq 1$ tel que $l^n \tilde{t}_i = 0$. Le sous-A-schéma ${}_{l^n}J_i$ noyau de la multiplication par l^n sur J_i est fini étale. Comme A_i est complet, la restriction induit un isomorphisme:

$${}_{l^n}J_i(A_i) \xrightarrow{\cong} {}_{l^n}J_i(\kappa_i).$$

Le diagramme commutatif:

$$\begin{array}{ccccccc}
{}_{l^n}J_i(A_i) & \subset & J_i(A_i) & \subset & \mathrm{Pic}(\mathscr{C}_i) & \longrightarrow & \mathrm{CH}^2(\mathscr{X}_i) \longrightarrow \mathrm{CH}^2(X_i) \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \quad \swarrow \sigma \\
{}_{l^n}J_i(\kappa_i) & \subset & J_i(\kappa_i) & \subset & \mathrm{Pic}(C_i) & \longrightarrow & \mathrm{CH}^2(Y_i)
\end{array}$$

montre que $\tilde{z}_i \in \mathrm{CH}^2(Y_i)$ est l'image par σ d'un élément de torsion l-primaire dans $\mathrm{CH}^2(X_i)$. Un argument de transfert ([F, Prop. 20.3]) montre alors que $\tilde{z} \in \mathrm{CH}^2(Y)\{l\}$ est dans l'image de $\mathrm{CH}^2(X)\{l\}$ par l'application σ_l.

Pour conclure, il faudrait en principe vérifier que les deux applications de spécialisation ρ_l et σ_l coïncident. Mais ceci est inutile. En effet, pour toute variété projective et lisse Y sur un corps fini, le groupe $\mathrm{CH}^2(Y)\{l\}$ est fini ([K/S], voir les rappels ci-dessous). Comme ρ_l et σ_l envoient $\mathrm{CH}^2(X)\{l\}$ dans $\mathrm{CH}^2(Y)\{l\}$ et que ρ_l est injective et σ_l surjective, chacune de ces applications est une bijection. □

Rappels 6.5 Pour une variété projective, lisse et géométriquement connexe Y sur un corps fini F, on sait que:

– Le groupe $A_0(Y)$ des zéro-cycles de degré zéro modulo l'équivalence rationnelle est un groupe fini (Kato et Saito [K/S], voir aussi [CT/R2] pour les surfaces).

– Le groupe $\mathrm{CH}^2(Y)_{\mathrm{tors}}$ est fini ([CT/S/S, Théorème 1]).

– Le groupe $A_0(Y)$ est isomorphe au groupe fondamental abélien géométrique de Y (qui classifie les revêtements abéliens de Y ne provenant pas de ceux du corps de base) (Kato et Saito [K/S], voir aussi [CT/S/S, § 2.4] et [CT/R2]). Le groupe $A_0(Y)$ est une extension du groupe $\underline{\mathrm{Alb}}_{Y/F}(F)$ des points rationnels de la variété d'Albanese de Y par un groupe déterminé par la torsion dans le groupe de Néron-Severi géométrique de Y (Kato et Saito, [K/S, Prop. 8]; pour Y une surface, voir aussi [CT/R2, Cor. 3]).

Démonstration du Lemme 6.4

(a) Si F est un corps parfait infini et Y/F une surface projective et lisse, géométriquement intègre, et si $\{P_j\}_{j\in J}$ est un ensemble fini de points fermés de Y, on peut trouver une courbe $C\subset Y$, projective, lisse, géométriquement intègre, contenant les points P_j et telle que de plus

$$H^1(Y, \mathcal{O}_Y(C))=0.$$

A l'exception de cette dernière condition, ce résultat de type Bertini est établi dans [Cr] et dans [A/K]. Comme C est obtenue comme section de $\mathcal{O}_Y(n)$ pour n entier suffisamment grand, on peut aussi bien choisir l'entier n tel que

$$H^1(Y, \mathcal{O}_Y(n))=0.$$

(b) Si, dans la situation ci-dessus, F est un corps fini, soit $F=\mathbb{F}_q$, de clôture algébrique $\bar{F}$, on peut choisir des entiers r et s premiers entre eux et considérer les corps parfaits infinis $F_{(1)}$, resp. $F_{(2)}$, union dans $\bar{F}$ des sous-extensions de degré une puissance de r, resp. de s. Appliquant alors le résultat de (*a*) on trouve des extensions *finies* F_1 et F_2 de F, de degrés premiers entre eux, et des courbes projectives, lisses, géométriquement intègres C_i/F_i sur Y_{F_i}, contenant z_{F_i}.

(c) Il reste à montrer qu'on peut relever les courbes C_i au-dessus de A_i. Pour établir ce point, oublions les indices i précédents, i.e. posons $F=F_i$, $A=A_i$ etc. Soit $A_n=A/\pi^n A$, pour π une uniformisante de A et n entier, $n\geqq 1$. L'hypothèse $H^2(Y, \mathcal{O}_Y)=0$ assure que l'on peut trouver un faisceau inversible $\mathscr{L}$ sur $\mathscr{X}$ tel que $\mathscr{L}/\pi=\mathscr{L}_Y$ soit isomorphe à $\mathcal{O}_Y(C)$ (Grothendieck, [Gk1, p. 11]). On cherche une section du faisceau $\mathscr{L}$ qui donne la section C du faisceau $\mathcal{O}_Y(C)$ par la flèche naturelle $H^0(\mathscr{X}, \mathscr{L})\to H^0(Y, \mathcal{O}_Y(C))$. Notons $\mathscr{X}_n=\mathscr{X}\times_A A/\pi^n$ et $\mathscr{L}_n=\mathscr{L}/\pi^n$. On va d'abord montrer qu'il existe une famille de diviseurs effectifs $\mathscr{C}_n\subset\mathscr{X}_n$ compatibles, avec $\mathscr{C}_1=C$, c'est-à-dire des sections $s_n\in H^0(\mathscr{X}_n, \mathscr{L}_n)=H^0(\mathscr{X}, \mathscr{L}_n)$ avec $s_n=s_{n+1}\otimes_A A/\pi^n$, telles que s_1 définisse la courbe $C\subset Y$. Supposons trouvées $s_1, \ldots, s_n$. On a la suite exacte de faisceaux cohérents sur X:

$$0\to\pi^{n+1}\mathscr{L}/\pi^{n+2}\mathscr{L}\to\mathscr{L}_{n+1}\to\mathscr{L}_n\to 0$$

et le faisceau $\pi^{n+1}\mathscr{L}/\pi^{n+2}\mathscr{L}$ est isomorphe au faisceau $\mathscr{L}/\pi$. L'obstruction à relever une section $s_n\in H^0(\mathscr{X}, \mathscr{L}_n)$ en $s_{n+1}\in H^0(\mathscr{X}, \mathscr{L}_{n+1})$ est donc à valeurs dans $H^1(\mathscr{X}, \mathscr{L}/\pi)=H^1(Y, \mathcal{O}_Y(C))$, et ce dernier groupe est nul d'après le (*a*). Comme $\mathscr{X}$ est propre sur A, la famille compatible des $\mathscr{C}_n\subset\mathscr{X}_n$ donne alors naissance à un diviseur $\mathscr{C}\subset\mathscr{X}$, qui est la section cherchée du faisceau $\mathscr{L}$: c'est l'algébrisation des modules formels ([Gk1, Cor. 1 p. 06]). □

Remarque 6.6 En suivant la même démarche, on peut établir (Bloch [Bl2]) une variante utile du lemme 6.4. Supposons donné un zéro-cycle z sur la fibre générique X/k. Ce zéro-cycle s'étend en un zéro-cycle relatif $\tilde{z}$ au-dessus de A. Soit $\tilde{z}_0$ le zéro-cycle induit sur Y. Alors, sous les hypothèses du lemme 6.4, et avec

les mêmes notations, on peut choisir les courbes $\mathscr{C}_i$ de telle sorte qu'elles contiennent le support des zéro-cycles $z_i = z \times_k k_i$.

(On introduit l'idéal cohérent $J \subset \mathscr{O}_{\mathscr{X}}$ définissant le support de $\tilde{z}$, dans (a), on choisit C telle que de plus $H^1(Y, J \otimes \mathscr{O}_Y(C)) = 0$, et dans (c), on cherche des sections compatibles des faisceaux $J \otimes \mathscr{L}_n$.)

Cette variante a la conséquence intéressante suivante. Pour une surface X/k comme dans le lemme (bonne réduction Y avec $H^2(Y, \mathscr{O}_Y) = 0$), pour toute extension non ramifiée K/k, l'application norme $N_{K/k} : A_0(X_K) \to A_0(X)$ est surjective.

Références

[A/K] Altman, A.B., Kleiman, S.L.: Bertini theorems for hypersurface sections containing a subscheme. Commun. Algebra **7**, 775–790 (1979)

[Ar] Artin, M.: Algebraic approximation of structures over complete local rings. Publ. Math. Inst. Hautes Étud. Sci. **36**, 23–58 (1969)

[Be] Beauville, A.: Surfaces algébriques complexes, 3ème éd. Astérisque **54** (1978)

[Bl1] Bloch, S.: Lectures on algebraic cycles. Duke Univ. Math. Ser. **4** (1980)

[Bl2] Bloch, S.: Letter to J.-L. Colliot-Thélène. September 3, 1980.

[Bl3] Bloch, S.: Torsion algebraic cycles, K_2 and Brauer groups of function fields. In: Kervaire, M., Ojanguren, M. (éds.) Groupe de Brauer. Séminaire, Les Plans-sur-Bex, Suisse 1980. (Lect. Notes Math., vol. 844) Berlin Heidelberg New York: Springer 1981

[Bl4] Bloch, S.: On the Chow groups of certain rational surfaces. Ann. Sci. Éc. Norm. Supér. **14**, 41–59 (1981)

[Bl5] Bloch, S.: Algebraic K-theory and class field theory for arithmetic surfaces. Ann. Math. **114**, 229–266 (1981)

[Bl6] Bloch, S.: A note on Gersten's conjecture in the mixed characteristic case. In: Dennis, K., Friedlander, E., Stein, M. (eds.) Applications of Algebraic K-Theory to Algebraic Geometry and Number Theory. Boulder 1983 (Contemp. Math. vol. 55, p. 75–78) Providence. Am. Math. Soc. 1986

[Bl/K/L] Bloch, S., Kas, A., Lieberman, D.: Zero-cycles on surfaces with $p_g = 0$. Compos. Math. **33**, 135–145 (1976)

[Bl/Sr] Bloch, S., Srinivas, V.: Remarks on correspondences and algebraic cycles. Am. J. Math. **105**, 1235–1253 (1983)

[B/M] Bombieri, E., Mumford, D.: Enriques' classification of surfaces in characteristic p. I. In: Global Analysis. Princeton: Princeton University Press 1969. II in: Complex Analysis and Algebraic Geometry. Cambridge: Cambridge University Press 1977. III. Invent. Math. **35**, 197–232 (1976)

[CT] Colliot-Thélène, J.-L.: Hilbert's Theorem 90 for K_2, with application to the Chow groups of rational surfaces. Invent. Math. **71**, 1–20 (1983)

[CT/R1] Colliot-Thélène, J.-L., Raskind, W.: K_2-cohomology and the second Chow group. Math. Ann. **270**, 165–199 (1985)

[CT/R2] Colliot-Thélène, J.-L., Raskind, W.: On the reciprocity law for surfaces over finite fields. J. Fac. Sci., Univ. Tokyo, Sect. IA **33**, 283–294 (1986)

[CT/S] Colliot-Thélène, J.-L., Sansuc, J.-J.: On the Chow groups of certain rational surfaces: a sequel to a paper of S. Bloch. Duke Math. J. **48**, 421–447 (1981)

[CT/S/S] Colliot-Thélène, J.-L., Sansuc, J.-J., Soulé, C.: Torsion dans le groupe de Chow de codimension deux. Duke Math. J. **50**, 763–801 (1983)

[Cb] Coombes, K.: The arithmetic of zero cycles on surfaces with geometric genus and irregularity zero. (preprint 1990)

[Cr] Coray, D.F.: Two remarks on the Bertini theorem. (unpublished typescript) 1980

[F] Fulton, W.: Intersection Theory. (Ergeb. Math. Grenzgeb., 3. Folge, Bd. 2) Berlin Heidelberg New York: Springer 1984

[Gi1] Gillet, H.: Riemann-Roch theorems for higher algebraic K-theory. Adv. Math. **40**, 203–289 (1981)

[Gi2] Gillet, H.: Gersten's conjecture for the K-theory with torsion coefficients of a discrete valuation ring. J. Algebra **103**, 377–380 (1986)

[Gi/Le] Gillet, H., Levine, M.: The relative form of Gersten's conjecture over a discrete valuation ring: the smooth case. J. Pure Appl. Algebra **46**, 59–71 (1987)
[Gn1] Grayson, D.: The K-theory of hereditary categories. J. Pure Appl. Algebra **11**, 67–74 (1977)
[Gn2] Grayson, D.: Universal exactness in algebraic K-theory. J. Pure Appl. Algebra **36**, 139–141 (1985)
[Gs] Gros, M.: 0-cycles de degré 0 sur les surfaces fibrées en coniques. J. Reine Angew. Math. **373**, 166–184 (1987)
[Gk1] Grothendieck, A.: Géométrie algébrique et géométrie formelle (exposé Bourbaki **182** (1958/1959)). Reproduit dans: Fondements de la géométrie algébrique. Paris: Secrétariat mathématique de l'Institut Henri Poincaré 1962
[Gk2] Grothendieck, A.: Technique de descente et théorèmes d'existence en géométrie algébrique, V. Les Schémas de Picard. Théorèmes d'existence (exposé Bourbaki **232** (1961/62)); VI. Les schémas de Picard. Propriétés générales. (exposé Bourbaki **236** (1961/62)). Reproduit in: Fondements de la géométrie algébrique. Paris: Secrétariat mathématique de l'institut Henri Poincaré 1962
[Gk3] Grothendieck, A.: Le groupe de Brauer, II. In: Dix exposés sur la cohomologie des schémas, J. Giraud et al. (Adv. Stud. Pure Math., vol. 3, pp. 67–87) Amsterdam: North-Holland 1968
[H] Hartshorne, R.: Algebraic Geometry. (Grad. Texts Math., vol. 52) Berlin Heidelberg New York: Springer 1977
[K] Kato, K.: A Hasse principle for two dimensional global fields. J. Reine Angew. Math. **366**, 142–181 (1986)
[K/S] Kato, K., Saito, S.: Unramified class field theory of arithmetical surfaces. Ann. Math. **118**, 241–275 (1983)
[Mi1] Milne, U.S.: Étale Cohomology. Princeton: Princeton University Press 1980
[Mi2] Milne, J.S.: Arithmetic Duality Theorems. (Perspect. Math. vol. I) Boston: Academic press 1986
[Mu] Mumford, D.: Abelian Varieties, Tata Institute of Fundamental Research Studies in Mathematics. Oxford: Oxford University Press 1970
[O] Ōkōchi, T.: On the Chow group of a conic bundle surface, Master's thesis. University of Tokyo 1984
[Q] Quillen, D.: Higher algebraic K-theory I. In: Bass, H. (ed.) Algebraic K-theory I. Proceedings, Seattle. (Lect. Notes Math., vol. 341) Berlin Heidelberg New York: Springer 1973
[R1] Raskind, W.: Algebraic K-theory, étale cohomology and torsion algebraic cycles. In: Stein, M.R. (ed.) Proceedings of the U.S.-Japan K-theory conference. Honolulu 1987. (Contemp. Math. vol. **83**, pp. 311–341) Providence: Am. Math. Soc. 1989
[R2] Raskind, W.: Torsion algebraic cycles on varieties over local fields. In: Jardine, J.F., Snaith, V.P. (ed.) Algebraic K-Theory: Connections with Geometry and Topology. Proceedings, Lake Louise 1987, pp. 343–388. Dordrecht Boston London: Kluwer Academic Publishers 1989
[R3] Raskind, W.: On K_1 of curves over global fields. Math. Ann. **288**, 179–193 (1990)
[Sa] Salberger, P.: Zero-cycles on rational surfaces over number fields. Invent. Math. **91**, 505–524 (1988)
[SGA1] Revêtements étales et groupe fondamental, Séminaire de Géométrie Algébrique du Bois Marie 1960/61, dirigé par A. Grothendieck. (Lect. Notes Math. vol. 224) Berlin Heidelberg New York: Springer 1971
[Sh1] Sherman, C.: K-cohomology of regular schemes. Commun. Algebra **7**, 999–1027 (1979)
[Sh2] Sherman, C.: Some theorems on the K-theory of coherent sheaves. Commun. Algebra **7**, 1489–1508 (1979)
[Sh3] Sherman, C.: K-theory of discrete valuation rings. J. Pure Appl. Algebra **61**, 79–98 (1989)
[So] Somekawa, M.: On Milnor K-groups attached to semi-abelian varieties. K-theory, **4**, 105–119 (1990)
[Su] Suslin, A.A.: Torsion in K_2 of fields. K-Theory **1**, 5–30 (1987)

Invent. math. 105, 247–272 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Intersection cohomology of cs-spaces and Zeeman's filtration

Nathan Habegger[1,*] **and Leslie Saper**[2,**]

[1] Département de Mathématiques, Université de Nantes, F-44072 Nantes, France and Department of Mathematics, University of Georgia, Athens, GA 30602, USA
[2] Department of Mathematics, Duke University, Durham, NC 27706, USA

Oblatum 10-XII-1990

0 Introduction

In this paper we generalize the notion of intersection homology in several directions.

The intersection homology of a singular space X was originally defined by Goresky and MacPherson simplicially [GM1] and then, using a construction suggested by Deligne, sheaf-theoretically [GM2]. It depends on the choice of an *ordinary perversity*, a function $p: \mathbb{Z}^+ \to \mathbb{Z}$ satisfying

$$p(0)=p(1)=p(2)=0, \tag{0.1}$$

and

$$p(k+1)=p(k) \quad \text{or} \quad p(k)+1 \qquad (k\in\mathbb{Z}^+). \tag{0.2}$$

The construction in [GM1] applied when X was a pl-pseudomanifold, while [GM2] considered topological pseudomanifolds (see § 1 for the definitions).

King [K] has shown using singular chains that a compactly supported intersection homology theory with trivial coefficients may be defined on the more general class of cs-spaces, that is, locally conelike topological stratified spaces (see § 1). Furthermore, he considers additional intersection homology theories associated to *generalized* perversities, ones that only satisfy (0.2). Finally, again using singular chains, Quinn [Q] has further extended intersection homology as a topological invariant of "manifold weakly stratified sets."

The viewpoint espoused in [GM2], that the intersection homology (or intersection cohomology as we prefer) is the hypercohomology of a certain complex of sheaves, viewed as an object of a derived category, has proved extraordinarily valuable. Consequently we extend in this paper the sheaf-theoretic definition

* Research supported in part by an NSF Postdoctoral Fellowship, a Yale University Junior Faculty Fellowship and NSF Grant DMS-88-02818
** Research supported in part by NSF Grants DMS-8504029, DMS-8705849, and DMS-8957216. The author is an Alfred P. Sloan Research Fellow.

of intersection cohomology to cs-spaces and prove topological invariance (that is, independence of cs-stratification). An advantage of working with a cs-space is that it possesses an intrinsic coarsest cs-stratification (due to Sullivan); generalizations of this are constructed in § 3 and will be essential to our proof of topological invariance.

As in [K], we will also consider generalized perversities. For an ordinary perversity, the intersection cohomology sheaves of [GM2] are determined by their restriction to an open dense submanifold whose complement has codimension $\geqq 2$. For a generalized perversity p, the intersection cohomology sheaves are determined by their restriction to an open subset (not necessarily a submanifold) whose complement has codimension $\geqq c_p$, where

$$c_p = \min\{k \in \mathbb{Z}^+ \mid p(k) \leqq k-2\} \cup \{\infty\}. \tag{0.3}$$

We formalize this notion in Definition 4.1 as a "codimension $\geq c$ intersection cohomology theory" (c-ICT), and it is in this possibly more general setting that we eventually prove topological invariance. Obviously the notion of "coefficients" must be extended for such theories; this is done in § 5.

Paralleling closely the exposition given by Borel [B, V.] for topological pseudomanifolds, we present in § 6 the main properties of intersection cohomology on cs-spaces, including topological invariance, duality and pairings. In particular, we find the dual theory and corresponding "cup product" associated to a generalized nonordinary perversity. This answers a question left open in [K].

As an application we extend McCrory's geometric analysis [M1, M2] of Zeeman's filtration on the homology of a triangulable space (that is, the filtration derived from the hypercohomology spectral sequence of the dualizing complex). Our main result here is that the hypercohomology filtration of the intersection cohomology of a cs-space equals the filtration induced by a particular decreasing sequence of generalized perversities. This is proved in § 8 after a brief note on the hypercohomology filtration in § 7.

In the Appendix we indicate how many of the results presented in [B, V.] concerning constructibility and duality of complexes of sheaves on topological pseudomanifolds extend to cs-spaces. We also present a proof of the fact that the dualizing complex on a cs-space is quasi-isomorphic to the classical sheaf complex of locally finite singular chains.

Acknowledgements. We would like to thank Clint McCrory who read an early preprint version of this work [HS] (essentially comprising §§ 2, 7, and 8) and made a number of helpful suggestions. Much of this research was carried out during visits by both authors to the University of California at San Diego; we would like to thank this institution for its hospitality.

Conventions and notation

In general we work in the setting and with the notation of [B, V.]. In particular, all sheaves are sheaves of R-modules, where R is a fixed commutative noetherian ring of finite cohomological dimension. The dimension of a topological space refers to cohomological dimension over R. All topological spaces are locally compact, locally completely paracompact, and of finite dimension.

The category of differential graded sheaves (DGS) or complexes of sheaves over a topological space X is denoted $\mathrm{DGS}(X)$; the associated derived category is $D(X)$. If we restrict to bounded complexes, we write $\mathrm{DGS}^b(X)$ and $D^b(X)$. A quasi-isomorphism between DGS is denoted as an equality.

For $\mathscr{S}^{\cdot} \in \mathrm{DGS}(X)$ and $s \in \mathbb{Z}$, one has $\mathscr{S}^{\cdot}[s]$ and $\tau_{\leqq s}\mathscr{S}^{\cdot} \in \mathrm{DGS}(X)$ (the *shift* and the *Deligne truncation*) defined by

$$(\mathscr{S}^{\cdot}[s])^i = \mathscr{S}^{i+s} \qquad (i \in \mathbb{Z}),$$

$$(\tau_{\leqq s}\mathscr{S}^{\cdot})^i = \begin{cases} \mathscr{S}^i & i < s, \\ \operatorname{Ker} \mathscr{S}^i \to \mathscr{S}^{i+1} & i = s, \\ 0 & i > s. \end{cases}$$

We denote the global sections of a sheaf $\mathscr{S}$ on X by $\Gamma(X, \mathscr{S})$ or, when the space is understood, by $\Gamma(\mathscr{S})$.

The Borel-Moore-Verdier dual of $\mathscr{S}^{\cdot}$ is denoted $\mathscr{D}_X^{\cdot}\mathscr{S}^{\cdot}$. The dualizing "sheaf" on X is $\mathscr{D}_X^{\cdot} = \mathscr{D}_X^{\cdot} R_X$, where R_X denotes the constant sheaf. Homology always refers to Borel-Moore homology (homology with closed supports) with coefficients in R, i.e., $H_i(X; R) = \mathbb{H}^{-i}(X; \mathscr{D}_X^{\cdot})$, where the right hand side denotes hypercohomology.

$\mathbb{Z}^+$ denotes the nonnegative integers. The subscript c is used in two ways, as an index of codimension and to indicate compact supports; the meaning should be clear from the context.

1 cs-Spaces and basic definitions

Definition 1.1 Let X be a topological space. We call X as cs-*space of dimension* n if there exists a filtration $\mathscr{X}$

$$X = X_n \supset X_{n-1} \supset \ldots \supset X_0 \supset X_{-1} = \emptyset$$

by closed subspaces such that

(i) The *stratum* $S_{n-k} = X_{n-k} \setminus X_{n-k-1}$ $(k \in \mathbb{Z}^+)$ is either empty or an $n-k$ dimensional topological manifold.
(ii) $S_n = X \setminus X_{n-1}$ is nonempty.
(iii) (*Local Normal Triviality*) For all $x \in S_{n-k}$ $(k \in \mathbb{Z}^+)$, there exists a neighborhood U of x in X (called a ($\mathscr{X}$-)*distinguished neighborhood*) and a homeomorphism

$$U = B \times \mathring{c}(L), \tag{1.1}$$

where B is an open ball of dimension $n-k$ and $\mathring{c}(L)$ is the open cone

$$\mathring{c}(L) = L \times [0, 1)/(x, 0) \sim (x', 0)$$
$$(\mathring{c}(\emptyset) = \text{ a point, denoted } v)$$

on a compact topological space L (called the *link*). L is assumed to have a filtration $\mathscr{L}$

$$L = L_{k-1} \supset L_{k-2} \supset \ldots \supset L_0 \supset L_{-1} = \emptyset$$

for which (1.1) becomes a filtered homeomorphism, i.e. $U \cap X_{n-j} = B \times \mathring{c}(L_{k-j-1})$ $(0 \leqq j \leqq k)$.

A filtration $\mathscr{X}$ as above is called a cs-*stratification* of X.

The $\mathscr{X}$-*depth* of X is the greatest difference in dimension between non-empty strata of $\mathscr{X}$. The *depth* of X is the least $\mathscr{X}$-depth, as $\mathscr{X}$ varies over all cs-stratification of X.

The above definition is due to Siebenmann [S] (in [S], X is called a "cs-set of finite formal dimension").

Definition 1.2 A *topological pseudomanifold* X with a *topological stratification* $\mathscr{X}$ is defined similarly to Definition 1.1, except that we also assume

(1) S_n is dense in X,
(2) $S_{n-1} = \emptyset$, and
(3) $(L, \mathscr{L})$ in Definition 1.1, (iii) is (inductively on dimension) a compact topological pseudomanifold of dimension $k-1$ with a topological stratification.

If we replace topological by pl throughout, we obtain the notion of a pl-*pseudomanifold* with pl-*stratification.*

It is clear that induction on dimension is well suited to local arguments on topological pseudomanifolds. The usual replacement for cs-spaces is induction on depth (e.g., see the appendix).

Notation 1.3 For any filtration $\mathscr{X}$ of a topological space X,

$$X = X_n \supset X_{n-1} \supset \ldots \supset X_0,$$

we will use the following notations:

$$\begin{aligned}
&U_k = X \setminus X_{n-k} \\
&S_{n-k} = X_{n-k} \setminus X_{n-k-1} = U_{k+1} \setminus U_k \\
&j_k\colon\ S_{n-k} \hookrightarrow U_{k+1} \\
&i_k\colon\ U_k \hookrightarrow U_{k+1} \qquad \text{(the inclusion maps)} \\
&f_x\colon\ \{x\} \hookrightarrow X
\end{aligned}$$

where $k \in \mathbb{Z}^+$ and $x \in X$. Similarly for $\mathscr{X}' = \{X'_{n-k}\}$, we write U'_k, etc.

We will need various notions of "constructibility" for DGS on topological spaces.

Definition 1.4 If X is a topological space and $\mathscr{S}^{\cdot} \in \mathrm{DGS}^b(X)$, $\mathscr{S}^{\cdot}$ is *cohomologically locally constant* (clc) if the sheaves $\mathscr{H}^{\cdot}(\mathscr{S}^{\cdot})$ are locally constant. If $\mathscr{X}$ is a filtration of X by closed subsets,

$$X = X_n \supset X_{n+1} \supset \ldots \supset X_0,$$

$\mathscr{S}^{\cdot}$ is $\mathscr{X}$-*cohomologically locally constant* ($\mathscr{X}$-clc) if the restriction $j_k^* \mathscr{S}^{\cdot}$ to each stratum is clc; $\mathscr{S}^{\cdot}$ is $\mathscr{X}$-*cohomologically constructible* ($\mathscr{X}$-cc) if it is $\mathscr{X}$-clc and the stalks of $\mathscr{H}^{\cdot}(\mathscr{S}^{\cdot})$ are finitely generated. If X is a cs-space, we shall say $\mathscr{S}^{\cdot}$ is cs-*constructible* if it is $\mathscr{X}$-cc for some cs-stratification $\mathscr{X}$.

Another condition on a DGS we shall use is *cohomological constructibility* (cc); it involves four conditions CC1–CC4, of which we shall only explicitly need the fourth (see [B, V., 3.3] for the full definition):

(1.2) CC4 (*Property* (P, Q) *of Wilder*). If $P \subset Q$ are open in X, $\bar{P} \subset Q$ and $\bar{P}$ compact, then the image of $\mathbb{H}_c^i(P; \mathscr{S}^\cdot)$ in $\mathbb{H}_c^i(Q; \mathscr{S}^\cdot)$ is finitely generated.

A variety of results on constructibility and duality for DGS on topological pseudomanifolds are collected in [B, V.]. In the appendix we indicate how to extend these to cs-spaces.

2 Intersection homology via geometric chains

In this section we motivate the concept of a c-ICT (§ 4) by extending the original construction of intersection homology on a pl-pseudomanifold (via geometric chains satisfying dimension restrictions on their intersection with singular strata; see [GM1] or [B, I.]) to generalized perversities. For simplicity we only consider trivial coefficients. As for ordinary cohomology the resulting sheaves can be characterized by a combination of vanishing and attaching conditions. However, we will find these conditions only make a difference on strata with codimension $\geqq c_p$.

Let X be a pl-pseudomanifold of dimension n with pl-stratification $\mathscr{X}$ and let p be a generalized perversity. For this section only, we assume

$$p(0) \geqq 0, \tag{2.1}$$

and so $c_p \geqq 2$; in fact this latter condition will later be forced upon us in order to acheive a nontrivial theory (Corollary 4.8).

Let $C_\cdot(X)$ denote the complex of locally finite geometric chains on X (with coefficients in R) [B, I., 2.1]. The *intersection chains of perversity* p are defined to be the subcomplex

$$\begin{aligned} I_p C_i(X) = \{\xi \in C_i(X) \mid & \dim(|\xi| \cap X_{n-k}) \leqq i-k+p(k), \\ & \dim(|\partial \xi| \cap X_{n-k}) \leqq i-1-k+p(k), k \in \mathbb{Z}^+\} \quad (i \in \mathbb{Z}). \end{aligned} \tag{2.2}$$

The presheaf $U \mapsto I_p C_\cdot(U)$ (resp. $U \mapsto C_\cdot(U)$) is actually a sheaf, denoted $\mathscr{I}_p \mathscr{C}_\cdot$ (resp. $\mathscr{C}_\cdot$), which is soft [B, II., § 5]. Define the *intersection homology*

$$I_p H_{n-i}(X; R) = \mathbb{H}^i(X; \mathscr{I}_p \mathscr{C}_{n-\cdot}) = H_{n-i}(I_p C_\cdot(X)) \quad (i \in \mathbb{Z}). \tag{2.3}$$

We will see later that this is independent of $\mathscr{X}$ and the pl-structure on X.

Theorem 6.1 of [B, II.] admits the following extension to generalized perversities:

Theorem 2.1 *The* DGS $\mathscr{S}^\cdot = \mathscr{I}_p \mathscr{C}_{n-\cdot}$ *satisfies:*

(1) $\mathscr{S}^\cdot$ *is a bounded complex, zero in negative degrees, and* $\mathscr{S}^\cdot|_{U_{c_p}}$ *is quasi-isomorphic to* $\mathscr{D}^\cdot_{U_{c_p}}[-n]$.

(2) $H^i(\mathscr{S}^\cdot_x) = 0$ *for* $x \in S_{n-k}$, $i > p(k)$, *and* $k \geqq c_p$.

(3) *The attachment map* $\mathscr{S}^\cdot|_{U_{k+1}} \to R i_{k*}(\mathscr{S}^\cdot|_{U_k})$ *is a quasi-isomorphism up to degree* $p(k)$ *for* $k \geqq c_p$.

Remark. If p is an ordinary perversity, then $c_p = 2$ and Theorem 2.1 reduces to Theorem 6.1 of [B, II.] since $\mathscr{D}^\cdot_{U_2}[-n]$ is just (a resolution of) the orientation sheaf. (N.B.: $\mathscr{D}^\cdot_X[-n] = \mathscr{C}_{n-\cdot}$.)

Remark. If $c_p \geqq n+1$, then Theorem 2.1 says that $\mathscr{I}_p\mathscr{C}_{n-\cdot}$ is quasi-isomorphic to $\mathscr{D}_X^{\cdot}[-n]$, so $I_pH_\cdot(X;R) = H_\cdot(X;R)$.

Proof. The proof is analogous to the proof of Theorem 6.1 in [B, II.]. We must first, however, modify the discussion of coning in [B, II., § 3] to deal with dimension zero cycles, since now it is possible that $k-p(k) \leqq 1$. Note that the cone on a zero dimensional cycle is only a cycle provided it is reduced (augments to zero). Thus the coning map $\xi \mapsto \mathring{c}\xi$, from chains on a space L to chains on $\mathring{c}(L)$, is only a map of complexes provided we use *reduced* chains on L. Thus for a generalized perversity p, Proposition 3.1 of [B, II.] should read (for $L \neq \emptyset$ a stratified pl-pseudomanifold of dimension $k-1$):

$$I_pH_i(\mathring{c}(L);R) = \begin{cases} \widetilde{I_pH}_{i-1}(L;R) & i \geqq k-p(k), \\ 0 & i < k-p(k). \end{cases} \tag{2.4}$$

Now for $x \in S_{n-k}$ $(k>0)$ with link L, (2.4) and a Künneth type argument [B, II., § 4] imply

$$H^i((\mathscr{I}_p\mathscr{C}_{n-\cdot})_x) = \begin{cases} \widetilde{I_pH}_{k-i-1}(L;R) & i \leqq p(k), \\ 0 & i > p(k), \end{cases} \tag{2.5}$$

whereas

$$H^i(\mathscr{D}_{U_{c_p}}^{\cdot}[-n]_x) = H^i((\mathscr{C}_{n-\cdot})_x) = \tilde{H}_{k-i-1}(L;R) \quad (i \in \mathbb{Z}). \tag{2.6}$$

For $k \geqq c_p$, $k-p(k)>1$. Thus $\widetilde{I_pH}$ in (2.5) may be replaced by I_pH since $k-i-1 \geqq k-p(k)-1$. This implies (2) and (3) as in [B, II.]. For (1), the required quasi-isomorphism is clear over U_2 by (2.1) and (2.2). So suppose inductively we have a quasi-isomorphism over U_k for some $k<c_p$. To extend this to U_{k+1}, we need to show the natural map between (2.5) and (2.6) is an isomorphism. But for $k<c_p$, $k-p(k) \leqq 1$ and thus (2.5) reduces to

$$H^i((\mathscr{I}_p\mathscr{C}_{n-\cdot})_x) = \widetilde{I_pH}_{k-i-1}(L;R) \quad (i \in \mathbb{Z}).$$

By our inductive assumption, the map $\widetilde{I_pH}_{k-i-1}(L;R) \to \tilde{H}_{k-i-1}(L;R)$ is an isomorphism, which completes the proof. □

It is clear from this theorem that $\mathscr{I}_p\mathscr{C}_{n-\cdot}$ may be constructed from $\mathscr{D}_{U_{c_p}}^{\cdot}[-n]$ by applying $\tau_{\leqq p(k)}Ri_{k*}$ successively for $k=c_p, \ldots, n$, that is, by a Deligne construction.

3 cs-Stratifications

Let X be a cs space of dimension n. In this section we construct various cs-stratifications of X associated to a DGS $\mathscr{S}^{\cdot}$ defined on an open set $\mathrm{Dom}(\mathscr{S}^{\cdot}) \subset X$. The constructions are inspired by Sullivan's construction of an intrinsic coarsest cs-stratification $\mathscr{X}^*$ of X [K, § 1]; we indicate below when in fact they reduce to $\mathscr{X}^*$. These cs-stratifications will be important for our proof of the topological invariance of intersection cohomology. If one is only interested in trivial coefficients, though, it suffices to use $\mathscr{X}^*$ and topological invariance becomes considerably simpler.

Definition 3.1 Let $\mathscr{X}$ and $\mathscr{X}'$ be cs-stratifications of X. $\mathscr{X}'$ is *coarser* than $\mathscr{X}$ (denoted $\mathscr{X}' > \mathscr{X}$) if each stratum of $\mathscr{X}'$ is a union of connected components of stata of $\mathscr{X}$.

If $\mathscr{X}' > \mathscr{X}$ and $\mathscr{X} > \mathscr{X}'$ then $\mathscr{X}' = \mathscr{X}$. If $\mathscr{X}' > \mathscr{X}$ then $U_k' \supset U_k$ and "$\mathscr{X}'$-clc$\Rightarrow\mathscr{X}$-clc". We will use these facts without reference.

Let $c \in \mathbb{Z}^+$ and let $\mathscr{S}^\cdot$ be as above. Define a symmetric and reflexive relation R_c by

$$x R_c y$$

if and only if there exists a homeomorphism

$$f\colon (U_x, x) \to (U_y, y) \tag{3.1}$$

of neighborhoods of x and y and a quasi-isomorphism

$$\mathscr{S}^\cdot|_V = (f^* \mathscr{S}^\cdot)|_V \tag{3.2}$$

for some open subset V of $U_x \cap \mathrm{Dom}(\mathscr{S}^\cdot) \cap f^{-1}(U_y \cap \mathrm{Dom}(\mathscr{S}^\cdot))$ whose complement has codimension $\geqq c$. Let $\sim_c$ denote the induced equivalence relation. If we need to specify the dependence on $\mathscr{S}^\cdot$, we will write $R_{c,\mathscr{S}^\cdot}$ and $\sim_{c,\mathscr{S}^\cdot}$. Note that if the complement of $\mathrm{Dom}(\mathscr{S}^\cdot)$ has codimension $\geqq c$, then R_c is already an equivalence relation (i.e., it is transitive). Also note that

$$x \sim_{n+1,\mathscr{S}^\cdot} y \Rightarrow x \sim_{c,\mathscr{S}^\cdot} y \Rightarrow x \sim_{c,\mathscr{S}^\cdot|_U} y \tag{3.3}$$

for $U \subset \mathrm{Dom}(\mathscr{S}^\cdot)$ open.

Denote the equivalence class of $x \in X$ relative to $\sim_c$ by $[x]$. Define

$$S_i' = \bigcup_{\dim[x]=i} [x],$$

$$X_i' = \bigcup_{j \leqq i} S_j',$$

and

$$U_k' = X \setminus X_{n-k}'.$$

The filtration $\{X_i'\}$ is denoted $\mathscr{X}_c(\mathscr{S}^\cdot)$; when $c \geqq n+1$ we simply write $\mathscr{X}(\mathscr{S}^\cdot)$. It clearly depends only on the quasi-isomorphism class of $\mathscr{S}^\cdot$.

Proposition 3.2 *Let $\mathscr{S}^\cdot \in \mathrm{DGS}(\mathrm{Dom}(\mathscr{S}^\cdot))$ and let $c \in \mathbb{Z}^+$.*

(1) *If $x \sim_c y$ there is a $\mathscr{X}_c(\mathscr{S}^\cdot)$-filtered homeomorphism of neighborhoods of x and y. (Namely, take the composition of the homeomorphisms from* (3.1) *that define a chain $x R_c x_1 \ldots x_N R_c y$.)*

(2) *Assume that R_c is transitive (for example, say that $X \setminus \mathrm{Dom}(\mathscr{S}^\cdot)$ has codimension $\geqq c$). Then $\mathscr{X}_c(\mathscr{S}^\cdot)$ is defined locally; that is, if $U \subset X$ is open, then*

$$\mathscr{X}_c(\mathscr{S}^\cdot)|_U = \mathscr{X}_{c,U}(\mathscr{S}^\cdot|_{U \cap \mathrm{Dom}(\mathscr{S}^\cdot)}),$$

where $\mathscr{X}_{c,U}$ denotes the above construction carried out in the cs-*space U.*

(3) *$\mathscr{X}_c(\mathscr{S}^\cdot)$ depends on $\mathscr{S}^\cdot$ only off a set of codimension $\geqq c$, that is, if $\mathscr{S}^\cdot$ and $\mathscr{S}'^\cdot$ are quasi-isomorphic on an open set $U \subset \mathrm{Dom}(\mathscr{S}^\cdot) \cap \mathrm{Dom}(\mathscr{S}'^\cdot)$ such that both $\mathrm{Dom}(\mathscr{S}^\cdot) \setminus U$ and $\mathrm{Dom}(\mathscr{S}'^\cdot) \setminus U$ have codimension $\geqq c$, then $\mathscr{X}_c(\mathscr{S}^\cdot) = \mathscr{X}_c(\mathscr{S}'^\cdot)$.*

The proofs are easy and left to the reader.

The filtration $\mathscr{X}_c(\mathscr{S}^\cdot)$ is not a cs-stratification in general; in fact it is not necessarily even a filtration by closed sets. However, if (for example) $\mathscr{S}^\cdot = R_{\mathrm{Dom}(\mathscr{S}^\cdot)}$ or $\mathscr{D}_{\mathrm{Dom}(\mathscr{S}^\cdot)}$, or else $c=0$, then condition (3.2) is vacuous and $\mathscr{X}_c(\mathscr{S}^\cdot)$ reduces to the coarsest cs-stratification $\mathscr{X}^*$ of X. More generally we have:

Theorem 3.3 *Let $c\in\mathbb{Z}^+$. Assume $\mathscr{S}^\cdot\in\mathrm{DGS}(\mathrm{Dom}(\mathscr{S}^\cdot))$ ($\mathrm{Dom}(\mathscr{S}^\cdot)\subset X$ open) extends to $\tilde{\mathscr{S}}^\cdot\in\mathrm{DGS}(X)$ which is $\mathscr{X}$-clc for some* cs*-stratification $\mathscr{X}=\{X_i\}$ of X. Then $\mathscr{X}_c(\mathscr{S}^\cdot)$ is a* cs*-stratification. Furthermore we have:*

(1) $\mathscr{X}_c(\mathscr{S}^\cdot)>\mathscr{X}$;
(2) $\mathscr{X}_c(\mathscr{S}^\cdot)=\mathscr{X}(\mathscr{S}^\cdot)$ *on* U_c *if* $U_c\subset\mathrm{Dom}(\mathscr{S}^\cdot)$;
(3) $\mathscr{S}^\cdot$ *is* $\mathscr{X}(\mathscr{S}^\cdot)$-clc *on* U_c *if* $U_c\subset\mathrm{Dom}(\mathscr{S}^\cdot)$.

Corollary 3.4 *Let $\mathscr{S}^\cdot\in\mathrm{DGS}(X)$ be $\mathscr{X}$-*clc *for some* cs*-stratifiction $\mathscr{X}$. Then $\mathscr{X}(\mathscr{S}^\cdot)$ is the unique coarsest* cs*-stratification for which $\mathscr{S}^\cdot$ is $\mathscr{X}(\mathscr{S}^\cdot)$-*clc.

We begin with some lemmas.

Lemma 3.5 *Let $\mathscr{S}^\cdot$ be as in Theorem 3.3. The equivalence classes of $\sim_c$ consist of unions of connected components of strata of $\mathscr{X}$.*

Proof. Since an equivalence class of $\sim_{c,\mathscr{S}^\cdot}$ is a union of equivalence classes for $\sim_{n+1,\tilde{\mathscr{S}}^\cdot}$, it suffices to assume that $c=n+1$ and that $\mathscr{S}^\cdot=\tilde{\mathscr{S}}^\cdot$ is defined on all of X.

Let $N=B\times\mathring{c}(L)$ be a $\mathscr{X}$-distinguished neighborhood of a point $x\in X$. We claim that $xR_c\,y$ for $y\in B\times\{v\}$. Consequently, within a $\mathscr{X}$-stratum the equivalence classes of $\sim_c$ are open and closed sets, which proves the lemma.

To see the claim, let $\tau\colon N\to\mathring{c}(L)$ be the projection. By Corollary A.2, $\mathscr{S}^\cdot=\tau^*\mathscr{S}'^\cdot$ on N for some DGS $\mathscr{S}'^\cdot$ on $\mathring{c}(L)$. Let $f\colon (N,x)\to(N,y)$ be a homeomorphism such that $\tau\circ f=\tau$; then $f^*\mathscr{S}^\cdot=f^*\tau^*\mathscr{S}'^\cdot=\tau^*\mathscr{S}'^\cdot=\mathscr{S}^\cdot$. □

Corollary 3.6 *Assume* $\dim[x]=i$ *for* $x\in X$. *Then* $[x]\cap S_i\neq\emptyset$ *and* $[x]\cap S_j=\emptyset$, $j>i$.

Proof. By the lemma, $\dim[x]=\max_{[x]\cap S_i\neq\emptyset} i$. □

Lemma 3.7 *Let $\mathscr{S}^\cdot$ be as in Theorem 3.3. Then $\mathscr{X}_c(\mathscr{S}^\cdot)$ is a* cs*-stratification and the equivalence classes of $\sim_c$ are unions of connected components of strata of $\mathscr{X}_c(\mathscr{S}^\cdot)$.*

Proof. It suffices to verify this locally near a point $x\in S'_{n-k}$; since x is equivalent to a point in S_{n-k} by Corollary 3.6, we can in fact assume $x\in S'_{n-k}\cap S_{n-k}$ by Proposition 3.2, (1). Let $U_x=B\times\mathring{c}(L)$ be a $\mathscr{X}$-distinguished neighborhood of x. By Lemma 3.5 and Corollary 3.6,

$$X'_{n-j}\cap U_x=\begin{cases}\emptyset & j>k,\\ B\times\{v\} & j=k,\\ B\times\mathring{c}(L'_{k-j-1}) & j<k,\end{cases}$$

where L'_{k-j-1} is a union of connected components of strata of L. These three cases show, respectively, that $\mathscr{X}_c(\mathscr{S}^\cdot)$ is a filtration by closed subsets, that the strata are manifolds, and that $\mathscr{X}_c(\mathscr{S}^\cdot)$ is locally conelike (an $\mathscr{X}_c(\mathscr{S}^\cdot)$-distinguished neighborhood of x is given by $U_x=B\times\mathring{c}(L)$, where L now has the filtration

$\{L'_{k-j-1}\}$). Finally we see $[x] \cap U_x = S'_{n-k} \cap U_x$, so that $[x]$ is open in S'_{n-k}; this proves the final statement of the lemma. □

Lemma 3.8 *Let $\sim$ and $\sim'$ be equivalence relations defining* cs-*stratifications $\mathscr{X}$ and $\mathscr{X}'$ as in this section. If $\sim$ is coarser than $\sim'$, then $\mathscr{X} > \mathscr{X}'$. In particular (see* (3.3)*),*

$$\mathscr{X}_c(\mathscr{S}^\cdot|_U) > \mathscr{X}_c(\mathscr{S}^\cdot) > \mathscr{X}(\mathscr{S}^\cdot)$$

for $\mathscr{S}^\cdot$ as in Theorem 3.3 and $U \subset \mathrm{Dom}(\mathscr{S}^\cdot)$ open.

Proof. The strata of $\mathscr{X}$ are unions of equivalence classes of $\sim$ by definition, the equivalence classes of $\sim$ are unions of equivalence classes of $\sim'$ since it is a coarser relation, and the equivalence classes of $\sim'$ are unions of connected components of strata of $\mathscr{X}'$ by Lemma 3.7. □

Proof of Theorem 3.3

$\mathscr{X}_c(\mathscr{S}^\cdot)$ is a cs-stratification by Lemma 3.7 and (1) follows from Lemma 3.5.

We now prove (2). Let $x \in U_c$ and let S' and S'' be the local strata through x for $\mathscr{X}_c(\mathscr{S}^\cdot)$ and $\mathscr{X}(\mathscr{S}^\cdot)$ respectively. By Lemma 3.8, we have $S'' \subset S'$. Suppose $S'' \neq S'$; then $x \sim_c y$ for some $y \in S' \setminus S''$. Thus there is a homeomorphism $f\colon (U_x, x) \to (U_y, y)$ with $\mathscr{S}^\cdot = f^* \mathscr{S}^\cdot$ off a codimension $\geqq c$ subset of U_x (recall we assume in (3) that $U_c \subset \mathrm{Dom}(\mathscr{S}^\cdot)$, so that R_c is transitive). By shrinking U_x if necessary, we may assume $U_y \cap S'' = \emptyset$. Since $x \in U_c$ and $\mathscr{X}(\mathscr{S}^\cdot) > \mathscr{X}$ (by (1)), S'' has codimension $< c$ and so $\mathscr{S}^\cdot = f^* \mathscr{S}^\cdot$ near some point $z \in S''$. But then $z \sim_{n+1} f(z) \notin S''$, which is a contradiction. Thus $S'' = S'$ which proves (2).

Finally we consider (3). Since we assume $U_c \subset \mathrm{Dom}(\mathscr{S}^\cdot)$, (2) implies that $\mathscr{X}_c(\mathscr{S}^\cdot) = \mathscr{X}(\mathscr{S}^\cdot)$ and $\mathscr{X}_c(\tilde{\mathscr{S}}^\cdot) = \mathscr{X}(\tilde{\mathscr{S}}^\cdot)$ on U_c. However $\mathscr{X}_c(\mathscr{S}^\cdot) = \mathscr{X}_c(\tilde{\mathscr{S}}^\cdot)$ by Proposition 3.2, (3), so $\mathscr{X}(\mathscr{S}^\cdot) = \mathscr{X}(\tilde{\mathscr{S}}^\cdot)$ on U_c. Consequently for (3) it suffices to assume $\mathrm{Dom}(\mathscr{S}^\cdot) = X$. Let $x \in S''_{n-k}$, a stratum of $\mathscr{X}(\mathscr{S}^\cdot)$. By Corollary 3.6, $x \sim_{n+1} y$ for some $y \in S_{n-k}$. So there is a homeomorphism $f\colon (U_x, x) \to (U_y, y)$ with $\mathscr{S}^\cdot = f^* \mathscr{S}^\cdot$ on U_x. Shrinking U_x if necessary, we can assume $S''_{n-k} \cap U_x = f^{-1}(S_{n-k} \cap U_y)$. Since $\mathscr{S}^\cdot$ is clc on $S_{n-k} \cap U_y$, $f^* \mathscr{S}^\cdot$ and thus $\mathscr{S}^\cdot$ is clc on $S''_{n-k} \cap U_x$. □

We note the following application of our theorem (which will not be needed in the sequel).

Corollary 3.9 *Let $\mathscr{S}^\cdot \in \mathrm{DGS}(X)$. Assume that for all $x \in X$, there exists a neighborhood U of x and a* cs-*stratification $\mathscr{X}$ of U such that $\mathscr{S}^\cdot|_U$ is $\mathscr{X}$-*clc*. Then $\mathscr{S}^\cdot$ is $\mathscr{X}$-*clc *for some* cs-*stratification $\mathscr{X}$ of X.*

Proof. Locally we can apply Theorem 3.3 to see that $\mathscr{X}_U(\mathscr{S}^\cdot|_U)$ is a cs-stratification and that $\mathscr{S}^\cdot|_U$ is $\mathscr{X}_U(\mathscr{S}^\cdot|_U)$-clc. By Proposition 3.2, (2), $\mathscr{X}(\mathscr{S}^\cdot)$ is the desired global cs-stratification. □

4 Codimension $\geqq c$ intersection cohomology theory

In order to emphasize what will be necessary for topological invariance (that is, independence of stratification), we abstract certain facts about intersection cohomology into the following definition. Let $c \in \mathbb{Z}^+$.

Definition 4.1 A *codimension $\geqq c$ intersection cohomology theory* (a c-ICT for short) is a set of axioms A for elements of $D^b(X)$, X any cs-space, which has

the following properties. Let $U \subset X$ be open and let $\mathscr{X}$ be a cs-stratification of X.

(i) *Naturality:* If $\mathscr{S}^{\cdot} \in D^b(X)$ satifies A on X, then $\mathscr{S}^{\cdot}|_U$ satisfies A on U.

(ii) *Extension:* If $U \supset U_c$ and $\mathscr{S}^{\cdot} \in D^b(U)$ is $\mathscr{X}$-clc (resp. $\mathscr{X}$-cc) and satisfies A on U, then there exists $\tilde{\mathscr{S}}^{\cdot} \in D^b(X)$, $\mathscr{X}$-clc (resp. $\mathscr{X}$-cc) and satisfying A on X, such that $\tilde{\mathscr{S}}^{\cdot}|_U = \mathscr{S}^{\cdot}$.

(iii) *Unique Morphism Extension:* If $U \supset U_c$ and $\mathscr{S}^{\cdot}$, $\mathscr{S}'^{\cdot} \in D^b(X)$ are χ-clc and satisfy A on X, then any morphism $\mathscr{S}^{\cdot}|_U \to \mathscr{S}'^{\cdot}|_U$ in $D^b(U)$ extends uniquely to a morphism $\mathscr{S}^{\cdot} \to \mathscr{S}'^{\cdot}$ in $D^b(X)$.

Remark 4.2 It follows from Definition 4.1, (iii) that the extension $\tilde{\mathscr{S}}^{\cdot}$ guaranteed by Definition 4.1, (ii) is unique up to a canonical isomorphism in $D^b(X)$.

As the notation suggests, a c-ICT is also a c'-ICT for $c' > c$. If A and A' are c-ICT and A' implies A, we say A' is *contained* in A ($A' \subset A$); in this case, A' merely restricts the allowable "coefficients" on U_c (see § 5), but does not change how they extend to X.

Definition 4.3 Let p be a generalized perversity and define

$$q(k) = k - 2 - p(k),$$
$$p^{-1}(i) = \min\{k \in \mathbb{Z}^+ \mid p(k) \geqq i\} \cup \{\infty\},$$

and

$$c_p = \min\{k \in \mathbb{Z}^+ \mid p(k) \leqq k-2\} \cup \{\infty\} = q^{-1}(0).$$

Let X be a cs-space of dimension n and $\mathscr{S}^{\cdot} \in D^b(X)$. We say that $\mathscr{S}^{\cdot}$ satisfies *the axioms* A_p if the *support condition*

$$\dim\{x \in X \mid H^i(\mathscr{S}^{\cdot}_x) \neq 0\} \leqq n - p^{-1}(i), \qquad i > c_p - 2, \tag{4.1}$$

and the *cosupport condition*

$$\dim\{x \in X \mid H^i(f_x^! \mathscr{S}^{\cdot}) \neq 0\} \leqq n - q^{-1}(n-i), \qquad i < n, \tag{4.2}$$

both hold. Here $f_x : \{x\} \to X$ is the inclusion map for $x \in X$. The sets in (4.1) and (4.2) are called the i^{th}-*support* and the i^{th}-*cosupport* of $\mathscr{S}^{\cdot}$ respectively.

Remark 4.4 It is easy to see that if p and p' are generalized perversities with $c_p = c_{p'}$ and $p(k) = p'(k)$ for $k \geqq c_p$, then $A_p = A_{p'}$.

We will see below that A_p is a c_p-ICT. In fact, all examples of c-ICT that we know of are contained in those generated from A_p by the next proposition. It is conceivable that these are all.

Proposition 4.5 (1) *A_p is a c_p-*ICT *for p a generalized perversity.*
(2) *Let $s \in \mathbb{Z}$. If A is a c-*ICT*, then so is $A[s]$, where $\mathscr{S}^{\cdot}$ satisfies $A[s]$ if and only if $\mathscr{S}^{\cdot}[-s]$ satisfies A.*
(3) *If A is a c-*ICT*, then so is $\mathscr{D}A$, provided we restrict to* cs-*constructible* DGS*. Here $\mathscr{S}^{\cdot}$ satisfies $\mathscr{D}A$ on a* cs-*space X if and only if $\mathscr{D}_X \mathscr{S}^{\cdot}$ satisfies A on X.*

We begin with two important lemmas.

Lemma 4.6 *Let X be a* cs-*space of dimension n. Assume $\mathscr{S}^{\cdot} \in \mathrm{DGS}(X)$ is $\mathscr{X}$-*clc *for some* cs-*stratification $\mathscr{X}$. Let p be a generalized perversity.*

(i) *The support condition* (4.1) *is equivalent to the* vanishing condition

$$H^i(\mathscr{S}^{\cdot}_x)=0, \quad \text{for } x\in S_{n-k}, \quad i>\max(p(k), c_p-2), \tag{4.3}$$

for all $k\in\mathbb{Z}^+$.

(ii) *The cosupport conditon* (4.2) *is equivalent to the* covanishing condition

$$H^i(f_x^! \mathscr{S}^{\cdot})=0, \quad \text{for } x\in S_{n-k}, \quad i<n-\max(q(k), 0), \tag{4.4}$$

for all $k\in\mathbb{Z}^+$.

(iii) *Assume the vanishing condition* (4.3) *holds. Then the covanishing condition* (4.4) *is equivalent, when* $k\geqq c_p$, *to the* attaching condition

(4.5) *the attachment map* $\mathscr{S}^{\cdot}|_{U_{k+1}} \to Ri_{k*}(\mathscr{S}^{\cdot}|_{U_k})$ *is a quasi-isomorphism up to degree* $p(k)$.

(iv) *The covanishing condition* (4.4) *is equivalent, when* $k<c_p$, *to*

(4.6) *the attachment map* $\mathscr{S}^{\cdot}|_{U_{k+1}} \to Ri_{k*}(\mathscr{S}^{\cdot}|_{U_k})$ *is a quasi-isomorphism up to degree* $k-2$ *and an injection on cohomology sheaves in degree* $k-1$.

Proof. By Proposition A.6 and [B, V., 3.10(d)], $j_k^! \mathscr{S}^{\cdot}$ is clc on S_{n-k} for $k\in\mathbb{Z}^+$. Then (i) and (ii) (resp. (iii) and (iv)) follow by an argument similar to [B, V., 4.9] (resp. [B, V., 4.3]). □

Lemma 4.7 *Let* Y *be a topological space and* $i: U \hookrightarrow Y$ *the inclusion of an open subset. Let* $\mathscr{A}^{\cdot}, \mathscr{B}^{\cdot}\in D(Y)$ *and* $m\in\mathbb{Z}$. *Assume that*

$$H^i(\mathscr{A}^{\cdot}_x)=0 \quad \text{for } i>m, \quad x\in Y,$$

and that

the attachment map $\mathscr{B}^{\cdot} \to Ri_* i^* \mathscr{B}^{\cdot}$ *is a quasi-isomorphism up to degree* m.

Then restriction induces an isomorphism

$$\mathrm{Mor}_{D(Y)}(\mathscr{A}^{\cdot}, \mathscr{B}^{\cdot}) \xrightarrow{\sim} \mathrm{Mor}_{D(U)}(i^* \mathscr{A}^{\cdot}, i^* \mathscr{B}^{\cdot}),$$

that is, every morphism $\mathscr{A}^{\cdot}|_U \to \mathscr{B}^{\cdot}|_U$ *extends uniquely to a morphism* $\mathscr{A}^{\cdot} \to \mathscr{B}^{\cdot}$.

Proof. Apply [B, V., 9.1]. □

Proof of Proposition 4.5 We first consider (1). Definition 4.1, (i) is obvious for A_p. In verifying Definition 4.1, (iii), we may assume that $U=U_{c_p}$. By Lemma 4.6, $\mathscr{S}^{\cdot}$ and $\mathscr{S}'^{\cdot}$ satisfy (4.3) for $k\geqq 0$ and (4.5) for $k\geqq c_p$. Thus we can apply Lemma 4.7 repeatedly, stratum by stratum, to see that a morphism extends uniquely from U_{c_p} to X. We now consider Definition 4.1, (ii). Define $\tilde{\mathscr{S}}^{\cdot}$ via a Deligne construction as

$$\tilde{\mathscr{S}}^{\cdot} = \tau_{\leqq p(n)} Ri_{n*} \dots \tau_{\leqq p(c_p)} Ri_{c_p*}(\mathscr{S}^{\cdot}|_{U_{c_p}}). \tag{4.7}$$

By Lemma 4.6, $\mathscr{S}^{\cdot}|_{U_{c_p}}$ (and therefore $\tilde{\mathscr{S}}^{\cdot}$) satisfies (4.3) and (4.4) for $k<c_p$. On the other hand, it is clear by construction that $\tilde{\mathscr{S}}^{\cdot}$ satisfies (4.3) and (4.5) for $k\geqq c_p$. By Proposition A.6 and [B, V., 3.9(b)] (resp. [B, V., 3.11(iii)]) $\tilde{\mathscr{S}}^{\cdot}$ is

$\mathscr{X}$-clc (resp. $\mathscr{X}$-cc) if $\mathscr{S}^{\cdot}$ is $\mathscr{X}$-clc (resp. $\mathscr{X}$-cc). Thus by Lemma 4.6, $\tilde{\mathscr{S}}^{\cdot}$ satisfies A_p. Finally we see by Definition 4.1, (iii) that $\tilde{\mathscr{S}}^{\cdot}$, *a priori* an extension of $\mathscr{S}^{\cdot}|_{U_{c_p}}$, is in fact an extension of $\mathscr{S}^{\cdot}$. This concludes the proof of (1).

Assertion (2) is clear, while (3) follows from the equality $\mathscr{D}_X^{\cdot}\mathscr{D}_X^{\cdot}\mathscr{S}^{\cdot}=\mathscr{S}^{\cdot}$ for cs-constructible DGS (Proposition A.7 and [B, V., 8.10]) and the fact that $\mathscr{D}_X^{\cdot}\mathscr{S}^{\cdot}$ is $\mathscr{X}$-clc (resp. $\mathscr{X}$-cc) whenever $\mathscr{S}^{\cdot}$ is (Proposition A.7 and [B, V., 8.7]). □

Remark. By the same reasoning one can show that A_p satifies Definition 4.1, (ii) with $U=U_c$ for $c<c_p$ (one extends from U_c to U_{c_p} by $\tau_{\leqq c_p-2}Ri_{c_p-1*}\dots Ri_{c*}$). It is the failure of Definition 4.1, (iii) that prevents A_p from being a c-ICT for $c<c_p$.

We have the following corollary of the proof.

Corollary 4.8 *If p is a generalized perversity with $c_p<2$ (in particular if $p(2)<0$), and $\mathscr{S}^{\cdot}\in D^b(X)$ is $\mathscr{X}$-clc and satisfies A_p, then $\mathscr{S}^{\cdot}=0$.*

Proof. For $c_p=0$, this is an immediate consequence of Definition 4.1, (iii) with $U=U_{c_p}=\emptyset$ and $\mathscr{S}^{\cdot}=\mathscr{S}'^{\cdot}$. Similarly for $c_p=1$, we only need to show that $\mathscr{S}^{\cdot}|_{U_1}=0$ if $\mathscr{S}^{\cdot}$ is $\mathscr{X}$-clc and satisfies A_p. Since $H^i(j_x^!\mathscr{S}^{\cdot})=H^{i-n}(\mathscr{S}_x^{\cdot})$ for $x\in U_1$ [B, V., 3.7(b)], this follows from (4.3) and (4.4). □

The primary result of this section is the following strengthening of Definition 4.1, (iii) in the case of an isomorphism.

Theorem 4.9 *Let A be a c-*ICT *and let X be a* cs-*space with* cs-*stratifications $\mathscr{X}$ and $\mathscr{X}'$. Suppose $\mathscr{S}^{\cdot}$, $\mathscr{S}'^{\cdot}\in D^b(X)$ are $\mathscr{X}$-*clc *and* $\mathscr{X}'$-*clc *respectively and satisfy A on X. Then any isomorphism*

$$\mathscr{S}^{\cdot}|_{U_c\cap U_c'}=\mathscr{S}'^{\cdot}|_{U_c\cap U_c'} \tag{4.8}$$

in $D^b(U_c\cap U_c')$ extends uniquely to an isomorphism

$$\mathscr{S}^{\cdot}=\mathscr{S}'^{\cdot}$$

in $D^b(X)$.

We first prove a lemma concerning the cs-stratifications constructed in § 3.

Lemma 4.10 *If $\mathscr{S}^{\cdot}\in D^b(X)$ is $\mathscr{X}$-*clc *and satisfies A on X, then $\mathscr{X}(\mathscr{S}^{\cdot}|_{U_c})=\mathscr{X}(\mathscr{S}^{\cdot})$.*

Proof. Since $\mathscr{X}'=\mathscr{X}(\mathscr{S}^{\cdot}|_{U_c})>\mathscr{X}(\mathscr{S}^{\cdot})$ by Lemma 3.8 and $\mathscr{X}(\mathscr{S}^{\cdot})$ is the coarsest cs-stratification for which $\mathscr{S}^{\cdot}$ is $\mathscr{X}(\mathscr{S}^{\cdot})$-clc by Corollary 3.4, it suffices to show $\mathscr{S}^{\cdot}$ is $\mathscr{X}'$-clc.

By Theorem 3.3, (3), $\mathscr{S}^{\cdot}$ is $\mathscr{X}$-clc on U_c; we will show inductively that $\mathscr{S}^{\cdot}$ is $\mathscr{X}'$-clc on U_k, for all k. Assume this is true on U_{k-1} and let $x\in S_{n-k}$. Let R and $\sim$ denote the relations of § 3 used in constructing $\mathscr{X}'=\mathscr{X}(\mathscr{S}^{\cdot}|_{U_c})$. Since $\mathscr{X}'>\mathscr{X}$, $x\in S'_{n-k'}$ for some $k'\leqq k$ and therefore by Corollary 3.6, $x\sim y$ for some $y\in S'_{n-k'}\cap S_{n-k'}$.

We first claim that $\mathscr{S}^{\cdot}$ is $\mathscr{X}'$-clc near y. For $k'<k$ this follows from the inductive hypothesis, so assume $k'=k$ and let U_y be a $\mathscr{X}$-distinguished neighborhood of y. $\mathscr{S}^{\cdot}$ is $\mathscr{X}'$-clc on $U_y\backslash(S_{n-k}\cap U_y)$ by the inductive hypothesis and is clc on $S'_{n-k}\cap U_y=S_{n-k}\cap U_y$ since it is $\mathscr{X}$-clc.

We now claim that if $x\sim y$ and $\mathscr{S}^{\cdot}$ is $\mathscr{X}'$-clc on a neighborhood U_y of y, then $\mathscr{S}^{\cdot}$ is $\mathscr{X}'$-clc on a neighborhood U_x of x. This will complete the proof of the lemma.

Clearly we can assume xRy, so let f be the homeomorphism of (3.1). By Proposition 3.2, (1), f preserves $\mathscr{X}'$ so by assumption $f^*\mathscr{S}^\cdot$ is $\mathscr{X}'$-clc on U_x and it will suffice to show $\mathscr{S}^\cdot = f^*\mathscr{S}^\cdot$ on U_x.

Now $\mathscr{S}^\cdot = f^*\mathscr{S}^\cdot$ on $U_x \cap U_c \cap f^{-1}(U_y \cap U_c)$ by (3.2), so $f^{-1*}\mathscr{S}^\cdot = \mathscr{S}^\cdot$ on $U_y \cap U_c \cap f(U_x \cap U_c)$. Also we have already noted that $\mathscr{S}^\cdot$ is $\mathscr{X}'$-clc on U_c, thus $f^{-1*}\mathscr{S}^\cdot$ is $\mathscr{X}'$-clc on $U_y \cap f(U_x \cap U_c)$ and hence $\mathscr{X}$-clc. So on $U_y \cap f(U_x \cap U_c)$ both $\mathscr{S}^\cdot$ and $f^{-1*}\mathscr{S}^\cdot$ are $\mathscr{X}$-clc and satisfy A (by Definition 4.1, (i)), while on $U_y \cap U_c \cap f(U_x \cap U_c)$ they are quasi-isomorphic. Consequently by Definition 4.1, (iii), $f^{-1*}\mathscr{S}^\cdot = \mathscr{S}^\cdot$ on $U_y \cap f(U_x \cap U_c)$, or $\mathscr{S}^\cdot = f^*\mathscr{S}^\cdot$ on $U_x \cap U_c$. By assumption, $f^*\mathscr{S}^\cdot$ is $\mathscr{X}'$-clc on U_x, hence $\mathscr{X}$-clc, so again by Definition 4.1, (iii), $\mathscr{S}^\cdot = f^*\mathscr{S}^\cdot$ on U_x. □

Proof of Theorem 4.9 We will show that (4.8) extends uniquely to an isomorphism

$$\mathscr{S}^\cdot|_{U_c} = \mathscr{S}''^\cdot|_{U_c} \tag{4.9}$$

in $D^b(U_c)$ and that

$$\mathscr{S}''^\cdot \quad \text{is} \quad \mathscr{X}\text{-clc.} \tag{4.10}$$

The theorem is then a consequence of Definition 4.1, (iii).

Equation (4.9) follows from (4.8) and Definition 4.1, (iii), applied to the cs-space U_c with the $\mathscr{X}'$ stratification, provided we show $\mathscr{S}^\cdot$ is $\mathscr{X}'$-clc on U_c. However, *restricted to* U_c,

$$\begin{aligned} \mathscr{X}(\mathscr{S}^\cdot) &= \mathscr{X}_c(\mathscr{S}^\cdot) && \text{(by Theorem 3.3, (2))} \\ &= \mathscr{X}_c(\mathscr{S}''^\cdot) && \text{(by Proposition 3.2, (3) and (4.8))} \\ &> \mathscr{X}' && \text{(by Theorem 3.3, (1)).} \end{aligned}$$

Since $\mathscr{S}^\cdot$ is $\mathscr{X}(\mathscr{S}^\cdot)$-clc by Theorem 3.3, (3), we see $\mathscr{S}^\cdot$ is $\mathscr{X}'$-clc on U_c.

To prove (4.10), we interchange the roles of $\mathscr{X}$ and $\mathscr{X}'$ in (4.9) to obtain

$$\mathscr{S}^\cdot|_{U_c} = \mathscr{S}''^\cdot|_{U_c}. \tag{4.11}$$

Then

$$\begin{aligned} \mathscr{X}(\mathscr{S}''^\cdot) &= \mathscr{X}(\mathscr{S}''^\cdot|_{U_c}) && \text{(by Lemma 4.10)} \\ &= \mathscr{X}(\mathscr{S}^\cdot|_{U_c}) && \text{(by (4.11))} \\ &> \mathscr{X} && \text{(by Theorem 3.3, (1)).} \end{aligned}$$

Since $\mathscr{S}''^\cdot$ is $\mathscr{X}(\mathscr{S}''^\cdot)$-clc by Theorem 3.3, (3), we see it is $\mathscr{X}$-clc. □

5 Codimension c coefficient systems

Let $c \in \mathbb{Z}^+$ and let A be a c-ICT as in Definition 4.1.

Definition 5.1 An *A-coefficient system* on a cs-space X is a bounded DGS $\mathscr{E}^\cdot$ defined on a dense open set $\mathrm{Dom}(\mathscr{E}^\cdot) \subset X$ of depth $< c$ whose complement has codimension $\geqq c$. We require that

(i) $\mathscr{E}^\cdot$ is $\mathscr{X}$-cc for some cs-stratification $\mathscr{X}$ of $\mathrm{Dom}(\mathscr{E}^\cdot)$ for which $\mathscr{X}$-depth $< c$, and

(ii) $\mathscr{E}^\cdot$ satisfies A on $\mathrm{Dom}(\mathscr{E}^\cdot)$.

Proposition 5.2 *Let $\mathscr{E}^\cdot$ be an A-coefficient system on X.*

(1) *If $A=A_p$ and $c=c_p$ for p a generalized perversity, then Definition* 5.1, *(ii) is equivalent to*

$$H^i(\mathscr{E}^\cdot_x)=0 \quad \text{for } i>c_p-2, \tag{5.1}$$

$$H^i(f_x^!\,\mathscr{E}^\cdot)=0 \quad \text{for } i<n, \tag{5.2}$$

for $x\in\mathrm{Dom}(\mathscr{E}^\cdot)$.

(2) *$\mathscr{E}^\cdot[s]$ is an $A[s]$-coefficient system.*

(3) *$\mathscr{D}^\cdot_{\mathrm{Dom}(\mathscr{E}^\cdot)}\mathscr{E}^\cdot$ is a $\mathscr{D}A$-coefficient system.*

Proof. (1) follows from the characterization of A_p in Lemma 4.6, (i) and (ii). (2) and (3) follow immediately from the definitions given that $\mathscr{D}^\cdot_{\mathrm{Dom}(\mathscr{E}^\cdot)}\mathscr{E}^\cdot$ is $\mathscr{X}$-cc if $\mathscr{E}^\cdot$ is $\mathscr{X}$-cc (Proposition A.7 and [B, V., 8.7]). □

As a corollary of Proposition 5.2, (1), we see that Definition 5.1 for $A=A_p$ does not depend on the full perversity p but merely on the integer c_p. Thus

Definition 5.3 $\mathscr{E}^\cdot$ is a *c-coefficient system* on a cs-space X if it is an A_p-coefficient system for any (and hence all) generalized perversities p with $c_p=c$.

Example 5.4 Let $\mathscr{X}$ be a cs-stratification of X. Assume $c\geqq 2$ and U_{c-1} is dense (*not* merely U_c). The following are c-coefficient systems:

(1) $\mathscr{E}^\cdot=R_{U_c}[-c+2]$. Clearly this is $\mathscr{X}$-cc and satisfies (5.1), so we only need to check (5.2). However, for $x\in S_{n-k}$, $k<c$,

$$H^i(f_x^!\,R_{U_c}[-c+2])=H^{i-n+k-c+2}((j_k^!\,R_{U_c})_x), \qquad (i\in\mathbb{Z}), \tag{5.3}$$

by [B, V., 3.7(b)] and the proof of [B, V., 3.11(ii)]. Let $R_{U_c}\to\mathscr{I}^\cdot$ be an injective resolution. Then $j_k^!\,R_{U_c}$ is the sheaf of germs of sections of $\mathscr{I}^\cdot$ supported on S_{n-k}. Since we can assume $\mathscr{I}^\cdot$ is zero in negative degrees, (5.3) vanishes for $i<n+(c-k-2)$. Since $k<c$, this proves (5.2) except when $k=c-1$ and $i=n-1$. In this case however, no local section of $\mathrm{Ker}(\mathscr{I}^0\to\mathscr{I}^1)=R_{U_c}$ is supported on S_{n-c+1} because U_{c-1} is dense, and so (5.3) vanishes again.

(2) $\mathscr{E}^\cdot=\mathscr{D}_{U_c}[-n]$. By Proposition A.7 and [B, V., 8.3] this is $\mathscr{X}$-cc. Now $f_x^!\,\mathscr{D}_{U_c}[-n]$ is the dualizing sheaf of a point shifted by $-n$ [B, V., 10.11(1)] so it is simply R in degree n. This verifies (5.2). For (5.1), let $x\in S_{n-k}$ with $k<c$ and apply Corollary A.12 (when $k=c-1$, note that $L\neq\emptyset$ since U_{c-1} is dense).

Remark 5.5 Let $\mathscr{E}^\cdot$ be a c-coefficient system. Suppose $c=2$ and X has no codimension 1 stratum (for some cs-stratification). Then $\mathrm{depth}(\mathrm{Dom}(\mathscr{E}^\cdot))<2$ is equivalent to $\mathrm{depth}(\mathrm{Dom}(\mathscr{E}^\cdot))=0$, so $\mathrm{Dom}(\mathscr{E}^\cdot)$ is an open submanifold. If $\mathscr{E}^\cdot$ is clc near x, $H^\cdot(f_x^!\,\mathscr{E}^\cdot)=H^{\cdot-n}(\mathscr{E}^\cdot_x)$ [B, V., 3.7(b)] and so (5.1) and (5.2) are equivalent to

$$H^i(\mathscr{E}^\cdot_x)=0 \quad \text{for } i\neq 0. \tag{5.4}$$

Therefore if $\mathscr{E}^\cdot$ is clc then it satisfies (by (5.4)) $\mathscr{E}^\cdot = \mathscr{H}^0(\mathscr{E}^\cdot)$. We thus recover the "usual" notion of a coefficient system for intersection cohomology: a locally constant sheaf with finitely generated stalks defined on a dense open submanifold whose complement has codimension $\geqq 2$. Note that Definition 5.1, (i) allows slightly more general $\mathscr{E}^\cdot$, however, which are $\mathscr{X}$-cc with $\mathscr{X}$-depth $= 1$.

Definition 5.6 A cs-stratification $\mathscr{X}$ of X is *adapted* to an A-coefficient system $\mathscr{E}^\cdot$ if $U_c \subset \mathrm{Dom}(\mathscr{E}^\cdot)$ and $\mathscr{E}^\cdot$ is $\mathscr{X}$-clc on U_c.

Proposition 5.7 *Let $\mathscr{E}^\cdot$ be an A-coefficient system on X and assume there exists a* cs*-stratification $\mathscr{X}$ adapted to $\mathscr{E}^\cdot$. Then*

(1) *$\mathscr{E}^\cdot$ has a unique largest extension $\tilde{\mathscr{E}}^\cdot$ which is an A-coefficient system;*

(2) *There is a unique coarsest* cs*-stratification $\tilde{\mathscr{X}}$ adapted to $\tilde{\mathscr{E}}^\cdot$.*

Proof. Let $\mathscr{X}$ be any cs-stratification adapted to $\mathscr{E}^\cdot$. Let $\mathscr{S}^\cdot$ be the $\mathscr{X}$-cc extension of $\mathscr{E}^\cdot|_{U_c}$ satisfying A from Definition 4.1, (ii) and let $\tilde{\mathscr{X}} = \mathscr{X}(\mathscr{S}^\cdot)$ be the coarsest stratification for which $\mathscr{S}^\cdot$ is $\tilde{\mathscr{X}}$-clc, by Corollary 3.4. By Theorem 4.9, $\mathscr{S}^\cdot$ and thus $\tilde{\mathscr{X}}$ is independent of $\mathscr{X}$.

We now claim that $\mathscr{S}^\cdot|_{\mathrm{Dom}(\mathscr{E}^\cdot)} = \mathscr{E}^\cdot$ and $\mathrm{Dom}(\mathscr{E}^\cdot) \subset \tilde{U}_c$. To see this, let $\mathscr{X}'$ be a cs-stratification on $\mathrm{Dom}(\mathscr{E}^\cdot)$ for which $\mathscr{E}^\cdot$ is $\mathscr{X}'$-cc and $\mathscr{X}'$-depth $< c$, as guaranteed by Definition 5.1, (i). By Theorem 4.9 applied to $(\mathscr{S}^\cdot|_{\mathrm{Dom}(\mathscr{E}^\cdot)}, \mathscr{X})$ and $(\mathscr{E}^\cdot, \mathscr{X}')$, we have $\mathscr{S}^\cdot|_{\mathrm{Dom}(\mathscr{E}^\cdot)} = \mathscr{E}^\cdot$. Thus $\mathscr{S}^\cdot|_{\mathrm{Dom}(\mathscr{E}^\cdot)}$ is $\mathscr{X}'$-clc and therefore

$$\begin{aligned} \tilde{\mathscr{X}}|_{\mathrm{Dom}(\mathscr{E}^\cdot)} &= \mathscr{X}_{\mathrm{Dom}(\mathscr{E}^\cdot)}(\mathscr{S}^\cdot|_{\mathrm{Dom}(\mathscr{E}^\cdot)}) \quad \text{(by Proposition 3.2, (ii))} \\ &> \mathscr{X}' \quad \text{(by Theorem 3.3, (1)).} \end{aligned}$$

This implies $\mathrm{Dom}(\mathscr{E}^\cdot) \subset \tilde{U}_c$ since $\mathscr{X}'$-depth $< c$.

By the claim, $\tilde{\mathscr{E}}^\cdot = \mathscr{S}^\cdot|_{\tilde{U}_c}$ is an A-coefficient system which extends $\mathscr{E}^\cdot$. Since the construction depended only on $\mathscr{E}^\cdot|_{U_c}$, it yields the same result for any extension of $\mathscr{E}^\cdot$, and thus a largest extension, which is assertion (1). We also see that $\tilde{\mathscr{X}} > \mathscr{X}$ for any $\mathscr{X}$ adapted to any extension of $\mathscr{E}^\cdot$, which is assertion (2). □

Remark. We do not know if the hypothesis of Proposition 5.7, that there exists a cs-stratification $\mathscr{X}$ adapted to $\mathscr{E}^\cdot$, is necessary for (1). In the situation of Remark 5.5, if $\mathscr{E}^\cdot = \mathscr{E}$ is a "usual" coefficient system then one can prove (1) without assuming $\mathscr{X}$ exists [B, V., 4.11].

Convention. We shall not distinguish $\mathscr{E}^\cdot$ and any A-coefficient system obtained from $\mathscr{E}^\cdot$ by restriction of domain. This has the effect of increasing the set of cs-stratifications adapted to $\mathscr{E}^\cdot$.

6 Intersection cohomology on cs-spaces

Let $c \in \mathbb{Z}^+$. Let A be a c-ICT, X a cs-space, and $\mathscr{E}^\cdot$ an A-coefficient system on $\mathrm{Dom}(\mathscr{E}^\cdot) \subset X$. We assume that there exists some cs-stratification $\mathscr{X}$ adapted to $\mathscr{E}^\cdot$.

Definition 6.1 Given a cs-stratification $\mathscr{X}$ of X which is adapted to $\mathscr{E}^\cdot$, we define the *A-intersection cohomology sheaf with coefficients in $\mathscr{E}^\cdot$*, $\mathscr{P}^\cdot_{A,\mathscr{X}}(\mathscr{E}^\cdot)$, to be the unique (up to canonical isomorphism in $D^b(X)$) $\mathscr{X}$-cc extension of $\mathscr{E}^\cdot|_{U_c}$ to all

of X which satisfies A, as guaranteed by Definition 4.1. The *A-intersection cohomology* $I_A H^i(X; \mathscr{E}^\cdot)$ $(i \in \mathbb{Z})$ is defined to be the hypercohomology $\mathbb{H}^i(X; \mathscr{P}^\cdot_{A,\mathscr{X}}(\mathscr{E}^\cdot))$.

Theorem 6.2 (Topological invariance) *$\mathscr{P}^\cdot_{A,\mathscr{X}}(\mathscr{E}^\cdot)$ and $I_A H^\cdot(X; \mathscr{E}^\cdot)$ are independent of the choice of* cs*-stratification $\mathscr{X}$ adapted to $\mathscr{E}^\cdot$.*

Proof. By Proposition 5.7, (2) there is a unique coarsest cs-stratification $\tilde{\mathscr{X}}$ adapted to $\mathscr{E}^\cdot$. Since $\tilde{\mathscr{X}} > \mathscr{X}$, $\tilde{\mathscr{X}}$-clc implies $\mathscr{X}$-clc and also $\tilde{U}_c \supset U_c$. Thus $\mathscr{P}^\cdot_{A,\tilde{\mathscr{X}}}(\mathscr{E}^\cdot) = \mathscr{P}^\cdot_{A,\mathscr{X}}(\mathscr{E}^\cdot)$ by Remark 4.2. □

We therefore write $\mathscr{P}^\cdot_A(\mathscr{E}^\cdot)$ for $\mathscr{P}^\cdot_{A,\mathscr{X}}(\mathscr{E}^\cdot)$. If $A = A_p[s]$ for a generalized perversity p and $s \in \mathbb{Z}$, we will replace $A_p[s]$ by $p[s]$ in the above notation. We have for any cs-stratification $\mathscr{X}$ adapted to $\mathscr{E}^\cdot$ the formula (see (4.7)):

$$\mathscr{P}^\cdot_p(\mathscr{E}^\cdot) = \tau_{\leq p(n)} R i_{n*} \dots \tau_{\leq p(c_p)} R i_{c_p *}(\mathscr{E}^\cdot|_{U_{c_p}}). \tag{6.1}$$

From (6.1) and Theorem 2.1 we clearly have the

Proposition 6.3 *For X a* pl*-pseudomanifold, $\mathscr{P}^\cdot_p(\mathscr{D}^\cdot_{U_{c_p}}[-n]) = \mathscr{I}_p \mathscr{C}_{n-}$. (defined in* §2*).*

Thus it is consistent with §2 to define

$$I_p H_i(X; R) = I_p H^{n-i}(X; \mathscr{D}^\cdot_{U_{c_p}}[-n]) \quad (i \in \mathbb{Z}).$$

Example 6.4 For the constant generalized perversity n we have

$$I_{n[n]} H^i(X; R_X) = H^i(X; R), \quad \text{and}$$
$$I_n H^i(X; \mathscr{D}^\cdot_X[-n]) = H_{n-i}(X; R),$$

ordinary (Borel-Moore) homology and cohomology. (We use perversity n rather than $n-1$ merely to avoid having to assume U_n is dense – see Example 5.4.)

Proposition 6.5 *$\mathscr{P}^\cdot_A(\mathscr{E}^\cdot)$ satisfies Wilder's (P, Q) condition* (1.2)*. For X compact, in particular, $I_A H^i(X; \mathscr{E}^\cdot)$ is finitely generated.*

Proof. By Proposition A.6 and [B, V., 3.10(e)], $\mathscr{P}^\cdot_A(\mathscr{E}^\cdot)$ $\mathscr{X}$-cc implies $\mathscr{P}^\cdot_A(\mathscr{E}^\cdot)$ cc and thus (1.2). □

For ordinary perversities p and q a morphism of coefficient systems induces a map $I_p H^i(X; \mathscr{E}) \to I_q H^i(X; \mathscr{F})$ provided $p \leqq q$ [B, V., 9.2]. For generalized perversities it is appropriate to consider in addition maps into $I_{q[s]} H^i$ since coefficient systems are no longer restricted to degree 0:

Proposition 6.6 *Let p and q be generalized perversities and let $s \in \mathbb{Z}$. Let $\mathscr{E}^\cdot$ and $\mathscr{F}^\cdot$ be A_p- and $A_q[s]$-coefficient systems respectively such that there exists a* cs*-stratification $\mathscr{X}$ adapted to both $\mathscr{E}^\cdot$ and $\mathscr{F}^\cdot$. Assume that*

$$\max(p(k), c_p - 2) \leqq \min(q(k), k-2) - s, \quad k \in [\min(c_p, c_q), n]. \tag{6.2}$$

Then any morphism

$$\mathscr{E}^\cdot|_{\mathrm{Dom}(\mathscr{E}^\cdot) \cap \mathrm{Dom}(\mathscr{F}^\cdot)} \to \mathscr{F}^\cdot|_{\mathrm{Dom}(\mathscr{E}^\cdot) \cap \mathrm{Dom}(\mathscr{F}^\cdot)}$$

extends uniquely to a morphism

$$\mathscr{P}^\cdot_p(\mathscr{E}^\cdot) \to \mathscr{P}^\cdot_{q[s]}(\mathscr{F}^\cdot)$$

in $D^b(X)$.

Proof. We may assume that $\mathrm{Dom}(\mathscr{E}^{\cdot}) \cap \mathrm{Dom}(\mathscr{F}^{\cdot}) = U_{\min(c_p, c_q)}$. Now apply Lemma 4.7 to extend stratum by stratum; the hypotheses of the lemma are fulfilled by (4.3), (4.5), (4.6), and (6.2). □

Corollary 6.7 *Under the above hypotheses, we have a natural map*

$$I_p H^i(X; \mathscr{E}^{\cdot}) \to I_{q[s]} H^i(X; \mathscr{F}^{\cdot}) = I_q H^{i+s}(X; \mathscr{F}^{\cdot}[-s]) \quad (i \in \mathbb{Z}).$$

Example. Let p be a generalized perversity so that $U_{c_p - 1}$ is dense (for some cs-stratification $\mathscr{X}$). There are natural maps

$$H^i(X; R) = I_{n[n]} H^i(X; R_X) \to I_{p[c_p - 2]} H^i(X; R_{U_{c_p}})$$

and

$$I_p H^i(X; \mathscr{D}^{\cdot}_{U_{c_p}}[-n]) \to I_n H^i(X; \mathscr{D}^{\cdot}_X[-n]) = H_{n-i}(X; R).$$

Definition 6.8 A cs-space X with cs-stratification $\mathscr{X}$ is called *normal* if U_2 is dense and every point $x \in X$ has a fundamental system of neighborhoods U with $U \cap U_2$ connected.

Remark 6.9 If X is normal, the links L of a stratum S_{n-k} are path-connected when $k \geqq 2$.

Proposition 6.10 *Let p be a generalized perversity ($c_p \geqq 2$) and define*

$$\tilde{p}(k) = \max(\min(p(k), k-2), 0) \quad (k \in \mathbb{Z}^+),$$

the associated ordinary perversity. *Let X be a* cs-*space of dimension n with* cs-*stratification $\mathscr{X}$. Assume U_1 is dense and U_{c_p} is normal. Then*

$$\mathscr{P}^{\cdot}_{\tilde{p}}(\mathscr{D}^{\cdot}_{U_{c_p}}[-n]) = \mathscr{P}^{\cdot}_{\tilde{p}}(\mathscr{D}^{\cdot}_{U_2}[-n]).$$

Proof. By Lemma 4.6 it suffices to check that $\mathscr{D}^{\cdot}_{U_{c_p}}[-n]$ satisfies the vanishing and covanishing conditions for $\tilde{p}$ at $x \in S_{n-k}$ with $k < c_p$. Since the links are path-connected when $2 \leqq k < c_p$ (by Remark 6.9), this follows from Corollary A.12 as in Example 5.4 (2). □

The definitions immediately imply the following duality theorem.

Theorem 6.11 *Let $\mathscr{E}^{\cdot}$ be an A-coefficient system and assume there exists a* cs-*stratification adapted to $\mathscr{E}^{\cdot}$. Then there exists a unique isomorphism in $D^b(X)$,*

$$\mathscr{D}^{\cdot}_X \mathscr{P}^{\cdot}_A(\mathscr{E}^{\cdot}) = \mathscr{P}^{\cdot}_{\mathscr{D}A}(\mathscr{D}^{\cdot}_{\mathrm{Dom}(\mathscr{E}^{\cdot})} \mathscr{E}^{\cdot}),$$

which extends the identity map of $\mathscr{D}^{\cdot}_{\mathrm{Dom}(\mathscr{E}^{\cdot})}(\mathscr{E}^{\cdot})$.

In order of Theorem 6.11 to be useful, we need to identify $\mathscr{D}A$.

Proposition 6.12 *Assume R is a field and p is a generalized perversity. Let*

$$q(k) = k - 2 - p(k) \quad (k \in \mathbb{Z}^+)$$

be the dual perversity. Then on cs-*constructible* DGS,

$$\mathscr{D}A_p = A_{q^{\cdot}}[c_p - 2 + n],$$

where q' is the generalized perversity

$$q'(k)=\max(q(k),0)+c_p-2 \qquad (k\in\mathbb{Z}^+). \tag{6.3}$$

If p is an ordinary perversity this reduces to

$$\mathcal{D}A_p=A_{q'}[n].$$

Proof. When R is a field and $\mathcal{S}^\cdot$ is a cs-constructible DGS, we have the formulas (see Proposition A.7 and [B, V., 9.7])

$$\begin{aligned} H^i((\mathcal{D}^\cdot_X\mathcal{S}^\cdot)_x)&=H^{-i}(f_x^!\mathcal{S}^\cdot)^* \\ H^i(f_x^!(\mathcal{D}^\cdot_X\mathcal{S}^\cdot))&=H^{-i}(\mathcal{S}^\cdot_x)^*, \end{aligned} \qquad (i\in\mathbb{Z},\ x\in X).$$

If we apply these to Definition 4.3, the proposition follows. The cutoff by zero and the shift by c_p-2 in the definition of q' are necessary to ensure that $c_{q'}=c_p$. □

Corollary 6.13 *Under the above assumptions, there is a natural isomorphism*

$$I_{q'[c_p-2]}H^i(X;\mathcal{D}^\cdot_{\mathrm{Dom}(\mathscr{E}^\cdot)}(\mathscr{E}^\cdot)[-n])=I_pH_c^{n-i}(X;\mathscr{E}^\cdot)^* \qquad (i\in\mathbb{Z}).$$

Proof. Over a field, we have an isomorphism $\mathbb{H}^i(X;\mathcal{D}^\cdot_X\mathcal{S}^\cdot)=\mathbb{H}_c^{-i}(X;\mathcal{S}^\cdot)^*$ for $\mathcal{S}^\cdot\in\mathrm{DGS}(X)$ (Proposition A.7 and [B, V., 9.7(1)]). □

Let $\overset{L}{\otimes}$ denote the left derived functor of $\otimes$ in $D^b(X)$ [B, V., 6.2].

Proposition 6.14 *Let p, q, and r be generalized perversities and let $s\in\mathbb{Z}$. Let $\mathscr{E}^\cdot$, $\mathscr{F}^\cdot$, and $\mathscr{G}^\cdot$ be A_p-, A_q-, and $A_r[s]$-coefficient systems respectively such that there exists a* cs*-stratification $\mathcal{X}$ adapted to all three. Assume that*

$$\begin{aligned} &\max(p(k),c_p-2)+\max(q(k),c_q-2) \\ &\qquad\leqq\min(r(k),k-2)-s, \qquad k\in[\min(c_p,c_q,c_r),n]. \end{aligned} \tag{6.4}$$

Let $U=\mathrm{Dom}(\mathscr{E}^\cdot)\cap\mathrm{Dom}(\mathscr{F}^\cdot)\cap\mathrm{Dom}(\mathscr{G}^\cdot)$. Then any morphism

$$\mathscr{E}^\cdot|_U\overset{L}{\otimes}\mathscr{F}^\cdot|_U\to\mathscr{G}^\cdot|_U$$

extends uniquely to a morphism

$$\mathscr{P}^\cdot_p(\mathscr{E}^\cdot)\overset{L}{\otimes}\mathscr{P}^\cdot_q(\mathscr{F}^\cdot)\to\mathscr{P}^\cdot_{r[s]}(\mathscr{G}^\cdot).$$

Proof. The proof is similar to that of Proposition 6.6. □

Corollary 6.15 *Under the above assumptions there is a natural pairing*

$$I_pH_c^i(X;\mathscr{E}^\cdot)\otimes I_qH^j(X;\mathscr{F}^\cdot)\to I_{r[s]}H_c^{i+j}(X;\mathscr{G}^\cdot).$$

Example. With q' defined as in (6.3), we have a pairing

$$I_p H_c^{n-i}(X; \mathscr{E}^\cdot) \otimes I_{q'[c_p-2]} H^i(X; \mathscr{D}^\cdot_{\mathrm{Dom}(\mathscr{E}^\cdot)}(\mathscr{E}^\cdot)[-n]) \to I_n H_c^n(X; \mathscr{D}^\cdot_X[-n]) = H_0^c(X; R).$$

When R is a field, this pairing, followed by the natural map $H_0^c(X; R) \to R$, induces the isomorphism of Corollary 6.13 (see [B, V., 9.16]).

7 The hypercohomology spectral sequence and the canonical filtration

In this section, which is independent of what has preceeded, we identify the hypercohomology filtration and the canonical filtration.

Let $\mathscr{S}^\cdot$ be a DGS which is bounded below and let $\mathscr{G}^\cdot = \bigoplus_p \mathscr{C}^p(\mathscr{S}^{\cdot -p})$ be the Godement canonical resolution. Define the filtration

$$F^k\mathscr{G}^\cdot = \bigoplus_{p \geqq k} \mathscr{C}^p(\mathscr{S}^{\cdot -p}); \tag{7.1}$$

the corresponding filtration on $\Gamma(X, \mathscr{G}^\cdot)$ induces the hypercohomology spectral sequence [G, 4.6.1] converging to $\mathbb{H}^\cdot(\mathscr{S}^\cdot) = H^\cdot(\Gamma(X, \mathscr{G}^\cdot))$. Since $\Gamma(X, \mathscr{C}^p(\cdot))$ is an exact functor, we see that $E_2^{p,q} = H^p(X; \mathscr{H}^q(\mathscr{S}^\cdot))$ for the hypercohomology spectral sequence.

On the other hand, given an arbitrary biregular filtration F on $\mathscr{S}^\cdot$ (i.e., F is a finite filtration on each $\mathscr{S}^i$), there is a spectral sequence converging to $\mathbb{H}^\cdot(\mathscr{S}^\cdot)$ and a corresponding filtration of hypercohomology defined as follows [D, 1.4.5]: Choose a *Γ-acyclic filtered resolution* $(\mathscr{S}'^\cdot, \mathscr{F}')$ of $(\mathscr{S}^\cdot, \mathscr{F})$; this is a filtered quasi-isomorphism $\mathscr{S}^\cdot \to \mathscr{S}'^\cdot$ into a bounded below complex with a biregular filtration such that $\mathrm{Gr}^k(\mathscr{S}'^q)$ is Γ-acyclic (i.e., $H^i(X; \mathrm{Gr}^k(\mathscr{S}'^q)) = 0$ for $i > 0$). Then the filtration F' on $\Gamma(X, \mathscr{S}'^\cdot)$ defines a spectral sequence converging to $\mathbb{H}^\cdot(\mathscr{S}^\cdot)$ with $E_1^{p,q} = \mathbb{H}^{p+q}(\mathrm{Gr}^p\mathscr{S}'^\cdot)$. Such resolutions exist and the spectral sequence is independent of the resolution.

In the case of the *canonical filtration* $\tau_{\leqq k}\mathscr{S}^\cdot$, we set $F^k\mathscr{S}^\cdot = \tau_{\leqq -k}\mathscr{S}^\cdot$ in order to obtain a decreasing filtration. The resulting spectral sequence has $E_1^{p,q} = H^{2p+q}(X; \mathscr{H}^{-p}(\mathscr{S}^\cdot))$.

Comparing this with the hypercohomology spectral sequence above suggests the following:

Proposition 7.1 *The hypercohomology spectral sequence is isomorphic to the spectral sequence induced by the canonical filtration after reindexing $E_{r+1}^{2p+q,-p} \to E_r^{p,q} (r \geqq 1)$. The corresponding filtrations on $\mathbb{H}^i(\mathscr{S}^\cdot)$ are isomorphic after reindexing $F^{k+i}\mathbb{H}^i(\mathscr{S}^\cdot) \to F^k\mathbb{H}^i(\mathscr{S}^\cdot)$.*

Proof. Define the *filtration décalée* [D, (1.3.3)] corresponding to (7.1) by

$$\mathrm{Dec}(F)^k\mathscr{G}^\cdot = \mathrm{Ker}(d: F^{k+\cdot}\mathscr{G}^\cdot \to \mathscr{G}^{\cdot+1}/F^{k+\cdot+1}\mathscr{G}^{\cdot+1}).$$

According to [D, (1.3.4)], the spectral sequence associated to F and $\mathrm{Dec}(F)$ on $\Gamma(X, \mathscr{G}^\cdot)$ are isomorphic under the reindexing $E_{r+1}^{2p+q,-p} \to E_r^{p,q} (r \geqq 1)$. It is easy to verify that $\mathrm{Dec}(F)^k\mathscr{G}^\cdot = \bigoplus_p \mathscr{C}^p(\tau_{\leqq -k}\mathscr{S}^{\cdot -p})$ and thus that $(\mathscr{G}^\cdot, \mathrm{Dec}(F))$ is a Γ-acyclic filtered resolution of $\mathscr{S}^\cdot$ with the canonical filtration $\tau_{\leqq -k}\mathscr{S}^\cdot$. □

Remark. In fact, [D] *defines* the hypercohomology spectral sequence to be the one induced by the canonical filtration, although we have not seen the proposition in the literature.

Corollary 7.2 *If F denotes the filtration from the hypercohomology spectral sequence,*

$$F^k\mathbb{H}^i(\mathscr{S}^\cdot)=\mathrm{Im}(\mathbb{H}^i(\tau_{\leqq i-k}\mathscr{S}^\cdot)\to\mathbb{H}^i(\mathscr{S}^\cdot)).$$

8 Filtration by perversity

Let X be a cs-space of dimension n and $\mathscr{X}$ a cs-stratification. In order to see how truncation affects the covanishing condition (4.4) we need the following lemma.

Lemma 8.1 *Let $\mathscr{S}^\cdot\in\mathrm{DGS}(X)$ be bounded from below and $\mathscr{X}$-clc and let $x\in S_{n-k}$ $(k\in\mathbb{Z}^+)$. The natural map*

$$H^i(f_x^!\,\tau_{\leqq m}\mathscr{S}^\cdot)\to H^i(f_x^!\,\mathscr{S}^\cdot) \tag{8.1}$$

is an isomorphism for $i\leqq n-k+m$ and an injection for $i=n-k+m+1$.

Proof. We have

$$\begin{aligned}(8.2)\quad H^i(f_x^!\,\tau_{\leqq m}\mathscr{S}^\cdot)&=H^{i-n+k}((j_k^!\,\tau_{\leqq m}\mathscr{S}^\cdot)_x) && \text{by the proof of [B, V., 3.11(ii)]}\\ &=\varinjlim_{U\ni x}\mathbb{H}^{i-n+k}(U;j_k^!\,\tau_{\leqq m}\mathscr{S}^\cdot)\\ &=\varinjlim_{U\ni x}\mathbb{H}^{i-n+k}_{U\cap S_{n-k}}(U;\tau_{\leqq m}\mathscr{S}^\cdot) && \text{by [B, V., 1.8(5)]},\end{aligned}$$

where the last line denotes hypercohomology with supports in $U\cap S_{n-k}$. Similarly,

$$H^i(f_x^!\,\mathscr{S}^\cdot)=\varinjlim_{U\ni x}\mathbb{H}^{i-n+k}_{U\cap S_{n-k}}(U;\mathscr{S}^\cdot). \tag{8.3}$$

The natural map between the hypercohomology spectral sequences associated to (8.2) and (8.3) is clearly an isomorphism on E_2^{pq} for $q\leqq m$ and an injection for $q=m+1$ (in fact the first spectral sequence is zero for $q>m$). By a general spectral sequence argument (see, for example, [Zu, (4.7)]), we have an isomorphism on E_∞^{pq} terms for $p+q\leqq m$ and an injection for $p+q=m+1$. The lemma follows since $i-n+k=p+q$. □

Theorem 8.2 *Let p be a generalized perversity and let $m\in\mathbb{Z}$. If $\mathscr{S}^\cdot\in D^b(X)$ is $\mathscr{X}$-clc and satisfies A_p, then $\tau_{\leqq m}\mathscr{S}^\cdot$ satisfies A_{p_m}, where*

$$p_m(k)=\min(p(k),m),\qquad (k\in\mathbb{Z}^+).$$

Proof. By Lemma 4.6, (4.3) and (4.4) hold for $\mathscr{S}^{\cdot}$ and p and we need to prove the analogous statements for $\tau_{\leqq m}\mathscr{S}^{\cdot}$ and p_m. By (4.3) and the definition of $\tau_{\leqq m}$ we have

$$H^i(\tau_{\leqq m}\mathscr{S}^{\cdot}_x)=0 \quad \text{for } x\in S_{n-k}$$

provided either $i>\max(p(k), c_p-2)$ or $i>m$, i.e. $i>\min(\max(p(k), c_p-2), m)$ $=\max(p_m(k), c_{p_m}-2)$. Similarly, by (4.4) and Lemma 8.1 we have

$$H^i(f_x^!\tau_{\leqq m}\mathscr{S}^{\cdot})=0 \quad \text{for } x\in S_{n-k}$$

provided $i<n-\max(q(k), 0)$ and $i<n-(k-2-m)$, i.e. $i<\min(n-\max(q(k), 0),$ $n-(k-2-m))=n-\max(k-2-p_m(k), 0)$. □

Corollary 8.3 *Let $\mathscr{E}^{\cdot}$ be a c_p-coefficient system and assume $\mathscr{X}$ is adapted to $\mathscr{E}^{\cdot}$. Then there is a unique isomorphism*

$$\tau_{\leqq m}\mathscr{P}_p^{\cdot}(\mathscr{E}^{\cdot})=\mathscr{P}_{p_m}^{\cdot}(\tau_{\leqq m}\mathscr{E}^{\cdot}|_{U_{c_{p_m}}})$$

in $D^b(X)$ which extends the identity map on $\tau_{\leqq m}\mathscr{E}^{\cdot}|_{U_{c_{p_m}}}$.

Proof. Apply the theorem, Definition 4.1, (i), and Remark 4.2. □

Remark. Modulo extension and restriction of domain, the c_{p_m}-coefficient system $\tau_{\leqq m}\mathscr{E}^{\cdot}|_{U_{c_{p_m}}}$ is independent of the choice of cs-stratification $\mathscr{X}$ adapted to $\mathscr{E}^{\cdot}$.

Corollary 8.4 *The filtration of $I_p H^i(X; \mathscr{E}^{\cdot})$ induced by the hypercohomology spectral sequence for $\mathscr{P}_p^{\cdot}(\mathscr{E}^{\cdot})$ is*

$$F^k I_p H^i(X;\mathscr{E}^{\cdot})=\operatorname{Im}(I_{p_{i-k}} H^i(X;\tau_{\leqq i-k}\mathscr{E}^{\cdot}|_{U_{c_{p_{i-k}}}})\to I_p H^i(X;\mathscr{E}^{\cdot})) \tag{8.4}$$

where the map is the natural one of Corollary 6.7.

Proof. Apply Corollaries 7.2 and 8.3. □

Remark 8.5 *If $\mathscr{E}^{\cdot}|_{U_{c_{p_{i-k}}}}$ is already a $c_{p_{i-k}}$-coefficient system, then by (5.1) it is equal to $\tau_{\leqq i-k}\mathscr{E}^{\cdot}|_{U_{c_{p_{i-k}}}}$. In this case then, the truncation in (8.4) may be omitted. This occurs, for example, when $\mathscr{E}^{\cdot}=R_{U_{c_p}}[-c_p+2]$ or $\mathscr{D}_{U_{c_p}}^{\cdot}[-n]$ and $U_{c_{p_{i-k}}-1}$ is dense (see Example 5.4).

Corollary 8.6 *Assume U_1 is dense for some* cs-*stratification. The Zeeman filtration of $H_i(X; R)$ (i.e., the filtration of the hypercohomology spectral sequence for $\mathscr{D}_X^{\cdot}[-n]$) is*

$$F^k H_i(X;R)=\operatorname{Im}(I_{n-i-k} H_i(X;R)\to H_i(X;R)), \tag{8.5}$$

where $n-i-k$ is the constant generalized perversity having that value.

Proof. Apply Corollary 8.4 and Remark 8.5 with p being the constant generalized perversity n and $\mathscr{E}^{\cdot}=\mathscr{D}_X^{\cdot}[-n]$. Then use Example 6.4. □

Corollary 8.7 *Assume U_1 is dense for some* cs-*stratification and X is normal. Then* (8.5) *holds with the generalized perversity $n-i-k$ replaced by the ordinary perversity t_{n-i-k} (t=top ordinary perversity).*

Proof. Apply Proposition 6.10. □

Remark. Results analogous to Corollaries 8.6 and 8.7 were obtained by McCrory for pl-pseudomanifolds by quite different means [M1, M2].

Appendix. Constructibility and duality on cs-spaces

For the sake of brevity, we will not restate and reprove all the facts from [B, V.] which we are extending to cs-spaces, but merely point out how the proofs need to be changed.

Lemma A.1 *Let M be a topological space with a filtration $\mathscr{M}$ by closed subsets and let Y be a connected compact contractible space. Let $X = Y \times M$ and let $\mathscr{X}$ be the filtration $Y \times \mathscr{M}$ of X. Let $y \in Y$, $M_y = \{y\} \times M$, and let $\tau\colon X \to M_y$ be the projection and $j\colon M_y \to X$ the inclusion. Let $\mathscr{S}^\cdot \in \mathrm{DGS}(X)$ be $\mathscr{X}$-clc. Then the natural map*

$$R\tau_* \mathscr{S}^\cdot \to j^* \mathscr{S}^\cdot \tag{A.1}$$

induced by restriction of sections is a quasi-isomorphism.

Proof. Let $z \in M_y$ and let $i\colon \tau^{-1}(z) \to X$ be the inclusion. It suffices to assume $\mathscr{S}^\cdot$ is injective (replace $\mathscr{S}^\cdot$ by an injective resolution) so $R\tau_* \mathscr{S}^\cdot = \tau_* \mathscr{S}^\cdot$. Since τ is proper, $\tau_* = \tau_!$, and we may compute (see [B, VI., 2.6])

$$H^\cdot((R\tau_* \mathscr{S}^\cdot)_z) = H^\cdot((\tau_! \mathscr{S}^\cdot)_z) = H^\cdot(\Gamma_c(\tau^{-1}(z); i^* \mathscr{S}^\cdot)). \tag{A.2}$$

However $\tau^{-1}(z)$ is compact and (because $\mathscr{S}^\cdot$ is injective) $i^* \mathscr{S}^\cdot$ is c-soft [Br, II., §§ 5.2, 9.2, 9.5]. Thus (A.2) equals the hypercohomology $\mathbb{H}^\cdot(\tau^{-1}(z); i^* \mathscr{S}^\cdot)$. But since $\mathscr{H}^\cdot \mathscr{S}^\cdot$ is locally constant on the strata of X and $\tau^{-1}(z)$ is connected, contractible, and contained in a single stratum of X, $\mathscr{H}^\cdot \mathscr{S}^\cdot$ is constant on $\tau^{-1}(z)$. Thus the spectral sequence for $\mathbb{H}^\cdot(\tau^{-1}(z); i^* \mathscr{S}^\cdot)$ collapses to just $(\mathscr{H}^\cdot \mathscr{S}^\cdot)_z = H^\cdot((j^* \mathscr{S}^\cdot)_z)$. □

Corollary A.2 *Under the above conditions, the natural map $\tau^* R\tau_* \mathscr{S}^\cdot \to \mathscr{S}^\cdot$ is a quasi-isomorphism.*

Proposition A.3 *The lemma* [B, V., 3.8] *remains valid if $(M, \mathscr{M})$ is merely assumed to be a topological space with a filtration by closed subsets, rather than a stratified topological pseudomanifold.*

Proof. Part (a) of [B, V., 3.8] follows immediately by taking hypercohomology with compact supports of (A.1). Then the given proof for part (b) of [B, V., 3.8] holds without change. □

In the applications, M will usually be a link L in a cs-space.

Lemma A.4 *In the situation of Lemma* A.1, *assume that Y is instead an open ball and that M is compact. Then if $\mathscr{S}^\cdot$ satisfies* (1.2) *(Wilder's (P, Q) condition), $\mathbb{H}_c^\cdot(Y \times M; \mathscr{S}^\cdot)$ is finitely generated.*

Proof. Let $\tilde{Y} \subset Y$ be a smaller open ball such that $\bar{\tilde{Y}} \subset Y$ is compact. The natural map

$$\mathbb{H}_c^\cdot(\tilde{Y} \times M; \mathscr{S}^\cdot) \to \mathbb{H}_c^\cdot(Y \times M; \mathscr{S}^\cdot) \tag{A.3}$$

is an isomorphic to $\mathbb{H}^{\cdot-\dim Y}(M_y; j^*\mathscr{S}^{\cdot})$ for $y\in\tilde{Y}$ (compatibly with (A.3)). However, by (1.2) the image of (A.3) is finitely generated. □

Corollary A.5 *Let X be a* cs-*space with* cs-*stratification $\mathscr{X}$. Let $\mathscr{S}^{\cdot}\in \mathrm{DGS}(U_k)$ be $\mathscr{X}$-clc and satisfy* (1.2) *(Wilder's (P, Q) condition). Let U be a distinguished neighborhood of $x\in S_{n-k}$ and let $Z=U\cap S_{n-k}$. Then $\mathbb{H}^{\cdot}_c(U\setminus Z; \mathscr{S}^{\cdot})$ is finitely generated.*

Proof. If $U=B\times \mathring{c}(L)$, then $U\setminus Z=B\times\mathbb{R}\times L$. Apply the lemma with $Y=B\times\mathbb{R}$ and $M=L$. □

Proposition A.6 *The results in* [B, V., 3.9–3.11] *remain valid for X a* cs-*space with a* cs-*stratification $\mathscr{X}$.*

Proof. Given Proposition A.3, we need only modify the proofs in [B, V.] for 3.10(e) and 3.11(iii). It is difficult to induct on dimension in cs-spaces since the links L may not be cs-spaces. However the sets $U\setminus Z$ (in the notation of Corollary A.5) *are* cs-spaces and of lower depth. Thus for 3.10(e), we replace induction on $\dim X$ by induction on depth and assume the result for $U_k (k\leqq n)$. We can then apply Corollary A.5 after which the proof concludes as in [B, V.]. Similarly in 3.11(iii) we replace $\mathbb{H}^{\cdot}(L; \mathscr{A}^{\cdot}|_L)$ by $\mathbb{H}^{\cdot+n-k+1}_c(U\setminus Z; \mathscr{A}^{\cdot})$ and apply Corollary A.5. □

Proposition A.7 *The results in* [B, V., 5.13, § 8 *and* 9.7] *(such as biduality) involving a stratified topological pseudomanifold $(X, \mathscr{X})$ remain valid for a* cs-*space with a* cs-*stratification.*

Proof. Given Proposition A.6, there is no change needed in the proofs in [B, V.]. □

Our final result is that when X is a cs-space, $\mathscr{D}^{\cdot}_X$ is quasi-isomorphic to the classical sheaf of locally finite singular chains. In particular, Borel-Moore homology vanishes in negative degrees for cs-spaces. Since we have not been able to find a proof for this in the literature, we provide one below. For $A\subseteq X$, let $\Delta_{\cdot}(X, A; R)$ (resp. $\Delta^c_{\cdot}(X, A; R)$) denote the complex of locally finite (resp. finite) singular chains in X (coefficients in R) modulo those lying in A. Denote its homology by ${}_{\mathrm{sing}}H_{\cdot}(X, A; R)$ (resp. ${}_{\mathrm{sing}}H^c_{\cdot}(X, A; R)$). Clearly $\Delta_{\cdot}(X, A; R)=\Delta^c_{\cdot}(X, A; R)$ when $X\setminus A$ is precompact in X. Let $\mathscr{A}^{\cdot}_X$ denote the sheaf generated by the presheaf $U\mapsto\Delta_{-\cdot}(X, X\setminus U; R)$; by [Br, I., Exercise 12], $\Gamma(\mathscr{A}^{\cdot}_X)\cong\Delta_{-\cdot}(X; R)$ and $\Gamma_c(\mathscr{A}^{\cdot}_X)\cong\Delta^c_{-\cdot}(X; R)$. By [Br, II., Exercise 32], $H^{\cdot}(H^p(X; \mathscr{A}^{\cdot}_X))=0$ for all $p>0$ and so by a spectral sequence argument, $\mathbb{H}^{\cdot}(X; \mathscr{A}^{\cdot}_X)\cong H^{\cdot}(\Gamma(\mathscr{A}^{\cdot}_X))\cong{}_{\mathrm{sing}}H_{-\cdot}(X; R)$ (similarly for compact supports).

Let $(\Delta^q)_{q=0}^{\infty}$ denote the standard q-simplicies. Fix compatible triangulations of $\mathbb{R}\times\Delta^q$ such that $[0, \infty)\times\Delta^q$ and $(-\infty, 0]\times\Delta^q$ are subcomplexes. This induces a suspension map of complexes $\Delta_{\cdot}(X; R)\xrightarrow{S}\Delta_{\cdot+1}(\mathbb{R}\times X; R)$.

Lemma A.8 *Let X be a topological space. Then the suspension map induces an isomorphism on homology*

$$ {}_{\mathrm{sing}}H_{\cdot}(X; R)\xrightarrow{\sim}{}_{\mathrm{sing}}H_{\cdot+1}(\mathbb{R}\times X; \mathbb{R}). $$

Proof (Compare [B, II., 2.1]). Let $\Delta_\cdot^{(-\varepsilon,\varepsilon)}$ (resp. $\Delta_\cdot^{<\varepsilon}$, $\Delta_\cdot^{>-\varepsilon}$) denote the subcomplex of chains in $\Delta_\cdot(\mathbb{R}\times X;R)$, all of whose singular simplicies lie in $(-\varepsilon,\varepsilon)\times X$ (resp. $(-\infty,\varepsilon)\times X$, $(-\varepsilon,\infty)\times X$). Note that the image of S lies in $\Delta'_\cdot=\Delta_\cdot^{<\varepsilon}+\Delta_\cdot^{>-\varepsilon}$. We choose $\varepsilon>0$ sufficiently small so that elements in the image of S involve no simplicies contained in $(-\varepsilon,\varepsilon)\times X$. By the usual subdivision argument ([Sp, 4.4.14]), the inclusion $\Delta'_\cdot\subseteq\Delta_\cdot(\mathbb{R}\times X;R)$ is a homotopy chain equivalence. It thus suffices to show that S induces an isomorphism ${}_{\text{sing}}H_\cdot(X;R)\xrightarrow{\sim}H_{\cdot+1}(\Delta'_\cdot)$.

Note that $\Delta_\cdot^{>-\varepsilon}$ (resp. $\Delta_\cdot^{<\varepsilon}$) is acyclic (one may homotope any cycle out to ∞ (resp. $-\infty$) – compare [B, II., Lemma 2.2]). Thus the connecting homomorphism associated to the short exact sequence

$$0\to\Delta_\cdot^{(-\varepsilon,\varepsilon)}\to\Delta_\cdot^{<\varepsilon}\oplus\Delta_\cdot^{>-\varepsilon}\to\Delta'_\cdot\to 0$$

yields an isomorphism $\partial_*\colon H_{\cdot+1}(\Delta'_\cdot)\xrightarrow{\sim}H_\cdot(\Delta_\cdot^{(-\varepsilon,\varepsilon)})$. But it is easy to check that the diagram

$$\begin{array}{ccc} H_{\cdot+1}(\Delta'_\cdot) & \xrightarrow{\partial_*} & H_\cdot(\Delta_\cdot^{(-\varepsilon,\varepsilon)}) \\ {\scriptstyle S_*}\uparrow & & \uparrow{\scriptstyle i_*} \\ {}_{\text{sing}}H_\cdot(X;R) & = & {}_{\text{sing}}H_\cdot(X;R) \end{array}$$

commutes, where i is the inclusion $x\mapsto(0,x)$. The lemma follows since i_* is an isomorphism as for finite chains. □

Corollary A.9 *Let X be a* cs*-space of dimension n and $\mathcal{X}$ a* cs*-stratification. Then for $x\in S_{n-k}$ with link L,*

$$H^i((\Delta'_\cdot)_x)\cong\begin{cases} {}_{\text{sing}}\tilde{H}^c_{k-i-n-1}(L;R) & \text{if } L\neq\emptyset, \\ R & \text{if } L=\emptyset \text{ and } i=n-k, \\ 0 & \text{otherwise.}\end{cases} \tag{A.4}$$

Proof. $\Delta'_\cdot{}_X$ is clearly $\mathcal{X}$-clc, so by [B, V., 3.10(b)] and Proposition A.6 the direct system $\mathbb{H}^i(U;\Delta'_\cdot{}_X)={}_{\text{sing}}\mathbb{H}_{-i}(U;R)$ is constant over distinguished neighborhoods $U=B\times\mathring{c}(L)$ of x. It thus suffices to prove (A.4) with the left-hand side replaced by ${}_{\text{sing}}H_{-i}(U;R)$.

By Lemma A.8 the $L=\emptyset$ case is immediate, while for $L\neq\emptyset$ we have ${}_{\text{sing}}H_{-i}(U;R)\cong{}_{\text{sing}}H_{k-i-n}(\mathring{c}(L);R)$. Let $\mathring{c}(L)'\Subset\mathring{c}(L)$ be a subcone; note that $\Delta_\cdot(\mathring{c}(L)\setminus\mathring{c}(L)';R)$ is acyclic. Then the long exact sequence associated to

$$0\to\Delta_\cdot(\mathring{c}(L)\setminus\mathring{c}(L)';R)\to\Delta_\cdot(\mathring{c}(L);R)\to\Delta_\cdot(\mathring{c}(L),\mathring{c}(L)\setminus\mathring{c}(L)';R)\to 0$$

shows that ${}_{\text{sing}}H_{k-i-n}(\mathring{c}(L);R)\cong{}_{\text{sing}}H_{k-i-n}(\mathring{c}(L),\mathring{c}(L)\setminus\mathring{c}(L)';R)$. We can use finite chains here since $\mathring{c}(L)'$ is compact, so the corollary now follows from homotopy invariance of ${}_{\text{sing}}H^c_\cdot$ and the usual exact sequence of the pair. □

Lemma A.10 *Let X be a* cs*-space of dimension n and $\mathcal{X}$ a* cs*-stratification. Then $\Delta'_\cdot{}_X$ is $\mathcal{X}$-*cc *and the complex of its cohomology sheaves is bounded.*

Proof. We just need to prove the finite generation and uniform boundedness of $H^i((\Delta'_\cdot{}_X)_x)$; let k be such that $x\in S_{n-k}$ and use induction on k. The case $k=0$

and more generally $L=\emptyset$ follows from Corollary A.9, so assume $k>0$ and $L\neq\emptyset$. Consider the exact sequence

$$\text{(A.5)} \qquad \ldots \to H^i((j_k^! \mathcal{A}_X^{\cdot})_x) \to H^i((\mathcal{A}_X^{\cdot})_x) \to H^i((R i_{k*} i_k^* \mathcal{A}_X^{\cdot})_x) \to \ldots$$

By [B, V., 3.11(iii)], Proposition A.6, and the inductive hypothesis, the third term is finitely generated and bounded. Now as in the proof of [B, V., 3.11(ii)], $H^i((j_k^! \mathcal{A}_X^{\cdot})_x) = H^{i+n-k}(f_x^! \mathcal{A}_X^{\cdot})$, which by Proposition A.6 and [B, V., 3.10(a)(c)] is just $\mathbb{H}_c^{i+n-k}(U; \mathcal{A}_X^{\cdot}) = {}_{\text{sing}}H_{k-i-n}^c(U; R)$ for a distinguished neighborhood U. Since U is contractible, this is just R in degree $k-n$. This implies the middle term of (A.5) is finitely generated and bounded. □

Theorem A.11 *Let X be a* cs*-space. Then $\mathcal{D}_X^{\cdot} = \mathcal{A}_X^{\cdot}$.*

Proof. For a contractible local neighborhood U, $\mathbb{H}_c^{-q}(U; \mathcal{A}_X^{\cdot}) = {}_{\text{sing}}H_q^c(U; R)$ is just R in degree 0. In view of the spectral sequence $\text{Ext}^p(\mathbb{H}_c^{-q}(U; \mathcal{A}_X^{\cdot}), R) \Rightarrow \mathbb{H}^{\cdot}(U; \mathcal{D}_X^{\cdot} \mathcal{A}_X^{\cdot})$ (see [B, V., 7.7(3)]), this implies $\mathcal{D}_X^{\cdot} \mathcal{A}_X^{\cdot} = R_X$. By Lemma A.10 and Proposition A.7 we can apply biduality ([B, V., 8.10]) to conclude that $\mathcal{A}_X^{\cdot} = \mathcal{D}_X^{\cdot}$. □

Corollary A.12 *Let X by a* cs*-space of dimension n with* cs*-stratification $\mathcal{X}$. For $x \in S_{n-k}$ with link L,*

$$H^i((\mathcal{D}_X^{\cdot}[-n])_x) = 0 \begin{cases} \text{for } i > k, \\ \text{for } i = k \text{ provided } L \neq \emptyset, \text{ and} \\ \text{for } i = k-1 \text{ provided } L \text{ is path-connected.} \end{cases}$$

Proof. Apply Corollary A.9 and Theorem A.11. □

References

[B] Borel, A. et al.: Intersection Cohomology (Progr. Math., vol. 50) Boston: Birkhauser 1984

[Br] Bredon, G.: Sheaf Theory. New York: McGraw-Hill 1967

[C] Cartan, H.: Séminaire de topologie algébrique de l'Ecole Normale Supérieure. Paris 1950/51, 2. éd. Paris: Secrétariat mathématique 1955

[D] Deligne, P.: Théorie de Hodge II. Publ. Math., Inst. Hautes Étud. Sci. **40**, 5–58 (1977)

[G] Godement, R.: Topologie algébrique et Théorie des faisceaux. (Actual. Sci. Ind. 1252) Paris: Hermann 1958

[GM1] Goresky, M., MacPherson, R.: Intersection Homology Theory. Topology **19**, 135–162 (1980)

[GM2] Goresky, M., MacPherson, R.: Intersection Homology II. Invent. Math. **72**, 77–130 (1983)

[HS] Habegger, N., Saper, L.: Intersection homology theory and Zeeman's filtration of homology. (Preprint, Fall 1986)

[I] Iverson, B.: Cohomology of Sheaves. Berlin Heidelberg New York: Springer 1986

[K] King, H.: Topological invariance of intersection homology without sheaves. Topology Appl. **20**, 149–160 (1985)

[M1] McCrory, C.: Zeeman's filtration of homology. Trans. Am. Math. Soc. **250**, 147–166 (1979)

[M2] McCrory, C.: Topology of the Deligne weight filtration. Proc. Symp. Pure Math. **40**, 217–226 (1983)

[Q] Quinn, F.: Intrinsic skeleta and intersection homology of weakly stratified sets. In:

McCrory, C., Shifrin, T. (eds.) Geometry and Topology: Manifolds, Varieties, and Knots (Lec. Notes Pure Appl. Math. vol. 105) New York: Marcel Dekker 1987

[S] Siebenmann, L.: Deformations of homeomorphisms on stratified sets. Comment. Math. Helv. **47**, 123–163 (1972)

[Sp] Spanier, E.: Algebraic Topology. New York: McGraw-Hill 1966

[Z] Zeeman, E.C.: Dihomology III. A generalization of the Poincaré duality for manifolds. Proc. Lond. Math. Soc., III. Ser. **13**, 153–183 (1963)

[Zu] Zucker, S.: L_2-cohomology of warped products and arithmetic groups. Invent. Math. **70**, 169–218 (1982)

Invent. math. 105, 273–303 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Classification of complete minimal surfaces in $\mathbf{R}^3$ with total curvature 12π

C.J. Costa *

Universidade Federal Fluminense, Departamento de Geometria, Rua São Paulo, sln°, 2400 Niterói-RJ, Brasil

Oblatum 15-XI-1989 & 27-VIII-1990

1. Introduction

A fundamental question for minimal surfaces in $\mathbf{R}^3$ is the classification of all the embedded complete minimal surfaces in $\mathbf{R}^3$ with finite total curvature. R. Ossermann [9] has shown that if M is a complete minimal surface in $\mathbf{R}^3$ with finite total curvature then M is conformally equivalent to a compact Riemann surface of genus γ, $\bar{M}_\gamma$, punctured at n points. Jorge and Meeks [7] have proved that if M is embedded then $C(M)=\int|K|\,dm=4\pi m$, $m=\gamma+n-1$, where $C(M)$ is the total curvature of M. R. Ossermann [9] has also proven that the plane and the catenoid are the only complete minimal surfaces embedded in $\mathbf{R}^3$ with $m=0$ and $m=1$, respectively. Recently some authors have shown the non-existence of complete embedded minimal surfaces with $C(M)<\infty$, for some values of γ and n. Lopez and Ros [8] have proven the non-existence of examples for $\gamma=0$ and $n\geqq 3$ and R. Schoen [10] has proven the non-existence of examples with $\gamma>0$ and $n=2$.

On the other hand, the author [1] has constructed an example of a complete minimnal immersion in $\mathbf{R}^3$ of genus one and three embedded ends. We will call this surface "Surface S_1". Later, D. Hoffman and W. Meeks [3] have proven that S_1 is embedded and, inspired by this surface, they were able to construct a one parameter family, S_y, $y>1$, of complete minimal surfaces embedded in $\mathbf{R}^3$. Each S_y is conformally equivalent to $\mathbf{C}/L(iy)-\{\pi(0), \pi(1/2), \pi(iy/2)\}$, where $L(iy)=\{m+niy\in\mathbf{C};\ m,\ n\in\mathbf{Z}\}$ and $\pi\colon \mathbf{C}\to\mathbf{C}/L(iy)$ is the canonical projection. We remark that S_1 has a flat end and two ends of catenoid types and S_y, $y>1$, has there ends of catenoid type. Also, the author [2] has proven that S_1 is the only complete minimal surface embedded in $\mathbf{R}^3$ with finite total curvature, conformally equivalent to a torus punctured at three points and with one flat end.

The main result in this work is the proof that S_y, $y\geqq 1$ are all complete minimal surfaces embedded in $\mathbf{R}^3$ with finite total curvature and with the topology of a torus minus three points. This result follows from Theorems A and B below. Theorem A is a theorem of uniqueness, but we need the result of

* Partially supported by CNPq-Brazil.

existence of Hoffman-Meeks [4] to conclude the classification. This result of existence can be obtained, through the maximum principle, as an easy consequence of the result developed in the proof of Theorem A (principally, the control of the logarithmic growth of the ends of catenoid type of S_y, $y>1$). Thus, to make this work self-contained, we present in Theorem B a proof of the existence of the surfaces S_y, $y>1$.

We say that two minimal surfaces M and M' are the same if there exists a rigid motion and a homothety that carries M onto M'. In this work we will prove the following theorems:

Theorem A (Uniqueness) *Let $X: M \mapsto \mathbf{R}^3$ be a complete minimal surface embedded in $\mathbf{R}^3$ of genus one and three ends of catenoid type. Then:*

(a) *M is conformally equivalent to $\mathbf{C}/L(iy) - \{\pi(0), \pi(1/2), \pi(iy/2)\}$, where $y>1$, $L(iy) = \{m + niy \in \mathbf{C};\ m, n \in \mathbf{Z}\}$ and $\pi: \mathbf{C} \to \mathbf{C}/L(iy)$ is the canonical projection,*
(b) *For each $y>1$ there exists at most one such surface X.*

Theorem B (Existence) *There exists a smooth one-parameter family, S_y, S_y: $M_y \mapsto \mathbf{R}^3$, $y \geqq 1$, of complete embedded minimal surfaces in $\mathbf{R}^3$ of genus one, three ends and finite total curvature such that:*

(a) *S_y is conformally equivalent to $\mathbf{C}/L(iy) - \{\pi(0), \pi(1/2), \pi(iy/2)\}$,*
(b) *S_1 has a flat end and two ends of catenoid types. That is, S_1 is the surface that appears in* [1], [2] *and* [3] *and*
(c) *S_y, $y>1$, thas three ends of catenoid type.*

Corollary B.1 *S_y, $y \geqq 1$, are all complete minimal surfaces embedded in $\mathbf{R}^3$ of genus one, three ends and finite total curvature.*

Corollary B.2 *The surfaces S_y, $y \geqq 1$, are all complete orientable minimal surface embedded in $\mathbf{R}^3$ with total curvature 12π.*

To prove Theorem A we will consider all the compact Riemann surfaces $\mathbf{C}/L(\tau)$, where $\tau = x + iy \in FM = \{\tau = x + iy \in \mathbf{C};\ x^2 + y^2 \geqq 1,\ y>0,\ |x| \leqq 1/2\}$ and $L(\tau) = \{m + n\tau \in \mathbf{C};\ m, n \in \mathbf{Z}\}$. First of all we will prove that if M is conformally equivalent to to $\mathbf{C}/L(\tau)$ punctures at three points and x is not zero then M cannot be embedded in $\mathbf{R}^3$ as a complete minimal surface with finite total curvature. This is contained in the Corollary 2.2 of Lemma 2 (see the Remark 2 after this corollary). By using this result we can consider M conformally equivalent to $\mathbf{C}/L(iy)$, $y \geqq 1$, punctured at three points q_1, q_2, q_3. Then, if M is a complete minimal surface embedded in $\mathbf{R}^3$ with finite total curvature, we will prove that M is, in fact, conformally equivalent to $\mathbf{C}/L(iy)$ punctured at the points $\pi(0)$, $\pi(1/2)$, $\pi(iy/2)$, where $\pi: \mathbf{C} \mapsto \mathbf{C}/L(iy)$ is the canonical projection and $y \geqq 1$. Afterwards, we will prove that for each $y>1$ there exists at most one embedding of $M_y = \mathbf{C}/L(iy) - \{\pi(0), \pi(1/2), \pi(iy/2)\}$ in $\mathbf{R}^3$ as a complete minimal surface with three ends of catenoid types. In fact, we find a smooth one-parameter family of complete minimal immersions $I_y: M_y \to \mathbf{R}^3$, $y \geqq 1$ and such that: (a) I_1 is an embedding, that is, I_1 is the surface S_1, (b) For $y>1$, each member of this family has all ends of catenoid type and (c) I_y is given by their Weierstrass representation (g_y, w_y) and we have control over the logarithmic growth of their ends.

To prove Theorem B we use the one-parameter family found in Theorem A. The properties of this family deduced during the proof of this theorem together

with the maximum principle imply that I_y is an embedding for every $y \geqq 1$. That is $I_y = S_y$ for every $y \geqq 1$.

The Corollary B.1 follows immediately from Theorems A and B. To prove Corollary B.2, we consider M a complete minimal surface embedded in $\mathbf{R}^3$ with total curvature 12π. Then M is conformally equivalent to a compact Riemann surface $\bar{M}_\gamma$, of genus γ punctured at n points. So, it follows from the formula of Jorge-Meeks [7] for the total curvature of a complete embedded minimal surface in $\mathbf{R}^3$ that $12\pi = -2\pi(2-2\gamma-2n)$ and therefore $\gamma+n=4$. But M cannot be embedded in $\mathbf{R}^3$ as a complete minimal surface of finite total curvature in the three following cases: (1) $M \cong S^2 - \{q_1, \ldots, q_n\}$, $3 \leqq n \leqq 5$, (2) $M \cong M_\gamma - \{q_1, q_2\}$, $\gamma > 0$ and (3) $M \cong M_\gamma - \{q_1\}$, $\gamma > 0$. Cases (1) and (2) were proven in [7] and [10] respectively and case (3) results from the fact that the coordinates of a minimal immersion are harmonic functions. Thus, $\gamma = 1$ and $n = 3$ and we use Corollary B.1 to conclude the proof of Corollary B.2.

In all this work we consider compact Riemann surfaces of genus one, $\bar{M}$, with complex structures induced by lattices $L(\tau)$ of $\mathbf{C}$ with $\tau \in FM$. Through π: $\mathbf{C} \mapsto \mathbf{C}:L(\tau)$ we identify elliptic functions and elliptic differentials of $L(\tau)$ with meromorphic functions and meromorphic differentials of $\bar{M}$, respectively. Then, we consider Weierstrass representation (g, w), of immersions X: $M \mapsto \mathbf{R}^3$ with the properties of Theorems A, were $\bar{M} = M - \{q_1, \ldots, q_n\}$ and $\bar{M} = \mathbf{C}/L(\tau)$.

Acknowledgements. The author wishes to thank the Université de Savoie and the Université de Grenoble I for their hospitality during the preparation of this work.

2. Complete minimal immersions in $\mathbf{R}^3$

In [9], we find the following representation of complete minimal surfaces, called Enneper-Ossermann-Weiserstrass representation:

Theorem 1 *Let X: $M \mapsto \mathbf{R}^3$ be a complete minimal immersion of finite total curvature. Then*

(a) *M is conformally equivalent to a compact Riemann surface of genus γ, $\bar{M}_\gamma$, punctured at n points.*

(b) *there exist a meromorphic function g and a meromorphic differential w in $\bar{M}_\gamma$ such that w is holomorphic in M and $q \in M$ is a pole of order m of g if and only if q is a zero of order $2m$ of w. Furthermore, g is the Gauss normal of the immersion,*

(c) *if α is a closed path in M, then*

$$\operatorname{Re} \int_\alpha gw = 0, \qquad \overline{\int_\alpha w} = \int_\alpha g^2 w \quad \textit{and}$$

(d) *every divergent path in M has infinite length.*

Conversely, let M be as in (a) and let g and w be a meromorphic function and a meromorphic differential in $\bar{M}_\gamma$, respectively. If (g, w) satisfies (b) and (c) then X: $R \to \mathbf{R}^3$,

$$X(q) = \operatorname{Re} \int_{q_0}^{q} ((1-g^2 w), i(1+g^2 w), 2gw), \quad q_0 \in M$$

is a minimal immersion with finite total curvature. Furthermore if X, satisfies (d), X is complete.

If $D_j \subset \bar{M}_\gamma$, $j = 1, \ldots, n$ is a small topological disk with $q_j \in D_j$, then $F_j = X(M \cap D_j)$ is an end of the immersion. Let $1 \leqq j \leqq n$ fixed. Since g is the Gauss

map of X we can suppose, after a rotation of X in $\mathbf{R}^3$, that $g(q_j)=0$. In this situation, see [7], F_j is an embedded end if and only if q_j is a pole of order two of w. Then, around q_j, we have the local expressions

$$g(z)=a_n z^n+o(z)^{n+1}, \quad a_n\neq 0, n\geqq 1 \quad \text{and} \quad w(z)=\frac{b}{z^2}+o(z)^{-1}, b\neq 0.$$

We say that the embedded end F_j is of *catenoid type* if $n=1$ and is a *flat end of order* $n-1$, *if* $n>1$. In the last case, the coordinate $X^3(q)=2\operatorname{Re}\int gw$, $q\in D_j\cap M$, is bounded and the immersion approaches a plane parallel to the plane $X^3=0$.

3. Elliptic functions and minimal immersions of genus one

Let $FM=\{\tau=x+iy\in\mathbf{C};\ x^2+y^2\geqq 1,\ y>0,\ |x|\leqq 1/2\}$ and let $L(\tau)$ be the lattice in $\mathbf{C}$ given by $L(\tau)=\{m+n\tau\in\mathbf{C};\ m,\ n\in\mathbf{Z}\}$ where $\tau\in FM$. Then $\mathbf{C}/L(\tau)$ with the complex structures induced by π: $\mathbf{C}\mapsto\mathbf{C}/L(\tau)$ are all Riemann surfaces of genus one. That is, see [5], we have the following theorem:

Theorem 2 *Let L be a lattice in $\mathbf{C}$, and let $\bar{M}=\mathbf{C}/L$ equipped with the complex structure induced by L. Then there exists $\tau\in FM$ such that $\mathbf{C}/L(\tau)$ is conformally equivalent to $\bar{M}$. Furthermore, if $\mathbf{C}/L(\tau)$ and $\mathbf{C}/L(\tau')$, τ, $\tau'\in F$ are conformally equivalent, then $\tau=\tau'$ or τ, $\tau'\in\partial F$.*

Remark 1 Henceforth, we shall consider only the compact Riemann surfaces of genus one, $\bar{M}=\mathbf{C}/L(\tau)$, where $L=L(\tau)$ is a lattice as in Theorem 2.

Two points z_1, $z_2\in\mathbf{C}$ are *L-congruent*, $z_1\equiv z_2$, if $z_1-z_2\in L$. Otherwise, they are *incongruent* and we shall write $z_1\not\equiv z_2$. A set $\{z_1, \ldots, z_n\}$ is *incongruent* if $z_j\not\equiv z_k$, $j\neq k$, $j, k=1, \ldots, n$. We define the middle-points of $L(\tau)$,

$$w_1=\frac{1}{2}, \quad w_2=-\frac{1+\tau}{2}, \quad w_3=\frac{\tau}{2}$$

and the numbers

$$e_k=P(w_k), \quad k=1, 2, 3$$

where P is the Weierstrass function of $L(\tau)$. We also have the quasielliptic function ζ: $\mathbf{C}\mapsto\mathbf{C}\cup\infty$ of $L(\tau)$ and the numbers $\eta_j\in\mathbf{C}$, $j=1, 2, 3$ such that,

$$\zeta'(z)=-P(z), \quad \zeta(w_j)=\eta_j, \quad \zeta(z+2w_j)=\zeta(z)+2\eta_j.$$

Furthermore, ζ has the following additive property:

$$\zeta(z_1\pm z_2)=\zeta(z_1)\pm\zeta(z_2)+\frac{P'(z_1)\mp P'(z_2)}{P(z_1)-P(z_2)}$$

and at $0\in\mathbf{C}$,

$$\zeta(z)=\frac{1}{z}+o(z)^3 \quad \text{and} \quad P(z)=\frac{1}{z^2}+o(z)^2.$$

4. Proof of Theorem A

To prove Theorem A we shall need Lemmas 1, 2 and 3 below.

Lemma 1 *Let X: $M=\mathbf{C}/L(\tau)-\{q_1, q_2, q_3\}\mapsto\mathbf{R}^3$ be a complete minimal surface embedded in $\mathbf{R}^3$ with three ends of catenoid type, where $\tau=x+iy\in FM$. Then, after a rotation, a translation and a homothety of X in $\mathbf{R}^3$ and a translation of coordinates of $\mathbf{C}$, there exist $z_1\in\mathbf{C}$, $z_1\not\equiv 0$, $z_1\not\equiv w_2$ and b_1, $b_2\in\mathbf{R}$ such that*

(a) $b_2 > -b_1 > 0$ *and* $\frac{1}{2} \leqq \frac{-b_1}{b_2} \leqq 1$,

(b) $\frac{\pi b_1^2}{P(z_1)-e_2} + \frac{\pi b_2^2}{P(z_1+w_2)-e_2} = y(2\eta_1+e_2)-\pi$,

(c) $y|2\eta_1+e_2|^2 - 2\pi \operatorname{Re}(2\eta_1+e_2) = 2\pi b_1 b_2$ *and*

(d) $H_\lambda(z_1) = E(z_1) - \lambda E(z_1+w_2) = 0$,

where $\lambda = -b_1/b_2$ *and* $E(z) = -2\eta_1 z + \zeta(z) + (2\pi/y)i \operatorname{Im} z$. *Furthermore,* E: $\mathbf{C} \mapsto \mathbf{C} \cup \infty$ *is a* C^∞*-periodic function of* $L(\tau)$.

Proof of (a). Let (g, w) be the Weierstrass representation of X. So, g is the Gauss normal of X and the orientation of M is induced by g. As M is a complete minimal surface embedded in $\mathbf{R}^3$ the vectors $g(q_j) \in S^2$, $j = 1, 2, 3$, are parallel. Furthermore, X is a proper embedding. So, $(\mathbf{R}^3 - M)$ has two connected components with common boundary M. Then, after a rotation of M in $\mathbf{R}^3$, we can suppose that

$$g(q_1) = g(q_2) = \infty, \qquad g(q_3) = 0$$

and

$$\lim_{q \to q_1} X^3(q) = \lim_{q \to q_3} X^3(q) = \infty, \qquad \lim_{q \to q_2} X^3(q) - = \infty, \tag{1}$$

where $X^3(q)$ is the third coordinate of the immersion. As the total curvature is 12π, g is an elliptic function in $L(\tau)$ of order 3. Let F_j be the ends of X associated to the points q_j, $j = 1, 2, 3$. As the ends are of catenoid type, q_1 and q_2 are simple poles of g and q_3 is a simple zero of g. Also, let $q' \in M$ be the only point of M where g has a simple poles. These facts and Theorem 1 imply that w is an elliptic differential in $L(\tau)$ of order two, w has a double pole at q_3, a double zero at q' and q' is the only zero of w. So, after a rotation of X around the x^3-axis in $\mathbf{R}^3$, a homothety in $\mathbf{R}^3$ and a translation of coordinates of $\mathbf{C}$, we find

$$w = (P(z) - e_s)\, dz, \qquad s \in \{1, 2, 3\}, \qquad \pi(0) = q_3 \quad \text{and} \quad \pi(w_s) = q'. \tag{2}$$

Let $z_j \in \mathbf{C}$ be such that

$$\pi(z_j) = q_j, \qquad j = 1, 2.$$

The Fig. 1 below shows g and w for the case $s = 2$.

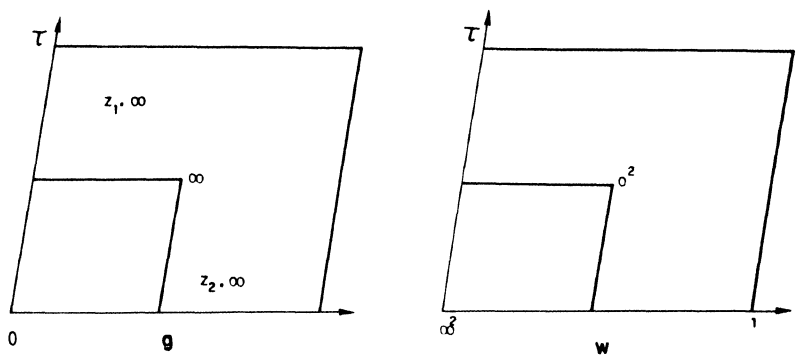


Fig. 1

Henceforth, we will use the following notation:

$$q=e^{i\pi\tau}, \quad \tau=x+iy\in FM \quad \text{and} \quad h=e^{-\pi y}. \tag{3}$$

We observe that $\{0, z_1, z_2, w_s\}$ is $L(\tau)$-incongruent and $P'(w_s)=0$. This implies that $P(z_j)-P(w_s)=P(z_j)-e_s\neq 0$, $j=1, 2$. So, by using this fact, we conclude that there exist $b_1, b_2\in\mathbf{C}-\{0\}$ such that

$$\begin{aligned} g(z)=a\zeta(z-w_s)&+\frac{b_1}{P(z_1)-e_s}\zeta(z-z_1)\\ &+\frac{b_2}{P(z_2)-e_s}\zeta(z-z_2)+a\zeta(w_s)+\frac{b_1}{P(z_1)-e_s}\zeta(z_1)\\ &+\frac{b_2}{P(z_2)-e^s}\zeta(z_2), \end{aligned} \tag{4}$$

where

$$a+\frac{b_1}{P(z_1)-e_s}+\frac{b_2}{P(z_2)-e_s}=0. \tag{5}$$

Furthermore, from (2), (4) and by using the local expression for ζ at $0\in\mathbf{C}$, we find that at $z=z_j, j=1, 2$,

$$\begin{aligned} g(z)&=\frac{b_j}{(P(z_j)-e_s)(z-z_j)}+o(z-z_j)^0 \quad \text{and}\\ gw(z)&=\left[\frac{b_j}{z-z_j}+o(z-z_j)^0\right]dz. \end{aligned} \tag{6}$$

As gw has simple poles at the points 0, z_1 and z_2 and a simple zero at w_s, from (6) and using (c) of Theorem 1 we obtain

$$\begin{aligned} gw(z)=[&-(b_1+b_2)\zeta(z)+b_1\zeta(z-z_1)+b_2\zeta(z-z_2)\\ &+(b_1+b_2)\zeta(w_s)-b_1\zeta(w_s-z_1)-b_2\zeta(w_s-z_2)]\,ds,\\ &b_j\in\mathbf{R}-\{0\}, \quad j=1, 2. \end{aligned} \tag{7}$$

Again, by using (c) of Theorem 1 and (2) we find

$$2\pi i \operatorname{res}_z g^2 w=-2\pi i \operatorname{res}_z w=0, \quad z\in\mathbf{C}. \tag{8}$$

Then $g^2 w$ is a elliptic differential in $L(\tau)$ with double poles at z_1 and z_2 and all the residues are zero. So, there exist $a_1, a_2\in\mathbf{C}-\{0\}$ and $a_3\in\mathbf{C}$ such that

$$g^2 w=[a_1 P(z-z_1)+a_2 P(z-z_2)+a_3]\,dz. \tag{9}$$

Now, by using (5) and the additive property of ζ and by comparison of the Laurent development of (9) and the Laurent development of (2) and of (4) at the points 0, z_1, z_2, respectively, we find

$$a_j=\frac{b_j^2}{P(z_j)-e_s}, j=1, 2 \quad \text{and} \quad a_3=2b_1 b_2-e_s\sum_{k=1}^{2}\frac{b_k^2}{P(z_k)-e_s}.$$

So, these results and (9) imply that

$$g^2 w(z)=\left\{\frac{b_1^2}{P(z_1)-e_s}[P(z-z_1)-e_s]\right.$$
$$\left.+\frac{b_2^2}{P(z_2)-e_s}[P(z-z_2)-e_s]+2b_1 b_2\right\} dz. \tag{10}$$

Now, we define the paths $\mu_j: [0, 1]\mapsto \mathbf{C}, j=1, 3$, by

$$\mu_1(t)=t+v_1\tau, \qquad \mu_3(t)=t\tau+v_3, \tag{11}$$

where $0<v_j<1$ are such that $\pi(\mu_j(t))\in M$. Then, from (2) and (10) we find

$$\int_{\mu_j} g^2 w=\left(\sum_{k=1}^{2}\frac{-2b_k^2}{P(z_k)-e_s}\right)(\eta_j+w_j e_s)+4w_j b_1 b_2$$

and

$$\int_{\mu_j} w=-2\eta_j-2w_j e_s, \qquad j=1, 3.$$

So, the equations above, the Legendre's relation, $2\eta_3=2\eta_1\tau-2\pi i$ and (c) of Theorem 1 imply

$$\left(\frac{b_1^2}{P(z_1)-e_s}+\frac{b_2^2}{P(z_2)-e_s}\right)(2\eta_1+e_s)-2b_1 b_2=\overline{2\eta_1+e_s}$$

and

$$y(\overline{2\eta_1+e_s})-\pi=\frac{\pi b_1^2}{P(z_1)-e_s}+\frac{\pi b_2^2}{P(z_2)-e_s}. \tag{12}$$

From the equations above we conclude that

$$y|2\eta_1+e_s|^2-2\pi\,\mathrm{Re}(2\eta_1+e_s)=2\pi b_1 b_2. \tag{13}$$

On the other hand, we observe that g is an elliptic function with simple poles at z_1, z_2 and w_s and g has a zero at $0\in\mathbf{C}$. These facts, with (2), (8), (12) and (13) show that $z_1+z_2\neq 0$ and g and w satisfy the hypothesis of Proposition 2 of [2]. Then

$$z_2\equiv z_1+w_s, \quad \text{and} \quad \frac{P'(z_1)}{P(z_1)-e_s}=-\frac{P'(z_2)}{P(z_2)-e_s}. \tag{14}$$

Now, we can prove (a) of Lemma 1. From (1) and by using local expressions of the differential that appears in (7) at the points 0, z_1 and z_2 we find that $b_1+b_2>0$, $b_1<0$ and $b_2>0$. In particular $-b_1/b_2\leqq 1$.

In order to prove that $-b_1/b_2\geqq 1/2$, let D_0 and D_1 be two small disks centered at $0\in\mathbf{C}$ and $z_1\in\mathbf{C}$, respectively. Let $\varphi_j: D_j\mapsto\mathbf{C}$ be defined by $\varphi_j(z)=[X^1(z)+iX^2(z)]^{-1}$ if $z\neq 0$, $j=0$, 1, and $\varphi_0(0)=\varphi_1(z_1)=0$, where X^1 and X^2

are, respectively, the first and second coordinate of X. From Theorem 1 we have that $X^1(z)+iX^2(z)=\int w-\int g^2 w$. So, from (2) and (10) we find

$$\lim_{z\mapsto 0}\frac{\varphi_0(z)}{z}=-1 \quad \text{and} \quad \lim_{z\mapsto z_1}\frac{\varphi_1(z)}{z-z_1}=\frac{P(z_1)-e_s}{b_1^2}\neq 0.$$

Thus, φ_0 and φ_1 are diffeomorphisms in a neighborhood of $0\in\mathbf{C}$ and $z_1\in\mathbf{C}$, respectively. So, we can define a diffeomorphism $\xi\colon D_0'\to D_1'$ by $\xi(z)=(\varphi_1)^{-1}(\varphi_0(z))$, where $D_0'\subseteq D_0$ and $D_1'\subseteq D_1$ are open disks centered at $0\in\mathbf{C}$ and $z_1\in\mathbf{C}$, respectively. Clearly, there exist $\delta>0$ and $L>0$ such that

$$\left|\frac{\xi(z)-z_1}{z}\right|\geqq L>0, \qquad 0<|z|<\delta \quad \text{and} \quad \xi(0)=z_1. \tag{15}$$

On the other hand, from (1) and by using that M separates $\mathbf{R}^3$ in two connected components we conclude

$$X^3(\xi(z))-X^3(z)>0, \qquad \text{for every } z\in D_0'-\{0\}. \tag{16}$$

Now we suppose that $-b_1/b_2<1/2$ and let $\varepsilon=2b_1+b_2>0$. Thus, from (7) and (15) we have

$$\begin{aligned}\tfrac{1}{2}\lim_{z\to 0}[X^3(\xi(z))-X^3(z)]&=\lim_{z\to 0}\{b_1\log|\xi(z)-z_1|+(b_1+b_2)\log|z|\}\\&=\lim_{z\to 0}\left\{b_1\log\left|\frac{\xi(z)-z_1}{z}\right|+\varepsilon\log|z|\right\}=-\infty.\end{aligned}$$

This result is a contradiction with (16). This finishes the proof of (a) of Lemma 1.

Now to finish the proof of Lemma 1 we need the following result:

Assertion 1 *Let $L(\tau)$ be a lattice of* $\mathbf{C}$, *where $\tau=x+iy\in FM$. Then (a) $y\,\mathrm{Re}(\eta_1+e_1)-2\pi>0$ and (b) $y|2\eta_1+e_3|^2-2\pi\,\mathrm{Re}(2\eta_1+e_3)>0$.*

Proof. Part (a) of assertion has been proved in [2, Proposition 3]. In order to prove (b) we will use the representation of $2\eta_1+e_3$ by the function $v_4(z)$ of Jacobi (see [11, vol. 1, pp. 16 and 32]):

$$2\eta_1+e_3=-\frac{v_4''(0)}{v_4(0)}, \qquad \text{where } v_4(z)=1+2\sum_{n=1}^{\infty}(-1)^n q^{n^2}\cos 2\pi n z.$$

So,

$$\begin{aligned}y|2\eta_1-e_3|^2-2\pi\,\mathrm{Re}(2\eta_1+e_3)=\frac{1}{|v_4(0)|^2}\{y|v_4''(0)|^2\\+2\pi\,\mathrm{Re}[v_4''(0)\,\overline{v_4(0)}]\}.\end{aligned}$$

We will use the notation (3). Then, for fixed $y\geqq\sqrt{3}/2$, it is sufficient to show that

$$4\pi y G_1(x)-G_2(x)>0,$$

where $G_1(x)$ and $G(x)$ are the functions

$$G_1(x)=\left(\sum_{n=1}^{\infty}(-1)^n n^2 h^{n^2}\cos \pi n^2 x\right)^2+\left(\sum_{n=1}^{\infty}(-1)^n n^2 h^{n^2}\sin \pi n^2 x\right)^2$$

and

$$G_2(x)=\left(\sum_{n=1}^{\infty}(-1)^n n^2 h^{n^2}\cos \pi n^2 x\right)\left(1+2\sum_{n=1}^{\infty}(-1)^n h^{n^2}\cos \pi n^2 x\right)$$
$$+2\left(\sum_{n=1}^{\infty}(-1)^n n^2 h^{n^2}\sin \pi n^2 x\right)\left(\sum_{n=1}^{\infty}(-1)^n h^{n^2}\sin \pi n^2 x\right).$$

By using the notation (3) and the fact that $h^{-1}=e^{\pi y}>15$ we have

$$\sum_{n=1}^{\infty} h^{n^2}\leqq\sum_{n=1}^{\infty} n^2 h^{n^2}\leqq\frac{h}{1-4h^3}.$$

Thus, if $|x|\leqq 1/3$, then from the expressions below we find, after an easy computation, that $G_1(x)>0$ and $G_2(x)<0$. Also, if $1/3\leqq|x|\leqq 1/2$ we have

$$G_1(x)\geqq\left(\sum_{n=1}^{\infty}(-1)^n n^2 h^{n^2}\sin \pi n^2 x\right)^2\geqq\left(\frac{\sqrt{3}}{2}h-\frac{4h^4}{1-4h^3}\right)^2>\frac{h^2}{2}$$

and

$$G_2(x)\leqq\left(\sum_{n=2}^{\infty} n^2 h^{n^2}\right)\left(1+2\sum_{n=1}^{\infty} h^{n^2}\right)+2\left(\sum_{n=1}^{\infty} n^2 h^{n^2}\right)^2\leqq 3h^2.$$

These inequalities prove that $4\pi y G_1(x)-G_2(x)>0$ for every $|x|\leqq 1/2$. This finishes the proof of the Assertion 1.

Now, the Eq. (13) together (a) of Lemma 1 and Assertion 1 prove that $s=2$. This result and (14) carried in the Eqs. (12) and (13) proves (b) and (c) of Lemma 1.

In order to prove (d) we proceed the integration of the differential (7) with respect to the paths (11). As $s=2$, then from (7), (14) and by using the additive property of ζ, we find

$$g w\,(z)=\frac{1}{2}\left[\sum_{j=1}^{2} b_j\left(\frac{P'(z)+P'(z_j)}{P(z)-P(z_j)}+\frac{P'(z_j)}{P(z_j)-e_2}\right)\right]dz$$

and

$$z_2\equiv z_1+w_2.$$

We have from (c) of Theorem 1 that the real part of the integral of the differential gw along to the rectilinear paths (11) is zero. So, by using the expression for integrals of a meromorphic differential along straight lines that appears in [11, vol. 3, p. 148], we conclude that

$$\sum_{j=1}^{2} b_j\,\mathrm{Re}\left[-2\eta_k z_j+2w_k\zeta(z_j)+w_k\frac{P'(z_j)}{P(z_j)-e_2}\right]=0,\qquad k=1,3.$$

By using Legendre's relation, we find tht these equations are equivalent to the complex equation

$$\sum_{j=1}^{2} b_j\left[E(z_j)+\frac{1}{2}\frac{P'(z_j)}{P(z)-e_2}\right]=0 \tag{17}$$

where $E(z)$ is the function defined in (d) of Lemma 1. Also Legendre's relation and the additive property of ζ implies that E is a $C^\infty L(\tau)$-periodic function.

On the other hand, let $z, z' \in \mathbf{C}$ such that $z' \equiv z + w_2$. Then, the additive property of ζ and Legendre's relation imply

$$E(z)+\frac{1}{2}\frac{P'(z)}{P(z)-e_2}=E(z').$$

Now, by using twice the expression above and (14) we conclude that Eqs. (17) are equivalent to the complex equation $H_\lambda(z_1)=E(z_1)-\lambda E(z_1+w_2)=0$, $\lambda=-b_1/b_2$. This completes the proof of Lemma 1.

Henceforth, we represent with the same letters the $L(\tau)$-periodic functions E and H_λ and their quotient functions $E, H_\lambda: \mathbf{C}/L(\tau) \mapsto \mathbf{C} \cup \infty$ that we obtain. Our next goal is to use Lemma 1 above and the Lemma 2 that we will prove below to show that there does not exist a complete minimal surface M embedded in $\mathbf{R}^3$ with M conformally equivalent to a torus $\mathbf{C}/L(\tau)$ punctured at three points, where $\tau = x + iy \in FM$ and $x \neq 0$. In order to prove this, first we need to prove that all the zeros of the function H_λ of Lemma 1 are $\{w_1, w_3\}+L(\tau)$ for every $1/2 \leqq \lambda \leqq 2$ and $\tau = x+iy \in FM$. To show this we will prove that: (1) ∞ is a regular value of H_λ, $(H_\lambda)^{-1}(\infty)=\{0, w_2\}+L(\tau)$ and $JH_\lambda(z)>0$ if $z=0$ or $z=w_2$, where $JH_\lambda(z)$ is the Jacobian of H_λ at the point z. That is, we will prove that the degree of H_λ is two ; (2) 0 is a regular value of H_λ and $JH_\lambda(z)>0$, if $H_\lambda(z)=0$. By using these results, the fact that $H_\lambda(w_j)=0$, $j=1, 3$ and that H_λ is an odd function, we conclude that $\{w_1, w_3\}+L(\tau)$ are all the zeros of H_λ. We will prove (1) and (2) through Lemma 2 above and their corollaries.

Let $z=u+iv$ be global analytic coordinates of $\mathbf{C}$ and let $\tau = x+iy \in FM$. Also, we write $z=u+i\beta y$ where $v=\beta y$. Then if $0 \leqq u \leqq 1$ and $|\beta| \leqq 1/2$ we have a fundamental domain for the $L(\tau)$-periodic function H_λ. We define the conditions C_k, $k=1, 2$,

(C_1) *The angle* $\pi(2u-x)$ *is such that* $\cos\pi(2u-x) \leqq \cos(3\pi/10)$ *and the angle* $\pi(2u+x)$ *is such that* $\cos(2u+x) \leqq 0$,

(C_2) *The angle* $\pi(2u-x)$ *is such that* $\cos\pi(2u-x) \leqq 0$
and the neighbourhood $U(\tau)$ *of* $w_1 = 1/2$,

$$U(\tau)=\{z=u+i\beta y \in \mathbf{C}; 0 \leqq u \leqq 1, |\beta| \leqq 1/10 \text{ and } z \text{ satisfies } (C_1) \text{ or } (C_2)\}.$$

Now we will prove the Lemma 2 above and their corollaries:

Lemma 2 *Let* $H_\lambda: \mathbf{C} \mapsto \mathbf{C} \cup \infty$ *be the function* $H_\lambda(z)=E(z)-\lambda E(z+w_2)$, *where* $E(z)$ *is the* $L(\tau)$*-periodic function defined in* (d) *of Lemma* 1, $\tau = x+iy \in FM$, $1/2 \leqq \lambda \leqq 2$ *and* $x \geqq 0$. *Then*

(a) $(H_\lambda)^{-1}(\infty)=\{0, w_2\}+L(\tau)$, $JH_\lambda(0)>0$ *and* $JH_\lambda(w_2)>0$, *where* $JH_\lambda(z)$ *is the Jacobian of* H_λ *at point* z.

(b) *If* $z=u+iv=u+i\beta y\in U(\tau)$ *is such that* $0\leqq\beta\leqq 1/10$ *and* $H_\lambda(z)=0$ *then* $JH_\lambda(z)>0$.
(c) *If* $z=u+iv=u+i\beta y\notin U(\tau)$ *where* $0\leqq u\leqq 1$ *and* $0\leqq\beta\leqq 1/4$ *then* $H_\lambda(z)\neq 0$.

Corollary 2.1 *We have that* $(H_\lambda)^{-1}\{0\}=\{w_1, w_3\}+L(\tau)$ *for every* $1/2\leqq\lambda\leqq 2$ *and* $\tau=x+iy\in FM$.

Corollary 2.2 *Let* M *be a complete minimal surface embedded in* $\mathbf{R}^3$ *of genus one, three ends and finite total curvature. We have to possibilities*:

(a) M *has a flat end and then* M *is the surface* S_1 *that appears in* [1], [2] *and* [3] *or*
(b) M *has three ends of catenoid type and then* M *is conformally equivalent to* $\mathbf{C}/L(iy)-\{\pi(0), \pi(1/2, \pi(iy/2)\}$, *where* $y>1$ *and* $\pi\colon \mathbf{C}\to\mathbf{C}/L(iy)$ *is the canonical projection.*

Remark 2 In particular Corollary 2.2 proves that if M is conformally equivalent to $\mathbf{C}/L(i\tau)$ punctured at three points where $\tau=x+iy\in FM$ and $x\neq 0$ then M cannot be embedded in $\mathbf{R}^3$ as a complete minimal surface with finite total curvature.

Proof of (a) *of Lemma 2.* We observe that $(H_\lambda)^{-1}(\infty)=\{0, w_2\}+L(\tau)$. First we wil prove that $JH_\lambda((0))>0$. For this, by using the Legendre's relation and the additive property of ζ, we find that $E(z+w_2)=-2\eta_1 z+\zeta(z+w_2)-\eta_2+(2\pi/y)i\,\mathrm{Im}(z)$. So, if $z=u+i\beta y$ with $|\beta|\leqq 1/2$ then, by using the trigonometric developments of the functions $\zeta(z+w_2)-\eta_2$ and $\zeta(z)$, ([11, vol. 1, p. 189 and vol. 3, p. 118]), we find

$$H_\lambda(z)=\pi\,\mathrm{cotg}\,\pi z+4\pi\sum_{n=1}^{\infty}R(\lambda, n)\sin 2\pi nz+\frac{2\pi}{y}(1-\lambda)i\,\mathrm{Im}\,z \tag{18}$$

where

$$R(\lambda, n)=\frac{q^{2n}-\lambda(-1)^n q^n}{1-q^{2n}},\qquad q\in e^{i\pi\tau}. \tag{19}$$

Let $\varphi\colon U\mapsto\mathbf{C}$ be the function $\varphi(z)=1/H_\lambda(z)$, where $U\subset\mathbf{C}$ is a neighbourhood of $0\in\mathbf{C}$. We have that $\varphi(0)=0$. Let $\alpha_j\colon(-\varepsilon, \varepsilon)\mapsto\mathbf{C}, j=1, 2$ be the paths $\alpha_1(t)=\varphi(t)$ and $\alpha_2(t)=\varphi(it)$ where $\varepsilon>0$. By using (18) and (19) we find, after an easy calculation, that $\alpha_1'(0)=(1, 0)$ and $\alpha_2'(0)=(0, 1)$. So, $\{\alpha_1'(0), \alpha_2'(0)\}$ is a positive basis of $\mathbf{R}^2\approx\mathbf{C}$. This proves that $JH_\lambda(0)>0$ for every $1/2\leqq\lambda\leqq 2$. On the other hand,

$$H_\lambda(z+w_2)=-\lambda H_\mu(z),\qquad 1/2\leqq\mu=1/\lambda\leqq 2$$

and $JH_\mu(0)>0$. So, we conclude that $JH_\lambda(w_2)>0$. This finishes the proof of (a) of Lemma 2.

Proof of (b) *of Lemma 2.* Let $\psi(z)$ be the meromorphic function

$$\psi(z)=\pi\cot\pi z+4\pi\sum_{n=1}^{\infty}R(\lambda, n)\sin 2\pi nz, \tag{20}$$

where $R(\lambda, n)$ is defined in (19). Then from (18), (19) and (20) we have

$$JH_\lambda(z)=|\psi'(z)|^2+\frac{2\pi}{y}(1-\lambda)\,\mathrm{Re}\,\psi'(z).$$

Then it is sufficient to show that

$$\mathrm{Re}\,\psi'(z)+\frac{2\pi}{y}(1-\lambda)<0,\ \lambda\leqq 1 \quad \text{and} \quad \mathrm{Re}\,\psi'(z)<0,\ 1\leqq\lambda. \tag{21}$$

In order to prove this, we have

$$\sum_{n=1}^{\infty} nR(\lambda, n)\cos 2\pi nz=(\lambda q+q^2)\left[1+\sum_{n=1}^{\infty} q^{2n}\right]\cos 2\pi z + \sum_{n=2}^{\infty} n\frac{q^{2n}-(-1)^n\lambda q^n}{1-q^{2n}}\cos 2\pi nz.$$

Let (u, x) fixed and let $z=u+iv=u+i\beta y$. We will consider the expressions above as a function of $(\beta, \lambda, y)\in Z$, where

$$Z=[0, 1/10]\times[1/2, 2]\times[\sqrt{3}/2, \infty). \tag{22}$$

So, if (u, x) satisfies C_j for $j=1$ or $j=2$, then

$$\mathrm{Re}(\lambda q+q^2)\cos 2\pi z\leqq\alpha_{2j-1} \quad \text{and} \quad \mathrm{Im}(\lambda q+q^2)\sin 2\pi z\leqq\alpha_{2j},$$

where $\alpha_k: Z\mapsto\mathbf{R}$, $k=1, 2, 3, 4$ are the functions

$$\alpha_k(\beta, \lambda, y)=\frac{\lambda}{2}h\left[h^{-2\beta}\cos\frac{3\pi}{10}+(-1)^{k+1}h^{2\beta}\cos\pi(2u+x)\right] + \frac{h^2}{2}(h^{-2\beta}+h^{2\beta}), \qquad k=1, 2$$

and

$$\alpha_k(\beta, \lambda, y)=\frac{\lambda}{2}h^{1+2\beta}+\frac{h^2}{2}(h^{-2\beta}+h^{2\beta}), \qquad k=3, 4.$$

So,

$$\mathrm{Re}\sum_{n=1}^{\infty} nR(\lambda, n)\cos 2\pi nz\leqq\varphi_{2j-1}(\beta, \lambda, y)$$

and

$$\mathrm{Im}\sum_{n=1}^{\infty} R(\lambda, n)\sin 2\pi nz\leqq\varphi_{2j}(\beta, \lambda, y),$$

where $\varphi_{2j-1}: Z \mapsto \mathbf{R}$ and $\varphi_{2j}: Z \mapsto \mathbf{R}$, $j=1, 2$ are the functions

$$\varphi_{2j-1}(\beta, \lambda, y)=\alpha_{2j-1}+\tfrac{1}{2}(h^{2\beta}+h^{-2\beta})(\lambda h+h^2)\sum_{n=1}^{\infty} h^{2n}$$

$$+\frac{1}{2(1-h^4)}\sum_{n=2}^{\infty} n(\lambda h^n+h^{2n})(h^{-2n\beta}+h^{2n\beta}), \qquad j=1, 2,$$

and

$$\varphi_{2j}(\beta, \lambda, y)=\alpha_{2j}+\tfrac{1}{2}(h^{2\beta}+h^{-2\beta})(\lambda h+h^2)\sum_{n=1}^{\infty} h^{2n}$$

$$+\frac{1}{2(1-h^4)}\sum_{n=2}^{\infty} (\lambda h^n+h^{2n})(h^{-2n\beta}+h^{2n\beta}), \qquad j=1, 2.$$

As $\beta \geqq 0$, $\operatorname{Im} H_\lambda(z)=0$ and (u, x) satisfies C_j, $j=1$ or $j=2$, then by using the results above and (18), we find that

$$0 \leqq -\operatorname{Im} \cot \pi z \leqq 4\varphi_{2j}(\beta, \lambda, y)+2(1-\lambda)\beta.$$

This result and (20) imply

$$-\operatorname{Re} \psi'(z) \geqq \pi^2-\pi^2 \Theta_j^*(\beta, \lambda, y),$$

where $\Theta_j^*: Z \mapsto \mathbf{R}$ are the functions

$$\Theta_j^*=[4\varphi_{2j}+2(1-\lambda)\beta]^2+8\varphi_{2j-1}, \qquad j \in \{1, 2\}.$$

Now we observe that

$$\Theta_j^*(\beta, \lambda, y) \leqq \Theta_j(\beta, \lambda, y),$$

where $\Theta_2^*=\Theta_2$ and the function Θ_1 is obtained from Θ_1^* by substitution of the constant $\cos \pi(2u+x)$ by zero. That is,

$$\Theta_1=[4\varphi_2+2\lambda h^{1+2\beta} \cos \pi(2u+x)+2(1-\lambda)\beta]^2$$
$$+8\varphi_1-4\lambda h^{1+2\beta} \cos \pi(2u+x), \qquad \Theta_2^*=\Theta_2.$$

It is easy to prove that $\Theta_1^* \leqq \Theta_1$. We know that $-1 \leqq \cos \pi(2u+x) \leqq 0$ (recall that $j=1$, that is, the condition (C_1) is satisfied). Now if we replace $\cos \pi(2u+x)$ by a variable value in the expression of Θ_1^* then this new function is an increasing function of this variable. To see this it is sufficient to calculate the first derivative. This derivative is positive if and only if

$$4\varphi_2(\beta, \lambda, y)+2(1-\lambda)\beta<1.$$

Now we observe that $4\varphi_2(\beta, \lambda, y)+2(1-\lambda)\beta$ has a non-negative second derivative in the variables β and λ, respectively, and a negative derivative in the variable y. So this function attains a maximum value at a point $p_0 \in \{(\beta, \lambda, y) \in Z;\ \beta=0,\ 1/10,\ \lambda=1/2,\ 2,\ y=\sqrt{3}/2\}$ (see (22) for the definition of Z). Now

a long and direct calculation shows that $p_0=(0, 2, \sqrt{3}/2)$ and $4\varphi_2(\beta, \lambda, y)+2(1-\lambda)\beta<1/2$. This proves the desired inequality. By using this result we find

$$-\operatorname{Re}\psi'(z)\geqq\pi^2-\pi^2\Theta_j(\beta, \lambda, y).$$

Then, to prove (21) is sufficient, from the inequality above, to show

$$\Theta_j(\beta, \lambda, y)+\frac{2}{\pi y}(1-\lambda)<1, (\beta, \lambda, y)\in Z, \lambda\leqq 1 \quad \text{and} \quad j=1, 2, \tag{23}$$

and

$$\Theta_j(\beta, \lambda, y)<1, (\beta, \lambda, y)\in Z, 1\leqq\lambda \quad \text{and} \quad j=1, 2. \tag{24}$$

To prove this we observe that as $h^{-1}>15$ then

$$(h^{2\beta}+h^{-2\beta})(\lambda h+h^2)\sum_{n=1}^{\infty}h^{2n}\leqq\tfrac{8}{5}h^{\frac{14}{5}},$$

$$\begin{aligned}\sum_{n=2}^{\infty}(\lambda h^n+h^{2n})(h^{-2n\beta}+h^{2n\beta})&=\sum_{k=1}^{2}\lambda^{2-k}h^{2k}\left(\frac{h^{4\beta}}{1-h^{k+2\beta}}+\frac{h^{-4\beta}}{1-h^{k-2\beta}}\right)\\&\leqq\frac{(\lambda h^2+h^4)(h^{4\beta}+h^{-4\beta})}{1-h^{1-2\beta}}\\&\leqq\tfrac{13}{10}\lambda(1-h^4)h^{\frac{8}{5}}\end{aligned}$$

and

$$\begin{aligned}\sum_{n=2}^{\infty}n(\lambda h^n+h^{2n})(h^{-2n\beta}+h^{2n\beta})&=\sum_{k=1}^{2}\lambda^{2-k}h^{2k}\left[\frac{(2-h^{k-2\beta})h^{-4\beta}}{(1-h^{k-2\beta})^2}+\frac{(2-h^{k+2\beta})h^{4\beta}}{(1-h^{k+2\beta})^2}\right]\\&\leqq\frac{2(\lambda h^2+h^4)(h^{4\beta}+h^{-4\beta})}{(1-h^{1-2\beta})^2}\leqq 3\lambda(1-h^4)h^{\frac{8}{5}}.\end{aligned}$$

On the other hand, we have

$$\frac{\partial^2\Theta_j}{\partial\beta^2}>0, \frac{\partial^2\Theta_j}{\partial\lambda^2}\geqq 0, \frac{\partial\Theta_j}{\partial y}<0 \quad \text{and} \quad \Theta_j(\beta, \lambda, y)>0, \quad j=1, 2.$$

These inequalities for partial derivatives imply that $\Theta_j+2(1-\lambda)/(\pi y)$ (respectively Θ_j) has a maximum at the point $p_j\in Z_1$ (respectively $q_j\in Z_2$) where

$$Z_1=\{(\beta, \lambda, \sqrt{3}/2)\in Z; \beta=0, 1/10, \lambda=1/2, 1\}$$

and

$$Z_2=\{(\beta, \lambda, \sqrt{3}/2)\in Z; \beta=0, 1/10, \lambda=1, 2\}.$$

Now an easy calculation, by using directly the formulas for Θ_j in the case $\beta=0$ and by using the inequalities above for $\beta=1/10$, we can prove that

$$\Theta_j(\beta, \lambda, y)+\frac{2}{\pi y}(1-\lambda)<\frac{7}{10}, (\beta, \lambda, y)\in Z_1, \quad j=1, 2,$$

and

$$\Theta_j(\beta, \lambda, y)<1, (\beta, \lambda, y)\in Z_2 \quad j=1, 2.$$

This proves (23) and (24) and then (21). This finishes the proof of (b) of Lemma 2.

Proof of (c) of Lemma 2. We will use the following expression for the function $H_\lambda(z)$ (see [11, vol. 2, pp. 16 and 17]):

$$H_\lambda(z)=\frac{v_0'(z)}{v_0(z)}-\lambda\frac{v_3'(z)}{v_3(z)}+2\pi i(1-\lambda)\beta,$$

where $2q^{\frac{1}{4}}v_0(z)=v_1(z)$ and $v_1(z)$, $v_3(z)$ are the Jacobi functions

$$v_1(z)=2q^{\frac{1}{4}}\left[\sin\pi z+\sum_{n=1}^{\infty}(-1)^n q^{n^2+n}\sin(2n+1)\pi z\right],$$

$$v_3(z)=1+2\sum_{n=1}^{\infty}q^{n^2}\cos 2\pi nz.$$

We observe that $v_0(z)$ and $v_3(z)$ are analytic function with simple zeros. The set of zeros of $v_0(z)$ and $v_3(z)$ are $L(\tau)$ and $\{w_2\}+L(\tau)$, respectively. Then to prove (c) it is sufficient to show that

$$V(z,\lambda)\neq 0, \text{ if } 1/2\leqq\lambda\leqq 2, z\notin U(\tau), 0\leqq u\leqq 1 \quad \text{and} \quad 0\leqq\beta\leqq 1/4, \tag{25}$$

where $z=u+iv=u+i\beta y$ and

$$V(z,\lambda)=v_0'(z)\,v_3(z)-\lambda v_0(z)\,v_3'(z)+2\pi i\beta(1-\lambda)\,v_0(z)\,v_3(z).$$

We define the functions $A_j(z)$, $j=0$, 3 and $\varepsilon_k(z)$, $k=1, 2, 3$, such that

$$A_0(z)=\sin\pi z-q^2\sin 3\pi z, \qquad A_3(z)=1+2q\cos 2\pi z$$

and

$$v_0'\,v_3=A_0'\,A_3+\varepsilon_1,\ v_0\,v_3'=A_0\,A_3'+\varepsilon_2 \quad \text{and} \quad v_0\,v_3=A_0\,A_3+\varepsilon_3.$$

We have that

$$\begin{aligned}|A_0'|&\leqq\frac{\pi}{2}\left[h^\beta+h^{-\beta}+3h^2(h^{-3\beta}+h^{3\beta})\right]\\&\leqq\frac{\pi}{2}(h^{\frac{1}{4}}+h^{-\frac{1}{4}}+3h^{\frac{5}{4}}+3h^{\frac{11}{4}}),\end{aligned}$$

$$\begin{aligned}|v_3-A_3|&\leqq\sum_{n=2}^{\infty}h^{n^2}(h^{2n\beta}+h^{-2n\beta})\\&\leqq\sum_{n=2}^{\infty}h^{2n}(h^{2n\beta}+h^{-2n\beta})\\&\leqq(1-h^{\frac{3}{2}})^{-1}(h^3+h^5),\\|A_3|&\leqq 1+h(h^{2\beta}+h^{-2\beta})\leqq 1+h^{\frac{3}{2}}+h^{\frac{1}{2}}\end{aligned}$$

and

$$|v_0' - A_0'| \leqq \frac{\pi}{2} \sum_{n=2}^{\infty} (2n+1)\, h^{n^2+n}(h^{(2n+1)\beta} + h^{-(2n+1)\beta})$$

$$\leqq \frac{\pi h}{2} \sum_{n=2}^{\infty} (2n+1)\, h^{2n+1}(h^{(2n+1)\beta} + h^{-(2n+1)\beta})$$

$$\leqq \frac{5\pi h^6 (h^{-5\beta} + h^{5\beta})}{2(1-h^{2-2\beta})^2}.$$

Then, as $y \geqq \sqrt{3}/2$, $h^{-1} > 15$ and $0 \leqq \beta \leqq 1/4$, an easy computation shows that

$$|\varepsilon_1(z)| \leqq \pi h^{\frac{11}{4}}.$$

By the same way we can prove that

$$\max\{3|\varepsilon_1(z)|, |\varepsilon_2(z)|, 3\pi|\varepsilon_3(z)|\} \leqq 3\pi h^{\frac{11}{4}}. \tag{26}$$

On the other hand, by using the notation (3) we have

$$\frac{2}{\pi} h^{\beta} \operatorname{Im} V(z, \lambda) = \Omega_1(u, \beta, \lambda) + R_1(u, \beta, \lambda), \tag{27}$$

where $\Omega_1 = \Omega_1(u, \beta, \lambda)$ and $R_1 = R_1(u, \beta, \lambda)$ are the functions

$$\begin{aligned}\Omega_1 = {} & \{[-1+2(1-\lambda)\beta] + [1+2(1-\lambda)\beta]\, h^{2\beta}\} \sin \pi u \\ & + h\{[-1-2\lambda-2(1-\lambda)\beta] \sin \pi(u-x) \\ & + [-1+2\lambda+2(1-\lambda)\beta]\, h^{-2\beta} \sin \pi(3u-x) \\ & + [1+2\lambda-2(1-\lambda)\beta]\, h^{2\beta} \sin \pi(u+x) \\ & + [1-2\lambda+2(1-\lambda)\beta]\, h^{4\beta} \sin \pi(3u+x)\}\end{aligned} \tag{28}$$

and

$$\begin{aligned}R_1 = {} & h^2\{[3-2(1-\lambda)\beta]\, h^{-2\beta} \sin \pi(3u-2x) \\ & - [3+2(1-\lambda)\beta]\, h^{4\beta} \sin \pi(3u+2x)\} \\ & + h^3\{[3+2\lambda-2(1-\lambda)\beta] \sin \pi(u-3x) \\ & - [3+2\lambda+2(1-\lambda)\beta]\, h^{2\beta} \sin \pi(u+3x) \\ & + [3-2\lambda-2(1-\lambda)\beta]\, h^{-4\beta} \sin \pi(5u-3x) \\ & + [-3+2\lambda-2(1-\lambda)\beta]\, h^{6\beta} \sin \pi(5u+3x)\} \\ & + h^{\beta}\left[\frac{2}{\pi} \operatorname{Im}(\varepsilon_1 - \lambda\varepsilon_2) + 4(1-\lambda)\beta \operatorname{Re} \varepsilon_3\right].\end{aligned} \tag{29}$$

Also,

$$\frac{2}{\pi} h^{\beta} \operatorname{Re} V(z, \lambda) = \Omega_2(u, \beta, \lambda) + R_2(u, \beta, \lambda), \tag{30}$$

where $\Omega_2=\Omega_2(u,\beta,\lambda)$ and $R_2=R_2(u,\beta,\lambda)$ are the functions

$$\begin{aligned}\Omega_2=&\{[1+2(1-\lambda)\beta]\,h^{2\beta}+[1-2(1-\lambda)\beta]\}\cos\pi u\\&+h\{[1-2\lambda-2(1-\lambda)\beta]\,h^{-2\beta}\cos\pi(3u-x)\\&+[1+2\lambda+2(1-\lambda)\beta]\cos\pi(u-x)\\&+[1+2\lambda-2(1-\lambda)\beta]\,h^{2\beta}\cos\pi(u+x)\\&+[1-2\lambda+2(1-\lambda)\beta]\,h^{4\beta}\cos\pi(3u+x)\}\end{aligned}\tag{31}$$

and

$$\begin{aligned}R_2=&h^2\{[-3+2(1-\lambda)\beta]\,h^{-2\beta}\cos\pi(3u-2x)\\&-[3+2(1-\lambda)\beta]\,h^{4\beta}\cos\pi(3u+2x)\}\\&+h^3\{-[3+2\lambda-2(1-\lambda)\beta]\sin\pi(u-3x)\\&-[3+2\lambda+2(1-\lambda)\beta]\,h^{2\beta}\sin\pi(u+3x)\\&+[-3+2\lambda+2(1-\lambda)\beta]\,h^{-4\beta}\sin\pi(5u-3x)\\&+[-3+2\lambda-2(1-\lambda)\beta]\,h^{6\beta}\cos\pi(5u+3x)\}\\&+h^{\beta}\left[\frac{2}{\pi}\operatorname{Re}(\varepsilon_1-\lambda\varepsilon_2)-4(1-\lambda)\beta\operatorname{Im}\varepsilon_3\right].\end{aligned}\tag{32}$$

Now, we define the function $S(\beta)$,

$$\begin{aligned}S(\beta)=&h^2[(3+2\beta)\,h^{-2\beta}+(3+\beta)\,h^{4\beta}]\\&+h^3[7+2\beta+(7-2\beta)\,h^{2\beta}+(2-\beta)\,h^{-4\beta}+(2+\beta)\,h^{6\beta}]\\&+15h^{\frac{11}{4}+\beta}.\end{aligned}$$

Then from (26), (29) and (32) we find that

$$|R_j(u,\beta,\lambda)|\leqq S(\beta),\ j=1,2\quad\text{and}\quad\frac{\partial^2 S(\beta)}{\partial\beta^2}>0,\ \frac{1}{2}\leqq\lambda\leqq 2,\ 0\leqq\beta\leqq\frac{1}{4}.\tag{33}$$

In order to prove (25) and then (c) of this lemma we define the compact sets

$$\begin{aligned}&I_1=[0,1/4]\times[0,1/2],&&I_2=[1/4,2/5]\times[1/5,1/2],\\&I_3=[1/4,2/5]\times[0,1/5],&&I_4=[3/4,1]\times[0,1/2],\\&I_5=[2/5,3/5]\times[0,1/2],\quad\text{and}&&I_6=[3/5,3/4]\times[0,1/2].\end{aligned}$$

We observe that

$$\cup I_j=[0,1]\times[0,1/2]=\{(u,x),0\leqq u\leqq 1,0\leqq x\leqq 1/2\}$$

and if $(u,x)\in I_5$ then $z=u+i\beta y$ satisfies, respectively, the condition (C_1) if $2/5\leqq u\leqq 1/2$ and (C_2) if $1/2\leqq u\leqq 3/5$. Also if $(u,x)\in I_6$ then $z=u+i\beta y$ satisfy the condition (C_2).

Also, we define the functions $a_{kj}(\beta)$, $k=1, 2$ and $j=1, \ldots, 4$, and $b_{kj}(\beta)$, $k=1, 2$ and $j=1, \ldots, 6$, by the expressions

$$a_{kj}(\beta)=c_{k0}(\beta)\, h^{2\beta}\cos\pi A_{1j}+c_{k5}(\beta)\, h^{1+4\beta}\cos\pi A_{5j}, \quad j=1, \ldots, 4, \tag{34}$$

$$b_{kj}(\beta)=c_{k1}(\beta)\cos\pi A_{1j}+\sum_{p=2}^{4} c_{kp}(\beta)\, h^{1+2(p-3)\beta}\cos\pi A_{p,j}, \quad j=1, \ldots 4 \tag{35}$$

and

$$\begin{aligned} b_{kj}(\beta)=&-c_{k0}(\beta)\, h^{2\beta}\sin\pi A_{1j}+c_{k1}(\beta)\sin\pi A_{1j} \\ &+\sum_{p=2}^{5} c_{kp}(\beta)\, h^{1+2(p-3)\beta}\sin\pi A_{pj}, \quad j=5, 6, \end{aligned} \tag{36}$$

where $c=c(\beta)=(c_{kp}(\beta))$ is the 2×6 matrix, $k=1, 2$ and $p=0, \ldots, 5$,

$$c=\begin{bmatrix} 1+\beta & 1-\beta & \beta & 2+\beta & 2-\beta & \beta \\ 1-2\beta & 1+2\beta & 3-2\beta & 5-2\beta & 5+2\beta & 3+2\beta \end{bmatrix} \tag{37}$$

and $A=(A_{pj})$ is the 5×6 matrix, $p=1, \ldots, 5$, $j=1, \ldots, 6$,

$$A=\begin{bmatrix} 1/4 & 2/5 & 2/5 & 1/4 & 2/5 & 1/4 \\ 1 & 3/4 & 1/2 & 1 & -3/10 & -1/4 \\ 1/2 & 1/4 & 2/5 & 3/4 & -1/10 & 1/10 \\ 3/4 & 9/10 & 3/5 & 1/2 & 3/2 & -2/5 \\ 1 & 1 & 1 & 1 & 3/2 & 3/2 \end{bmatrix}. \tag{38}$$

We observe that

$$\begin{aligned} &a'_{kj}(\beta)<0, \ j=1, \ldots, 4, \ k=1, 2 \quad \text{and} \quad b''_{kj}(\beta)<0, \\ &\qquad j=1, \ldots, 6, k=1, 2. \end{aligned} \tag{39}$$

We will finish the proof of (c) of Lemma 2 by considering, for $1/2\leqq\lambda\leqq 2$, three cases. Case 1: $(u, x)\in I_1\cup I_2\cup I_3$ and $0\leqq\beta\leqq 1/4$, case 2: $(u, x)\in I_4$ and $0\leqq\beta\leqq 1/4$ and case 3: $(u, x)\in I_5\cup I_6$ and $1/10\leqq\beta\leqq 1/4$.

Proof of case 1. If $(u, x)\in I_1$ then $\cos\pi u\geqq\cos\dfrac{\pi}{4}$, $\cos\pi(u-x)\geqq\cos\dfrac{\pi}{2}$ and $\cos\pi(u+x)\geqq\cos\dfrac{3\pi}{4}$. Thus,

$$\Omega_2(u, \beta, \lambda_k)+R_2(u, \beta, \lambda_k)\geqq a_{k1}(\beta)+b_{k1}(\beta)-S(\beta), \quad k=1, 2, \tag{40}$$

where $\lambda_1=1/2$ and $\lambda_2=2$. We have

$$a_{11}(\beta)=(1+\beta)\, h^{2\beta}\cos\frac{\pi}{4}+\beta h^{1+4\beta}\cos\pi$$

and

$$\begin{aligned} b_{11}(\beta)=&(1-\beta)\cos\frac{\pi}{4}+\beta h^{1-2\beta}\cos\pi+(2+\beta)h\cos\frac{\pi}{2} \\ &+(2-\beta)\, h^{1+2\beta}\cos\frac{3\pi}{4}. \end{aligned}$$

So, by using that $h^{-1}=e^{\pi y}>15$ and the results above we find that

$$a_{11}(\beta)\geqq a_{11}(\tfrac{1}{4})>\tfrac{1}{5}$$

and

$$b_{11}(\beta)-S(\beta)\geqq\min\{b_{11}(0)-S(0),\, b_{11}(\tfrac{1}{4})-S(\tfrac{1}{4})\}>\tfrac{3}{10}.$$

Thus,

$$\Omega_2(u,\beta,\tfrac{1}{2})+R_2(u,\beta,\tfrac{1}{2})>\tfrac{1}{2},\,(u,x)\in I_1,\,0\leqq\beta\leqq\tfrac{1}{4}.$$

By the same way, we find that

$$\Omega_2(u,\beta,2)+R_2(u,\beta,2)>\tfrac{1}{5},\,(u,x)\in I_1,\,(u,x)\in I_1,\,0\leqq\beta\leqq\tfrac{1}{4}.$$

So, as Ω_2+R_2 is a linear function of λ, then the inequalities above and (40) imply that

$$\Omega_2(u,\beta,\lambda)+R_2(u,\beta,\lambda)>\tfrac{1}{5},\,(u,x)\in I_1,\,0\leqq\beta\leqq\tfrac{1}{4}. \tag{41}$$

If $(u,\,x)\in I_2$ then we have $\cos\pi u\geqq\cos\dfrac{2\pi}{5}$, $\cos\pi(u-x)\geqq\cos\dfrac{\pi}{4}$, $\cos\pi(u+x)\geqq\cos\dfrac{9\pi}{10}$ and $-\cos\pi(3u-x)\geqq\cos\dfrac{3\pi}{4}$.

If $(u,\,x)\in I_3$ then we have $\cos\pi u\geqq\cos\dfrac{2\pi}{5}$, $\cos\pi(u-x)\geqq\cos\dfrac{2\pi}{5}$, $\cos\pi(u+x)>\cos\dfrac{3\pi}{5}$ and $-\cos\pi(3u-x)\geqq\cos\dfrac{\pi}{2}$.

Thus, for $(u,\,x)\in I_2\cup I_3$ and $0\leqq\beta\leqq 1/4$ we conclude, by using the results above and (31), ..., (38), that

$$\Omega_2(u,\beta,\lambda_k)+R_2(u,\beta,\lambda_k)\geqq a_{kj}(\beta)+b_{kj}(\beta)-S(\beta),\qquad k=1,2,j=2,3, \tag{42}$$

where $\lambda_1=1/2$ and $\lambda_2=2$. Also from (33) and (39) we have that $a_{kj}(\beta)$ attains a minimum value at the point $\beta=1/4$ and $b_{kj}(\beta)-S(\beta)$ attains a minumum value at the point $\beta=0$ or $\beta=1/4$. So, as $y\geqq\sqrt{3}/2$ and $h^{-1}=e^{\pi y}>15$ then an easy, but long calculation show that $a_{kj}(\beta)>0$ and $b_{kj}(\beta)-S(\beta)>0$, $k=1,\,2$ and $j=2,\,3$. So,

$$\Omega_2(u,\beta,\lambda_k)+R_2(u,\beta,\lambda_k)>0,\,(u,x)\in I_2\cup I_3,\,0\leqq\beta\leqq 1/4.$$

This result together with (30), (41) and (42) and the fact that $\operatorname{Re}V(z,\,\lambda)$ is a linear function of λ, proves, in case 1, that $\operatorname{Re}V(z,\,\lambda)>0$ if $1/2\leqq\lambda\leqq 2$ and $0\leqq\beta\leqq 1/4$. This proves (25) in the first case.

Proof of case 2. If $(u,\,x)\in I_4$ then $-\cos\pi u\geqq\cos\dfrac{\pi}{4}$, $-\cos\pi(u-x)\geqq\cos\dfrac{3\pi}{4}$ and $-\cos\pi(u+x)\geqq\cos\dfrac{\pi}{2}$. So, we have from (31), ..., (38), that

$$-\Omega_2(u,\beta,\lambda_k)-R_2(u,\beta,\lambda_k)\geqq a_{k4}(\beta)+b_{k4}(\beta)-S(\beta),\qquad k=1,2, \tag{43}$$

where $\lambda_1 = 1/2$ and $\lambda_2 = 2$. Also from (33) and (39) we have that $a_{k4}(\beta)$ attains a minimum value at the point $\beta = 1/4$ and $b_{k4}(\beta) - S(\beta)$ attains a minimum value at the point $\beta = 0$ or $\beta = 1/4$. So, as $y \geqq \sqrt{3}/2$ and $h^{-1} = e^{\pi y} > 15$ then an easy, but long calculation shows that $a_{k4}(\beta) > 0$ and $b_{k4}(\beta) - S(\beta) > 0$, $k = 1, 2$. This result and (43) shows

$$-\Omega_2(u, \beta, \lambda_k) - R_2(u, \beta, \lambda_k) > 0, (u, x) \in I_4, \quad k = 1, 2.$$

The inequality above together with (30) and the fact that $\operatorname{Re} V(z, \lambda)$ is a linear function of λ, proves that $\operatorname{Re} V(z, \lambda) < 0$, if $1/2 \leqq \lambda \leqq 2$, $0 \leqq \beta \leqq \frac{1}{4}$ and $(u, x) \in I_4$. This proves (25) in the second case.

Proof of case 3. If $(u, x) \in I_5$ then $\sin \pi u \geqq \sin \frac{2\pi}{5}$, $-\sin \pi(u - x) \leqq \sin \frac{\pi}{10}$, and $\sin \pi(3u - x) \geqq \sin \frac{-3\pi}{10}$. Also, if $(u, x) \in I_6$ then $\sin \pi u \geqq \sin \frac{\pi}{4}$, $-\sin \pi(3u - x) \geqq \sin \frac{-\pi}{4}$, $\sin \pi(u - x) \geqq \sin \frac{\pi}{10}$ and $-\sin \pi(u + x) \geqq \sin \frac{-2\pi}{5}$. So, for $(u, x) \in I_5 \cup I_6$ we have from (28), (29), (33), (36), (37) and (38), that

$$-\Omega_1(u, \beta, \lambda_k) - R_1(u, \beta, \lambda_k) \geqq b_{kj}(\beta) - S(\beta), \quad k = 1, 2, \tag{44}$$

where $\lambda_1 = 1/2$ and $\lambda_2 = 2$. Also from (33) and (39) we have that $b_{kj}(\beta) - S(\beta)$ attains a minimum value at the point $\beta = 1/10$ or $\beta = 1/4$. So, as $y \geqq \sqrt{3}/2$ and $h^{-1} = e^{\pi y} > 15$ then an easy, but long calculation shows that $b_{kj}(\beta) - S(\beta) > 0$, $k = 1, 2$. This result and (44) shows that

$$-\Omega_1(u, \beta, \lambda_k) - R_1(u, \beta, \lambda_k) > 0, \quad k = 1, 2, (u, x) \in I_5 \cup I_6.$$

The inequality above together with (27) and the fact that $\operatorname{Im} V(z, \lambda)$ is a linear function of λ, proves that $\operatorname{Im} V(z, \lambda) < 0$, if $1/2 \leqq \lambda \leqq 2$, $\frac{1}{10} \leqq \beta \leqq \frac{1}{4}$ and $(u, x) \in I_5 \cup I_6$. This proves (25) in the third case. This finishes the proof of Lemma 2.

Proof of Corollary 2.1. Let $L(\tau)$ and $L(\tau^*)$ be lattices of $\mathbf{C}$ where $\tau = x + iy \in FM$ and $\tau^* = -x + iy$. Also let $H = H_\lambda$ and $H^* = H_\lambda^*$ be, respectively, the $L(\tau)$ and $L(\tau^*)$-periodic functions as defined in Lemma 1. We observe that $H(z) = 0$, $z \in \mathbf{C}$ if and only if $H^*(\bar{z}) = 0$ (see [11, vol. 3, p. 118 and to use the definition of H_λ]). So, by using these results is sufficient to prove the lemma for $x \geqq 0$. In order to prove this, let H_λ be the $L(\tau)$-periodic functions as defined in Lemma 1, where $\tau = x + ix \in FM$, $x \geqq 0$ and $1/2 \leqq \lambda \leqq 2$. It is clear, by using the additive property of ζ and Legendre's relations that $H_\lambda(mw_1 + nw_3) = 0$ for every $m, n \in \mathbf{Z}$. Also, as H_λ is $L(\tau)$-periodic, we consider $H_\lambda : \mathbf{C}/L(\tau) \to \mathbf{C} \cup \infty$ the quotient function. Observe that we denote with the same letter the function H_λ and its quotient function. Then H_λ is a function between compact surfaces and by (a) of Lemma 2 the degree of H_λ is 2. Also, $H_\lambda(w_1) = 0$ and w_1 satisfies the condition (C_1). So, $JH_\lambda(w_1) > 0$. On the other hand

$$H_\lambda(z + w_2) = -\lambda\left[E(z) - \frac{1}{\lambda} E(z + w_2)\right] = -\lambda H_\mu(z), \quad \mu = \lambda^{-1}. \tag{45}$$

So, as $w_1 + w_2 \equiv w_3$ and $1/2 \leqq \mu \leqq 2$ we conclude from (45) and Lemma 2 that $H_\lambda(w_3) = 0$ and $JH_\lambda(w_3) > 0$. That is, we have that $H_\lambda(w_j) = 0$, $JH_\lambda(w_j) > 0$,

$j=1, 3$, $H(z)$ is an odd function and, from (a) of Lemma 2, the degree of H is 2. So, to finish the proof it is sufficient to show that 0 is a regular value of H_λ and $JH_\lambda(z)>0$, for every $z\in\mathbf{C}$ such that $H(z)=0$. To prove this, let $z\in\mathbf{C}$ such that $H(z)=0$. As H_λ is periodic and an odd function there exist u, α, $\beta\in\mathbf{R}$, $0\leqq u\leqq 1$ and $0\leqq\beta\leqq 1/2$ such that $z\equiv\alpha+\beta\tau=u+i\beta y$ or $-z\equiv u+i\beta y$. We consider two cases: $0\leqq\beta\leqq 1/4$ and $1/4\leqq\beta\leqq 1/2$. If $0\leqq\beta\leqq 1/4$ we have from (b) and (c) of Lemma 2 that $JH_\lambda(z)>0$. On the other hand, if $1/4\leqq\beta\leqq 1/2$ we consider $z'=z+w_2$. Then there exist $0\leqq u'\leqq 1$ and $0\leqq\beta'\leqq 1/4$ such that $z'\equiv u'+i\beta' y$ or $-z'\equiv u'+i\beta' y$. By using (45), the fact that H_λ is an odd function and that $H_\lambda(z)=0$, we conclude that $H_\mu(z')=0$. Then, from (b) and (c) of Lemma 2 we conclude that $J_\mu H(z')>0$. Thus, by using again (45) we find that $JH_\lambda(z)>0$. This finishes the proof of Corollary 2.1.

Proof of Corollary 2.2. Let $M=\mathbf{C}/L(\tau)-\{q_1, q_2, q_3\}$ be a Riemann surface conformally equivalent to a compact Riemann surface $\mathbf{C}/L(\tau)$ punctured at three points, where $\tau=x+iy\in FM$ and let $X: M\mapsto\mathbf{R}^3$ be a complete minimal embedding in $\mathbf{R}^3$ with finite total curvature. We suppose first that X has a flat end. In this case, from [2, Theorem 1], we have that $x=0$, $y=1$ and X is the surface S_1 that appears in [1], [2] and [3]. On the other hand, if the ends of X are all of catenoid type then after a rigid motion and a homothety of X in $\mathbf{R}^3$ and a translation of coordinates of $\mathbf{C}$, there exist $z_1\in\mathbf{C}$ and b_1, $b_2\in\mathbf{R}$ such that $b_2>-b_1>0$, $q_1=\pi(z_1)$, $q_2=\pi(z_1+w_2)$, $q_3=\pi(0)$ and the Eqs. (b), (c) and (d) of Lemma 1 are satisfied. Also, from (d) of Lemma 1 and Corollary 2.1 we conclude that $z_1\equiv w_1$ or $z_1\equiv w_3$. So, these results and (b) of Lemma 1 show that

$$\frac{\pi b_1^2}{e_k-e_2}+\frac{\pi b_2^2}{e_t-e_2}=y\overline{(2\pi_1+e_2)}-\pi, \tag{46}$$

where (k, t) is a permutation of $(1, 3)$ and $b_j\in\mathbf{R}-\{0\}$, $j=1, 2$.

On the other hand, the Propositions 3 and 4 that appear in [2] show that: (1) If $x>0$ then the imaginary part of $2\eta_1+e_2$ is positive and the imaginary part of e_1-e_2 and e_3-e_2 are negative, respectively; (2) if $x<0$ then the imaginary part of $2\eta_1+e_2$ is negative and the imaginary part of e_1-e_2 and e_3-e_2 are positive, respectively. This result carried in Eq. (46) shows that $x=0$. This finishes the proof of Corollary 2.2.

Lemma 3 *Let $L(iy)=\{m+niy\in\mathbf{C};\ m, n\in\mathbf{Z}\}$ be a lattice, where $y\geqq 1$ and let $b_1(y)=b_1$, $b_2(y)=b_2$ be a solution of the equations*

$$\frac{\pi b_1^2}{e_1-e_2}+\frac{\pi b_2^2}{e_3-e_2}=y(2\eta_1+e_2)-\pi \tag{47}$$

and

$$y|2\eta_1+e_2|^2-2\pi(2\eta_1+e_2)=2\pi b_1 b_2 \tag{48}$$

with the condition $b_2\geqq -b_1>0$. Then,

(a) *The funcions $b_1, b_2 : [1, \infty)\to\mathbf{R}$ are differentiable,*

(b) $b_2=-b_2=\sqrt{\dfrac{\pi}{2}}$ *if* $y=1$,

(c) $\dfrac{-b_1}{b_2}<1$, *if* $y>1$, *and*

(d) $\dfrac{-b_1}{b_2} > \dfrac{1}{2}$, *if* $y > 1$.

In order to prove Lemma 3 we need the following proposition:

Proposition 1 *Let* $L(i\,y)$ *as in Lemma* 3. *Then*

(a) $\dfrac{y(2\eta_i + e_2) - \pi}{e_2} \geqq \dfrac{1}{2(1-3h)(1+h)^2}$,

(b) $\dfrac{e_2 - e_3}{2\eta_1 + e_2} \geqq 2(1-h^2)(1+h^2)^4$ *and*

(c) $e_1 - e_2 \geqq \pi^2(1-2h)^4$.

Proof. We define the functions $f_1, f_2 : [1, \infty] \mapsto \mathbf{R}$ by $f_1(y) = e_2$ and $f_2(y) = y(2\eta_1 + e_2) - \pi$. Then from [11, vol. 3, pp. 139 and vol. 2, p. 28], we have

$$2\eta_1 + e_2 = -8\pi^2 \sum_{n=1}^{\infty} (-1)^n \frac{n h^n}{1-h^{2n}} \leqq \frac{8\pi^2 h}{1-h^2}, \tag{49}$$

$$e_2 - e_3 = 16\pi^2 h \left[1 + \sum_{n=1}^{\infty} h^{n(n+1)}\right]^4 \geqq 16\pi^2 h(1+h^2)^4, \tag{50}$$

$$e_1 - e_2 = \pi^2 \left[1 + 2\sum_{n=1}^{\infty} (-1)^n h^{n^2}\right]^4 \geqq \pi^2(1-2h)^4 \tag{51}$$

and

$$e_2 = -\frac{\pi^2}{3} - 8\pi^2 \sum_{n=1}^{\infty} (-1)^n \frac{n h^n}{1+(-1)^n h^n}. \tag{52}$$

Also, from (52) and (49) we have

$$0 > f_1'(y) \geqq 8\pi^3 (-1)^n \frac{n^2 h^n}{(1+(-1)^n h^n)^2} \geqq \frac{-8\pi^3 h}{(1-h)^2}(1-3h) \tag{53}$$

and

$$f_2'(y) \leqq 8\pi^3 y \left\{ \sum_{n=1}^{2} (-1)^n \frac{n^2(h^n + h^{3n})}{(1-h^{2n})^2} - \frac{1}{\pi y} \sum_{n=1}^{3} (-1)^n \frac{n h^n}{1-h^{2n}} \right\}$$

$$\leqq -4\pi^3 y \frac{h}{(1-h^2)^2}. \tag{54}$$

Then,

$$\frac{f_2'(y)}{f_1'(y)} \geqq \frac{1}{2(1-3h)(1+h)^2}. \tag{55}$$

On the other hand, we have that $f_1(1) = f_2(1) = 0$ and from (53) we have $f_1'(y) < 0$. Then the functions f_1 and f_2 satisfy the hypotheses of Proposition 7 of [2]. So, from (55) we find

$$\frac{f_2(y)}{f_1(y)} = \frac{y(2\eta_1 + e_2) - \pi}{e_2} \geqq \frac{1}{2(1-3h)(1+h)^2}.$$

The inequality above, together with (49), (50) and (51) prove Proposition 1.

Proof of Lemma 3. From (50), (51), (54) and from Proposition 1 we conclude that $y(2\eta_1+e_2)-\pi\leqq 0$, $e_1-e_2>0$ and $e_3-e_2<0$ for every $y\geqq 1$ and $y(2\eta_1+e_2)-\pi=0$ if and only if $y=1$. So, the Eqs. (47) and (48) are, for each $y>1$, a pair of transversal hyperbolas and when $y=1$ the first hyperbola degenerates to a pair of straight lines. From (48) we conclude that if $b_1=b_1(y)$ and $b_2=b_2(y)$ is a solution of this equation then $b_1 b_2<0$. Let b_1 and b_2 be the only solution of (47) and (48) such that $b_1<0$ and $b_2>0$ (See Fig. 2, below). The proof of (a) and (b) follows immediately from the observations above.

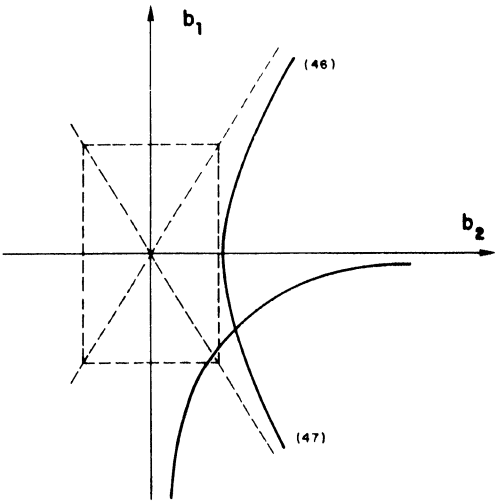


Fig. 2

In order to prove (c) of Lemma 3, we reason by contradiction. We suppose that $-b_1\geqq b_2>0$. From (47) we have

$$b_1^2=b_2^2\left(\frac{B}{A}-\frac{B}{\pi b_2^2}\right),$$

where $B=(y(2\eta_1+e_2)-\pi)(e_2-e_1)$ and $A=(y(2\eta_1+e_2)-\pi)(e_3-e_2)$. We observe that $B>A>0$. Then, as $-b_1\geqq b_2>0$, from (48) and the inequality above we find that

$$2\pi b_2^2\leqq(2\eta_1+e_2)[2\pi-y(\eta_1+e_2)] \quad \text{and} \quad b_2^2\geqq\frac{AB}{\pi(B-A)}.$$

These equations imply that

$$\frac{2[y(2\eta_1+e_2)-\pi](e_2-e_3)(e_1-e_2)}{3e_2(2\eta_1+e_2)}\leqq 2\pi-y(2\eta_1+e_2).$$

By using this inequality, Proposition 1 and (49) we find

$$\frac{\pi}{3}\frac{(1-h^2)(1-2h)^4}{(1-3h)(1+h)^2} \leqq 1-4\pi y\left(\frac{h}{1-h^2}-\frac{2h^2}{1-h^4}\right) \leqq 1-12h\frac{(1-h)^2}{1-h^4}.$$

As, $y>1$ and $h^{-1}=e^{\pi y}>23$, then an easy computation shows that the inequality above in a contradiction. This proves (c) of Lemma 3.

In order to prove (d) of Lemma 3, let $b_1=b_1(y)$, $b_2=b_2(y)$ be the solution of Eqs. (47) and (48) in the lattice $L(iy)$ with $b_1<0$ and $b_2>0$. We define the function $a:(1,\infty)\to(0,\infty)$ by

$$a(y)=\frac{-b_1(y)}{b_2(y)}.$$

We know from (c) of this lemma that $a(y)<1$. Now we will show that

$$a(y)>\tfrac{1}{2}. \tag{56}$$

In order to prove this we find from (47) and (48) that

$$\frac{a^2(y)}{e_1-e_2}+\frac{1}{e_3-e_2}+2a(y)[1-\alpha(y)]\frac{1}{2\eta_1+e_2}=0,$$

where

$$\alpha(y)=\frac{\pi}{2\pi-y(2\eta_1+e_2)}.$$

So, as $a(y)>0$ we conclude that

$$a(y)=\frac{e_1-e_2}{2\eta_1+e_2}\left\{-1+\alpha(y)+\sqrt{[1-\alpha(y)]^2+\frac{(2\eta_1+e_2)^2}{(e_1-e_2)(e_2-e_3)}}\right\}.$$

Then to prove (56) it is sufficient to show that

$$\frac{2\eta_1+e_2}{e_2-e_3}-\frac{2\eta_1+e_2}{4(e_2-e_2)}+\frac{\pi}{2\pi-y(2\eta_1+e_2)}>1. \tag{57}$$

To prove (57) it is sufficient, by using (49), (50) and (51) to show that

$$\sum_{j=1}^{3}\varphi_j(y)>1, \tag{58}$$

where $\varphi_j:(1,\infty)\to(0,\infty)$ are the functions:

$$\varphi_1(y)=\frac{(1-h)^2}{2(1+2h^2)^4}, \qquad \varphi_2(y)=\frac{-2h(1-2h+5h^2)}{(1-2h)^4}$$

and

$$\varphi_3(y)=\frac{1}{2-22yh}.$$

We remark that to define the function φ_2 and φ_3 we use that $h^{-1}=e^{\pi y}>23$ and that, respectively,

$$\frac{1-2h+4h^2}{1-h^4}\leqq 1-2h+5h^2 \quad \text{and} \quad 8\pi\left(\frac{1}{1-h^2}-\frac{2h}{1-h^4}\right)>22.$$

We have

$$\frac{\varphi_1'(y)+\varphi_2'(y)}{\pi h}\leqq 1+8h+\frac{2(1+3h)}{(1-2h)^5}<5 \tag{59}$$

and

$$\frac{\varphi_3'(y)}{\pi h}=\frac{22(\pi^{-1}-y)}{(2-22yh)^2}<-5. \tag{60}$$

It easy to prove the last inequality. It is sufficient to show that it is true for $1\leqq y\leqq\frac{5}{4}$ where $yh\geqq\frac{1}{50}$ and for $y\geqq\frac{5}{4}$. On the other hand,

$$\sum_{j=1}^{3}\varphi_j(1)>1 \quad \text{and} \quad \lim_{y\to\infty}\sum_{j=1}^{3}\varphi_j(y)=1.$$

This result together with (59) and (60) proves (58). This finishes the proof of (d) and completes the proof of Lemma 3.

Proof of Theorem A. The proof of (a) of Theorem A follows immediately from (b) of Corollary 2.2. In order to prove (b), let $X\colon M\mapsto\mathbf{R}^3$ be a complete minimal surface embedded in $\mathbf{R}^3$ of genus one and three ends of catenoid type, where M is conformally equivalent to $\mathbf{C}/L(1, iy)-\{\pi(0), \pi(1/2), \pi(iy/2)\}$ and $y>1$. Then it follows from Lemma 1, Lemma 2 and Corollary 2.1, that there exist $b_1=b_1(y)$, $b_2=b_2(y)\in\mathbf{R}$ such that $b_2>-b_1>0$, $1\geqq -b_1/b_2\geqq 1/2$,

$$\frac{\pi b_1^2}{e_k-e_2}+\frac{\pi b_2^2}{e_t-e_2}=y(2\eta_1+e_2)-\pi \tag{61}$$

and

$$y|2\eta_1+e_2|^2-2\pi(2\eta_1+e_2)=2\pi b_1 b_2, \tag{62}$$

where (k, t) is a permutation of (1, 3). We observe that (61) and (62) when $k=1$ and $t=3$ are the same equations when $k=3$ and $t=1$ if we change the variable b_1 for the variable b_2. Now we use Lemma 3 to show that if $k=1$ and $t=3$ then for each $y>1$ there exists one and only one solution b_1, b_2 of these equations with the property $b_2>0$ and $b_1<0$. Furthermore, $1/2<-b_1/b_2<1$. Thus, if $k=3$ and $t=1$, the observations and results above show that there do not exist solutions b_1, b_2 of these equations with the desired properties. This proves (b) and finishes the proof of Theorem A.

In order to prove Theorem B, we need some results and notations. Let $X\colon D\to\mathbf{R}^3$ be an embedded complete minimal surface of finite total curvature where $D=\{z\in\mathbf{C};\ 0<|z|\leqq 1\}$ and let $N\colon D\to S^2\cong\mathbf{C}\cup\{\infty\}$ be the Gauss normal of X. Suppose that $N(z)\to p\in\{0, \infty\}$, if $z\to 0$. Then there exist a punctured neighbourhood D' of $0\in\mathbf{C}$, $R_1>0$ and $z_1\in\mathbf{C}$ such that $F=X(D')\subset\mathbf{R}^3$ is a graph over $\{z\in\mathbf{C};\ |z-z_1|\geqq R_1\}\subset\mathbf{C}\cong\mathbf{R}^2$. That is, $X(D')$ an be parametrized by

$$X^3(z)=a\log R+b+o(R^{-1}), \qquad |z-z_1|=R\geqq R_1,$$

where a, $b \in \mathbf{R}$ and $o(R^{-1})$ is such that $Ro(R^{-1}) \to 0$ if $R \to \infty$. The number a is called the logarithmic growth of the end F. If $a \neq 0$ ($a=0$) we have a end of catenoid type (a flat end).

Proposition 2 *Let $X_j^3(z)$, $j=1, 2$ be two minimal graphs with logarithmic growth a_j over $A_j=\{z \in \mathbf{C};\ |z-z_j| \geqq R_j\}$, where $z_j \in \mathbf{C}$, $R>0$ and $a_j \in \mathbf{R}$. Suppose that the Gauss normals, $N_j(z)$, of $X_j^3(z)$ satisfy $N_j(z) \to p \in \{0, \infty\}$ if $z \to \infty$ and that $a_2 > |a_1| \geqq 0$. Then*

$$X_2^3(z) - X_1^3(z) \to \infty, \quad \text{if } z \to \infty.$$

Proof. If $a_2 > 0$ and $a_1 < 0$, the proof is trivial. Suppose that $a_2 > a_1 \geqq 0$ and let $z \in \mathbf{C}$ be such that $|z| = R$. Then if R is large enough we have that $|z - z_j| = c_j R$, $1/2 < c_j < 3/2$. So,

$$X_j^3(z) = a_j \log R + a_j \log c_j + b_j + o((c_j R)^{-1}).$$

The expression above and the fact that c_j is bounded proves Proposition 2.

Proof of Theorem B. Let $y \geqq 1$ and let M_y be the Riemann surfaces

$$M_y = \mathbf{C}/L(iy) - \{\pi(0), \pi(1/2), \pi(iy/2)\}. \tag{63}$$

We have proved in Theorem A and in Lemma 3 that there exists a pair of smooth real functions $b_j(y)$, $y \geqq 1$, $j=1, 2$, that are solutions of Eqs. (47) and (48) and such that

$$g = g_y = \frac{P'(z)}{2}\left[\frac{a(y)}{P(z)-e_2} + \frac{b_1(y)}{(e_1-e_2)(P(z)-e_1)} + \frac{b_2(y)}{(e_3-e_2)(P(z)-e_3)}\right] \tag{64}$$

and

$$w = w_y = (P(z)-e_2)\,dz \tag{65}$$

is a Weierstrass representation, (g_y, w_y), of a complete minimal surface S_y: $M_y \to \mathbf{R}^3$,

$$S_y(z) = (S_y^1(z), S_y^2(z), S_y^3(z)) = \operatorname{Re} \int_{w_2}^{z} ((1-g^2), i(1+g^2), 2g)w, \tag{66}$$

where $w_2 = -(1+iy)/2$, $P(z)$ is the Weierstrass P-function of $L(iy)$, $e_j = P(w_j)$, $j=1, 2, 3$, and

$$a(y) + \frac{b_1(y)}{e_1-e_2} + \frac{b_2(y)}{e_3-e_2} = 0, \quad b_2 > -b_1 > 0 \quad \text{and} \quad \frac{1}{2} < \frac{-b_1}{b_2}. \tag{67}$$

We observe that $g = g_y$ is a meromorphic function of order three in $\mathbf{C}/L(iy)$, with simple poles at $\pi(w_j)$, $j=1, 2, 3$ and a zero at $\pi(0)$. $w = w_y$ has a double pole at $\pi(0)$ and a double zero at $\pi(w_2)$. The ends of the surface are the points

$\pi(0)$, $\pi(w_1)$ and $\pi(w_3)$. The point $\pi(w_2)$ is a regular point of the surface. A easy computation (see (7) of Lemma 1 and Theorem A) shows that

$$gw=\frac{P'(z)}{2}\left[\frac{b_1}{P(z)-e_1}+\frac{b_2}{P(z)-e_3}\right]dz$$
$$=[-(b_1+b_2)\zeta(z)+b_1\zeta(z-w_1)+b_2\zeta(z-w_3)+b_1\zeta(w_1)+b_2\zeta(w_3)]\,dz \quad (68)$$

and

$$g^2w=\left[\frac{b_1^2}{e_1-e_2}(P(z-w_1)-e_2)+\frac{b_2^2}{e_3-e_2}(P(z-w_3)-e_2)+2b_1b_2\right]dz. \quad (69)$$

We remark that if $y=1$, then, from Lemma 3, $b_2=-b_2=\sqrt{(\pi/2)}$ and S_1, up to a homothety of $\mathbf{R}^3$ and a rotation around the x^3-axis in $\mathbf{R}^3$, is the surface that appears in [1], [2] and [3]. So, S_1 is embedded. We will prove that (67) is a sufficient condition for the embeddedness of S_y for every $y>1$.

In order to prove this, we observe that S_y, $y>1$, has three ends of catenoid type. Let F_0, F_1 and F_2 be the ends of S_y associated to the points $\pi(0)$, $\pi(w_1)$ and $\pi(w_3)$, respectively. Then from (67) and (68) we have

$$S_y^3(z)\to-\infty, \text{ if } z\to w_3 \quad \text{and} \quad S_y^3(z)\to+\infty, \text{ if } z\to 0 \text{ or } z\to w_1.$$

Then outside of a large compact set in $\mathbf{R}^3$, F_0 and F_1 are contained in the half-space $x^3>0$ and F_3 is contained in the half-space $x^3<0$. On the other hand, in a neighborhood of $0\in\mathbf{C}$, we have

$$\zeta(z)=\frac{1}{z}-\sum_{n=1}^{\infty}c_{2n+2}z^{2n+1}, \quad (70)$$

where

$$c_n=\sum_{s\in L(iy)-\{0\}}\left(\frac{1}{s^n}\right). \quad (71)$$

Then an easy computation by using (68) and (70) (see for example [10, p. 801]) shows that the logarithmic growth of the ends F_j, $j=0, 1, 2$, are, respectively, b_1+b_2, $-b_1$ and $-b_2$. So, the ends F_j, $j=0, 1, 2$ can be parametrized as a graph over the exterior of a disk of $\mathbf{C}$ by

$$X^3(z)=-b_j\log R+a_j+o(R^{-1}), |z-z_j|=R>R', \quad j=0, 1, 2, \quad (72)$$

where $z_j\in\mathbf{C}$, $a_j\in\mathbf{R}$, $R'>0$ and

$$b_0=-(b_1+b_2). \quad (73)$$

We remark that from (67) and (73) we have

$$b_2(y)>-b_1(y)>-b_0(y)>0. \quad (74)$$

By using (63) and (66), we define the smooth functions α: $[1, \infty)\times\mathbf{C}\to\mathbf{C}$ and φ: $[1, \infty)\times M_1\to\mathbf{C}\times\mathbf{R}=\mathbf{R}^3$ by

$$\alpha(y, u+iv)=u+iyv \quad (75)$$

and

$$\varphi(y, p)=\varphi_y(p)=S_y(\alpha(y, z)),\ p=\pi(z),\ z=u+iv. \quad (76)$$

Then φ describes the one-parameter family of minimal surfaces, S_y as a one-parameter family of differentiable surfaces φ_y. Now we will prove two assertions.

Assertion 2 *Let (y_n, p_n) be a convergent sequence in $[1, \infty) \times M_1$, such that $y_n \to y_0$ and $p_n \to p_0 = \pi(z_0)$, $z_0 \in \{0, 1/2, i/2\}$ and let z_n be a convergent sequence in* $\mathbf{C}$ *such that $z_n \to z_0$ and $\pi(z_n) = p_n$. Then there exists a convergent sequence z_n^* in* $\mathbf{C}$ *such that $z_n^* \to z_0$ and*

$$|(\varphi^1 + i\varphi^2)(y_n, p_n) - (\varphi^1 + i\varphi^2)(y_0, p_n^*)| \to 0, \quad \text{if } n \to \infty,$$

where $\varphi = (\varphi^1 + i\varphi^2, \varphi^3)$ and $p_n^* = \pi(z_n^*)$.

Proof. From (65), (66), (69) and (76) we find

$$\begin{aligned}(\varphi^1 + i\varphi^2)(y, \pi(z)) &= (S_y^1 + iS_y^2)(\alpha(y, z)) \\ &= \int_{\alpha(y, w_2)}^{\alpha(y, z)} w - \int_{\alpha(y, w_2)}^{\alpha(y, z)} g^2 w = \frac{b_1^2}{e_1 - e_2}[\zeta(z - w_1) + e_2 z] \\ &\quad + \frac{b_2^2}{e_3 - e_2}[\zeta(z - w_3) + e_2 z] + 2b_1 b_2 z - \overline{\zeta(z) + e_2 z}\,|_{\alpha(y, w_2)}^{\alpha(y, z)},\end{aligned}$$

where $\zeta = \zeta_y$, $w_j = w_j(y)$, $e_j = e_j(y)$, $b_j = b_j(y)$, $j = 1, 2, 3$ are related to the lattice $L(iy)$. We observe that ζ_{y_n} is uniformly convergent to ζ_{y_0} over compact sets K with $K \cap [L(iy_n) \cup L(iy_0)] = \phi$. So, if $z_n \to 0$, we can choose z_n^* such that $\alpha(y_n, z_n) = \alpha(y_0, z_n^*)$ and if $z_n \to w_j$, $j = 1, 3$, we choose z_n^* such that

$$\frac{b_j^2(y_n)}{e_j(y_n) - e_2(y_n)} \frac{1}{\alpha(y_n, z_n) - w_j(y_n)} = \frac{b_j^2(y_0)}{e_j(y_0) - e_2(y_0)} \frac{1}{\alpha(y_0, z_n^*) - w_j(y_0)}.$$

Now an easy computation, by using (70), (71) and the facts above proves Assertion 2.

Assertion 3 *The function $\varphi\colon [1, 1 + a] \times M_1 \to \mathbf{C} \times \mathbf{R}$ is a proper function for every $a > 0$, where we denote with the same letter the function defined in* (76) *and its restriction.*

Proof. Let K be a compact set in $\mathbf{C} \times \mathbf{R}$ and let (y_n, p_n) be a sequence in $\varphi^{-1}(K)$. We reason contradiction. Suppose that there does not exists a convergent subsequence of (y_n, p_n). Then we can suppose that $y_n \to y_0$, $1 \leqq y_0 \leqq 1 + a$ and $p_n \to \pi(z_0)$, $z_0 \in \{0, 1/2, i/2\}$. Let z_n be such that $z_n \to z_0$ and $\pi(z_n) = p_n$. Then by using Assertion 2 we find a sequence $z_n^* \to z_0$ such that

$$|(\varphi^1 + i\varphi^2)(y_n, p_n) - (\varphi^1 + i\varphi^2)(y_0, p_n^*)| \to 0, \quad \text{if } n \to \infty,$$

where $p_n^* = \pi(z_n^*)$. As $\varphi(y_n, p_n) \in K$ and $|(\varphi^1 + i\varphi^2)(y_0, p_n^*)| \to \infty$ if $n \to \infty$ we find a contradiction. This proves Assertion 3.

Assertion 4 *Suppose that $S_a\colon M_a \to \mathbf{C} \times \mathbf{R} = \mathbf{R}^3$ is an embedding for $a \geqq 1$. Then there exist $\varepsilon > 0$ such that $S_y\colon M_y \to \mathbf{C} \times \mathbf{R}$ is an embedding for every $a \leqq y < a + \varepsilon$.*

Proof. We know that $\varphi_a\colon M_1 \to \mathbf{R}^3$ is an embedding. It is sufficient to prove that there exists $\varepsilon > 0$ such that φ_y is an embedding for every $a \leqq y < a + \varepsilon$. Let K_j, $j = 1, 2$ be the solid cylinders $K_j = \{(z, x^3) \in \mathbf{C} \times \mathbf{R};\ |z| \leqq R_j,\ |x^3| \leqq H_j\}$ and let P and Q be the planes $P = \{(z, x^3) \in \mathbf{C} \times \mathbf{R},\ x^3 = H_2\}$ and $Q = \{(z, x^3) \in \mathbf{C} \times \mathbf{R};\ x^3 = -H_2\}$, where $0 < R_1 < R_2$ and $0 < H_1 < H_2$. We can choose R_j and H_j large enough such that:

(a) $\varphi_a(M_1) \cap (\mathbf{R}^3 - K_1)$ is formed by three graphs F_0, F_1 and F_2 (the ends of S_a associated, respectively, to the points $\pi(0)$, $\pi(w_1)$ and $\pi(w_3)$) over the exterior of open disks of $\mathbf{C}$.

(b) There exists $\varepsilon^* > 0$ such that $\varphi_y(M_1) \cap (\mathbf{R}^3 - K_2)$ is again formed by three graphs $F_j(y)$, over the exterior of open disks of $\mathbf{C}$, for every $a \leqq y < a + \varepsilon^*$, where $F_j(y) = (\varphi_y(\varphi_a^{-1}(F_j))) \cap (\mathbf{R}^3 - K_2)$, $j = 0, 1, 2$.

(c) If $a = 1$, then for every $a \leqq y < a + \varepsilon^*$, the border of $F_0(y)$, ∂F_0, is a simple closed curve in the cylinder $C = \{(z, x^3) \in K_2;\ |z| = R_2 \text{ and } |x^3| < H_2/10\}$ and $\partial F_1(y)$, $\partial F_2(y)$ are simple closed curves in the interior of the compact disks $D_1 = K_2 \cap P$ and $D_2 = K_2 \cap Q$, respectively.

(d) If $a > 1$, then $\partial F_2(y)$ is a simple closed curve in the interior of the compact disc $D_2 = K_2 \cap Q$ and $\partial F_0(y) \cup \partial F_1(y)$ is contained in the interior of the compact disc $D_1 = K_2 \cap P$. Furthermore, $\partial F_0(y)$ and $\partial F_1(y)$ are simple closed curves and $\partial F_1(y)$ is contained in the interior of the bounded connected component of $(P - \partial F_0(y))$. We observe that this last property of $\partial F_0(y)$ and $\partial F_1(y)$ is verified. In fact, as S_a is embedded then for R_2 and H_2 large enough and ε^* small, we can suppose that $\partial F_0(a)$ and $\partial F_1(a)$ are simple closed curves and $\partial F_0(a) \cap \partial F_1(a) = \phi$. Also $\partial F_1(a)$ is contained in the interior of the bounded connected component of $(P - \partial F_0(a))$ (to see this use (72), (74) and Proposition 2). Then, by using the smoothness of the family S_y and the fact that $F_0(y)$, $F_1(y)$ are graphs for every $a \leqq y < a + \varepsilon^*$, we conclude the proof of the property desired (see Fig. 3 below).

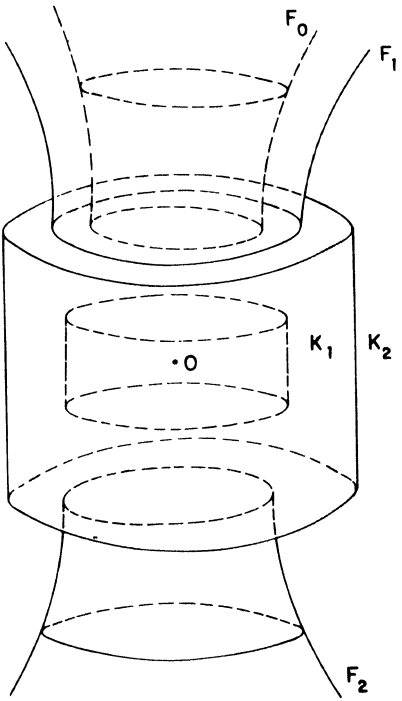


Fig. 3

Now we can prove:

(i) $\varphi_y(M_1)\cap(\mathbf{R}^3-K_2)$ *is embedded for every* $a\leqq y<a+\varepsilon^*$.

Proof. We reason by contradiction. Suppose that there exist $k, j\in\{0, 1, 2\}$, $k<j$ and $a\leqq y<a+\varepsilon^*$ such that $F_k(y)\cap F_j(y)\neq\phi$. Then by using Proposition 2, (72) and (74) we find that $F_k(y)\cap F_j(y)$ is compact and from (c) if $a=1$ or from (d) if $a>1$ of this assertion we have that $\partial F_k(y)\cap\partial F_j(y)=\phi$. Let $\mathscr{F}^s(y)$ be $\mathscr{F}^s(y)=\{p+(0, 0, s)\in\mathbf{R}^3;\ p\in F_k(y)\}$, for every $s\in\mathbf{R}$. Notice that $\mathscr{F}^0(y)=F_k(y)$. Now if $k=0$ and $j=1$, we conclude by using (67), (c) and (b) above and Proposition 2, that there exists $s<0$ such that $\mathscr{F}^s(y)\cap F_1(y)=\phi$. Let $s^*=\sup\{s<0;\ \mathscr{F}^s(y)\cap F_1(y)=\phi\}$. We observe that $\partial\mathscr{F}^{s^*}(y)\cap\partial F_1(y)=\phi$. But, using the maximum principle we conclude that $\mathscr{F}^{s^*}(y)=F_1(y)$. This is a contradiction and proves (i) in the case $k=0$ and $j=1$. Also, by using similar arguments, we can find contradictions in the other cases, $k=0$ and $j=2$ and $k=1$ and $j=2$. This completes the proof of (i).

(ii) *There exist* $\varepsilon>0$, $\varepsilon<\varepsilon^*$ *such that* $\varphi_y(M_1)\cap K_2$ *is embedded for every* $a\leqq y<a+\varepsilon$.

Proof. Suppose by contradiction that there exists a convergent sequence y_n, $y_n\geqq a$, $y_n\to a$ and such that $\varphi_{y_n}(M_1)\cap K_2$ is not embedded for every n. Then there exist sequences p_n and q_n in M_1 with $p_n\neq q_n$ and such that $\varphi(y_n, p_n)=\varphi(y_n, q_n)\in K_2$. As $\varphi/[1, a+\varepsilon^*]\times M_1$ is a proper function (see Assertion 3) we can suppose that $p_n\to p$ and $q_n\to q$, $p, q\in M_1$. If $p\neq q$, we have a contradiction because $\varphi(y_n, p_n)\to\varphi(a, p)$, $\varphi(y_n, q_n)\to\varphi(a, q)$ and $\varphi(a, p)\neq\varphi(a, q)$. If $p=q$, let U be a neighbourhood of $p\in M_1$ and $\delta>0$, such that $\varphi_y(U)$ are embedded for every $a\leqq y<a+\delta$. As $p_n, q_n\in U$ for large n we find a contradiction. This completes the proof of (ii). With (i) and (ii) we finish the proof of Assertion 4.

From Assertion 4 and by using that S_1 is an embedding, we can define $1<y^*\leqq\infty$ such that $y^*=\sup\{s>1;\ S_y$ is an embedding for every $1\leqq y<s\}$. So, to complete the proof of Theorem B it is sufficient to show that $y^*=\infty$. Suppose that $y^*<\infty$. Then, by using the Assertion 3, we conclude that φ_{y^*} is not embedding. In this case, let $p\neq q$ be such that $\varphi_{y^*}(p)=\varphi_{y^*}(q)$ and let U, V be neighbourhoods of p and q respectively, $U\cap V=\phi$ and such that φ_{y^*}/U, φ_{y^*}/V are embedding. Furthermore, we can choose U and V such that φ_{y^*}/U is not transverse to φ_{y^*}/V at the point $\varphi_{y^*}(p)$. In fact, if φ_{y^*}/U is transverse to φ_{y^*}/V at the point $\varphi_{y^*}(p)$ then there exists $\delta>0$ such that $(\varphi_y/U)\cap(\varphi_y/V)\neq\phi$ for every $y^*-\delta<y\leqq y^*$. This is a contradiction with the definition of y^*. Then φ_{y^*}/U is not transverse to φ_{y^*}/V and it follows from the maximum principle (see [10], for example) that $N=\varphi_{y^*}(M_1)\subset\mathbf{R}^3$ is an orientable embedded minimal surface of finite total curvature. So, N is conformally equivalent to $\bar{N}-\{p_1, \ldots, p_k\}$, where $\bar{N}$ is a compact Riemann surface. Moreover, $S_{y^*}: M_{y^*}\to N$ (see (66) and (75)) is a Riemann covering and therefore it has an holomorphic extension, $S_{y^*}: \mathbf{C}/L(iy^*)\to\bar{N}$, to the compact surfaces. As S_{y^*} has embedded ends we have that S_{y^*} is also unbranched at the points $\pi(0)$, $\pi(1/2)$, $\pi(iy^*/2)$. So, S_{y^*} has degree one. This is a contradiction with the fact that φ_{y^*} is not an embedding and proves that $y^*=\infty$. This completes the proof of Theorem B.

Proof of Corollary B.1. This follows immediately from Theorems A and B.

Proof of Corollary B.2. Let $X: M\mapsto\mathbf{R}^3$ be a complete minimal surface M embedded in $\mathbf{R}^3$ with finite total curvature 12π. Then, from Theorem 1, M is conformal-

ly equivalent to a compact Riemann surface $\bar{N}_\gamma$ of genus γ punctured at n points. Since M is embedded it follows from the formulas of Jorge-Meeks [7] that $12\pi = -2\pi(2-2\gamma-2n)$. So, $\gamma+n=4$. But, M cannot be embedded in $\mathbf{R}^3$ as a complete minimal surface of finite total curvature in the three following cases: (1) $M \cong S^2-\{q_1, \ldots, q_n\}$, $3 \leqq n \leqq 5$, (2) $M \cong M_\gamma-\{q_1, q_2\}$, $\gamma>0$ and (3) $M \cong \bar{M}_\gamma-\{q_1\}$, $\gamma>0$. Cases (1) and (2) follows from [7] and [10] respectively and the case (3) follows from the fact that the coordinates of a minimal immersion are harmonic functions. Thus, $\gamma=1$ and $n=3$. On the other hand, Corollary B.1 shows that S_y, $y \geqq 1$ are all complete minimal surfaces embedded in $\mathbf{R}^3$ of genus one, three ends and finite total curvature. This completes the proof of Corollary B.2.

References

1. Costa, C.J.: Example of a complete minimum immersion of genus one an three embedded ends. Bol. Soc. Bras. Mat. **15**, 47–54 (1984)
2. Costa, C.J.: Uniqueness of complete minimal surfaces in $\mathbf{R}^3$ of genus one and three ends. J. Differ. Geom. **30**, 597–618 (1989)
3. Hoffman, D., Meeks, W.: A complete embedded minimal surface in $\mathbf{R}^3$ with genus one and three ends. J. Differ. Geom. **21**, 109–127 (1985)
4. Hoffman, D., Meeks, W.: One parameter families of embedded minimal surface. (in preparation)
5. Lang, S.: Elliptic functions. Reading, MA: Addison Wesley 1973
6. Greenberg, M., Harper, S.: Algebraic Topology: A First Course. Reading, MA: Benjamin-Cummings 1981
7. Jorge, J.P., Meeks, W.: The topology of complete minimal surfaces of finite total Gaussian curvature. Topology **22**, (1983)
8. Lopez, F.J., Ros, A.: On embedded complete minimal surfaces of genus zero. (Preprint)
9. Osserman, R.: A survey of minimal surfaces. van Nostrand Reinhold 1969
10. Schoen, R.: Uniqueness, simetry and embeddedness of minimal surfaces. J. Differ. Geom. **18** (1983)
11. Tannery, J., Molk, R.: Éléments de la théorie des fonctions elliptiques. Bronx, New York: Chelsea 1972

Invent. math. 105, 305–346 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Coefficients généralisés de séries principales sphériques et distributions sphériques sur $G_{\mathbb{C}}/G_{\mathbb{R}}$ ⋆

P. Delorme

Département de Mathématique-Informatique, Faculté des Sciences de Luminy, 163, avenue de Luminy, 13288-Marseille Cedex 9, Unité de Recherche Associée au C.N.R.S. n° 225

Oblatum 13-XII-1990

0 Introduction

Soit G un groupe de Lie semi-simple complexe, connexe, simplement connexe, d'algèbre de Lie $\mathfrak{g}$. Soit $\mathfrak{h}$ une forme réelle de $\mathfrak{g}$ et σ la conjugaison par rapport à $\mathfrak{h}$. On note aussi σ l'involution correspondante de G et H le sous-groupe de G formé des points fixes de σ. On suppose que H *admet un sous-groupe de Cartan compact* T. Soit θ une involution de Cartan de G commutant à σ, K le groupe des points fixes de θ. On peut choisir T inclus dans K. Si $\mathfrak{t}$ est l'algèbre de Lie de T, on note $\mathfrak{a}=i\mathfrak{t}\subset\mathfrak{g}$, $A=\exp\mathfrak{a}$. Alors TA est un sous-groupe de Cartan de G. Soit $P=TAN$ un sous-groupe de Borel de G (qui est $\sigma\theta$-stable). Pour $\lambda\in\mathfrak{a}'_{\mathbb{C}}$, on introduit la série principale sphérique C^∞ de G, (π_λ, I_λ). Par restriction des fonctions à K, I_λ est isomorphe à $I=C^\infty(K/T)$. On notera $\tilde{\pi}_\lambda$ la représentation de G sur I déduite de π_λ par transport de structure. Notons ρ la demi-somme des racines de $\mathfrak{a}$ dans l'algèbre de Lie de N (regardée comme réelle). Pour $\mathrm{Re}(\lambda-\rho)$ strictement dominant, la fonction ξ_λ sur G définie par:

$$\xi_\lambda(g)=0 \quad \text{si } g\notin HP$$
$$\xi_\lambda(htan)=a^{\lambda-\rho}, \quad \forall h\in H, t\in T, a\in A, n\in N$$

est une fonction continue sur G (cf. [1, 13] et § 3.2), et permet de définir sur I_λ une forme linéaire continue, notée encore ξ_λ, par:

$$\forall f\in I_\lambda, \quad \langle\xi_\lambda, f\rangle=\int_K \xi_\lambda(k)\, f(k)\, dk.$$

Par transport sur I, la famille de formes linéaires $\lambda\to\tilde{\xi}_\lambda$ se prolonge en une fonction méromorphe sur $\mathfrak{a}'_{\mathbb{C}}$ à valeurs dans $\mathscr{D}'(K/T)$ (cf. [1, 13]). Par ailleurs ξ_λ est invariante par H, G agissant sur I'_λ, dual topologique de I_λ, par la représentation contragrédienté de π_λ, π'_λ. Lorsque ξ_λ et $\xi_{-\lambda}$ sont définies, il n'est pas

⋆ À ma mère

difficile de voir que, pour $\varphi \in C_c^\infty(G/H)$, $\pi'_\lambda(\varphi)\ \xi_\lambda$ s'identifie à un élément de $I_{-\lambda}$ et l'expression

$$\langle \pi'_\lambda(\varphi)\ \xi_\lambda, \xi_{-\lambda} \rangle = \Theta_\lambda(\varphi)$$

a un sens. On définit ainsi une distribution sur G/H, Θ_λ, qui est invariante à gauche par H et propre sous l'algèbre des opérateurs différentiels G-invariants sur G/H, qui dépend méromorphiquement de λ.

L'objet de cet article est d'établir certaines propriétés de Θ_λ.

On sait d'après P. Harinck ([10]) que Θ_λ est une fonction localement intégrable sur G/H. De plus, notant A^{reg} l'ensemble des éléments réguliers de A, Θ_λ est une fonction analytique sur l'ouvert $HA^{\mathrm{reg}}H$ qui vérifie (au moins pour λ dans un ouvert de $\mathfrak{a}'_{\mathbb{C}}$):

$$(*) \qquad \forall a \in A^{\mathrm{reg}}, \qquad \Theta_\lambda(aH) = \frac{\sum_{x \in W} d(x, \lambda)\, a^{-x\lambda}}{\Delta_{\mathscr{A}}(a)}$$

où W est le groupe de Weyl de (G, A), $\Delta_{\mathscr{A}}(a)$ est une fonction indépendante de λ (voir définition au § 7.2), et les $d(x, \lambda)$ sont des complexes dépendant méromorphiquement de λ.

Le problème résolu dans cet article est la détermination des $d(x, \lambda)$. Notons $W_H = \{w \in W \mid w \text{ a un représentant dans } H\}$. Alors on montre en particulier que $d(x, \lambda) = 0$ si $x \notin W_H$.

En outre, on montre que:

$$\forall w \notin W_H, \qquad \forall a \in A^{\mathrm{reg}}, \qquad \Theta_\lambda(waH) = 0.$$

Ces résultats font l'objet du Théorème 3 (§ 7.1) et permettent de comparer Θ_λ aux distributions construites par P. Harinck dans [10].

L'idée pour établir ces propriétés est d'étudier $\Theta_\lambda(waa_t H)$ lorsque $t \to +\infty$, où $a_t = \exp tX_0$ et X_0 est un élément régulier de $\mathfrak{a}$.

Mais Θ_λ étant présentée comme distribution, il faut passer par les fonctions tests. On se limitera dans la suite de cette introduction à $w = 1$.

On construit d'abord (§ 4.1) des fonctions C_c^∞ sur G/H portées par HAH. Pour cela on note C une chambre de Weyl dans $\mathfrak{a}$ pour W (pas nécessairement égale à la chambre C^+ correspondant à N). On pose $\mathscr{C} = \exp C \subset A$. On note $\Omega(C) = H\mathscr{C}H$ qui est un ouvert dans G/H. Si $\psi \in C_c^\infty(H/T)$ et $\varphi \in C_c^\infty(\mathscr{C})$ $(\subset C_c^\infty(\mathfrak{a}))$ on définit $\Phi = \Phi(\psi, \varphi) \in C_c^\infty(G/H)$ par:

$$\Phi(gH) = 0 \quad \text{si } gH \notin \Omega(C)$$

$$\Phi(haH) = \psi(h)\, \varphi(a) \qquad \forall h \in H, \qquad \forall a \in \mathscr{C}.$$

On prend $X_0 \in -C$ et on définit $\varphi_t \in C_c^\infty(\mathscr{C})$ par:

$$\varphi_t(a) = 0 \quad \text{si } aa_t \notin \mathscr{C}$$

$$\varphi_t(aa_t) = \varphi(aa_t) \quad \text{si } aa_t \in \mathscr{C}.$$

Lorsque $x \in W$ et $C = -xC^+$, on étudie de deux façons différentes:

$$\lim_{t \to +\infty} a_t^{-x(\lambda+\rho)} \Theta_\lambda(\Phi(\psi, \varphi_t))$$

(pour des valeurs convenables de λ).

La première méthode consiste à partir de (*) et à utiliser la formule intégrale (établie à la Proposition 3, § 2.7):

$$\forall f \in C_c(\Omega(C))\,(\subset C_c(G/H)),$$
$$\int_{\Omega(C)} f(gH)\,d\dot{g} = C_G \int_{H/T} \int_{\mathscr{C}} f(haH)\,D(a)\,da\,d\dot{h},$$

où C_G est une constante dépendant des normalisations des mesures et D une fonction explicite sur A.

On trouve facilement (Lemme 18, § 7):

$$\lim_{t \to +\infty} a_t^{-x(\lambda+\rho)} \Theta_\lambda(\Phi(\psi, \varphi_t)) = \varepsilon(x)\,d(x, \lambda) \Big(\int_{H/T} \psi(hT)\,d\dot{h} \Big) \Big(\int_{\mathscr{C}} \varphi(a)\,a^{-x(\lambda+\rho)}\,da \Big)$$

où $\varepsilon(x) = \pm 1$.

La deuxième méthode d'étude de cette limite consiste à écrire la définition de Θ_λ et à étudier d'abord $\lim_{t \to +\infty} a_t^{-x(\lambda+\rho)} \pi'_\lambda(\Phi(\psi, \varphi_t))\,\xi_\lambda$, qui est un élément de $I_{-\lambda}$ (identifié à un sous-espace de I'_λ).

On démontre (Théorème 1, § 4.3) que cette limite dans $I_{-\lambda}$ existe et est égale à:

$$0 \quad \text{si } x \notin W_H$$
$$C_G\,u_{\lambda,x} \quad \text{si } x \in W_H$$

où

$$u_{\lambda,x} = \int_{H/T} \int_{\mathscr{C}} \psi(hT)\,\varphi(a)\,a^{-2x\rho}\,\pi'_\lambda(hwa)\,\eta_{\lambda,x}\,d\dot{h}\,da$$

et

$$\eta_{\lambda,x} = A'(w_0, \lambda)\,\delta_{xw_0}^{w_0\lambda}.$$

Ici $A(w_0, \lambda)$ est l'intégrale d'entrelacement de I_λ vers $I_{w_0\lambda}$ (qui converge avec nos restrictions sur λ), $A'(w_0, \lambda)$ est son transposé et $\delta_{xw_0}^{w_0\lambda}$ est l'élément de $I'_{w_0\lambda}$ défini par:

$$\forall f \in I_{w_0\lambda}, \quad \langle \delta_{xw_0}^{w_0\lambda}, f \rangle = f(xw_0)$$

(la formule pour $\eta_{\lambda,x}$ étant établie au Lemme 11, § 5).

Précisons que ce résultat est obtenu par un argument de convergence dominée, les majorations nécessaires étant obtenues au § 4.

On a alors:

$$\lim_{t \to +\infty} a_t^{-x(\lambda+\rho)} \Theta_\lambda(\Phi(\psi, \varphi_t)) = 0 \quad \text{si } x \notin W_H$$
$$= C_G \langle u_{\lambda,x}, \xi_{-\lambda} \rangle \quad \text{si } x \in W_H.$$

On utilise alors les propriétés des intégrales d'entrelacement (entrelacement, propriétés relatives à la transposition) et on fait passer l'intégrale d'entrelacement à droite (en transposant) pour obtenir:

$$\forall x \in W_H, \quad \langle u_{\lambda,x}, \xi_{-\lambda}\rangle = \langle v_{\lambda,x}, A'(w_0, -w_0\lambda)\,\xi_{-\lambda}\rangle$$

où

$$v_{\lambda,x} = \int_{H/T} \int_{\mathscr{C}} \psi(hT)\,\varphi(a)\,a^{-x(\lambda+\rho)}\,\pi'_{w_0\lambda}(h)\,\delta_{xw_0}^{w_0\lambda}\,dh\,da.$$

On calcule alors l'effet des intégrales d'entrelacement sur $\xi_{-\lambda}$ (Théorème 2, § 5) en se réduisant aux groupes de rang 1 ($\mathrm{SL}(2,\mathbb{C})$), pour lesquels les calculs sont faits explicitement (§ 6).

On voit alors que le support de $v_{\lambda,x}$ est contenu dans un ouvert sur lequel $A'(w_0, -w_0\lambda)\,\xi_{-\lambda}$ est une fonction C^∞. Le calcul s'achève par l'emploi d'une formule intégrale. La comparaison des résultats obtenus par les 2 méthodes donne le théorème cherché.

Situons notre travail par rapport à ceux d'autres auteurs.

Dans le cas $\mathrm{SL}(2,\mathbb{C})/\mathrm{SL}(2,\mathbb{R})$ le résultat principal (Théorème 3) est dû explicitement à P. Harinck ([20, Th. 3.4]) et repose sur les résultats de Faraut ([8]). Le cas $\mathrm{GL}(n,\mathbb{C})/U(p,q)$ a été également traité par N. Bopp et P. Harinck (travail en cours). Ce résultat joue un rôle important dans la dérivation de la formule de Plancherel pour ces espaces symétriques, en permettant d'identifier, dans les formules d'inversion du type de celles obtenues par S. Sano ([21]), des distributions sphériques à des coefficients généralisés (cf. [20, § 4]). Notre méthode diffère de celles de ces auteurs.

Remerciements. Je remercie vivement Pascale Harinck pour d'utiles conversations dans lesquelles elle m'a expliqué ses résultats sur $\mathrm{SL}(2,\mathbb{C})$. Je remercie également Jacques Carmona pour m'avoir patiemment écouté et fait bénéficier de ses critiques éclairées pendant l'élaboration de ce travail.

1 Notations-Conventions

1.1

Si X est un ensemble, on notera $|X|$ son cardinal.

Si E est un espace vectoriel réel, on notera $E_{\mathbb{C}}$ son complexifié.

Si E est un espace vectoriel complexe et F un sous-espace vectoriel réel de E avec $F \cap iF = \{0\}$ on identifiera parfois $F_{\mathbb{C}}$ à $F + iF \subset E$.

Si E est un espace vectoriel localement convexe séparé, on notera E' son dual topologique. Le crochet de dualité sera noté $\langle\ ,\ \rangle$ et parfois $\langle\ ,\ \rangle_{E,E'}$. Si A est un opérateur continu de E dans F (avec E, F *e.l.c.s*), on notera A' le transposé de A qui envoie F' dans E'.

Si l'on a une représentation π d'un groupe G dans E, *e.l.c.s*, par des opérateurs continus, on notera π' la représentation contragrédiente définie par $\pi'(g) = (\pi(g^{-1}))'$.

Si E est un espace de Hilbert (réel ou complexe) on notera $(\ ,\ )$ le produit scalaire (linéaire en la 1ère variable).

1.2

Si L est un groupe de Lie, on notera $\mathfrak{l}$ son algèbre de Lie, $U(\mathfrak{l})$ l'algèbre enveloppante de $\mathfrak{l}_{\mathbb{C}}$. L'action régulière gauche (resp. droite) de $D \in U(\mathfrak{l})$ (resp. $g \in G$) sur f, fonction C^∞ sur L, sera notée $L_D f$ (resp. $R_D f$, resp. $L_g f$, $R_g f$).

1.3

Pour les généralités qui suivent nous renvoyons à [3, § 2], [4, § 1.3] et [17].

Si X est une variété, on notera $\mathscr{D}(X) = C_c^\infty(X)$, $\mathscr{E}(X) = C^\infty(X)$ munis de leur topologie usuelle et $\mathscr{D}'(X)$ (resp. $\mathscr{E}'(X)$) l'espace des distributions (resp. distributions à support compact) sur X. Si X est un groupe de Lie (resp. un espace homogène d'un groupe de Lie) muni d'une mesure de Haar (resp. invariante) à gauche, on identifie $\mathscr{E}(X)$ (resp. $\mathscr{D}(X)$) à un sous-espace de $\mathscr{D}'(X)$ (resp. $\mathscr{E}'(X)$). Si G est un groupe de Lie (unimodulaire pour simplifier) muni d'une mesure de Haar dg, pour φ, fonction sur G, on note $\check{\varphi}$ la fonction sur G définie par $\check{\varphi}(g) = \varphi(g^{-1})$. On étend cette opération aux distributions (par transposition). Si (π, I) est une représentation C^∞ de G sur un espace localement convexe séparé complet, il existe une représentation de l'algèbre de convolution $\mathscr{E}'(G)$ dans I, notée π, telle que:

$$\forall \tau \in \mathscr{E}'(G), \quad \forall f \in I, \quad \forall T \in I', \tag{1}$$
$$\langle \pi(\tau) f, T \rangle_{I, I'} = \langle \tau, c_{f,T} \rangle_{\mathscr{E}', \mathscr{E}}$$

où $c_{f,T}$ est le coefficient de π défini par

$$c_{f,T}(g) = \langle \pi(g) f, T \rangle.$$

Il existe aussi, sous les mêmes hypothèses, une représentation de $\mathscr{E}'(G)$ sur I', notée π' telle que:

$$\forall \tau \in \mathscr{E}'(G), \quad f \in I, \quad T \in I' \tag{2}$$
$$\langle f, \pi'(\tau)\, T \rangle = \langle \tau, \check{c}_{f,T} \rangle.$$

Alors (1) et (2) impliquent:

$$\forall \tau \in \mathscr{E}'(G), \quad (\pi(\tau))' = \pi'(\check{\tau}) \tag{3}$$

2 Quelques faits sur la structure de G/H et une formule intégrale

2.1

Soit G un groupe de Lie semi-simple complexe, connexe et simplement connexe d'algèbre de Lie $\mathfrak{g}$.

Soit $\mathfrak{h}$ une forme réelle de $\mathfrak{g}$. On note σ la conjugaison de $\mathfrak{g}$ par rapport à $\mathfrak{h}$, ainsi que l'involution de G qu'on en déduit. On note H l'ensemble des

points fixes de σ dans G. On suppose dans tout cet article que H *admet un sous-groupe de Cartan compact.*

Soit $\mathfrak{q}=\{X\in\mathfrak{g}\,|\,\sigma(X)=-X\}$. On a $\mathfrak{g}=\mathfrak{h}\oplus\mathfrak{q}$ et $\mathfrak{q}=i\mathfrak{h}$. On se fixe une involution de Cartan, θ, de G qui commute à σ. On note K le sous-groupe des points fixes de θ et $\mathfrak{p}=i\mathfrak{k}=\{X\in\mathfrak{g}\,|\,\theta(X)=-X\}$.

On se fixe un produit scalaire sur $\mathfrak{g}$ (regardée comme réelle), $\mathscr{B}$, défini par:

$$\forall X,\ Y\in\mathfrak{g},\qquad \mathscr{B}(X,Y)=-\mathrm{Tr}_{\mathbb{R}}(\mathrm{ad}\,X\ \mathrm{ad}(\theta(Y))).$$

On note T un sous-groupe de Cartan compact de H contenu dans K. On note $\mathfrak{a}=it\subset\mathfrak{g}$ et A le sous-groupe analytique de G d'algèbre de Lie $\mathfrak{a}$. Si P est un sous-groupe parabolique minimal de G contenant TA on notera N_P (ou N) son radical unipotent. On note W le quotient du normalisateur dans K de $\mathfrak{a}$, $N_K(\mathfrak{a})$ par son centralisateur dans K (qui est égal à T). C'est aussi le groupe de Weyl de (K, T). On note W_H le quotient du normalisateur dans $K\cap H$ de $\mathfrak{a}$ par T qui est aussi le groupe de Weyl de $(K\cap H, T)$. On a $W_H\subset W$. Si $w\in W$ on notera de la même façon un représentant de w dans $N_K(\mathfrak{a})$. Noter que, comme $T\subset H$, on a $w\in W_H$ si et seulement si tout représentant de w est dans H.

Remarque 1 Le fait de ne pas distinguer entre éléments de W et représentants ne conduira pas dans cet article à des problèmes fâcheux. En effet, à chaque fois qu'une construction fera intervenir un représentant de $w\in W$, on verra que celle-ci ne dépend pas du choix de celui-ci. Cela permettra un certain allègement des notations.

2.2

On se fixe un parabolique minimal de G, P, contenant TA. On note Δ l'ensemble des racines de $\mathfrak{t}_{\mathbb{C}}=\mathfrak{t}\oplus\mathfrak{a}$ dans $\mathfrak{g}$ et Δ^+ (ou Δ_P^+) l'ensemble des racines de $\mathfrak{t}_{\mathbb{C}}$ dans $\mathfrak{n}$. On notera $\rho_{\mathbb{C}}$ (ou $\rho_{\mathbb{C},P}$) la demi-somme des éléments de Δ^+. On notera δ_i, $i=1,\ldots,\ell$ les poids fondamentaux de Δ (relativement à Δ^+) et pour chaque i, on note (π_i, V_i) la représentation irréductible holomorphe de G de plus haut poids δ_i, que l'on munit d'un produit scalaire invariant par K. On notera e_{δ_i} un vecteur de plus haut poids de π_i et de norme 1, et $e_i^1=e_{\delta_i}$, e_i^2, …, $e_{\delta_i}^{n_i}$ une base orthonormée de V_i formée de vecteurs-poids sous $\mathfrak{t}_{\mathbb{C}}$ (e_i^j étant de poids μ_j). Dans $V_i\otimes V_i'$ on considère le vecteur $v_H^i=\sum_{j=1}^{n_i} e_i^j\otimes(e_i^j)'$ où $((e_i^j)')$ est la base duale de (e_i^j). On munit V_i' du produit scalaire déduit naturellement de celui sur V_i. Notez que v_H^i correspond à l'identité dans l'identification de $V_i\otimes V_i'$ avec $\mathrm{End}_{\mathbb{C}}(V_i)$ et ne dépend pas du choix de la base orthonormée choisie. On définit alors une fonction ε_i sur G par:

$$\forall g\in G,\qquad \varepsilon_i(g)=((\pi_i(g)\otimes\pi_i'(\sigma(g)))(e_{\delta_i}\otimes e'_{\delta_i}), v_H^i).$$

On a aussi

$$\forall g\in G,\qquad \varepsilon_i(g)=\langle\pi_i(g)\,e_{\delta_i}, \pi_i'(\sigma(g))\,e'_{\delta_i}\rangle.$$

La fonction ε_i est analytique sur G et vérifie

$$\forall h\in H, g\in G, t\in T, a\in A, n\in N,\qquad \varepsilon_i(hgtan)=a^{2\delta_i}\varepsilon_i(g).$$

Ici on a noté $a \to a^{2\delta_i}$ le quasi-caractère de A de différentielle $2\delta_{i|\mathfrak{a}}$. Plus généralement, si λ est une forme linéaire sur $\mathfrak{t}_{\mathbb{C}}$ ou $\mathfrak{a}$, on note $x \to x^{\lambda}$ le quasicaractère de TA ou A de différentielle λ. Notons que la demi-somme des poids de $\mathfrak{a}$ dans $\mathfrak{n}$ (regardé comme réel), ρ (ou ρ_P) est égal à $2\rho_{\mathbb{C}|\mathfrak{a}}$.

2.3

Lemme 1 *Si w est un représentant d'un élément de W, $\sigma(w^{-1})\, w$ est un élément de T indépendant du représentant choisi, et noté t_w. On a $t_w^2 = 1$ et $\sigma(w^{-1})\, w = w^{-1}\sigma(w) = t_w$. De plus $t_w = 1$ si et seulement si $w \in W_H$.*

Démonstration. Si $t \in T$, wtw^{-1} est dans T et donc fixé par σ. D'où $wtw^{-1} = \sigma(w)\, t\sigma(w)^{-1}$. Donc $\sigma(w)^{-1} w$ centralise T et c'est un élément, t_w, de T. D'autre part, si $t \in T$, $\sigma(wt)^{-1} wt = t^{-1}\sigma(w)^{-1} wt = t^{-1} t_w t = t_w$. Donc t_w est indépendant du représentant choisi. On a $\sigma(t_w) = t_w$ puisque $t_w \in T$.

Alors:

$$t_w^2 = t_w\, \sigma(t_w) = \sigma(w)^{-1} w(\sigma(\sigma(w)^{-1} w))$$
$$= \sigma(w)^{-1} w w^{-1} \sigma(w) = 1.$$

D'où:

$$t_w = t_w^{-1} \quad \text{et} \quad \sigma(w)^{-1} w = w^{-1}\sigma(w) = t_w.$$

Maintenant:

$$t_w = 1 \Leftrightarrow \sigma(w)^{-1} w = 1$$
$$\Leftrightarrow w = \sigma(w).$$

Donc:

$$t_w = 1 \Leftrightarrow w \in W_H. \quad \square$$

2.4

On rappelle que les doubles classes ouvertes de G modulo (H, P) sont exactement les ensembles $HwTAN = HwAN$ où $w \in W$ (notez que wT ne dépend que de $w \in W$) (cf. [1, Prop. B.1]).

Etudions les fonctions ε_i sur ces doubles classes.

On a, avec les notations évidentes:

$$\varepsilon_i(hwtan) = a^{2\delta_i}\, \varepsilon_i(w).$$

Or:

$$\varepsilon_i(w) = \langle \pi_i(w)\, e_{\delta_i}, \pi_i'(\sigma(w))\, e'_{\delta_i} \rangle$$
$$= \langle \pi_i(\sigma(w)^{-1} w)\, e_{\delta_i}, e'_{\delta_i} \rangle.$$

D'où $\varepsilon_i(w) = t_w^{\delta_i}$.

Comme $t_w^2 = 1$, on a $t_w^{\delta_i} = \pm 1$. Comme G est simplement connexe, les caractères $t \to t^{\delta_i}$ engendrent le groupe des caractères de T. Donc $t_w^{\delta_i} = 1$ pour tout i équivaut à $t_w = 1$, i.e. $w \in W_H$.

Proposition 1 (i) *Les fonctions ε_i sont réelles et de signe constant sur chaque double classe ouverte de G modulo (H, P).*

(ii) $G - \bigcup_{w \in W} HwAN = \left\{ g \in G \mid \prod_{i=1}^{\ell} \varepsilon_i(g) = 0 \right\}$.

(iii) $HAN = \{g \in G \mid \varepsilon_i(g) > 0, \forall i = 1, \ldots, \ell\}$.

Démonstration. On a vu (i). Pour (ii), voir la démonstration du Théorème 2.5 de [13]. Les fonctions considérées par Olafson ne sont pas exactement les mêmes que les nôtres ou du moins sont présentées de façon différente. Pour (iii), on voit que, si $g \in HAN$, les $\varepsilon_i(g)$ sont strictement positifs. Réciproquement, si $\varepsilon_i(g) > 0$ pour tout i, de (ii) on déduit que $g \in HwAN$ pour un $w \in W$. D'après ce que l'on a vu, $\varepsilon_i(w) > 0$ pour tout i implique $w \in W_H$ et $HwAN = HAN$. □

2.5

On note que $\mathfrak{a}$ est un sous-espace de Cartan de $\mathfrak{q}$ (cf. [10, Def. 2.2]). On note $\mathscr{A}$ le sous-ensemble de Cartan de $G|H$ correspondant (cf. [10, Def. 2.2]), i.e.:

$$\mathscr{A} = \{gH \in G/H \mid g\sigma(g)^{-1} \text{ centralise } \mathfrak{a}\}.$$

Proposition 2 *(i)* $\mathscr{A} = \bigcup_{w \in W} AwH = \bigcup_{w \in W} wAH$.

(ii) On a $AwH \cap Aw'H \neq \emptyset$ si et seulement si $wW_H = w'W_H$ auquel cas $AwH = Aw'H$.

Démonstration. On sait d'après [14, § 3], que $\mathscr{A} = \bigcup_{i=1}^{m} Ak_iH$ pour des $k_i \in K$.

Soit $k \in K$ tel que $k\sigma(k)^{-1}$ centralise $\mathfrak{a}$ (donc T). Alors:

$$(\mathrm{Ad}\, k)\, \mathrm{Ad}(\sigma(k)^{-1})(T) = T$$
$$\mathrm{Ad}(\sigma(k))^{-1}(T) = \mathrm{Ad}(k^{-1})(T).$$

Notons $T_1 = \mathrm{Ad}(k^{-1})\, T$. C'est un sous-groupe de Cartan de K. Montrons que T_1 est inclus dans H. En effet, si $t_1 \in T_1$, on a $t_1 = k^{-1}tk$ pour un $t \in T$. Alors

$$t_1\, \sigma(t_1)^{-1} = k^{-1}tk\, \sigma(k)^{-1}t^{-1}\sigma(k).$$

Et comme $k\sigma(k)^{-1}$ centralise T:

$$t_1\, \sigma(t_1)^{-1} = k^{-1}tt^{-1}k\sigma(k)^{-1}\sigma(k) = 1.$$

Donc $t_1 \in H$ et T_1 est un sous-groupe de Cartan de $K \cap H$. Il existe $h \in H \cap K$ tel que:

$$\mathrm{Ad}\, h(T_1) = T, \quad \text{soit encore}$$
$$\mathrm{Ad}(hk^{-1})\, T = T.$$

Donc hk^{-1} normalise T ainsi que kh^{-1}. Alors $kH = wH$ pour un $w \in W$. D'où:

$$\mathscr{A} \subset \bigcup_{w \in W} AwH.$$

Réciproquement, si $g=awh$ avec $a\in A$, $h\in H$ et $w\in N_K(\mathfrak{a})=N_K(T)$, on a:

$$g\sigma(g)^{-1}=awa\sigma(w)^{-1}a^{-1}.$$

Or $\sigma(w)^{-1}w=t_w$, i.e. $w=\sigma(w)\,t_w$. Donc $w\sigma(w)^{-1}=\sigma(w)\,t_w\,\sigma(w)^{-1}=\sigma(wt_w w^{-1})$. Comme w normalise T et que σ fixe T, on a: $w\sigma(w)^{-1}=wt_w w^{-1}\in T$. Comme A et T commutent, on a:

$$g\sigma(g)^{-1}=wt_w w^{-1}\in T,$$

et $g\sigma(g)^{-1}$ centralise T. Ceci achève de prouver (i).

Prouvons (ii). Soient w, $w'\in N_K(\mathfrak{a})$. Comme w, w' normalisent A, on a:

$$AwH\cap Aw'H\neq\emptyset\Leftrightarrow w^{-1}w'\in AH.$$

Or $H=\exp(\mathfrak{p}\cap\mathfrak{h})\,(K\cap H)$ puisque $\mathfrak{h}=(\mathfrak{k}\cap\mathfrak{h})\oplus(\mathfrak{p}\cap\mathfrak{h})$ est une décomposition de Cartan de $\mathfrak{h}$. Alors:

$$AwH\cap Aw'H\neq\emptyset\Leftrightarrow\exists a_0\in A,\quad X\in\mathfrak{p}\cap\mathfrak{h},$$
$$k\in K\cap H,\quad w^{-1}w'=a_0(\exp X)\,k.$$

Comme $w^{-1}w'\in K$, en appliquant θ aux deux membres de cette dernière égalité, on obtient:

$$a_0(\exp X)\,k=a_0^{-1}(\exp -X)\,k,$$

soit encore

$$a_0^2=\exp(-2X).$$

Or $A\cap H$ est réduit à l'élément neutre et exp est injective sur $\mathfrak{p}$. D'où $a_0=1$ et $X=0$. Finalement

$$AwH=Aw'H\Leftrightarrow w^{-1}w'\in K\cap H.\quad\square$$

2.6 *Choix des mesures de Haar*

Le produit scalaire euclidien $\mathscr{B}$ sur $\mathfrak{g}$ introduit en 2.1 permet de sélectionner sur les sous-groupes de Lie de G (et sur G lui-même) une mesure de Haar à gauche qu'on appellera mesure standard (cf. [11, § 7]).

Si $L\subset G$ est compact, la mesure ainsi sélectionnée ne sera pas en général de masse totale 1. On notera $\mathrm{vol}(L)$ cette masse. On appellera mesure normalisée la mesure de Haar sur L de masse totale 1. A l'exception des sous-groupes K et T de G pour lesquels on utilisera les mesures normalisées (notées dk, dt), on utilisera les mesures à gauche standard (sauf mention expresse du contraire) (notée $d\ell$ pour un sous-groupe L). Si $Q\subset L$ sont des sous-groupes de Lie de G et qu'il existe une mesure invariante à gauche sur L/Q, on notera $d\dot{\ell}$ celle qui vérifie:

$$\forall f\in C_c(L),\ \int_L f(\ell)\,d\ell=\int_{L/Q}\left(\int_Q f(\ell q)\,dq\right)d\dot{\ell}.$$

2.7 *Une formule intégrale*

Avec ces choix, on a la proposition suivante:

Proposition 3 *Soit $w\in W$. Soit C une chambre de Weyl dans $\mathfrak{a}$ relativement à Δ (pas nécessairement celle correspondant à $\Delta^+=\Delta_P^+$). On note $\mathscr{C}=\exp C$. Alors:*

(i) l'application F_w: $H/T\times\mathscr{C}\to G/H$ *définie par* $F_w(hT, a)=hwaH$ *est un difféomorphisme de* $H/T\times\mathscr{C}$ *sur son image notée* $\Omega_w(C)$.

(ii) Si φ *est continue sur* G/H *à support compact dans* $\Omega_w(C)$, *on a*:

$$\int_{G/H}\varphi(gH)\,d\dot{g}=C_G\int_{H/T}\int_{\mathscr{C}}\varphi(hwaH)\,D_w(a)\,da\,d\dot{h}$$

où:

$$\begin{aligned}D_w(a)&=\prod_{\alpha\in\Delta^+}|a^{-\alpha}-a^{\alpha}t_w^{\alpha}|^2\\&=\prod_{\alpha\in\Delta^+}|a^{\alpha}-a^{-\alpha}t_w^{-\alpha}|^2\\&=\prod_{\alpha\in\Delta}|a^{-\alpha}-a^{\alpha}t_w^{\alpha}|\end{aligned}$$

et $C_G=(\frac{1}{2})^{|\Delta|}(\text{vol } T)$.

Démonstration. Prouvons l'injectivité de F_w.

Supposons $hwaH=h'wa'H$ avec h, $h'\in H$ et a, $a'\in\mathscr{C}$. Si $xH=yH$, on a $x\sigma(x^{-1})=y\sigma(y^{-1})$. Ici cela donne: $hwa^2\sigma(w^{-1})\ h^{-1}=h'w(a')^2\sigma(w^{-1})\ (h')^{-1}$. Mais $\sigma(w)^{-1}w=t_w$. D'où:

$$hwa^2t_w\,w^{-1}h^{-1}=h'w(a')^2t_w\,w^{-1}(h')^{-1}.$$

Donc a^2t_w et a'^2t_w sont conjugués sous G. Or ce sont des éléments réguliers du sous-groupe de Cartan TA de G. Donc leurs composantes déployées sont conjuguées par un élément de W. Mais a^2 et a'^2 sont des éléments de $\mathscr{C}$. Donc $a^2=a'^2$, ce qui implique $a=a'$. Alors $h'^{-1}h$ centralise l'élément régulier de TA, $wa^2t_w\,w^{-1}$. Donc $h'^{-1}h$ centralise TA. C'est donc un élément de TA et même de T puisque $TA\cap H=T$. D'où $hT=h'T$ et l'injectivité de F_w est prouvée. Etudions maintenant la régularité de F_w. Pour cela, si $h\in H$, $a\in\mathscr{C}$, on identifie l'espace tangent à G/H en $hwaH$ à $\mathfrak{q}$ en associant à $X\in\mathfrak{q}$ le vecteur tangent en $t=0$ à la courbe dans G/H, $t\to(hwa\exp tX)\,H$. De façon similaire, l'espace tangent en (hT, a) à $H/T\times\mathscr{C}$ s'identifie à $\mathfrak{t}^{\perp}\times\mathfrak{a}$ où $\mathfrak{t}^{\perp}$ désigne l'orthogonal dans $\mathfrak{h}$ de $\mathfrak{t}$ (pour $\mathscr{B}$). Pour chaque $\alpha\in\Delta^+$, on choisit $X_\alpha\in\mathfrak{g}^{\alpha}$ avec $\|X_\alpha\|=\dfrac{1}{\sqrt{2}}$ et l'on pose

$$\begin{aligned}Y_\alpha&=(\mathrm{Ad}\,w)(X_\alpha)+\sigma((\mathrm{Ad}\,w)X_\alpha)\\Z_\alpha&=i(\mathrm{Ad}\,w)(X_\alpha)+\sigma(i(\mathrm{Ad}\,w)(X_\alpha))\end{aligned}$$

Alors $(Y_\alpha)_{\alpha\in\Delta^+}$, $(Z_\alpha)_{\alpha\in\Delta^+}$ est une base orthonormée de $\mathfrak{t}^{\perp}$. Comme $\sigma(w)=wt_w$ (cf. Lemme 1), on a (en notant wX_α au lieu de $(\mathrm{Ad}\,w)(X_\alpha)$, etc. ...):

$$\begin{aligned}Y_\alpha&=wX_\alpha+t_w^{\alpha}w\sigma(X_\alpha)\\Z_\alpha&=i(wX_\alpha-t_w^{\alpha}w\sigma(X_\alpha)).\end{aligned}$$

Des égalités pour $X\in\mathfrak{t}^{\perp}$:

$$\begin{aligned}h(\exp X)wa&=hwa(a^{-1}w^{-1}\exp Xwa)\\&=hwa\exp(\mathrm{Ad}(a^{-1}w^{-1})X)\end{aligned}$$

on déduit, en désignant par $P_{\mathfrak{q}}$ la projection orthogonale de $\mathfrak{g}$ sur $\mathfrak{q}$,

$$(dF_w)_{(hT,a)}(X,0)=P_{\mathfrak{q}}(\mathrm{Ad}(a^{-1}w^{-1})X).$$

De même on voit que:

$$\forall Y\in\mathfrak{a},\qquad (dF_w)_{(hT,a)}(0,Y)=Y.$$

Calculons $(dF_w)_{(hT,a)}(Y_\alpha,0)=Y'_\alpha$.
On a:

$$\begin{aligned}Y'_\alpha&=P_{\mathfrak{q}}(a^{-\alpha}X_\alpha+t_w^\alpha a^\alpha\sigma(X_\alpha))\\&=\tfrac{1}{2}(a^{-\alpha}X_\alpha+t_w^\alpha a^\alpha\sigma(X_\alpha)-\sigma(a^{-\alpha}X_\alpha+t_w^\alpha a^\alpha\sigma(X_\alpha)))\\&=\tfrac{1}{2}(a^{-\alpha}X_\alpha+t_w^\alpha a^\alpha\sigma(X_\alpha)-a^{-\alpha}\sigma(X_\alpha)-t_w^\alpha a^\alpha X_\alpha))\\&=\tfrac{1}{2}(a^{-\alpha}-a^\alpha t_w^\alpha)(X_\alpha-\sigma(X_\alpha)).\end{aligned}$$

Soit

$$Y'_\alpha=\tfrac{1}{2}(a^{-\alpha}-a^\alpha t_w^\alpha)U_\alpha\quad\text{où}\quad U_\alpha=X_\alpha-\sigma(X_\alpha).$$

De même on calcule $Z'_\alpha=(dF_w)_{(hT,a)}(Z_\alpha,0)$ et l'on a:

$$Z'_\alpha=\tfrac{1}{2}(a^{-\alpha}-a^\alpha t_w^\alpha)V_\alpha\quad\text{où}\quad V_\alpha=i(X_\alpha+\sigma(X_\alpha)).$$

Si $H_1,\ldots,H_\ell$ est une base orthonormée de $\mathfrak{a}$, on voit que la famille des H_i, U_α, V_α forme une base orthonormée de $\mathfrak{q}$. La régularité de F_w en résulte. Par ailleurs, la mesure sur G/H, (resp. H/T, A) est donnée par une forme différentielle ω (resp. ω_1, ω_2) sur G/H, (resp. H/T, A) qui vaut 1 (resp. $(\mathrm{vol}\,T)^{-1}$, 1) sur une base orthonormée de $T_{hwaH}(G/H)$ identifiée à $\mathfrak{q}$ comme ci-dessus (resp. $T_{hT}(H/T)$ identifie à $\mathfrak{t}^\perp$ comme ci-dessus, resp. $T_a(A)$ identifiée à $\mathfrak{a}$).

Expliquons le facteur $(\mathrm{vol}\,T)^{-1}$. Il vient de l'utilisation de la mesure normalisée sur T pour la normalisation de la mesure sur H/T. Le calcul de la différentielle de F_w en (hT,a) montre que l'image réciproque de ω, forme sur $\Omega_w(C)$, par F_w, $F_w^*\,\omega$ vérifie l'égalité:

$$F_w^*\,\omega=\mathrm{vol}\,T\prod_{\alpha\in\Delta^+}\left|\frac{a^{-\alpha}-a^\alpha t_w^\alpha}{2}\right|^2\omega_1\wedge\omega_2.$$

D'où la formule intégrale. On obtient les autres expressions de D_w en remarquant que:

$$|a^{-\alpha}-a^\alpha t_w^\alpha|=|t_w^{-\alpha}|\,|a^{-\alpha}-t_w^\alpha a^\alpha|$$

puisque $t_w^\alpha=\pm1$. Donc

$|a^{-\alpha}-a^\alpha t_w^\alpha|=|a^\alpha-a^{-\alpha}t_w^{-\alpha}|$. □

3 Série principale et vecteurs distributions H-sphériques

3.1

Soit $\lambda\in\mathfrak{a}'_{\mathbb{C}}$. On considère $(\pi_\lambda^P, I_\lambda^P)$ la représentation de la série principale sphérique de G induite C^∞ du caractère de P défini par $tan\to a^\lambda$. Plus précisément $I_\lambda=\{f:$

$G \to \mathbb{C} | f$ est C^∞ et $f(x\,tan) = a^{-\lambda-\rho} f(x)$, $\forall x \in G$, $t \in T$, $a \in A$, $n \in N_P\}$ et G agit par représentation régulière gauche.

On omettra souvent les indices P lorsqu'aucune confusion ne sera à craindre. On notera $I = C^\infty(K/T)$ muni de sa topologie usuelle. Si $f \in I_\lambda$ on notera $\tilde{f}$ sa restriction à K qui est invariante à droite par T. On regardera $\tilde{f}$ comme élément de I. Alors $f \to \tilde{f}$ définit une bijection de I_λ sur I et on munit I_λ de la topologie de I transportée par cet isomorphisme. Si $f \in I$ on notera f_λ l'élément de I_λ tel que $\tilde{f}_\lambda = f$.

On notera $\tilde{\pi}_\lambda$ la représentation de G sur I définie par transport de structure de π_λ par l'isomorphisme $f \to \tilde{f}$.

Grâce à la mesure invariante sur K/T, le dual $I' = \mathscr{D}'(K/T)$ contient $I = C^\infty(K/T)$ et même $C(K/T)$. De même le dual I'_λ de I_λ contient $I_{-\lambda}$ et même $C_{-\lambda} = \{f : G \to \mathbb{C} | f$ continue, $f(g\,tan) = a^{\lambda-\rho} f(g)\}$ sur lequel agit G par représentation régulière gauche.

De plus π'_λ laisse stable $I_{-\lambda}$ et $C_{-\lambda}$ et coïncide sur $I_{-\lambda}$ avec π_λ et sur $C_{-\lambda}$ avec l'action régulière gauche. De même $\tilde{\pi}'_\lambda$ coïncide sur I avec $\tilde{\pi}_{-\lambda}$.

Le transposé de l'opérateur $I_\lambda \to I$ (resp. $I \to I_\lambda$) défini par $f \to \tilde{f}$ (resp. $f \to f_\lambda$) sera noté $T \in I' \to T_{-\lambda} \in I'_\lambda$ (resp. $T \in I'_\lambda \to \tilde{T} \in I'$), notations qui sont compatibles avec les identifications et notations antérieures.

On a la propriété de régularisation suivante: Si $\varphi \in C_c^\infty(G)$ et $T \in I'_\lambda$ on a:

$$\pi'_\lambda(\varphi)\, T \in I_{-\lambda}.$$

Remarque 2 On sait que π_λ est à croissance modérée au sens de [5], comme espace des vecteurs C^∞ d'une représentation Banachique. Alors si l'on note $(I_\lambda)_{(K)}$ le (g, K)-module des vecteurs K-finis de I_λ, tout opérateur d'entrelacement de (g, K)-module entre $(I_\lambda)_{(K)}$ et $(I_{\lambda'})_{(K)}$ se laisse prolonger en un opérateur d'entrelacement continu de G-modules entre I_λ et $I_{\lambda'}$. En particulier lorsque le prolongement méromorphe des intégrales d'entrelacement est défini, il définit un opérateur continu entre I_λ et $I_{w\lambda}$ et on peut considérer son transposé (voi plus loin).

3.2

On écrit $\lambda \in \mathfrak{a}'_{\mathbb{C}}$ dans la base de $\mathfrak{a}'_{\mathbb{C}}$ formée des restrictions $\bar{\delta}_i$ des δ_i à $\mathfrak{a}$:

$$\lambda = 2 \sum_{i=1}^{\ell} \lambda_i \bar{\delta}_i.$$

On a
$$\rho_{\mathbb{C}} = \sum_{i=1}^{\ell} \delta_i \quad \text{et} \quad \rho = 2 \sum_{i=1}^{\ell} \bar{\delta}_i.$$

On suppose $\lambda - \rho$ strictement dominant i.e. $\mathrm{Re}(\lambda_i - 1) > 0$ pour tout i. On introduit alors des fonctions $\xi_\lambda^{w,P}$, $w \in W$ définies par

$$\xi_\lambda^{w,P}(g) = \prod_{i=1}^{\ell} |\varepsilon_i(g)|^{\lambda_i - 1} \quad \text{si } g \in HwP$$

$$= 0 \quad \text{si } g \notin HwP.$$

Alors il résulte de la Proposition 1 que les fonctions $\xi_\lambda^{w,P}$ sont continues sur G et d'après les propriétés des ε_i on a:

$$\xi_\lambda^{w,P} \in C_{-\lambda} \subset I'_\lambda.$$

C'est un élément H-invariant de I'_λ sous π'_λ qui est propre sous l'action de $U(\mathfrak{g})^{\mathfrak{h}}$, centralisateur dans $U(\mathfrak{g})$ de $\mathfrak{h}$ (cf. [13, Th. 5-1]). Sa restriction à K (ou K/T) est une distribution $\tilde{\xi}_\lambda^{w,P}$. La fonction $\lambda \to \tilde{\xi}_\lambda^{w,P}$ définie pour $\lambda \in \mathfrak{a}'_{\mathbb{C}}$ avec $\lambda - \rho$ strictement Δ^+-dominant et à valeurs dans $\mathscr{D}'(K/T) = I'$ s'étend en une fonction méromorphe cf. [1] ou [13]. Ici on peut même utiliser [2] pour voir que les singularités de $\tilde{\xi}_\lambda^{w,P}$ sont situées sur une famille localement finie d'hyperplans de $\mathfrak{a}'_{\mathbb{C}}$ (de complémentaire connexe).

On notera encore de la même façon les prolongements méromorphes des $\tilde{\xi}_\lambda^{w,P}$ et $\xi_\lambda^{w,P}$ désignera (lorsque $\tilde{\xi}_\lambda^{w,P}$ est défini) l'élément de I'_λ tel que

$$(\xi_\lambda^{w,P})^\sim = \tilde{\xi}_\lambda^{w,P}.$$

Lorsque $\xi_\lambda^{w,P}$ et $\xi_{-\lambda}^{w,P}$ sont définis, pour tout $\varphi \in C_c^\infty(G/H)$ on a $\pi'_\lambda(\varphi)\ \xi_\lambda^{w,P} \in I_{-\lambda}$ et l'expression

$$\langle \pi'_\lambda(\varphi)\ \xi_\lambda^{w,P}, \xi_{-\lambda}^{w,P} \rangle_{I_{-\lambda}, I'_{-\lambda}}$$

a un sens et permet de définir une distribution $\Theta_\lambda^{w,P}$ que nous proposons d'étudier. On notera ξ_λ ou ξ_λ^P pour $\xi_\lambda^{1,P}$ et Θ_λ ou Θ_λ^P pour $\Theta_\lambda^{1,P}$.

4 Etude du comportement asymptotique des distributions Θ_λ

4.1 Définition de fonctions sur G/H à support dans $\Omega_w(C)$

On se donne une chambre de Weyl de $\mathfrak{a}$ relativement à Δ. On reprend les notations de la Proposition 3. On se donne $\psi \in C_c^\infty(H/T)$ et $\varphi \in C_c^\infty(\mathscr{C})$. On définit une fonction $\Phi = \Phi^w(\psi, \varphi)$ sur G/H par:

$$\Phi(gH) = 0 \quad \text{si } gH \notin \Omega_w(C)$$
$$\Phi(hwaH) = \psi(h)\,\varphi(a) \quad \text{si } h \in H, \quad a \in A.$$

On se fixe maintenant $X_0 \in -C$ et on note $a_t = \exp tX_0$, $t > 0$. On définit pour $t > 0$, $\varphi_t \in C_c^\infty(\mathscr{C})$ par:

$$\varphi_t(a) = 0 \quad \text{si } a \in \mathscr{C} \quad \text{et} \quad aa_t \notin \mathscr{C}$$
$$\varphi_t(a) = \varphi(aa_t) \quad \text{si } a \in \mathscr{C} \quad \text{et} \quad aa_t \in \mathscr{C}.$$

On pose: $\Phi_t^w = \Phi^w(\psi, \varphi_t)$.

4.2

Soit $x \in W$ et $i \in \{1, \ldots, \ell\}$. Notons $e_{x\delta_i}$ un vecteur dans l'espace de π_i de poids $x\delta_i$ sous $\mathfrak{t}_{\mathbb{C}}$ et de norme 1, et $e'_{x\delta_i}$ un vecteur dans l'espace de π'_i de poids $-x\delta_i$

tel que $\langle e_{x\delta_i}, e'_{x\delta_i}\rangle = 1$. Alors $e_{x\delta_i}\otimes e'_{x\delta_i}$ est indépendant du choix de $e_{x\delta_i}$. On définit une fonction C^∞ sur G par:

$$\eta_{i,x}(g) = ((\pi_i(g)\otimes\pi'_i(\sigma(g)))\,(e_{\delta_i}\otimes e'_{\delta_i}), e_{x\delta_i}\otimes e'_{x\delta_i}).$$

Lemme 2 *(i)* $\forall g\in G, a\in A, n\in N, t, t'\in T, \eta_{i,x}(t'gtan) = a^{2\delta_i}\eta_{i,x}(g)$
(ii) $\forall k\in K, \eta_{i,x}(k) = t_x^{\delta_i}|(\pi_i(k)\, e_{\delta_i}, e_{x\delta_i})|^2$
(ii) $\forall g\in G, \eta_{i,x}(g) = t_x^{\delta_i}|(\pi_i(g)\, e_{\delta_i}, e_{x\delta_i})|^2$.

Démonstration. (i) résulte immédiatement de la définition. Identifions l'espace V'_i de π'_i avec l'espace de π_i par l'application antilinéaire $e\to e'$ de V_i dans V'_i définie par $e'(v) = (v, e)$. La notation $(e_i^j)'$ est cohérente avec cette définition. Dans cette correspondance on a:

$$\forall k\in K, \forall e, f\in V_i, (e', f') = (\overline{e, f}) \quad \text{et} \quad \pi'_i(k)\, e' = (\pi_i(k)\, e)'.$$

Donc: $(\pi'_i(k)\, e', f') = (\overline{\pi_i(k)\, e, f})$.

Par ailleurs, comme σ agit trivialement sur T on voit que π'_i et $\pi'_i\circ\sigma$ définissent des représentations équivalentes (et irréductibles) de K. Il existe donc un unique opérateur unitaire $T: V'_i\to V'_i$ tel que T entrelace π'_i et $\pi'_i\circ\sigma$ restreintes à K et tel que:

$$Te'_{\delta_i} = e'_{\delta_i}.$$

Supposons (ce qui est possible) que $e_{x\delta_i} = \pi_i(x)\, e_{\delta_i}$.

Alors $e'_{x\delta_i} = \pi'_i(x)\, e'_{\delta_i}$ et:

$$\begin{aligned} Te'_{x\delta_i} &= T\pi'_i(x)\, e'_{\delta_i} \\ &= \pi'_i(\sigma(x))\, Te'_{\delta_i} \\ &= \pi'_i(x\, t_x)\, e'_{\delta_i}, \end{aligned}$$

puisque $\sigma(x) = x\, t_x$ (cf. Lemme 1) et $Te'_{\delta_i} = e'_{\delta_i}$.

D'où

$$\begin{aligned} Te'_{x\delta_i} &= t_x^{-\delta_i}\,\pi'_i(x)\, e'_{\delta_i} \\ &= t_x^{\delta_i}(\pi_i(x)\, e_{\delta_i})', \qquad \text{car } t_x^2 = 1, \\ &= t_x^{\delta_i}\, e'_{x\delta_i}. \end{aligned}$$

Alors

$$\begin{aligned} (\pi'_i(\sigma(k))\, e'_{\delta_i}, e'_{x\delta_i}) &= ((\pi'_i(\sigma(k))\,(Te'_{\delta_i}), e'_{x\delta_i}) \\ &= (T\pi'_i(k)\, e'_{\delta_i}, e'_{x\delta_i}) \\ &= (\pi'_i(k)\, e'_{\delta_i}, T^{-1}\, e'_{x\delta_i}) \\ &= (\pi'_i(k)\, e'_{\delta_i}, t_x^{-\delta_i}\, e'_{x\delta_i}) \\ &= t_x^{\delta_i}(\pi'_i(k)\, e'_{\delta_i}, e'_{x\delta_i}) \\ &= t_x^{\delta_i}(\overline{\pi_i(k)\, e_{\delta_i}, e_{x\delta_i}}). \end{aligned}$$

D'où $\eta_{i,x}(k) = t_x^{\delta_i}|(\pi_i(k)\, e_{\delta_i}, e_{x\delta_i})|^2$ ce qui prouve (ii). Alors (iii) résulte de (i), (ii) et de la décomposition d'Iwasawa de G. □

Donc si $x \in W_H$, $\eta_{i,x}$ est positive ou nulle sur G et si $\lambda - \rho$ est strictement Δ^+-dominant on peut définir

$$\eta_{\lambda,x} = \prod_{i=1}^{\ell} \eta_{i,x}^{\lambda_i - 1}$$

qui est continue sur G et élément de $C_{-\lambda} \subset I'_\lambda$.

4.3

Théorème 1 *On se fixe w, $x \in W$. On suppose que $C = -x(C^+)$ où C^+ est la chambre de Weyl définie par Δ_P^+. Donc $X_0 \in x(C^+)$. Soient $\psi \in C_c^\infty(H/T)$, $\varphi \in C_c^\infty(\mathscr{C})$. On suppose toujours que $\lambda - \rho$ est strictement Δ^+-dominant.*

(i) si $x \notin W_H$ on a:

$$\lim_{t \to +\infty} a_t^{-x(\lambda+\rho)} \pi'_\lambda(\Phi_t^w(\psi, \varphi))\, \xi_\lambda = 0.$$

(ii) Si $x \in W_H$, l'élément de I'_λ, $u^w_{\lambda,x}$ défini par

$$u^w_{\lambda,x} = \int_{H/T} \int_{\mathscr{C}} \psi(hT)\, \varphi(a)\, a^{-2x\rho}\, \pi'_\lambda(hwa)\, \eta_{\lambda,x}\, dh\, da$$

est en fait dans $I_{-\lambda}$ et

$$\lim_{t \to +\infty} a_t^{-x(\lambda+\rho)} \pi'_\lambda(\Phi_t^w(\psi, \varphi))\, \xi_\lambda = C_G\, u^w_{\lambda,x}.$$

Dans (i) et (ii) les convergences ont lieu dans $I_{-\lambda}$.

4.4

Remarquons qu'il suffit pour démontrer le théorème de s'intéresser aux restrictions à K des éléments de $C_{-\lambda}$ considérés, d'après la définition de la topologie sur $I_{-\lambda}$. Nous utiliserons, entre autres, le lemme suivant:

Lemme 3 *Soit $(f_t)_{t \in \mathbb{R}^+}$ une famille d'éléments de $C^\infty(K)$ telle que*

(i) $\forall D \in U(\mathfrak{k})$, $L_D f_t$ admet une limite simple lorsque t tend vers $+\infty$.

(ii) $\forall D \in U(\mathfrak{k})$, $\exists C_D > 0$, $\forall k \in K$, $\forall t \in \mathbb{R}^+$, $|L_D f_t(k)| \leqq C_D$.

Alors f_t converge uniformément vers $f \in C^\infty(K)$ et pour tout $D \in U(\mathfrak{k})$, $(L_D f_t)$ converge uniformément vers $L_D f$ lorsque t tend vers $+\infty$.

Démonstration. D'après [15, Lemme 5.1], il existe $D_1, \ldots, D_n \in U(\mathfrak{k})$ et $C > 0$ tels que

$$\forall f \in C^\infty(K), \quad |f(e)| \leqq C \sum_{i=1}^{n} \int_K |(L_{D_i} f)(k)|\, dk$$

et en appliquant cette inégalité aux fonctions $R_y f$, $y \in K$ et l'invariance de la mesure dk, on a

$$\forall f \in C^\infty(K), \quad \forall g \in K, \quad |f(g)| \leqq C \sum_{i=1}^{n} \int_K |L_{D_i} f(k)|\, dk.$$

Le théorème de convergence dominée de Lebesgue et nos hypothèses impliquent que les fonctions $L_{D_i} f_t$ convergent dans $L^1(K)$ lorsque t tend vers $+\infty$. L'inégalité ci-dessus implique la convergence uniforme de (f_t). On procède de même pour les dérivées des f_t. □

4.5

On va poser $f_t = (a_t^{-x(\lambda+\rho)}\, \pi'_\lambda(\Phi_t^w)\, \xi_\lambda)^\sim \in I$.

Pour pouvoir appliquer ce lemme, on va étudier $L_D f_t$ pour $D \in U(\mathfrak{t})$. Posons

$$F_t = \pi'_\lambda(\Phi_t^w)\, \xi_\lambda \in I_{-\lambda}.$$

Alors

$$\begin{aligned} L_D F_t &= \pi_{-\lambda}(D)\, F_t \\ &= \pi'_\lambda(D)\, \pi'_\lambda(\Phi_t^w)\, \xi_\lambda \\ &= \pi'_\lambda(L_D \Phi_t^w)\, \xi_\lambda. \end{aligned}$$

Pour cette raison on s'intéresse à $L_D \Phi_t^w(\psi, \varphi)$. On introduit $\mathscr{R}_x$ l'algèbre à élément unité des fonctions sur A^{reg} (ensemble des éléments réguliers de A), engendrée par les fonctions $\dfrac{1}{1-ua^{2\alpha}}$, a^α où $\alpha \in x(\Delta^+)$ et $u \in \mathbb{C}$ avec $|u| = 1$.

Proposition 4 *Soit $D \in U(\mathfrak{g})$. Alors il existe $p \in \mathbb{N}^*$,*

$$U_1, \ldots, U_p \in U(\mathfrak{h}), \qquad Y_1, \ldots, Y_p \in U(\mathfrak{a}),$$
$$\psi_1, \ldots, \psi_p \in C^\infty(H/T), \qquad \varphi_1, \ldots, \varphi_p \in \mathscr{R}_x$$

tels que:

$$\forall \psi \in C_c^\infty(H/T), \qquad \forall \varphi \in C_c^\infty(\mathscr{C}),$$

$$L_D \Phi^w(\psi, \varphi) = \sum_{i=1}^{p} \Phi^w(\psi_i(L_{U_i} \psi), \quad \varphi_i(L_{Y_i} \varphi)).$$

Pour démontrer cette proposition nous allons établir plusieurs lemmes.

Lemme 4 *Si $Y \in U(\mathfrak{a})$ et $\varphi_1 \in \mathscr{R}_x$, alors $L_Y \varphi_1 \in \mathscr{R}_x$.*

Démonstration. Si $Y \in \mathfrak{a}$, $\alpha \in \Delta^+$, $L_Y(a^\alpha) = -\alpha(Y)\, a^\alpha \in \mathscr{R}_x$

$$L_Y\left(\frac{1}{1-ua^{2\alpha}}\right) = (L_Y a^{2\alpha}) \frac{u}{(1-ua^{2\alpha})^2} \in \mathscr{R}_x.$$

Une récurrence sur le degré de $Y \in U(\mathfrak{a})$ donne le résultat voulu. □

Lemme 5 *Si* $X\in\mathfrak{g}$, *il existe* $V_1, \ldots, V_r\in\mathfrak{h}_w=\mathrm{Ad}\,w^{-1}(\mathfrak{h})$, $Y\in\mathfrak{a}$, $Z_1, \ldots, Z_s\in\mathfrak{h}$ *et* $\tau_1, \ldots, \tau_r, \nu_1, \ldots, \nu_s\in\mathscr{R}_x$ *tels que*:

$$\forall a\in A^{\mathrm{reg}},\qquad X=\sum_{i=1}^{r}\tau_i(a)\,V_i+Y+\sum_{j=1}^{s}\nu_j(a)(\mathrm{Ad}\,a)(Z_i).$$

Démonstration. Il suffit de prouver le lemme pour des éléments d'une base de $\mathfrak{g}$. Si $X\in\mathfrak{h}_w$ ou $X\in\mathfrak{a}$ le lemme est évident. Notons $\mathfrak{n}_x=\mathrm{Ad}\,x(\mathfrak{n})$. Alors comme

$$\mathfrak{g}=\mathfrak{h}\oplus\mathfrak{a}\oplus\mathfrak{n}$$

on a aussi

$$\mathfrak{g}=\mathfrak{h}_w\oplus\mathfrak{a}\oplus\mathfrak{n}_x$$

(en changeant de forme réelle et de parabolique minimal).

Il reste donc à prouver le lemme pour $X\in\mathfrak{g}^\alpha$ avec $\alpha\in x(\Delta^+)$. On note $\sigma'=\mathrm{Ad}\,w^{-1}\circ\sigma\circ\mathrm{Ad}\,w$. C'est une involution antilinéaire de $\mathfrak{g}$ dont l'ensemble des points fixes est $\mathfrak{h}_w$. Il est clair que $\sigma\sigma'(\mathfrak{g}^\alpha)=\mathfrak{g}^\alpha$. Comme $\sigma\sigma'$ est $\mathbb{C}$-*linéaire*, $\sigma\sigma'$ agit sur $\mathfrak{g}^\alpha$ par un scalaire u_α. Mais, σ,σ' sont des transformations orthogonales de $\mathfrak{g}$ (relativement à $\mathscr{B}$) donc

$$u_\alpha=\pm 1$$

et si $X\in\mathfrak{g}^\alpha$ on a:

$$\sigma'(X)=u_\alpha\,\sigma(X).$$

Posons

$$Z=X+\sigma(X)\in\mathfrak{h}$$

et

$$V=X+\sigma'(X)\in\mathfrak{h}_w.$$

Alors $\sigma(X)$ et $\sigma'(X)=u_\alpha\,\sigma(X)$ sont de poids $-\alpha$ sous $\mathfrak{a}$ et l'on a:

$$(\mathrm{Ad}\,a)(Z)=a^\alpha X+a^{-\alpha}\sigma(X).$$

Or

$$V=X+u_\alpha\,\sigma(X).$$

D'où

$$X=-\frac{u_\alpha\,a^{-\alpha}}{a^\alpha-u_\alpha\,a^{-\alpha}}\,V+\left(\frac{1}{a^\alpha-u_\alpha\,a^{-\alpha}}\right)(\mathrm{Ad}\,a)(Z)$$

i.e.

$$X=\frac{V}{(1-u_\alpha\,a^{2\alpha})}-\frac{u_\alpha\,a^\alpha}{(1-u_\alpha\,a^{2\alpha})}(\mathrm{Ad}\,a)(Z)$$

et X a la forme voulue. □

Démonstration de la Proposition 4. Un argument de récurrence sur le degré D et le Lemme 3, montre qu'il suffit de démontrer la proposition pour $D\in\mathfrak{g}$. Soit donc $X\in\mathfrak{g}$. D'abord $L_X\,\Phi^w(\psi,\varphi)$ est clairement nulle en dehors de $\Omega_w(C)$. Maintenant, en utilisant une base (X_ℓ) de $\mathfrak{g}$, $\mathrm{Ad}(hw)^{-1}X=\sum_{\ell=1}^{\dim\mathfrak{g}}\nu_\ell(h)\,X_\ell$ où les ν_ℓ sont des éléments de $C^\infty(H)$. En appliquant le lemme précédent aux X_l on en déduit

l'existence de $V_i \in \mathfrak{h}_w$, $Y_j \in \mathfrak{a}$, $Z_k \in \mathfrak{h}$, ω_i, ω'_j, $\omega''_k \in C^\infty(H)$, φ_i, φ'_j, $\varphi''_k \in \mathscr{R}_x$ tels que: $\forall a \in A^{\mathrm{reg}}, \forall h \in H$

$$\mathrm{Ad}(hw)^{-1} X = \sum_{i=1}^{r} \omega_i(h)\, \varphi_i(a)\, V_i + \sum_{j=1}^{s} \omega'_j(h)\, \varphi'_j(a)\, Y_j + \sum_{k=1}^{t} \omega''_k(h)\, \varphi''_k(a)\, (\mathrm{Ad}\, a)(Z_k).$$

En tenant compte de $\exp(-tX)\, hwa = hw(\exp -t\mathrm{Ad}(hw)^{-1} X)\, a$ il vient facilement, pour $h \in H$ et $a \in \mathscr{C}$

$$(*) \qquad (L_X \Phi^w(\psi, \varphi))(hwaH) = -\sum_{i=1}^{r} \omega_i(h)\, \varphi_i(a)\, (R_{U_i} \psi)(h)\, \varphi(a) + \sum_{j=1}^{s} \omega'_j(h)\, \varphi'_j(a)\, (L_{Y_j} \varphi)(a)$$

où $U_i = (\mathrm{Ad}\, w)(V_i)$.

En effet, si $V \in \mathfrak{h}_w$, $U = (\mathrm{Ad}\, w)(V) \in \mathfrak{h}$ et si $a \in \mathscr{C}$ on a:

$$\frac{d}{dt} \Phi^w(\psi, \varphi)(hw(\exp tV)\, aH)_{|t=0} = \frac{d}{dt} \Phi^w(\psi, \varphi)(h(\exp tU)\, waH)_{|t=0} = (R_U \psi)(h)\, \varphi(a).$$

De même, si $Y \in \mathfrak{a}$,

$$\frac{d}{dt} \Phi^w(\psi, \varphi)(hw(\exp tY)\, aH)_{|t=0} = \psi(h)(-L_Y \varphi)(a)$$

et si $Z \in \mathfrak{h}$

$$\frac{d}{dt} \Phi^w(\psi, \varphi)(hw(\exp(t(\mathrm{Ad}\, a)(Z))\, aH)_{|t=0} = 0.$$

Mais pour $U \in \mathfrak{h}$, en utilisant une base (T_k) de $\mathfrak{h}$, il existe des fonctions $\theta_k \in C^\infty(H)$ telles que

$$(R_U \psi)(h) = \sum_{k=1}^{\dim \mathfrak{h}} \theta_k(h)\, (L_{T_k} \psi)(h).$$

En introduisant ces formules pour les $R_{U_i} \psi$ dans (*) on arrive à une expression de $L_X \Phi(\psi, \varphi)$ qui est du type de celui de la proposition, à ceci près que les fonctions ψ_i sont dans $C^\infty(H)$ et non dans $C^\infty(H/T)$ i.e.

$$(**) \qquad L_X \Phi^w(\psi, \varphi)(hwaH) = \sum_{i=1}^{p} \psi_i(h)\, (L_{U_i} \psi)(h)\, \varphi_i(a)\, (L_{Y_i} \varphi)(a)$$

où $U_i \in U(\mathfrak{h})$, $Y_i \in U(\mathfrak{a})$, $\varphi_i \in \mathscr{R}_x$, $\varphi_i \in C^\infty(H)$.

Mais $\Phi^w(\psi, \varphi)(htwaH) = \Phi^w(\psi, \varphi)(hwaH)$ pour $t \in T$, et $\psi(ht) = \psi(h)$. On intègre alors les deux membres de l'égalité (**) sur $t \in T$ (en ayant remplacé h par ht) et on obtient l'expression voulue. □

4.6

On revient à l'étude de $L_D f_t$, $D \in U(\mathfrak{f})$ (c.f. début de 4.5). D'après la Proposition 4, calculant $L_D \Phi^w(\psi, \varphi)$, et la formule intégrale de la Proposition 3, on a:

$$(L_D f_t)(k) = C_G\, a_t^{-x(\lambda+\rho)} \sum_{i=1}^{p} \int_{H/T} \int_{\mathscr{C}} \psi_i(hT)\,(L_{D_i}\psi)(hT) \cdot \varphi_i(a)\,(L_{Y_i}\varphi_t(a))\, D_w(a)\,(\pi'_\lambda(hwa)\,\xi_\lambda)(k)\,da\,d\dot{h}$$

où $D_i \in U(\mathfrak{h})$, $Y_i \in U(\mathfrak{a})$, $\psi_i \in C^\infty(H/T)$, $\varphi_i \in \mathscr{R}_x$.

Mais $L_{Y_i}\varphi_t = (L_{Y_i}\varphi)_t$. Or il est clair que si $\varphi \in C_c^\infty(\mathscr{C})$ et $v \in C(\mathscr{C})$ on a pour tout $t>0$

$$\int_{\mathscr{C}} \varphi_t(a)\, v(a)\, da = \int_{\mathscr{C}} \varphi(a)\, v(aa_{-t})\, da$$

(notez que $a \in \mathscr{C}$ et $t>0$ impliquent $aa_{-t} \in \mathscr{C}$). D'où

$$(L_D f_t)(k) = C_G\, a_t^{-x(\lambda+\rho)} \sum_{i=1}^{p} \int_{H/T} \int_{\mathscr{C}} (\psi_i L_{D_i}\psi)(hT)\,\varphi_i(aa_{-t})\,(L_{Y_i}\varphi)(a) \cdot D_w(aa_{-t})\,(\pi'_\lambda(hwa)\,\pi'_\lambda(a_{-t})\,\xi_\lambda)(k)\,da\,d\dot{h}$$

formule que l'on préfère réécrire sous la forme

$$(L_D f_t)(k) = C_G \sum_{i=1}^{p} \int_{H/T} \int_{\mathscr{C}} (\psi_i L_{D_i}\psi)(hT)\,\varphi_i(aa_{-t})\,(L_{Y_i}\varphi)(a) \cdot a_t^{-2x\rho}\, D_w(aa_{-t})\,(\pi'_\lambda(hwa)\, a_t^{-x(\lambda-\rho)}\,\pi'_\lambda(a_{-t})\,\xi_\lambda)(k)\,da\,d\dot{h}.$$

On considère $(L_D f_t)(k)$ comme somme d'intégrales sur $H/T \times \mathscr{C}$ dépendant du paramètre t auxquelles on va appliquer le théorème de convergence dominée pour voir que les hypothèses du Lemme 3 sont vérifiees.

Lemme 5 *On conserve les hypothèses du Théorème 1.*

(i) $\exists C_1>0, \forall k \in K, \forall t>0, |a_t^{-x(\lambda-\rho)}(\pi'_\lambda(a_{-t})\,\xi_\lambda)(k)| \leqq C_1$

(ii) $\forall k \in K, \lim_{t\to+\infty} a_t^{-x(\lambda-\rho)}(\pi'_\lambda(a_{-t})\,\xi_\lambda)(k) = \begin{cases} 0 & \text{si } x \notin W_H \\ \eta_{\lambda,x}(k) & \text{si } x \in W_H. \end{cases}$

Démonstration. On étudie d'abord $\varepsilon_i(a_t\, k)$:

$$\varepsilon_i(a_t\, k) = ((\pi_i(a_t\, k) \otimes \pi'_i(\sigma(a_t\, k)))\,(e_{\delta_i} \otimes e'_{\delta_i}), \sum_{j=0}^{n_i} e_i^j \otimes (e_i^j)').$$

En utilisant le fait que $\theta(a_{-t}) = a_t$, $\sigma(a_t) = a_{-t}$, que e_i^j (resp. $(e_i^j)'$) est de poids μ_j (resp. $-\mu_j$) sous $\mathfrak{t}_{\mathbb{C}}$ et l'invariance des produits scalaires sous K, on en déduit

$$\varepsilon_i(a_t\, k) = \sum_{j=0}^{n_i} a_t^{2\mu_j}(\pi_i(k)\, e_{\delta_i}, e_i^j)\,(\pi'_i(\sigma(k))\, e'_{\delta_i}, (e_i^j)').$$

Comme $a_t = \exp tX_0$ avec $X_0 \in x(C^+)$, on a, pour tout j et tout $t>0$: $a_t^{2\mu_j} \leqq a_t^{2x\delta_i}$, puisque π_i est de plus haut poids δ_i.

D'où l'on déduit, en utilisant l'unitarité de π_i et π'_i comme représentations de K:

$$\forall t>0, \quad \forall k\in K, \quad |a_t^{-2x\delta_i}\,\varepsilon_i(a_t k)|\leqq n_i. \tag{4}$$

D'autre part on a:

$$\lim_{t\to+\infty} a_t^{-2x\delta_i}\, a_t^{2\mu_j}=0 \quad \text{sauf si } \mu_j=x\delta_i$$

(ce qui n'arrive que pour un indice j).

Alors:

$$\lim_{t\to+\infty} a_t^{-2x\delta_i}\,\varepsilon_i(a_t k)=(\pi_i(k)\,e_{\delta_i}, e_{x\delta_i})\times(\pi'_i(\sigma(k))\,e'_{\delta_i}, e'_{x\delta_i})$$

i.e.

$$\lim_{t\to+\infty} a_t^{-2x\delta_i}\,\varepsilon_i(a_t k)=\eta_{i,x}(k) \tag{5}$$

(d'après la définition de $\eta_{i,x}$, c.f. 4.2).

Or

$$(\pi'_\lambda(a_{-t})\,\xi_\lambda)(k)=\xi_\lambda(a_t k) \quad (\text{car } \xi_\lambda\in C_{-\lambda})$$

et

$$\xi_\lambda=\prod_{i=1}^{\ell}\operatorname{Sup}(\varepsilon_i,0)^{\lambda_i-1} \tag{6}$$

d'après la définition de ξ_λ et la Proposition 1 (iii).

Alors (4) implique la majoration du lemme en prenant $C_1=\prod_{i=1}^{\ell}(n_i)^{\operatorname{Re}\lambda_i-1}$.

D'autre part (5) implique que pour tout i:

$$\lim_{t\to+\infty} a_t^{-2x\delta_i}\operatorname{Sup}(\varepsilon_i(a_t k),0)=\operatorname{Sup}(\eta_{i,x}(k),0)$$

et (6) donne alors:

$$\lim_{t\to+\infty} a_t^{-x(\lambda-\rho)}\,\xi_\lambda(a_t k)=\prod_{i=1}^{\ell}(\operatorname{Sup}(\eta_{i,x}(k),0))^{\lambda_i-1}.$$

Or, si $x\notin W_H$, l'une des fonctions $\eta_{i,x}$ est toujours négative ou nulle d'après le Lemme 2 (iii) et l'on a bien la limite voulue. Si $x\in W_H$, cette fois, toujours d'après le Lemme 2 (iii), toutes les fonctions $\eta_{i,x}$ sont positives et la limite est bien égale à $\eta_{\lambda,x}(k)$. □

Lemme 6 *Soit U un compact de G. Il existe une constante $C_U>0$ (dépendant de λ) telle que*:

$$\forall f\in C_{-\lambda}, \quad \sup_{(u,k)\in U\times K}|f(uk)|\leqq C_U\sup_{k\in K}|f(k)|.$$

Démonstration. En effet UK est un compact de G, donc contenu dans un sous-ensemble de G de la forme KU_1U_2 avec U_2 compact de A et U_2 compact de

N. Utilisant la propriété $f(kan)=a^{\lambda-\rho}f(k)$, on voit que l'on peut prendre $C_U=\sup_{a\in U_1}|a^{\lambda-\rho}|$ qui est fini puisque U_1 est compact. □

Appliquons les deux lemmes précédents en prenant:

$$U=\{hwa\,|\,hT\in\mathrm{Supp}\,\psi,\ a\in\mathrm{Supp}\,\varphi\}$$

et posant $C_{\varphi,\psi}=C_U\,C_1$ on a

Lemme 7 *(i)* $\exists C_{\varphi,\psi}$, $\forall k\in K$, $t>0$, $h\in\mathrm{Supp}\,\psi$, $a\in\mathrm{Supp}\,\varphi$

$$|a_t^{-x(\lambda-\rho)}(\pi'_\lambda(hwaa_{-t})\,\xi_\lambda)(k)|\leqq C_{\varphi,\psi}$$

(ii) $\forall k\in K$, $t>0$, $h\in\mathrm{Supp}\,\psi$, $a\in\mathrm{Supp}\,\varphi$

$$\lim_{t\to+\infty}a_t^{-x(\lambda-\rho)}(\pi'_\lambda(hwaa_{-t})\,\xi_\lambda)(k)=\begin{cases}0 & \text{si } x\notin W_H\\ (\pi'_\lambda(hwa)\,\eta_{\lambda,x})(k) & \text{si } x\in W_H\end{cases}.$$

Il reste pour pouvoir appliquer le théorème de convergence dominée à contrôler $D_w(aa_{-t})$ et $\varphi_i(aa_{-t})$ pour $t>0$ et $a\in$ Supp φ.

A cette fin on va établir:

Lemme 8 *(i) Soit $v\in\mathscr{R}_x$. La famille de fonctions $a\to v(aa_{-t})$ converge simplement (sur $\mathscr{C}$) quand $t\to+\infty$. De plus, il existe $C_{v,\varphi}$ telle que*

$$\forall a\in\mathrm{Supp}\,\varphi,\qquad \forall t>0,\qquad |v(a_{-t}a)|\leqq C_{v,\varphi}$$

(ii) $\forall a\in\mathscr{C}(=\exp(-x(C^+)))$,

$$\lim_{t\to+\infty}a_t^{-2x\rho}\,D_w(aa_{-t})=a^{-2x\rho}$$

et $\exists C_{w,\varphi}$, $\forall a\in\mathrm{Supp}\,\varphi$, $|a_t^{-2x\rho}\,D_w(aa_{-t})|\leqq C_{w,\varphi}$.

Démonstration. Soit $\alpha\in x\Delta^+$. Alors $(aa_{-t})^\alpha=a^\alpha(\exp-t\alpha(X_0))$. Mais $X_0\in x(C^+)$. D'où $\alpha(X_0)>0$ et $\lim_{t\to+\infty}(aa_{-t})^\alpha=0$. De plus

$$\exists C_{1,\varphi},\qquad C_{2,\varphi}\in]0,1[,\qquad \forall a\in\mathrm{Supp}\,\varphi,$$
$$C_{1,\varphi}\leqq(aa_{-t})^\alpha\leqq C_{2,\varphi}.$$

En effet $a^\alpha_{-t}\leqq 1$ et si $a\in\mathscr{C}=\exp(-x\Delta^+)$ on a $a^\alpha<1$. Par compacité de Supp φ on en déduit l'existence de $C_{1,\varphi}$ et $C_{2,\varphi}$. Ceci implique facilement (i) pour les générateurs de $\mathscr{R}_x$, a^α et $\dfrac{1}{1-ua^{2\alpha}}$, $\alpha\in x\Delta^+$, $|u|=1$, et donc pour tous les éléments de $\mathscr{R}_x$.

Prouvons (ii). D'après la Proposition 3 on a:

$$D_w(a)=\prod_{\alpha\in\Delta}|a^{-\alpha}-a^\alpha t_w^\alpha|$$
$$=\prod_{\alpha\in x\Delta^+}|a^{-\alpha}-a^\alpha t_w^\alpha|^2$$

car $|a^{-\alpha}-a^\alpha t_w^\alpha|=|a^\alpha-a^{-\alpha}t_w^{-\alpha}|$(car $t_w^\alpha=\pm1=t_w^{-\alpha}$).

Donc

$$D_w(a)=\prod_{\alpha\in x\Delta^+} a^{-2\alpha}|1-a^{2\alpha}t_w^{\alpha}|^2$$

et

$$D_w(aa_{-t})=a_{-t}^{-2x\rho}\,a^{-2x\rho}\prod_{\alpha\in x\Delta^+}|1-(aa_{-t})^{2\alpha}t_w^{\alpha}|^2.$$

D'où

$$\lim_{t\to+\infty} a_t^{-2x\rho}\,D_w(aa_{-t})=a^{-2x\rho}$$

(car $\lim_{t\to+\infty}(aa_{-t})^{2\alpha}=0$).

Par ailleurs la majoration de $(aa_{-t})^{\alpha}$, pour $t>0$, $\alpha\in x\Delta^+$ et $a\in\mathrm{Supp}\,\varphi$, donnée ci-dessus permet de majorer $a_t^{-2x\rho}D_w(aa_{-t})$ comme désiré. □

4.7 Fin de la démonstration du Théorème 1

Le Lemme 7 et le Lemme 8 permettent d'appliquer le théorème de convergence dominée à l'expression de $(L_D f_t)(k)$ (cf. début de 4.6) et d'appliquer le Lemme 3. Le Théorème 1 en résulte.

5 Intégrales d'entrelacement

5.1

Dans ce paragraphe nous aurons à considérer d'autres paraboliques minimaux de G que le parabolique de départ, P. Lorsque nous en aurons besoin nous utiliserons P en indice (supérieur ou inférieur) pour indiquer la dépendance en P de nos constructions. Nous l'omettrons lorsque seul P interviendra. On note $V_P=\theta(N_P)$. Si $w\in W$ on normalise la mesure de Haar sur $V_P\cap w^{-1}N_P w$ comme dans [12], ce qui est un choix différent des mesures standard (et des choix de [6]). Ces mesures seront encore notées dv.

Pour $w\in W$ on note $D_P(w)=\{\lambda\in\mathfrak{a}_{\mathbb{C}}'\,|\,\mathrm{Re}(\lambda,\alpha)>0,\ \forall\alpha\in S_P(w)\}$ où $S_P(w)=\Delta_P^+\cap -w^{-1}(\Delta_P^+)$.

On définit alors pour $\lambda\in D_P(w)$ l'intégrale d'entrelacement $A_P(w,\lambda)$: $I_\lambda^P\to I_{w\lambda}^P$ par la formule suivante (cf. [12, Formules 1.7, 1.9]):

$$\forall f\in I_\lambda,\quad \forall x\in G,(A_P(w,\lambda)f)(x)=\int_{V_P\cap w^{-1}N_P w} f(xwv)\,dv$$

où l'intégrale converge absolument.

Remarque 3 On peut considérer le transposé $A_P'(w,\lambda)$ de $A_P(w,\lambda)$. Mais les intégrales d'entrelacement vérifient des propriétés de dualité et l'on a d'après [12, Prop. 7.8 (ii)] et la continuité des A_P:

$$A_P'(w,\lambda)_{|I_{-w\lambda}^P}=A_P(w^{-1},-w\lambda).$$

D'autre part, comme on est dans le cas sphérique, $A_P(w,\lambda)$ ne dépend pas du choix du représentant de w utilisé.

Pour $w\in W$ on définit un opérateur, $R(w)$, de $C(G/T)$ dans lui-même par:

$$R(w)f(g)=f(gw).$$

Il est clair que $R(w)$ envoie I_λ^P dans $I_{w\lambda}^{wPw^{-1}}$ et entrelace π_λ^P et $\pi_{w\lambda}^{wPw^{-1}}$. Encore une fois $R(w)$ ne dépend pas du choix du représentant de w.

Le lemme suivant est aisé à établir et laissé au lecteur.

Lemme 9 *Soit $w\in W$, $\alpha\in\Delta$ et $s_\alpha\in W$ la symétrie correspondante. On notera $\beta=w\alpha$. Alors $s_\beta=ws_\alpha w^{-1}$. Soit $\lambda\in\mathfrak{a}_{\mathbb{C}}'$. Si $s_\alpha\lambda\in D_P(s_\alpha)$ on a $s_\beta w\lambda=ws_\alpha\lambda\in D_{wPw^{-1}}(s_\beta)$ et le diagramme suivant est commutatif*:

$$\begin{array}{ccc} I_{s_\alpha\lambda}^P & \xrightarrow{A_P(s_\alpha,s_\alpha\lambda)} & I_\lambda^P \\ {\scriptstyle R(w)}\downarrow & & \downarrow{\scriptstyle R(w)} \\ I_{ws_\alpha\lambda}^{wPw^{-1}} & \xrightarrow{A_{wPw^{-1}}(s_\beta,ws_\alpha\lambda)} & I_{w\lambda}^{wPw^{-1}} \end{array}$$

Lemme 10 *Soit $\lambda\in\mathfrak{a}_{\mathbb{C}}'$ avec $\mathrm{Re}(\lambda-\rho_P)$ strictement Δ_P^+-dominant. Alors, pour $w\in W$, on a*:

$$R(w)'\,\xi_{w\lambda}^{1,wPw^{-1}}=\xi_\lambda^{w,P}$$

où $R(w)'$ est le transposé de l'opérateur $R(w)$: $I_\lambda^P\to I_{w\lambda}^{wPw^{-1}}$.

Démonstration. Il est clair que $R(w)'$ agit sur $C_{-w\lambda}^{wPw^{-1}}\subset(I_{w\lambda}^{wPw^{-1}})$ par $R(w^{-1})$, car pour tout f, $g\in C(K/T)$,

$$\int_{K/T} f(kw)\,g(k)\,d\dot{k}=\int_{K/T} f(k)\,g(kw^{-1})\,d\dot{k}.$$

L'égalité à prouver se réduit alors à

$$\forall g\in G,\qquad \xi_{w\lambda}^{1,wPw^{-1}}(gw^{-1})=\xi_\lambda^{w,P}(g).$$

Or, si $g\notin HwP$, $\xi_\lambda^{w,P}(g)=0$ et $gw^{-1}\notin HwPw^{-1}$ implique $\xi_\lambda^{1,wPw^{-1}}(gw^{-1})=0$. Donc:

$$\forall g\notin HwP,\qquad \xi_\lambda^{w,P}(g)=\xi_{w\lambda}^{wPw^{-1}}(gw^{-1})=0.$$

Maintenant, si $g\in HwP$, on a $g=hwan$ avec $h\in H$, $a\in A$, $n\in N_P$. Donc:

$$\xi_\lambda^{w,P}(g)=a^{\lambda-\rho_P}$$

et

$$\xi_{w\lambda}^{1,wPw^{-1}}(gw^{-1})=\xi_{w\lambda}^{1,wPw^{-1}}(hw^{-1}waw^{-1}wnw^{-1}).$$

Mais $wnw^{-1}\in N_{wPw^{-1}}$, $waw^{-1}\in A$. D'où:

$$\xi_{w\lambda}^{1,wPw^{-1}}(gw^{-1})=(waw^{-1})^{w\lambda-\rho_{wPw^{-1}}}.$$

Mais $\rho_{wPw^{-1}}=w\rho_P$. D'où l'égalité voulue. □

5.2

Rappelons (cf. § 3.1) que l'on a un isomorphisme $T \to \tilde{T}$ (resp. $T \to T_{-\lambda}$) de $(I_\lambda^P)'$ sur $I' = \mathscr{D}'(K/T)$ (resp. de I' sur $(I_\lambda^P)'$).

Si $w \in W$ on note $\delta_w \in \mathscr{D}'(K/T)$ la distribution de Dirac en $wT \in K/T$. On notera δ_w^λ au lieu de $(\delta_w)_{-\lambda}$ (ou $\delta_w^{\lambda,P}$ si on veut faire apparaître la dépendance en P). Alors on a:

Lemme 11 *Soit $x \in W_H$. Soit w_0 l'élément de W tel que $w_0(\Delta_P^+) = -\Delta_P^+$ (on a $w_0^2 = 1$). Si $\mathrm{Re}(\lambda - \rho_P)$ est strictement Δ_P^+-dominant, on a:*

$$A'_P(w_0, \lambda)\, \delta_{xw_0}^{w_0\lambda, P} = \eta_{\lambda,x}^P.$$

Démonstration. On omet les indices P. On a $\lambda \in D(w_0)$ et $A(w_0, \lambda)$: $I_\lambda \to I_{w_0\lambda}$ est bien défini. Donc $A'(w_0, \lambda)$ envoie $I'_{w_0\lambda}$ sur I'_λ et comme $\eta_{\lambda,x} \in C_{-\lambda} \subset I'_\lambda$, l'égalité du lemme a un sens. Pour la prouver il suffit de comparer, pour $f \in I_\lambda$, $\langle A'(w_0, \lambda)\, \delta_{xw_0}^{w_0\lambda}, f\rangle_{I_\lambda, I_\lambda}$ et $\langle \eta_{\lambda,x}, f\rangle_{I_\lambda, I_\lambda}$. Mais

$$\begin{aligned}\langle A'(w_0, \lambda)\, \delta_{xw_0}^{w_0\lambda}, f\rangle &= \langle \delta_{xw_0}^{w_0\lambda}, A(w_0, \lambda) f\rangle \\ &= (A(w_0, \lambda) f)(xw_0) \\ &= \int_{\theta(N)} f(xw_0 w_0 v)\, dv \\ &= \int_{\theta(N)} f(xv)\, dv.\end{aligned}$$

Par ailleurs

$$\begin{aligned}\langle \eta_{\lambda,x}, f\rangle &= \int_K \eta_{\lambda,x}(k) f(k)\, dk \\ &= \int_K \eta_{\lambda,x}(xk) f(xk)\, dk.\end{aligned}$$

Mais avec notre normalisation des mesures de Haar sur K et $\theta(N)$, on a pour tout $F \in C_{-2\rho}$

$$\int_K F(k)\, dk = \int_{\theta(N)} F(v)\, dv$$

(cf. [18, 8.4.7]; la normalisation de dv dans [18] coïncide avec la nôtre).

Or $\eta_{\lambda,x}(x.) f(x.) \in C_{-2\rho}$. D'où

$$\langle \eta_{\lambda,x}, f\rangle = \int_{\theta(N)} \eta_{\lambda,x}(xv) f(xv)\, dv.$$

Mais $\eta_{\lambda,x} = \prod_{i=1}^{\ell} \eta_{i,x}^{\lambda_i - 1}$ et comme $x \in W_H$,

$$\eta_{i,x}(g) = |(\pi_i(g) e_{\delta_i}, e_{x\delta_i})|^2 \qquad \text{(cf. Lemme 2).}$$

Prenant $e_{x\delta_i} = \pi_i(x) e_{\delta_i}$ on a donc

$$\begin{aligned}\eta_{i,x}(xv) &= |(\pi_i(v) e_{\delta_i}, e_{\delta_i})|^2 \\ &= |(e_{\delta_i}, \pi_i(\theta(v^{-1})) e_{\delta_i}|^2\end{aligned}$$

et comme $\theta(v^{-1}) \in N$, on a

$$\eta_{i,x}(xv) = 1.$$

D'où l'on déduit:

$$\langle \eta_{\lambda,x}, f \rangle = \int_{\theta(N)} f(xv)\, dv = \langle A'(w_0, \lambda)\, \delta_{xw_0}^{w_0\lambda}, f \rangle. \quad \square$$

D'après le lemme précédent on a, avec les notations et hypothèses du Théorème 1:

$$u_{\lambda,x}^{w} = A'(w_0, \lambda)\, v_{\lambda,x}^{w}$$

avec

$$v_{\lambda,x}^{w} = \int_{H/T} \int_{\mathscr{C}} \psi(hT)\, \varphi(a)\, a^{-2x\rho}\, \pi'_{w_0\lambda}(hwa)\, \delta_{xw_0}^{w_0\lambda}\, da\, d\dot{h}$$

(on a utilisé la propriété d'entrelacement de $A'(w_0, \lambda)$).

Lemme 12 *On a* $v_{\lambda,x}^{w} \in I_{-w_0\lambda}(\subset I'_{w_0\lambda})$ *et*

$$v_{\lambda,x}^{w}(g) = 0 \quad \text{si } g \notin Hwxw_0 AN,$$

$$v_{\lambda,x}^{w}(hwxw_0) = \frac{1}{c_0} \left(\int_{\mathscr{C}} \varphi(a)\, a^{-x(\lambda+\rho)}\, da \right) \psi(h),$$

où c_0 *est une constante que nous préciserons dans la démonstration.*

Démonstration. Par dualité, on obtient facilement:

$$\pi'_{w_0\lambda}(wa)\, \delta_{xw_0}^{w_0\lambda} = a^{-x\lambda+x\rho}\, \delta_{wxw_0}^{w_0\lambda}.$$

En effet, si $f \in I_{w_0\lambda}$,

$$\begin{aligned} \langle \pi'_{w_0\lambda}(wa)\, \delta_{xw_0}^{w_0\lambda}, f \rangle &= \langle \delta_{xw_0}^{w_0\lambda}, \pi_{w_0\lambda}((wa)^{-1}) f \rangle \\ &= f(waxw_0) \\ &= f(wxw_0(xw_0)^{-1} axw_0) \\ &= f(wxw_0)((xw_0)^{-1} axw_0)^{-w_0\lambda-\rho}. \end{aligned}$$

Mais

$$\begin{aligned} ((xw_0)^{-1} axw_0)^{-w_0\lambda-\rho} &= a^{-xw_0(w_0\lambda+\rho)} \\ &= a^{-x\lambda+x\rho}. \end{aligned}$$

D'où l'égalité voulue. On a alors

$$v_{\lambda,x}^{w} = \int_{H/T} \int_{\mathscr{C}} \psi(hT)\, \varphi(a)\, a^{-x(\lambda+\rho)}\, \pi'_{w_0\lambda}(h)\, \delta_{wxw_0}^{w_0\lambda}\, da\, d\dot{h}.$$

Donc, si $f \in I_{w_0\lambda}$, en utilisant (2) (§ 1), on a:

$$\langle v_{\lambda,x}^{w}, f \rangle = \left(\int_{\mathscr{C}} \varphi(a)\, a^{-x(\lambda+\rho)}\, da \right) \int_{H/T} \psi(hT)\, f(hwxw_0)\, d\dot{h}.$$

On introduit alors une fonction sur G:

$$\Phi(g) = 0 \quad \text{si } g \notin Hwxw_0 AN$$

et

$$\Phi(hwxw_0an)=a^{w_0\lambda-\rho}\psi(hT), \quad \forall h\in H, \quad a\in A, \quad n\in N.$$

Il est facile de voir que Φ est bien définie, C^∞ sur G (en utilisant le fait simple à prouver que $(h,a,n)\to hwxw_0an$ est un difféomorphisme de $H\times A\times N$ sur un ouvert de G) et qu'en outre $\Phi\in I_{-w_0\lambda}$. Calculons alors

$$\langle\Phi,f\rangle=\int_{K/T}\Phi(kT)f(kT)\,d\dot{k}.$$

Mais, d'après [13, Lemme 1.3]:

$$\int_{K/T}\Phi(k)f(k)\,d\dot{k}=c_0\sum_{y\in W_H\backslash W}\int_{H/T}\Phi(k(hy))f(k(hy))\,a(hy)^{-2\rho}\,d\dot{h}.$$

Ici, pour $g\in G$, $k(g)\in K$, $a(g)\in A$, $n(g)\in N$ sont définis par l'égalité $g=k(g)\,a(g)\,n(g)$ (décomposition d'Iwasawa) et c_0 est la constante de l'énoncé.

Or $\Phi(hy)=\Phi(k(hy))\ a(hy)^{w_0\lambda-\rho}$ est nul si $y\notin W_H wxw_0$ et égal à $\psi(h)$ si $y=wxw_0$. Par ailleurs

$$f(hy)=f(k(hy))\,a(hy)^{-w_0\lambda-\rho}.$$

D'où

$$\langle\Phi,f\rangle=c_0\int_{H/T}\psi(h)f(hwxw_0)\,d\dot{h}.$$

Finalement

$$\forall f\in I_{w_0\lambda}, \quad \langle v^w_{\lambda,x},f\rangle=\frac{1}{c_0}\int_{\mathscr{C}}\varphi(a)\,a^{-x(\lambda+\rho)})\,da\langle\Phi,f\rangle.$$

D'où l'égalité voulue. □

5.3

On note, pour $\alpha\in\Delta$, H_α la coracine de $\alpha_{|\mathfrak{a}}$ (i.e. $H_\alpha\in\mathfrak{a}$). Pour $\lambda\in\mathfrak{a}'_{\mathbb{C}}$, on pose $\lambda_\alpha=\lambda(H_\alpha)$. Pour $w\in W$, rappelons qu'on a posé $S_P(w)=\{\beta\in\Delta_P^+\mid w\beta\in\Delta_P^+\}$. On note $T(w)=S(w^{-1})$, i.e. $T(w)=\{\beta\in\Delta^+\mid w^{-1}\beta\in\Delta^+\}$.

Pour $\alpha\in\Delta$, on note $G(\alpha)$ le sous-groupe analytique de G d'algèbre de Lie $\mathfrak{g}(\alpha)$ engendrée par $\mathfrak{g}^\alpha$ et $\mathfrak{g}^{-\alpha}$. On dit que α est compacte si $G(\alpha)\cap H$ est compact et non compacte sinon. On note Δ_c (resp. Δ_n) l'ensemble des racines non compactes. On note

$$\varepsilon(\alpha)=-1 \quad \text{si } \alpha \text{ est compacte}$$
$$\varepsilon(\alpha)=1 \quad \text{si } \alpha \text{ est non compacte.}$$

Théorème 2 *Soient w, $y\in W$. On a l'égalité de fonctions méromorphes en $\lambda\in\mathfrak{a}'_{\mathbb{C}}$ à valeurs dans $\mathscr{D}'(K/T)$:*

$$(A'(w,w^{-1}\lambda)\,\xi^y_\lambda)^\sim=\left(\prod_{\alpha\in T(w)}\varepsilon(y\alpha)\left(\frac{2}{\lambda_\alpha}\right)\right)(\xi^{yw}_{w^{-1}\lambda})^\sim$$

où $^\sim$ désigne l'isomorphisme $I'_\lambda\to I'=\mathscr{D}'(K/T)$.

Démonstration du Théorème 2. On procède par récurrence sur la longueur de w, $\ell(w)$, relativement à $\Delta^+ = \Delta_P^+$. Si $w=1$ il n'y a rien à démontrer. On suppose le théorème démontré pour tous les éléments de longueur inférieure ou égale à $p \geqq 0$. Soit alors $w \in W$ avec $\ell(w)=p$ et α une racine simple de Δ^+ telle que $\ell(ws_\alpha)=\ell(w)+1$.

On se place en un point λ où toutes les fonctions méromorphes (en nombre fini) qui interviendront sont définies. C'est certainement le cas pour λ dans un ouvert de $\{\mu \in \mathfrak{a}_{\mathbb{C}}' \mid \mathrm{Re}(\mu-\rho)$ est strictement Δ^+-dominant$\}$. Alors:

$$A(ws_\alpha, s_\alpha w^{-1}\lambda) = A(w, w^{-1}\lambda)\, A(s_\alpha, s_\alpha w^{-1}\lambda)$$

d'après les propriétés des intégrales d'entrelacement (cf. [12] ou [6]) et en transposant on en déduit:

$$A'(ws_\alpha, s_\alpha w^{-1}\lambda)\, \xi_\lambda^y = A'(s_\alpha, s_\alpha w^{-1}\lambda)\, A'(w, w^{-1}\lambda)\, \xi_\lambda^y.$$

L'hypothèse de récurrence appliquée à w montre que

$$A'(ws_\alpha, s_\alpha w^{-1}\lambda)\, \xi_\lambda^y = \left(\prod_{\beta \in T(w)} \varepsilon(y\beta)\left(\frac{2}{\lambda_\beta}\right)\right) A'(s_\alpha, s_\alpha w^{-1}\lambda)\, \xi_{w^{-1}\lambda}^{yw}.$$

Mais $T(ws_\alpha)=S(w^{-1}s_\alpha)$ et d'après [6, p. 51], on a $S(s_\alpha w^{-1})=S(w^{-1}) \cup wS(s_\alpha)$ (union disjointe). Or α est simple, donc $S(s_\alpha)=\{\alpha\}$. D'où $T(ws_\alpha)=T(w) \cup \{w\alpha\}$ (union disjointe).

Il reste donc à prouver que:

$$A'(s_\alpha, s_\alpha w^{-1}\lambda)\, \xi_{w^{-1}\lambda}^{yw} = \varepsilon(yw\alpha)\left(\frac{2}{\lambda_{w\alpha}}\right) \xi_{(ws_\alpha)^{-1}\lambda}^{yws_\alpha}.$$

Comme tous les objets sont définis cela résultera de l'égalité de fonction méromorphe en λ:

$$(7) \qquad (A'(s_\alpha, s_\alpha w^{-1}\lambda)\, \xi_{w^{-1}\lambda}^{yw})^{\sim} = \varepsilon(yw\alpha)\left(\frac{2}{\lambda_{w\alpha}}\right)(\xi_{(ws_\alpha)^{-1}\lambda}^{yws_\alpha})^{\sim}.$$

Pour cela, on va établir:

Lemme 13 On a l'égalité de fonctions méromorphes en λ

$$(A'(s_\alpha, s_\alpha \lambda)\, \xi_\lambda^1)^{\sim} = \varepsilon(\alpha)\left(\frac{2}{\lambda_\alpha}\right)(\xi_{s_\alpha\lambda}^{s_\alpha})^{\sim}$$

$(\varepsilon(\alpha) = -1$ si α est compacte, 1 sinon).

Démonstration du Lemme 13 D'après [1, Lemme 7.4], ce lemme se réduit au même résultat pour $G(\alpha)$, qui dans ce cas sera démontré plus loin (§ 6).

Suite de la démonstration du Théorème 2 Grâce à ce lemme on peut calculer $A'(s_\alpha, s_\alpha w^{-1}\lambda)\, \xi_{w^{-1}\lambda}^{yw}$. En effet par prolongement méromorphe des identités on déduit du Lemme 10:

$$\xi_{w^{-1}\lambda}^{yw,P} = R(yw)'\, \xi_{y\lambda}^{1;ywPw^{-1}y^{-1}}.$$

D'où:

$$A'_P(s_\alpha, s_\alpha w^{-1}\lambda)\,\xi^{ywP}_{w^{-1}\lambda} = A'_P(s_\alpha, s_\alpha w^{-1}\lambda)\,R(yw)'\,\xi^{1,ywPw^{-1}y^{-1}}_{y\lambda}.$$

On pose alors $w' = yw$ et $\beta' = w'\alpha$ qui est simple pour $\Delta^+_{P'}$ où $P' = w'Pw'^{-1}$, $\lambda' = w^{-1}\lambda$. D'après le Lemme 9 on obtient par transposition:

$$A'_P(s_\alpha, s_\alpha \lambda')\,R(w')' = R(w')'\,A'_{P'}(s_{\beta'}, s_{\beta'} w'\lambda').$$

D'où

$$A'_P(s_\alpha, s_\alpha w^{-1}\lambda)\,\xi^{yw,P}_{w^{-1}\lambda} = R(w')'\,A'_{P'}(s_{\beta'}, s_{\beta'} w'\lambda')\,\xi^{1,P'}_{y\lambda}.$$

On applique le lemme précédent à P' et β' et l'on obtient:

$$A'_P(s_\alpha, s_\alpha w^{-1}\lambda)\,\xi^{yw,P}_{w^{-1}\lambda} = \varepsilon(\beta')\left(\frac{2}{(y\lambda)_{\beta'}}\right) R(w')'\,\xi^{s_{\beta'},P'}_{s_{\beta'}y\lambda} \tag{8}$$

avec

$$\begin{aligned}\varepsilon(\beta') &= -1 \quad \text{si } G(\beta')\cap H \text{ est compact}\\ &= +1 \text{ sinon.}\end{aligned}$$

Mais $\beta' = yw\alpha$. Donc $(y\lambda)_{\beta'} = (y\lambda)_{yw\alpha} = \lambda_{w\alpha}$ et $\varepsilon(\beta') = \varepsilon(yw\alpha)$. Enfin, le Lemme 10 montre que (par prolongement méromorphe):

$$R(w')'\,\xi^{s_{\beta'},P'}_{s_{\beta'}y\lambda} = R(w')'\,R(s_{\beta'})'\,\xi^{1,s_{\beta'}P's_{\beta'}}_{y\lambda}.$$

Or $R(s_{\beta'})\circ R(w') = R(s_{\beta'}w')$ et par transposition on a

$$R(w')'\,\xi^{s_{\beta'},P'}_{s_{\beta'}y\lambda} = R(s_{\beta'}w')'\,\xi^{1,s_{\beta'}P's_{\beta'}}_{y\lambda}.$$

On applique à nouveau le Lemme 10, en tenant compte du fait que $s_{\beta'}P's_{\beta'} = s_{\beta'}w'P(s_{\beta'}w')^{-1}$:

$$R(w')'\,\xi^{s_{\beta'},P'}_{s_{\beta'}y\lambda} = \xi^{s_{\beta'}w',P}_{(s_{\beta'}w')^{-1}y\lambda}.$$

Mais $s_{\beta'}w' = w's_\alpha w'^{-1}w' = yws_\alpha$ et $(s_{\beta'}w')^{-1}y\lambda = s_\alpha w^{-1}\lambda = (ws_\alpha)^{-1}\lambda$.

Finalement, en tenant compte de ces relations et l'égalité précédente, et en remplaçant dans (8) $R(w')'\,\xi^{s_{\beta'},P'}_{s_{\beta'}y\lambda}$ par sa valeur ainsi obtenue, on a:

$$A'_P(s_\alpha, s_\alpha w^{-1}\lambda)\,\xi^{yw,P}_{w^{-1}\lambda} = \varepsilon(yw\alpha)\left(\frac{2}{\lambda_{w\alpha}}\right)\xi^{yws_\alpha,P}_{(ws_\alpha)^{-1}\lambda}$$

égalité valable sur un ouvert de $\mathfrak{a}'_{\mathbb{C}}$ (où toutes les fonctions méromorphes rencontrées sont définies). L'égalité (7) en résulte, ainsi que le théorème. □

6 Démonstration du Lemme 13 en rang 1

6.1 Cas où H est compact

Alors $G \simeq \mathrm{SL}(2,\mathbb{C})$, $H = K \simeq \mathrm{SU}(2)$, ξ_λ^1 est un vecteur invariant par K, et on sait comment il se transforme par les intégrales d'entrelacement. En fait $\xi_\lambda^1 \in I_{-\lambda}$. Alors, d'après la Remarque 3 on a:

$$A'(s_\alpha, s_\alpha \lambda)\, \xi_\lambda^1 = A(s_\alpha, -\lambda)\, \xi_\lambda^1$$

et d'après [6, Prop. 3.7], on a:

$$A(s_\alpha, -\lambda)\, \xi_\lambda^1 = -\frac{2}{\lambda_\alpha}\, \xi_{s_\alpha \lambda}^1$$

(le 2 provient d'une normalisation différente des mesures de Haar). Mais $W_H = W$ dans ce cas et $\xi_{s_\alpha \lambda}^{s_\alpha} = \xi_{s_\alpha \lambda}^1$. Finalement

$$A'(s_\alpha, s_\alpha \lambda)\, \xi_\lambda^1 = -\frac{2}{\lambda_\alpha}\, \xi_{s_\alpha \lambda}^{s_\alpha} = \varepsilon(\alpha) \frac{2}{\lambda_\alpha}\, \xi_{s_\alpha \lambda}^{s_\alpha}$$

et le Lemme 13 est démontré dans le cas où G est de rang 1 et H compact.

6.2 Cas où H est non compact. Préliminaires

Dans ce cas G peut s'identifier à $\mathrm{SL}(2,\mathbb{C})$ et H à $\mathrm{SL}(2,\mathbb{R})$, K à $\mathrm{SU}(2)$. Mais les calculs nous ont semblé un peu malaisés dans ce contexte. Nous préférerons identifier G/H à $SO_e(3,1)/SO_e(2,1)$. L'homomorphisme naturel de $\mathrm{SL}(2,\mathbb{C})$ dans $SO_e(3,1)$ n'est pas un isomorphisme, mais tous les calculs se passent modulo le centre de $\mathrm{SL}(2,\mathbb{C})$. Dans la suite nous ferons donc comme si $G = SO_e(3,1)$ et

$$H = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & a \end{pmatrix} \middle| a \in SO_e(2,1) \right\}. \quad \text{Alors}$$

$$K = \left\{ \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \middle| a \in SO(3) \right\}$$

$$T = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{pmatrix} \; a \in SO(2) \right\}$$

$$A = \left\{ \begin{pmatrix} cht & 0 & 0 & sht \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ sht & 0 & 0 & cht \end{pmatrix} \middle| \; t \in \mathbb{R} \right\}$$

$$N = \left\{ \begin{pmatrix} 1 - \frac{1}{2}(x_2^2 + x_3^2) & -x_2 & -x_3 & \frac{1}{2}(x_2^2 + x_3^2) \\ x_2 & 1 & 0 & -x_2 \\ x_3 & 0 & 1 & -x_3 \\ -\frac{1}{2}(x_2^2 + x_3^2) & -x_2 & -x_3 & 1 + \frac{1}{2}(x_2^2 + x_3^2) \end{pmatrix} \middle| \; x_2, x_3 \in \mathbb{R}. \right\}$$

En posant

$$X_0=\begin{pmatrix}0&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&0\end{pmatrix}$$

et $a_t=\exp tX_0$ on a $(a_t)^\alpha=e^t$, $\rho=\alpha$. Le poids fondamental δ_1 vérifie $\delta_{1|\mathfrak{a}}=\frac{1}{2}\alpha$. Si $\lambda=2s\delta_1\in\mathfrak{a}'_{\mathbb{C}}$ avec $s\in\mathbb{C}$, on a:

$$\lambda_\alpha=2s,\qquad (a_t)^\lambda=e^{st}$$

$$I_\lambda=\{\varphi\colon G\to\mathbb{C}\mid\varphi(gma_tn)=e^{-(s+1)t}\varphi(g),\ \forall g\in G,\ m\in T,\ n\in N,\ t\in\mathbb{R}\}$$

On choisit pour représentant de s_α:

$$\begin{pmatrix}-1&0&0&0\\0&-1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}.$$

On a $W_H=\{e\}$. Identifions la fonction ε_1. C'est un coefficient d'une représentation de G, (π,H_π), de dimension 4, (à savoir $\pi_1\otimes(\pi'_1\circ\sigma)$), $g\mapsto(\pi(g)e,v)$ où v est un vecteur invariant par H et e un vecteur invariant par N et de poids $2\delta_{1|\mathfrak{a}}=\alpha_{|\mathfrak{a}}$ sous A. De plus $\varepsilon_1(1)=1$. Ces propriétés de π montrent que π est la représentation naturelle de $SO_e(3,1)$ dans $\mathbb{R}^4$ (ou plutôt sa complexifiée) munie du produit scalaire naturel. On note (e_1,e_2,e_3,e_4) la base canonique de $\mathbb{R}^4$. Alors e_1 est invariant sous H et e_1-e_4 est invariant par N et de poids α sous A. On a donc $\varepsilon_1(g)=(g(e_1-e_4),e_1)$. Alors, si $\mathrm{Re}(\lambda-\rho)$ est strictement Δ^+-dominant (i.e. $\mathrm{Re}(s-1)>0$), on a

$$\xi^1_\lambda(g)=(\mathrm{Sup}(\varepsilon_1(g),0))^{s-1}$$

et

$$\xi^{s_\alpha}_\lambda(g)=\xi^1_\lambda(s_\alpha g).$$

On va étudier $(\xi^1_\lambda)^\sim$ et $(\xi^{s_\alpha}_\lambda)^\sim$ qui sont des distributions sur K/T que l'on identifie naturellement à S^2 dans $\mathbb{R}e_1\oplus\mathbb{R}e_2\oplus\mathbb{R}e_3$ (la classe de T correspondant à e_1). D'autre part $G/P\approx K/T$. Avec ces identifications, H a 2 orbites ouvertes dans $S^2(\approx G/P)$,

$$O^\pm=\{(x_1,x_2,x_3)\in S^2\mid\pm x_1>0\}.$$

Alors $\varepsilon_1(k)=(k(e_1-e_4),e_1)=(ke_1,e_1)$. D'où

$$\tilde{\varepsilon}_1(x_1,x_2,x_3)=x_1$$

et

$$(\xi^1_\lambda)^\sim=(\mathrm{Sup}(x_1,0))^{s-1}\qquad\text{pour }\mathrm{Re}(s-1)>0.$$

De même on voit que $\eta=\eta_{1,1}$ et $\eta_\lambda=\eta_{\lambda,1}$ vérifient

$$\eta(g)=\tfrac{1}{2}(g(e_1-e_4),e_1-e_4)$$

et

$$\eta_\lambda(g)=\frac{1}{2^{s-1}}(g(e_1-e_4),e_1-e_4)^{s-1} \quad \text{pour } \mathrm{Re}(s-1)>0.$$

Alors $\eta_\lambda(k)=\frac{1}{2^{s-1}}((ke_1,e_1)+1)^{s-1}$ et $\eta_\lambda^{\sim}(x_1,x_2,x_3)=\frac{1}{2^{s-1}}(x_1+1)^{s-1}$ (toujours pour $\mathrm{Re}(s-1)>0$).

Avec les notations du Théorème 1 on notera:

$$u^1_{\lambda,1}=u^+_\lambda, \qquad u^{s_\alpha}_{\lambda,1}=u^-_\lambda\ (1\in W_H).$$

On va calculer de deux façons $\langle u^\pm_\lambda,\xi^1_{-\lambda}\rangle_{I_{-\lambda},I'_{-\lambda}}$ pour λ tel que $\mathrm{Re}(s-1)>0$ et où toutes les fonctions méromorphes qui interviendront (en nombre fini) sont définies, ce qui arrivera donc sur un ouvert de $a'_{\mathbb{C}}$.

6.3 *Premier calcul de* $\langle u^\pm_\lambda,\xi^1_{-\lambda}\rangle_{I_{-\lambda},I'_{-\lambda}}$

Avec les notations du Lemme 12, on pose

$$v^1_{\lambda,1}=v^+_\lambda \quad \text{et} \quad v^{s_\alpha}_{\lambda,1}=v^-_\lambda.$$

On sait (cf. Lemme 12 et ce qui le précède) que $u^\pm_\lambda=A'(w_0,\lambda)\,v^\pm_\lambda$ et ici $w_0=s_\alpha$. Mais, d'après le Lemme 12, $v^\pm_\lambda\in I_{-w_0\lambda}=I_\lambda$.

Alors la Remarque 3 permet de conclure que $u^\pm_\lambda=A(s_\alpha,\lambda)\,v^\pm_\lambda$. Donc, en posant $\xi^1_{-\lambda}=\xi_{-\lambda}$, on a:

$$\begin{aligned}\langle u^\pm_\lambda,\xi_{-\lambda}\rangle&=\langle A(s_\alpha,\lambda)\,v^\pm_\lambda,\xi_{-\lambda}\rangle\\&=\langle v^\pm_\lambda,A'(s_\alpha,\lambda)\,\xi_{-\lambda}\rangle.\end{aligned}$$

Mais, d'après [1, Proposition 6.1], on a:

$$A'(s_\alpha,\lambda)\,\xi_{-\lambda}=B^+(\lambda)\,\xi_\lambda+B^-(\lambda)\,\xi^{s_\alpha}_\lambda$$

où B^+ et B^- sont méromorphes en λ. Donc

$$\langle u^\pm_\lambda,\xi_{-\lambda}\rangle=B^+(\lambda)\,\langle v^\pm_\lambda,\xi_\lambda\rangle+B^-(\lambda)\,\langle v^\pm_\lambda,\xi^{s_\alpha}_\lambda\rangle.$$

Mais nous avons déterminé $v^\pm_\lambda$ au Lemme 12. En particulier v^+_λ est à support contenu dans Hw_0P et v^-_λ est à support dans HP. De plus, comme ici $x=1$, on a $C=-C^+$ et $\mathscr{C}=\exp(-C^+)$ que l'on notera $\mathscr{C}^-$. Alors

$$v^-_\lambda(h)=v^+_\lambda(hw_0)=\frac{1}{c_0}\psi(h)\int_{\mathscr{C}^-}a^{-\lambda-\rho}\varphi(a)\,da.$$

On utilise à nouvea la formule intégrale de [13, Lemme 1.3] pour calculer:

$$\langle v_\lambda^\pm, \xi_\lambda^\pm \rangle = \int_{K/T} v_\lambda^\pm(k)\, \xi_\lambda^\pm(k)\, dk,$$

où l'on a noté $\xi_\lambda^+ = \xi_\lambda^1$ et $\xi_\lambda^- = \xi_\lambda^{s_\alpha}$. Alors, pour des raisons de support, on a:

$$\langle v_\lambda^+, \xi_\lambda^+ \rangle = 0, \quad \langle v_\lambda^-, \xi_\lambda^- \rangle = 0$$

et (la constante c_0 disparaît):

$$\langle v_\lambda^+, \xi_\lambda^- \rangle = \left(\int_{\mathscr{C}^-} a^{-\lambda-\rho} \varphi(a)\, da \right) \times \left(\int_{H/T} \psi(h)\, \xi_\lambda^-(hs_\alpha)\, d\dot{h} \right)$$

$$\langle v_\lambda^+, \xi_\lambda^- \rangle = \left(\int_{\mathscr{C}^-} a^{-\lambda-\rho} \varphi(a)\, da \right) \times \int_{H/T} \psi(h)\, d\dot{h}$$

et

$$\langle v_\lambda^-, \xi_\lambda^+ \rangle = \left(\int_{\mathscr{C}^-} a^{-\lambda-\rho} \varphi(a)\, da \right) \times \int_{H/T} \psi(h)\, d\dot{h}.$$

D'où

Lemme 14 $\langle u_\lambda^\pm, \xi_{-\lambda} \rangle = B^\mp(\lambda) \left(\int_{\mathscr{C}^-} a^{-\lambda-\rho} \varphi(a)\, da \right) \left(\int_{H/T} \varphi(h)\, d\dot{h} \right)$ (*pour des valeurs convenables de* λ).

6.4 Deuxième calcul de $\langle u_\lambda^\pm, \xi_{-\lambda} \rangle$

On notera:

$$O = O^+ \cup O^- (\subset S^2)$$

$$E = \{x \in S^2 \,|\, |x_1| < 1\}$$

et pour $\varepsilon > 0$

$$O_\varepsilon = \{x \in S^2 \,|\, |x_1| > \varepsilon\}$$

$$E_\varepsilon = \{x \in S^2 \,|\, |x_1| < \varepsilon\}.$$

Enfin on note

$$\mathscr{U} = \{\tau \in \mathscr{D}'(S^2) \,|\, \tau_{|E} \text{ est } C^\infty\}$$

$$\mathscr{V} = \{\xi \in \mathscr{D}'(S^2) \,|\, \xi_{|O} \text{ est } C^\infty\}.$$

Définissons une forme bilinéaire entre $\mathscr{U}$ et $\mathscr{V}$. Soit (φ_1, φ_2) une partition de l'unité C^∞ subordonnée au recouvrement ouvert de S^2 par $E_{3/4}$ et $O_{1/4}$, i.e.: $\varphi_1, \varphi_2 \in \mathscr{D}(S^2)$, Supp $\varphi_1 \subset E_{3/4}$, Supp $\varphi_2 \subset O_{1/4}$ et $\varphi_1 + \varphi_2 \equiv 1$. En particulier $\varphi_2 \equiv 1$ sur $O_{3/4}$ (puisque $\varphi_1 \equiv 0$ sur $O_{3/4}$). Alors si $\tau \in \mathscr{U}$ et $\xi \in \mathscr{V}$, $\varphi_1 \tau$ est C^∞ sur S^2 et $\langle \varphi_1 \tau, \xi \rangle_{\mathscr{D}, \mathscr{D}'}$ est défini. De même $\varphi_2 \xi$ est C^∞ sur S^2 et $\langle \tau, \varphi_2 \xi \rangle_{\mathscr{D}', \mathscr{D}}$ est défini. On définit alors:

$$\langle\!\langle \tau, \xi \rangle\!\rangle = \langle \varphi_1 \tau, \xi \rangle_{\mathscr{D}, \mathscr{D}'} + \langle \tau, \varphi_2 \xi \rangle_{\mathscr{D}', \mathscr{D}}.$$

Si $\tau \in \mathscr{D}(S^2) \subset \mathscr{U}$, il est clair que

$$\langle\langle \tau, \xi \rangle\rangle = \langle \tau, \xi \rangle_{\mathscr{D}, \mathscr{D}'}. \tag{9}$$

Lemme 15 *Soit $\tau \in \mathscr{U}$ avec* $\operatorname{Supp} \tau \subset O_{(3/4)+\varepsilon}(\varepsilon > 0)$ *et soit $\varphi_3 \in \mathscr{D}(S^2)$ avec $\varphi_3 \equiv 1$ sur $O_{(3/4)+\varepsilon}$ et* $\operatorname{Supp} \varphi_3 \subset O_{3/4}$. *Alors, pour tout $\xi \in \mathscr{V}$, $\langle\langle \tau, \xi \rangle\rangle = \langle \tau, \varphi_3 \xi \rangle_{\mathscr{D}', \mathscr{D}}$ (qui est bien défini puisque $\varphi_3 \xi \in \mathscr{D}(S^2)$).*

Démonstration. On a par définition:

$$\langle\langle \tau, \xi \rangle\rangle = \langle \varphi_1 \tau, \xi \rangle_{\mathscr{D}, \mathscr{D}'} + \langle \tau, \varphi_2 \xi \rangle_{\mathscr{D}', \mathscr{D}}.$$

Mais $\operatorname{Supp} \varphi_1 \subset E_{3/4}$ et $\operatorname{Supp} \tau \subset O_{(3/4)+\varepsilon}$. Comme $E_{3/4} \cap O_{(3/4)+\varepsilon} = \emptyset$, on a $\varphi_1 \tau = 0$ et:

$$\langle\langle \tau, \xi \rangle\rangle = \langle \tau, \varphi_2 \xi \rangle_{\mathscr{D}', \mathscr{D}}.$$

Mais $\varphi_3 \equiv 1$ sur $O_{(3/4)+\varepsilon}$ et $\operatorname{Supp} \tau \subset O_{(3/4)+\varepsilon}$. Donc $\varphi_3 \tau = \tau$ et

$$\begin{aligned} \langle\langle \tau, \xi \rangle\rangle &= \langle \varphi_3 \tau, \varphi_2 \xi \rangle_{\mathscr{D}', \mathscr{D}} \\ &= \langle \tau, \varphi_2 \varphi_3 \xi \rangle_{\mathscr{D}', \mathscr{D}}. \end{aligned}$$

Mais $\varphi_2 \equiv 1$ sur $O_{3/4}$ et $\operatorname{Supp} \varphi_3 \subset O_{3/4}$. Donc $\varphi_2 \varphi_3 = \varphi_3$ et l'on a bien

$$\langle\langle \tau, \xi \rangle\rangle = \langle \tau, \varphi_3 \xi \rangle_{\mathscr{D}', \mathscr{D}}. \quad \square$$

Soit $\psi \in C_c^\infty(H)$, que l'on regarde comme distribution à support compact sur G (à l'aide de la mesure de Haar standard sur H). On considère les actions de G, $\tilde{\pi}'_\lambda$ et $\tilde{\pi}'_{-\lambda}$ sur $I' = \mathscr{D}'(S^2)$ (transportées des actions de G sur I'_λ et $I'_{-\lambda}$ par les isomorphismes $\tilde{}$).

Lemme 16 *Soit $\lambda \in \mathfrak{a}'_{\mathbb{C}}$. Soient $\tau \in \mathscr{U}$ et $\psi \in C_c^\infty(H)$, tels que $\tilde{\pi}'_\lambda(\psi)\tau \in \mathscr{U}$. Soit $\xi \in \mathscr{V}$ que l'on suppose H-invariant par $\tilde{\pi}'_{-\lambda}$. Alors*

$$\langle\langle \tilde{\pi}'_\lambda(\psi)\tau, \xi \rangle\rangle = \int_H \psi(h)\, dh \langle\langle \tau, \xi \rangle\rangle$$

pour ψ à support assez petit (i.e. *tel que* $\operatorname{Supp} \psi$. $O_{7/8} \subset O_{(3/4)+\varepsilon}$ *pour un* $\varepsilon > 0$).

Démonstration. On écrit $\tau = \tau_1 + \tau_2$ avec $\tau_1 \in \mathscr{D}(S^2)$ et $\tau_2 \in \mathscr{D}'(S^2)$ avec $\operatorname{Supp} \tau_2 \subset O_{7/8}$. Alors

$$\langle\langle \tilde{\pi}'_\lambda(\psi)\tau, \xi \rangle\rangle = \langle\langle \tilde{\pi}'_\lambda(\psi)\tau_1, \xi \rangle\rangle + \langle\langle \tilde{\pi}'_\lambda(\psi)\tau_2, \xi \rangle\rangle.$$

Mais $\tau_1 \in \mathscr{D}(S^2)$, et

$$\tilde{\pi}'_\lambda(\psi)_{|\mathscr{D}(S^2)} = \tilde{\pi}_{-\lambda}(\psi).$$

De plus: $\tilde{\pi}_{-\lambda}(\psi)\tau_1 \in \mathscr{D}(S^2)$. Donc, d'après (9):

$$\langle\langle \tilde{\pi}'_\lambda(\psi)\tau_1, \xi \rangle\rangle = \langle \tilde{\pi}_{-\lambda}(\psi)\tau_1, \xi \rangle_{\mathscr{D}, \mathscr{D}'}.$$

Mais: $(\pi_{-\lambda}(\psi))' = \pi'_{-\lambda}(\check{\psi})$ d'après (3) (cf. § 1.3) et par transport par $\tilde{}$ on en déduit:

$$\langle\langle \tilde{\pi}'_\lambda(\psi)\tau_1, \xi \rangle\rangle = \langle \tau_1, \tilde{\pi}'_{-\lambda}(\check{\psi})\xi \rangle_{\mathscr{D}, \mathscr{D}'},$$

et en utilisant l'invariance par H de ξ on a finalement:

$$\langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau_1, \xi\rangle\!\rangle = \Big(\int_H \psi(h)\,dh\Big)\,\langle \tau_1, \xi\rangle_{\mathscr{D},\mathscr{D}'} = \Big(\int_H \psi(h)\,dh\Big)\,\langle\!\langle \tau_1, \xi\rangle\!\rangle. \tag{10}$$

Etudions maintenant $\langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau_2, \xi\rangle\!\rangle$. La condition de support sur ψ et τ_2 montre que l'on a $\mathrm{Supp}\,\tilde{\pi}'_\lambda(\psi)\,\tau_2 \subset O_{(3/4)+\varepsilon}$ et l'on a d'après le Lemme 15:

$$\langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau_2, \xi\rangle\!\rangle = \langle \tilde{\pi}'_\lambda(\psi)\,\tau_2, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}}.$$

Soit $\tau_n \in \mathscr{D}(S^2)$, une suite convergeant faiblement vers τ_2, avec de plus $\mathrm{Supp}\,\tau_n \subset O_{7/8}$ (on se ramène, pour exhiber τ_n, à un problème sur $\mathbb{R}^2$ et on utilise la convolution, cf. [16]). Alors on a encore $\mathrm{Supp}(\tilde{\pi}'_\lambda(\psi)\,\tau_n) \subset O_{(3/4)+\varepsilon}$. D'autre part, d'après (3), § 1.3, on a

$$\forall f \in \mathscr{D}(S^2), \qquad \langle \tilde{\pi}'_\lambda(\psi)\,\tau_n, f\rangle_{\mathscr{D}',\mathscr{D}} = \langle \tau_n, \tilde{\pi}_\lambda(\check{\psi})\,f\rangle_{\mathscr{D}',\mathscr{D}}.$$

D'où $\tilde{\pi}'_\lambda(\psi)\,\tau_n$ converge faiblement dans $\mathscr{D}'$ vers $\tilde{\pi}'_\lambda(\psi)\,\tau_2$. Donc

$$\langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau_2, \xi\rangle\!\rangle = \lim_{n\to+\infty} \langle \tilde{\pi}'_\lambda(\psi)\,\tau_n, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}}. \tag{11}$$

Comme $\mathrm{Supp}(\tilde{\pi}'_\lambda(\psi)\,\tau_n) \subset O_{(3/4)+\varepsilon}$, le Lemme 15 implique:

$$\langle \tilde{\pi}'_\lambda(\psi)\,\tau_n, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}} = \langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau_n, \xi\rangle\!\rangle$$

et comme $\tilde{\pi}'_\lambda(\psi)\,\tau_n \in \mathscr{D}$ on a d'après (9):

$$\langle \tilde{\pi}'_\lambda(\psi)\,\tau_n, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}} = \langle \tilde{\pi}'_\lambda(\psi)\,\tau_n, \xi\rangle_{\mathscr{D},\mathscr{D}'}.$$

Mais $\tau_n \in \mathscr{D}$, donc $\tilde{\pi}'_\lambda(\psi)\,\tau_n = \tilde{\pi}_{-\lambda}(\psi)\,\tau_n$ et en utilisant (3), § 1.3 on a:

$$\langle \tilde{\pi}_{-\lambda}(\psi)\,\tau_n, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}} = \langle \tau_n, \tilde{\pi}'_{-\lambda}(\check{\psi})\,\xi\rangle_{\mathscr{D},\mathscr{D}'}.$$

En utilisant l'invariance de ξ sous H on a donc:

$$\langle \tilde{\pi}_\lambda(\psi)\,\tau_n, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}} = \Big(\int_H \psi(h)\,dh\Big)\,\langle \tau_n, \xi\rangle_{\mathscr{D},\mathscr{D}'}.$$

Mais d'après (9) et le Lemme 15 on a:

$$\langle \tau_n, \xi\rangle_{\mathscr{D},\mathscr{D}'} = \langle\!\langle \tau_n, \xi\rangle\!\rangle = \langle \tau_n, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}}.$$

Finalement:

$$\langle \tilde{\pi}'_\lambda(\psi)\,\tau_n, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}} = \Big(\int_H \psi(h)\,dh\Big)\,\langle \tau_n, \varphi_3\,\xi\rangle_{\mathscr{D},\mathscr{D}'}$$

et en passant à la limite, grâce à (11) on a:

$$\langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau_2, \xi\rangle\!\rangle = \Big(\int_H \psi(h)\,dh\Big)\,\langle \tau_2, \varphi_3\,\xi\rangle_{\mathscr{D}',\mathscr{D}}.$$

On utilise encore le Lemme 15 pour finalement obtenir:

$$\langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau_2, \xi\rangle\!\rangle = \Big(\int_H \psi(h)\,dh\Big)\,\langle\!\langle \tau_2, \xi\rangle\!\rangle. \tag{12}$$

En ajoutant les relations (10) et (12) on obtient

$$\langle\!\langle \tilde{\pi}'_\lambda(\psi)\,\tau, \xi\rangle\!\rangle = \Big(\int_H \psi(h)\,dh\Big)\,\langle\!\langle \tau, \xi\rangle\!\rangle. \quad \square$$

Revenons au calcul de $\langle u_\lambda^\pm, \xi_{-\lambda}\rangle = \langle \tilde{u}_\lambda^\pm, \tilde{\xi}_{-\lambda}\rangle_{\mathscr{D},\mathscr{D}'}$. On a $\tilde{u}_\lambda^\pm \in \mathscr{D}(S^2)$ et d'après la définition de $u_\lambda^\pm$ (cf. Théorème 1 et § 6.2) et les propriétés de transformations sous A de η_λ (Lemme 2) on a:

$$u_\lambda^+ = \Big(\int_{\mathscr{C}^-} \varphi(a)\,a^{-(\lambda+\rho)}\,da\Big)\,(\pi'_\lambda(\psi)\,\eta_\lambda)$$

et

$$u_\lambda^- = \Big(\int_{\mathscr{C}^-} \varphi(a)\,a^{-(\lambda+\rho)}\,da\Big)\,(\pi'_\lambda(\psi)\,\pi'_\lambda(s_\alpha)\,\eta_\lambda)$$

(on a intégré l'action de A sur η_λ). On notera $\eta_\lambda^+ = \eta_\lambda$, $\eta_\lambda^- = \pi'_\lambda(s_\alpha)\,\eta_\lambda$. Alors, comme $u_\lambda^\pm$ est C^∞ sur S^2, on a:

$$\langle u_\lambda^\pm, \xi_{-\lambda}\rangle = \langle\!\langle \tilde{u}_\lambda^\pm, \tilde{\xi}_{-\lambda}\rangle\!\rangle$$

et on peut appliquer le Lemme 16 (pour Supp ψ assez petit) et l'on a:

$$\langle u_\lambda^\pm, \xi_{-\lambda}\rangle = \Big(\int_{\mathscr{C}^-} \varphi(a)\,a^{-(\lambda+\rho)}\,da\Big)\,\Big(\int_H \psi(h)\,dh\Big)\,\langle\!\langle \tilde{\eta}_\lambda^\pm, \tilde{\xi}_{-\lambda}\rangle\!\rangle.$$

Il ne reste plus qu'à calculer $\langle\!\langle \tilde{\eta}_\lambda^\pm, \tilde{\xi}_{-\lambda}\rangle\!\rangle$.

Rappelons que l'on a choisi λ avec $\mathrm{Re}(s-1)>0$. Donc $\tilde{\xi}_{-\lambda}$ n'est pas en général une fonction. Par contre $\tilde{\eta}_\lambda^+$ est la fonction sur la sphère définie par:

$$x=(x_1,x_2,x_3) \mapsto \frac{1}{2^{s-1}}(x_2+1)^{s-1}.$$

Alors $\langle\!\langle \tilde{\eta}_\lambda^\pm, \tilde{\xi}_{-\lambda}\rangle\!\rangle$ est la valeur en $t=s$ du prolongement méromorphe de la fonction $g^\pm$ définie par

$$t \to \int_{K/T} \tilde{\xi}_{-\nu}(k)\,\eta_\lambda^\pm(k)\,dk = g^\pm(t), \quad \text{où} \quad \nu = 2t\delta \text{ avec } -\mathrm{Re}(t+1)>0.$$

Donc, pour $\mathrm{Re}(t+1)<0$, on a:

$$g^+(t) = \int_{S^2} (\sup(x_1,0))^{-t-1}\,\frac{(x_1+1)^{s-1}}{2^{s-1}}\,dx.$$

On prend sur la sphère des coordonnées cylindriques (x_1, θ) et l'on a $dx = dx_1 \frac{d\theta}{4\pi}$ (on veut que S^2 soit de surface 1). D'où

$$g^+(t) = \frac{1}{2} \int_0^1 x_1^{-t-1} \frac{(x_1+1)^{s-1}}{2^{s-1}} dx_1.$$

Considérons la fonction h de 2 variables complexes (u, v) définie par

$$h(u, v) = \frac{1}{2} \int_0^1 x_1^{-u-1} \frac{(x_1+1)^{v-1}}{2^{v-1}} dx_1 \quad \text{pour } \mathrm{Re}(u) < 0 \quad \text{et} \quad v \in \mathbb{C}.$$

Comme la distribution $(x_+)^z$ n'a des pôles que pour $z \in -\mathbb{N}^*$, h est méromorphe sur $\mathbb{C}^2$ avec au pire des pôles sur les plans d'équation $u = p \in \mathbb{N}$, et l'on a, pour s non entier:

$$g(s) = h(s, s).$$

Calculons

$$h(u, u) \quad \text{pour } \mathrm{Re}\, u < 0.$$

D'après [7, Formule 1.1, p. 10], on a:

$$\int_0^1 t^{x-1}(1-t)^{y-1}(1+bt)^{-x-y} dt = (1+b)^{-x} B(x, y)$$

$$\text{pour } b > -1, \quad \mathrm{Re}\, x > 0, \quad \mathrm{Re}\, y > 0.$$

On prend ici $x = -u$, $y = 1$, $b = 1$ et l'on trouve

$$h(u, u) = \frac{1}{2^u} \times 2^u B(-u, 1) = \frac{\Gamma(-u)\,\Gamma(1)}{\Gamma(-u+1)} = -\frac{1}{u}.$$

D'où, par prolongement holomorphe,

$$g^+(s) = -\frac{1}{s}$$

et

$$\langle\langle \tilde{\eta}_\lambda^+, \tilde{\xi}_{-\lambda} \rangle\rangle = -\frac{1}{s}.$$

Calculons maintenant:

$$\langle\langle \tilde{\eta}_\lambda^-, \tilde{\xi}_{-\lambda} \rangle\rangle.$$

On a pour $\mathrm{Re}(t+1) < 0$:

$$g^-(t) = \frac{1}{2} \int_0^1 x_1^{-t-1} \frac{(1-x_1)^{s-1}}{2^{s-1}} dx_1$$

et, par définition de la fonction B, pour $\mathrm{Re}\, t < 0$ et $\mathrm{Re}\, s > 0$ (cf. [7, Formule 1, p. 9]):

$$g^-(t) = \frac{1}{2^s} B(-t, s) = \frac{1}{2^s} \frac{\Gamma(-t)\,\Gamma(s)}{\Gamma(s-t)}.$$

Comme $\lim_{t \to s} \frac{1}{\Gamma(s-t)} = 0$, on voit que pour s non entier:

et donc

$$\langle\langle \tilde{\eta}_{\lambda}^{-}, \tilde{\xi}_{-\lambda} \rangle\rangle = 0.$$

Finalement on a obtenu:

Lemme 17 *Pour ψ à support assez petit (voisin de e) et* $\mathrm{Re}(s-1) > 0$:

$$\langle u_{\lambda}^{+}, \xi_{-\lambda} \rangle = -\frac{1}{s} \left(\int_{\mathscr{C}^-} \varphi(a)\, a^{-(\lambda+\rho)} da \right) \int_H \psi(h)\, dh$$

$$\langle u_{\lambda}^{-}, \xi_{-\lambda} \rangle = 0.$$

6.5 Fin de la démonstration du Lemme 13 en rang 1

Par comparaison des Lemmes 14 et 17 on a

$$B^{-}(\lambda) = -\frac{1}{s}, \qquad B^{+}(\lambda) = 0$$

soit encore:

$$A'(s_\alpha, \lambda)\, \xi^1_{-\lambda} = -\frac{2}{\lambda_\alpha} \xi^{s_\alpha}_{\lambda}$$

puis en changeant λ en $-\lambda = s_\alpha \lambda$

$$A'(s_\alpha, s_\alpha \lambda)\, \xi^1_{\lambda} = \frac{2}{\lambda_\alpha} \xi^{s_\alpha}_{s_\alpha \lambda} = \varepsilon(\alpha) \frac{2}{\lambda_\alpha} \xi^{s_\alpha}_{s_\alpha \lambda}$$

et ceci achève la démonstration du Lemme 13 dans le cas où G est de rang 1.

Notre démonstration du Théorème 2 est maintenant complète. □

7 Détermination des distributions Θ_λ

7.1

On rappelle que pour $\varphi \in C_c^{\infty}(G/H)$ on a posé $\Theta_\lambda(\varphi) = \langle \pi'_\lambda(\varphi)\, \xi_\lambda . \xi_{-\lambda} \rangle$ dès que ξ_λ et $\xi_{-\lambda}$ sont définis (ce qui arrive sur le complémentaire d'une famille localement finie d'hyperplans de $\mathfrak{a}'_{\mathbb{C}}$). La distribution Θ_λ est propre sous l'algèbre des opérateurs différentiels invariants sur G/H, $D(G/H)$ (pour le paramétre $-\lambda$, dans les identifications habituelles), d'après les propriétés des vecteurs ξ_λ (voir § 3.2) et invariante sous l'action à gauche de H. On sait, d'après [10, Th. 4.2], qu'une telle distribution est une fonction localement intégrable sur G/H, analytique sur l'ouvert des éléments réguliers de G/H ([10, Def. 2.5]) que l'on note $(G/H)'$. Si $E \subset G/H$, on note E' l'ensemble des éléments réguliers de E. Alors on a:

Théorème 3 *Pour les valeurs régulières de λ où ξ_λ et $\xi_{-\lambda}$ sont définies on a:*
(i) Soit $w \in W$. Soit $a \in A$. Si $w \notin W_H$ et awH est régulier dans G/H, on a

$$\Theta_\lambda(awH)=0.$$

(ii) Si aH est régulier dans G/H:

$$\Theta_\lambda(aH)=(-1)^{|\Delta_n^+|}2^{|\Delta^+|}\Big(\prod_{\alpha\in\Delta^+}(\lambda_\alpha)^{-1}\Big)\times\frac{\sum_{x\in W_H}(-1)^{\ell(x)}a^{-x\lambda}}{\prod_{\alpha\in\Delta^+}(a^{-\alpha}-a^{\alpha})},$$

où $\ell(x)$ est la longueur de x comme élément de W.

7.2

On commence la démonstration du Théorème 3. On introduit la fonction $\Delta_{\mathscr{A}}$ sur $\mathscr{A}$ comme dans [10, Formule (4.6)], qui est définie par:

$$\forall w\in W,\quad \forall a\in A,\quad \Delta_{\mathscr{A}}(waH)=(wa^2\sigma(w)^{-1})^{-\rho_{\mathbb{C}}}\prod_{\alpha\in\Delta^+}(1-(wa^2\sigma(w^{-1}))^{\alpha}).$$

Mais $\sigma(w)^{-1}=t_w w^{-1}$. On pose alors $a_w=waw^{-1}$, $t'_w=wt_w w^{-1}$. Alors $t'^2_w=1$ et:

$$\Delta_{\mathscr{A}}(waH)=t'^{-\rho_{\mathbb{C}}}_w a_w^{-2\rho_{\mathbb{C}}}\prod_{\alpha\in\Delta^+}(1-t'^{\alpha}_w a_w^{2\alpha})$$

$$\Delta_{\mathscr{A}}(waH)=t'^{\rho_{\mathbb{C}}}_w\prod_{\alpha\in\Delta^+}(a_w^{-\alpha}-t'^{\alpha}_w a_w^{\alpha}).$$

En particulier

$$\Delta_{\mathscr{A}}(aH)=\prod_{\alpha\in\Delta^+}(a^{-\alpha}-a^{\alpha}).$$

Par ailleurs, on sait d'après [10, Théorème 4.8], que, sous les hypothèses du théorème, la fonction $\Psi_{\mathscr{A}}$, définie sur $\mathscr{A}'$ par $\Delta_{\mathscr{A}}\Theta_\lambda$, se prolonge en une fonction analytique sur $\mathscr{A}$ (il n'y a pas de racines réelles de t dans $\mathfrak{h}$ et donc, avec les notations de [10], $\mathscr{A}'_{\mathbb{R}}=\mathscr{A}$).

En outre, le fait que Θ_λ soit propre dans $D(G/H)$ implique que, pour les valeurs régulières de λ, cette fonction peut s'écrire sous la forme:

$$(13)\qquad \forall a\in A,\quad \forall w\in W,\quad (\Delta_{\mathscr{A}}\Theta_\lambda)(waH)=\sum_{x\in W}d(w,x,\lambda)\,a^{-x\lambda}.$$

Clairement les $d(w,y,\lambda)$ sont méromorphes en λ et on va les calculer pour λ tel que $\mathrm{Re}(\lambda-\rho)$ soit strictement Δ^+-dominant, en supposant en outre que les fonctions méromorphes qui interviendront dans les calculs seront définies. Au bout du compte cela déterminera ces fonctions méromorphes sur un ouvert, donc partout. Pour cela on va calculer de deux façons différentes $\langle u^w_{\lambda,x},\xi_{-\lambda}\rangle$ (cf. définition de $u^w_{\lambda,x}$ au Théorème 1) (on posera $u^w_{\lambda,x}=0$ si $x\notin W_H$).

7.3 Premier calcul de $\langle u^w_{\lambda,x}, \xi_{-\lambda}\rangle$

D'après le Théorème 1 on a:

$$\langle u^w_{\lambda,x}, \xi_{-\lambda}\rangle = \frac{1}{C_G} \lim_{t\to+\infty} a_t^{-x(\lambda+\rho)} \Theta_\lambda(\Phi_t^w(\psi,\varphi)).$$

On va calculer $\Theta_\lambda(\Phi_t^w(\psi,\varphi))$ en utilisant (13). Comme Θ_λ est une fonction localement intégrable sur G/H, invariante à gauche par H, la formule (13) et la formule intégrale de la Proposition 3 montrent que:

$$\frac{1}{C_G}\Theta_\lambda(\Phi_t^w(\psi,\varphi)) = \int_{H/T} \psi(hT)\, d\dot{h} \int_{\mathscr{C}} \varphi_t(a) \frac{\sum_{y\in W} d(w,y,\lambda)\, a^{-y\lambda}}{\Delta_{\mathscr{A}}(waH)} D_w(a)\, da.$$

Soit encore:

$$\frac{1}{C_G}\Theta_\lambda(\Phi_t^w(\psi,\varphi))$$
$$= \int_{H/T} \psi(hT)\, d\dot{h} \times \int_{\mathscr{C}} \varphi(a) \frac{\sum_{y\in W} d(w,y,\lambda)\,(aa_{-t})^{-y\lambda}}{\Delta_{\mathscr{A}}(waa_{-t}H)} D_w(aa_{-t})\, da.$$

Pour $a\in\mathscr{C}$, on a vu au Lemme 8

$$D_w(aa_{-t}) \sim a_t^{2x\rho}\, a^{-2x\rho}.$$

De même, comme on suppose que $\mathrm{Re}(\lambda-\rho)$ est Δ^+-dominant et que $X_0\in x(C^+)$ on a

$$\lim_{t\to+\infty} a_t^{-x\lambda} \sum_{y\in W} d(w,y,\lambda)\,(aa_{-t})^{-y\lambda} = d(w,x,\lambda)\, a^{-x\lambda}.$$

Enfin

$$\Delta_{\mathscr{A}}(waa_{-t}H) \sim \varepsilon(w,x)\, a^{-x\rho} a_t^{x\rho}$$

où $\varepsilon(w,x)$ est un signe qui ne nous intéressera que lorsque $w=1$. Dans ce cas:

$$\Delta_{\mathscr{A}}(aH) = \prod_{\alpha\in\Delta^+} (a^{-\alpha}-a^{\alpha})$$
$$= (-1)^{|\{\alpha\in x\Delta^+ \,|\, \alpha\in-\Delta^+\}|} \prod_{\alpha\in x\Delta^+} (a^{-\alpha}-a^{\alpha}).$$

Mais de la définition de la longueur il vient

$$|\{\alpha\in x\Delta^+ \,|\, \alpha\in-\Delta^+\}| = \ell(x).$$

Donc

$$\Delta_{\mathscr{A}}(aH) = (-1)^{\ell(x)} \prod_{\alpha\in x\Delta^+} (a^{-\alpha}-a^{\alpha})$$

et l'on en déduit

$$\Delta_{\mathscr{A}}(aa_{-t}H) \sim (-1)^{\ell(x)} a^{-x\rho} a_t^{x\rho}.$$

D'où

$$\lim_{t\to+\infty} \varphi(a)\, a_t^{-x(\lambda+\rho)} \frac{\sum_{y\in W} d(w,y,\lambda)\,(aa_{-t})^{-y\lambda}}{\Delta_{\mathscr{A}}(waa_{-t}H)} D_w(aa_{-t})$$
$$= d(w,x,\lambda)\,\varepsilon(w,x)\,a^{-x(\lambda+\rho)}\varphi(a)$$

avec $\varepsilon(w,x)=\pm 1$ et $\varepsilon(1,x)=(-1)^{\ell(x)}$.

Il n'est pas difficile de voir que la convergence est dominée (voir démonstration du Lemme 8). Finalement on a obtenu:

Lemme 18 *Soit $\lambda\in\mathfrak{a}_{\mathbb{C}}'$, régulier, avec* $\mathrm{Re}(\lambda-\rho)$ strictement Δ^+*-dominant et tel que $\xi_{-\lambda}$ soit défini. Alors*:

$$\langle u_{\lambda,x}^w, \xi_{-\lambda}\rangle = \varepsilon(w,x)\,d(w,x,\lambda)\int_{H/T}\psi(hT)\,d\dot{h}\times\int_{\mathscr{C}}\varphi(a)\,a^{-x(\lambda+\rho)}\,da.$$

(*rappelons que $u_{\lambda,x}^w=0$ si $x\notin W_H$*).

7.4 Deuxième calcul de $\langle u_{\lambda,x}^w, \xi_{-\lambda}\rangle$

Lemme 19 *Pour λ dans un ouvert non vide de $\{\mu\in\mathfrak{a}_{\mathbb{C}}'\,|\,\mathrm{Re}(\mu-\rho)$ est Δ^+-dominant$\}$, on a, pour $x\in W_H$*:

(i) $\langle u_{\lambda,x}^w, \xi_{-\lambda}\rangle = 0$ *si* $w\notin W_H$

(ii) $\langle u_{\lambda,x}^1, \xi_{-\lambda}\rangle = B(\lambda)\int_{\mathscr{C}}\varphi(a)\,a^{-x(\lambda+\rho)}\,da\int_{H/T}\varphi(hT)\,d\dot{h}$ *où* $B(\lambda)=(-1)^{|\Delta_n^+|}\prod_{\alpha\in\Delta^+}\left(\frac{2}{\lambda_\alpha}\right)$.

Démonstration. On a vu (cf. Lemme 12 et ce qui précède) que $u_{\lambda,x}^w=A'(w_0,\lambda)\,v_{\lambda,x}^w$ avec $v_{\lambda,x}^w\in I_{-w_0\lambda}$. D'après la Remarque 3 on a donc:

$$u_{\lambda,x}^w = A(w_0,-w_0\lambda)\,v_{\lambda,x}^w.$$

Donc

$$\langle u_{\lambda,x}^w,\xi_{-\lambda}\rangle = \langle A(w_0,-w_0\lambda)\,v_{\lambda,x}^w,\xi_{-\lambda}\rangle$$
$$= \langle v_{\lambda,x}^w, A'(w_0,-w_0\lambda)\,\xi_{-\lambda}\rangle.$$

Mais d'après le Théorème 2 on a l'égalité de fonctions méromorphes

$$A'(w_0,-w_0\lambda)\,\xi_{-\lambda} = B(\lambda)\,\xi_{-w_0\lambda}^{w_0}$$

où

$$B(\lambda)=\prod_{\alpha\in\Delta^+}\varepsilon(\alpha)\left(\frac{-2}{\lambda_\alpha}\right)=(-1)^{|\Delta_n^+|}\prod_{\alpha\in\Delta^+}\left(\frac{2}{\lambda_\alpha}\right).$$

Donc:

$$\langle u_{\lambda,x}^w,\xi_{-\lambda}\rangle = B(\lambda)\,\langle v_{\lambda,x}^w,\xi_{-w_0\lambda}^{w_0}\rangle.$$

Mais on sait d'après le Lemme 12 que $v^w_{\lambda,x}$ est à support contenu dans $Hwxw_0AN$ et, par définition, $\xi^{w_0}_{-w_0\lambda}$ est à support dans Hw_0AN. Donc

$$\langle v^w_{\lambda,x}, \xi^{w_0}_{-w_0\lambda}\rangle = 0 \quad \text{si } Hwxw_0AN \neq Hw_0AN$$

i.e. (comme $x \in W_H$) si $w \notin W_H$. D'où la partie (i) du lemme.

On prend maintenant $w=1$. Alors

$$\langle u^1_{\lambda,x}, \xi_{-\lambda}\rangle = B(\lambda)\langle v^1_{\lambda,x}, \xi^{w_0}_{-w_0\lambda}\rangle.$$

Or, d'après le Lemme 12, $v^1_{\lambda,x}$ a son support inclus dans Hxw_0AN et

$$v^1_{\lambda,x}(hxw_0) = \frac{1}{c_0}\left(\int_{\mathscr{C}} \varphi(a)\, a^{-x(\lambda+\rho)}\, da\right)\psi(h).$$

On applique alors la formule intégrale de [13, Lemme 1.3] (avec la constante c_0: voir la démonstration du Lemme 12) et l'on a (puisque l'on peut travailler dans l'ouvert Hxw_0AN où $\xi^{w_0}_{-\lambda}$ est défini par une fonction)

$$\langle v^1_{\lambda,x}, \xi_{-\lambda}\rangle = \int_{\mathscr{C}} \varphi(a)\, a^{-x(\lambda+\rho)}\, da \int_{H/T} \psi(hT)\, d\dot{h}.$$

7.5 Fin de la démonstration du Théorème 3

D'après le Lemme 18 on voit que $d(w,x,\lambda)\equiv 0$ si $x \notin W_H$ (puisque $u^w_{\lambda,x}=0$ si $x \notin W_H$).

Maintenant, en comparant le Lemme 18 et le Lemme 19 (i), on voit aussi que $d(w,x,\lambda)=0$ si $w \notin W_H$ et $x \in W_H$. Donc $d(w,x,\lambda)=0$ pour tout $x \in W$, si $w \notin W_H$ et ceci prouve la partie (i) du Théorème 3. Maintenant, en comparant les Lemmes 18 et 19 (ii), on voit que:

$$d(1,x,\lambda) = (-1)^{\ell(x)} B(\lambda).$$

Donc, d'après la définition des fonctions d, on a:

$$\Delta_{\mathscr{A}}(aH)\,\Theta_\lambda(aH) = B(\lambda) \sum_{x \in W_H} (-1)^{\ell(x)} a^{-x\lambda}.$$

D'où, pour $aH \in (G/H)'$

$$\Theta_\lambda(aH) = B(\lambda)\frac{\sum_{x \in W_H} (-1)^{\ell(x)} a^{-x\lambda}}{\prod_{\alpha \in \Delta^+} (a^{-\alpha} - a^{\alpha})}$$

avec $B(\lambda) = (-1)^{|\Delta_n^+|} \prod_{\alpha \in \Delta^+} \left(\frac{2}{\lambda_\alpha}\right)$. □

Références

[1] van den Ban, E.: The principal series for a reductive symmetric space I. Ann. Sci. Éc. Norm. Supér. **21**, 359–412 (1988)

[2] Bernstein, I.N., Gelfand, S.I.: Meromorphic property of the functions P^{λ}, Funkts. Anal. Prilozh. **3**, 68–69 (1969)

[3] Bruhat, F.: Sur les représentations induites des groupes de Lie. Bull. Soc. Math. Fr. **84**, 97–205 (1956)

[4] Cartier, P.: Vecteurs différentiables dans les représentations unitaires des groupes de Lie, exposé au séminaire Bourbaki n° 454. (Lect. Notes Math., vol. 514) Berlin Heidelberg New York: Springer 1976

[5] Casselman, W.: Canonical extensions of Harish-Chandra modules to representations of *G*. (Preprint 1987)

[6] Duflo, M.: Représentations irréductibles des groupes semi-simples complexes. (Lect. Notes Math., vol. 497, pp. 26–87) Berlin Heidelberg New York: Springer 1975

[7] Erdelyi, A.: Higher transcendental functions, vol. I. New York Toronto: Mc Graw-Hill 1953

[8] Faraut, J.: Distributions sphériques sur les espaces hyperboliques. J. Math. Pures Appl. **58**, 369–414 (1979)

[9] Gelfand, I.M., Chilov, G.E.: Les distributions, vol. I. Paris: Dunod 1962

[10] Harinck, P.: Fonctions généralisées sphériques sur $G_{\mathbb{C}}/G_{\mathbb{R}}$. Ann. Sci. Éc. Norm. Supér. **23**, 1–38 (1990)

[11] Harish-Chandra: Harmonic analysis on real reductive groups I. J. Funct. Anal. **19**, 104–204 (1975)

[12] Knapp, A., Stein, E.: Intertwining operators for semi-simple Lie group II. Invent. Math. **60**, 9–84 (1980)

[13] Olafson, G.: Fourier and Poisson transformation associated to a semi-simple symmetric space. Invent. Math. **90**, 605–629 (1987)

[14] Oshima, T., Matsuki, T.: Orbits on affine symmetric spaces under the action of the isotropy subgroup. J. Math. Soc. Japan **32**, 399–414 (1980)

[15] Poulsen, N.: On C^{∞}-vectors and intertwining bilinear forms for representations of Lie groups. J. Funct. Anal. **9**, 87–120 (1972)

[16] Schwartz, L.: Théorie des distributions, vol. I. Paris: Hermann 1950

[17] Schwartz, L.: Sous-espaces hilbertiens et noyaux associés, applications aux représentations des groupes de Lie. In: Centre Belge de Recherches Mathématiques (ed.) Deuxième Colloque CBRM sur l'analyse fonctionnelle. Liège 1964, pp. 153–163. Paris: Gauthier-Villars 1964

[18] Wallach, N.: Harmonic analysis on homogeneous spaces. New York: Dekker 1973

[19] Warner, G.: Harmonic analysis on semi-simple Lie groups, vol. I. Berlin Heidelberg New York: Springer 1973

[20] Harinck, P.: Fonctions généralisées sphériques induites sur $G_{\mathbb{C}}/G_{\mathbb{R}}$ et applications. (Preprint)

[21] Sano, S.: Invariant spherical distributions and the Fourier inversion formula on $\mathrm{GL}(n, \mathbb{C})/\mathrm{GL}(n, \mathbb{R})$, J. Math. Soc. Japan **36**, 191–218 (1984)

Invent. math. 105, 347–373 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Rational points on K3 surfaces: A new canonical height

Joseph H. Silverman *

Brown University, Box 1917, Providence, RI 02912, USA

Oblatum 3-IX-1990 & 26-XI-1990

Introduction

A fundamental tenet of Diophantine Geometry is that the geometric properties of an algebraic variety should determine its basic arithmetic properties. This is certainly true for curves, where the sign of the Euler characteristic of C determines whether the set of rational points on C is finite ($\chi(C)<0$), a finitely generated group ($\chi(C)=0$), or parametrizable ($\chi(C)>0$). For higher dimensional varieties there are some precise conjectures due to Bombieri, Lang, and Vojta [15] which predict when the rational points on a variety should be finite or degenerate (i.e. not Zariski dense), and some conjectures of Manin et al. [2, 5] on the distribution of rational points in those cases when they are Zariski dense. But except for abelian varieties, their subvarieties, and some Fano varieties (varieties for which the anticanonical bundle is ample), there are very few general theorems.

In this paper we will study the rational points on a certain class of K3 surfaces defined over a number field K. The moduli space of marked algebraic K3 surfaces is a countable union of 19 dimensional quasi-projective varieties. We are going to look at the 18 dimensional family studied by Wehler [17]. Wehler's family consists of K3 surfaces S whose automorphism group $\mathrm{Aut}(S)$ contains a subgroup $\mathscr{A}$ isomorphic to the free product $\mathbb{Z}_2 * \mathbb{Z}_2$ of two cyclic groups of order 2. We will use the geometric information provided by this infinite automorphism group to study the K-rational points on S.

For any point $P \in S$, we can look at the orbit of P under the action of $\mathscr{A}$,

$$\mathscr{C} = \mathscr{C}(P) = \{\phi P : \phi \in \mathscr{A}\}.$$

We call such an orbit a *chain*. Then the study of the K-rational points on S is divided into two parts: (i) Describe the points in a given chain $\mathscr{C} \subset S(K)$. (ii) Describe the collection of chains in $S(K)$.

The chains themselves naturally separate into two sorts, those with finitely many elements and those with infinitely many elements. This is analogous to the points on an abelian variety, which generate either finite or infinite sub-

* Research partially supported by NSF DMS-8842154 and a Sloan Foundation Fellowship

groups. Our first main result says that $S(K)$ contains only finitely many finite chains. (Just as an abelian variety contains only finitely many points of finite order defined over any given number field.) Our second main result, which we describe more precisely below, gives the counting function for the number of points in a chain with height less than a given bound.

A basic tool in the study of Diophantine equations is the theory of height functions. On abelian varieties Néron and Tate developed a theory of canonical heights which have especially nice transformation properties. If A is an abelian variety and if $D \in \mathrm{Pic}(A)$ is a symmetric divisor class, then the canonical height $\hat{h}_D$ interacts with the group law on A via the formula $\hat{h}_D(nP) = n^2 \hat{h}_D(P)$. (See, e.g. [9].) In a similar way we will develop a theory of canonical heights on our K3 surface S which transforms canonically relative to the automorphism group $\mathscr{A}$. More precisely, we will prove the existence of two canonical heights $\hat{h}^+$ and $\hat{h}^-$ on S which satisfy the transformation formulas

$$\hat{h}^+(\phi P) = (2+\sqrt{3})^{\ell(\phi)} \hat{h}^{\chi(\phi)}(P)$$

and

$$\hat{h}^-(\phi P) = (2+\sqrt{3})^{-\ell(\phi)} \hat{h}^{-\chi(\phi)}(P)$$

for every $\phi \in \mathscr{A}$. Here $\ell(\phi) \in \mathbb{Z}$ and $\chi(\phi) \in \{\pm 1\}$ are functions that we will precisely describe in Sect. 2. The functions $\hat{h}^\pm$ have many properties similar to those enjoyed by canonical heights on abelian varieties. For example, $\hat{h}^\pm(P) \geqq 0$ for all points $P \in S(\bar{K})$, and

$$\hat{h}^+(P) = 0 \Leftrightarrow \hat{h}^-(P) = 0 \Leftrightarrow \mathscr{C}(P) \text{ is finite.}$$

In order to draw arithmetic conclusions, one ultimately must relate everything to heights relative to a projective embedding. So we also show that the function

$$\hat{h} = \tfrac{1}{2}(\sqrt{3}-1)(\hat{h}^+ + \hat{h}^-)$$

is a Weil height corresponding to a very ample divisor. These and other properties of the canonical heights $\hat{h}^+$ and $\hat{h}^-$ are described in Theorem 1.1 in Sect. 1.

From the transformation formulas it is clear that the product $\hat{h}^+ \hat{h}^-$ takes the same value at P and at ϕP for any automorphism $\phi \in \mathscr{A}$. This allows us to define a canonical height of a chain $\hat{H}(\mathscr{C}) = \hat{h}^+(P)\hat{h}^-(P)$ which can be calculated by choosing any point P in the chain. Notice that $\hat{H}(\mathscr{C}) = 0$ if and only if the chain $\mathscr{C}$ is finite; it is this observation which ultimately enables us to prove that $S(K)$ contains only finitely many finite chains. In the case that $\hat{H}(\mathscr{C}) > 0$, its value determines the arithmetic complexity of the chain. For example, $\hat{H}(\mathscr{C})$ appears in the following counting formula which we will prove in Sect. 4:

$$\#\{P \in \mathscr{C} : \hat{h}(P) \leqq B\} = \kappa \log(B^2/\hat{H}(\mathscr{C})) + O(1) \quad \text{as } B \to \infty.$$

Here $\kappa = \kappa(\mathscr{C}) = 0.3796\ldots$ or $0.7593\ldots$ depending on whether the chain $\mathscr{C}$ is one or two sided, and the big-O constant is absolute (it is even independent of the surface S.)

The organization of this paper is as follows. We begin in Sect. 1 by setting notation and stating our main results. We also raise several questions for our K3 surfaces which are analogous to theorems and conjectures on abelian varie-

ties. Section 2 is devoted to geometry. We determine how the automorphisms in $\mathscr{A}$ act on $\mathrm{Pic}(S)$, and use this information to draw various conclusions about chains, ample and effective divisors, and curves of low genus. In Sect. 3 we prove the existence and basic properties of the canonical height, and in Sect. 4 we apply our results to bound the number of finite chains and count the number of points in infinite chains. To illustrate our general theory, we present a numerical example in Sect. 5, including a list of the first few points in a particular infinite chain. This list has an unexpected arithmetic property. In the final section we show how a strong form of Vojta's conjecture [15] can be used to explain this property.

1 Notation and the main theorems

In this section we will state our main results and raise a number of interesting questions meriting further study. We begin by setting some notation which will remain fixed throughout this paper.

K a number field.

S/K a smooth surface contained in $\mathbb{P}^2 \times \mathbb{P}^2$ given by the intersection of two effective divisors, one of type (1, 1) and one of type (2, 2). In other words, S is the locus described by two equations of the form

$$\sum_{i,j=1}^{3} a_{ij} x_i y_j = \sum_{i,j,k,l=1}^{3} b_{ijkl} x_i x_j y_k y_l = 0$$

p_1, p_2 the projections $p_j: S \to \mathbb{P}^2$ induced by the natural projections

$$p_j: \mathbb{P}^2 \times \mathbb{P}^2 \to \mathbb{P}^2.$$

D_1, D_2 $\in \mathrm{Pic}(S)$, where D_j is the divisor class of $p_j^* H$, where $H \in \mathrm{Pic}(\mathbb{P}^2)$ is a hyperplane section.

σ_1, σ_2 $\in \mathrm{Aut}(S)$, where σ_j is the involution of S induced by the double cover $p_j: S \to \mathbb{P}^2$.

$\mathscr{A}$ $\subset \mathrm{Aut}(S)$. The subgroup of $\mathrm{Aut}(S)$ generated by σ_1 and σ_2.

α $= 2 + \sqrt{3}$.

E^+, E^- $\in \mathrm{Pic}(S) \otimes \mathbb{R}$, given by the formulas

$$E^+ = \alpha D_1 - D_2, \qquad E^- = -D_1 + \alpha D_2.$$

The surface S is a K3 surface. By varying the coefficients of the polynomials defining S, one can produce an 18 dimensional family of such surfaces (up to isomorphism over $\mathbb{C}$). For basic facts about K3 surfaces, see [3, 6]. This particular family of surfaces is studied by Wehler in [17]; he shows that the general member of this family has automorphism group exactly equal to $\mathscr{A}$. For our purposes it is enough to have $\mathscr{A}$ as a subgroup of $\mathrm{Aut}(S)$, so we will not need to worry about the fact that Wehler's proof does not guarantee the existence of even a single $S/\bar{\mathbb{Q}}$ with $\mathrm{Aut}(S) = \mathscr{A}$. (This is a failing common to all theorems in algebraic geometry which assert that the general member of a family has a certain property. For example, the Noether-Lefschetz theorem [7, 10] and the Dimension and Smoothness theorems in Brill-Noether theory [1, p. 214] have this form. Note that the term "general" means except for countably many proper Zariski closed subsets of the relevant moduli space. Since

$\bar{\mathbb{Q}}$ itself is countable, there is nothing to prevent every $\bar{\mathbb{Q}}$ point in the moduli space from being deleted!).

Our first result says that there are height functions $\hat{h}^+$ and $\hat{h}^-$ on $S(\bar{K})$ which transform "canonically" relative to the group of automorphisms $\mathscr{A}$.

Theorem 1.1 *There is a unique pair of functions*

$$\hat{h}^+, \hat{h}^- : S(\bar{K}) \to \mathbb{R}$$

satisfying the following two properties:

(i) $$\hat{h}^{\pm} = h_{E^{\pm}} + O(1).$$

(ii) $$\hat{h}^{\pm} \circ \sigma_1 = \alpha^{\mp 1}\, \hat{h}^{\mp}.$$
$$\hat{h}^{\pm} \circ \sigma_2 = \alpha^{\pm 1}\, \hat{h}^{\mp}.$$

The functions $\hat{h}^+$ and $\hat{h}^-$ have the following additional properties:

(iii) $$\hat{h}^{\pm} \circ \phi = \alpha^{\pm \ell(\phi)}\, \hat{h}^{\pm \chi(\phi)} \quad \text{for all } \phi \in \mathscr{A}.$$

[*Here $\chi: \mathscr{A} \to \{\pm 1\}$ and $\ell: \mathscr{A} \to \mathbb{Z}$ are functions we will describe in Sect. 2; and $\hat{h}^{\pm 1}$ is an alternative notation for $\hat{h}^{\pm}$.*]

(iv) *Define* $$\hat{h} = \hat{h}^+ + \hat{h}^-.$$

Then $\hat{h}$ is a Weil height function for the ample divisor class $(\alpha - 1)(D_1 + D_2)$.

(v) *The function $\hat{h}^+ \hat{h}^-$ is $\mathscr{A}$-invariant. That is,*

$$\hat{h}^+(\phi P)\hat{h}^-(\phi P) = \hat{h}^+(P)\hat{h}^-(P) \quad \text{for all } \phi \in \mathscr{A} \quad \text{and all} \quad P \in S(\bar{K}).$$

(vi) $$\hat{h}^{\pm}(P) \geqq 0 \quad \text{for all } P \in S(\bar{K}).$$

(vii) *Let $P \in S(\bar{K})$. Then*

$$\hat{h}^+(P) = 0 \Leftrightarrow \hat{h}^-(P) = 0 \Leftrightarrow \{\phi P : \phi \in \mathscr{A}\} \text{ is finite.}$$

The automorphisms in $\mathscr{A}$ act on the points of S. We will call the orbit of a point P a chain, since it can be pictured as

$$\ldots \overset{\sigma_2}{\leftrightarrow} \sigma_1\sigma_2 P \overset{\sigma_1}{\leftrightarrow} \sigma_2 P \overset{\sigma_2}{\leftrightarrow} P \overset{\sigma_1}{\leftrightarrow} \sigma_1 P \overset{\sigma_2}{\leftrightarrow} \sigma_2\sigma_1 P \overset{\sigma_1}{\leftrightarrow} \sigma_1\sigma_2\sigma_1 P \overset{\sigma_2}{\leftrightarrow} \ldots$$

Of course, such a chain may loop back on itself.

Definition. A *chain* $\mathscr{C} \subset S(\bar{K})$ is a set of points of the form

$$\mathscr{C}(P) = \{\phi P : \phi \in \mathscr{A}\}$$

for some $P \in S(\bar{K})$. We will say that a chain $\mathscr{C}$ is *K-rational* if $\mathscr{C} \subset S(K)$. We define the (*canonical*) *height of a chain* $\mathscr{C}$ to be

$$\hat{H}(\mathscr{C}) = \hat{h}^+(P)\hat{h}^-(P) \quad \text{for any } P \in \mathscr{C}.$$

(Note that Theorem 1.1(v) says that $\hat{H}(\mathscr{C})$ is independent of the choice of $P \in \mathscr{C}$.)

Since $S(K)$ is a disjoint union of chains, we can break up the study of $S(K)$ into two questions:

What do the points in a K-rational chain look like?

What does the set of K-rational chains look like?

We will be able to answer the first question quite well. For the second, we will be able to limit the number of finite K-rational chains; but at present we do not have any good description of the infinite K-rational chains. We begin with our description of the finite chains.

Theorem 1.2 (a) *$\mathscr{C}$ is finite $\Leftrightarrow \hat{H}(\mathscr{C})=0 \Leftrightarrow \hat{h}(P)=0$ for all $P\in\mathscr{C}$.*

(b) *For any constant B, the set*

$$\{\text{chains } \mathscr{C}\subset S(K)\colon \hat{H}(\mathscr{C})\leqq B\}$$

is finite. In particular, there are only finitely many chains $\mathscr{C}\subset S(K)$ with $\#\mathscr{C}<\infty$.

Remark. We can actually prove something a little bit stronger concerning the set of finite chains. For any $P\in S(\bar{K})$, let $\mathscr{C}(P)$ denote the chain containing P. Then we can prove that

$$\{P\in S(\bar{K})\colon \#\mathscr{C}(P)<\infty\}$$

is a set of bounded height. Thus this set contains only finitely many points defined over fields of bounded degree.

Next we investigate the heights of the points in a given infinite chain. We give an estimate for the point of smallest height and count the number of points with height less than a given bound.

Theorem 1.3 *Let $\mathscr{C}\subset S(K)$ be a chain with $\#\mathscr{C}=\infty$.*

(a)

$$2\sqrt{\hat{H}(\mathscr{C})}\leqq \min_{P\in\mathscr{C}} \hat{h}(P)\leqq 2\alpha\sqrt{\hat{H}(\mathscr{C})}.$$

(b) *Let*

$$\mu(\mathscr{C})=\#\{\sigma\in\mathscr{A}\colon \sigma Q=Q\}$$

be the order of the stabilizer of Q for any point $Q\in\mathscr{C}$, so $\mu(\mathscr{C})$ equals 1 or 2. If $B^2\geqq 4\hat{H}(\mathscr{C})$, then

$$\left|\#\{P\in\mathscr{C}\colon \hat{h}(P)\leqq B\}-\frac{1}{\mu(\mathscr{C})}\log_\alpha\frac{B^2}{4\hat{H}(\mathscr{C})}\right|\leqq 4.$$

(Note that from (a), *if $B^2<4\hat{H}(\mathscr{C})$, then no point $P\in\mathscr{C}$ satisfies $\hat{h}(P)\leqq B$.*)

(c) *For any ample divisor $D\in\mathrm{Pic}(S)$,*

$$\#\{P\in\mathscr{C}\colon h_D(P)\leqq B\}=\frac{1}{\mu(\mathscr{C})}\log_\alpha\frac{B^2}{\hat{H}(\mathscr{C})}+O(1)\quad \text{as } B\to\infty.$$

The $O(1)$ constant depends only on D and the choice of Weil height function h_D.

Using Theorem 1.3, we can reduce the question of counting points in $S(K)$ to the question of counting rational chains.

Theorem 1.4 *Let*

$$S(K)_{\text{fin}} = \{P \in S(K) : \hat{h}(P) = 0\} = \{P \in S(K) : \mathscr{C}(P) \text{ is finite}\},$$

and let $\mu(\mathscr{C})$ *be as in Theorem 3. Then*

$$\#\{P \in S(K) : \hat{h}(P) \leqq B\} = \#S(K)_{\text{fin}} + \sum_{\substack{\mathscr{C} \subset S(K) \\ 0 < 4\hat{H}(\mathscr{C}) \leqq B^2}} \left\{ \frac{1}{\mu(\mathscr{C})} \log_\alpha \frac{B^2}{4\hat{H}(\mathscr{C})} + \varepsilon(\mathscr{C}) \right\},$$

where $\#S(K)_{\text{fin}}$ *is finite, and* $|\varepsilon(\mathscr{C})| \leqq 4$ *for all chains* $\mathscr{C}$.

We have defined a K-rational chain to be a chain whose points are all defined over K. As the referee has pointed out, it is also natural to look at those chains $\mathscr{C}$ which are Galois invariant. We can use the canonical height to quickly give a very explicit description of all such chains.

Corollary 1.5 *Let* $P \in S(\bar{K})$ *be a point whose chain* $\mathscr{C}(P)$ *is stable under the action of* $\text{Gal}(\bar{K}/K)$. *Then* P *satisfies one of the following three conditions.*

(i) $P \in S(K)$. (*I.e.* $\mathscr{C}(P)$ *is* K*-rational.*)

(ii) $P \in S_{\text{fin}}$. (*I.e.* $\mathscr{C}(P)$ *is finite.*)

(iii) $[K(P) : K] = 2$. *More precisely, there is a* $\tau \in \mathscr{A}$ *and a* $j \in \{1, 2\}$ *so that* $p_j(\tau P) \in \mathbb{P}^2(K)$. *Further, in this case there are no non-trivial elements of* $\mathscr{A}$ *that fix* P.

The results stated above raise a number of interesting questions which are very much analogous to problems which have been extensively studied on abelian varieties.

Question 1 (K 3 Analogue of Néron's Theorem) Describe the counting function for rational chains,

$$\#\{\mathscr{C} \subset S(K) : \hat{H}(\mathscr{C}) \leqq B\},$$

as $B \to \infty$. Bogomolov and Mumford (cf. the appendix to [11]) have shown that a K 3 surface always contains (singular) curves of genus 0 and 1, so this number will grow rapidly, at least if K is large enough. Batyrev and Manin [2] have a precise conjecture which predicts that "most" of the rational points on S will lie on curves of genus 0. However, it is an easy consequence of a result of Wehler [18] that a general surface of this type contains no smooth curves of genus less than 2. (See Corollary 2.6.) It seems to be a difficult question to determine all of the singular curves of genus 0 and 1. We define

$$S(K)^* = \{P \in S(K) : P \text{ lies on no curve of genus 0 or 1 in } S\},$$

and ask for the order of growth of

$$\#\{\mathscr{C} \subset S(K)^* : \hat{H}(\mathscr{C}) \leqq B\}.$$

(A bold conjecture of Bogomolov (cf. [2, p. 35]) asserts that $S(\bar{K})^* = \emptyset$, which would make the answer to our last question very easy!)

Question 2 (K 3 Analogue of Torsion Conjecture) Is there a bound for $S(K)_{\text{fin}}$ that depends only on K, independent of the choice of K 3 surface $S/K \subset \mathbb{P}^2 \times \mathbb{P}^2$?

Question 3 (K3 Analogue of Raynaud's Theorem) Let $C \subset S$ be an irreducible curve which is not fixed by any non-trivial element of $\mathscr{A}$. Is it true that $C \cap S_{\text{fin}}$ is finite? (Here S_{fin} is an abreviation for $S(\bar{K})_{\text{fin}}$, the set of all points of S having finite chains.)

We observe that if C_1 and C_2 denote the branch loci of the projections p_1 and p_2 respectively, then for any $\phi \in \mathscr{A}$, every point in $C_1 \cap \phi C_2$ will be in S_{fin}. Now varying ϕ will undoubtedly yield infinitely many points in $C_1 \cap S_{\text{fin}}$. This is why we must restrict ourselves to curves not fixed by elements of $\mathscr{A}$.

Question 4 (K3 Analogue of Lang Conjecture) Is there a positive lower bound for $\hat{h}(P)$, independent of S, valid for points $P \in S(K) \setminus S(K)_{\text{fin}}$? More generally, let $h(S)$ be the height of the point in moduli space corresponding to S, where we fix a projective embedding of the moduli space. Is there a constant $c > 0$, depending only on K, so that for all S/K,

$$\hat{h}(P) \geqq c h(S) \quad \text{for all } P \in S(K),\ P \notin S(K)_{\text{fin}}?$$

Question 5 (K3 Analogue of Lehmer Conjecture) Are there constants $c = c(S) > 0$ and $d = d(S)$ so that for all extension fields L/K,

$$\hat{h}(P) \geqq \frac{c}{[L:K]^d} \quad \text{for all } P \in S(L),\ P \notin S(L)_{\text{fin}}?$$

More precisely, can one take $d = \frac{1}{2}$, as is expected to be true for abelian surfaces?

Question 6 (K3 Analogue of Serre's "Image of Galois" Theorem) Let

$$\mathscr{A}_P = \{\phi \in \mathscr{A} : \phi P = P\}$$

denote the stabilizer of a point $P \in S$; and for any subgroup $\mathscr{B} \subset \mathscr{A}$, let

$$S[\mathscr{B}] = \{P \in S(\bar{K}) : \mathscr{A}_P = \mathscr{B}\}.$$

Thus if $P \in S[\mathscr{B}]$, then there is a bijection $\mathscr{A}/\mathscr{B} \leftrightarrow \mathscr{C}(P)$.

Assume now that $[\mathscr{A} : \mathscr{B}] < \infty$. Then every element in $S[\mathscr{B}] \cap S(K)$ generates a finite K-rational chain, so Theorem 1.2(b) implies that $S[\mathscr{B}] \cap S(K)$ is a finite set. The question we pose is to describe the Galois group $\text{Gal}(K(S[\mathscr{B}])/K)$ as a subgroup of the symmetric group of $S[\mathscr{B}]$. In particular, is the index of $\text{Gal}(K(S[\mathscr{B}])/K)$ in the symmetric group of $S[\mathscr{B}]$ bounded independently of $\mathscr{B}$, subject always to the assumption that $[\mathscr{A} : \mathscr{B}] < \infty$?

2 The geometry of the K3 surface S

Let $S \subset \mathbb{P}^2 \times \mathbb{P}^2$ be a smooth surface as described in Sect. 1. Then the projections $p_j : S \to \mathbb{P}^2$ each exhibit S as a double cover of $\mathbb{P}^2$ branched along a smooth sextic curve. The subgroup $\mathscr{A}$ of $\text{Aut}(S)$ generated by σ_1 and σ_2 is isomorphic to the free product $\mathbb{Z}_2 * \mathbb{Z}_2$ of two groups of order 2. (See [17].) In this section we are going to study the geometry of S. We begin with a simple geometric calculation

Lemma 2.1 (a)

$$\sigma_1^* D_1 = D_1, \qquad \sigma_1^* D_2 = 4D_1 - D_2,$$

$$\sigma_2^* D_1 = -D_1 + 4D_2, \qquad \sigma_2^* D_2 = D_2,$$

(b)

$$\sigma_1^* E^+ = \alpha^{-1} E^-, \qquad \sigma_1^* E^- = \alpha E^+,$$

$$\sigma_2^* E^+ = \alpha E^-, \qquad \sigma_2^* E^- = \alpha^{-1} E^+.$$

Proof. (a) First, letting H be a generator of $\mathrm{Pic}(\mathbb{P}^2)$, we find that

$$\sigma_j^* D_j = \sigma_j^* p_j^* H = (p_j \sigma_j)^* H = p_j^* H = D_j.$$

This proves two of the desired equalities.

Next, from the definition of σ_j, we see that

$$p_j^*(p_j P) = (P) + (\sigma_j P).$$

(This is an equality of zero-cycles.) Hence for any $D \in \mathrm{Pic}(S)$,

$$p_j^* p_{j*} D = D + \sigma_{j*} D = D + \sigma_j^* D.$$

(Since σ_j is an involution, we have $\sigma_{j*} = \sigma_j^*$.) Now we let $H_1 = H \times \mathbb{P}^2$ and $H_2 = \mathbb{P}^2 \times H$ be a basis for $\mathrm{Pic}(\mathbb{P}^2 \times \mathbb{P}^2)$ and compute an intersection number:

$$\begin{aligned}
(p_{2*} D_1) \cdot H &= (p_{2*} p_1^* H) \cdot H \\
&= (p_1^* H) \cdot (p_2^* H), \quad \text{by "push-pull" formula,} \\
&= S \cdot H_1 \cdot H_2, \quad \text{intersection in } \mathbb{P}^2 \times \mathbb{P}^2, \\
&= (H_1 + H_2) \cdot (2H_1 + 2H_2) \cdot H_1 \cdot H_2, \quad \text{since } S \text{ is the intersection of a (1, 1) form and a (2, 2) form,} \\
&= 4.
\end{aligned}$$

Since $\mathrm{Pic}(\mathbb{P}^2) = \mathbb{Z}H$ and $H^2 = 1$, we see that $p_{2*} D_1 = 4H$. Substituting above gives

$$\sigma_2^* D_1 = p_2^* p_{2*} D_1 - D_1 = p_2^*(4H) - D_1 = 4D_2 - D_1.$$

This is the third equality, and the fourth is proven similarly.

(b) These all follow from (a), the definition of $E^\pm$, and the fact that $4 - \alpha = \alpha^{-1}$. For example,

$$\sigma_1^* E^+ = \sigma_1^*(\alpha D_1 - D_2) = \alpha D_1 - (4D_1 - D_2) = -\alpha^{-1} D_1 + D_2 = \alpha^{-1} E^-.$$

The others are similar. □

It will actually be useful to have a complete description of how the automorphisms in $\mathscr{A}$ act on E^+ and E^-. To do this we need to describe two maps. The first is the unique character

$$\chi: \mathscr{A} \to \{\pm 1\}$$

of order 2 on $\mathscr{A}$. This is most simply defined by

$$\chi(\sigma_{i_1}\sigma_{i_2}\sigma_{i_3}\ldots\sigma_{i_n})=(-1)^n.$$

The kernel of χ is the cyclic subgroup of $\mathscr{A}$ generated by $\sigma_1\sigma_2$.

The other function

$$\ell\colon \mathscr{A}\to\mathbb{Z}$$

is defined inductively by the rules:

$$\ell(e)=0,\quad \ell(\sigma_1\phi)=-1-\ell(\phi),\quad \ell(\sigma_2\phi)=1-\ell(\phi).$$

The following brief table should make the definition clear, while the subsequent lemma explains why it is natural to look at function ℓ.

ϕ	...	$\sigma_1\sigma_2\sigma_1$	$\sigma_1\sigma_2$	σ_1	e	σ_2	$\sigma_2\sigma_1$	$\sigma_2\sigma_1\sigma_2$	...
$\ell(\phi)$	...	-3	-2	-1	0	1	2	3	...
$\chi(\phi)$	...	-1	1	-1	1	-1	1	-1	...

Proposition 2.2 (a) *The map $\ell\colon \mathscr{A}\to\mathbb{Z}$ is a bijection of sets.*

(b)
$$\ell(\phi\psi)=\ell(\phi)+\chi(\phi)\ell(\psi)\quad \textit{for all } \phi,\psi\in\mathscr{A}.$$

Thus ℓ represents a (non-trivial) element of the cohomology group $H^1(\mathscr{A},\mathbb{Z})$, where we make $\mathbb{Z}$ into an $\mathscr{A}$-module via the action $\chi\colon \mathscr{A}\to\{\pm1\}=\mathrm{Aut}(\mathbb{Z})$.

(c)
$$\phi^* E^{\pm}=\alpha^{\pm\ell(\phi)}E^{\pm\chi(\phi)}.$$

(*Here $E^{\pm 1}$ means the same thing as $E^{\pm}$.*)

Proof. (a) This is clear from the table; or more rigorously, one can easily check by induction that if $i_1, i_2, \ldots, i_n$ is an alternating sequence of 1's and 2's, then

$$\ell(\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_n})=\begin{cases} n, & \text{if } i_1=2;\\ -n, & \text{if } i_1=1.\end{cases}$$

(b) This, too, is easily checked by induction. It is clearly true if $\phi=e$. Assume now it is true for ϕ; we need to check it for $\sigma_1\phi$ and $\sigma_2\phi$. We will do the former; the latter is done similarly.

$$\begin{aligned}\ell(\sigma_1\phi\psi)&=-1-\ell(\phi\psi),\quad \text{definition of }\ell,\\ &=-1-\{\ell(\phi)+\chi(\phi)\ell(\psi)\},\quad \text{induction hypothesis},\\ &=\ell(\sigma_1\phi)-\chi(\phi)\ell(\psi),\quad \text{definition of }\ell,\\ &=\ell(\sigma_1\phi)+\chi(\sigma_1\phi)\ell(\psi),\quad \text{since }\chi(\sigma_1)=-1.\end{aligned}$$

We have just verified the cocycle relation, so ℓ represents an element of the cohomology group $H^1(\mathscr{A},\mathbb{Z})$. To see it is non-trivial, we note that $\ker(\chi)\cong\mathbb{Z}$

via the identification $\sigma_2\sigma_1 \leftrightarrow 1$, and that $\ker(\chi)$ acts trivially on $\mathbb{Z}$. Hence under the restriction map

$$H^1(\mathscr{A}, \mathbb{Z}) \xrightarrow{\text{res}} H^1(\ker(\chi), \mathbb{Z}) \cong \text{Hom}(\mathbb{Z}, \mathbb{Z}) \cong \mathbb{Z},$$

the image of ℓ is $\ell(\sigma_2\sigma_1)=2$; so ℓ represents a non-trivial cohomology class.
(c) Again the proof is by induction. The desired result is clearly true for $\phi = e$. Assuming it is true for ϕ, we must check it for $\sigma_1\phi$ and $\sigma_2\phi$. Thus

$$\begin{aligned}
(\sigma_1\phi)^* E^{\pm} &= \phi^*(\sigma_1^* E^{\pm}) \\
&= \phi^*(\alpha^{\mp 1} E^{\mp}), \quad \text{from Lemma 5.1(b)}, \\
&= \alpha^{\mp 1}(\alpha^{\mp \ell(\phi)} E^{\mp \chi(\phi)}), \quad \text{induction hypothesis}, \\
&= \alpha^{\mp(1+\ell(\phi))} E^{\mp \chi(\phi)} \\
&= \alpha^{\pm \ell(\sigma_1\phi)} E^{\pm \chi(\sigma_1\phi)}, \quad \text{definition of } \ell \text{ and } \chi.
\end{aligned}$$

The verification for $\sigma_2\phi$ is done similarly. □

We are now going to describe several geometric properties of our surface S. These results are relevant to our discussion in Sect. 1, and should prove useful in further studying the arithmetic of S. However, they are not actually needed for the construction and applications of the canonical height, so the reader who is mainly interested in arithmetic results can skip directly to Sect. 3.

As a first geometric application of Prop. 2.2, we show that an infinite chain is always Zariski dense in S.

Corollary 2.3 *Let $\mathscr{C} \subset S$ be a chain with $\#\mathscr{C} = \infty$. Then $\mathscr{C}$ is Zariski dense in S.*

Proof. Suppose that $\mathscr{C}$ is not Zariski dense, and let Γ be the one-dimensional part of the Zariski closure of $\mathscr{C}$. Since $\#\mathscr{C} = \infty$, we know that $\Gamma \neq \emptyset$. Write

$$\Gamma = \bigcup_{i=1}^{n} C_i$$

as a union of irreducible curves $C_i \subset S$. By construction, $\#(\mathscr{C} \cap C_i) = \infty$ for each $1 \leqq i \leqq n$; and the C_i's are the only irreducible curves on S with this property.

For any $\phi \in \mathscr{A}$, we have $\phi\mathscr{C} = \mathscr{C}$. Since ϕ is one-to-one, it follows that

$$\infty = \#(\mathscr{C} \cap C_i) = \#(\phi\mathscr{C} \cap C_i) = \#(\mathscr{C} \cap \phi^{-1} C_i).$$

Hence $\phi^{-1} C_i$ is one of the C_j's. In other words, we get a map (clearly a homomorphism) from $\mathscr{A}$ to the symmetric group $\mathscr{S}_n$,

$$\pi: \mathscr{A} \to \mathscr{S}_n, \qquad \text{determined by } \phi(C_i) = C_{\pi_\phi(i)}$$

The kernel of π has finite index in $\mathscr{A}$, so we can find an integer $k \geqq 1$ such that $\pi((\sigma_2\sigma_1)^k) = e$.

In particular, we see that $(\sigma_2\sigma_1)^k C_1 = C_1$. On the other hand, when we compute intersection indices using Prop. 2.2, we find

$$\begin{aligned} E^{\pm} \cdot C_1 &= E^{\pm} \cdot (\sigma_2\sigma_1)^k C_1 \\ &= ((\sigma_2\sigma_1)^k)^* E^{\pm} \cdot C_1 \\ &= \alpha^{\pm 2k} E^{\pm} \cdot C_1. \end{aligned}$$

Since $k \neq 0$ and $\alpha = 2 + \sqrt{3}$, we conclude that $E^+ \cdot C_1 = E^- \cdot C_1 = 0$. Hence

$$0 = C_1 \cdot (E^+ + E^-) = (\alpha - 1)(C_1 \cdot (D_1 + D_2)).$$

But $D_1 + D_2$ is ample and C_1 is a curve, so this is a contradiction. Therefore $\mathscr{C}$ is Zariski dense in S. □

Remark. We just proved that an infinite chain $\mathscr{C} \subset S$ is Zariski dense in S. If $\mathscr{C} \subset S(\mathbb{Q})$, it is natural to ask how $\mathscr{C}$ is distributed in $S(\mathbb{R})$. The map that takes a point $P \in S(\mathbb{R})$ to its chain $\mathscr{C}(P) \subset S(\mathbb{R})$ is a (non-linear) dynamical system which certainly merits further study. Preliminary computer calculations reveal some interesting patterns, but at present we have nothing definitive to say.

It is an easy exercise to classify the subgroups of $\mathscr{A}$; in particular, the subgroups of finite index are precisely the subgroups which contain $(\sigma_2\sigma_1)^k$ for some non-zero integer k. So a point $P \in S$ will generate a finite chain (i.e. in our earlier notation, $P \in S_{\text{fin}}$) if an only if it is fixed by some $(\sigma_2\sigma_1)^k$. We can use Prop. 2.2 to show that each $(\sigma_2\sigma_1)^k$ has only a finite number of fixed points. More generally, we prove the following.

Corollary 2.4 *For any $\phi \in \mathscr{A}$, let*

$$\mathscr{F}_\phi = \{P \in S : \phi P = P\}$$

be the set of fixed points of ϕ. Also let $C_i \subset S$ be the ramification locus of p_i: $S \to \mathbb{P}^2$ for $i = 1, 2$.

(a) *If $\phi = \tau\sigma_i\tau^{-1}$ for some $\tau \in \mathscr{A}$, then $\mathscr{F}_\phi = \tau C_i$.*

(b) *If $\phi = (\sigma_1\sigma_2)^k$ for some $k \in \mathbb{Z}$, $k \neq 0$, then $\mathscr{F}_\phi$ is finite.*

Proof. (a) The definition of the σ_i's shows immediately that $\mathscr{F}_{\sigma_i} = C_i$. Hence

$$\begin{aligned} P \in \mathscr{F}_\phi &\Leftrightarrow \tau\sigma_i\tau^{-1} P = P \Leftrightarrow \sigma_i(\tau^{-1} P) = \tau^{-1} P \\ &\Leftrightarrow \tau^{-1} P \in \mathscr{F}_{\sigma_i} = C_i \Leftrightarrow P \in \tau C_i. \end{aligned}$$

(b) Suppose $\#\mathscr{F}_\phi = \infty$. Since $\mathscr{F}_\phi$ is a Zariski closed subset of S, it follows that there is an irreducible curve $C \subset \mathscr{F}_\phi$. Then $\phi|_C$ is the identity map on C, so $\phi^* C = C$. Since ϕ has degree 1, it follows that for any divisor $D \in \mathrm{Div}(S)$,

$$C \cdot D = \phi^* C \cdot \phi^* D = C \cdot \phi^* D.$$

Taking $D = E^{\pm}$ and $\phi = (\sigma_1\sigma_2)^k$, we observe that $\chi(\phi) = 1$, so

$$C \cdot E^{\pm} = C \cdot \phi^* E^{\pm} = C \cdot (\alpha^{\pm \ell(\phi)} E^{\pm \chi(\phi)}) = C \cdot (\alpha^{\pm \ell(\phi)} E^{\pm}).$$

Hence

$$(1 - \alpha^{\pm \ell(\phi)})(C \cdot E^{\pm}) = 0.$$

Now $\phi=(\sigma_1\sigma_2)^k\neq e$, so $\ell(\phi)\neq 0$, which means that $\alpha^{\pm\ell(\phi)}\neq 1$. Therefore $C\cdot E^+ = C\cdot E^- = 0$. Now we obtain a contradiction by observing that C is effective and that $E^+ + E^- = (\alpha-1)(D_1+D_2)$ is in the ample cone of $\mathrm{Pic}(S)\otimes\mathbb{R}$. Therefore $\mathscr{F}_\phi$ contains no curves, so it is a finite set of points. □

Remark. For any subgroup $\mathscr{B}\subset\mathscr{A}$, let $S[\mathscr{B}]$ be as in Question 6 of Sect. 1. It follows immediately from Corollary 2.4 and the remarks preceding it that if $[\mathscr{A}:\mathscr{B}]<\infty$, then $S[\mathscr{B}]$ is finite, since $S[\mathscr{B}]$ will be contained in $\mathscr{F}_{(\sigma_1\sigma_2)^k}$ for some non-zero integer k. It is an interesting problem to try to determine the order of $S[\mathscr{B}]$ for particular subgroups.

For example, let $\mathscr{B}_\tau$ be the subgroup generated by σ_1 and $\tau\sigma_2\tau^{-1}$ for some $\tau\in\mathscr{A}$. Then one easily sees that $S[\mathscr{B}_\tau]$ is contained in $C_1\cap\tau C_2$. We can compute the intersection index $C_1\cdot\tau C_2$ as follows: Each p_i is ramified over a smooth curve of degree 6, so $C_2=6D_i$ in $\mathrm{Pic}(S)$. Hence

$$C_1\cdot\tau C_2=\tau^*(6D_1)\cdot(6D_2)=\frac{432\alpha^2}{(\alpha^2-1)^2}(\alpha^{\ell(\tau)+(\chi(\tau)+1)/2}+\alpha^{-\ell(\tau)-(\chi(\tau)+1)/2}).$$

This certainly suggests that $\#S[\mathscr{B}_\tau]\to\infty$ as $|\ell(\tau)|\to\infty$, but a rigorous proof would need to take multiplicities into account, as well as the possibility that some points in $C_1\cap\tau C_2$ might have stabilizer strictly larger than $\mathscr{B}_\tau$.

Next let $\mathscr{B}_\tau$ be the (cyclic) subgroup of $\mathscr{A}$ generated by $(\sigma_1\sigma_2)^k$ for some integer $k\neq 0$. Alan Landman has sketched for me a proof that the Lefschetz number of $(\sigma_1\sigma_2)^k$ is

$$\alpha^{2k}+\alpha^{-2k}+22,$$

which again suggests that $\#S[\mathscr{B}_k]\to\infty$ as $k\to\infty$. We briefly indicate Landman's proof.

Consider the action of σ_1 on the Hodge diamond $H^*=H^{i,j}(S,\mathbb{C})$ of S. Since σ_1 is an involution, very eigenvalue of σ_1 on H^* is ± 1, say n^+ plus ones and n^- minus ones. By the generalized Lefschetz fixed point formula,

$$n^+-n^-=L(\sigma_1)=(\text{Euler characteristic of fixed point set of }\sigma_1)=-18,$$

since σ_1 fixes the branch locus of p_1, which is a smooth sextic curve in $\mathbb{P}^2$. On the other hand, $n^++n^-=\dim H^*=24$. Hence $n^+=3$ and $n^-=21$. Now σ_1 acts as $+1$ on H^0 and H^4, and it also acts as $+1$ on the subspace of $H^{1,1}$ spanned by the image of D_1, since $\sigma_1^*D_1=D_1$. Hence it acts as -1 on the rest of H^*. In particular, if we let V be the image of $\mathrm{Pic}(S)\otimes\mathbb{C}$ in $H^{1,1}$ and $V^\perp$ its orthogonal complement, then σ_1 acts as -1 on $V^\perp$, since it is diagonalizable and all of it's eigenvalues are -1. Further, σ_1 acts as -1 on $H^{0,2}\cong\mathbb{C}$ and $H^{2,0}\cong\mathbb{C}$. By a similar argument, the same holds for σ_2. Hence $(\sigma_1\sigma_2)^k$ acts as $+1$ on all of H^* except for V; in other words, $H^*\cong V\oplus\mathbb{C}^{22}$ as a representation space for $(\sigma_1\sigma_2)^k$. Using the basis $V=\mathbb{C}E^+\oplus\mathbb{C}E^-$, we find that

$$L((\sigma_1\sigma_2)^k)=\mathrm{Trace}(\sigma_1\sigma_2)^k|_V+\mathrm{Trace}(\sigma_1\sigma_2)^k|_{\mathbb{C}^{22}}=\alpha^{2k}+\alpha^{-2k}+22,$$

which is Landman's formula for the Lefschetz number of $(\sigma_1\sigma_2)^k$.

We can also describe the effective and ample divisors in $\mathrm{Pic}(S)$, or at least in that part of $\mathrm{Pic}(S)$ spanned by D_1 and D_2.

Proposition 2.5 *Let $D \in \mathrm{Pic}(S)$ be a divisor which can be written in the form*

$$D = n_1 D_1 + n_2 D_2.$$

Then the following are equivalent:

(i) *D is effective.*

(ii) *D is ample.*

(iii) *$D \cdot E^+ > 0$ and $D \cdot E^- > 0$.*

(iv) *$n_1 > -\alpha n_2$ and $n_2 > -\alpha n_1$.*

In its entirety, this proposition only holds for divisors in $\mathrm{Span}\{D_1, D_2\}$; *it may not be true for all divisors in* $\mathrm{Pic}(S)$. *However, the following implications are true for every* $D \in \mathrm{Pic}(S)$:

$$\text{(ii)} \Rightarrow \text{(i)} \Rightarrow \text{(iii)}$$

Proof. Before starting the proof, we note that $D_1^2 = D_2^2 = 2$ and $D_1 \cdot D_2 = 4$. So using the definition of $E^{\pm}$ and $\alpha^2 = 4\alpha - 1$, one easily computes

$$E^+ \cdot E^+ = E^- \cdot E^- = 0 \quad \text{and} \quad E^+ \cdot E^- = 12\alpha.$$

We begin by showing that (i) implies (iii) for every divisor $D \in \mathrm{Pic}(S)$. So let D be any effective divisor on S. We write

$$D = t_1 E^+ + t_2 E^- + \Gamma \in \mathrm{Pic}(S) \otimes \mathbb{R}$$

with

$$t_1 = \frac{D \cdot E^-}{12\alpha}, \qquad t_2 = \frac{D \cdot E^+}{12\alpha}, \quad \text{and} \quad \Gamma \cdot E^+ = \Gamma \cdot E^- = 0.$$

Since E^+ and E^- span the same subspace of $\mathrm{Pic}(S) \otimes \mathbb{R}$ as D_1 and D_2, we note that $\Gamma \cdot D_1 = \Gamma \cdot D_2 = 0$.

For any integer k, let $\phi_k = (\sigma_2 \sigma_1)^k \in \mathrm{Aut}(S)$. Since $D_1 + D_2$ is very ample and ϕ_k is an automorphism, the divisor $\phi_k^*(D_1 + D_2)$ is very ample. Hence its intersection with the effective divisor D is a positive integer. Thus

$$\begin{aligned} 1 \leqq D \cdot \phi_k^*(D_1 + D_2), &\quad \text{intersection of effective and very ample,} \\ = (\alpha - 1)(D \cdot \phi_k^*(E^+ + E^-)), &\quad \text{definition of } E^{\pm}, \\ = (\alpha - 1)(t_1 E^+ + t_2 E^-) \cdot (\alpha^{2k} E^+ + \alpha^{-2k} E^-), &\quad \text{from Prop. 2.2(c),} \\ = (\alpha - 1)\, 12\alpha (\alpha^{-2k} t_1 + \alpha^{2k} t_2). \end{aligned}$$

This inequality holds for all $k \in \mathbb{Z}$, which implies that $t_1 > 0$ and $t_2 > 0$. We conclude that $D \cdot E^+ > 0$ and $D \cdot E^- > 0$.

Next we show that (ii) implies (i), again for arbitrary divisors on S. We will use the following general facts that hold on all K3 surfaces. (See, e.g., [3] or [6].)

$$\begin{aligned} h^2(\mathcal{O}_S) = h^0(\Omega_S) = p_g(S) = 1, \\ h^1(\mathcal{O}_S) = q = b_1(S) = 0, \\ p_a(S) = h^2(\mathcal{O}_S) - h^1(\mathcal{O}_S) = 1, \\ \mathcal{K}_S = 0, \quad (\mathcal{K}_S \text{ the canonical class on } S.) \end{aligned}$$

Let $D \in \mathrm{Pic}(S)$ be any divisor. The Riemann-Roch theorem [6, p. 472], [8, V.1.6] for surfaces says that

$$h^0(D) - h^1(D) + h^0(\mathscr{K}_S - D) = \tfrac{1}{2}(D^2 - D \cdot \mathscr{K}_S) + p_a(S) + 1.$$

So for a K3 surface we find that

$$h^0(D) + h^0(-D) = h^1(D) + \tfrac{1}{2} D^2 + p_a(S) + 1 \geqq \tfrac{1}{2} D^2 + 2.$$

By assumption, D is ample, so $h^0(-D) = 0$ and $D^2 > 0$. Hence $h^0(D) > 2$, so D is effective.

The proof that (iii) and (iv) are equivalent is immediate once one calculates

$$D \cdot E^+ = 2\sqrt{3}(n_1 + \alpha n_2) \quad \text{and} \quad D \cdot E^- = 2\sqrt{3}(\alpha n_1 + n_2).$$

We have now proven (ii)$\Rightarrow$(i)$\Rightarrow$(iii)$\Leftrightarrow$(iv), so it remains to show that (iii) implies (ii). We will use the Nakai-Moishezon criterion [8, V.1.10]. Write $D = t_1 E^+ + t_2 E^-$ with $t_1, t_2 \in \mathbb{R}$ as above. Note this is where we use the assumption that D is a linear combination of D_1 and D_2. Our assumption (iii) implies that $t_1 > 0$ and $t_2 > 0$. Hence

$$D^2 = 24 \alpha t_1 t_2 > 0.$$

Next let $C \subset S$ be any irreducible curve. Then C gives an effective divisor, so from (i) implies (iii) we deduce that $C \cdot E^+ > 0$ and $C \cdot E^- > 0$. (Remember that we proved (i)$\Rightarrow$(iii) for every divisor in $\mathrm{Pic}(S)$.) Hence

$$C \cdot D = t_1 (C \cdot E^+) + t_2 (C \cdot E^-) > 0.$$

By the Nakai-Moishezon criterion, D is ample. □

If $\mathrm{Pic}(S)$ has rank 2, we can also pick out the curves on S of small arithmetic genus. Wehler's generalization of the Noether-Lefschetz theorem [18, Theorem 5.6] says that a general surface like S satisfies this rank 2 condition, although it does not appear to be known how to find such surfaces defined over $\mathbb{Q}$ or over a number field. And even if one had such a surface, in order to describe the rational points on S, one really needs to find the curves of *geometric* genus 0 and 1; this seems in general to be a difficult problem. (See also [14, § 5] for a computation of all curves of arithmetic genus 0 on a particular K3 surface having a very large Picard group, and a discussion of the problem of singular curves of geometric genus 0.)

Corollary 2.6 *Assume that* $\mathrm{Pic}(S)$ *has rank* 2. *Let* $C \subset S$ *be an irreducible curve on* S. *Then the arithmetic genus of* C *satisfies*

$$p_a(C) \geqq 2.$$

Further, if $p_a(C) = 2$, *then there is an automorphism* $\phi \in \mathscr{A}$ *such that*

$$\phi^* C = D_1 \text{ or } D_2 \quad \text{in } \mathrm{Pic}(S).$$

Proof. Write $C=n_1D_1+n_2D_2$ in $\mathrm{Pic}(S)$, where $n_1, n_2\in\mathbb{Z}$. Then the adjunction formula [8, V.1.5 and Ex. V.1.3] gives (note S has trivial canonical bundle)

$$\begin{aligned} p_a(C)&=\tfrac{1}{2}(C^2+C\cdot\mathcal{K}_S)+1\\ &=n_1^2+4n_1n_2+n_2^2+1=\frac{(\alpha n_1+n_2)(n_1+\alpha n_2)}{\alpha}+1. \end{aligned}$$

Since C is effective, Prop. 2.5 implies that $\alpha n_1+n_2>0$ and $n_1+\alpha n_2>0$, so we see that $p_a(C)>1$ (strict inequality.) Further, $p_a(C)$ is an integer, from which it follows that $p_a(C)\geqq 2$. This proves the first assertion.

For the second assertion, we observe that

$$p_a(C)=2\Leftrightarrow n_1^2+4n_2n_2+n_2^2=1.$$

As above, we factor the quadratic form as

$$1=n_1^2+4n_1n_2+n_2^2=(n_1+\alpha n_2)(n_1+\alpha^{-1}n_2)$$

in the ring $\mathbb{Z}[\alpha]=\mathbb{Z}[\sqrt{3}]$. Each factor is a unit, and the fundamental unit in $\mathbb{Z}[\alpha]$ is α, so there is an integer k such that

$$n_1+\alpha n_2=\alpha^k \quad\text{and}\quad n_1+\alpha^{-1}n_2=\alpha^{-k}.$$

Next we write C in terms of the basis $\{E^+, E^-\}$ for $\mathrm{Pic}(S)\otimes\mathbb{R}$.

$$\begin{aligned} C&=\left(\frac{1}{\alpha^2-1}\right)((\alpha n_1+n_2)E^+ +(n_1+\alpha n_2)E^-)\\ &=\left(\frac{1}{\alpha^2-1}\right)(\alpha^{1-k}E^+ +\alpha^kE^-). \end{aligned}$$

Using Prop. 2.2, we can find an automorphism $\phi\in\mathcal{A}$ such that $\ell(\phi)=k$. Note that $\phi^*E^\pm=\alpha^{\pm\ell(\phi)}E^{\pm\chi(\phi)}$. Hence

$$\begin{aligned} \phi^*C&=\left(\frac{1}{\alpha^2-1}\right)(\alpha\cdot\alpha^{-\ell(\phi)}\phi^*E^+ +\alpha^{\ell(\phi)}\phi^*E^-)\\ &=\left(\frac{1}{\alpha^2-1}\right)(\alpha E^{\chi(\phi)}+E^{-\chi(\phi)})\\ &=\begin{cases} D_1 & \text{if } \chi(\phi)=1,\\ D_2 & \text{if } \chi(\phi)=-1. \end{cases} \quad\square \end{aligned}$$

3 Existence of the canonical height

In this section we are going to prove that the canonical heights $\hat{h}^+$ and $\hat{h}^-$ exist, are unique, and satisfy properties (i)–(vii) of Theorem 1.1. We start with

uniqueness. Suppose that $\hat{h}^{\pm}$ and $\hat{g}^{\pm}$ satisfy (i) and (ii). Using (i) we see that $\hat{f}^{\pm}=\hat{h}^{\pm}-\hat{g}^{\pm}$ is bounded, and then using (ii) repeatedly gives

$$\hat{f}^{\pm}((\sigma_2\sigma_1)^{\pm n}P)=\alpha^{2n}\hat{f}^{\pm}(P).$$

The left-hand-side is bounded independently of n, so dividing by α^{2n} and letting $n\to\infty$ shows that $\hat{f}^{\pm}(P)=0$. Hence $\hat{h}^{\pm}=\hat{g}^{\pm}$.

To show the $\hat{h}^{\pm}$ exist, we use Tate's averaging method. From Prop. 2.1(b) we have the divisor class relations

$$(\sigma_2\sigma_1)^*E^+=\alpha^2E^+, \qquad (\sigma_1\sigma_2)^*E^-=\alpha^2E^-;$$

and these give the height relations

$$h_{E^+}(\sigma_2\sigma_1P)=\alpha^2h_{E^+}(P)+O(1), \qquad h_{E^-}(\sigma_1\sigma_2P)=\alpha^2h_{E^-}(P)+O(1).$$

Now Tate's method shows that the limits

$$\hat{h}^+(P)=\lim_{n\to\infty}\alpha^{-2n}h_{E^+}((\sigma_2\sigma_1)^nP)$$

and

$$\hat{h}^-(P)=\lim_{n\to\infty}\alpha^{-2n}h_{E^+}((\sigma_1\sigma_2)^nP)$$

exist and satisfy

(i) $\hat{h}^+=h_{E^+}+O(1)$, $\hat{h}^-=h_{E^-}+O(1)$.

(ii)′ $\hat{h}^+(\sigma_2\sigma_1P)=\alpha^2\hat{h}^+(P)$, $\hat{h}^-(\sigma_1\sigma_2P)=\alpha^2\hat{h}^-(P)$.

We briefly sketch the argument for $\hat{h}^+$. (See, e.g., [12, VIII.9.1] or [9, Chap. 5] for the analogous construction on abelian varieties.)

Thus let $P\in S(\bar{K})$. Then for any $n\geqq m$ we have

$$\begin{aligned}
&|\alpha^{-2n}h_{E^+}((\sigma_2\sigma_1)^nP)-\alpha^{-2m}h_{E^+}((\sigma_2\sigma_1)^mP)|\\
&\quad=\left|\sum_{i=m+1}^{n}\alpha^{-2i}h_{E^+}((\sigma_2\sigma_1)^iP)-\alpha^{-2i+2}h_{E^+}((\sigma_2\sigma_1)^{i-1}P)\right|\\
&\quad\leqq\sum_{i=m+1}^{n}\alpha^{-2i}|h_{E^+}(\sigma_2\sigma_1Q_i)-\alpha^2h_{E^+}(Q_i)|, \qquad \text{where } Q_i=(\sigma_2\sigma_1)^{i-1}P,\\
&\quad\leqq\sum_{i=m+1}^{\infty}\alpha^{-2i}\kappa, \qquad \text{where } |h_{E^+}\circ(\sigma_2\sigma_1)-\alpha^2h_{E^+}|\leqq\kappa=\kappa(S),\\
&\quad=\frac{\alpha^{-2m}\kappa}{\alpha^2-1}.
\end{aligned}$$

Hence the sequence $\alpha^{-2n}h_{E^+}((\sigma_2\sigma_1)^nP)$ is Cauchy, so the limit defining $\hat{h}^+$ exists. Further, taking $m=0$ and letting $n\to\infty$ in the above inequality shows that

$$|\hat{h}^+(P)-h_{E^+}(P)|\leqq\frac{\kappa}{\alpha^2-1}.$$

We now have functions $\hat{h}^{\pm}$ satisfying (i); and in place of (ii) we have the conditions we denoted above by (ii)′. Using (i) and Lemma 2.1(b) we have

$$\begin{aligned}\hat{h}^{+}(\sigma_2 P) &= h_{E^+}(\sigma_2 P)+O(1)=h_{\sigma_2^* E^+}(P)+O(1)\\ &=\alpha h_{E^-}(P)+O(1)=\alpha\hat{h}^{-}(P)+O(1).\end{aligned}$$

Now replace P by $(\sigma_1\sigma_2)^n P$ and use (ii)′ to get

$$\begin{aligned}\alpha^{2n}\hat{h}^{+}(\sigma_2 P) &= \hat{h}^{+}((\sigma_2\sigma_1)^n\sigma_2 P)\\ &=\hat{h}^{+}(\sigma_2(\sigma_1\sigma_2)^n P)\\ &=\alpha\hat{h}^{-}((\sigma_1\sigma_2)^n P)+O(1)\\ &=\alpha^{2n+1}\hat{h}^{-}(P)+O(1).\end{aligned}$$

Dividing by α^{2n} and letting $n\to\infty$ gives the equality $\hat{h}^{+}(\sigma_2 P)=\alpha\hat{h}^{-}(P)$. This is one of the equations in (ii); the other three are proven similarly.

Next we verify (iii) by induction on ϕ. It is clearly true for $\phi=e$, so we assume it is true for ϕ and verify it for $\sigma_1\phi$ and $\sigma_2\phi$. Thus

$$\begin{aligned}\hat{h}^{\pm}\circ(\sigma_1\phi) &= \alpha^{\mp 1}\hat{h}^{\mp}\circ\phi, && \text{from (ii)},\\ &=\alpha^{\mp 1\mp\ell(\phi)}\hat{h}^{\mp\chi(\phi)}, && \text{induction hypothesis},\\ &=\alpha^{\pm\ell(\sigma_1\phi)}\hat{h}^{\pm\chi(\sigma_1\phi)}, && \text{definition of } \ell \text{ and } \chi.\end{aligned}$$

This proves (iii) for $\sigma_1\phi$, and $\sigma_2\phi$ is proven similarly.

Property (iv) follows immediately from (i) and the definition of $E^{\pm}$:

$$\begin{aligned}\hat{h}&=\hat{h}^{+}+\hat{h}^{-}=h_{E^+}+h_{E^-}+O(1)\\ &=h_{\alpha D_1-D_2}+h_{-D_1+\alpha D_2}+O(1)=(\alpha-1)h_{D_1+D_2}+O(1).\end{aligned}$$

Similarly, property (v) is immediate from (iii):

$$\hat{h}^{+}(\phi P)\hat{h}^{-}(\phi P)=\{\alpha^{\ell(\phi)}\hat{h}^{\chi(\phi)}(P)\}\cdot\{\alpha^{-\ell(\phi)}\hat{h}^{-\chi(\phi)}(P)\}=\hat{h}^{+}(P)\hat{h}^{-}(P).$$

To prove property (vi), we use (iv), which implies that the function $\hat{h}=\hat{h}^{+}+\hat{h}^{-}$ is bounded below. (A Weil height corresponding to an ample divisor is always bounded below.) Hence

$$\hat{h}^{+}=\hat{h}-\hat{h}^{-}\geqq-\hat{h}^{-}-O(1);$$

so for any point $P\in S(\bar{K})$ and any $n\geqq 0$,

$$\begin{aligned}\hat{h}^{+}(P) &= \alpha^{-2n}\hat{h}^{+}((\sigma_2\sigma_1)^n P)\\ &\geqq\alpha^{-2n}\{-\hat{h}^{-}((\sigma_2\sigma_1)^n P)-O(1)\}\\ &=-\alpha^{-4n}\hat{h}^{-}(P)-O(\alpha^{-2n}).\end{aligned}$$

Letting $n\to\infty$ gives $\hat{h}^{+}(P)\geqq 0$. A similar argument gives $\hat{h}^{-}(P)\geqq 0$.

It remains to prove property (vii). Suppose first that $\hat{h}^+(P)=0$. Then using (iii) and (iv) we find that

$$\begin{aligned}\hat{h}((\sigma_2\sigma_1)^n P)&=\hat{h}^+((\sigma_2\sigma_1)^n P)+\hat{h}^-((\sigma_2\sigma_1)^n P)\\&=\alpha^{2n}\hat{h}^+(P)+\alpha^{-2n}\hat{h}^-(P)\\&=\alpha^{-2n}\hat{h}^-(P).\end{aligned}$$

Therefore $\{(\sigma_2\sigma_1)^n P: n=1, 2, \ldots\}$ is a set of bounded height for the height $\hat{h}$. Since $\hat{h}$ is a height relative to an ample divisor, it follows that this is a finite set. Hence there is an integer $n\geqq 1$ such that $(\sigma_2\sigma_1)^n P=P$, and so

$$\frac{\mathscr{A}}{(\sigma_2\sigma_1)^{n\mathbb{Z}}}\to\{\phi P: \phi\in\mathscr{A}\},\qquad \phi\mapsto\phi P,$$

is a well-defined surjective map. Since the subgroup $(\sigma_2\sigma_1)^{n\mathbb{Z}}$ has finite index in $\mathscr{A}$, (precisely, it has index $2n$,) it follows that the set of ϕP's is finite.

Conversely, suppose that there are only finitely many ϕP's. Then $\hat{h}^+(\phi P)$ is bounded independently of ϕ, so

$$\hat{h}^+(P)=\alpha^{-2n}\hat{h}^+((\sigma_2\sigma_1)^n P)\xrightarrow[n\to\infty]{}0.$$

This completes the proof that

$$\hat{h}^+(P)=0\Leftrightarrow\{\phi P: \phi\in\mathscr{A}\}\text{ is finite.}$$

The proof for $\hat{h}^-$ is done similarly. □

4 Applications of the canonical height

In this section we use canonical height functions to prove our main theorems describing the rational points of S.

Proof of Theorem 1.2. (a) Since $\hat{H}(\mathscr{C})=\hat{h}^+(P)\hat{h}^-(P)$ for any point $P\in\mathscr{C}$, this is just a restatement of Theorem 1.1(vii).

Proof of Theorem 1.3. (a) Fix any $Q\in\mathscr{C}$. Then

$$\begin{aligned}\min_{P\in\mathscr{C}}\hat{h}(P)&=\min_{\phi\in\mathscr{A}}\hat{h}(\phi Q)\\&=\min_{\phi\in\mathscr{A}}\hat{h}^+(\phi Q)+\hat{h}^-(\phi Q)\\&=\min_{\phi\in\mathscr{A}}\alpha^{\ell(\phi)}\hat{h}^{\chi(\phi)}(Q)+\alpha^{-\ell(\phi)}\hat{h}^{-\chi(\phi)}(Q),\quad\text{from Theorem 1.1(iii),}\\&=\min_{n\in\mathbb{Z}}\alpha^n\hat{h}^{\chi(n)}(Q)+\alpha^{-n}\hat{h}^{-\chi(n)}(Q).\end{aligned}$$

Here we are writing $\chi\colon\mathbb{Z}\to\{\pm 1\}$ for $\chi(n)=(-1)^n$. We have also used the fact that $\ell\colon\mathscr{A}\to\mathbb{Z}$ is surjective (Prop. 2.2(a)).

If we let

$$f(n)=\alpha^n\hat{h}^{\chi(n)}(Q)+\alpha^{-n}\hat{h}^{-\chi(n)}(Q),$$

then we must find the minimum of $f(n)$ over all $n\in\mathbb{Z}$. The lower bound is immediate:

$$\min_{n\in\mathbb{Z}} f(n)\geqq\inf_{t\in\mathbb{R}}\{t\hat{h}^+(Q)+t^{-1}\hat{h}^-(Q)\}=2\sqrt{\hat{h}^+(Q)\hat{h}^-(Q)}=2\sqrt{\hat{H}(\mathscr{C})}.$$

For the upper bound, we let

$$m=2\left[\frac{1}{2}\log_\alpha\sqrt{\frac{\hat{h}^-(Q)}{\hat{h}^+(Q)}}+\frac{1}{2}\right].$$

(Here $[\cdot]$ denotes greatest integer.) Then m is even and

$$\alpha^{-1}\sqrt{\frac{\hat{h}^-(Q)}{\hat{h}^+(Q)}}\leqq\alpha^m\leqq\alpha\sqrt{\frac{\hat{h}^-(Q)}{\hat{h}^+(Q)}},$$

so

$$\min_{n\in\mathbb{Z}} f(n)\leqq f(m)\leqq 2\alpha\sqrt{\hat{h}^+(Q)\hat{h}^-(Q)}=2\alpha\sqrt{\hat{H}(\mathscr{C})}.$$

This completes the proof of Theorem 1.3(a). □

Proof of Theorem 1.2. (b) Let $\mathscr{C}\subset S(K)$ be a chain with $\hat{H}(\mathscr{C})\leqq B$. From Theorem 1.3(a), there is a point $P\in\mathscr{C}$ with $\hat{h}(P)\leqq 2\alpha\sqrt{B}$. Hence

$$\#\{\mathscr{C}\subset S(K)\colon \hat{H}(\mathscr{C})\leqq B\}\leqq\#\{P\in S(K)\colon \hat{h}(P)\leqq 2\alpha\sqrt{B}\}.$$

But $\hat{h}$ is a Weil height with respect to an ample divisor class (Theorem 1.1(iv)), so this last set is finite. This proves the first half of Theorem 1.2(b); and the second half follows from (a), which says that

$$\{\mathscr{C}\subset S(K)\colon \#\mathscr{C}<\infty\}=\{\mathscr{C}\subset S(K)\colon \hat{H}(\mathscr{C})=0\}.\quad\square$$

Proof of Theorem 1.3. (b) Let

$$\nu(\mathscr{C},\hat{h};B)=\#\{P\in\mathscr{C}\colon \hat{h}(P)\leqq B\}$$

be the counting function of $\mathscr{C}$ relative to the canonical height $\hat{h}$. Fix any point $Q\in\mathscr{C}$. Then the map

$$\mathscr{A}\to\mathscr{C},\qquad \phi\mapsto\phi Q,$$

is surjective and $\mu(\mathscr{C})$-to-1, so

$$\begin{aligned}\mu(\mathscr{C})\,\nu(\mathscr{C},\hat{h};B)&=\#\{\phi\in\mathscr{A}\colon \hat{h}(\phi Q)\leqq B\}\\&=\#\{\phi\in\mathscr{A}\colon \alpha^{\ell(\phi)}\hat{h}^{\chi(\phi)}(Q)+\alpha^{-\ell(\phi)}\hat{h}^{-\chi(\phi)}(Q)\leqq B\}\\&=\#\{n\in\mathbb{Z}\colon \alpha^n\hat{h}^{\chi(n)}(Q)+\alpha^{-n}\hat{h}^{-\chi(n)}(Q)\leqq B\}\\&=\#\{m\in\mathbb{Z}\colon (\alpha^2)^m\hat{h}^+(Q)+(\alpha^2)^{-m}\hat{h}^-(Q)\leqq B\}\\&\quad+\#\{m\in\mathbb{Z}\colon (\alpha^2)^m(\alpha\hat{h}^+(Q))+(\alpha^2)^{-m}(\alpha^{-1}\hat{h}^-(Q))\leqq B\}.\end{aligned}$$

To conclude the proof of Theorem 1.3(b), we will apply the following elementary counting lemma to each of the last two terms.

Lemma. *Let $a, b, B>0$ and $\lambda>1$ be constants, and let*

$$\Sigma(B)=\{m\in\mathbb{Z}\colon \lambda^m a+\lambda^{-m}b\leqq B\}.$$

If $B^2<4ab$, then $\Sigma(B)=\emptyset$. If $B^2\geqq 4ab$, then

$$-1\leqq \#\Sigma(B)-\log_\lambda\frac{B^2}{4ab}\leqq 1+\log_\lambda(4).$$

Proof. Notice that

$$\begin{aligned} m\in\Sigma(B)&\Leftrightarrow a\lambda^{2m}-B\lambda^m+b\leqq 0\\ &\Leftrightarrow \frac{2b}{B+\sqrt{B^2-4ab}}\leqq\lambda^m\leqq\frac{B+\sqrt{B^2-4ab}}{2a}.\end{aligned}$$

In particular, if $B^2<4ab$, then there are no values of m, so $\#\Sigma(B)=0$.

Now we assume that $B^2\geqq 4ab$. Then every $m\in\Sigma(B)$ satisfies the inequality $b/B\leqq\lambda^m\leqq B/a$, so

$$\#\Sigma(B)\leqq\#\left\{m\in\mathbb{Z}\colon \log_\lambda\frac{b}{B}\leqq m\leqq\log_\lambda\frac{B}{a}\right\}\leqq 1+\log_\lambda\frac{B^2}{ab}.$$

Similarly, if $2b/B\leqq\lambda^m\leqq B/2a$, then m will be in $\Sigma(B)$, so

$$\#\Sigma(B)\geqq\#\left\{m\in\mathbb{Z}\colon \log_\lambda\frac{2b}{B}\leqq m\leqq\log_\lambda\frac{B}{2a}\right\}\geqq -1+\log_\lambda\frac{B^2}{4ab}.$$

This concludes the proof of the Lemma.

Resuming the proof of Theorem 1.3(b), we use the Lemma twice (and the assumption that $B^2\geqq 4\hat{H}(\mathscr{C})$) to conclude that

$$-2\leqq\mu(\mathscr{C})\,v(\mathscr{C},\hat{h};B)-2\log_{\alpha^2}\frac{B^2}{4\hat{H}(\mathscr{C})}\leqq 2+2\log_{\alpha^2}(4).$$

(Note that $\hat{H}(\mathscr{C})=\hat{h}^+(Q)\hat{h}^-(Q)$ by definition.) Since $2\log_{\alpha^2}4=\log_\alpha 4\approx 1.05$, this is stronger than the desired result.

(c) Since $\hat{h}$ and h_D are both Weil height functions with respect to ample divisors, there is a constant $c>0$ such that

$$c^{-1}h_D(P)\leqq\hat{h}(P)\leqq ch_D(P)\quad\text{for all } P\in S(\bar{K}).$$

Hence continuing with the notation from (b), we have

$$v(\mathscr{C},h_D;c^{-1}B)\leqq v(\mathscr{C},\hat{h};B)\leqq v(\mathscr{C},h_D;cB),$$

so (c) follows immediately from (b). □

Proof of Theorem 1.4. From Theorem 1.2(a), a point has canonical height 0 if and only if it is in $S(K)_{\text{fin}}$; and from Theorem 1.3(a), a chain $\mathscr{C}$ has no points with canonical height less than $2\sqrt{\hat{H}(\mathscr{C})}$. Hence using Theorem 1.3(b), we find

$$\begin{aligned}\#\{P\in S(K)\colon\hat{h}(P)\leqq B\}&=\#S(K)_{\text{fin}}+\sum_{\substack{\mathscr{C}\subset S(K)\\0<4\hat{H}(\mathscr{C})\leqq B^2}}\ \sum_{\substack{P\in\mathscr{C}\\\hat{h}(P)\leqq B}}1\\&=\#S(K)_{\text{fin}}+\sum_{\substack{\mathscr{C}\subset S(K)\\0<4\hat{H}(\mathscr{C})\leqq B^2}}\left\{\frac{1}{\mu(\mathscr{C})}\log_\alpha\frac{B^2}{4\hat{H}(\mathscr{C})}+\varepsilon(\mathscr{C})\right\}.\end{aligned}$$

From Theorem 1.2(b), the set $S(K)_{\text{fin}}$ is finite; and from Theorem 1.3(b), the error $\varepsilon(\mathscr{C})$ satisfies $|\varepsilon(\mathscr{C})|\leqq 4$. □

Proof of Corollary 1.5 Let $P \in S(\bar{K})$ be a point whose chain $\mathscr{C}(P)$ is stable under the action of $\mathrm{Gal}(\bar{K}/K)$, and let

$$\mathscr{A}_P = \{\phi \in \mathscr{A} : \phi P = P\}$$

be the stabilizer of P. We consider two cases, according to whether or not $\mathscr{A}_P$ is trivial.

Suppose first that $\#\mathscr{A}_P > 1$. If it contains an element of the form $(\sigma_1 \sigma_2)^k$, $k \neq 0$, then $\mathscr{A}_P$ has finite index inside $\mathscr{A}$. Since $\mathscr{A}/\mathscr{A}_P \xleftrightarrow{\text{one-to-one}} \mathscr{C}(P)$, this implies that $\mathscr{C}(P)$ is finite, so (ii) holds. Otherwise $\mathscr{A}_P$ must contain an element of the form $\tau \sigma_j \tau^{-1}$, since these are the only other elements in $\mathscr{A}$. Let $\psi \in \mathscr{A}_P$ be such an element of order 2, and notice that $\chi(\psi) = -1$.

Now if $P \in S(K)$, then P satisfies (i) and we're done with this case; so we may as well assume that there is some $\gamma \in \mathrm{Gal}(\bar{K}/K)$ with $\gamma P \neq P$. On the other hand, we know that $\gamma P \in \mathscr{C}(P)$, say $\gamma P = \phi P$ for some $\phi \in \mathscr{A}$. Since $\psi P = P$, we also have $\gamma P = \phi\psi P$. Using the fact that $\chi(\phi\psi) = \chi(\phi)\chi(\psi) = -\chi(\phi)$ and replacing ϕ by $\phi\psi$ if necessary, we may assume that

$$\gamma P = \phi P, \quad \chi(\phi) = 1, \quad \text{and} \quad \ell(\phi) \neq 0.$$

(Note $\ell(\phi) = 0 \Rightarrow \phi = e \Rightarrow \gamma P = \phi P = P$, contrary to our assumptions.)

We next observe that height functions are Galois invariant. (See, e.g., [12, VIII.5.10].) Hence

$$\hat{h}^+(P) = \hat{h}^+(\gamma P) = \hat{h}^+(\phi P) = \alpha^{l(\phi)} \hat{h}^+(P), \quad \text{since } \chi(\phi) = 1.$$

But $\alpha^{\ell(\phi)} \neq 1$, so we conclude that $\hat{h}^+(P) = 0$, and from Theorem 1.1(vii) that $\mathscr{C}(P)$ is finite. Thus P again satisfies (ii).

It remains to consider the case that $\mathscr{A}_P$ is trivial. In this case, for every $\gamma \in \mathrm{Gal}(\bar{K}/K)$ there is a *unique* $\phi_\gamma \in \mathscr{A}$ such that $\gamma P = \phi_\gamma P$. We thus get a well-defined homomorphism

$$\mathrm{Gal}(\bar{K}/K) \to \mathscr{A}, \quad \gamma \mapsto \phi_\gamma,$$

whose kernel is precisely $\mathrm{Gal}(\bar{K}/K(P))$. Hence $K(P)$ is a Galois extension of K, and its Galois group injects into $\mathscr{A}$. But the only finite subgroups of $\mathscr{A}$ are groups of order 1 or 2. Therefore either $K(P) = K$, which is condition (i); or else $[K(P): K] = 2$, which is the first part of condition (iii). It remains to show that in this last case, the other parts of condition (iii) are true.

If we write $\mathrm{Gal}(K(P)/K) = \{e, \gamma\}$, then we know that ϕ_γ has order 2, so $\phi_\gamma = \tau^{-1} \sigma_j \tau$ for some $\tau \in \mathscr{A}, j \in \{1, 2\}$. It follows that

$$\gamma(\tau P) = \tau(\gamma P) = \tau \phi_\gamma P = \sigma_j(\tau P), \text{ so}$$

$$\gamma(p_j(\tau P)) = p_j(\gamma(\tau P)) = p_j(\sigma_j(\tau P)) = p_j(\tau P).$$

Therefore $p_j(\tau P)$ is fixed by $\mathrm{Gal}(K(P)/K)$, hence $p_j(\tau P) \in \mathbb{P}^2(K)$. This proves the second part of condition (iii); and the third part is immediate, since we are considering the case where $\#\mathscr{A}_P = 1$. □

5 A numerical example

In this section we will choose a particular K3 surface and calculate the heights of some of its rational points. By changing coordinates, we may as well asssume that our bilinear form gives the usual flag manifold in $\mathbb{P}^2 \times \mathbb{P}^2$,

$$L(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

We intersect the flag manifold with a "randomly chosen" surface of type (2, 2) given by the equation

$$\begin{aligned} Q(\mathbf{x}, \mathbf{y}) = {} & x_1^2 y_1^2 + 3 x_1 x_2 y_1^2 + x_2^2 y_1^2 + 4 x_1^2 y_1 y_2 + 3 x_1 x_2 y_1 y_2 - 2 x_3^2 y_1 y_2 \\ & - x_1^2 y_2^2 + 2 x_2^2 y_2^2 - x_1 x_3 y_2^2 - 4 x_2 x_3 y_2^2 + 5 x_1 x_3 y_1 y_3 \\ & - 4 x_2 x_3 y_1 y_3 + 7 x_1^2 y_2 y_3 + 4 x_2^2 y_2 y_3 + x_1 x_2 y_3^2 + 3 x_3^2 y_3^2. \end{aligned}$$

Then our surface is the locus in $\mathbb{P}^2 \times \mathbb{P}^2$ given by the two equations

$$S\colon L(\mathbf{x}, \mathbf{y}) = Q(\mathbf{x}, \mathbf{y}) = 0.$$

A brief (5 hours on a Macintosh IIcx) search reveals that there are twelve rational points on S with multiplicative height less than or equal to 40.

$$\begin{aligned} \{ & ([0, 1, 1], [1, 1, -1]), ([1, 0, 0], [0, 0, 1]), ([0, 1, 0], [0, 0, 1]), \\ & ([1, 0, -1], [0, 1, 0]), ([0, 0, 1], [0, 1, 0]), ([0, 0, 1], [1, 0, 0]), \\ & ([3, 1, 3], [-3, 3, 2]), ([1, 0, 0], [0, 7, 1]), ([8, 6, 9], [-6, 5, 2]), \\ & ([1, 0, -1], [9, 1, 9]), ([3, 8, 11], [1, 1, -1]), ([12, 1, -20], [2, -4, 1])\} \end{aligned}$$

These twelve points lie in six distinct chains, each of which is (undoubtedly) infinite, although two of them are one sided. More precisely, ([0, 1, 0], [0, 0, 1]) is on the ramification locus of π_1, so is fixed by σ_1; and ([0, 0, 1], [1, 0, 0]) is on the ramification locus of π_2, so is fixed by σ_2. To illustrate the rapid growth of points in a given chain, we will look at the chain generated by ([0, 1, 0], [0, 0, 1]), which begins

$$P_1 \xrightarrow{\sigma_2} ([1, 0, 0], [0, 0, 1]) \xrightarrow{\sigma_1} ([1, 0, 0], [0, 7, 1])$$

$$\xrightarrow{\sigma_2} ([1645, -344, 2408], [0, 7, 1]) \xrightarrow{\sigma_1} .$$

In Table 1 we have listed the first few points in the chain generated by P_1. As expected, the coordinates grow extremely rapidly.

Table 1

ϕ	ϕP_1
e	$([0, 1, 0], [0, 0, 1])$
σ_2	$([1, 0, 0], [0, 0, 1])$
$\sigma_1\sigma_2$	$([1, 0, 0], [0, 7, 1])$
$\sigma_2\sigma_1\sigma_2$	$([1645, -344, 2408], [0, 7, 1])$
$(\sigma_1\sigma_2)^2$	$([1645, -344, 2408], [-1.3\cdot 10^{13}, 5.6\cdot 10^{12}, 9.7\cdot 10^{12}])$
$\sigma_2(\sigma_1\sigma_2)^2$	$([2.2\cdot 10^{49}, -3.0\cdot 10^{49}, 4.6\cdot 10^{49}], [-1.3\cdot 10^{13}, 5.6\cdot 10^{12}, 9.7\cdot 10^{12}])$
$(\sigma_1\sigma_2)^3$	$([2.2\cdot 10^{49}, -3.0\cdot 10^{49}, 4.6\cdot 10^{49}], [2.2\cdot 10^{186}, 1.6\cdot 10^{186}, 6.4\cdot 10^{184}])$
$\sigma_2(\sigma_1\sigma_2)^3$	$([-7.9\cdot 10^{695}, 1.0\cdot 10^{696}, 1.5\cdot 10^{695}], [2.2\cdot 10^{186}, 1.6\cdot 10^{186}, 6.4\cdot 10^{184}])$

Table 2

ϕ	$\ell(\phi)$	$\chi(\phi)$	$h_{E^+}(\phi P_1)$	$h_{E^-}(\phi P_1)$	$\alpha^{-\ell} h_{E^\chi}$ $(\ell(\phi)\geqq 0)$	$\alpha^{\ell} h_{E^{-\chi}}$ $(\ell(\phi)\leqq 0)$
e	0	1	0	0	0	0
σ_2	1	−1	0	0	0	
$\sigma_1\sigma_2$	−2	1	−1.94591	7.26224		0.5214
$\sigma_2\sigma_1\sigma_2$	3	−1	27.1139	−0.524316	0.5216	
$(\sigma_1\sigma_2)^2$	−4	1	−1.13758	104.912		0.5408
$\sigma_2(\sigma_1\sigma_2)^2$	5	−1	396.612	−1.66497	0.5478	
$(\sigma_1\sigma_2)^3$	−6	1	−2.22648	1486.82		0.5503
$\sigma_2(\sigma_1\sigma_2)^3$	7	−1	5552.05	−1.44477	0.5506	

For any point $P=(\mathbf{x}, \mathbf{y})\in\mathbb{P}^2\times\mathbb{P}^2$, we normalize our height functions by taking the usual Weil height on $\mathbb{P}^2$ and setting

$$h_{E^+}(P)=\alpha h(\mathbf{x})-h(\mathbf{y}) \quad \text{and} \quad h_{E^-}(P)=-h(\mathbf{x})+\alpha h(\mathbf{y}).$$

In Table 2 we have computed the heights of the first few elements in the chain generated by P_1.

From Table 2 we see that

$$\hat{h}^+(P_1)\approx 2.1 \quad \text{and} \quad \hat{h}^-(P_1)\approx 0.55.$$

We are computing these values by the formula

$$\hat{h}^{\pm}(P)=\alpha^{\mp\ell(\phi)}\hat{h}^{\pm\chi(\phi)}(\phi P)=\alpha^{\mp\ell(\phi)}h_{E^{\pm\chi(\phi)}}(\phi P)+O(\alpha^{\mp\ell(\phi)}).$$

In order to make the error term on the right-hand-side of this equation small, we need to choose the plus/minus sign so that $\mp\ell(\phi)<0$. So for each ϕ, Table 2 provides an approximation only to $\hat{h}^-(P_1)$, not to $\hat{h}^+(P_1)$. However, since our point P_1 satisfies $\sigma_1(P_1)=P_1$, it follows from the general theory that

$$\hat{h}^+(P_1)=\hat{h}^+(\sigma_1 P_1)=\alpha\hat{h}^-(P_1).$$

This allows us to approximate both canonical heights from Table 2.

In order to compute $\hat{h}^-(P_1)$ to d decimal places, we would need to choose ϕ so that $\alpha^{-|\ell(\phi)|}\ll 10^{-d}$. Taking $\phi=(\sigma_1\sigma_2)^n$, this means that we would need to compute $(\sigma_1\sigma_2)^n P_1$ with $n\geqq 0.87\,d+O(1)$ in order to get d decimal places of accuracy. This is completely infeasible for even moderate values of d, since (some of) the coordinates of $(\sigma_1\sigma_2)^n P_1$ will have around $\alpha^{2n}\approx 10^{1.14n}$ digits. It seems worthwhile to consider some other method of computing the canonical heights $\hat{h}^+$ and $\hat{h}^-$, possibly by decomposing them into canonical local height functions, as is done for abelian varieties. (See [9, Chap. 11] for the general theory on abelian varieties, and [13] for the Silverman-Tate algorithm which allows one to efficiently compute canonical heights on elliptic curves.) We will consider this question of local decomposition and a computational algorithm in a subsequent paper.

6 Integral points and Vojta's conjecture

A brief look at Table 1 in the previous section reveals a somewhat surprising pattern. Suppose we agree to write points $P \in S(\mathbb{Q})$ in the normalized form

$$(*) \qquad P=([x_1,x_2,x_3],[y_1,y_2,y_3]) \quad \text{with } x_1,x_2,x_3,y_1,y_2,y_3 \in \mathbb{Z}$$
$$\text{and} \quad \gcd(x_1,x_2,x_3)=\gcd(y_1,y_2,y_3)=1.$$

Then the larger points in Table 1 seem to satisfy

$$|x_1| \approx |x_2| \approx |x_3| \quad \text{and} \quad |y_1| \approx |y_2| \approx |y_3|.$$

This is similar to a phenomenon which occurs for elliptic curves. Thus suppose we take an elliptic curve

$$E: y^2z = 4x^3 - g_2xz^2 - g_3z^3$$

defined over $\mathbb{Q}$ and a sequence of distinct points $P_n=[x_n, y_n, z_n] \in E(\mathbb{Q})$ written as above in normalized form. Such a point is integral if $|z_n|=1$. A famous theorem of Siegel says that there are only finitely many integral points. But Siegel actually proved the considerably stronger result that

$$\lim_{n\to\infty} \frac{\log \min\{|x_n|,|y_n|,|z_n|\}}{\log \max\{|x_n|,|y_n|,|z_n|\}} = 1.$$

(See [12, IX.3.3].)

The proof of Siegel's theorem uses the Thue-Siegel-Roth theorem. The following higher dimensional analogue of the Thue-Siegel-Roth theorem has been conjectured by Paul Vojta.

Conjecture (Vojta, [15]) *Let K be a number field, Σ a finite set of places of K, V/K a smooth projective variety, $A \in \mathrm{Div}(V)$ an ample divisor, $D \in \mathrm{Div}(V)$ an effective divisor with at worst normal crossings. There is a proper Zariski closed subset $Z=Z(V,D) \subset V$ so that for every $\varepsilon>0$,*

$$\sum_{v\in\Sigma} \lambda_D(P,v) + h_{\mathcal{K}_V}(P) \leqq \varepsilon h_A(P) + O(1) \qquad \text{for all } P \in V(K),\ P \notin Z.$$

Here $\mathcal{K}_V$ is a canonical divisor on V, and $\lambda_D(\cdot, v)$ is a local height function on V for the divisor D and absolute value v. (Cf. [9, Chap. 10].) *The $O(1)$ constant depends on K, Σ, V, A, D, ε.*

Remark. Vojta states his conjecture with $Z=Z(K,\Sigma,V,A,D,\varepsilon)$, and says that the components of Z, other than individual points, can be chosen to depend only on V and D. We have merely adjusted the $O(1)$ to accomodate the finitely many isolated points in the original Z.

Taking the special case of Vojta's conjecture applied to K3 surfaces, we can prove the following result, which (conjecturally) explains the pattern in Table 1.

Proposition 6.1 *Let $S/\mathbb{Q}$ be a* K3 *surface as described in Sect. 1, and normalize the coordinates of $P \in S(\mathbb{Q})$ as described above in* (*). *Let h be a height function*

on S corresponding to any ample divisor. Assume that Vojta's conjecture is true for S. Then there is a finite union of curves $Z \subset S$ so that

$$\lim_{\substack{P \in S(\mathbb{Q}) \setminus Z \\ h(P) \to \infty}} \frac{\log \min\{|x_1(P)|, |x_2(P)|, |x_3(P)|\}}{\log \max\{|x_1(P)|, |x_2(P)|, |x_3(P)|\}} = 1.$$

A similar statement holds with x_1, x_2, x_3 replaced by y_1, y_2, y_3.

Proof. Let $f \in \mathbb{Q}[X_1, X_2, X_3]$ be a non-zero linear form, and let $H = \{f = 0\} \subset \mathbb{P}^2$ be the corresponding hyperplane section. We observe that the local height of a point $\mathbf{t} = [t_1, t_2, t_3] \in \mathbb{P}^2$ relative to the divisor H is given by the simple formula

$$\lambda_{\mathbb{P}^2, H}(\mathbf{t}, v) = \log \max\left\{\left|\frac{t_1}{f(\mathbf{t})}\right|_v, \left|\frac{t_2}{f(\mathbf{t})}\right|_v, \left|\frac{t_3}{f(\mathbf{t})}\right|_v\right\}.$$

(See [15, p. 8].) On the other hand, if $\mathbf{t} \in \mathbb{P}^2(\mathbb{Q})$ is normalized, then

$$h_{\mathbb{P}^2, H}(\mathbf{t}) = \log \max\{|t_1|, |t_2|, |t_3|\} + O(1),$$

where the $O(1)$ depends only on the linear form f defining H.

We are going to apply Vojta's conjecture with

$$V = S, \quad A = D_1 + D_2, \quad \text{and} \quad D = \pi_1^* H.$$

We observe first that

$$\alpha h_{D_1} = h_{D_2} + h_{E^+} + O(1) = h_{D_2} + \hat{h}^+ + O(1) \geqq h_{D_2} + O(1) \quad \text{since } \hat{h}^+ \geqq 0.$$

(Note we could not conclude directly from general principles that $h_{D_1} \gg h_{D_2}$, since D_1 is not ample.) Hence if we write $P = (\mathbf{x}, \mathbf{y})$ normalized as above, then

$$\begin{aligned} h_{S,A}(P) &= h_{S,D_1}(P) + h_{S,D_2}(P) + O(1) \\ &\leqq (\alpha + 1) h_{S,D_1}(P) + O(1) \\ &\leqq 4 h_{S,\pi_1^* H}(P) + O(1), \quad \text{since } D_1 \sim \pi_1^* H, \\ &= 4 h_{\mathbb{P}^2, H}(\mathbf{x}) + O(1) \\ &= 4 \log \max\{|x_1|, |x_2|, |x_3|\} + O(1). \end{aligned}$$

Now we apply Vojta's conjecture, using the field $\mathbb{Q}$ and the set consisting of the archimedean absolute value $\{\infty\}$. The canonical bundle on a K3 surface is trivial, so $\mathcal{K}_S = 0$. Hence there is a proper Zariski closed subset $Z \subset S$ so that

$$\lambda_{S,\pi_1^* H}(P, \infty) \leqq \varepsilon h_{S,A}(P) + O(1) \quad \text{for all } P \in S(\mathbb{Q}),\ P \notin Z.$$

Now $\lambda_{S,\pi_1^* H} = \lambda_{\mathbb{P}^2, H} \circ \pi_1 + O(1)$ by functoriality. So combining Vojta's inequality with the formula for $\lambda_{\mathbb{P}^2, H}$ and the estimate for $h_{S,A}(P)$ derived above, we find that

$$(1 - 4\varepsilon) \log \max\{|x_1|, |x_2|, |x_3|\} \leqq \log |f(\mathbf{x})| + O(1).$$

This holds for all $P=(\mathbf{x}, \mathbf{y})\notin Z$, where $Z\subset S$ is fixed, and the $O(1)$ depends on S, f, and ε.

Applying this result three times, taking successively $f=X_1$, $f=X_2$, and $f=X_3$, we conclude that

$$(1-4\varepsilon)\log\max\{|x_1|,|x_2|,|x_3|\}$$
$$\leqq\log\min\{|x_1|,|x_2|,|x_3|\}+O_\varepsilon(1), \quad \text{for all } P\in S(\mathbb{Q}),\ P\notin Z.$$

Hence

$$1-4\varepsilon-O_\varepsilon\left(\frac{1}{h(P)}\right)\leqq\frac{\log\min\{|x_1|,|x_2|,|x_3|\}}{\log\max\{|x_1|,|x_2|,|x_3|\}}\leqq 1, \quad \text{for all } P\in S(\mathbb{Q}),\ P\notin Z.$$

Now let $h(P)\to\infty$ with $P\in S(\mathbb{Q})$, $P\notin Z$. This gives

$$1-4\varepsilon\leqq\liminf_{\substack{P\in S(\mathbb{Q})\setminus Z\\ h(P)\to\infty}}\frac{\log\min\{|x_1|,|x_2|,|x_3|\}}{\log\max\{|x_1|,|x_2|,|x_3|\}}$$
$$\leqq\limsup_{\substack{P\in S(\mathbb{Q})\setminus Z\\ h(P)\to\infty}}\frac{\log\min\{|x_1|,|x_2|,|x_3|\}}{\log\max\{|x_1|,|x_2|,|x_3|\}}\leqq 1.$$

Since this holds for every $\varepsilon>0$, the limit exists and equals 1. □

Remark. A proof of Vojta's conjecture, even for K3 surfaces, seems difficult, although possibly not hopeless in view of the recent advances made by Vojta [16] and Faltings [4]. A more tractable problem might be to show that on the K3 surfaces studied in this paper, if one fixes a chain $\mathscr{C}$ of rational points, then the conclusion of Proposition 6.1 holds if the limit is taken over points in $\mathscr{C}$. Note that this does not reduce to Siegel's theorem (i.e. integral points on a curve), since we have shown that $\mathscr{C}$ is Zariski dense in S. But one might hope to use the additional structure provided by the group of automorphisms $\mathscr{A}$ acting on $\mathscr{C}$ and $\mathrm{Pic}(S)$, just as the group structure on an abelian variety is exploited in the proofs of Faltings and Vojta.

Acknowledgements. I would like to thank Alan Landman, Yuri Manin, Gisbert Wüstholz, and the referee for their valuable assistance and suggestions.

References

1. Arbarello, E., Cornalba, M., Griffiths, P.A., Harris, J.: Geometry of algebraic curves, vol. 1. (Grund. math. Wiss., vol. 267.) Berlin Heidelberg New York: Springer 1985
2. Batyrev, V.V., Manin, Yu.: Sur le nombre des points rationnels de hauteur borné des variétés algébriques. Math. Ann. **286**, 27–44 (1990)
3. Beauville, A.: Complex Algebraic Surfaces. (Lond. Math. Soc. Lect. Note Ser., vol. 68.) Cambridge: Cambridge University Press 1983
4. Faltings, G.: Diophantine approximation on abelian varieties. Ann. Math. (to appear)
5. Franke, J., Manin, Yu., Tschinkel, Yu.: Rational points of bounded height on Fano varieties. Invent. Math. **95**, 421–435 (1989)
6. Griffiths, P., Harris, J.: Algebraic Geometry. New York: Wiley Interscience 1978

7. Griffiths, P., Harris, J.: On the Noether-Lefschetz theorem and some remarks on codimension-two cycles. Math. Ann. **271**, 31–51 (1985)
8. Hartshorne, R.: Algebraic Geometry. Berlin Heidelberg New York: Springer 1977
9. Lang, S.: Fundamentals of Diophantine Geometry. Berlin Heidelberg New York: Springer 1983
10. Lefschetz, S.: L'analysis situs et la géométrie algébrique. Paris: Gauthier-Villars 1924
11. Mori, S., Mukai, Sh.: The uniruledness of the moduli space of curves of genus 11 (Lect. Notes Math., vol. 1016, pp. 334–353). Berlin Heidelberg New York: Springer 1983
12. Silverman, J.H.: The Arithmetic of Elliptic Curves. Berlin Heidelberg New York: Springer 1986
13. Silverman, J.H.: Computing heights on elliptic curves. Math. Comput. **51**, 339–358 (1988)
14. Swinnerton-Dyer, H.P.F.: Applications of algebraic geometry to number theory. Proc. Symp. Pure Math., vol. 20) Providence: Am. Math. Soc. 1971
15. Vojta, P.: Diophantine approximations and value distribution theory (Lect. Notes Math., vol. 1239). Berlin Heidelberg New York: Springer 1987
16. Vojta, P.: Siegel's theorem in the compact case. Ann. Math. (to appear)
17. Wehler, J.: K 3-surfaces with Picard number 2. Arch. Math. **50**, 73–82 (1988)
18. Wehler, J.: Hypersurfaces of the Flag Variety. Math. Z. **198**, 21–38 (1988)

Invent. math. 105, 375–400 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Prescribing harmonic measure on convex domains

David Jerison *

MIT, Department of Mathematics, Cambridge, MA 02139, USA

Oblatum 20-XI-1990

Introduction

Let Ω be a convex, open subset of $\mathbb{R}^N$. The Gauss map $g: \partial\Omega \to S^{N-1}$, defined almost everywhere with respect to surface measure $d\sigma$, is given by $g(x)=\theta$, where θ is the outer unit normal at $x \in \partial\Omega$. The classical Minkowski problem asks for which measures ν on S^{N-1} one can find a convex Ω such that $g_*(d\sigma)=d\nu$, that is, $\nu(E)=\sigma(g^{-1}(E))$ for all Borel sets $E \subset S^{N-1}$. Our goal in this paper is to study the analogous problem with harmonic measure in place of surface measure.

Suppose that the origin 0 belongs to Ω. Harmonic measure for Ω at 0 is the measure ω on $\partial\Omega$ such that for all continuous functions f on $\partial\Omega$,

$$u(0)=\int_{\partial\Omega} f\,d\omega,$$

where u solves the Dirichlet problem: $\Delta u=0$ in Ω, $u=f$ on $\partial\Omega$. (If Ω is unbounded, we suppose further that f tends to zero at infinity and u is the unique solution that tends in zero at infinity.)

Convexity implies, in particular, that $\partial\Omega$ is locally Lipschitz. It follows from Dahlberg's theorem [D] that $d\omega$ and $d\sigma$ are mutually absolutely continuous. Thus for any convex open Ω containing the origin we can define a measure μ on S^{N-1} by $\mu(E)=\omega(g^{-1}(E))$ for all Borel sets $E \subset S^{N-1}$ (i.e., $g_*(d\omega)=d\mu$). Harmonic measure has total mass 1, so the problem of prescribing harmonic measure on a convex domain can be stated: *Given a probability measure μ on S^{N-1}, find a convex open set Ω such that $g_*(d\omega)=d\mu$.*

When $N=2$, a complete existence, uniqueness, and regularity theory for this problem follows from a continuous version of the Schwarz-Christoffel conformal mapping formula. For $N \geqq 2$, we construct a solution in the smooth case by the method of continuity and obtain the general solution by a limiting procedure. Our proof depends not only on the regularity theory in the Minkow-

* Supported in part by NSF grant DMS-8804582 and a Presidential Young Investigator Award DMS-8451770. This research was begun at the Mathematical Sciences Research Institute at Berkeley.

ski problem due to Alexandrov, Pogorelov, Nirenberg, Calabi, and Cheng-Yau [CY1, CY2] but also on recent sharp estimates for the Monge-Ampère equation due to L. Caffarelli [C1, C2, C3, C4]. The proof also depends on a new doubling property for harmonic measure in convex domains (Theorem 2).

Theorem 1 (for large k) was announced in [J1], but the proof there was much more complicated. Motivated in part by our original proof of Theorem 1, Caffarelli [C3] has greatly simplified his own proof of $C^{1,\alpha}$ regularity for solutions to Monge-Ampère equations. As a result, our proof of Theorem 1 can be simplified so that the only estimate on harmonic measure we need is the doubling property. The other results announced in [J1] but no longer needed here, namely, estimates for harmonic measure in convex and strictly convex C^1 domains and L^p estimates for second derivatives of solutions to Monge-Ampère equations will appear in [J2].

Acknowledgements. I would like to thank Luis Caffarelli for explaining his results to me and for his encouragement. I would also like to thank Björn Dahlberg, Charles Epstein, and Pierre Pansu for helpful comments. Finally, I would like to thank L.N. Trefethen for disproving the conjecture in Sect. 7(b).

1 Main results and an overview

The main result is existence, uniqueness, and regularity in Hölder classes.

Theorem 1 *For any non-negative integer k and any* α, $0<\alpha<1$, *let* $R\in C^{k,\alpha}(S^{N-1})$ *be positive and satisfy* $\int_{S^{N-1}} R\,d\theta=1$, *where* $d\theta$ *is the standard volume element on the unit sphere* S^{N-1}. *There exists a bounded, strongly convex open subset* Ω *of* $\mathbb{R}^N$, *containing the origin, with* $C^{k+2,\alpha}$ *boundary, such that for all Borel sets* $E\subset S^{N-1}$,

$$\omega(g^{-1}(E))=\int_E R\,d\theta, \tag{1.1}$$

where g is the Gauss map and ω *is harmonic measure for* Ω *at* 0. *Moreover the domain* Ω *is unique up to dilation.*

Corollary. *Let* μ *be a probability measure on* S^{N-1}. *There exists a convex, open set* $\Omega\subset\mathbb{R}^N$, *such that* $\omega(g^{-1}(E))=\mu(E)$ *for all Borel sets* $E\subset S^{N-1}$.

We have not established uniqueness up to dilation in the case of a general measure when $N\geqq 3$. (For the case $N=2$ see Sect. 7.)

The proof of Theorem 1 has two parts. The first part is a global calculation to invert a linearized equation in which the key step is an integration by parts formula (4.1). The second part consists of *a priori* estimates saying that the $C^{k,\alpha}$ norm of R controls the $C^{k+2,\alpha}$ norm of $\partial\Omega$. In order to explain these estimates, we need to write down the equation we are solving.

We can assume that if U is a sufficiently small neighborhood of a point of $\partial\Omega$, then, after rotation,

$$U\cap\Omega=U\cap\{(x_0,x)\in\mathbb{R}\times\mathbb{R}^n: x_0>w(x)\} \tag{1.2}$$

for a convex function $w(x_1, \ldots, x_n)$ on $\mathbb{R}^n$, $n=N-1$. Suppose that $w\in C^{2,\alpha}$ and that Ω is *strongly convex*, i.e., $\partial^2 w/\partial x_i \partial x_j$ is positive definite. Recall that Gauss curvature K is the Jacobian determinant of the Gauss map. Using the Gauss map to identify $\partial\Omega$ with S^n, we may write K as a function on S^n instead of $\partial\Omega$. Thus $g_*(\mathrm{d}\sigma)=(1/K)\,\mathrm{d}\theta$ and

$$\det(\partial^2 w(x)/\partial x_i \partial x_j)=(1+|\nabla w(x)|^2)^{\frac{N+1}{2}} K(\theta), \tag{1.3}$$

where $\theta=(-1, \nabla w(x))/\sqrt{1+|\nabla w(x)|^2}$.

Equation (1.3) is a local presentation of the Minkowski equation for which a rather complete existence and regularity theory has been developed.

Let h denote the normal derivative of Green's function for Ω with pole at 0. Then $\mathrm{d}\omega/\mathrm{d}\sigma=h$ and Eq. (1.1) can be written locally as

$$\det(\partial^2 w(x)/\partial x_i \partial x_j)=(1+|\nabla w(x)|^2)^{\frac{N+1}{2}} h(w(x), x)/R(\theta). \tag{1.4}$$

Because the factor h is a first derivative of Green's function, it should be viewed as a nonlinear first-order pseudodifferential operator on w. (One can also see this from the linearization of 1.4 in Sect. 3.) The regularity properties of h are only slightly worse than those of R and $1+|\nabla w|^2$. In particular, it is easy to show that a positive lower bound on R implies an upper bound on $|\nabla w|$, and a comparison of Ω with a half-space shows that h is bounded above. But for arbitrary convex domains h may vanish. The key is to find an appropriate lower bound which fits the regularity theory of the Monge-Ampère equation.

The appropriate estimate on h (or $\omega=h\,\mathrm{d}\sigma$) is expressed in terms of the sets $\{x\in\mathbb{R}^n: w(x)\leqq a\cdot x+b\}$ for $a\in\mathbb{R}^n$, $b\in\mathbb{R}$. These sets correspond to the intersection of $\partial\Omega$ with half-spaces.

For $a\in\mathbb{R}^N$, $b\in\mathbb{R}$, denote

$$F=F(a, b)=\{x\in\partial\Omega: a\cdot x>b\},$$

$\pi=$*the orthogonal projection of* $\mathbb{R}^N$ *onto the hyperplane* $x\cdot a=b$,

$$E=\{x\in\bar{\Omega}: a\cdot x=b\}$$

$$\tfrac{1}{2}F=\{x\in F: \pi x\in\tfrac{1}{2}E\}.$$

$B_r=$the open ball centered at 0 in $\mathbb{R}^N$ of radius r.

The doubling property for a measure μ on $\partial\Omega$ is

$$\mu(F)\leqq M\,\mu(\tfrac{1}{2}F) \tag{1.5}$$

for every $F=\{x\in\partial\Omega: a\cdot x>b\}$ *with* $a\in\mathbb{R}^N$ *and* $b\geqq 0$.

Theorem 2 *Let* Ω *be a convex, open subset of* $\mathbb{R}^N$. *Suppose that* $B_r\subset\Omega\subset B_1$. *Let* ω *be harmonic measure for* Ω *at* 0. *Then* ω *satisfies the doubling property* 1.5 *with a constant* M *depending only on* N *and* $r>0$.

The paper is organized as follows. In Sect. 2 we discuss the Minkowski support function and regularity for Monge-Ampère equations. In Sect. 3 we calculate the linearized operator L for our prescribed curvature problem. In

Sect. 4 we find the image and kernel of L. In Sect. 5 we prove the doubling property for harmonic measure, and in Sect. 6 we deduce the main existence and uniqueness theorem. In Sect. 7 we show how the Schwarz-Christoffel formula solves our problem completely in $\mathbb{R}^2$. Next we show that a concavity property conjectured in [J1] is false. Finally, we discuss an exterior problem which may have even deeper parallel structure with the Minkowski problem.

2 Monge-Ampère equations and Gauss curvature

Let k be a non-negative integer, and let $0<\alpha<1$. Let Ω be a bounded, strongly convex open subset of $\mathbb{R}^N$ with $C^{k+2,\alpha}$ boundary. The Gauss map is invertible, so we can define the Minkowski support function $u\colon S^n\to\mathbb{R}$ ($n=N-1$) by $u(\theta)=x\cdot\theta$, where x is the unique point of $\partial\Omega$ with outer normal θ. Extend u to a function $\bar{u}(y)=|y|\,u(y/|y|)$, which is homogeneous of degree 1 in $\mathbb{R}^N$. Then $V\bar{u}(y)$ is homogeneous of degree 0. We denote $F(\theta)=V\bar{u}(\theta)$, the restriction of $V\bar{u}$ to S^n. One can check that F maps S^n to $\partial\Omega$ and inverts g. (See [CY1].) So far we only know that $u\in C^{k+1,\alpha}(S^n)$, but we shall see eventually that $u\in C^{k+2,\alpha}(S^n)$.

In addition to the formula (1.3) for $K(\theta)$, we need to express K in two other ways. Following [CY1], we can express K in the lower hemisphere of S^n in terms of the restriction of $\bar{u}(y_0, y_1, \ldots, y_n)$ to $y_0=1$.

$$F(\theta)=\left(\sum_{j=1}^{n} y_i\frac{\partial\bar{u}}{\partial y_i}(-1,y)-\bar{u}(-1,y),\frac{\partial\bar{u}}{\partial y_1}(-1,y),\ldots,\frac{\partial\bar{u}}{\partial y_n}(-1,y)\right), \tag{2.1}$$

where $\theta=(-1,y)/\sqrt{1+|y|^2}$, $y=(y_1,\ldots,y_n)\in\mathbb{R}^n$. Assuming that $u\in C^2(S^n)$,

$$\det\left(\frac{\partial^2\bar{u}}{\partial y_i\,\partial y_j}(-1,y_1,\ldots,y_n)\right)(1+|y|^2)^{\frac{n}{2}+1}=1/K(\theta). \tag{2.2}$$

Evidently, a similar equation holds in any other hemisphere.

Let $\nabla_i u$, $\nabla_{ij}u$ denote covariant derivatives on S^n with respect to an orthonormal frame. Then (assuming $u\in C^2(S^n)$)

$$\det(\nabla_{ij}u+u\,\delta_{ij})=1/K. \tag{2.3}$$

The strong convexity of w implies that (1.3) is a uniformly elliptic (fully nonlinear) equation. Likewise $\partial^2\bar{u}/\partial y_i\,\partial y_j$ and $\nabla_{ij}u+u\,\delta_{ij}$ are positive definite, and (2.2) and (2.3) are uniformly elliptic, under the hypothesis of strong convexity.

Proposition 1 *Let k be a non-negative integer, $0<\alpha<1$. If Ω is a bounded, strongly convex $C^{k+2,\alpha}$ domain, then its support function u belongs to $C^{k+2,\alpha}(S^n)$ and $\nabla_{ij}u+u\,\delta_{ij}\geqslant 0$. Conversely, if $u\in C^{k+2,\alpha}(S^n)$ and $\nabla_{ij}u+u\,\delta_{ij}\geqslant 0$, then u defines a strongly convex $C^{k+2,\alpha}$ domain (via the map $F=V\bar{u}$). Moreover, the Gauss map is a $C^{k+1,\alpha}$ diffeomorphism between $\partial\Omega$ and S^n.*

Proof. Suppose that $w\in C^{k+2,\alpha}$. Then u and $\bar{u}$ belong to $C^{k+1,\alpha}$ and $K\in C^{k,\alpha}$. If $k\geqq 1$, then Eq. (2.2) holds and using standard Schauder theory we find that $\bar{u}\in C^{k+2,\alpha}$. (Differentiate or take difference quotients; see [GT, Lemma 17.16].)

Conversely, if $u \in C^{k+2,\alpha}$, then $\bar{u} \in C^{k+2,\alpha}$, $K \in C^{k,\alpha}$, and $w \in C^{k+1,\alpha}$. If $k \geqq 1$, then we can use standard Schauder theory in 1.3 to deduce that $w \in C^{k+2,\alpha}$.

In the case $k=0$, the same outline works, but it depends on a recent regularity result of Caffarelli for viscosity solutions to the Monge-Ampère equation. A *convex viscosity solution* v to the equation $\det(\partial^2 v/\partial x_i \partial x_j) = f$ is a convex continuous function v such that *for any* $\varphi \in C^2$, $\det(\partial^2 \varphi/\partial x_i \partial x_j) \geqq f(x)$ (resp. $\leqq f(x)$) *implies* $v - \varphi$ *cannot have an interior minimum (resp. maximum).* Caffarelli [C2, Theorem 2b'] has shown that convex viscosity solutions v with $f \in C^\alpha$, $f > 0$, and for which the level sets of v are bounded have the optimal regularity $v \in C^{2,\alpha}$.

It is easy to check that uniform limits of classical solutions are viscosity solutions. Let $\partial\Omega$ be $C^{2,\alpha}$; then we may choose a sequence $\partial\Omega_j$ in C^∞ converging to $\partial\Omega$ in $C^{2,\alpha}$. Equation (2.2) is valid for the associated functions $\bar{u}_j$. By (1.3), the Gauss curvature K_j tends to K in C^α and $\bar{u}_j$ tends to $\bar{u}$ in $C^{1,\alpha}$ (locally). In particular, we have uniform convergence, so we have optimal regularity $\bar{u} \in C^{2,\alpha}$. A similar argument shows that $u \in C^{2,\alpha}$ implies $\partial\Omega$ belongs to $C^{2,\alpha}$. This proves the proposition.

Let $\mathscr{P}_1$ be the family of degree 1 spherical harmonics, that is, the restriction to S^n of any homogeneous degree 1 polynomial. If $v \in \mathscr{P}_1$, then $\bar{v}$ is homogeneous of degree 1 on $\mathbb{R}^N$ and $\nabla \bar{v}$ is a constant vector. Thus $\nabla(\bar{u} + \bar{v})$ is a translate of $\nabla \bar{u}$. Furthermore, since $\partial^2 \bar{v}(-1, y)/\partial y_i \partial y_j = 0$, Eq. (2.2) is unchanged. It is easy to check that $\nabla_{ij} v + v\delta_{ij} = 0$ for all $i, j = 1, \ldots, n$, and all $v \in \mathscr{P}_1$. Thus Eq. (2.3) is also unchanged. We can always translate Ω so that the origin 0 belongs to Ω. This will be true if and only if the support function u for Ω is positive.

Let us now describe some lower order regularity estimates due to Cafarelli. Let V be an open subset of $\mathbb{R}^n$ and let $w \in C^{2,\alpha}(V)$ be convex. If $E = \{x \in V: w(x) - a \cdot x - b \leqq 0\}$ is closed, then it is convex. Let $\frac{1}{2}E$ be the dilate of E by the factor $\frac{1}{2}$ with origin the center of mass of E. We will say that $w \in V(C)$ if $\det(\partial^2 w/\partial x_i \partial x_j) = f$ satisfies

$$\int_E f(x)\,dx \leqq C \int_{\frac{1}{2}E} f(x)\,dx \tag{2.4}$$

for all closed sets E as above.

Theorem 3 [C3, Theorem 1'] *Suppose that w_k is a sequence of functions in $V(C)$ tending uniformly to w. Suppose that $w(x^0) = a \cdot x^0 + b$ and $w(x) - a \cdot x - b \geqq 0$ for all $x \in V$. Then the set $\{x \in V: w(x) = a \cdot x + b\}$ is either the single point x^0 or else it has no extremal points in V.*

A domain Ω is *strictly convex* if its boundary contains no line segments. (This is to be distinguished from the stronger property "strong convexity".) We define a modulus of strict convexity η: $[0, \infty) \to [0, \infty)$ as a monotone increasing function such that $\eta(s) > 0$ for $s > 0$ and for which $\inf_S \max_{x \in S} d(x, \partial\Omega) \geqq \eta(s)$, where the infimum is taken over all closed line segments $S \subset \bar{\Omega}$ of length s.

Theorem 3' *Let Ω be a bounded, strongly convex, $C^{2,\alpha}$ open set in $\mathbb{R}^N$. Suppose that $B_r \subset \Omega \subset B_1$ and suppose that $K\,d\sigma$ satisfies the doubling property* (1.5) *with constant M. Then Ω has a modulus of strict convexity η depending only on M, N, and $r > 0$.*

Proof. $B_r \subset \Omega \subset B_1$ implies that for every $x \in \partial\Omega$ there is a rotation so that $B_{r/10}(x) \cap \Omega$ is represented as in (1.2) by the region above the graph of a convex Lipschitz function w with Lipschitz norm $|\nabla w| \leqq 2/r$. It follows from (1.3) that in every ball of radius $r/10$, w belongs to the class $V(C)$ for an appropriate open set V and C depends only on M and r. The claim is not entirely obvious because the hyperplanes in which we defined $\frac{1}{2}E$ in (2.4) and $\frac{1}{2}E$ preceding the doubling property (1.5) need not be parallel. To explain why we can still make this claim we need several geometric remarks. First of all the factor $\frac{1}{2}$ we chose was unimportant: any fixed factor between 0 and 1 suffices in all situations. Second, the bound on ∇w does imply that the two hyperplanes form an angle bounded away from perpendicular. Finally, for any convex set E of $\mathbb{R}^n$ with non-empty interior, there are concentric ellipsoids E_1 and E_2 so that E_2 is the dilate of E_1 by a dimensional constant and $E_1 \subset E \subset E_2$. (F. John's lemma; see [dG].) Thus any convex subset E' of E with comparable eccentricity and comparable size is such a comparable dilate of E' contains E. Using the doubling property (1.5) for a larger factor than, it is now easy to see that W belongs to $V(C)$.

If Theorem 3′ were false, then there would exist a sequence Ω_k and $S_k \subset \bar{\Omega}_k$ of line segments of some fixed lengths such that $\max_{x \in S_k} d(x, \partial\Omega_k) \to 0$. But then a subsequence would converge uniformly and this contradicts Theorem 3.

Suppose that P_0 is the tangent plane to $\partial\Omega$ at x and P_1 and P_2 are parallel planes intersecting Ω at distance s and $2s$ from P_0. Suppose that $s < r/4$. Strict convexity implies that $(\partial\Omega) \cap P_1$ does not intersect the cone from x to $\bar{\Omega} \cap P_2$. A limiting argument gives a quantitative version of this statement and an iteration yields and estimate on the $C^{1,\alpha'}$ norm of $\partial\Omega$:

Theorem 4 [C3, Theorem 2] *Under the hypotheses of Theorem* 3′, *there is* $\alpha' > 0$ *such that* α' *and the* $C^{1,\alpha'}$ *norm of* $\partial\Omega$ *depend only on* M, N, *and* r.

The passage from $C^{1,\alpha}$ to $C^{2,\alpha}$ and higher regularity (which is essential to the existence and uniqueness as well as the regularity asserted in Theorem 1) still requires a more complicated iteration argument, which in turn depends on the entire development of regularity in the Minkowski problem.

Theorem 5 [C2, Theorem 4b′, C4] *Let* $k \geqq 0$, $0 < \alpha < 1$. *Let* $K > 0$ *be a* $C^{k,\alpha}$ *function on* $\partial\Omega \times S^N$ *such that* $K(x, g(x))$ *is the Gauss curvature for a strictly convex region* Ω *in the viscosity sense* (*i.e., Eq.* (1.3) *is satisfied in the viscosity sense*), *then* Ω *is strongly convex and* $C^{k+2,\alpha}$ *with norm depending only on* N, *the lower bound for* K *and its* $C^{k,\alpha}$ *norm.*

3 First variation

Denote $\mathcal{U} = \{u \in C^{k+2,\alpha}(S^n) : u > 0 \text{ and } \nabla_{ij} u + u\,\delta_{ij} \geqslant 0\}$. We showed in Sect. 2 that $\mathcal{U}$ is a parametrization for the set of bounded, strongly convex $C^{k+2,\alpha}$ domains containing the origin; $\partial\Omega$ is the image under $F = \nabla \bar{u}$ of S^n. Denote $\mathcal{R} = \{R \in C^{k,\alpha}(S^n) : \int R\,d\theta = 1\}$. Given $u \in \mathcal{U}$, let harmonic measure at 0 for the domain Ω with support function u be denoted ω. Define $\mathcal{F}$ by $\mathcal{F}(u) = R$, where $g_*(d\omega) = R\,d\theta$. The main result of this section is the computation of the first variation of $\mathcal{F}$.

Let $v \in C^{k+2,\alpha}(S^n)$ and $\bar{v}(y) = |y| v(y/|y|)$ for $y \in \mathbb{R}^N$. For $t \in \mathbb{R}$, denote

$$F(t, \theta) = \nabla(\bar{u} + t\bar{v})(\theta).$$

For t near 0, let Ω^t be the strongly convex domain with support function $u + tv$. ($\partial\Omega^t = F(t, S^n)$; $\Omega = \Omega^0$.) Denote by c_{ij} the cofactor matrix of $\nabla_{ij} u + u\delta_{ij}$, i.e.,

$$c_{ij}(\nabla_{j\ell} u + u\delta_{j\ell}) = \delta_{i\ell} \det(\nabla_{pq} u + u\delta_{pq}) = \delta_{i\ell}/K. \tag{3.1}$$

Here and henceforth we use the summation convention: the repeated index j denotes summation $j = 1, \ldots, n$. The Neumann operator Λ: $C^{k+1,\alpha}(\partial\Omega) \to C^{k,\alpha}(\partial\Omega)$ is given by $(\Lambda f)(x) = \theta \cdot \nabla w(x)$ for $x \in \partial\Omega$, where w satisfies $\Delta w = 0$ in Ω and $w = f$ on $\partial\Omega$. We prove

Proposition 2 *The mapping $\mathscr{F}$ is continuously differentiable from $\mathscr{U}$ to $\mathscr{R}$ (with the usual $C^{k+2,\alpha}$ and $C^{k,\alpha}$ norms respectively). Moreover, its directional derivative is*

$$(\mathrm{d}/\mathrm{d}t)\,\mathscr{F}(u + tv)|_{t=0} = Lv,$$

where $L = L_u$ is given by

$$Lv = c_{ij} \nabla_i (h \nabla_j v) - \frac{1}{K} \Lambda(hv).$$

(From here onward we will often identify functions on S^n and $\partial\Omega$ using the Gauss map correspondence. Thus h is initially defined on $\partial\Omega$ and v on S^n, but we can take covariant derivatives of h on S^n and operate by Λ on v.)

The proof will occupy the remainder of this section. Let us begin by explaining why R belongs to $C^{k,\alpha}(S^n)$ when u belongs to $C^{k+2,\alpha}(S^n)$. Recall the classical estimate

$$\|w\|_{C^{k+2,\alpha}(\bar{\Omega})} \leqq C(\|\Delta w\|_{C^{k,\alpha}(\bar{\Omega})} + \|w\|_{C^{k+2,\alpha}(\partial\Omega)}), \tag{3.2}$$

with C depending only on the $C^{k+2,\alpha}$ norm of $\partial\Omega$.

Let G be Green's function for Ω with pole at 0: $G(x) = -C_N |x|^{2-N} + A(x)$, where

$$\Delta A(x) = 0 \quad \text{for } x \in \Omega, \qquad A(x) = C_N |x|^{2-N} \quad \text{for } x \in \partial\Omega,$$

and $C_N = 1/(N-2)\,\mathrm{vol}(S^n)$, $N \geqq 3$; when $N = 2$

$$G(x) = (1/2\pi) \log |x| + A(x),$$

$\Delta A(x) = 0$ in Ω, $A(x) = -(1/2\pi)\log|x|$ for $x \in \partial\Omega$. By 3.2, $A \in C^{k+2,\alpha}(\bar{\Omega})$ and $h = \theta \cdot \nabla G(x) = |\nabla G(x)| \in C^{k+1,\alpha}(\partial\Omega)$. Since $\mathrm{d}\omega/\mathrm{d}\sigma = h$ and $g_*(\mathrm{d}\sigma) = (1/K)\,\mathrm{d}\theta$, we see that

$$R(\theta) = h(F(\theta))/K(\theta), \tag{3.3}$$

$h \circ F \in C^{k+1,\alpha}(S^n)$ and from (1.3), $K \in C^{k,\alpha}(S^n)$, so that $R \in C^{k,\alpha}$.

Now we can use (2.3) to write a global equation for R in terms of u:

$$\mathscr{F}(u)(\theta) = h(g^{-1}(\theta)) \det(\nabla_{ij} u(\theta) + u(\theta)\delta_{ij}) = R(\theta). \tag{3.4}$$

Lemma A *Let $e_1, \ldots, e_n$ be an orthonormal frame for S^n for which we define covariant derivatives. Then*

$$\nabla^2 G(F(\theta))(e_i, e_j) = K h c_{ij} \tag{3.5}$$

$$\nabla^2 G(F(\theta))(e_i, \theta) = K c_{ij} \nabla_j h \tag{3.6}$$

$$\nabla^2 G(F(\theta))(\theta, \theta) = -K h \operatorname{trace}(c_{ij}). \tag{3.7}$$

(Here ∇^2 denotes the Hessian in $\mathbb{R}^N$ as a bilinear form on $\mathbb{R}^N$, and the vectors are viewed as vectors in $\mathbb{R}^N$ orthogonal to θ, i.e. tangent vectors to $\partial\Omega$ at $F(\theta)$.)

Remark. $K \operatorname{trace}(c_{ij})$ is the mean curvature of $\partial\Omega$ at $g^{-1}(\theta)$.

Proof. (3.7) follows immediately (3.5) since the equation $\Delta G = 0$ is valid on $\partial\Omega$. For the calculation we will assume that $u \in C^3(S^n)$. The lemma remains true for $u \in C^{2,\alpha}(S^n)$ by a routine limiting argument using the fact from Sect. 2 that if $u^\varepsilon \to u$ in $C^{2,\alpha}$, then the corresponding boundaries $\partial\Omega^\varepsilon$ approach $\partial\Omega$ in $C^{2,\alpha}$.

It suffices to make the calculation at a single point. We can assume after rotation that the point is the South Pole $P = (-1, 0) \in \mathbb{R} \times \mathbb{R}^n$. We use local coordinates $y \in \mathbb{R}^n$, $\theta = (-1, y)/\sqrt{1+|y|^2}$ and F given as in (2.1). Moreover, we can assume that at $y = 0$ our orthonormal basis is the standard basis $e_1 = (0, 1, 0, \ldots, 0)$, $e_2 = (0, 0, 1, 0, 0), \ldots, e_n = (0, \ldots, 0, 1)$. Then

$$\nabla_i u(P) = (\partial/\partial y_i)\, u(\theta)|_{y=0}$$

$$\nabla_{ij} u(P) = (\partial^2/\partial y_i\, \partial y_j)\, u(\theta)|_{y=0},$$

and since $\bar{u}(-1, y) = \sqrt{1+|y|^2}\, u(\theta)$,

$$(\partial/\partial y_i)\, \bar{u}(-1, 0) = \nabla_i u(P)$$

$$(\partial^2/\partial y_i\, \partial y_j)\, \bar{u}(-1, 0) = \nabla_{ij} u(P) + u(P)\, \delta_{ij}.$$

Abbreviate $a_{ij} = \nabla_{ij} u(P) + u(P)\, \delta_{ij}$.

$$\partial F^0/\partial y_k = y_i (\partial^2/\partial y_i\, \partial y_k)\, \bar{u}(-1, y),$$

$$\partial F^j/\partial y_k = y_i (\partial^2/\partial y_j\, \partial y_k)\, \bar{u}(-1, y), \qquad j = 1, \ldots, n,$$

$$\partial^2 F^0/\partial y_k\, \partial y_\ell = y_i (\partial^3/\partial y_i\, \partial y_k\, \partial y_\ell)\, \bar{u}(-1, y) + (\partial^2/\partial y_k\, \partial y_\ell)\, \bar{u}(-1, y).$$

Thus at $y = 0$,

$$(\partial/\partial y_k)\, F = a_{kj} e_j \quad \text{and} \quad (\partial^2/\partial y_k\, \partial y_\ell)\, F^0 = a_{k\ell}.$$

Differentiating $G(F(\theta)) = 0$ with respect to y twice we find

$$(\partial^2 F/\partial y_k\, \partial y_\ell) \cdot \nabla G(F(\theta)) + \nabla^2 G(F(\theta)))(\partial F/\partial y_k, \partial F/\partial y_\ell) = 0.$$

Recalling that at $y = 0$, $\theta = P$ and $\nabla G = hP$, we rewrite this equation as

$$-h a_{k\ell} + \nabla^2 G(F(P))(a_{kj} e_j, a_{\ell i} e_i) = 0.$$

To obtain (3.5), we multiply by $c_{kp}c_{\ell q}$ and sum. Finally, differentiating the equation $h(\theta)=\theta\cdot\nabla G(F(\theta))$ with respect to y_k at $y=0$, we obtain $\nabla_k h(P)=\nabla^2 G(F(P))(P, a_{kj}e_j)$. (3.6) follows if we multiply by c_{ik} and sum.

Let $G(t, x)$ be Green's function for Ω^t with pole at 0. Thus, for $N\geqq 3$, $G(t, x)=-c_N|x|^{2-N}+A(t, x)$ where $A(t, x)$ is harmonic in x for $x\in\Omega^t$ and $A(t, x)=c_N|x|^{2-N}$ for $x\in\partial\Omega^t$.

Lemma B *$(\partial/\partial t)A(t, x)$ is a continuous function of t with values in $C^{k+1,\alpha}(\bar{\Omega}^t)$. In particular, $\nabla_x(\partial/\partial t)A(t, x)$ is continuous in (t, x), so that $\nabla_x(\partial/\partial t)A(t, x)=(\partial/\partial t)\nabla_x A(t, x)$.*

Proof. Denote $H=g^{ij}\partial^2/\partial x_i\partial x_j+g^i\partial/\partial x_i$ an elliptic differential operator on $\bar{\Omega}$ with $g^{ij}\in C^{k,\alpha}(\bar{\Omega})$, $g^i\in C^{k-1,\alpha}(\bar{\Omega})$ $(k\geqq 1)$.

We have the boundary Schauder estimate [GT].

$$\|w\|_{C^{k+1,\alpha}(\bar{\Omega})}\leqq C(\|Hw\|_{C^{k-1,\alpha}(\bar{\Omega})}+\|w\|_{C^{k+1,\alpha}(\partial\Omega)}). \tag{3.8}$$

For $k=0$ we need a space to substitute for $C^{-1,\alpha}$. Fix s, $0<s<\alpha$. Define

$$B^{\alpha-1}(\Omega)=\{f\in C^s(\Omega): |f(x)-f(x')|\leqq C|x-x'|^s\max(\mathrm{d}(x, \partial\Omega)^{\alpha-1-s}, \mathrm{d}(x', \partial\Omega)^{\alpha-1-s})\}.$$

Suppose that $g^{ij}\in C^\alpha(\bar{\Omega})$, $g^i\in B^{\alpha-1}(\Omega)$. We then have [J3, L]

$$\|w\|_{C^{1,\alpha}(\bar{\Omega})}\leqq C(\|Hw\|_{B^{\alpha-1}(\Omega)}+\|w\|_{C^{1,\alpha}(\partial\Omega)}). \tag{3.9}$$

(The constants in 3.8 and 3.9 depend only on the ellipticity constant of g^{ij} and the norms of g^{ij} and g^i in the appropriate spaces.)

Let I be an interval of $\mathbb{R}$ containing 0. Let $\varphi: I\times\bar{\Omega}\to\mathbb{R}^N$ be such that $\varphi(t, \cdot)$ is a $C^{k+2,\alpha}$ diffeomorphism of $\bar{\Omega}$ onto $\bar{\Omega}^t$. We may also suppose that φ is continuously differentiable in t and

$$\varphi\in C^0(I; C^{k+2,\alpha}(\bar{\Omega})),$$

$$(\partial/\partial t)\nabla_x^j\varphi\in C^0(I; C^{k+1-j,\alpha}(\bar{\Omega})), \quad j=0, 1, 2.$$

When $k=0$, we need to replace the case $j=2$ by

$$(\partial/\partial t)\nabla_x^j\varphi\in C^0(I; B^{\alpha-1}(\Omega)).$$

(In fact, one can readily construct φ so that

$$|(\partial/\partial t)\nabla_x^j\varphi(t, x)|\leqq C\,\mathrm{d}(x, \partial\Omega)^{\alpha+1-j}, j=2, 3.)$$

If Δ^t denotes the pull-back by $\varphi(t, \cdot)$ of Δ, $\Delta^t A(t, \varphi(t, x))=0$ in Ω and

$$A(t, \varphi(t, x))=c_N|\varphi(t, x)|^{2-N}\, x\in\partial\Omega.$$

$\Delta^t = g^{ij}\,\partial^2/\partial x_i\,\partial x_j + g^j\,\partial/\partial x_j$, where

$$g^{ij} \in C^0(I;\, C^{k+1,\alpha}(\bar{\Omega})),\ (\partial/\partial t)\, g^{ij} \in C^0(I;\, C^{k,\alpha}(\bar{\Omega}))$$

$$g^j \in C^0(I;\, C^{k+\alpha}(\bar{\Omega})),\ (\partial/\partial t)\, g^j \in C^0(I;\, C^{k-1,\alpha}(\bar{\Omega}))$$

($B^{\alpha-1}(\Omega)$ replaces $C^{k-1,\alpha}(\bar{\Omega})$ when $k=0$.)

As $t' \to t$,

$$\Delta^t\left[\frac{A(t', \varphi(t', x)) - A(t, \varphi(t, x))}{t-t'}\right] = \left(\frac{\Delta^t - \Delta^{t'}}{t-t'}\right) A(t', \varphi(t', x))$$

tends to

$$((\partial g^{ij}/\partial t)\,\partial^2/\partial x_i\,\partial x_j + (\partial g^j/\partial t)\,\partial x_j)\, A(t, \varphi(t, x))$$

in $C^{k-1,\alpha}(\bar{\Omega})$ (or $B^{\alpha-1}(\Omega)$ if $k=0$). The boundary values $A(t', \varphi(t', x)) - A(t, \varphi(t, x))$ tend to $C_N(\partial/\partial t)\,|\varphi(t, x)|^{2-N}$ in $C^{k+1,\alpha}(\partial\Omega)$. Therefore by (3.8) (or (3.9)) $(\partial/\partial t)\, A(t, \varphi(t, x))$ exists in $C^{k+1,\alpha}(\bar{\Omega})$. Finally, because $(\partial\varphi/\partial t) \in C^0(I;\, C^{k+1,\alpha}(\bar{\Omega}))$ and $A(t, \varphi(t, x)) \in C^0(I;\, C^{k+2,\alpha}(\bar{\Omega}))$ it follows that as $t' \to t$, $(t'-t)^{-1}(A(t', z) - A(t, z))$ tends to $(\partial/\partial t)\, A(t, z)$ in $C^{k+1,\alpha}(\bar{\Omega}^t)$. (We can define $A(t', z)$ for $z \in \bar{\Omega}^t$ by taking any extension of from $z \in \bar{\Omega}^{t'}$ that preserves the smoothness class.)

Denote $\dot{G}(t, x) = (\partial/\partial t)\, G(t, x)$. $\nabla G(t, x)$ denotes the gradient in the x variable. Note that $\dot{G}(t, x) = \dot{A}(t, x)$ for all $x \in \bar{\Omega}^t \setminus \{0\}$, so that we can extend $\dot{G}(t, x)$ to be a harmonic function in Ω^t, including the origin. We will view h as a function on S^n, not $\partial\Omega^t$:

$$h(t, \theta) = \theta \cdot \nabla G(t, F(t, \theta)), \tag{3.10}$$

where $F(t, \cdot)$ is the inverse of the Gauss map on $\partial\Omega^t$. Differentiating the equation $G(t, F(t, \theta)) = 0$ with respect to t, we find

$$\dot{G}(0, F(0, \theta)) + \nabla G(0, F(0, \theta)) \cdot (\partial/\partial t)\, F(0, \theta) = 0. \tag{3.11}$$

It is easy to calculate that

$$(\partial/\partial t)\, F(0, \theta) = (\nabla_i v)\, e_i + v\theta. \tag{3.12}$$

Recall that $e_i \cdot \nabla G(0, \theta) = 0$, so (3.11) becomes

$$\dot{G}(0, F(0, \theta)) = -h(0, \theta)\, v(\theta). \tag{3.13}$$

(3.13) is the well-known Hadamard variational formula [G, Chap. 15].

Because $\dot{G}(0, x)$ is harmonic with boundary values $-hv$, we have

$$\theta \cdot \nabla \dot{G}(0, F(0, \theta)) = -\Lambda(hv). \tag{3.14}$$

Differentiating Eq. (3.10) with respect to t we find, with the help of Lemma B,

$$(\partial/\partial t)\, h(0, \theta) = \theta \cdot \nabla \dot{G}(0, F(0, \theta)) + \nabla^2 G(0, F(0, \theta))(\theta, (\partial/\partial t)\, F(0, \theta)).$$

Using (3.14), (3.12), (3.6), and (3.7), we have

$$(\partial/\partial t)\, h(0, \theta) = -\Lambda(hv) + K c_{ij} (\nabla_i h)(\nabla_j v) - K v h\, \mathrm{trace}(c_{ij}). \tag{3.15}$$

It is not hard to calculate [CY1]

$$(\partial/\partial t)\det(\nabla_{ij}(u+tv)+(u+tv)+(u+tv)\delta_{ij})|_{t=0}=c_{ij}(\nabla_{ij}v+v\delta_{ij}).$$

Consequently,

$$\begin{aligned}(\mathrm{d}/\mathrm{d}t)\mathscr{F}(u+tv)|_{t=0}&=((\partial/\partial t)h(0,\theta)/K+hc_{ij}(\nabla_{ij}v+v\delta_{ij})\\&=c_{ij}\nabla_i(h\nabla_j v)-\frac{1}{K}\Lambda(hv)\\&=L_u v.\end{aligned}$$

The fact that the operator bounded operator L_u: $C^{k+2,\alpha}(S^n)\to C^{k,\alpha}(S^n)$ varies continuously in operator norm as u varies in $C^{k+2,\alpha}(S^n)$ follows from another limiting argument along the lines of Lemma B, using (3.8), (3.9) and Proposition 1 in Sect. 2.

4 Local inversion

Every dilate of Ω gives the same density R; equivalently $\mathscr{F}(tu)=\mathscr{F}(u)$ for every constant $t>0$. Thus we can restrict $\mathscr{F}$ to $\mathscr{U}_1=\{u\in\mathscr{U}:\max_{S^n}u=1\}$. Our goal is to prove

Lemma C *$\mathscr{F}$ is a local homeomorphism from $\mathscr{U}_1$ to $\mathscr{R}$.*

We do this by proving

Proposition 3 *Let $k\geqq 0$, $0<\alpha<1$. Let Ω be a strongly convex $C^{k+2,\alpha}$ domain containing the origin. The operator L given in Proposition 2 by $Lv=(\mathrm{d}/\mathrm{d}t)\mathscr{F}(u+tv)|_{t=0}$ is a Fredholm operator L: $C^{k+2,\alpha}(S^n)\to C^{k,\alpha}(S^n)$ with null space $\ker(L)=\mathrm{span}\{u\}$ and image*

$$\mathrm{image}(L)=\{f\in C^{k,\alpha}(S^n):\int f\,\mathrm{d}\theta=0\}.$$

Once the proposition is proved, Lemma C follows immediately from the implicit function theorem.

Note that $Lu=(\mathrm{d}/\mathrm{d}t)\mathscr{F}((1+t)u)=0$, so $\mathrm{span}\{u\}\subset\ker L$. Also $\int\mathscr{F}(u+tv)\,\mathrm{d}\theta=1$ implies $\int Lv\,\mathrm{d}\theta=(\mathrm{d}/\mathrm{d}t)\int\mathscr{F}(u+tv)\,\mathrm{d}\theta|_{t=0}=0$. Therefore $\mathrm{image}(L)\subset\{f\in C^{k,\alpha}(S^n):\int f\,\mathrm{d}\theta=0\}$. Our task is to prove the opposite inclusions.

Motivated by the equations $Lu=0$ and $L^*1=0$ (where L^* is the formal adjoint of L), we consider the quadratic form $\int vL(uv)\,\mathrm{d}\theta$. Let β and V be the harmonic functions in Ω with boundary values $\beta|_{\partial\Omega}=uh$ and $V|_{\partial\Omega}=v$. (As usual, we identify functions on S^n with functions on $\partial\Omega$ via the Gauss map.) The crux of the proof is the formula

$$\begin{aligned}(4.1)\qquad\int_{S^n}vL(uv)\,\mathrm{d}\theta=&-\int_{S^n}hc_{ij}u(\nabla_i v)(\nabla_j v)\,\mathrm{d}\theta\\&-\int_\Omega\beta(x)|\nabla V(x)|^2\,\mathrm{d}x,\quad\textit{for all } v\in C^{2,\alpha}(S^n).\end{aligned}$$

It suffices to prove (4.1) for $u \in C^\infty(S^n)$. Both the left and right hand sides are continuous under passage to the limit in $C^{k+2,\alpha}(S^n)$.

By an integration by parts,

$$\int h c_{ij}(\nabla_i u)(\nabla_j v) v \, d\theta = -\int v^2 \nabla_j(h c_{ij} \nabla_i u) \, d\theta - \int h c_{ij}(\nabla_i u)(\nabla_j v) v \, d\theta.$$

Hence,

$$\int h c_{ij}(\nabla_i u)(\nabla_j v) v \, d\theta = -\tfrac{1}{2}\int v^2 \nabla_j(h c_{ij} \nabla_i u) \, d\theta = -\tfrac{1}{2}\int v^2 c_{ij} \nabla_j(h \nabla_i u) \, d\theta$$

$$= -\tfrac{1}{2}\int v^2 \frac{1}{K} \Lambda(hu) \, d\theta.$$

The second equality holds because $\nabla_j c_{ij} = 0$. (This is equivalent to self adjointness of the linearized Minkowski operator; see [CY1].) The third equality holds because $Lu = 0$.

Therefore,

$$\int v L(uv) \, d\theta = \int v c_{ij} \nabla_i(h \nabla_j(uv)) \, d\theta - \int \frac{v}{K} \Lambda(huv) \, d\theta$$

$$= -\int h c_{ij}(\nabla_i v)(\nabla_j(uv)) \, d\theta - \int \frac{v}{K} \Lambda(huv) \, d\theta$$

$$= -\int h c_{ij} v(\nabla_i v)(\nabla_j u) \, d\theta - \int h c_{ij} u(\nabla_i v)(\nabla_j v) \, d\theta - \int \frac{v}{K} \Lambda(huv) \, d\theta$$

$$= -\int h c_{ij} u(\nabla_i v)(\nabla_j v) \, d\theta + \int \left(\frac{1}{K} \Lambda(hu) \frac{v^2}{2} - \frac{v}{K} \Lambda(huv)\right) d\theta.$$

Consider the harmonic functions W and Z in Ω with boundary values $W|_{\partial\Omega} = v^2$ and $Z|_{\partial\Omega} = x \cdot \nabla(V^2)|_{\partial\Omega}$, with V as above. Recall that $(1/K)\, d\sigma = g_*(d\sigma)$ and that Green's formula implies Λ is selfadjoint on $L^2(\partial\Omega, d\sigma)$. Thus we can transplant our integral to $\partial\Omega$ via g:

$$\int_{S^n} \left(\frac{1}{K} \Lambda(hu) \frac{v^2}{2} - \frac{v}{K} \Lambda(huv)\right) d\theta$$

$$= \int_{\partial\Omega} \left(\Lambda(hu) \frac{v^2}{2} - v\Lambda(huv)\right) d\sigma = \int_{\partial\Omega} \left(hu\Lambda\left(\frac{v^2}{2}\right) - huv\Lambda(v)\right) d\sigma.$$

Next, for $x \in \partial\Omega$ and $\theta = g(x)$, observe that $x - u\theta$ is tangent to $\partial\Omega$. Since $W = V^2$ on $\partial\Omega$;

$$x \cdot \nabla W - x \cdot \nabla(V^2) = u\theta \cdot \nabla W + (x - u\theta) \cdot \nabla W - [u\theta \cdot \nabla(V^2) + (x - u\theta) \cdot \nabla(V^2)]$$

$$= u\theta \cdot \nabla W - u\theta \cdot \nabla(V^2) = u\Lambda(v^2) - 2uv\Lambda(v).$$

Also, $h\, d\sigma = d\omega$, so

$$\int_{\partial\Omega} (hu\Lambda(v^2/2) - huv\Lambda(v)) \, d\sigma = \tfrac{1}{2} \int_{\partial\Omega} (x \cdot \nabla W - x \cdot \nabla(V^2)) \, d\omega$$

$x\cdot\nabla W$ is a harmonic function in Ω whose value at $x=0$ is 0. Thus $\int_{\partial\Omega} x\cdot\nabla W\,d\omega = 0$. Moreover, $\Delta Z=0$ implies

$$\begin{aligned}\Delta(Z-x\cdot\nabla(V^2)) &= \Delta(-x\cdot\nabla(V^2))\\ &= -2(2|\nabla V|^2+x\cdot\nabla|\nabla V|^2).\end{aligned}$$

Therefore,

$$\begin{aligned}-\tfrac{1}{2}\int_{\partial\Omega} x\cdot\nabla(V^2)\,d\omega &= -\tfrac{1}{2}Z(0)=\int_\Omega G(x)(2|\nabla V|^2+x\cdot\nabla|\nabla V|^2)\,dx\\ &= -\int_\Omega [\operatorname{div}(xG(x))-2G(x)]\,|\nabla V|^2\,dx\\ &= -\int_{\partial\Omega}[x\cdot\nabla G(x)+(N-2)G(x)]\,|\nabla V|^2\,dx.\end{aligned}$$

We claim that $\beta(x)=x\cdot\nabla G(x)+(N-2)G(x)$. Recall that $G(x)=-c_N|x|^{2-N}+A(x)$. Since $(x\cdot\nabla+N-2)|x|^{2-N}=0$, we see that $x\cdot\nabla G(x)+(N-2)G(x)=(x\cdot\nabla+N-2)A(x)$ is harmonic in Ω including the origin. Finally,

$$(x\cdot\nabla G(x)+(N-2)G(x))|_{\partial\Omega}=x\cdot\nabla G(x)|_{\partial\Omega}=uh.$$

This proves (4.1).

If $L(uv)=0$, then $\int vL(uv)\,d\theta=0$. This implies v is constant because u and h are positive, so that β is positive. Therefore, $\ker L=\operatorname{span}\{u\}$.

In order to complete the proof of Proposition 3 we need to check that L is a Fredholm operator of index zero.

Denote $Tv=c_{ij}\nabla_j(h\nabla_j v)-v$. T: $C^{k+2,\alpha}(S^n)\to C^{k,\alpha}(S^n)$ is a bounded operator. Let $Tv=f$. If $v(p)=\max_{S^n} v$, then

$$\nabla_j v(p)=0 \quad\text{and}\quad c_{ij}\nabla_i(h\nabla_j v)(p)\leqq 0.$$

Therefore $f(p)=Tv(p)\leqq -v(p)$ or $\max v=v(p)\leqq -f(p)$. Similarly, $\min v\geqq -\max f$, so that $\max|v|\leqq\max|f|$. It now follows from the method of Theorem 6.8 [GT] that T has a bounded inverse B: $C^{k,\alpha}(S^n)\to C^{k+2,\alpha}(S^n)$. (Use the continuity method from the operator $v\mapsto\delta_{ij}\nabla_{ij}v-v$.) Next, $BL=I+E$, where I is the identity map on $C^{k+2,\alpha}(S^n)$ and $Ev=Bv-B\left(\frac{1}{K}A(hv)\right)$. If v_j is a bounded sequence in $C^{k+2,\alpha}$, then there is a convergent subsequence in $C^{k+1,\alpha}$. Since $h\in C^{k+1,\alpha}$, $1/K\in C^{k,\alpha}$, and A is bounded from $C^{k+1,\alpha}$ to $C^{k,\alpha}$, we see that Ev_j has a convergent subsequence in $C^{k+2,\alpha}$. Thus E is compact on $C^{k+2,\alpha}$.

The map T is invertible, so the equations

$$(I+E)w=\psi \quad\text{and}\quad Lw=T\psi$$

for w, $\psi\in C^{k+2,\alpha}(S^n)$ are equivalent. In particular, $\ker(I+E)=\ker(L)=\operatorname{span}\{u\}$, so that $\dim\ker(I+E)=1$. By the Fredholm-Riesz theorem, image$(I+E)$ has codimension 1 in $C^{k+2,\alpha}(S^n)$. It follows that image(L) has codimension 1 in $C^{k,\alpha}(S^n)$. But $\int Lw\,d\theta=0$ for all $w\in C^{k+2,\alpha}(S^n)$, so image$(L)=\{f\in C^{k,\alpha}(S^n)$: $\int_{S^n} f\,d\theta=0\}$. This concludes the proof of Proposition 3 and Lemma C.

5 The doubling property for harmonic measure

Recall that for a convex open subset $\Omega \subset \mathbb{R}^N$ with $B_r \subset \Omega \subset B_1$, r controls the Lipschitz constant of $\partial\Omega$. Thus by a well known result of Hunt and Wheeden [HW], for all $x \in \partial\Omega$, $s > 0$

$$\omega((\partial\Omega) \cap B_{2s}(x)) \leqq C\, \omega((\partial\Omega) \cap B_s(x)) \tag{5.1}$$

with a constant C depending only on N and r. The difficulty in the proof of the doubling property (1.5) for ω is that the sets F may not be comparable to balls. This difficulty is overcome by

Lemma D *Suppose $B_r \subset \Omega \subset B_1$. Let $x, y \in \Omega$, $x \notin B_{r/2}$. There is a constant C depending only on r and N such that for any t, $1/4 \leqq t \leqq 1$,*

$$|G(x)| \leqq C\, |G(t\,x + (1-t)\,y)|\,.$$

Proof. Harnack's inequality says that if W is a positive harmonic function in $B_{2s}(x)$, then

$$\max_{\bar{B}_s(x)} W \leqq 3 \min_{\bar{B}_s(x)} W. \tag{5.2}$$

Let U be the region above a Lipschitz graph:

$$U = \{(x_0, x') \in \mathbb{R} \times \mathbb{R}^n : x_0 > w(x')\} \quad \text{and} \quad |\nabla w(x')| \leqq M_0\,.$$

Let $x \in \partial U$ and $s > 0$. There exists a point $z \in U$ such that $|x - z| = s$ and $\operatorname{dist}(x, \partial U) > s/2M_0$. Dahlberg's boundary Harnack inequality [D, JK] says that if V_1 and V_2 are positive harmonic functions in $B_{2s}(x) \cap U$ which vanish on $B_{2s}(x) \cap \partial U$ and $V_1(z) = V_2(z)$, then

$$C^{-1}\, V_1(y) \leqq V_2(y) \leqq C\, V_1(y) \quad \textit{for all } y \in B_s(x) \cap \bar{U}, \tag{5.3}$$

where C depends only on M_0 and N. Let W be a positive harmonic function in $B_{2s}(x) \cap U$ such that $W(y) \geqq 1$ for all $y \in B_s(x) \cap \partial U$. Then [HW2, Lemma 2.1].

$$W(z) \geqq C^{-1} \tag{5.4}$$

with z as above and C a constant depending only on M_0 and N.

Now suppose $x, y \in \Omega$, $x \notin B_{r/2}$, and denote $|x - y| = 10\,s/r$. Note that $\Omega \subset B_1$ implies $s \leqq r/5$. Denote $p = t\,x + (1-t)\,y$ for some $t \in [1/4, 1]$. Let $\Phi(z) = t(z - y) + y$, that is, Φ is dilation by the factor t with origin at y. Note that $\Phi(x) = p$. If $\operatorname{dist}(x, \partial\Omega) > s$, then by convexity, $B_{ts}(p) = \Phi(B_s(x)) \subset \Omega$. Thus $\operatorname{dist}(p, \partial\Omega) > t\,s$. Since $|x - p| < 10\,s/r$, we can form a chain of overlapping balls centered on the segment from x to p and use Harnack's inequality 5.2 to show that $|G(x)| \leqq C\, |G(p)|$ for a constant C depending only on r.

If $\operatorname{dist}(x, \partial\Omega) < s$, then consider the region $T = \Gamma \cap \Omega \cap B_{8s}(x)$, where Γ is the open cone from the origin with central axis the ray from 0 to x and opening

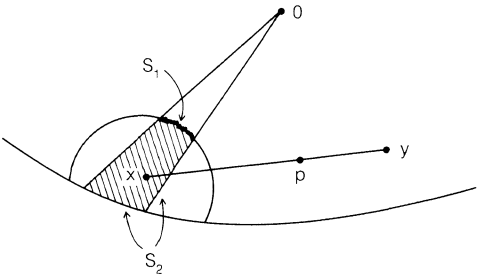


Fig. 1

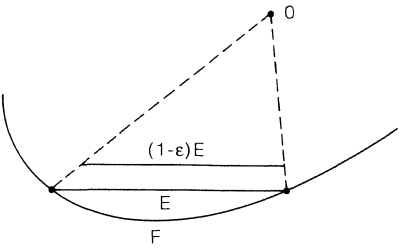


Fig. 2

angle sr. The boundary of T can be divided into a top S_1 and a union of sides and bottom S_2:

$$S_1 = \Gamma \cap \partial B_{8s}(x), \qquad S_2 = (\partial U) \setminus \bar{S}_1 .$$

Define V as the harmonic function in T with boundary values $V(z) = \max_{S_1} |G|$ for $z \in S_1$ and $V(z) = 0$ for $z \in S_2$. Choose z_1 on the ray from 0 to x so that $\operatorname{dist}(z_1, \partial\Omega) = 5s$. Then (5.2) (applied to a suitable chain of balls of radius comparable to s) and (5.4) imply $|G(z)| \leqq CV(z_1)$. Therefore, using (5.3) and (5.2) we have $|G(x)| \leqq CV(x)$ for a new constant C depending only on r and N. Now define $V_1(z) = V(\Phi^{-1}(z))$ for all $z \in \Phi(\bar{T})$. By convexity, $\Phi(T) \subset \Omega$. A suitable chain of balls of radius comparable to s links any point of S_1 to any point of $\Phi(S_1)$. Therefore, (5.2) implies $V_1(z) \leqq C|G(z)|$ for all $z \in \Phi(S_2)$, so by the maximum principle, $V_1(z) \leqq C|G(z)|$ for all $z \in \Phi(T)$. Hence $C^{-1}|G(x)| \leqq V(x) = V_1(p) \leqq C|G(p)|$. This proves Lemma D.

Proof of Theorem 2. Let $F = \{x \in \partial\Omega : a \cdot x \geqq b\}$, $b \geqq 0$. Let $E = \{x \in \bar{\Omega} : a \cdot x = b\}$. Let ε be the distance from E to the tangent plane to $\partial\Omega$ of the form $\{x : a \cdot x = b'\}$ with $b' > b$. Since $b \geqq 0$, the half-space $a \cdot x > b$ does not contain the origin. If $\varepsilon \geqq r/10$, then it is easy to see that there is a point $x \in \partial\Omega$ such that $B_{r/100}(x) \cap \partial\Omega \subset \frac{1}{2}F$. It follows from (5.1) that $\omega(F) \leqq \omega(\partial\Omega) \leqq C\,\omega(\frac{1}{2}F)$.

Now suppose that $\varepsilon < r/10$. For $x \in E$, $\operatorname{dist}((1-\varepsilon)x, \partial\Omega) \geqq \varepsilon r/2$. On the other hand, every $X \in F$ is within 2ε of $(1-\varepsilon)E$ (Fig. 2). Choose a collection of points $z_k \in (1-\varepsilon)E$, $k = 1, \ldots, N_0$ such that $|z_k - z_\ell| \geqq \varepsilon$ for $k \neq \ell$ and $\bigcup_k B_{c_0\varepsilon}(z_k) \supset F$. Because $\sum_k \sigma(B_{c_0\varepsilon}(z_k) \cap \partial\Omega)$ is comparable to $\sigma(F)$, it follows that N_0 is comparable to $\sigma(E)/\varepsilon^{N-1}$. As we saw earlier, we can choose similar ellipsoids E_1 and E_2 of comparable size such that $E_1 \subset E \subset E_2$. The axes of E_1 all exceed $\varepsilon r/C_1$ for some dimensional constant C_1. Therefore we can choose points $p_k \in \frac{1}{2}E_1$, $k = 1, \ldots, N_1$ such that $|p_k - p_\ell| \geqq \varepsilon r/C_1 C_2$ for $k \neq \ell$, $d(p_k, \partial(\frac{1}{2}E_1)) \geqq \varepsilon r/C_1 C_2$ and $\bigcup_k B_{\varepsilon r/C_1}(p_k) \supset \frac{1}{4}E_1$. It follows that N_1 is comparable to $\sigma(E_1)/\varepsilon^{N-1}$, which is comparable to N_0.

Dahlberg's comparison theorem for Lipschitz domains [D] applied to Ω says that for any s, $0 < s < 1$, and any $x \in \partial\Omega$, and any $z \in \Omega \setminus B_{r/2}$, such that $s/M_1 \leqq |x - z| \leqq M_1 s$, and $\operatorname{dist}(z, \partial\Omega) \geqq s/M_1$,

$$|G(z)|\, s^{N-2}/C \leqq \omega(B_s(x) \cap \partial\Omega) \leqq C\, |G(z)|\, s^{N-2}, \tag{5.5}$$

where C depends only on r, M_0, M_1, and N. Lemma D implies

$$|G(z)| \leqq C\, |G(p)| \tag{5.6}$$

for any $z \in (1-\varepsilon)E$ and any $p \in (1-\varepsilon)(\frac{1}{2}E_1)$. In particular,

$$\max_{k=1}^{N_0} |G(z_k)| \leqq C \min_{k=1}^{N_1} |G((1-\varepsilon)p_k)|.$$

Let $x_k \in \frac{1}{2}F$ be such that $\pi x_k = p_k$. Then $F_k = B_{\varepsilon r/2C_1C_2}(x_k) \cap \partial\Omega$ is a pairwise disjoint family and $F_k \subset \frac{1}{2}F$. By (5.5), $\omega(F_k) \geqq |G((1-\varepsilon)p_k)|\, \varepsilon^{N-2}/C$. Thus

$$\omega(\tfrac{1}{2}F) \geqq N_1\, \varepsilon^{N-2} \min |G((1-\varepsilon)p_k)|.$$

On the other hand, by (5.5) and (5.1),

$$\omega(B_{c_0\varepsilon}(z_k) \cap \partial\Omega) \leqq C\, |G(z_k)|\, \varepsilon^{N-2},$$

so that $\omega(F) \leqq C N_0\, \varepsilon^{N-2} \max_k |G(z_k)|$. Since N_0 is comparable to N_1, we have the inequality we want: $\omega(F) \leqq C\, \omega(\frac{1}{2}F)$ for a constant C depending only on N and r.

6 Existence and uniqueness

We can now prove Theorem 1 by the method of continuity. Fix $R \in C^{k,\alpha}(S^n)$, $R > 0$, $\int R\, d\theta = 1$. Let $R_0 = 1/\operatorname{vol}(S^n)$. Evidently $\mathscr{F}(1) = R_0$, corresponding to the case in which Ω is the unit ball. Let $S = \{t \in [0,1]$: There exists $u \in \mathscr{U}_1$ such that $\mathscr{F}(u) = (1-t)R_0 + tR\}$. S is nonempty since $0 \in S$. Also S is open because $\mathscr{F}$ is locally invertible (Lemma C). We need to show that S is closed.

Lemma E *Let $u\in\mathscr{U}_1$, $\mathscr{F}(u)=R$. Suppose that $R\geqq c_0>0$. There is a constant C depending only on c_0 and $\|R\|_{C^{k,\alpha}(S^n)}$ such that*

$$u\geqq 1/C, \tag{6.1}$$

$$\|u\|_{C^{k+2,\alpha}(S^n)}\leqq C, \tag{6.2}$$

and

$$\nabla_{ij}u+u\,\delta_{ij}\geqslant\delta_{ij}/C. \tag{6.3}$$

Proof. The normalization $\max u=1$ is the same as $\Omega\subset B_1$ and $(\partial\Omega)\cap\partial B_1\neq\emptyset$. Let r be the largest radius so that $B_r\subset\Omega$, then $\min u=r$. We will show that r is bounded below by a positive constant, depending only on c_0.

Choose $x_0\in(\partial\Omega)\cap\partial B_r$. Let H be the smallest hemisphere containing Ω with center x_0:

$$H=\{x:|x-x_0|<s \text{ and } (x-x_0)\cdot\theta_0<0\},$$

where $\theta_0=g(x_0)$. Since $\Omega\subset B_1$, $s\leqq 1+r\leqq 2$. But since $(\partial\Omega)\cap(\partial B_1)\neq\emptyset$, $s\geqq\sqrt{1-r^2}$. We may as well assume $r<1/4$, so $1/\sqrt{2}\leqq s\leqq 2$.

Choose $x\in(\partial\Omega)\cap B_s(x_0)$. Denote by G_1 Green's function for H with pole at 0. By the maximum principle $|G_1|\geqq|G|$ on Ω. Hence $\theta\cdot\nabla G_1(x)\geqq\theta\cdot\nabla G(x)=h(\theta)$ with $\theta=g(x)$. Furthermore, because x is a point of tangency of $\partial\Omega$ and $\partial B_s(x_0)$, the Gauss curvature of $\partial\Omega$ at x, $K(x)\geqq s^{1-N}\geqq 2^{1-N}$. But then $c_0\leqq R(\theta)\leqq h(x)/K(x)\leqq 2^{N-1}\,\theta\cdot\nabla G_1(x)$. It is easy to compute Green's function for the hemisphere explicitly by reflection and easy to check that as $r\to 0$, $\theta\cdot\nabla G_1(x)$ tends to 0 uniformly in s, $1/\sqrt{2}\leqq s\leqq 2$ and uniformly in $x\in(\partial H)\cap\partial B_s(x_0)$. This yields the lower bound for r in terms of c_0.

Now since $K=h/R$ and R is bounded above and below by positive constants Theorem 2 implies that $K\,d\sigma$ satisfies the doubling property (1.5). It follows from Caffarelli's Theorem 4 that we have a bound on the $C^{1,\alpha'}$ norm of $\partial\Omega$ for some $\alpha'>0$ depending only on c_0 and upper and lower bounds for R. It follows from this that h has a positive lower bound and we have control of $\|h\|_{C^{\alpha'}(\partial\Omega)}$. From now on the regularity of h is no different from that of the term $(1+|\nabla w(x)|^2)^{\frac{N+1}{2}}$ appearing in (1.4). In particular, Theorem 5 implies control of the $C^{2,\alpha'}$ norm of $\partial\Omega$ and that $\partial\Omega$ is strongly convex. Thus (6.3) holds. Finally, using Proposition 1 and (3.2) we see that $\|h\|_{C^{k+1,\alpha}(\partial\Omega)}$ is controlled by $\|u\|_{C^{k+2,\alpha}(S^n)}$. Thus since $K=h/R$ we can obtain the desired bound (6.2) by a standard bootstrap argument using Theorem 5.

Now consider a sequence $u_j\in\mathscr{U}_1$, such that $\mathscr{F}(u_j)=(1-t_j)R_0+t_jR$ and $t_j\to t$. By Lemma E, the sequence u_j is bounded in $C^{k+2,\alpha}(S^n)$ and hence has a convergent subsequence in $C^{k+2,\alpha'}$ $\alpha'<\alpha$. The limit u satisfies $\mathscr{F}(u)=(1-t)R_0+tR$ and $u\in C^{k+2,\alpha}(S^n)$ by Theorem 5 and Proposition 1. We conclude that $S=[0,1]$. In particular, $1\in S$, so $\mathscr{F}$ is surjective, and we have proved existence.

By a slight variant of the argument just given, using Lemma C, there is a unique continuous mapping $\gamma\colon[0,1]\to\mathscr{U}_1$ such that $\mathscr{F}(\gamma(t))=(1-t)R_0+tR$ and $\gamma(0)=u_0\equiv 1$. Define $\Psi(R)=\gamma(1)$, then $\mathscr{F}\circ\Psi\colon\mathscr{R}\to\mathscr{R}$ is the identity map and it is easy to check that Ψ is continuous on $\mathscr{R}$. One can also check that the image of Ψ in $\mathscr{U}_1$ is open and closed. Since $\mathscr{U}_1$ is connected, $\mathscr{F}$ is invertible with inverse Ψ. This proves uniqueness.

It remains to deduce the corollary to Theorem 1. The procedure is similar to [CY1, Sect. 5]. For an arbitrary open, convex set Ω in $\mathbb{R}^N$ we give a more precise definition of the Gauss map g on $\partial\Omega$ as a set-valued function:

$$g(x)=\{\theta\in S^n: (x'-x)\cdot\theta<0 \text{ for every } x'\in\Omega\}.$$

In other words, $g(x)$ is the set of outer normals θ of the support planes of Ω at x. The map g is single-valued except on a set of surface measure zero (and hence harmonic measure zero). For $E\subset S^n$, define

$$g^{-1}(E)=\{x\in\partial\Omega: g(x)\cap E\neq\emptyset\}.$$

Suppose that $0\in\Omega$. Denote harmonic measure for Ω at 0 by ω. Then ω defines a Borel measure μ on S^n by

$$\mu(E)=\omega(g^{-1}(E)).$$

Let $\tilde{\mu}$ be an arbitrary probability measure on S^n. Choose $R_j\in C^\infty(S^n)$, $R_j>0$ such that $d\mu_j=R_j\,d\theta$ tends weakly to $\tilde{\mu}$. Theorem 1 implies that there exists a C^∞ strongly convex domain Ω_j containing the origin so that if g_j denotes the Gauss map for Ω_j and ω_j denotes harmonic measure at 0 for Ω_j, $g_{j*}(d\omega_j)=d\mu_j$. Since every dilate of Ω_j has the same property, we may assume after dilation that $B_1\subset\Omega_j$ and $(\partial B_1)\cap\partial\Omega_j\neq\emptyset$.

The Minkowski distance between compact subsets K_1 and K_2 of $\mathbb{R}^N$ is defined as

$$d(K_1, K_2)=\max_{x\in K_1} d(x, K_2)+\max_{x\in K_2} d(x, K_1).$$

It is well known that the set of compact subsets of a fixed compact set is compact in the metric d. (Blaschke selection theorem.) Therefore by a diagonal procedure we can replace Ω_j by a subsequence and find a convex open set Ω such that

$$d(\bar{\Omega}_j\cap\bar{B}_M, \bar{\Omega}\cap\bar{B}_M)\leqq 1/j \quad \text{for } j\geqq M.$$

Evidently, $B_1\subset\Omega$ and $\partial B_1\cap\partial\Omega\neq\emptyset$.

Our goal is to show that $\mu=\tilde{\mu}$. We begin by showing that we can assume $\Omega_j\subset\Omega$.

If u_j^M is the support function for $\Omega_j\cap B_M$ and u^M is the support function for $\Omega\cap B_M$, then $1\leqq u_j^M\leqq u^M+1/j$. Therefore,

$$(1-1/j)\,u_j^M\leqq u^M.$$

Thus we can replace Ω_j by the dilate $(1-1/j)\,\Omega_j$ and retain all the same convergence properties with the additional property $\Omega_j\subset\Omega$.

Lemma F *Suppose that Ω is a convex, open set in $\mathbb{R}^N$ and $d(0,\partial\Omega)\leqq 1$. Then there is a constant $C=C_N$ such that*

$$\omega((\partial\Omega)\setminus B_M)\leqq C\,M^{-1}.$$

Proof. Let $x_0\in\partial\Omega$ with $|x_0|=d(0,\partial\Omega)\leqq 1$. Let H be the hemisphere

$$H=\{x\in\mathbb{R}^N: |x-x_0|<M-2, (x-x_0)\cdot\theta<0\}.$$

(We can assume $M \geqq 3$.) Let W be the harmonic function in H with boundary values

$$W(x)=1 \quad \text{for } |x-x_0|=M-2,\ (x-x_0)\cdot\theta<0,$$

and

$$W(x)=0 \quad \text{for } |x-x_0|<M-2,\ (x-x_0)\cdot\theta=0.$$

The maximum principle implies $\omega((\partial\Omega)\setminus B_M)\leqq W(0)$. A routine calculation with the Poisson kernel for the hemisphere shows that $W(0)\leqq C M^{-1}$.

Remark. Suppose that $g_j(x_j)=\theta_j$, θ_j tends to θ and x_j tends to x. Then $x\in\partial\Omega$ and $\theta\in g(x)$.

Proof. The fact that $x\in\bar{\Omega}$ is obvious. Suppose $x\in\Omega$. Choose $s>0$ such that $B_s(x)\subset\Omega$ and choose j sufficiently large that $|x-x_j|<s/2$. Then $\Omega_j\subset\{x'\colon (x'-x_j)\cdot\theta_j<0\}$, so that $\mathrm{d}(x_j+(s/2)\theta_j, \Omega_j)\geqq s/2$. However, $x_j+(s/2)\theta_j\in B_s(x)\subset\Omega$. Thus for all large j, $\mathrm{d}(\Omega_j\cap B_M, \Omega\cap B_M)\geqq s/2$ if $x\in B_M$, a contradiction. Hence $x\in\partial\Omega$.

Next suppose $\theta\notin g(x)$. Then there exists $x'\in\Omega$ such that $(x'-x)\cdot\theta=s>0$. For large j, $(x'-x_j)\cdot\theta_j\geqq s/2$, so that $\mathrm{dist}(x', \Omega_j)\geqq s/2$. Again this is a contradiction, so $\theta\in g(x)$.

Let E be a closed subset of S^n. It follows immediately from the remark that for any $\varepsilon>0$ and any M there exists j_0 such that for any $j\geqq j_0$, $x\in(\partial\Omega_j)\cap B_M$, if $g_j(x)\in E$, then $\mathrm{d}(x, g^{-1}(E))<\varepsilon$. It is easy to check that $g^{-1}(E)$ is closed. Let U be an open subset of $\partial\Omega$, $U\supset g^{-1}(E)$. Let W be the bounded harmonic function in Ω with boundary values $W=1$ on U, $W=0$ on $(\partial\Omega)\setminus\bar{U}$. Since W is continuous on U, it follows that for any $\varepsilon>0$ and any M there exists j_0 such that for $j\geqq j_0$, $W(x)\geqq 1-\varepsilon$ for all $x\in g_j^{-1}(E)\cap B_M$. Let ω_j^M denote harmonic measure at 0 for $\Omega_j\cap B_M$. The maximum principle in $\Omega_j\cap B_M$ and Ω and Lemma F for $\Omega_j\cap B_M$ imply

$$\begin{aligned}\omega_j(g_j^{-1}(E)\cap B_M) &\leqq \omega_j^M((g_j^{-1}(E)\cap B_M)\cup(\Omega_j\cap\partial B_M))\\ &\leqq \omega(U)/(1-\varepsilon)+\omega_j^M(\Omega_j\cap\partial B_M)\leqq\omega(U)/(1-\varepsilon)+C M^{-1}.\end{aligned}$$

On the other hand Lemma F for Ω_j implies

$$\mu_j(E)=\omega_j(g_j^{-1}(E))\leqq\omega_j(g_j^{-1}(E)\cap B_M)+C M^{-1}.$$

Therefore,

$$\limsup_{j\to\infty}\mu_j(E)\leqq\omega(U)/(1-\varepsilon)+2C M^{-1}.$$

Since $U\supset g^{-1}(E)$, $\varepsilon>0$, and M are arbitrary,

$$\limsup_{j\to\infty}\mu_j(E)\leqq\omega(g^{-1}(E))=\mu(E), \quad \textit{for any closed } E\subset S^n.$$

Next, if V is any open subset of S^n, $\mu_j(V)=1-\mu_j(S^n\setminus V)$ and $\mu(V)=1-\mu(S^n\setminus V)$ imply

$$\liminf_{j\to\infty}\mu_j(V)=\mu(V) \quad \textit{for any open } V\subset S^n. \tag{6.4}$$

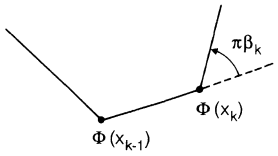


Fig. 3

For any closed $E \subset S^n$, $\varepsilon > 0$, choose V_1, V_2 open sets so that $E \subset V_1$, $\bar{V}_1 \subset V_2$ and $\tilde{\mu}(V_2) \leqq \tilde{\mu}(E) + \varepsilon$. Let $\varphi \in C^0(S^n)$ be such that $0 \leqq \varphi \leqq 1$, $\varphi = 1$ on $\bar{V}_1$ and $\varphi = 0$ on $S^n \setminus V_2$. By (6.4),

$$\tilde{\mu}(E) + \varepsilon \geqq \tilde{\mu}(V_2) \geqq \int \varphi \, d\tilde{\mu} = \lim_{j \to \infty} \int \varphi \, d\mu_j \geqq \liminf_{j \to \infty} \mu_j(V_1) = \mu(V_1) \geqq \mu(E).$$

Therefore, $\mu(E) \leqq \tilde{\mu}(E)$ for any closed $E \subset S^n$ and since $\mu(S^n) = \tilde{\mu}(S^n) = 1$, $\tilde{\mu}(V) \leqq \mu(V)$ for any open set $V \subset S^n$. Thus

$$\tilde{\mu}(E) = \inf_{\substack{V \text{ open} \\ V \supset E}} \tilde{\mu}(V) \leqq \inf_{\substack{V \text{ open} \\ V \supset E}} \mu(V) = \mu(E)$$

and our proof is complete.

7 Further remarks

(a) The Schwarz-Christoffel formula

Let $x_1 < x_2 < \ldots < x_m$, $r > 0$, $\psi \in \mathbb{R}$, and $0 < \beta_k < 2$. The Schwarz-Christoffel formula

$$\Phi'(z) = r e^{i\psi} \prod_{k=1}^{m} (z - x_k)^{-\beta_k}, \qquad \Phi(i) = 0, \tag{7.1}$$

defines an analytic function Φ on the upper half-plane $H = \{z \in \mathbb{C}: \operatorname{Im} z > 0\}$. (For $z \in H$, $0 < \arg(z - x_k) < \pi$.) Φ extends to a continuous function on the real axis. Let $x_0 = -\infty$, $x_{m+1} = +\infty$. $\Phi([x_k, x_{k+1}])$ is the line segment from $\Phi(x_k)$ to $\Phi(x_{k+1})$, and the segments that meet at $\Phi(x_k)$ form an exterior angle $\pi \beta_k$ (Fig. 3). In general it is hard to decide for which choices of $x_1, \ldots, x_m, \beta_1, \ldots, \beta_m$, $\Phi(\mathbb{R} \cup \{\infty\})$ is a simple closed curve, so that Φ is the conformal mapping of H to a polygon. However, in the convex case one can choose $x_1, \ldots, x_m$ arbitrarily. As we shall see presently, this shows that one can solve the prescribed harmonic measure problem in $\mathbb{R}^2$ for any finite sum of point masses. The general case then follows from a continuous version of (7.1).

Proposition 4 *Let $0 < \beta_k < 1$, $k = 1, \ldots, m$. Suppose that $\sum_{k=1}^{m} \beta_k < 2$. Then for any real numbers $x_1 < x_2 < \ldots < x_m$, (7.1) defines a conformal mapping from H to the interior of a convex polygon with vertices $\Phi(x_1), \ldots, \Phi(x_m)$, $\Phi(\infty)$ and exterior*

angles $\pi\beta_1, \ldots, \pi\beta_m$, $\pi\beta_\infty = \pi\left(2 - \sum_{k=1}^{m} \beta_k\right)$. *If* $\sum_{k=1}^{m} \beta_k \leqq 1$, *then the polygon is unbounded, i.e.,* $\Phi(\infty) = \infty$. *In that case the exterior angle* $\pi\beta_\infty$ *at* ∞ *is measured on the Riemann sphere.*

Proof. We need only show that Φ restricted to $\mathbb{R} \cup \{\infty\}$ is a simple closed curve in $\mathbb{C} \cup \{\infty\}$. The conclusion then follows from the argument principle. Let V be the unbounded connected component of $\mathbb{C} \setminus \Phi(\mathbb{R})$. (There will be two such components if $\Phi(\infty) = \infty$. In that case, choose the component to the right of the $\Phi(\mathbb{R})$ as x increases to ∞.) The boundary ∂V is a polygon with vertices of three types A, B, C. A vertex of type A is a vertex of ∂V equal to one of $\Phi(x_k)$ $k = 0, 1, \ldots, m$ for which $\Phi(\mathbb{R} \cup \{\infty\})$ is simple near the vertex. A vertex of type B is one at which $\Phi(\mathbb{R} \cup \{\infty\})$ is not simple and is contained in the interior of some segment $\Phi([x_k, x_{k+1}])$ $k = 0, \ldots, m$. Vertices of type C are all remaining vertices of ∂V. Each vertex of type C is equal to $\Phi(x_k)$ for at least two values of k, $k = 0, 1, \ldots, m$. ∂V is simple, so the sum of its exterior angles is 2π. But exterior angles at type B vertices are all nonpositive and exterior angles at vertices of type C are less than or equal to the minimum of the corresponding $\pi\beta_k$. It follows from $\sum_{k=0}^{m} \beta_k = 2$ and $\beta_k > 0$ that there can be no vertices of type C and that each $\Phi(x_k)$, $k = 0, \ldots, m$ arises as a vertex of type A. It then follows that the exterior angles of type B are all strictly negative, so there can be no vertices of type B and $\Phi(\mathbb{R} \cup \{\infty\})$ is simple.

Let $\alpha_k \in [0, 2\pi)$, $\theta_k = (\cos\alpha_k, \sin\alpha_k)$, $k = 1, \ldots, m+1$. Let μ be the measure on S^1, $\mu = \sum_{k=1}^{m+1} c_k = \delta_{\theta_k}$ with $c_k > 0$ and $\sum_{k=1}^{m+1} c_k = 1$. Choose $-\infty = x_0 < x_1 < \ldots < x_m < x_{m+1} = +\infty$ so that c_k is the harmonic measure for H at i of the interval (x_{k-1}, x_k), $k = 1, \ldots, m+1$. In other words, choose $x_1 < \ldots < x_m$ so that

$$c_k = \frac{1}{\pi} \int_{x_{k-1}}^{x_k} \frac{dt}{1+t^2}, \qquad k = 1, \ldots, m+1.$$

Next, choose ψ and β_k, $k = 1, \ldots, m$ so that

$$\alpha_k = \psi - \pi/2 - \sum_{j=k}^{m} \pi\beta_j.$$

We claim that the convex polygon $\Phi(H)$ solves the prescribed harmonic measure problem for μ. Indeed, Φ preserves harmonic measure, so that if ω denotes harmonic measure for $\Phi(H)$ at 0, $\omega(\Phi((x_{k-1}, x_k))) = c_k$, $k = 1, \ldots, m+1$. (7.1) implies that

$$\arg \Phi'(x) = \psi - \sum_{j=k}^{m} \pi\beta_j \quad \text{for } x \in (x_{k-1}, x_k) \tag{7.2}$$

$k = 1, \ldots, m+1$. So the outer normal to the side $\Phi((x_{k-1}, x_k))$ is θ_k.

In general, for an arbitrary probability measure μ, define the monotone increasing function $U: \mathbb{R} \to [0, 2\pi)$ by

$$U(x) = \min \left\{ \alpha : \mu([0, \alpha]) \geqq \frac{1}{\pi} \int_{-\infty}^{x} \frac{dt}{1+t^2} \right\}.$$

For any $E \subset [0, 2\pi)$,

$$\mu(\{(\cos\alpha, \sin\alpha) : \alpha \in E\}) = \frac{1}{\pi} \int_{U^{-1}(E)} \frac{dt}{1+t^2}.$$

The polygonal case corresponds to a step function U and we see from (7.2) that the equation we need to solve is

$$\arg \Phi'(x) - \pi/2 = U(x). \tag{7.3}$$

Indeed, define Φ by $\Phi(z) = 0$ and

$$\Phi'(z) = \exp(-V(z) + i(U(z) + \pi/2)),$$

where $U(z)$ is the Poisson integral of U in H and $V(z)$ is the harmonic conjugate of U. Then Φ' is an analytic function and a continuous version of Proposition 4 shows that Φ is a conformal mapping onto a convex region $\Omega = \Phi(H)$. It is easy to check that $g_*(d\omega) = d\mu$.

For uniqueness observe that $g_*(d\omega) = d\mu$ is equivalent to (7.3). V is uniquely determined up to an additive constant, so Ω is unique up to dilation. Finally, suppose $d\mu = R\, d\theta$ with $R \in C^{k,\alpha}(S^1)$, $R > 0$, then U belongs to $C^{k+1,\alpha}_{\text{loc}}(\mathbb{R}) \cap L^\infty(\mathbb{R})$ and by standard boundedness properties of the Poisson integral and Hilbert transform, $\Phi' \in C^{k+1,\alpha}_{\text{loc}}(\bar{H})$. Hence $\Phi' \in C^{k+1,\alpha}_{\text{loc}}(\bar{H})$ and $\partial\Omega$ is $C^{k+2,\alpha}$ except possibly near $\Phi(\infty)$. On the other hand, it is easy to show that $\Phi(\infty) \neq \infty$, and we can use a linear fractional transformation to trade ∞ in H with a finite boundary point and obtain $C^{k+2,\alpha}$ regularity for $\partial\Omega$ near $\Phi(\infty)$ as well.

(b) Concavity of $\int uR\, d\theta$ fails

The well-known Brunn-Minkowski formula [BF] says that if Ω_0 and Ω_1 are convex open sets in $\mathbb{R}^N$ and $\Omega_t = (1-t)\Omega_0 + t\Omega_1$, then $\text{vol}(\Omega_t)^{1/N}$ is a concave function of t, $0 \leqq t \leqq 1$. Moreover it is linear if and only if Ω_1 is a translate and dilate of Ω_0. Both existence and uniqueness proofs in Minkowski's problem can be based on this inequality.

Recall that

$$\text{vol}(\Omega) = \int_{\partial\Omega} u\, d\sigma. \tag{7.4}$$

Given the formal similarities between our problem and Minkowski's, we asked in [J1] whether

$$F(\Omega) = \int_{\partial\Omega} u\, d\omega$$

has the property $F(\Omega_t)$ is concave in t. (The power 1 rather than $1/N$ is dictated by homogeneity under dilation.) We will show that $F(\Omega_t)$ need not be concave. A numerical proof was first provided by L.N. Trefethen. In part (c) we will speculate about what we hope is a better candidate for a Brunn-Minkowski type inequality.

Let Ω be a fixed convex domain. Let ω^P denote harmonic measure for $\partial\Omega$ at P. If $\Omega_0=\Omega+P_0$ and $\Omega_1=\Omega+P_1$, $P_t=(1-t)P_0+tP_1$, then $F(\Omega_t)=\int_{\partial\Omega}\theta\cdot(x-P_t)\,d\omega^{P_t}(x)$. We will calculate $F(\Omega_0)$ when Ω is an isosceles triangle and P_t is on the central axis. For ζ in the upper half-plane H, we assume $0<\arg(1+\zeta)<\pi$ and $-\pi<\arg(1-\zeta)<0$. Define $\Phi(z)=\int_0^z(1-\zeta)^{-\beta}(1+\zeta)^{-\beta}\,d\zeta$. Assume $1/2<\beta<1$. Denote $a=\Phi(1)=\int_0^1(1-x^2)^{-\beta}\,dx$, $b=\int_0^\infty(1+x^2)^{-\beta}\,dx$. Then $\Omega=\Phi(H)$ is the isosceles triangle with vertices $\pm a=\Phi(\pm1)$, and $ib=\Phi(\infty)$, $\Phi(0)=0$ and $\Phi(iy)$, $0<y<\infty$ travels up the central axis. Let $\Phi(iy)=ih$. The harmonic measure of the base of the triangle at ih is $p(y)=\frac{1}{\pi}\int_{-1}^{1}\frac{y}{x^2+y^2}\,dy$. Therefore the harmonic measure of each of the equal sides is $\frac{1}{2}(1-p(y))$. For $P=ih$, denote

$$\begin{aligned}F(h)&=\int_{\partial\Omega}\theta\cdot(x-P)\,d\omega^P(x)\\&=g\,p(y)+(b-h)\frac{a}{\sqrt{a^2+b^2}}(1-p(y)).\end{aligned}$$

We are going to show that for h near b, $F''(h)>0$, which contradicts concavity ($F''(h)\leqq 0$). In fact,

$$F'(h)=p+h\frac{dp}{dh}-\frac{a}{\sqrt{a^2+b^2}}(1-p)-(b-h)\frac{a}{\sqrt{a^2+b^2}}\frac{dp}{dh}.$$

$$F''(h)=2\left(1+\frac{a}{\sqrt{a^2+b^2}}\right)\frac{dp}{dh}+\left(h-(b-h)\frac{a}{\sqrt{a^2+b^2}}\right)\frac{d^2p}{dh^2}.$$

$$dp/dh=(1+y^2)^\beta p'(y)\simeq-(2/\pi)\,y^{2\beta-2}\quad\text{as}\quad y\to\infty.$$

$$\begin{aligned}d^2p/dh^2&=2y\,\beta(1+y^2)^{\beta-1}p'(y)\,dy/dh+(1+y^2)^\beta p''(y)\,dy/dh\\&\simeq\frac{4}{\pi}(1-\beta)\,y^{4\beta-3}\quad\text{as}\quad y\to\infty.\end{aligned}$$

As $y\to\infty$, $h\to b$, so $F''(h)\simeq\frac{4b}{\pi}(1-\beta)\,y^{4\beta-3}$, which is positive.

(c) An exterior problem

One of the differences between our interior harmonic measure problem and the Minkowski problem is that in the Minkowski problem there is uniqueness up to translation and $1/K$ satisfies the N compatibility conditions

$$\int_{S^n}\theta^k(1/K(\theta))\,d\theta=0\qquad k=0,1,\ldots,n \tag{7.5}$$

where θ^k is the restriction to S^n of the k-th coordinate function ($n=N-1$). By contrast, we had only one compatibility condition

$$\int R\,d\theta=1 \tag{7.6}$$

and uniqueness up to a one-dimensional family of transformations (dilations). This difference was also reflected in the fact that our linearized operator was not selfadjoint, ($\ker(L)=\text{span}\{u\}$, $\ker(L^*)=\text{span}\{1\}$) whereas the linearized Minkowski operator is selfadjoint. (See [CY1].)

Let us now consider an exterior problem. Let Ω be a bounded convex region in $\mathbb{R}^N$. Define Green's function for the complement of Ω at ∞ as the unique harmonic function G in $\mathbb{R}^N\setminus\bar{\Omega}$ such that

$$G(x)=0 \quad \text{on } \partial\Omega$$

$$\lim_{x\to\infty} G(x)=1 \quad \text{when} \quad N\geqq 3$$

$G(x)-\log|x|$ is bounded as $x\to\infty$ when $N=2$. Define $h(x)=\theta\cdot\nabla G(x)$, where θ is the outer unit normal to Ω at x.

In the exterior problem we could try to prescribe $h\,d\sigma$ as a function of θ. Because the pole of Green's function is fixed at ∞, the solution is at best unique up to translation. However, it is easy to see using the Schwarz-Christoffel formula that the compatibility conditions analogous to (7.6) are complicated and nonlinear. The right problem appears to be to try to prescribe $h^2\,d\sigma$:

Given $S\in C^\infty(S^n)$, find a convex Ω so that $g_*(h^2\,d\sigma)=S\,d\theta$ with $h=\theta\cdot\nabla G$ and G is Green's function for $\mathbb{R}^N\setminus\Omega$ with pole at ∞.

Here are some of the reasons why this appears to be a natural problem. First of all, if $g_*(h^2\,d\theta)=S\,d\theta$, then S satisfies compatibility conditions

$$\int_{S^n} \theta^k S(\theta)\,d\theta=0, \quad k=0,\ldots,n \tag{7.7}$$

identical to (7.5). Second, the linearized operator associated with the problem is selfadjoint. Third, $1-G(x)$ is the equilibrium potential for Ω and for $N\geqq 3$, the capacity of Ω defined as the constant $c=\text{cap}(\Omega)$ such that

$$1-G(x)=c\,|x|^{2-N}+O(|x|^{1-N}) \quad \text{as} \quad x\to\infty$$

satisfies an analogue of (7.4)

$$\text{cap}(\Omega)=\int u h^2\,d\sigma, \quad \text{if } N\geqq 3. \tag{7.8}$$

This leads us to conjecture an analogue of the Brunn-Minkowski inequality

Conjecture. *For $N\geqq 3$, $\text{cap}((1-t)\Omega_0+t\Omega_1)^{\frac{1}{N-2}}$ is a concave function of $t\in[0,1]$. It is linear if and only if Ω_1 is a dilate and translate of Ω_0.*

Because the linearized operator is selfadjoint it is rather easy to calculate the second variation. The conjectured concavity property $(\mathrm{d}^2/\mathrm{d}t^2)\,\mathrm{cap}(\Omega_t)^{\frac{1}{N-2}} \leqq 0$ at $t=0$, for instance, is equivalent to the inequality

$$\int u S\,\mathrm{d}\theta \int v L v\,\mathrm{d}\theta \leqq (N-3)\left(\int v S\,\mathrm{d}\theta\right)^2 \tag{7.9}$$

where $S = g_*(h^2\,\mathrm{d}\sigma)$ for the domain Ω_0, u is the support function for Ω_0, L is the 1st variation of $u \mapsto h^2\,\mathrm{d}\sigma$ at u, and v is the support function for Ω_1. In particular in $\mathbb{R}^3$, the inequality is equivalent to $L \leqq 0$. A similar inequality for the first variation arises in $\mathbb{R}^2$.

If this conjecture were true, then one could hope to solve the problem of prescribing $h^2\,\mathrm{d}\sigma$ as follows. Given $S \in C^\infty(S^n)$ such that $S>0$ and S satisfies (7.7), minimize $\int u S\,\mathrm{d}\theta$ subject to the constraint $\mathrm{cap}(\Omega)=1$, where u is the support function of Ω. If the minimum is λ, then one would expect that the minimizing domain satisfies $g_*(h^2\,\mathrm{d}\sigma) = \lambda^{-1} S$. When $N \geqq 4$, one could then dilate to solve $g_*(h^2\,\mathrm{d}\sigma) = S$. When $N=3$, $g_*(h^2\,\mathrm{d}\sigma)$ is invariant under dilation as well as translation, so one can only expect to solve for $\lambda^{-1} S$ for a single $\lambda > 0$. The minimum problem is entirely analogous to Minkowski's approach to the Minkowski problem with volume replaced by capacity [BF].

There is a large literature on capacity. Isoperimetric inequalities for capacity were studied by Poincaré and Szego, among others (See [P]). However our conjecture seems to be complementary to earlier work. The conjectured concavity property should be contrasted with a convexity theorem, or lower bound for $(\mathrm{d}^2/\mathrm{d}t^2)\,\mathrm{cap}(\Omega_t)$, for certain perturbations due to Garabedian and Schiffer [GS].

References

[BF] Bonnesen, T., Fenchel, W.: Theory of Convex Sets.

[C1] Caffarelli, L.A.: A localization property of viscosity solutions to Monge-Ampère equation and their strict convexity. Ann. Math. **131**, 129–134 (1990)

[C2] Caffarelli, L.A.: Interior $W^{2,p}$ estimates for solutions of Monge-Ampère equation. Ann. Math. **131**, 135–150 (1990)

[C3] Caffarelli, L.A.: Some regularity properties of solutions to Monge-Ampère equation. Commun. Pure Appl. Math. (to appear)

[C4] Caffarelli, L.A.: Interior a priori estimates for solutions of fully non-linear equations. Ann. Math. **131**, 189–213 (1989)

[CY1] Cheng, S.-Y., Yau, S.-T.: On the regularity of the solution of the n-dimensional Minkowski problem. Commun. Pure Appl. Math. **29**, 495–516 (1976)

[CY2] Cheng, S.-Y., Yau, S.-T.: On the regularity of the Monge-Ampère equation $\det(\partial^2 u/\partial x_i\,\partial x_j) = F(x, u)$. Commun. Pure Appl. Math. **30**, 41–68 (1977)

[D] Dahlberg, B.E.J.: Estimates for harmonic measure. Arch. Ration. Mech. Anal. **65**, 275–283 (1977)

[dG] deGuzman, M.: Differentiation of Integrals in $\mathbb{R}^n$. (Lect. Notes Math., vol. 481). Berlin Heidelberg New York: Springer 1975

[G] Garabedian, P.R.: Partial Differential Equations. New York: Wiley 1964

[GS] Garabedian, P.R., Schiffer, M.: Convexity of domain functionals. J. Anal. Math. **2**, 281–368 (1952–53)

[GT] Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order, 2nd ed. Berlin Heidelberg New York: Springer 1983

[HW1] Hunt, R., Wheeden, R.W.: On the boundary values of harmonic functions. Trans. Am. Math. Soc. **132**, 307–322 (1968)

[HW2] Hunt, R., Wheeden, R.W.: Positive harmonic functions on Lipschitz domains. Trans. Am. Math. Soc. **147**, 507–527 (1970)

[J1] Jerison, D.: Harmonic measure in convex domains. Bull. Am. Math. Soc. **21**, 255–260 (1989)

[J2] Jerison, D.: Sharp estimates for harmonic measure on convex domains. In: Dahlberg, B., Fabes, E., Fefferman, R., Jerison, D., Kenig, C. (eds.) I.M.A. Series Springer-Verlag P.D.E. Workshop at U. of Chicago (to appear)

[J3] Jerison, D.: Regularity of the Poisson kernel and free boundary problems. In: Duda, R. (ed). Colloquium Publications volume in honor of A. Zygmund. 1991 (to appear)

[JK] Jerison, D., Kenig, C.E.: Boundary behavior of harmonic functions in nontangentially accessible domains. Adv. Math. **46**, 80–147 (1982)

[L] Lieberman, G.M.: The quasilinear Dirichlet problem with decreased regularity at the boundary. Commun. Partial Differ. Equations **6**, 437–497 (1981)

[P] Payne, L.E.: Isoperimetric inequalities and their applications. SIAM Rev. **9**, 453–488 (1967)

Invent. math. 105, 401–414 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Curvature, tangentiality, and controlled topology

Steven C. Ferry[1] **and Shmuel Weinberger**[2]*

[1] Department of Mathematical Sciences, Suny at Birghamton, Birghamton, NY 13901, USA
[2] Department of Mathematics University of Chicago, Chicago, IL 60637, USA

Oblatum 29-X-1990

1 Introduction

This paper is the first of several where we will combine ideas related to curvature with those of controlled topology to get information about unstable tangent bundles. Our approach is closely related to that of [12], where the Novikov Conjecture was proved for closed nonpositively curved manifolds. Our methods improve those of [12] in that we dispense with compactness (or even finite volume) restrictions and obtain unstable results. This last is desirable because fibered smoothing theory then implies splitting theorems for diffeomorphism groups. These results are not obviously attainable by the earlier (stable) approach.

To emphasize the simplicity of the method, this paper only deals with manifolds of nonnegative sectional curvature. Later papers in the series will deal with group actions (orbifolds), stratified spaces, and other curvature hypotheses. While the technology involved in these applications is more complicated, the basic method is well illustrated in the present paper.

Here is a somewhat loose statement of our main geometrical result. A more precise version is stated as Theorem 2.2 in the next section.

Theorem 1 *Let W be a complete Riemannian manifold of nonpositive curvature and dimension $\geqq 4$. Suppose that $f : W \to W'$ is a proper homotopy equivalence and a homeomorphism in the complement of some compact set. Then f is canonically covered by an isomorphism of unstable topological tangent bundles. The isomorphism produced agrees with the isomorphism given by f outside of a, perhaps larger, compact set.*

The first author was partially supported by NSF Grant DMS – 8702991, while the second was partially supported by an NSF grant and an NSF Presidential Young Investigator Award.

* *Current address:* Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012, USA

This result implies, by Novikov's theorem [25] on the topological invariance of rational Pontrjagin classes, that the rational Pontrjagin classes of W and W' are the same. By surgery theory (see Sec. 1) this implies the usual Novikov conjecture:

Corollary 1.1 *Let W^n be a complete Riemannian manifold of nonpositive curvature and dimension $n \geqq 4$. If $f : M^n \to M'^n$ is a proper homotopy equivalence which is a homeomorphism near infinity and $g : M \to W$ is a proper map then $(g \circ f)_*(L(M) \cap [M]) = g_*(L(M') \cap [M'])$ in rational homology $H_*(W; \mathbb{Q})$. In fact, the classes of the signature operators (which, away from 2, are the Sullivan orientations [29], [24]) for M and M' agree when pushed into W.*

The above corollary is due to Kasparov [21] by a rather different method. There is a more refined version at the prime 2 of the equality of characteristic classes (see [33]) that follows from our theorem. Algebraically, this can be stated as:

Corollary 1.2 *The Ranicki-Quinn surgery assembly map for $\pi_1(W)$ is split injective.*

This refinement implies the following:

Corollary 1.3 (Stable Homeomorphism) *Let W be a complete Riemannian manifold of nonpositive curvature, and suppose that $f : W \to W'$ is a proper homotopy equivalence which is a homeomorphism in the complement of some compact set. Then $f \times \mathrm{id} : W \times \mathbb{R}^3 \to W' \times \mathbb{R}^3$ is properly homotopic to a homeomorphism.*

Farrell and Jones, [13], have recently announced that closed manifolds of nonpositive curvature are rigid, that is, that the above corollary is true for closed manifolds without any stabilization. Oddly, their powerful methods use stable techniques and therefore do not seem to imply our theorems on splitting unstable diffeomorphism groups. In particular, they get calculations of such homeomorphism groups in the concordance stable range for closed manifolds of dimension at least 11.

Our construction is slightly more general than stated above in that compact support is not really necessary. We can prove stable homeomorphism whenever the point-inverses $f^{-1}(x)$ don't grow too quickly as x approaches infinity.

Corollary (of method) 1.4 *Let W be a complete Riemannian manifold of nonpositive curvature, and suppose that $f : W \to W'$ is a proper homotopy equivalence for which $\mathrm{diam}(\exp^{-1} f^{-1}(f(x)))$ is uniformly bounded for all x, then $f \times \mathrm{id} : W \times \mathbb{R}^3 \to W' \times \mathbb{R}^3$ is properly homotopic to a homeomorphism.*

This was misstated in [17, 33]. If the injectivity radius of W is growing at infinity one can use the boundedness of $\mathrm{diam}(f^{-1}(f(x))$.

Recall the main result of smoothing theory [19]: A manifold is smoothable if and only if its (stable) tangent microbundle can be given a vector bundle structure. Consequently, our main theorem shows that if W is complete with nonpositive curvature and W' is homotopy equivalent to W by a homotopy equivalence which is a homeomorphism near infinity, then W' is smoothable. Fibered smoothing theory requires similar (but unstable) bundle information, so a fibered version of this observation yields the following:

Theorem 2 (i) *If W^n, $n \geqq 5$, is a complete noncompact Riemannian manifold of nonpositive curvature, then the forgetful map $\mathrm{Diffeo}^c_\circ(W) \to \mathrm{Homeo}^c_\circ(W)$ has a*

homotopy section. Here $\mathrm{Diffeo}^c_\circ(W)$ *and* $\mathrm{Homeo}^c_\circ(W)$ *are the identity components of the groups of diffeomorphisms and homeomorphisms of* M *with compact support.*

(ii) *If* $W^n, n \geqq 5$, *is a compact irreducible locally symmetric space of noncompact type (e.g. a compact hyperbolic manifold) then* $\mathrm{Diffeo}^c_\circ(M) \to \mathrm{Homeo}^c_\circ(M)$ *has a homotopy section. In addition, given* $\varepsilon > 0$, *these sections can be chosen to provide diffeomorphisms within* ε *of the original homeomorphisms.*

Stabilizing this result, and applying Waldhausen's K-theory, we obtain:

Theorem 3. *If* W *is a complete Riemannian manifold of nonpositive curvature, then the A-theory assembly map* $W_+ \wedge A(*) \to A(W)$ *is split injective.*

Recently, Bökstedt, et al. (see [2]) used entirely different methods to prove a rational version of this result for a much larger class of groups.

In future papers we will generalize this method to split other assembly maps. Heuristically, the idea is to build a model for the tangent bundle for which the differential has good control properties. Bounded topology then determines the bundle. In the present case, the key result from controlled topology is the α-approximation Theorem, which deforms maps with small point-inverse images to homeomorphisms.

A similar philosophy occurs in the C^*-algebraic setting. Roe [28] has reformulated some of the ideas of [8] in an analytic setting similar to our controlled one. There is a basic cocycle calculation that provides an analogue to the α-approximation theorem, (see [34] for a translation of this into an analytic proof of Novikov's theorem). One uses this calculus in parametrized ways to deduce Novikov-type results. While our topological argument gives more precise geometric topological and algebraic conclusions, these analytical methods apply to a variety of elliptic operators and hence have applications in differential and holomorphic geometry.

Acknowledgemets. Some of the results of this series have been announced in [17]. We would like to thank Frank Connolly, Jonathan Rosenberg, Mel Rothenberg, and Bruce Williams for usefull conversations. The first author would like to thank University of Chicago for its hospitality during a visit in Spring '87 when this research was begun, and both authors would like to thank the Courant Institute for its support during several visits since.

2 Controlled topology and Kasparov's Theorem

We begin by proving that a homotopy equivalence from a topological manifold to a nonpositively curved manifold is unstably tangential. As a corollary, we obtain an integral version of the Novikov Conjecture for complete Riemannian manifolds of nonpositive curvature. Related results are proven by C^*-algebra techniques in [21].

Definitions 2.1 If M^n is a topological manifold, a *topological tangent bundle* τ_M for M is a neighborhood U of the diagonal $\triangle M \subset M \times M$ such that the projection $p_1 : U \to M$ is a topological $\mathbb{R}^n$-bundle. Here, $p_1(x, y) = x$. Note that $p_2 : \tau_M \to M$ gives a topological analog of the exponential map. The Kister-Mazur Theorem, [23], guarantees that every neighborhood of $\triangle M$ in $M \times M$ contains a topological tangent bundle for M and that any two topological tangent

bundles for M are unique up to bundle isomorphism relative to any subbundle where they already agree.

For the purposes of this paper, we will extend the definition of "topological tangent bundle" slightly by defining a topological tangent bundle for M to be a topological $\mathbb{R}^n$-bundle together with a fiber-preserving embedding $h : U \to E$, where $U \subset M \times M$ is a neighborhood of M as above. The section $h| \circ \triangle : M \to \triangle M \to E$ is called the zero section.

A homeomorphism $f : M \to N$ between topological manifolds induces an isomorphism of tangent bundles as follows: If $V \subset N \times N$ is a topological tangent bundle for N, let $U \subset M \times M$ be $(f \times f)^{-1}(V)$. One easily verifies that $p_1 : U \to M$ is the pullback of $p_1 : V \to N$. The bundle map $f \times f : U \to V$ is called the *differential* of f. More generally, we will call a fiber-preserving map $\hat{f}$ between topological tangent bundles a differential of a homeomorphism f if $\hat{f}$ covers f and it $\hat{f}$ agrees with $f \times f$ in a neighborhood of the zero section.

If $f : M \to N$ is a homotopy equivalence of topological manifolds, f is said to be *unstably tangential* if there are topological tangent bundles τ_M and τ_N for M and N and a map $\hat{f} : \tau_M \to \tau_N$ covering f such that the restriction of $\hat{f}$ to each fiber is a homeomorphism. This is equivalent to saying that the induced bundle $f^*\tau_N$ is fiber-preserving homeomorphic to τ_M. The map f is *unstably normal* if there is such an $\hat{f}$ which is properly homotopic to a homeomorphism $h : \tau_M \to \tau_N$.

Theorem 2.2 *Let $M^n, n \geqq 4$, be a complete Riemannian manifold with nonpositive curvature and let $f : M \to N$ be a homotopy equivalence which is a homeomorphism near infinity. Then f is unstably normal. The maps $\hat{f}$ and h, as well as the homotopy between them, can be taken to be the differential of f outside of $p_1^{-1}(C)$, where $C \subset M$ is a compact set.*

Corollary 2.3 (Novikov conjecture) *If π is the fundamental group of a manifold M^n of nonpositive curvature, then the Ranicki-Quinn assembly map $H_n(\pi; \mathbb{L}) \to L_n^K(\mathbb{Z}\pi)$ is a split monomorphism. Here, the decoration K denotes an arbitrary subgroup of $Wh(\mathbb{Z}\pi)$.*

Proof. We first consider the case in which M^n is orientable and $n \geqq 5$. Consider the surgery exact sequence with compact supports:

$$\to \mathcal{S}_c^K(N \times \mathbb{R}^1) \to [N \times \mathbb{R}^1, \infty; F/\mathrm{TOP}, *] \to L_{n+1}^K(\mathbb{Z}\pi)$$
$$\to \mathcal{S}_c^K(N) \to [N, \infty; F/\mathrm{TOP}, *] \to L_n^K(\mathbb{Z}\pi).$$

Since N is orientable, $[N, \infty; F/\mathrm{TOP}, *] \cong H_c^0(N; \mathbb{L}) \cong H_n(N; \mathbb{L})$, so the last map in the sequence gives a map $H_n(N; \mathbb{L}) \to L_n^K(\mathbb{Z}\pi)$. This is the assembly map in question.

If a map $f : M \to N$ gives a structure on N with compact support, construct $\hat{f}$ and h as in Theorem 2.2 above. Making the proper homotopy from f to h transverse to the zero section in N relative to a neighborhood of ∞ gives a normal cobordism from M to N. This implies that the map $\mathcal{S}_c^K(N) \to [N, \infty; F/\mathrm{TOP}, *]$ is zero. See [3, p. 49]. The map $[N, \infty; F/\mathrm{TOP}, *] \to L_n^K(\mathbb{Z}\pi)$ is therefore a monomorphism.

If $n \leqq 4$, we do not have a surgery exact sequence, but we note that the entire surgery sequence above is periodic with period 4, so the result follows upon replacing N by $N \times \mathbb{R}^4$. Note that for compactly generated spaces X and Y, $[X \times \mathbb{R}^1, \infty; Y, *] \cong [X, \infty; \Omega Y, *]$.

If M is a nonorientable complete Riemannian manifold of nonpositive curvature, then there is a free orientation-preserving action of $\pi = \pi_1(M)$ on $\widetilde{M} \times \widetilde{M}$. Applying the argument above to the orientable manifold $\widetilde{M} \times_\pi \widetilde{M}$ gives a split monomorphism $H_{2n}(\pi; \mathbb{L}) \to L^K_{2n}(\mathbb{Z}\pi)$. Periodicity and crossing with $\mathbb{R}^1$ give the result in other dimensions as above. This completes the proof that the assembly map is a monomorphism. An easy way to see the splitting is to note that the argument given above is sufficiently canonical that it works for surgery spaces, showing that the map from the structure space to the normal maps is nullhomotopic. This will be proven in detail in a later paper in this series. A splitting based on this idea is constructed in a similar situation in [12].

Definitions 2.4 (i) If $p : E \to X$ is a vector bundle, then we will denote the open disk-bundle of radius t centered at the zero section by E_t. Similarly, B^n_t will denote a ball of radius t in $\mathbb{R}^n$ which is centered at the origin. We will frequently write B^n in place of B^n_1.
(ii) A map $f : X \to Y$ from a metric space X to a space Y is called an *ε-map* if $\operatorname{diam}(f^{-1}(y)) < \varepsilon$ for each $y \in Y$.

Here is the main technical result used in proving Theorem 2.2:

Theorem 2.5 *Let $n \geqq 4$ and k be given. Then for every $\varepsilon > 0$ and r with $0 < r < 1$ there is a $\delta > 0$ such that if:*

(1) *X^k is a k-dimensional simplicial complex.*
(2) *$p : E \to X^k$ is a vector bundle with fiber $\mathbb{R}^n$.*
(3) *$p' : E' \to X$ is a topological $\mathbb{R}^n$-bundle.*
(4) *$f : E_1 \to E'_1$ is a fiber-preserving δ-map.*

Then there is an open neighborhood U of $f(E_r)$ in E' and a fiber-preserving open embedding $h : U \to E$ with $d(h \circ f | E_r, \mathrm{id}) < \varepsilon$. If, in addition, Y is a closed subcomplex of X and $f | : E_1 \to E'$ is a fiber-preserving open embedding over Y, then h can be taken equal to f^{-1} over Y. Here the term "over" means "when restricted to the inverse image of."

Corollary 2.6 *Let E and E' be a vector bundle and a topological $\mathbb{R}^n$-bundle over a finite-dimensional complex X as above. If there exists a fiber-preserving δ-map $f : E \to E'$ for any $\delta > 0$, then f is properly fiber-preserving isotopic to a fiberwise homeomorphism $h : E \to E'$.*

Proof. Let $\varepsilon > 0$ be given and let $\delta' > 0$ be as in the statement of Theorem 2.5 for $r = \frac{1}{2}$. Let $s : E \to E$ be multiplication by $\frac{\delta}{\delta'}$. Then applying Theorem 2.5 to $f \circ s$, there is a fiber-preserving open embedding $h^{-1} : E_{1/2} \to E'$. The uniqueness statement in the Kister-Mazur Theorem shows that E and E' are isomorphic as topological $\mathbb{R}^n$-bundles.

Corollary 2.7 *Let $n \geqq 4$ and k be given. Then there is a $\delta > 0$ such that if X^k is a k-dimensional complex, $p : E \to X^k$ is a vector bundle with fiber R^n, $p' : E' \to X'$ is a topological R^n-bundle, and $f : E \to E'$ is a bundle map such*

that the restriction of f to each fiber is a δ-map, then $f^*(E')$ is isomorphic to E as a topological R^n-bundle.

Proof. The map f induces a fiber-preserving map $E \to f^*(E')$. Corollary 2.7 then follows immediately from Corollary 2.6.

Theorem 2.5 follows easily by induction from the following:

Proposition 2.8 *Let $n \geqq 4$ and k be given. Then for every $\varepsilon > 0$ and r with $0 < r < 1$ there is a $\delta > 0$ such that if $\triangle^k$ is a k-dimensional simplex and $f : \mathbb{R}^n \times \triangle^k \to \mathbb{R}^n \times \triangle^k$ is a fiber-preserving map such that $f|B^n \times \triangle^k$ has point-inverses of diameter $< \delta$ and such that $f|B^n \times \partial\triangle^k$ is an open embedding, then there is an open neighborhood U of $f(B^n_r \times \triangle^k)$ in $\mathbb{R}^n \times \triangle^k$ and a fiber-preserving open embedding $h : U \to \mathbb{R}^n \times \triangle^k$ such that $B^n_r \times \triangle^k \subset h(U)$ with $d(h \circ f|B^n_r \times \triangle^k, \mathrm{id}) < \varepsilon$ and $h|\partial\triangle^k = f^{-1}$. Moreover, $f|(B^n_r \times \triangle^k)$ and $h^{-1}|B^n_r \times \triangle^k$ are homotopic rel $\partial\triangle^k$ through 2ε-maps.*

The proof of Proposition 2.8 relies heavily on the following result of Černavskii [5] and Edwards and Kirby [10]:

Theorem (Local contractibility for embeddings) *Let n, r, and s be given, with $0 < r < s$. Then for each $k > 0$ and $\varepsilon > 0$ there exists a $\delta > 0$ such that if X^k is any k-dimensional polyhedron and $f : \mathrm{int}(B^n_s) \times X \to \mathbb{R}^n \times X$ is a fiber-preserving family of open embeddings δ-close to the inclusion, then there is an isotopy $H_t : \mathrm{int}(B^n_r) \times X \to R^n \times X$ of fiber-preserving open embeddings such that $d(H_t, \mathrm{id}) < \varepsilon$ for all t and such that $H_0 = f|\mathrm{int}(B^n_r) \times X$ and $H_1 =$ inclusion.*

Proof (Proposition 2.8). The proof is by induction on k.

Step I. $k = 0$. This case says that given $n \geqq 4$, $\varepsilon > 0$, and r with $0 < r < 1$ there is a $\delta > 0$ such that if $f : \mathbb{R}^n \to \mathbb{R}^n$ is a map such that $f|B^n$ has point-inverses of diameter $< \delta$, then there is an open neighborhood U of $f(B^n_r)$ in $\mathbb{R}^n$ and an open embedding $h : U \to \mathbb{R}^n$ such that $B^n_r \subset h(U)$ with $d(h \circ f|B^n_r, \mathrm{id}) < \varepsilon$. Moreover, $f|B^n_r$ and $h^{-1}|B^n_r$ are homotopic through 2ε-maps.

For $n \geqq 5$, this is proved in § 3 of [15], using results from several sources. We will give a proof in case $n = 4$ below. Here are some relevant definitions and theorems:

Definition 2.9 If X and Y are metric spaces, a map $f : X \to Y$ is called an *ε-domination* if there exist a map $g : Y \to X$ and a homotopy $k_t : f \circ g \sim \mathrm{id}$ so that $d(k_t, \mathrm{id}) < \varepsilon$ for all t. An ε-domination is an *ε-equivalence* if there exists an additional homotopy $h_t : g \circ f \sim \mathrm{id}$ so that $d(f \circ h_t, \mathrm{id}) < \varepsilon$ for all t.

Theorem (Eilenberg [11]) *If X is a compact* ANR, *then for every $\varepsilon > 0$ there is a $\delta > 0$ so that if $g : X \to Y$ is any surjective δ-map, then there is a map $f : Y \to X$ so that $f \circ g$ is ε-homotopic to the identity, i.e., g is the right inverse to an ε-domination.*

Theorem ([15, Corollary 3.7]) *If M^n is a closed topological manifold, then for each $\varepsilon > 0$ there is a $\delta > 0$ so that if $f : M \to N$ is a δ-domination, then f is an ε-equivalence.*

Theorem (α-approximation theorem [7]) *If M^n is a closed topological manifold, then for every $\varepsilon > 0$ there is a $\delta > 0$ so that if $f : M \to N$ is a δ-equivalence, then f is ε-homotopic to a homeomorphism.*

Combining these gives:

Theorem ([15, Theorem 1]) *If M is a closed n-manifold, $n \geqq 5$, and $\varepsilon > 0$ is given, then there is a $\delta > 0$ so that if N is an n-manifold and $g : M \to N$ is a δ-map, then there is a homeomorphism $h : N \to M$ so that $h \circ g$ is ε-homotopic to the identity.*

The case $k = 0$ of Proposition 2.8 is a local version of this last theorem. Such a local theorem is easily proven using the same handle theorems that are used in proving the closed case. The arguments above are dimension-free, except for the α-approximation Theorem. Thus we need to prove:

Theorem 4 *The α-approximation Theorem is true in dimension 4.*

Proof (Theorem 4). This follows from the theorem in dimension 5 and Quinn's Thin h-Cobordism Theorem [27]. If $f : M \to N$ is an ε-equivalence between 4-manifolds, $f \times \mathrm{id} : M \times S^1 \to N \times S^1$ is an ε-equivalence between 5-manifolds. This gives a controlled homeomorphism $h : M \times S^1 \to N \times S^1$ and a thin h-cobordism from M to N. By the Thin h-Cobordism Theorem, M is controlled homeomorphic to N. This result has also been obtained in [1].

Step II. Assume that Proposition 2.8 is true for $k \leqq l - 1$. We will prove it for $k = l$.

Let $\varepsilon > 0$ and r, $0 < r < 1$ be given. Divide the interval $[r, 1]$ into thirds, letting $r_0 = r$, $r_1 = \frac{2r}{3} + \frac{1}{3}$, $r_2 = \frac{r}{3} + \frac{2}{3}$. Let $\varepsilon' = \min(\frac{\varepsilon}{2}, \frac{1-r}{6})$ and choose $\varepsilon_1 < \varepsilon'$ so that if X is any polyhedron of dimension $\leqq k$ and $f : \mathrm{int}(B^n_{r_1}) \times X \to \mathbb{R}^n \times X$ is a fiber-preserving family of open embeddings $2\varepsilon_1$-close to the inclusion, then there is an isotopy $H_t : \mathrm{int}(B^n_{r_0}) \times X \to \mathbb{R}^n \times X$ of fiber-preserving open embeddings such that $d(H_t, \mathrm{id}) < \varepsilon'$ for all t and such that $H_0 = f|\mathrm{int}(B^n_{r_0}) \times X$ and $H_1 =$ inclusion.

By the induction hypothesis, we can choose $\delta > 0$ so that if $p \leqq l - 1$ and $f : \mathbb{R}^n \times \triangle^p \to \mathbb{R}^n \times \triangle^p$ is a fiber-preserving map such that $f|B^n \times \triangle^p$ has point-inverses of diameter $< \delta$ and such that $f|B^n \times \partial\triangle^p$ is an open embedding, then there is an open neighborhood U of $f(B^n_{r_2} \times \triangle^p)$ in $\mathbb{R}^n \times \triangle^p$ and a fiber-preserving open embedding $h : U \to \mathbb{R}^n \times \triangle^p$ such that $B^n_{r_2} \times \triangle^p \subset h(U)$ with $d(h \circ f|B^n_{r_2} \times \triangle^p, \mathrm{id}) < \varepsilon_1$ and $h|\partial\triangle^p = f^{-1}$.

Inducting on cells, we can choose $\delta > 0$ so that if Y is a complex with $\dim(Y) \leqq l - 1$ and $f : \mathbb{R}^n \times Y \to \mathbb{R}^n \times Y$ is a fiber-preserving map such that $f|B^n \times Y$ has point-inverses of diameter $< \delta$ and such that f is an open embedding over some subcomplex $Y_0 \subset Y$, then there is an open neighborhood U of $f(B^n_{r_2} \times Y)$ in $\mathbb{R}^n \times Y$ and a fiber-preserving open embedding $h : U \to \mathbb{R}^n \times Y$ such that $B^n_{r_2} \times Y \subset h(U)$ with $d(h \circ f|B^n_{r_2} \times Y, \mathrm{id}) < \varepsilon_1$ and $h|Y_0 = f^{-1}$.

Let $f : \mathbb{R}^n \times \triangle^l \to \mathbb{R}^n \times \triangle^l$ be a fiber-preserving map as in the statement of Proposition 2.8 which restricts to a δ-map on $B^n \times \triangle^l$. By Step I above, for each $t \in \triangle^l$, we can find a neighborhood $U_t \subset \mathbb{R}^n$ so that $U_t \times \{t\}$ is a neighborhood of $f(B^n_{r_2} \times \{t\})$ and so that there is an open embedding $h_t : U_t \times \{t\} \to \mathbb{R}^n \times \{t\}$ so that $d(h_t \circ f|B^n_{r_2} \times \{t\}, \mathrm{id}) < \varepsilon_1$. For each $t \in \triangle^l$ there is a neighborhood $O_t \subset \triangle^l$ so that $f(B^n_{r_2} \times \{s\}) \subset U_t \times \{s\}$ and $d(h_t \circ f|B^n_{r_2} \times \{s\}, \mathrm{id}) < \varepsilon_1$ for each $s \in O_t$.

Let $\gamma > 0$ be a Lebesgue number for the covering $\{O_t\}$ of $\triangle^l$ and let T be a triangulation of $\triangle^l$ with $\text{mesh}(T) < \gamma$.

By induction as above, there is an open neighborhood U of $f(B^n_{r_2} \times T^{(l-1)})$ and a fiber-preserving open embedding $h : U \to \mathbb{R}^n \times T^{(l-1)}$ so that $d(h \circ f, \text{id}) < \varepsilon_1$. Let $\triangle^l_1$ be a typical l-simplex of T. Since $\triangle^l_1 \subset O_t$ for some t, there is an open neighborhood U' of $f(B^n_{r_2} \times \{t\})$ and an open embedding $h' : U' \to \mathbb{R}^n$ so that $d((h' \times \text{id}) \circ f, \text{id}) < \varepsilon_1$ on $B^n_{r_2} \times \triangle^l_1$. Since U and U' are neighborhoods of $f(B^n_{r_2} \times \partial\triangle^l_1)$, $h \circ (h' \times \text{id})^{-1}$ is defined on $(h' \times \text{id}) \circ f(B^n_{r_2} \times \partial\triangle^l_1)$. Since $d((h' \times \text{id}) \circ f, \text{id})$, $h \circ (h' \times \text{id})^{-1}$ is defined on $B^n_{r_1} \times \partial\triangle^l_1$ and is $2\varepsilon_1$-close to id. Therefore, there is an isotopy of open embeddings $H_t : B^n_{r_0} \times \partial\triangle^l_1 \to \mathbb{R}^n \to \partial\triangle^l_1$ with $H_0 = h \circ (h' \times \text{id})^{-1} | (B^n_{r_0} \times \partial\triangle^l_1)$ and $H_1 = \text{id}$ such that $d(H_t, \text{id}) < \varepsilon'$.

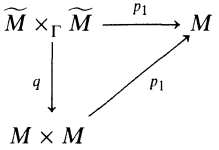
Choose a collar neighborhood of $\partial\triangle^l_1$ in $\triangle^l_1$ and note that $H_0 \circ (h' \times \text{id}) = h$ and $H_1 \circ (h' \times \text{id}) = h' \times \text{id}$. We can therefore use the isotopy $H_t \circ (h' \times \text{id})$ along the collar to construct a fiber-preserving open embedding $\bar{h} : f(B^n_{r_0} \times \triangle^l_1) \to \mathbb{R}^n \times \triangle^l_1$ which is equal to h over $\partial\triangle^l_1$ and $h' \times \text{id}$ over the core of $\triangle^l_1$. Doing this for each l-simplex of T gives us the desired open embedding $\bar{h}$ over all of T. We have $d(\bar{h} \circ f, \text{id}) < \varepsilon$, since

$$\begin{aligned} d(H_t \circ (h' \times \text{id}) \circ f, \text{id}) &\leqq d(H_t \circ (h' \times \text{id}) \circ f, (h' \times \text{id}) \circ f) + d((h' \times \text{id}) \circ f, \text{id}) \\ &\leqq \varepsilon' + \varepsilon_1 < \varepsilon . \end{aligned}$$

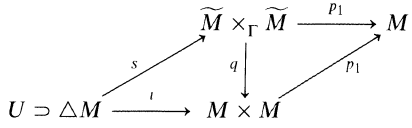
This completes the proof of Proposition 2.8. Theorem 2.5 now follows by induction on skeleta. There is one tricky point here – the fact that $p : E \to X$ is a vector bundle is used to produce trivializations of E over simplices which are isometries on the individual fibers. This is important for applying Proposition 2.8, since the δ is measured in the standard metric on $\mathbb{R}^n \times \triangle^k$. The argument would work for a topological $\mathbb{R}^n$-bundle E only if E had a metric which restricted to the standard metric on each $\mathbb{R}^n$ fiber.

Remark 2.10 Strictly speaking, the use of ε-maps in the preceding was unnecessary. Since a homotopy equivalence $f : N \to N$ comes equipped with a homotopy inverse, we could have kept track of lifts of both maps and used the α-approximation Theorem directly.

The topological exponential cover Let $\widetilde{M}$ be the universal cover of M. Then $\Gamma = \pi_1(M)$ acts on $\widetilde{M}$. Let Γ act diagonally on $\widetilde{M} \times \widetilde{M}$ and form the balanced product $\widetilde{M} \times_\Gamma \widetilde{M}$. Projection onto the first factor induces a map $p_1 : \widetilde{M} \times_\Gamma \widetilde{M} \to \widetilde{M}/\Gamma = M$. Similarly, we have a map $q : \widetilde{M} \times_\Gamma \widetilde{M} \to M \times M$ defined by projecting onto both factors mod Γ. We have a commuting diagram:



Note that there is a map $s : \triangle M \to \widetilde{M} \times_\Gamma \widetilde{M}$ defined by choosing a representative $\tilde{m}$ for m in $\widetilde{M}$ and sending m to $[\tilde{m}, \tilde{m}]$ in $\widetilde{M} \times_\Gamma \widetilde{M}$. This map extends to a neighborhood U of $\triangle M$ in $M \times M$. Given a pair (m_1, m_2) with m_1 close to m_2, we choose $(\tilde{m}_1, \tilde{m}_2)$ with $\tilde{m}_1$ close to $\tilde{m}_2$. If the distances are small enough, these maps are well-defined. This gives a commuting diagram:

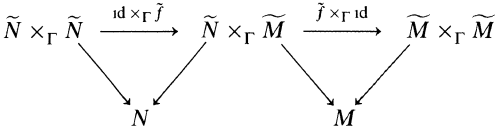


Now consider the case in which $\widetilde{M} \cong \mathbb{R}^n$. In this case, $\widetilde{M} \times_\Gamma \widetilde{M} \to M$ is an $\mathbb{R}^n$-bundle and $\widetilde{M} \times_\Gamma \widetilde{M} \to M$ is a topological tangent bundle for M in the sense of Definition 2.1. The map $p_2 : \widetilde{M} \times_\Gamma \widetilde{M} \to M$ is now a topological analog of the exponential map $\tau_M \to M$ which restricts to a universal covering on the tangent space at each point. The zero section is given by $s(m) = [\tilde{m}, \tilde{m}]$ as above.

Let N be a Riemannian manifold and let $f : N \to M$ be a map inducing an isomorphism $f_* : \pi_1(N) \to \pi_1(M) = \Gamma$. Let $\tilde{f} : \widetilde{N} \to \widetilde{M}$ be a lifting of f and consider $\tilde{f} \times_\Gamma \tilde{f} : \widetilde{N} \times_\Gamma \widetilde{N} \to \widetilde{M} \times_\Gamma \widetilde{M}$. Note that $\tilde{f} \times_\Gamma \tilde{f}$ is independent of the choice of lifting $\tilde{f}$. Define a fiberwise exponential map $\exp : TN \to \widetilde{N} \times_\Gamma \widetilde{N}$ by identifying TN_n with the tangent space to $p_1^{-1}(n)$, which is a copy of $\widetilde{N}$, at $[\tilde{n}, \tilde{n}]$ and using the ordinary exponential map. If N is complete with nonpositive sectional curvature, a theorem of Hadamard [18], shows that $\exp : TN_n \to p_1^{-1}(n)$ is a homeomorphism for each n. In fact, Hadamard's theorem shows that $d(\exp(x), \exp(y)) \geqq d(x, y)$ for each $x, y \in TN_n$. (We are using TN and τ_N to distinguish differentiable tangent bundles and honest exponential maps from microbundles and topological exponential covers). The map $\tilde{f} \times_\Gamma \tilde{f} : \widetilde{M} \times_\Gamma \widetilde{M} \to \widetilde{N} \times_\Gamma \widetilde{N}$ consists of copies of $\tilde{f}$ mapping the fiber of $\widetilde{N} \times_\Gamma \widetilde{N} \to N$ over n to the fiber of $\widetilde{M} \times_\Gamma \widetilde{M} \to M$ over $f(n)$.

We are now ready to prove Theorem 2.2. If N is complete with nonpositive sectional curvature and $f : N \to M$ is a homotopy equivalence which is a homeomorphism near infinity, construct $\tilde{f} \times_\Gamma \tilde{f} \circ \exp : TN \to \widetilde{M} \times_\Gamma \widetilde{M}$ as above. By Hadamard's Theorem, $\widetilde{N}$ is diffeomorphic to $\mathbb{R}^n$. Since $\tilde{f} : \widetilde{N} \to \widetilde{M}$ is a proper homotopy equivalence, it follows for $n \geqq 4$ that $\widetilde{M}$ is homeomorphic to $\mathbb{R}^n$, as well. The point-inverses of $\tilde{f}$ cover the point-inverses of f. Since the map $f^{-1}(m) \to M$ is trivial on π_1, the covering is trivial and the map $\tilde{f}^{-1}(\tilde{m}) \to f^{-1}(m)$ is a homeomorphism for each $m \in M$ and $\tilde{m} \in \widetilde{M}$ with $p_1(\tilde{m}) = m$. In particular, the point-inverses of f are compact. Since f is a homeomorphism off of a compact set, it follows easily that there is a $K > 0$ so that $\operatorname{diam}(\tilde{f}^{-1}(\tilde{m})) \leqq K$ for all $\tilde{m} \in \widetilde{M}$. It follows from Corollary 2.7 that f is unstably tangential.

That f is unstably normal follows from the fact that f may be factored as:



Corollary 2.7 applies directly to show that $(\mathrm{id} \times_\Gamma \tilde{f}) \circ \exp^{-1} : TN \to \tilde{N} \times_\Gamma \widetilde{M}$ is properly homotopic to a fiber-preserving homeomorphism, so composition with exp gives a fiber-preserving homeomorphism $\tilde{N} \times_\Gamma \tilde{N} \to \tilde{N} \times_\Gamma \widetilde{M}$ properly homotopic to $\mathrm{id} \times_\Gamma \tilde{f}$.

To apply the same trick to $\tilde{f} \times_\Gamma \mathrm{id}$, we must construct a similar fiber-preserving exponential homeomorphism from a vector bundle E to the $\mathbb{R}^n$-bundle $\tilde{N} \times_\Gamma \widetilde{M} \to M$. Here is the construction: Γ acts on $T\tilde{N}$ and $\exp : T\tilde{N} \to \tilde{N}$ is equivariant, so exp induces a map $\exp : T\tilde{N} \times_\Gamma \widetilde{M} \to \tilde{N} \times_\Gamma \widetilde{M}$. The vector bundle projection $T\tilde{N} \to \tilde{N}$ induces a vector bundle projection $T\tilde{N} \times_\Gamma \widetilde{M} \to \tilde{N} \times_\Gamma \widetilde{M}$. Since the map $\tilde{N} \times_\Gamma \widetilde{M} \to M$ is an $\mathbb{R}^n$-bundle, there is a section $s : M \to \tilde{N} \times_\Gamma \widetilde{M}$. The required vector bundle $E \to M$ is the restriction of $T\tilde{N} \times_\Gamma \widetilde{M} \to \tilde{N} \times_\Gamma \widetilde{M}$ to the zero section s and the required exponential homeomorphism is $\exp|E$. That $\tilde{f}$ is unstably normal now follows from Corollary 2.7.

The statement of Theorem 2.2 is stronger than what we have shown in that it requires that the constructions be done relative to a neighborhood of infinity. This is important in the noncompact Novikov Conjecture. See [20] for a description of some pitfalls in the noncompact case. Here is a version of Corollary 2.6 which works "rel infinity."

Proposition 2.11 *Let E be an n-dimensional vector bundle and let E' be a topological $\mathbb{R}^n$ bundle, $n > 4$, over a finite-dimensional complex X. If there exists a fiber-preserving δ-map $f : E \to E'$ for any $\delta > 0$, then f is properly fiber-preserving isotopic to a fiberwise homeomorphism $h : E \to E'$. For each $\delta > 0$ there is an $r > 0$ depending on $\dim X, n$, and δ, such that if Y is a subcomplex of X, $s : Y \to E$ is a section, and $f|E_r$ is a fiber-preserving embedding, then we can take h to be equal to f on some neighborhood V of $s(Y)$. Here, E_r is the subbundle of $E|Y$ such that $(E_r)_y$ is the ball of radius r around $s(y)$.*

Proof. This follows quickly from Theorem 2.5. Extend the section s to a section s over all of X and consider E to be a vector bundle with s as zero-section. If $R > 0$ is a real number, then $f(Rv)$ is a fiber-preserving $\frac{\delta}{R}$-map which is an open embedding when restricted to the $\frac{r}{R}$-sphere. For R and r sufficiently large, Theorem 2.5 guarantees that $v \to f(Rv)$ is fiber-preserving homotopic to an open embedding h on $E_{\frac{1}{2}}$ with $h|E_{\frac{1}{2}} = f(Rv)|E_{\frac{1}{2}}$ over Y. Then $h(\frac{1}{R}v)$ is an open embedding equal to f on a neighborhood of $s(Y)$. The Kister-Mazur Theorem gives us a fiber-preserving homeomorphism $h : E \to E'$ which is equal to h in a neighborhood of the zero-section.

The proof of Theorem 2.2 is now immediate.

Remarks 2.12 There are two or three approaches to proving the α-approximation Theorem. The first, found in [7], is directly modeled on Siebenmann's proof [29] of the CE Approximation Theorem, which says that CE maps are uniform limits of homeomorphisms. This approach uses surgery and torus geometry. A second proof is constructed using Chapman's squeezing process [6] to approximate an α-equivalence by a CE map and then using Siebenmann's theorem or Edwards' Disjoint Disk Theorem [9]. This last is interesting to geometry aficionados in that it replaces the reliance on surgery theory by a direct geometric argument.

One can also maintain a purely bounded point of view. (Actually, this is a little more cumbersome for our results on noncompact manifolds). The key result, replacing, and, in fact, equivalent to α-approximation, is that the space (in an appropriate sense) of bounded homotopy structures of Euclidean space is contractible. In view of bounded surgery theory [16], this boils down to the statement that the bounded surgery theory of Euclidean space is just a shifted version of usual surgery theory. In general, our method can be viewed as a descent from bounded topology on universal covers to ordinary topology on spaces.

Of course, to detect the image of the assembly map one does not really need isomorphism results; injectivity (which is sufficiently canonical) suffices. Working rationally, for instance, one can prove this by detecting boundedly the codimension n signature for a manifold equiped with a map to Euclidean space. This is, in effect, the upshot of the basic cyclic cocyle calculation in Roe's [28] reformulation of [8].

3 Splitting the A-theory assembly map

Waldhausen, see [31], has defined a functor $A(X)$ from spaces to spaces which is analogous to the algebraic K-theory functor. For each space X, there is an assembly map $BX_+ \wedge A(*) \to A(X)$. We now prove Theorem 3 of the introduction, which says that this assembly map splits when X is a complete Riemannian manifold of nonpositive sectional curvature.

Proof. By Theorem 3.1 of [31], $BX_+ \wedge A(*)$ is the homotopy fiber of the natural map $\varrho : A(X) \to Wh^{\mathrm{PL}}(X)$ and the inclusion map $i : BX_+ \wedge A(*) \to A(X)$ is the assembly map. Here, $Wh^{\mathrm{PL}}(X)$ is the PL Whitehead space of X. We produce a map $\sigma : Wh^{\mathrm{PL}}(X) \to A(X)$ which is a right inverse for ϱ. Using the loop structure on $A(X)$ to multiply maps, we obtain a map $i \times \sigma : (BX_+ \wedge A(*)) \times Wh^{\mathrm{PL}}(X) \to A(X)$ which is a homology isomorphism and therefore a homotopy equivalence. Thus, to prove Theorem 2 we need only produce the map σ.

By [31], there is a homotopy splitting $A(X) \cong Wh^{\mathrm{DIFF}}(X) \times \Omega^\infty S^\infty(X_+)$ so that $\varrho | Wh^{\mathrm{DIFF}}(X)$ is the forgetful map $\varrho' : Wh^{\mathrm{DIFF}}(X) \to Wh^{\mathrm{PL}}(X)$. Thus, it suffices to construct a homotopy right inverse for ϱ'. By Theorem 1 of [32], $Wh^{\mathrm{CAT}}(X)$ is the stabilized CAT h-cobordism space, $\varinjlim \mathscr{H}^{\mathrm{CAT}}(X \times I^n)$, so we need only construct a suitably natural homotopy right inverse for the forgetful map $\varrho'' : \mathscr{H}^{\mathrm{DIFF}}(X) \to \mathscr{H}^{\mathrm{PL}}(X)$. This map is constructed using parameterized smoothing theory.

A k-simplex of $\mathscr{H}^{\mathrm{CAT}}(X)$ is a CAT fiber bundle over Δ^k with fiber $X \times I$ and a given product structure on $(X \times 0 \cup \partial X \times I) \times \Delta^k$. Thus, $\mathscr{H}^{\mathrm{CAT}}(X)$ is a classifying space for CAT concordances. Fibered smoothing theory ([4], [22, Essay V]) says that smoothings of such a bundle are in $1-1$ correspondence with homotopy classes of reductions of the unstable tangent bundle from CAT_n to $O(n)$ which restrict to reductions from CAT_{n-1} to $O(n-1)$ on the boundary.

If $E \to B$ is a fiber bundle with fiber $X \times I$ and structure group the CAT concordances, there is a strong deformation retraction from E to the boundary "$X \times \{1\} \times B$". The quotes indicate that "$X \times \{1\} \times B$" need not be a product

bundle. This enables us to define a fiber homotopy equivalence of triples,

$$(X \times I \times B, X \times \{0\} \times B, X \times \{1\} \times B) \to (E, X \times \{0\} \times B, \text{“}X \times \{1\} \times B\text{”})$$

This extends to a fiber-homotopy equivalence:

$$(X \times \mathbb{R} \times B, X \times (-\infty, 0] \times B, X \times [1, \infty) \times B) \to (E, X \times (-\infty, 0] \times B, \text{“}X \times \{1\} \times B\text{”}[1, \infty))$$

Applying Theorem 2.5 to this fibered homotopy equivalence gives a fibered equivalence of topological tangent bundles covering the map f above. Moreover, near $\pm\infty$ this equivalence is induced by an equivalence of topological tangent bundles of $X \times 1 \times B$ and “$X \times 1 \times B$”. Thus, fibered smoothing theory gives a homotopy section to the forgetful map $\mathscr{H}^{\mathrm{DIFF}}(X) \to \mathscr{H}^{\mathrm{TOP}}(X)$. By Essay V of [22], $\mathscr{H}^{\mathrm{TOP}}(X)$ is homotopy equivalent to $\mathscr{H}^{\mathrm{PL}}(X)$, so stabilizing this map and delooping gives the required homotopy section of $Wh^{\mathrm{DIFF}}(X) \to Wh^{\mathrm{PL}}(X)$. As before, it is convenient in our context to stabilize compact concordances (or noncompact concordances with compact support) by crossing with $\mathbb{R}$ and using compact support, rather than by crossing with I rel boundary.

One way to achieve the delooping of the section $\mathscr{H}^{\mathrm{DIFF}}(X) \to \mathscr{H}^{\mathrm{TOP}}(X)$ is to use the same technique applied to $X \times \mathbb{R}$, and use versions of $\mathscr{H}$ that are bounded in the $\mathbb{R}$- direction. It is then a general fact that $Wh^{\mathrm{CAT}}(X)$ is homotopy equivalent to $\mathscr{H}^{\mathrm{CAT,BDD}}(X)$.

4 Splitting diffeomorphism groups

Lemma 4.1 *For any complete nonpositively curved W^n with $n \geqq 5$,*
(i) *There is a map $s' : F^c(W)/\mathrm{Homeo}^c(W) \to B\mathrm{Diffeo}$ so that $\varrho \circ s' = j$ in the diagram below.*
(ii) *There is a map $s : \Omega(F^c(W)/\mathrm{Homeo}^c(W)) \to \Omega(F^c(W)/\mathrm{Diffeo}^c(W))$ so that $\Omega\varrho' \circ s = \mathrm{id}$.*

$$\begin{array}{ccccccc}
F^c(W) & \xrightarrow{i} & \dfrac{F^o(W)}{\mathrm{Homeo}^c(W)} & \xrightarrow{j} & B\,\mathrm{Homeo}^c(W) & \xrightarrow{k} & BF^c(W) \\
\| & & \varrho' \uparrow \downarrow s'' & \searrow^{s'} & \uparrow \varrho & & \| \\
F^c(W) & \xrightarrow{i'} & \dfrac{F^c(W)}{\mathrm{Diffeo}^c(W)} & \xrightarrow{j'} & B\,\mathrm{Diffeo}^c(W) & \xrightarrow{k'} & BF^c(W) \\
 & & \uparrow & & \uparrow & & \\
 & & \dfrac{\mathrm{Diffeo}^c(W)}{\mathrm{Homeo}^c(W)} & = & \dfrac{\mathrm{Diffeo}^c(W)}{\mathrm{Homeo}^c(W)} & &
\end{array}$$

Proof. We have already seen that any fiber-homotopically trivial topological W-bundle $E \to B$ has topological tangent bundle fiberwise homeomorphic over B to $\tau_W \times B$. Applying this to the universal bundle over $F^c(W)/\mathrm{Homeo}^c(W)$, we see that the map $j : F^c(W)/\mathrm{Homeo}^c(W) \to B\mathrm{Homeo}^c(W)$ has a lift s' to $B\mathrm{Diffeo}^c(W)$.

Since $k \circ s'$ is nullhomotopic, s' has a lift

$$s'' : F^c(W)/\mathrm{Homeo}^c(W) \to F^c(W)/\mathrm{Diffeo}^c(W).$$

Unfortunately, s'' need not be a section to ϱ'. However, we do have $j \circ (\varrho' \circ s'') = j$. If we loop the entire diagram, it follows that $\Omega j \circ (\mathrm{id} - \Omega(\varrho' \circ s''))$ is nullhomotopic, so there is a map $\alpha : \Omega(F^c(W)/\mathrm{Homeo}^c(W)) \to \Omega F^c(W)$ with $\Omega i \circ \alpha$ homotopic to $\mathrm{id} - \Omega(\varrho' \circ s'')$. It follows that $s = (\Omega i' \circ \alpha) + \Omega s''$ is a homotopy section to ϱ'.

Corollary 4.2 (i) *If $W^n, n \geqq 5$, is a complete noncompact Riemannian manifold of nonpositive curvature, then the forgetful map* $\mathrm{Diffeo}^c_\circ(W) \to \mathrm{Homeo}^c_\circ(W)$ *has a homotopy section. Here* $\mathrm{Diffeo}^c_\circ(W)$ *and* $\mathrm{Homeo}^c_\circ(W)$ *are the identity components of the groups of diffeomorphisms and homeomorphisms of M with compact support.* (ii) *If $W^n, n \geqq 5$, is a compact irreducible locally symmetric space of noncompact type (e.g. a compact hyperbolic manifold) then* $\mathrm{Diffeo}^c_\circ(M) \to \mathrm{Homeo}^c_\circ(M)$ *has a homotopy section.*

In addition, given $\varepsilon > 0$, these sections can be chosen to provide diffeomorphisms within ε of the original homeomorphisms.

Proof. In case (i), applying Lemma 4.1 is trivial because obstruction theory shows that $F^c_\circ(W)$ is contractible.

To see (ii), realize that under the above hypotheses $\pi_1(M)$ is centerless (see, e.g. [24] for a stronger theorem). Therefore, $F(M)$ is homotopically the discrete group $\mathrm{Out}(\pi_1(M))$ of outer automorphisms of $\pi_1(M)$.

Consequently, to apply the lemma, it suffices to show that $\mathrm{Diffeo}(M) \to \mathrm{Out}(\pi_1(M))$ is onto. This is an immediate consequence of Mostow's rigidity theorem, because each outer automorphism is realized geometrically by an isometry.

In each case, the existence of a controlled section follows from "concordance implies isotopy" as in [22, p.78].

Corollary 4.3 *If $M^n, n \geqq 5$, is a closed Riemannian manifold of nonpositive curvature, then $\pi_k(\mathrm{Homeo}(M))$ is a direct summand of $\pi_k(\mathrm{Diffeo}(M))$ for all $k \geqq 2$.*

Proof. This follows easily from Lemma 4.1 and the fact that $\pi_k(F(M)) = 0$ for $k \geqq 2$. We therefore have $\pi_k(\mathrm{Homeo}(M)) \cong \pi_{k+1}(F(M)/\mathrm{Homeo}(M))$ and $\pi_k(\mathrm{Diffeo}(M)) \cong \pi_{k+1}(F(M)/\mathrm{Diffeo}(M))$ for $k \geqq 2$. The result then follows, since the homotopy groups of $F(M)/\mathrm{Homeo}(M)$ are direct summands of the homotopy groups of $F(M)/\mathrm{Homeo}(M)$ by Lemma 4.1.

Remarks 4.4 (i) The problem in proving Corollary 4.3 in case $k = 1$ is that $\pi_1(F(M))$ is isomorphic to the center of $\pi_1(M)$. It seems conceivable that there are nontrivial elements of $\pi_1(F(M))$ which are represented by loops of homeomorphisms but not by loops of diffeomorphisms.

(ii) Farrell and Jones have shown, [14], that $\mathrm{Homeo}(M)$ has relatively little homotopy in the concordance-stable range. Little seems to be known about the homotopy groups of homeomorphism groups outside of this range.

Bibliography

1. Au, K.K.: Thesis, University of California at San Diego, 1990
2. Bökstedt, M., Hsiang, W.-C., Madsen, I.: The cyclotomic trace and algebraic K-theory of spaces. Århus (preprint)
3. Browder, W.: Surgery on simply connected manifolds. Berlin Heidelberg New York: Springer 1970

4. Burghelea, D., Lashof, R.: The homotopy type of the space of diffeomorphisms I. Trans. Am. Math. Soc. **196,** 1–36 (1974)
5. Černavskii, A.V.: Local contractibility of the homeomorphism group of a manifold. Sov. Math. Dok. **9,** 1171–1174 (1968)
6. Chapman, T.A.: Approximation results in topological manifolds. Mem. Am. Math. Soc. **251,** (1981)
7. Chapman, T.A., Ferry, S.C.: Approximating homotopy equivalences by homeomorphisms. Am. J. Math. **101,** 583–607 (1979)
8. Connes, A., Moscovici, H.: Cyclic cohomology, the Novikov conjecture, and hyperbolic groups. Topology **29,** 345–388 (1990)
9. Daverman, R.J.: Decompositions of manifolds. New York Academic Press, 1986
10. Edwards, R.D., Kirby, R.C.: Deformations of spaces of embeddings. Ann. Math. **93,** 63–88 (1971)
11. Eilenberg, S.: Sur les transformations a petites tranches. Fundam. Math. **30,** 92–95 (1938)
12. Farrell, F.T., Hsiang, W.-C.: On Novikov's Conjecture for nonpositively curved manifolds, I. Ann. Math. **113,** 197–209 (1981)
13. Farrell, F.T., Jones, L.E.: Rigidity and other aspects of compact nonpositively curved manifolds. Bull. Am. Math. Soc. **22,** 59–64 (1990)
14. Farrel, F.T., Jones, L.E.: Computations of stable pseudoisotopy spaces (preprint)
15. Ferry, S.C.: Homotoping ε-maps to homeomorphisms, Am. J. Math. **101,** 567–582 (1979)
16. Ferry, S.C., Pedersen, E.K.: Epsilon Surgery Theory, I (preprint)
17. Ferry, S.C., Rosenberg, J., Weinberger, S.: Equivariant topological rigidity phenomena, C. R. Acad. Sci. Paris Série I **306,** 777–782 (1988)
18. Hicks, N.: Notes on differential geometry. London, Van Nostrand Reinhold Co. 1971
19. Hirsch, M., Mazur, B.: Smoothings of Piecewise Linear manifolds. Princeton: Princeton University Press 1974
20. Hsiang, W.-C., Rees, H.: Miscenko's work on Novikov's Conjecture. Contemp. Math. **10** (1982)
21. Kasparov, G.: Equivariant KK-theory and the Novikov Conjecture. Invent. Math. **91,** 147–202 (1988)
22. Kirby, R.C., Siebenmann, L.C.: Foundational essays on topological manifolds, smoothings and triangulations. Princeton: Princeton University Press 1977
23. Kister, J.M.: Microbundles are fiber bundles. Ann. Math. **80,** 190–199 (1964)
24. Lawson, H.B., Yau, S.T.: Compact manifolds of nonpositive curvature. J. Differ. Geo. **7,** 211–228 (1972)
25. Madsen, I., Milgram, R.J.: The classifying spaces for surgery and cobordism of manifolds. Princeton: Princeton University Press
26. Novikov, S.P.: The algebraic construction and properties of hermetian analogues of K-theory for rings with involution from the point of view of Hamiltonian formalism. Some applications to differential topology and the theory of characteristic classes, I, II. Izv. Nauk SSSR ser. mat. **34,** 253–288, 478–500 (1970)
27. Quinn, F.: Ends of maps, III, J. Differ. Geom. **17,** 503–521 (1982)
28. Roe, J.: Exotic cohomology and index theory for complete Riemannian manifolds (preprint)
29. Siebenmann, L.C.: Approximating cellular maps by homeomorphisms. Topology **11,** 271–294 (1972)
30. Sullivan, D.: Geometric Topology Part I: Localization, Periodicity, and Galois Symmetry. MIT (preprint)
31. Waldhausen, F.: Algebraic K-theory of topological spaces. I. In: Proceedings of Symposia in Pure Mathematics. AMS, 1978, Volume 32 pp. 35–60
32. Waldhausen, F.: Algebraic k-theory of spaces, a manifold approach. In: Canadian Mathematical Society Conference Proceedings, 1982
33. Weinberger, S.: Aspects of the Novikov Conjecture. In: Contemp. Math. Am. Math. Soc., 1990, pp. 281–297

Invent. math. 105, 415–430 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system

P.L. Lions [1] **and B. Perthame** [2]

[1] CEREMADE, Université Paris IX-Dauphine, Place de Lattre de Tassigny, F-75775 Paris Cedex 16
[2] Département de Mathématiques, Université d'Orléans, BP 6759, F-45067 Orléans Cedex 2

Oblatum 14-I-1991

Summary. We prove that, if the initial data has moments in v higher than three, then the solution of Vlasov-Poisson has also moments in v higher than three. We deduce from this different regularity results on the local density, the force field or the solution itself. Also we give a new uniqueness result, and new regularity results for solutions satisfying only the energy and L^∞ bounds. Our proofs are based on a new representation formula and logarithmic estimates for the force field.

Résumé. Nous montrons que, si la donnée initiale possède des moments en v plus élevés que trois, alors la solution de l'Equation de Vlasov-Poisson a aussi des moments plus élevés que trois. Nous en déduisons différents résultats de régularité sur la densité locale, le champ de force ou la solution elle-même. Nous donnons également un nouveau résultat d'unicité et de nouveaux résultats de régularité pour les solutions vérifiant uniquement les estimations d'énergie et les bornes L^∞. Nos démonstrations sont fondés sur une nouvelle formule de représentation et des estimées logarithmiques sur le champ de force.

1 Introduction

In this paper, we will discuss several questions concerning the three dimensional Vlasov-Poisson system: one looks for $f(t, x, v) \geqq 0 (x \in \mathbb{R}^3, v \in \mathbb{R}^3, t \geqq 0)$ solution of

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + E \cdot \nabla_v f = 0 \tag{1}$$

$$f(0, x, v) = f_0(x, v)$$

with

$$E(t, x) = \nabla_x \Phi(t, x) \tag{2}$$

$$\Phi = \frac{1}{4\pi} \frac{1}{|x|} * (\sigma \rho), \qquad \rho(t, x) = \int f(t, x, v)\, dv. \tag{3}$$

The initial data for (1), $f_0(x, v)$, is a given nonnegative integrable and bounded function.

The system (1)–(3) appears in different physical models; in astrophysics it represents the evolution of the density of massive particles and then $\sigma = -1$, in plasma physics or semi-conductor devices it represents the evolution of charged particles and then $\sigma = +1$. In both cases $f(x, v, t)$ is the density of particles located at the point x, with velocity v. E is the force field corresponding to the potential Φ. Since we will always consider local densities $\rho(x, t)$ in $L^{6/5}(\mathbb{R}^3)$, (3) can be written equivalently

$$-\Delta_x \Phi = \rho, \qquad \Phi \to 0 \quad \text{if } |x| \to +\infty. \tag{4}$$

First, we are going to prove the propagation of moments higher than 3. More precisely, if $|v|^m f_0 \in L^1(\mathbb{R}^6)$ for all $m < m_0$, $m_0 > 3$, then we build a solution of (1)–(3) satisfying $|v|^m f(t, x, v) \in L^1(\mathbb{R}^6)$ for any $t \geqq 0$. In particular, for $m_0 > 6$, we deduce from Sobolev injections that $E \in L^\infty(\mathbb{R}^3 \times [0, T])$ for any $t > 0$, and, following Horst [14] f is a smooth function if $f_0(x, v)$ is smooth. This improves a recent result of Pfaffelmöser [19] who has shown independently of our work the existence of a classical solution when f_0 is, roughly speaking, Lipschitz continuous in the space $L^1_v(L^\infty_x)$. Let us also mention that the existence and uniqueness of classical solutions have been built by Iordanskii [16] in dimension 1, Ukai and Okabe [21] in dimension 2, Bardos-Degond [3] in dimensions 3 for small data. The case of (nearly) symmetric data has been treated by Batt [4], Wollman [22], Horst [12], J. Schaeffer [20] and the relativistic case has been treated by Glassey-Schaeffer [11].

The second result improves the known regularity of $\rho(x, t)$ and $E(x, t)$ for weak solutions satisfying only the natural energy estimate

$$\int |v|^2 f_0(x, v)\, dx\, dv < +\infty. \tag{5}$$

For instance, we obtain $E \in L^\infty([0, T]; L^p(\mathbb{R}^3))$ for any $p < 5$ while the classical result for the weak solutions built by Horst and Hunze [13] is $E \in L^\infty([0, T]; L^{15/4}(\mathbb{R}^3))$. We will also improve the local regularity results of Perthame [18]. We will use however the boundedness of f and thus we do not cover the weak solutions built by Di Perna-Lions [9, 10] which only satisfy the minimal regularity $f_0 \log f_0 \in L^1(\mathbb{R}^6)$. Finally we also would like to quote that weak solutions where previously built by Arsenev [1, 2], Illner-Neunzert [15].

The paper is organized as follows. In Sect. 2, we show the propagation of moments and we state the subsequent regularity results. In Sect. 3, we show a uniqueness result in the space of solutions such that $E \in L^\infty([0, T] \times \mathbb{R}^3)$, provided that f_0 is roughly speaking Lipschitz continuous in $L^\infty_x(L^1_v)$. Finally, Sect. 4 is devoted to the regularity of ρ and E for initial data satisfying (5).

The main results of this paper where announced in Lions-Perthame [17].

2 Propagation of moments

Our first result is

Theorem 1 *Let $f_0 \geqq 0$, $f_0 \in L^1 \cap L^\infty(\mathbb{R}^3 \times \mathbb{R}^3)$. We assume that*

$$\int_{\mathbb{R}^3 \times \mathbb{R}^3} |v|^m f_0(x, v)\, dx\, dv < +\infty \qquad \text{if } m < m_0 \tag{6}$$

where $3<m_0$. *Then, there exists a solution* $f\in C(\mathbb{R}^+; L^p(\mathbb{R}^3\times\mathbb{R}^3))\cap L^\infty(\mathbb{R}^+; L^\infty(\mathbb{R}_3\times\mathbb{R}^3))$ (*for all* $1\leqq p<+\infty$) *of* (1)–(3) *satisfying*

$$\sup_{t\in[0,T]}\int_{\mathbb{R}^3\times\mathbb{R}^3}|v|^m f(t,x,v)\,dx\,dv<+\infty, \tag{7}$$

for all $T<+\infty$, $m<m_0$.

Corollary 2 *The solution of Theorem 1 satisfies*

$$\rho\in C(\mathbb{R}^+; L^q(\mathbb{R}^3)) \quad \text{for} \quad 1\leqq q<\frac{3+m_0}{3} \tag{8}$$

$$E\in C(\mathbb{R}^+; L^q(\mathbb{R}^3)) \quad \text{for} \quad \frac{3}{2}<q<\frac{3(3+m_0)}{6-m_0} \quad \text{if } m_0\leqq 6. \tag{9}$$

$$E\in C(\mathbb{R}^+; C^{0,\alpha}(\mathbb{R}^3)) \quad \text{with} \quad \alpha<\frac{m_0-6}{3+m_0} \quad \text{if } m_0>6. \tag{9'}$$

Corollary 3 *Let* $m_0>6$ *in Theorem 1 and assume that* f_0 *satisfies*

$$\forall R>0, \quad \forall T>0,$$
$$\operatorname{supess}\{f_0(y+vt,w), |y-x|\leqq Rt^2, |v-w|\leqq Rt\}\in L^\infty((0,T)\times\mathbb{R}^3_x; L^1(\mathbb{R}^3_v)). \tag{10}$$

Then, the above solution satisfies

$$\forall T\in(0,\infty), \quad \rho\in L^\infty([0,T]\times\mathbb{R}^3_x). \tag{11}$$

Remark. These results can easily be extended to the Vlasov-Poisson system set in domains of $\mathbb{R}^N$ or involving different species of particles.

Before going into the proof of these results, let us point out that the assumption (6) implies that $E(x,0)$ belongs to $L^2(\mathbb{R}^3)$ and thus for $t=0$

$$\mathscr{E}(t)=\int_{\mathbb{R}^3\times\mathbb{R}^3}|v|^2 f(t)\,dx\,dv+\sigma\int_{\mathbb{R}^3}|E(t)|^2\,dx<+\infty. \tag{12}$$

Therefore the solutions we build have finite energy i.e. (12) holds for any $t>0$, as in the usual theory of weak solutions (see [13] or the review papers by Batt [5, 6] or Degond [7]). In particular, (1) holds in $\mathscr{D}'$ writing

$$E\cdot\nabla_x f=\operatorname{div}_v(Ef),$$

and one has

$$\int_{\mathbb{R}^3\times\mathbb{R}^3}\beta(f(t))\,dx\,dv=\int_{\mathbb{R}^3\times\mathbb{R}^3}\beta(f_0)\,dx\,dv \quad \forall t\geqq 0, \tag{13}$$

for any nonnegative continuous function β. This is a consequence of the theory of renormalized solutions of Di Perna-Lions [8, 9]; it also shows that f is constant along the characteristics which are defined almost everywhere. Note that

the regularity of the force field E in Corollary 2 is not enough to prove the existence of classical solutions of the characteristic system

$$\dot{X}=V, \quad \dot{V}=E(t, X(t)).$$

For $m_0>6$, the Corollary 2 shows in particular that $E\in L^\infty([0, T]\times\mathbb{R}^3)$ for $T>0$. By a classical result (see [14]), this implies that initially smooth data remain smooth (say C^∞ with compact support) for all times. Then, the solution is unique. In fact, we will give a uniqueness result in a larger class in Sect. 3.

We begin with the proof of the corollaries.

Proof of Corollary 2. (8) is classically deduced from the "interpolation" inequality

$$\rho(t, x)\leqq \int_{\mathbb{R}^3} |v|^k f(t)\,dv/R^k+R^3\|f\|_\infty, \quad \forall R.$$

By choosing $R(x, t)$ so as to minimize the r.h.s. of this inequality we find more precisely

$$\|\rho(t)\|_{L^{\frac{3+k}{3}}(\mathbb{R}^3)}\leqq C\Big(\int_{\mathbb{R}^3\times\mathbb{R}^3} |v|^k f(t)\,dv\,dx\Big)^{\frac{3}{3+k}}, \quad \forall k\geqq 0, \tag{14}$$

for some constant C depending only on $\|f\|_\infty$. (14) clearly gives (8). And (9) and (9′) are the usual regularity results for E given through (2)–(3). Let us also mention that we have

$$\|E(t)\|_q\leqq C\|\rho(t)\|_{\frac{m+3}{3}}\leqq C\Big(\int_{\mathbb{R}^3\times\mathbb{R}^3} |v|^m f(t)\,dx\,dv\Big)^{\frac{3}{3+m}} \tag{15}$$

with

$$q=\frac{3(3+m)}{6-m} \quad \text{if } m<6, \qquad q=\infty \quad \text{if } m>6. \tag{16}$$

Proof of Corollary 3. Since the Theorem 1 gives in fact a force field $E\in L^\infty([0, T]; W^{1,q}(\mathbb{R}^3))$, for all $T>0$, and for all q, $1<q<+\infty$, we may use the weak characteristics built in [8]. Thus for a.e. (x, v) we have

$$f(t, x, v)=f_0(X(0), V(0)), \tag{17}$$

where (X, V) solves

$$\dot{X}(s)=V(s), \quad \dot{V}(s)=E(s, X(s)), \quad X(t)=x, \quad V(t)=v. \tag{18}$$

Since E is bounded on $[0, T]\times\mathbb{R}^3$ by the Corollary 2, we have for $s\in[0, T]$

$$|V(s)-v|\leqq Rt, \quad |X(0)-x-vt|\leqq Rt^2. \tag{19}$$

(17) and (19) together yield

$$f(t, x, v)\leqq \sup\{f_0(y+vt, w);\ |y-x|\leqq Rt^2, |w-v|\leqq Rt\}. \tag{20}$$

The assumption (10) just asserts that the r.h.s. of (20) belongs to $L^\infty(\mathbb{R}^3_x; L^1(\mathbb{R}^3_v))$ for all $R>0$, $t\in(0, T)$ and thus $\rho(t, x)$ is bounded.

Remark. Of course, many variants of Corollary 2 are possible. For instance, if we replace (10) by

$$\forall R\in(0,\infty) \quad \sup\{f(y,v);|y-x|\leqq R\}\in L^1(\mathbb{R}^3_x;L^\infty(\mathbb{R}^3_v)\} \tag{21}$$

we get the weaker conclusion

$$\forall t\in(0,T), \quad t^3\rho(t,x)\in L^\infty([0,T]\times\mathbb{R}^3_x). \tag{22}$$

Indeed, we now deduce from (17)–(19)

$$\rho(t,x)\leqq\int_{\mathbb{R}^3}\sup\{f_0(x+vt+z,w);w\in\mathbb{R}^3,|z|\leqq R\}\,dv \tag{23}$$

or

$$\rho(t,x)\leqq\frac{1}{t^3}\int_{\mathbb{R}^3}\sup\{f_0(y+z,w);w\in\mathbb{R}^3,|z|\leqq R\}\,dy \tag{24}$$

and one concludes using (21).

We now turn to the proof of Theorem 1.

Proof of Theorem 1. In order to separate the different estimates and difficulties, we divide the proof in four steps. First of all, we prove some general estimates on E. Then, we conclude assuming that the time interval $(0,T)$ on which we solve the Vlasov-Poisson system is small enough and that E is bounded in $L^{3/2}(\mathbb{R}^3_x)$, in fact which is false in general but holds true for bounded domains of $\mathbb{R}^3_x$. We relax this assumption on E in a fourth and final step.

Also, to simplify the exposition, we only prove Theorem 1 assuming we deal with smooth solutions (say C^∞ with compact support) and we obtain a priori estimates which depend only on $\iint|v|^m f_0\,dx\,dv$, $\|f_0\|_\infty$. This is in fact enough by a standard approximation method (see [5, 6]) involving a straightforward regularization of the Coulomb potential: indeed, it will be clear from the proof below that the estimates are uniform for the approximated systems. We thus have to pass to the limit keeping the desired estimates by standard arguments.

First step. General estimates. Since we deal with classical solutions, we may treat $E\cdot\nabla_v f$ as a source term in (1), and thus

$$\begin{aligned} f(t,x,v) &= -\int_0^t(\operatorname{div}_v Ef)(t-s,x-vs,v)\,ds+f_0(x-vt,v)\\ &= -\int_0^t\operatorname{div}_v[Ef(t-s,x-vs,v)]\,ds\\ &\quad -\int_0^t s\operatorname{div}_x[Ef(t-s,x-vs,v)]\,ds+f_0(x-vt,v). \end{aligned}$$

Integrating with respect to v then gives

$$\rho(t,x) = -\operatorname{div}_x \int_0^t s \int_{\mathbb{R}^3} [Ef(t-s, x-vs, v)]\, dv\, ds + \int_{\mathbb{R}^3} f_0(x-vt, v)\, dv \tag{25}$$

and we denote

$$h(t,x) = \int_{\mathbb{R}^3} f_0(x-vt, v)\, dv. \tag{26}$$

Following the proof of Corollary 2, we have

$$\|h(t)\|_{\frac{m+3}{3}} \leqq C\Big(\int_{\mathbb{R}^3 \times \mathbb{R}^3} |v|^m f_0\, dx\, dv\Big)^{\frac{3}{m-3}}, \qquad \forall m < m_0. \tag{27}$$

We deduce from (25) that

$$\|E(t)\|_p \leqq C \left\| \int_0^t s \int_{\mathbb{R}^3} Ef(t-s, x-vs, v)\, dv\, ds \right\|_p + \|h\|_r$$

with

$$\frac{1}{p} = \frac{1}{r} - \frac{1}{3} \quad \text{and} \quad \frac{3}{2} < p < +\infty.$$

We will use this estimate for $p = m+3$ which gives $r = 3(3+m)/(6+m)$. Since $m \geqq 3$, we get

$$\|h\|_r \leqq C\, \|h(t)\|_{\frac{m+3}{3}} \leqq C\Big(\int_{\mathbb{R}^3 \times \mathbb{R}^3} |v|^m f_0\, dx\, dv\Big)^{\frac{3}{m+3}} \leqq C$$

for some constants C which only depend on the initial data. We finally obtain

$$\|E(t)\|_{m+3} \leqq C + C \left\| \int_0^t s \int_{\mathbb{R}^3} Ef(t-s, x-vs, v)\, dv\, ds \right\|_{m+3}. \tag{28}$$

On the other hand, we deduce from (1) the estimates on the moments

$$\begin{aligned} \left| \frac{d}{dt} \int_{\mathbb{R}^3 \times \mathbb{R}^3} |v|^k f(t)\, dx\, dv \right| &\leqq \int_{\mathbb{R}^3 \times \mathbb{R}^3} k\, |E(t)|\, f(t)\, |v|^{k-1}\, dx\, dv \\ &\leqq C\, \|E(t)\|_{3+k} \Big\| \int_{\mathbb{R}^3} |v|^{k-1} f(t)\, dv \Big\|_{(3+k)/(2+k)} \\ &\leqq C\, \|E(t)\|_{3+k} \Big(\int_{\mathbb{R}^3 \times \mathbb{R}^3} |v|^k f(t)\, dv\, dx \Big)^{\frac{k+2}{k+3}}, \end{aligned}$$

this last inequality is obtained thanks to the "interpolation inequality" used to prove (14). Now we set

$$M_k(t) = \sup_{0 \leqq s \leqq t} \int |v|^k f(s)\, dx\, dv,$$

and we find

$$\frac{d}{dt} M_k(t) \leqq \left| \frac{d}{dt} \int |v|^k f(t)\, dx\, dv \right| \leqq C \|E(t)\|_{3+k} M_k(t)^{\frac{k+2}{k+3}}, \tag{29}$$

which in turn, implies

$$M_k(t) \leqq C\{M_k(0) + (t \sup_{s\in(0,t)} \|E(s)\|_{3+k})^{k+3}\}. \tag{30}$$

Second step. Small time estimates. We first show that for any m, $3<m<m_0$, $M_m(t)$ and $M_{m+\alpha}(t)$ are bounded on a small time interval $(0, t_0)$ for some $\alpha>0$. To do so, we estimate from above the integral term of (28) (integrated on $(0, t_0)$ rather than $(0, t)$) for any $r>3/2$ and $t_0 \leqq t \leqq T$, by

$$\left\| \int_0^{t_0} s\, ds \sup_{\tau\in(0,T)} \left(\int_{\mathbb{R}^3} |E|^r(\tau, y) \frac{dy}{s^3} \right)^{1/r} \left(\int_{\mathbb{R}^3} f(t-s, x-vs, v)\, dv \right)^{1/r'} \right\|_{m+3} \|f\|_\infty^{(r'-1)/r'} \leqq C t_0^{2-3/r} \sup_{\tau\in(0,T)} \|E(\tau)\|_r \sup_{s,\tau\in(0,t)} \left\| \int_{\mathbb{R}^3} f(\tau, x-vs, v)\, dv \right\|_{\frac{m+3}{r'}}^{1/r'} \tag{31}$$

where r' is the conjugate exponent of r, $1/r+1/r'=1$, $1 \leqq r' < 3$ and C only depends upon the initial data. Since we know from the classical theory of Vlasov-Poisson Equation that $M_2(t)$ is bounded on any interval $(0, T)$, we deduce from (14) and (16) that $E(\tau)$ is bounded in any L^r, $\frac{3}{2}<r\leqq\frac{15}{4}$. Here we will choose r close enough to $3/2$ to be sure that $r \leqq 15/4$. Therefore, using (14) – which also holds for $f := f(\tau, x-vs, v)$ – with the value k such that $\dfrac{k+3}{3} = \dfrac{(m+3)}{r'}$ we estimate (31) from above by

$$t_0^{2-3/r} C(r)(M_k(t))^{\frac{1}{m+3}} \leqq t_0^{2-3/r} C(r)(M_k(0) + \sup_{s\in(0,t)} \|E(s)\|_{3+k}^{k+3})^{\frac{1}{m+3}} \tag{32}$$

thanks to (30). Notice that $k>m$ because $r'<3$, but k may be chosen as close to m as we wish and thus satisfying $m_0>k>m$. Also since the value $m>3$ in (16), leads to a value $q>(m+3)$, we may again choose r close enough to $3/2$ so that $3+k \leqq q$. Actually the limitation $m_0>3$ in Theorem 1 just comes from that requirement. Finally, for such a choice of r (and k) we find by (16)

$$\|E(s)\|_{3+k} \leqq C(M_m(s))^{\frac{3}{3+m}}$$

and (31) is bounded by

$$C t_0^\gamma (1 + M_m(t)^\delta) \tag{33}$$

for some $\delta, \gamma > 0$; $\gamma = 2-3/r$ and $\delta = 3(k+3)/(m+3)^2$, $m_0>k>m$.

Coming back to (28), and choosing $t_0 = t$, we have obtained

$$\|E(t)\|_{m+3} \leqq C(1 + t^\gamma M_m(t))^\delta.$$

This inequality combined with (29) for $k=m$, just gives

$$\frac{d}{dt} M_m(t) \leqq C(M_m(t)^{\frac{m+2}{m+3}} + t^\gamma M_m(t)^{\delta + \frac{m+2}{m+3}}),$$

and thus $M_m(t)$ is bounded on a small time interval $[0, t_0]$. Then, (16) shows that $|E(t)|_{L^q}$ is bounded on $[0, t_0]$ for $\frac{3}{2} < q \leqq q_1$ where q_1 is larger than $3+m$ (since $m>3$). Finally (30) shows that $M_k(t)$ is bounded on $[0, t_0]$ for the values of k such that $0 \leqq k \leqq q_1 - 3 > m$.

For the end of the proof, only (33) will be used. However we have pointed out the possibility of proving small time estimates because it could be useful elsewhere.

Third step. The case when $E \in L^{3/2}$. Let us assume in this part of the proof that $E \in L^\infty((0, T); L^{3/2}(\mathbb{R}^3))$. Of course this does not hold in general, only the weaker regularity $L^\infty((0, T); L^{3/2,\infty}(\mathbb{R}^3))$ is true in $\mathbb{R}^3$. In a bounded domain however the $L^{3/2}$ regularity would hold and we will show in the fourth step how to relax the assumption later on. Then we may estimate $\|E(t)\|_{m+3}$ in (28) by

$$\|E(t)\|_{m+3} \leqq C + C \left\| \int_0^{t_0} \ldots \right\|_{m+3} + C \left\| \int_{t_0}^{t} \ldots \right\|_{m+3},$$

and the first integral term is bounded using the bound (33) of the second step on. Therefore, we may again bound $\|E(t)\|_{m+3}$ by

$$\begin{aligned} \|E(t)\|_{m+3} &\leqq C + C t_0^\gamma M_m(t)^\delta + C \left\| \int_{t_0}^{t} s\, ds \sup_{\tau \in (0,T)} \left(\int_{\mathbb{R}^3} |E|_{3/2}(\tau, y) \frac{dy}{s^3} \right)^{2/3} \right. \\ &\quad \left. \cdot \left(\int_{\mathbb{R}^3} f(t-s, x-vs, v)\, dv \right)^{1/3} \right\|_{m+3} \|f\|_\infty^{\frac{2}{3}} \\ &\leqq C + C t_0^\gamma M_m(t)^\delta + C |\log t_0| \sup_{\tau \in (0,T)} \|E(\tau)\|_{3/2} \\ &\quad \cdot \sup_{s, \tau \in (0,t)} \| \int_{\mathbb{R}^3} f(\tau, x-vs, v)\, dv \|_{m+3}^{\frac{1}{3}} \end{aligned}$$

and, using (14), this is finally bounded by

$$\|E(t)\|_{m+3} \leqq C + C t_0^\gamma M_m(t)^\delta + C |\log t_0| \sup_{\tau \in (0,T)} \|E(\tau)\|_{3/2}\, M_m(t)^{\frac{1}{m+3}}. \tag{34}$$

(34) combined with (29) just yields, choosing $t_0^\gamma = M_m(t)^{1-\delta}$ (which is allowed when $M_m(t)$ is large, the only case we have to consider)

$$\frac{d}{dt} M_m(t) \leqq C(1 + M_m(t) |\log M_m(t)|) \tag{35}$$

and Gronwall lemma again allows to conclude that $M_m(t)$ is bounded on any interval $(0, T)$.

Fourth step. The general case. When E is not bounded in $L^{3/2}$, we modify the above strategy in decomposing the Coulomb potential in two parts

$$E = E_1 + F \tag{36}$$

$$E_1 = \frac{\alpha}{4\pi}\left(\chi_R(x)\, \nabla \frac{1}{|x|}\right) * \rho$$

where $0 \leqq \chi_R \leqq 1$ is a smooth function such that $\chi_R(x) = 1$ if $|x| \leqq R$ and $\chi_R(x) = 0$ if $|x| \geqq 2R$. Since $\chi_R(x)\, \nabla \frac{1}{|x|}$ is bounded in $L^1 \cap L^{3/2}$ we have

$$\|E_1(t)\|_{L^{3/2} \cap L^{15/4}} \leqq C(R), \tag{37}$$

and the steps 2 and 3 can be applied to E_1. On the other hand, we have

$$\|F\|_{L^\infty},\ \|\nabla F\|_{L^\infty},\ \|D^2 F\|_{L^\infty} \leqq C/R. \tag{38}$$

We now write the Vlasov Equation as

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + F \cdot \nabla_v f = -E_1 \cdot \nabla_v f, \tag{39}$$

and we are going to show that the characteristic field associated with (39) allows the same estimates of (28) than the field $(x, v) \mapsto (x - vs, v)$. This characteristic field is given by

$$\dot{X}(s) = -V(s), \qquad X(0) = x,$$

$$\dot{V}(s) = -F(t-s, X(s)), \qquad V(0) = v, \quad 0 \leqq s \leqq T,$$

and we have

$$f(t, x, v) = \int_0^t \nabla_V (E_1 f)(t-s, X(s), V(s))\, ds + f_0(X(t), V(t))$$

(we omit the dependence of (X, V) upon (t, x, v)). At this point we notice that the application $(x, v) \mapsto (X(s), V(s))$ is invertible (its Jacobian is 1) and, thanks to (38), its second derivative is bounded as well as that of its inverse: $(y, w)(s)$ depending on $(X, V)(x)$ for a given t.

We get

$$\rho(t, x) = \int_0^t \int_{\mathbb{R}^3} \left\{ \nabla_x [E_1 f(t-s, X), V(s))] \frac{\partial y}{\partial V} + \nabla_v [E_1 f(t-s, X(s), V(s))] \frac{\partial w}{\partial V} \right\} dv\, ds$$

$$+ \int_{\mathbb{R}^3} f_0(X(t), V(t))\, dv$$

$$= \int_0^t \int_{\mathbb{R}^3} \left\{ \nabla_x \left[E_1 f(\ldots) \frac{\partial y}{\partial V} \right] - E_1 f(\ldots) \frac{\partial}{\partial x}\left(\frac{\partial y}{\partial V}\right) \right.$$

$$\left. - E_1 f(\ldots) \frac{\partial}{\partial v}\left(\frac{\partial w}{\partial V}\right) \right\} dv\, ds + \int_{\mathbb{R}^3} f_0(\ldots)\, dv.$$

We deduce from this expression a formula close to (28) and it is clear that the most difficult term to treat is $\nabla_x[E_1 F \dots$. This yields, as in establishing (28),

$$\|E(t)\|_{m+3} \leqq C + C\left\| \int_0^t \int_{\mathbb{R}^3} \left(\frac{\partial y}{\partial V}\right) E_1 f(t-s, X(s), V(s))\, ds\, dv \right\|_{m+3}. \tag{28'}$$

We have to prove that the same estimates as in steps 2 and 3 hold for (28').

First of all, since $\left(\frac{\partial y}{\partial V}\right)$ is lipschitz continuous, and $\left(\frac{\partial y}{\partial V}\right)(s=0)=0$, we have

$$\left|\left(\frac{\partial y}{\partial V}\right)(s)\right| \leqq Cs. \tag{40}$$

Then, following the step 2, we compute

$$\int_{\mathbb{R}^3} |E_1(t-s, X(s))|^r\, dv = \int_{\mathbb{R}^3} |E_1(t-s, X)|^r J(x, v, t, s)\, dX,$$

where for (t, s, x) fixed

$$J = \left|\det \frac{\partial X}{\partial v}\right|^{-1}.$$

The matrices $\frac{\partial X}{\partial v}, \frac{\partial V}{\partial v}$ satisfy

$$\frac{\partial \dot{X}}{\partial v}(s) = \frac{\partial V}{\partial v}(s), \qquad \frac{\partial \dot{V}}{\partial v}(s) = DF(t-s, X)\frac{\partial X}{\partial V}(s),$$

$$\frac{\partial X}{\partial v}(0) = 0, \qquad \frac{\partial V}{\partial v} = \mathrm{Id},$$

and thus, choosing R large enough, (38) allows us to obtain $\frac{\partial V}{\partial v}(s)$ as close as we wish to Id, uniformly in x, v, t, s for $0 \leqq s, t \leqq T$. Such a choice of R leads to

$$J \leqq C/s^3 \quad \text{for } x, v \in \mathbb{R}^3,\ 0 \leqq s, t \leqq T. \tag{41}$$

This gives exactly the corresponding estimate in step 2. Then, we just notice that, as in (14),

$$\left\| \int_{\mathbb{R}^3} f(\tau, X, V)\, dv \right\|_{\frac{3+\rho}{3}} \leqq C\left(\int_{\mathbb{R}^6} |v|^\rho f(\tau, X, V)\, dx\, dv \right)^{\frac{3}{3+\rho}}$$

and this can also be bounded by

$$C\left(\int_{\mathbb{R}^6} (1+|V|\rho) f(\tau, X, V)\, dv\, dx \right)^{\frac{3}{3+\rho}} \leqq C(1 + M_\rho(\tau)^{\frac{3}{3+\rho}})$$

because, $|V(s)-v|$ is uniformly bounded. This concludes the corresponding estimates of step 2 and the same arguments also apply to step 3.

This shows that the fundamental inequality (35) holds for the modified transport equation and thus we have proved Theorem 1.

3 Uniqueness

In this section we are going to prove the uniqueness of the solutions to Vlasov-Poisson Equation such that the force field E is bounded (L^∞). We will need some extra regularity results that require further assumptions on f_0

$$\forall R, T>0, \quad \sup\{|\nabla f_0(y+vt, w)|; |y-x|\leqq R, |w-v|\leqq R\} \in L^\infty((0, T)\times \mathbb{R}^3_x; L^1(\mathbb{R}^3_v)\cap L^2(\mathbb{R}^3_v)), \tag{42}$$

where ∇f_0 refers to the x and v gradients of f_0. Notice that the L^2 bound of (42) is satisfied if the L^1 bound holds and f_0 is lipschitz continuous in x and v. Our argument relies on the representation formula (28) and on the following regularity results.

Lemma 4 *If $\rho(x, t)\in L^\infty((0, T); L^\infty(\mathbb{R}^3_x)\cap L^1(\mathbb{R}^3_x))$, there is $\alpha>0$ ($\alpha=\varepsilon^{-CT}$ for some $C>0$) such that the characteristic curves*

$$\dot{X}_i(s)=V_i(s), \quad \dot{V}_i(s)=E(X_i(s), s), \qquad X_i(t)=x_i, \quad V_i(t)=v_i, \quad i=1 \text{ or } 2 \tag{43}$$

are uniquely defined and satisfy

$$|X_1(s)-X_2(s)|+|V_1(s)-V_2(s)|\leqq C(|x_1-x_2|^\alpha+|v_1-v_2|^\alpha) \quad \forall t, s\in[0, T], \tag{44}$$

whenever E is given by ρ through Poisson Equation (2)–(3).

Corollary 5 *With the assumption* (42), *any solution to Vlasov-Poisson Equation such that ρ belongs to $L^\infty((0, T); L^\infty(\mathbb{R}^3_x)\cap L^1(\mathbb{R}^3_x))$ satisfies $\rho\in L^\infty((0, T); C^{0,1}(\mathbb{R}^3_x))$ and*

$$E\in L^\infty((0, T); C^{1,\beta}(\mathbb{R}^3_x)) \quad \text{for all } \beta<1. \tag{45}$$

Theorem 6 *We make the assumptions of Theorem* 1, (10), (42), *then the solution of Vlasov-Poisson Equation such that $\rho\in L^\infty((0, T)\times\mathbb{R}^3_x)$ is unique.*

Remarks. 1. The boundedness of ρ in L^∞ implies, by Corollary 5, that the solutions of Vlasov-Poisson Equation are smooth. Thus Theorem 6 applies to classical solutions.

2. Notice that (42) and $\rho\in L^\infty((0, T)\times\mathbb{R}^3_x)$ implies (10) and the assumptions of Theorem 6 could be slightly improved.

3. Of course the difficulty in Lemma 3 is that E is not lipschitz continuous in x.

We now turn to the proof of these results.

Proof of Lemma 4. In order to prove Lemma 4, we just need to prove

$$|E(x,t)-E(y,t)| \leqq C|x-y| \log \frac{1}{|x-y|}, \tag{46}$$

say for $|x-y| \leqq \frac{1}{2}$. Indeed, the existence of solutions to (43) just follows from the continuity of E in x. And the estimate follows from a Gronwall type argument on $Y_i=(X_i, V_i)$; we have for $0 \leqq s \leqq t \leqq T$

$$\frac{d}{ds}|Y_1-Y_2| \leqq C|Y_1-Y_2| \log(|Y_1-Y_2|)^{-1}$$

And thus, as long as $|Y_1-Y_2|(s)<1/2$, we have

$$\log|Y_1-Y_2|(s) \leqq \log|Y_1-Y_2|(t)\, e^{-CT}$$

which is enough to conclude that (45) holds.

To prove (46), we use the same decomposition of E as in (36), $E=E_\varepsilon+F_\varepsilon+E_2$ with

$$E_\varepsilon=\frac{\alpha}{4\pi}\left(\nabla \frac{1}{|x|}\chi_{\{|x| \leqq \varepsilon\}}\right)*\rho,$$

$$F_\varepsilon=\frac{\alpha}{4\pi}\left(\nabla \frac{1}{|x|}\chi_{\{\varepsilon \leqq |x| \leqq 1\}}\right)*\rho,$$

$$E_2=\frac{\alpha}{4\pi}\left(\nabla \frac{\alpha}{4\pi}\left(\nabla \frac{1}{|x|}\chi_{\{|x| \geqq 1\}}\right)*\rho\right).$$

For $|x| \geqq 1$, $\nabla \dfrac{1}{|x|}$ is lipschitz continuous and thus

$$|E_2(x,t)-E_2(y,t)| \leqq C|x-y| \int_{\mathbb{R}^3} \rho(x,t)\,dx \qquad \forall t \leqq T.$$

We also have

$$\begin{aligned}|F_\varepsilon(x,t)-F_\varepsilon(y,t)| &\leqq C|x-y|\left(\frac{1}{|z|^3}1_{\{\varepsilon \leqq |z| \leqq 1\}}\right)*\rho \\ &\leqq C|\log\varepsilon| \sup_z \rho(z,t)\,|x-y|\end{aligned}$$

because

$$\int_{t \leqq |z| \leqq 1} \frac{1}{|z|^3}\,dz = 2\pi|\log\varepsilon|.$$

Finally

$$|E_\varepsilon(x,t)-E_\varepsilon(y,t)| \leqq |E_\varepsilon(x,t)|+|E_\varepsilon(y,t)| \leqq C\varepsilon \sup_z \rho(z,t).$$

Choosing $\varepsilon=|\log|z-y||$ in these inequalities we find (46) and Lemma 4 is proved.

Proof of Corollary 5. Thanks to the representation formula of f with the characteristic curves we find, for $|x_1-x_2|\leqq 1/2$,

$$|\rho(x_1,t)-\rho(x_2,t)|\leqq \int_{\mathbb{R}^3_v} |f_0(X_1(0),V_1(0))-f_0(X_2(0),V_2(0))|\,dv$$

where (X_i, V_i) satisfies (43) with $v_1=v_2=v$. Since E is bounded in L^∞, we have, following (19),

$$|X_i(0)-vt-x_i|\leqq R, \qquad |V_i(0)-v|\leqq R,$$

and thus we may estimate the above terms by

$$\int_{\mathbb{R}^3_v} \sup\{|\nabla f_0|(y+vt,w);|y-x_1|\leqq R,|w-v|\leqq R\}\,dv$$
$$\cdot \sup_{v\in\mathbb{R}^3}\{|X_1(0)-X_2(0)|+|V_1(0)-V_2(0)|\}\leqq C(R,T)\,|x_1-x_2|^\alpha$$

thanks to (42) and (44). This shows that $\rho(\cdot,t)$ is Hölder continuous and the $C^{1,\alpha}$ regularity of $E(\cdot,t)$ follows from Schauder estimates. Then, we obtain that the characteristic curves are Lipschitz continuous (we may take $\alpha=1$ in (45)), and thus $E(\cdot,t)$ belongs to $C^{1,\beta}(\mathbb{R}^3)$ for all $\beta<1$.

We may now conclude the uniqueness proof.

Proof of Theorem 6. First, let us notice that an elementary modification of the proof of Corollary 5 gives, thanks to the L^2 bound in (42),

$$\nabla_v f(t,x,v)\in L^\infty((0,T)\times\mathbb{R}^3_x;L^2(\mathbb{R}^3_v)), \tag{48}$$

for any solution f. Secondly, we set

$$D(t)=\sup_{0\leqq s\leqq t}\|(f_1-f_2)\|_{L^2(\mathbb{R}^6)},$$

for two possible solutions f_1, f_2 of (1)–(3) and we claim that

$$\frac{d}{dt}D(t)\leqq C(T)\,\|(E_1-E_2)(t)\|_{L^2(\mathbb{R}^3)}. \tag{49}$$

Indeed Vlasov equations give

$$\frac{\partial}{\partial t}(f_1-f_2)^2+v\cdot\nabla_x(f_1-f_2)^2+E_1\cdot\nabla_v(f_1-f_2)^2\leqq 2\,|E_2-E_1|\cdot|f_1-f_2|\cdot|\nabla_v f_2|$$

and thus, using (48),

$$\frac{d}{dt}D(t)^2\leqq 2\int_{\mathbb{R}^3_x}|(E_2-E_1)(t)|\,\|(f_1-f_1)(t,x,v)\|_{L^2(\mathbb{R}^3_v)}\|\nabla_v f_2(t,x,v)\|_{L^2(\mathbb{R}^3_v)}\,dx$$
$$\leqq C(T)\,\|(E_1-E_2)(t)\|_{L^2(\mathbb{R}^3_x)}D(t),$$

which clearly proves (49). Finally, we use formula (28) which gives

$$\|(E_1-E_2)(t)\|_{L^2(\mathbb{R}^3_x)}$$
$$\leqq C\left\|\int_0^t\int_{\mathbb{R}^3}(E_1-E_2)(x-vs,t-s)\,f_1(t-s,x-vs,v)s\,ds\,dv\right\|_{L^2(\mathbb{R}^3_x)}$$
$$+C\left\|\int_0^t\int_{\mathbb{R}^3}E_2(x-vs,t-s)(f_2-f_1)(t-s,x-vs,v)s\,ds\,dv\right\|_{L^2(\mathbb{R}^3_x)}.$$

The first term in the r.h.s. of this inequality may be estimated by

$$\left\|\int_0^t\frac{s\,ds}{s^{3/2}}\|(E_1-E_2)(y,t-s)\|_{L^2(\mathbb{R}^3_y)}\Big(\int_{\mathbb{R}^3_v}f_1^2(t-s,x-vs,v)dv\Big)^{1/2}\right\|_{L^2(\mathbb{R}^3_x)}$$
$$\leqq\int_0^t\frac{ds}{s^{1/2}}\|f_1(t-s)\|_{L^2(\mathbb{R}^6)}\|(E_1-E_2)(t-s)\|_{L^2(\mathbb{R}^3)}$$
$$\leqq Ct^{1/2}\sup_{s\leqq t}\|(E_1-E_2)(s)\|_{L^2(\mathbb{R}^3)}.$$

Processing the other term in the same way yields

$$\|(E_2-E_1)(t)\|_{L^2(\mathbb{R}^3)}\leqq\tfrac{1}{2}\sup_{0\leqq s\leqq t}|(E_2-E_1)(s)\|_{L^2(\mathbb{R}^3)}+C(T)\,D(t),$$

for $t\leqq t_0$ small enough. From this one easily deduces that

$$\|(E_2-E_1)(t)\|_{L^2(\mathbb{R}^3)}\leqq C(T)\,D(t),$$

and combining this inequality with (49) just shows that $D(t)=0$ by a Gronwall argument. Therefore $f_1=f_2$ for $t\leqq t_0$ and Theorem 6 is proved.

4 Regularity results using the energy and L^∞ bounds. Open questions

If we just want to use the natural energy and L^∞ bounds on f_0

$$\int_{\mathbb{R}^3\times\mathbb{R}^3}|v|^2f_0\,dx\,dv+\int_{\mathbb{R}^3}|E|^2(0)\,dx<+\infty,\qquad\|f_0\|_\infty<+\infty,\tag{50}$$

then, Theorem 1 cannot be applied. We find the usual ([13]) global bounds

$$\|\rho(t)\|_p<+\infty\quad\forall t>0,\qquad 1\leqq p\leqq 5/3,$$
$$\|E(t)_p<+\infty\quad\forall t>0,\qquad 3/2<p\leqq 15/4.\tag{51}$$

An improvement of this regularity was given in [18], but the new estimates were only local. The method of proof of Theorem 1 however allows to improve these results. We find

Theorem 7 *Weak solutions of the Vlasov-Poisson system* (1)–(3), *with an initial data satisfying only the energy and* L^∞ *estimate* (50), *have the following regularity*

$$\|E(t)\|_p < +\infty, \quad \forall t > 0 \qquad \text{for } \tfrac{3}{2} < p < 5, \tag{52}$$

$$\int_0^T \int_{B_R \times \mathbb{R}^3_v} |v|^m f \, dx \, dv < +\infty, \quad \forall T, R > 0 \qquad \text{for } 0 < m < \tfrac{5}{2}, \tag{53}$$

$$\int_0^T \|\rho(t)\|_{L^p(B_R)} dt < +\infty, \quad \forall T, R > 0 \qquad \text{for } 1 \leqq p < \tfrac{11}{6}, \tag{54}$$

$$\int_0^T \int_{B_R} |E(t)|^p \, dx \, dt < +\infty, \quad \forall T, R > 0 \qquad \text{for } 1 \leqq p < \tfrac{11}{2}, \tag{55}$$

as soon as $\int f_0(x - vt, v) \, dv \in L^\infty(\mathbb{R}^+; L^{15/8}(\mathbb{R}^3))$.

Proof of Theorem 7. The regularity of E can be proved using the estimate (28), upper bounded by (31), (32) in the second step of the proof of Theorem 1. For $m < 2$, this gives

$$\|E(t)\|_{m+3} \leqq C(m)(M_k(t))^{1/m+3} \tag{56}$$

where $C(m) \to +\infty$ if $m \to 2$, and $k := k(m)$ can be chosen as close as necessary to m, and thus we may choose $k = 2$. This proves (52).

Next, we notice that for any $\alpha < 1$

$$\sup_{t \in [0,T]} \int_{\mathbb{R}^3 \times \mathbb{R}^3} |v|^\alpha f(t) \, E(t) \, dx \, dv < +\infty, \quad \forall T > 0 \tag{57}$$

indeed, following the proof of (29), we have

$$\int_{\mathbb{R}^2 \times \mathbb{R}^3} |v|^\alpha f(t) \, E(t) \, dx \, dv \leqq C(\alpha) \|E(t)\|_{4+\alpha} \| \int_{\mathbb{R}^3} |v|^\alpha f(t) \, dv \|_{\frac{4+\alpha}{3+\alpha}}$$

$$\leqq C(\alpha) \|E(t)\|_{4+\alpha} \Big(\int_{\mathbb{R}^3 \times \mathbb{R}^3} |v|^2 f(t) \, dv \, dx + 1 \Big)$$

Applying the higher moment lemma ([18]) we find (53). Finally, (54) is obtained as (14) and (55) is a local version of (56).

We would like to conclude this section with some questions concerning the results of this paper. Of course the first one is to know if moments in v between 2 and 3 are propagated by Vlasov-Poisson Equation. Concerning the uniqueness proof, it seems possible to improve the assumption (42) by just requirering that the quantity in (42) belongs to $L^\infty((0, T) \times \mathbb{R}^3_x; L^1(\mathbb{R}^3_v) \cap L^{3/2}(\mathbb{R}^3_v))$. But it is certainly possible to get a uniqueness result without any assumption on the gradient of the initial data. For solutions satisfying only the energy and L^∞ bounds of Sect. 4, the conservation of the total energy is still not solved. Also it seems possible to get the L^5 bound in (52). Of course any L^p estimate on E for p larger than 5 would automatically allow a boot-strap argument leading to further regularity results.

References

1. Arsenev, A.A.: Global existence of a weak solution of Vlasov's system of equations. U.S.S.R. Comput. Math. Math. Phys. **15**, 131–143 (1975)
2. Arsenev, A.A.: Some estimates for the solution of the Vlasov Equation (Russian) Zh. Vychiol. Mat. i Mat. Fiz. 25 (1985), n° 1, 80–87
3. Bardos, C., Degond, P.: Global existence for the Vlasov-Poisson Equation in 3 space variables with small initial data. Ann. Inst. Henri Poincaré, Anal. Non Linéaire **2**, 101–118 (1985)
4. Batt, J.: Global symmetric solutions of the initial value problem of stellar dynamics. J. Differ. Equations **25**, 342–364 (1977)
5. Batt, J.: The nonlinear Vlasov-Poisson system of Partial Differential Equations in stellar dynamics. Publ. CNER Math. Pures Appl. **5**, 1–30 (1983)
6. Batt, J.: The present state of the existence theory of the Vlasov-Poisson and Vlasov-Maxwell system of partial differential equations in plasma physics. In: Boffi, V., Neunzert, H. (eds.) Applications of mathematics in technology. Proceedings Rome 1984, pp. 375–385. Stuttgart: Teubner 1984
7. Degond, P.: Régularité de la solution des équations cinétiques en physique des plasmas. Sémin. Équations Dériv. Partielles. Palaiseau: Ecole Polytechnique 1985–1986
8. Di Perna, R.J., Lions, P.L.: Ordinary differential equations transport equations and Sobolev spaces. Invent. Math. **98**, 511–547 (1989); Semin. Équations Dériv. Partielles. Palaiseau: Ecole Polytechnique 1988–1989
9. Di Perna, R.J., Lions, P.L.: Solution globales d'équations du type Vlasov-Poisson. C.R. Acad. Sci. Paris **307**, 655–658 (1988) and paper in preparation
10. Di Perna, R.J., Lions, P.L.: Global weak solutions of kinetic equations. Rend. Semin. Mat., Torino (to appear)
11. Glassey, R.T., Schaeffer, J.: On symmetric solutions of the relativistic Vlasov-Poisson system. Commun. Math. Phys. **101**, 459–473 (1985)
12. Horst, E.: On the classical solutions of the initial value problem for the unmodified nonlinear Vlasov Equations. I: General theory. II: Special cases. Math. Methods Appl. Sci. **3**, 229–248 (1981); **4**, 19–32 (1982)
13. Horst, E., Hunze, R.: Weak solutions of the initial value problem for the unmodified nonlinear Vlasov equation. Math. Methods Appl. Sci. **6**, 262–279 (1984)
14. Horst, E.: Global strong solutions of Vlasov's Equation. Necessary and sufficient conditions for their existence. Partial differential equations. Banach Cent. Publ. **19**, 143–153 (1987)
15. Illner, R., Neunzert, H.: An existence theorem for the unmodified Vlasov Equation. Math. Meth. Appl. Sci. **1**, 530–544 (1979)
16. Iordanskii, S.V.: The Cauchy problem for the kinetic equation of plasma. Transl., II. Ser., Am. Math. Soc. **35**, 351–363 (1964)
17. Lions, P.L., Perthame, B.: Régularité des solutions du système de Vlasov-Poisson en dimension 3. C.R. Acad. Sci., Paris, Sér. I t. 311 (1990)
18. Perthame, B.: Higher moments for kinetic equations: application to Vlasov-Poisson and Fokker-Planck equations. Math. Meth. Appl. Sci. **13**, 441–452 (1990)
19. Pfaffelmoser, K.: Global classical solutions of the Vlasov-Poisson system in three dimensions for general initial date. (Preprint)
20. Schaeffer, J.: Global existence for the Poisson-Vlasov system with nearly symmetric data. J. Diff. Eq. **69**, 111–148 (1987)
21. Ukai, S., Okabe, T.: On classical solutions in the large in time of two dimensional Vlasov's equation. Osaka J. Math. **15**, 245–261 (1978)
22. Wollman, S.: Global in time solutions of the two dimensional Vlasov-Poisson system. Commun. Pure Appl. Math. **33**, 173–197 (1980)

Invent. math. 105, 431–441 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Théorie de Hovanskii et problème de Dulac

R. Moussu et C. Roche

Laboratoire de Topologie, Université de Bourgogne, CNRS 755,
BP 138, F-21004 Dijon, France

Oblatum 30-IX-1989 & 29-XI-1990

Ce travail a pour but de montrer comment la théorie des variétés pfaffiennes de Hovanskii [H, R] permet de résoudre géométriquement le problème de Dulac dans un cas particulier [D, M, EMMR].

I Introduction et résultats

Soit $\vartheta = a(x, y)\,dx + b(x, y)\,dy = 0$ une équation différentielle analytique sur un ouvert de $\mathbf{R}^2$, à points singuliers isolés. Un *polycycle* P de $\vartheta = 0$ est une union finie, connexe de points singuliers (*ses sommets*) et de courbes intégrales (*ses côtés*) qui possède une *application retour* unilatérale f. C'est-à-dire, il existe une courbe analytique

$$T\colon [0, 1] \to \mathbf{R}^2, \qquad T(0) \in P,$$

transverse aux courbes intégrales de $\vartheta = 0$ telle que la courbe intégrale passant par $T(t)$ recoupe une première fois $T([0, 1])$ en $T(f(t))$ si t est assez petit. Evidemment f fixe 0 et est analytique sur un ouvert $]0, \varepsilon[$ mais, n'est pas en général analytique en 0.

Problème de Dulac. *Montrer que si f n'est pas l'identité, 0 est un point fixe isolé de f.*

Ce problème est résolu par Il'Yashenko [I_1] lorsque les sommets a_k, $k = 1, 2, \ldots, n$ de P sont non dégénérés dans le sens suivant: les courbes $a = 0$, $b = 0$ se coupent transversalement aux points a_k. Dans le cas général, une esquisse de preuve a été proposée dans [EMMR, Y]. Récemment Il'Yashenko [I_2] et Écalle [E_1, E_2] ont fourni des démonstrations complètes et indépendantes.

Tous ces travaux reprennent et complètent la méthode de Dulac: on montre tout d'abord que f possède un développement asymptotique $\hat{f}$ (transasymptotique dans [EMMR]) tel qui si f n'est pas l'identité, $\hat{f}$ n'est pas l'identité. On conclut en remarquant que, dans ce cas, $f(x) - x$ possède une partie principale non oscillante. Nous allons démontrer, par un argument de nature complètement différente, le théorème suivant:

Théorème 1 *Soit f l'application retour d'un polycycle P d'une équation différentielle analytique $\vartheta=0$ sur un ouvert de $\mathbf{R}^2$. Si f n'est pas l'identité, 0 est un point fixe isolé de f lorsque P vérifie l'hypothèse $(\mathscr{I})$ suivante: ϑ possède un facteur intégrant analytique en chacun de ses sommets.*

Rappelons qu'un *facteur intégrant* analytique F de ϑ en un point a est un germe de fonction analytique en a, non constant, tel que

$$\vartheta \wedge dF + F d\vartheta = 0.$$

Bien que très restrictive, la condition $(\mathscr{I})$ n'est pas complètement arbitraire. En effet, après désingularisations, les sommets de P possèdent toujours un facteur intégrant formel. Nous reviendrons sur ceci à la fin de II. D'autre part signalons que la preuve du Théorème 1 s'applique à une équation différentielle sur une variété analytique réelle de dimension deux.

La preuve du théorème repose sur une propriété de finitude du nombre de composantes connexes des variétés pfaffiennes de Hovanskii. Rappelons tout d'abord leur définition, ou plus exactement celle que nous utiliserons.

Soit M un ouvert semi-analytique de $\mathbf{R}^n$, ω une 1-forme différentielle analytique sur un voisinage de l'adhérence $\bar{M}$ de M dans $\mathbf{R}^n$ et soit V une sous-variété connexe, de codimension 1, de M. Nous dirons que V est une *hypersurface séparante* de M associée à ω, ce que nous écrirons sous la forme d'un triplet (V, ω, M) si

(i) $M \setminus V$ est la réunion de deux ouverts connexes dont V est la frontière commune dans M;
(ii) en tout point x de V, $\operatorname{Ker}\omega(x)$ est l'espace tangent $T_x V$ de V en x.

Théorème 2 *Soient X un ensemble semi-analytique de $\mathbf{R}^n$ et (V_k, ω_k, M) pour $k = 1, 2, \ldots, q$ des hypersurfaces séparantes. Si M est relativement compact le nombre de composantes connexes de $X \cap V_1 \cap V_2 \cap \ldots \cap V_q$ et fini.*

Ainsi, du point de vue de la finitude du nombre de composantes connexes, les hypersurfaces séparantes se comportent comme les sous-ensembles semi-analytiques. En fait, elles héritent d'autres propriétés des ensembles semi-analytiques: finitude des groupes d'homologie, structure conique.... Ceci sera étudié dans un travail ultérieur dans un cadre plus général.

Cependant, les hypersurfaces séparantes ne sont pas, en général, semi-analytiques en un point de leur bord. Par exemple, le graphe de $x \mapsto e^{-1/x}$ est une hypersurface séparante de $x^2 dy - y dx = 0$ dans $\mathbf{R}_+^* \times \mathbf{R}$. De cette remarque et du Théorème 2 on peut déduire un démonstration géométrique du théorème sur les *fonctions de Liouville*. Plus précisément, le Théorème 2 permet de montrer un théorème des zéros isolés pour une classe de fonctions plus grande que celle de Liouville.

Nous ne sommes pas certains que le résultat du Théorème 2 soit complétement original. Dans [T], J-C. Tougeron étudie des questions du même type et surtout Hovanskii dans [H] énonce un théorème dont le Théorème 2 n'est qu'un aspect. Mais, nos hypothèses semblent moins restrictives et la méthode que nous utilisons pour le prouver est différente de celles utilisées dans [T]. Par contre nous utilisons l'argument basique de Hovanskii, la variante suivante du théorème de Rolle classique:

Si c: $[\alpha, \beta] \to M$ est une courbe C^1 qui coupe une hypersurface séparante (V, ω, M) en $c(\alpha)$, $c(\beta)$, il existe $\tau \in [\alpha, \beta]$ tel que $\omega(c(\tau)) \cdot (c'(\tau)) = 0$.

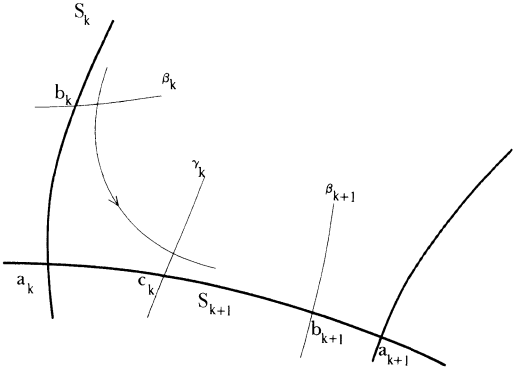


Fig. 1

Dans le chapitre suivant nous prouverons le Théorème 1 en admettant le Théorème 2. Ce dernier est démontré dans le chapitre III en utilisant deux lemmes dont la démonstration sera faite dans IV.

II Démonstration du Théorème 1

Soit $\vartheta = a\,dx + b\,dy$ une équation différentielle analytique sur un ouvert de $\mathbf{R}^2$, à points singuliers isolés et soit P un polycycle de $\vartheta = 0$ à côtés S_k et sommets $a_k = S_k \cap S_{k+1}$, $k = 1, 2, \ldots, n$, avec la convention $n+1 \equiv 1$. Nous supposons que P vérifie la condition $(\mathscr{I})$: il existe une fonction analytique non constante, F_k, sur un voisinage U_k de a_k pour $k = 1, 2, \ldots, q$ telle que:

$$(\mathscr{I}_k) \qquad \vartheta \wedge dF_k + F_k\, d\vartheta = 0.$$

Choisissons, pour chaque k, des points $b_k \in S_k \cap (U_k - a_k)$, $c_k \in S_{k+1} \cap (U_k - a_k)$ et des courbes analytiques (comme sur la Fig. 1), transverses à $\vartheta = 0$ notées

$$\beta_k\colon\ x_k \in]-1, 1[\mapsto \beta_k(x_k), \qquad \beta_k(0) = b_k,$$

$$\gamma_k\colon\ y_k \in]-1, 1[\mapsto \gamma_k(y_k), \qquad \gamma_k(0) = c_k.$$

La courbe intégrale de $\vartheta = 0$ passant par $\beta_k(x_k)$ pour $x_k \in]0, \varepsilon_k[$ coupe une première fois γ_k au point $\gamma_k(y_k)$. L'application $\bar{g}_k\colon x_k \mapsto y_k$ ainsi définie est analytique et se prolonge en 0 par continuité. Montrons tout d'abord l'assertion suivante:

Le graphe $\bar{G}_k$ de $\bar{g}_k$ est une courbe intégrale d'une équation différentielle analytique $\bar{\omega}_k = 0$ sur un voisinage de a_k.

De $(\mathscr{I}_k)$ on déduit que les composantes irréductibles de $F_k = 0$ sont des courbes intégrales de $\vartheta = 0$. Ainsi F_k ne s'annule pas sur la composante connexe U_k' de $U_k \setminus S_k \cup S_{k+1}$ qui contient des points de β_k, γ_k. On choisit U_k de façon que

U_k' soit simplement cónnexe. Les courbes intégrales de $\vartheta=0$ dans U_k' sont les courbes de niveau d'une fonction analytique:

$$I_k\colon U_k' \to \mathbf{R} \quad \text{avec} \quad dI_k = \frac{\vartheta}{F_k}.$$

En particulier $\bar{g}_k$ est défini implicitement par l'équation $I_k(\beta_k(x_k)) = I_k(\gamma_k(y_k))$, $y_k = \bar{g}_k(x_k)$. Ainsi $\bar{g}_k$ est solution de l'équation différentielle:

$$\left(\frac{\vartheta}{F_k}\right)_{\beta_k(x_k)} (\beta_k'(x_k)) = \left(\frac{\vartheta}{F_k}\right)_{\gamma_k(y_k)} (\gamma_k'(y_k)) \cdot \frac{dy_k}{dx_k}.$$

Son graphe $\bar{G}_k$ est une courbe intégrale de l'équation analytique

$$\bar{\omega}_k = F_k(\gamma_k(y_k))\ \vartheta(\beta_k(x_k)) \cdot (\beta_k'(x_k))\, dx_k - F_k(\beta_k(x_k))\ \vartheta(\gamma_k(y_k)) \cdot (\gamma_k'(y_k))\, dy_k$$

sur un voisinage de 0. Soit $\bar{\omega}_k$ une 1-forme analytique sur le même voisinage de 0 telle que $\bar{\omega}_k \wedge \bar{\omega}_k = 0$ et dont 0 est un zéro isolé.

La courbe intégrale de $\vartheta=0$ passant par $\gamma_k(y_k)$, pour y_k assez petit, coupe une première fois β_{k+1} en $\beta_{k+1}(x_{k+1})$. L'application $h_k\colon y_k \mapsto x_{k+1}$ ainsi définie est un difféomorphisme analytique qui fixe 0. Le graphe G_k du composé

$$g_k = h_k \circ \bar{g}_k\colon\ x_k \mapsto x_{k+1}, \qquad x_k > 0$$

est une courbe intégrale de l'équation $0 = \omega_k = H_k^*(\bar{\omega}_k)$ où H_k est le difféomorphisme

$$H_k\colon\ (x_k, x_{k+1}) \mapsto (x_k, h_k^{-1}(x_{k+1})).$$

D'autre part, l'application retour f de P évaluée sur γ_1 est le composé $f = g_n \circ g_{n-1} \circ \cdots \circ g_1$.

Considérons les ω_k, $k=1, 2, \ldots, n$ comme des 1-formes (analytiques) sur des voisinages de 0 dans $\mathbf{R}^n = \{x = (x_1, x_2, \ldots, x_n)\}$. Soit $\rho > 0$ tel ques les ω_k soient analytiques sur la boule $B_0(\rho) = \{x \in \mathbf{R}^n / \|x\| < \rho\}$. Notons M le sous-ensemble $B_0(\rho) \cap (\mathbf{R}_+^*)^n$. Pour $k=1, 2, \ldots, n$, (V_k, ω_k, M) est une hypersurface séparante où

$$V_k = \{x \in M / (x_k, x_{k+1}) \in G_k\}.$$

Le graphe (dans M) de $f_{n-1} = g_{n-1} \circ \ldots \circ g_1$ est la courbe $\Gamma = \bigcap_{k=1}^{n-1} V_k$ et l'ensemble $\mathrm{Fix}\, f$ des points fixes de f est la projection $\pi_1(\Gamma \cap V_n)$ avec $\pi_1\colon x \mapsto x_1$. D'après le Théorème 2, $\Gamma \cap V_n$ a un nombre fini de composantes connexes et si

$$\#\,\mathrm{Fix}\, f = \#\bigcap_{k=1}^{n} V_k = \infty,$$

Γ est contenue dans V_n et f est l'identité.

Remarque. Les travaux précédents sur le problème de Dulac (utilisation des développements asymptotiques) commencent par la désingularisation des sommets du polycycle. Dans notre approche, cette première étape n'a pas été nécessaire, compte tenu de l'hypothèse ($\mathscr{I}$). Par contre, il est clair qu'elle le redevient

sans l'hypothèse ($\mathscr{I}$); c'est le seul moyen d'avoir des renseignements sur les applications g_k.

Supposons que P soit un polycycle de $\vartheta=0$ à sommets a_k, $k=1, 2, \ldots, n$, réduits: ces points sont des points singuliers hyperboliques ou semi-hyperboliques de $\vartheta=0$ (voir [M]). Dans les deux cas $\vartheta=0$ possède un facteur intégrant formel $\hat{F}_k$ au point a_k (voir [C-M]). Dans des coordonnées analytiques (u, v) au point a_k, bien choisies, $\hat{F}$ s'écrit

$$\hat{F}(u, v)=u^p \cdot \sum_{n\geqq 0} \hat{a}_n(u) v^n, \qquad p\geqq 1,\ \hat{a}_n\in \mathbf{R}[\![u]\!].$$

Selon les cas, les $\hat{a}_n$ sont analytiques, resommables. Compte tenu de ces propriétés, on peut espérer que notre approche géométrique du problème de Dulac puisse être encore utilisée sans l'hypothèse ($\mathscr{I}$). Ceci nécessiterait évidemment une théorie des ensembles définis par des équations non nécessairement analytiques mais resommables.

III Stratification adaptée et preuve du Théorème 2

La démonstration du Théorème 2 repose sur la notion de stratification adaptée à X et à la collection de 1-formes, $\Omega=\{\omega_1, \omega_2, \ldots, \omega_q\}$, si (V_k, ω_k, M) pour $k=1, 2, \ldots, q$ sont les hypersurfaces séparantes de l'énoncé. La définition, l'existence et les propriétés de telles stratifications sont l'objet du premier paragraphe. Dans le second nous en déduierons la preuve du théorème.

1 Stratifications adaptées

Nous utiliserons dans toute la suite la notation suivante: soient a un point de $\mathbf{R}^n$, ρ un réel >0 et A un sous ensemble de $\mathbf{R}^n$,

$$A_a(\rho)=A\cap B_a(\rho) \quad \text{avec} \quad B_a(\rho)=\{x\in\mathbf{R}^n/\|x-a\|<\rho\}.$$

Un sous-ensemble semi-analytique, connexe, N de $\mathbf{R}^n$ est appelé *un feuillet* si c'est une sous-variété analytique de $\mathbf{R}^n$. On dira que N est un *feuillet normal de la boule* $B_a(\rho)$ s'il existe des fonctions h_l, $l=1, 2, \ldots, s$ analytiques sur $B_a(\rho)$ telles que:

$$N_a(\rho)=\{x\in B_a(\rho)/h_1(x)=\ldots=h_{n-p}(x)=0,\ h_{n-p+1}(x)>0, \ldots, h_s(x)>0\}$$

$$dh_1\wedge dh_2\wedge\ldots\wedge dh_{n-p}(x)\neq 0 \quad \text{si} \quad x\in N_a(\rho).$$

Toutes les stratifications considérées dans la suite sont finies, analytiques et semi-analytiques. C'est-à-dire que leurs strates, en nombre fini, sont des feuillets de $\mathbf{R}^n$.

Théorème de Lojasiewicz [L] *Soient X un sous-ensemble semi-analytique de $\mathbf{R}^n$, a un point adhérent à X et $f_1, f_2, \ldots, f_t$ des fonctions analytiques sur un voisinage de a dans $\mathbf{R}^n$. Il existe $\rho>0$ et une stratification $\mathscr{N}$ de $\bar{B}_a(\rho)$ telle que:*

(i) *$X_a(\rho)$ et $\partial X_a(\rho)$ sont des réunions de strates de $\mathscr{N}$.*

(ii) *Le signe* $(+, 0, -)$ *de chaque* f_l $l=1, 2, \ldots, t$ *est constant sur chaque strate de* $\mathcal{N}$.

(iii) *Les strates de* $\mathcal{N}$ *adhérentes à a sont des feuillets normaux de* $B_a(\rho)$, *transverses aux sphères* $\partial B_a(\rho)$ *pour* $r>0$.

Une stratification vérifiant (i), (ii) (resp. (i), (ii), (iii)) est appelée une *stratification* (resp. *normale*) *de* $\bar{B}_a(\rho)$ *adaptée à* X *et aux* f_l.

Soient ω_k, $k=1, 2, \ldots, q$ des 1-formes différentielles analytiques sur un voisinage de $a\in\mathbf{R}^n$ et soit $\Omega=\{\omega_k/k=1, \ldots, q\}$. Si J est un sous-ensemble de $\{1, 2, \ldots, q\}$ nous notons Ω_J la collection des ω_j pour $j\in J$, $\Omega_J(x)$ le sous-espace vectoriel de $T_x^*\mathbf{R}^n$ engendré par les $\omega_j(x)$, $j\in J$ et $\operatorname{Ker}\Omega_J(x)$ l'intersection des sous-espaces $\operatorname{Ker}\omega_j(x)$ de $T_x\mathbf{R}^n$ pour $j\in J$. Si N est une sous-variété de $\mathbf{R}^n$ nous dirons que Ω_J est transverse à N (et nous écrirons $\Omega_J \pitchfork N$) si en tout point x de N

$$\dim(\operatorname{Ker}\Omega_J(x)\cap T_xN)=\dim N-\#J.$$

Si de plus, pour tout $x\in N$, $\Omega(x)+T_x^0N$ est engendré par $\Omega_J(x)+T_x^0N$ nous dirons que Ω_J est *une base de* Ω *le long de* N, où T_x^0N est l'orthogonal de T_xN.

Proposition 1 *Soient* X *un sous-ensemble semi-analytique de* $\mathbf{R}^n$ *et* $\Omega=\{\omega_1, \omega_2, \ldots, \omega_q\}$ *des 1-formes analytiques sur un voisinage d'un point* $a\in\mathbf{R}^n$. *Il existe* $\rho>0$, *une stratification normale* $\mathcal{N}$ *de* $\bar{B}_a(\rho)$ *adaptée à* X *et une application* J *de* $\mathcal{N}$ *dans l'ensemble des parties de* $\{1, 2, \ldots, q\}$ *telle que pour toute strate* $N\in\mathcal{N}$, $\Omega_{J(N)}$ *soit une base de* Ω *le long de* N.

Nous dirons qu'une telle stratification est *adaptée* à X et Ω. Dans toute la suite nous notons W l'ensemble $\bigcap_{k=1}^{q} V_k$.

Lemme 1 *Soient* (V_k, ω_k, M) $k=1, \ldots, q$ *des hypersurfaces séparantes et* N *un feuillet de* M. *Si* Ω_J *est une base de* Ω *le long de* N *toute composante connexe de* $W\cap N$ *est une composante connexe de* $\bigcap_{j\in J} V_j\cap N$.

Lemme 2 *Soit* N *un feuillet normal dans une boule* $B_a(\rho_0)$, *et soient* (V_k, ω_k, M) $k=1, \ldots, q$ *des hypersurfaces séparantes telles que* $\Omega=\{\omega_1, \ldots, \omega_q\}$ *soit transverse à* N *et* $p=\dim X>q$. *Il existe* $\rho>0$ *et un sous-ensemble semi-analytique* $Y\subset X$ *de dimension strictement plus petite que* p *tel que toute composante connexe de* $W\cap N\cap\bar{B}_a(\rho)$ *rencontre* Y.

2 *Démonstration du Théorème 2*

La démonstration se fait par une double récurrence sur p et q avec $p=\dim X$, $q=\#\Omega$, $\Omega=\{\omega_1, \ldots, \omega_q\}$.

Lorsque $q=0$, le théorème est le résultat classique [L] sur la finitude du nombre de composantes connexes des ensembles semi-analytiques relativement compacts. Lorsque $p=0$, pour le même raison, $X\cap M$ est un ensemble fini de points et le résultat est encore vrai.

A titre d'exemple montrons le théorème lorsque $p=1$. Puisqu'un ensemble semi-analytique fermé de dimension 1 est une union finie d'adhérences de cour-

bes analytiques lisses et de points isolés, il suffit de montrer que $\#V\cap X=\infty$ entraine $X\subset V$ lorsque (V, ω, M) est une variété séparante et X est l'image du segment $[0, 1]$ par une application analytique:

$$\gamma\colon\ ]-\varepsilon, 1+\varepsilon[\to \mathbf{R}^n, \qquad \gamma'(t)\neq 0 \quad \text{et} \quad t\neq 0.$$

Supposons donc $\#(V\cap X)=\infty$ et soit $\{t_n\}$ une suite monotone de points de $[0, 1]$ qui converge vers un point t_0 telle que les points $\gamma(t_n)$ appartiennent à $V\cap X$. On peut toujours supposer que $\{t_n\}$ est strictement décroissante. D'après le théorème de Rolle-Hovanskii, il existe une suite $\{t'_n\}$, telle que

$$t_{n+1}<t'_n\leqq t_n, \qquad g(t'_n)=0$$

où $g(t)=\omega(\gamma(t))\cdot(\gamma'(t))$. Puisque g est analytique au voisinage de t_0, g est identiquement nulle et X est contenu dans V.

Abordons maintenant la récurrence. Soit X un ensemble semi-analytique de dimension p et soit W l'intersection des q hypersurfaces séparantes (V_k, ω_k, M) $k=1, 2, \ldots, q$. Supposons (hypothèse de récurrence) que l'intersection d'un ensemble semi-analytique de dimension p' et de q' hypersurfaces séparantes de M n'a qu'un nombre fini de composantes connexes dès que: $p'<p$ (et q' quelconque) ou $q'<q$ (et p' quelconque). Compte tenu de la relative compacité de M pour montrer le théorème il suffit de montrer l'affirmation suivante:

(*) *Pour chaque point $a\in\bar{M}$ il existe $\rho>0$ tel que $W\cap X_a(\rho)$ a un nombre fini de composantes connexes.*

Soit, comme dans la proposition, une stratification normale $\mathcal{N}$ d'une boule fermée $\bar{B}_a(\rho)$ adaptée à X et $\Omega=\{\omega_1, \ldots, \omega_q\}$. Puisque $X_a(\rho)$ est une union finie de feuillets de $\mathcal{N}$ il suffit de prouver (*) lorsque $X_a(\rho)$ est un feuillet N de $\mathcal{N}$. Notons $J=J(N)$ le sous-ensemble de $\{1, 2, \ldots, q\}$ tel que Ω_J est une base de Ω le long de N.

Si Ω_J est strictement inclus dans Ω, d'après le Lemme 1, toute composante connexe de $W\cap N$ est une composante connexe de $\bigcap_{j\in J} V_j\cap N$. D'après l'hypothèse de récurrence ($\#J=q'<q$) cet ensemble a un nombre fini de composantes connexes. D'autre part, si $\Omega_J=\Omega$ et $\dim N=p'<p$, on sait que (*) est encore vrai par hypothèse de récurrence. Supposons Ω transverse à N, $\dim N=p$ et distinguons deux cas:

● $p=q$: soient $\Omega'=\{\omega_1, \ldots, \omega_{q-1}\}$ et $W'=\bigcap_{k=1}^{q-1} V_k$. Il est clair que Ω' est encore transverse à N. D'après l'hypothèse de récurrence ($\#\Omega'=q'<q$), $W'\cap N$ est une union finie de courbes lisses $\{C_l\}_{l\in L}$. Écrivons $W\cap N$ sous la forme

$$W\cap N=(W'\cap N)\cap V_q.$$

Deux points distincts d'une courbe C_l ne peuvent pas appartenir simultanément à V_q: en effet, sinon, d'après le théorème de Rolle-Hovanskii il existerait un point $c\in C_l$ tel que la tangente à C_l en c soit dans le noyau de $\omega_q(c)$; ce qui contredirait la transversalité de Ω et N.

● $p>q$: le feuillet N vérifie les hypothèses du Lemme 2. Toute composante connexe de $W\cap N$ rencontre un ensemble semi-analytique Y de dimension $<p$.

Ainsi $W \cap N$ a moins de composantes connexes que $W \cap Y$ et d'après l'hypothèse de récurrence $W \cap Y$ en a un nombre fini.

Ceci achève la démonstration du Théorème 2.

IV Preuve de la Proposition 1 et des Lemmes 1 et 2

Soient $\mathcal{N}_1$, $\mathcal{N}_2$ deux stratifications d'une boule $\bar{B}_a(\rho)$ de $\mathbf{R}^n$. Supposons qu'une strate N de $\mathcal{N}_1$ soit réunion de strates de $\mathcal{N}_2$. La partition $\mathcal{N}_3$ de $\bar{B}_a(\rho)$ obtenue en remplaçant dans $\mathcal{N}_1$ les strates contenues dans $\bar{N}$ par les composantes connexes de leurs intersections avec les strates de $\mathcal{N}_2$ est encore une stratification. Nous dirons que $\mathcal{N}_3$ est *le raffinement* de $\mathcal{N}_1$ par $\mathcal{N}_2$ sur N.

Nous utiliserons très souvent dans la suite (et sans le signaler) la remarque élémentaire suivante: si C_1, C_2 sont deux composantes connexes distinctes d'un espace topologique A et si $B \subset A$, alors $C_1 \cap B$ ne rencontre aucune composante connexe de $C_2 \cap B$.

Dans toute la suite $\omega_1, \omega_2, \ldots, \omega_q$ designent des 1-formes analytiques sur une boule de centre $a \in \mathbf{R}^n$ et (V_k, ω_k, M), $k = 1, 2, \ldots, q$ des hypersurfaces séparantes qui leur sont associées. On pourra remarquer que la propriété «séparante» de V_k n'intervient pas dans les preuves de la proposition et des lemmes; seul le fait que V_k est une solution de codimension 1 de $\omega_k = 0$ est utilisé.

1 Démonstration de la Proposition 1

Elle se fait par une «récurrence descendante» sur la dimension des strates d'une stratification adaptée à X. Soit $\mathcal{N}$ une stratification normale de $B_a(\rho)$ adaptée à X telle que:

$(\dagger)_{p+1}$ il existe une application J de l'ensemble des strates $N' \in \mathcal{N}$ de dimension $\geqq p+1$ dans les parties de $\{1, 2, \ldots, q\}$ telle que $\Omega_{J(N')}$ soit une base de Ω le long de N'.

Montrons l'existence de $\rho' \leqq \rho$ et d'une stratification normale $\mathcal{N}''$ de $\bar{B}_a(\rho')$ adaptée à X telle que $(\dagger)_p$ soit vrai, c'est-à-dire J est définie sur l'ensemble des strates de dimension $\geqq p$ de $\mathcal{N}''$.

Soit N une strate de $\mathcal{N}$ de dimension p. Elle est définie par des fonctions analytiques $h_1, h_2, \ldots, h$, telles que:

- $N = \{x / h_1(x) = \ldots = h_{n-p} = 0 \text{ et } h_i(x) > 0 \text{ si } i > n-p\}$
- $\sigma(x) = dh_1 \wedge \ldots \wedge dh_{n-p}(x) \neq 0$ si $x \in N$.

Un système de coordonnées analytiques centré en a étant fixé, on note f_J la somme des carrés des coefficients de la $(n-p+\#J)$-forme $\sigma \wedge \pi_J$, où $\pi_J = \bigwedge_{j \in J} \omega_j$.

C'est une fonction analytique et $\Omega_J = \{\omega_j / j \in J\}$ est transverse à N en x si et seulement si $f_J(x) \neq 0$.

Soit $\mathcal{N}'_N$ une stratification normale d'une boule fermée $\bar{B}_a(\rho'_N)$ adaptée à X, N et aux fonctions f_J lorsque J décrit les sous-ensembles de $\{1, 2, \ldots, q\}$. Définissons une application, notée encore J de l'ensemble des strates $N' \in \mathcal{N}'_N$

de dimension p dans ces sous-ensembles qui vérifie la propriété (†): $J(N')$ est choisi de telle façon que

- $f_{J(N')}$ ne s'annule pas sur N'; ainsi $\Omega_{J(N')}$ est transverse à N',
- si J' est un sous-ensemble de $\{1, 2, \ldots, q\}$ tel que si $\#J' > \#J(N')$, alors $f_{J'}$ est identiquement nulle sur N'; en particulier si $j \notin J(N')$

$$\sigma \wedge \pi_{J(N') \cup \{j\}}(x) = 0 \quad \text{si} \quad x \in N'.$$

Ainsi $\omega_j(x)$ appartient au sous-espace de $T_x^* \mathbf{R}^n$ engendré par $\Omega_{J(N')}(x) + T_x^0 N'$ et $\Omega_{J(N')}$ est une base de Ω le long de N'.

Soit $\mathcal{N}_N''$ le raffinement de $\mathcal{N}$ sur N par $\mathcal{N}_N'$. Il est important de remarquer que dans ce raffinement, N est la seule strate de dimension $\geqq p$ de $\mathcal{N}$ qui est changée. Soit $\mathcal{N}''$ la stratification intersection des $\mathcal{N}_N''$ lorsque N décrit les strates de dimension p de $\mathcal{N}$. Par construction c'est une stratification normale d'une boule $\bar{B}_a(\rho'')$ adaptée à X telle que à chaque strate $N'' \in \mathcal{N}''$, de dimension $\geqq p$, correspond un sous-ensemble $\Omega_{J(N'')}$ de Ω qui est une base de Ω le long de N'' où $J(N'')$ est défini de la façon suivante:

- $J(N'') = J(N)$ si $\dim N'' > p$ où N est la strate de $\mathcal{N}$ contenant N''.
- $J(N'') = J(N')$ si $\dim N'' = p$ où N' est une strate de dimension p de $\mathcal{N}_N'$ contenant N''.

2 Démonstration du Lemme 1. Rappelons qu'il s'agit de prouver que: si Ω_J est une base de Ω le long d'un feuillet N, alors toute composante connexe de $W \cap N$, $\left(W = \bigcap_{j=1}^{q} V_j\right)$ est une composante connexe de $\bigcap_{j \in J} V_i \cap N$. Pour le montrer il suffit de prouver l'assertion suivante: soit C une composante connexe de $\bigcap_{j \in J} V_j \cap N$ et soit $j_0 \in \{1, 2, \ldots, q\} \setminus J$, alors C est contenue dans V_{j_0} si V_{j_0} rencontre C. La condition, Ω_J est une base de Ω le long de N implique, que pour tout $x \in N$ on ait:

$$T_x N \cap \operatorname{Ker} \Omega_J(x) \subset \operatorname{Ker} \omega_{j_0}(x).$$

Or ces espaces sont respectivement l'espace tangent à C et V_{j_0} au point x et, ainsi, C est une sous-variété de V_{j_0}.

3 Démonstration du Lemme 2. On peut toujours supposer, (quitte a raffiner) que N est un feuillet normal de $B_a(\rho)$, que

$$\bar{N} \setminus N = \bigcup_{i=0}^{l} N_i \cup Y_0$$

où les N_i $i = 0, 1, \ldots, l$ sont des feuillets normaux de $B_a(\rho)$ avec la convention $N_0 = \{a\}$ et Y_0 est un sous-ensemble semi-analytique de $\partial B_a(\rho)$ de dimension $< p$ (par exemple $Y_0 = \bar{N} \cap \partial \bar{B}_a(\rho)$ si $\bar{N}$ est transverse aux sphères de centre a et de rayon $\leqq \rho$). On note η la $(n-p)$-forme: $dg_1 \wedge \ldots \wedge dg_{n-p}$, où $g_1 = \ldots = g_{n-p} = 0$ et $g_l > 0$, $l > n-p$ sont les équations de N.

A chaque strate N_i on associe la fonction analytique sur $B_a(\rho)$

$$\delta_i = h_1^2 + h_2^2 + \ldots + h_s^2$$

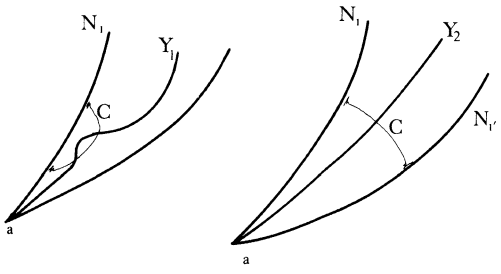


Fig. 2

où $h_1 = h_2 = \ldots = h_s = 0$, $h_{s+1}, \ldots, h_\alpha > 0$ sont les équations de N_i. Le sous-ensemble E_i de N constitué des points x tels que

$$d\delta_i \wedge \eta \wedge \omega_1 \wedge \omega_2 \wedge \ldots \wedge \omega_q(x) = 0$$

contient l'ensemble des points où $W \cap N$ n'est pas transverse aux niveaux de δ_i. Quitte à multiplier les h_i par des unités convenables de la boule on peut supposer que E_i est un sous-ensemble semi-analytique de N de dimension $< p$. Ainsi

$$Y_1 = \bigcap_{i=1}^{l} \bar{E}_i \cap Y_0$$

est encore semi-analytique, de dimension $< p$.

Soit C une composante connexe de $W \cap N$. Puisque Ω est transverse à N, C est une sous-variété analytique lisse de dimension $p-q$ et $\bar{C} \setminus C$ est contenu dans $\bar{N} \setminus N$. Montrons tout d'abord l'assertion suivante:

(*) Si C ne rencontre pas Y_1, alors $\bar{C}$ rencontre deux strates N_i, N_i' de $\bar{N} \setminus N$ qui ne sont pas incidentes, c'est-à-dire $N_i \not\subset \bar{N}_i'$ et $N_i' \not\subset \bar{N}_i$.

Remarquons tout d'abord que, si C est une sous-variété fermée de $B_a(\rho)$, alors C n'est pas transverse aux sphères de centre a, c'est-à-dire à δ_0 et C coupe Y_1.

Supposons que C ne rencontre pas Y_1. Alors $\bar{C} \setminus C$ est non vide et est contenu dans la réunion des N_i pour $i \geqq 0$. Pour prouver l'assertion, il suffit de montrer (compte tenu des propriétés de frontière des stratifications) que $\bar{C} \setminus N$ n'est pas contenu dans l'adhérence $\bar{N}_{i_0}$ d'une de ces strates. Si $\bar{C} \setminus C$ était contenu dans $\bar{N}_{i_0}$, la restriction de δ_{i_0} à C posséderait un maximum puisque

$$\delta_{i_0}(\bar{C} \setminus C) = 0 \quad \text{et} \quad \delta_{i_0}|_C \neq 0.$$

Ce point serait un point de $E_{i_0} \subset Y_1$.

Notons, pour $i = 0, 1, \ldots, l$, $\psi_i(x)$ la distance euclidienne d'un point x à N_i. D'après [L, p. 137], il existe une fonction continue, semi-analytique sur $\mathbf{R}^n$: φ_i telle que

$$\psi_i(x) \leqq 4\,\varphi_i(x) \leqq 16\,\psi_i(x).$$

L'ensemble Y_2' des points $x \in N$ pour les quels il existe deux indices i, i' distincts tels que $\varphi_i(x) = \varphi_{i'}(x)$ est semi-analytique. Soit Y_2 la réunion des strates de Y_2' de dimension $< p$.

D'après l'assertion (*), si C est une composante connexe de $W \cap N$ qui ne rencontre pas Y_1, $\bar{C}$ rencontre deux strates N_i, $N_{i'}$ de ∂N non incidentes et ainsi on a, de la même façon que pour $\psi_i - \psi_{i'}$:

$$(\varphi_i - \varphi_{i'})(N_i) < 0 \quad \text{et} \quad (\varphi_i - \varphi_{i'})(N_{i'}) > 0.$$

Il existe nécessairement un point de C où $\varphi_i - \varphi_{i'}$ s'annule. Ainsi toute composante connexe de $W \cap N$ coupe le sous-ensemble $Y = Y_1 \cup Y_2$ qui est semi-analytique, de dimension $< p$.

Remarque. De la démonstration de ce lemme et du Théorème 2 on déduit aisément que pour tout $\rho > 0$ assez petit $W \cap N_a(\rho)$ coupe transversalement les sphères $\partial B_a(r)$, $r < \rho$.

Bibliographie

[C-M] Cerveau, D., Mattei, J.F.: Formes intégrables holomorphes singulières. Astérisque **97** (1982)

[D] Dulac, H.: Sur les cycles limites. Bull. Soc. Math. Fr. **51**, 45–188 (1923)

[E_1] Écalle, J.: Finitude des cycles-limites et accéléro-sommation de l'application de retour. In: Françoise, J.-P., Roussarie, R. (eds.) Bifurcations of planar vector fields. Proceedings, Luminy 1989. (Lect. Notes Math., vol. 1445) Berlin Heidelberg New York: Springer 1990

[E_2] Écalle, J.: Livre à paraitre dans la collection Travaux En Cours. Paris: Hermann 1991

[EMMR] Écalle, J., Martinet, J., Moussu, R., Ramis, J.-P.: Non-accumulation de cycles limites. C.R. Acad. Sci., Paris, Sér. I. **304,** (I) 375–378, (II) 431–434 (1987)

[H] Khovanskii, A.G.: Real analytic varieties with the finiteness property and complex abelian integrals. Funct. Anal. Appl. **18**, 119–127 (1984)

[I_1] Il'Yashenko, Y.S.: Limit cycles of polynomial vector fields with nondegenerate singular points on the real plane. Funct. Anal. Appl., **18**, 199–207 (1984)

[I_2] Il'Yashenko, Y.S.: Théorèmes de finitude pour les cycles limites I. (En russe.) Usp. Mat. Nauk. **45**, 143 (1989) (Et II à paraitre)

[L] Lojasiewicz, S.: Ensembles semi-analytiques. I.H.E.S Preprint, Bures-Sur-Yvette (1965)

[M] Moussu, R.: Le problème de la finitude du nombre de cycles limites. Séminaire Bourbaki, N° 655 (1985); Astérisque **145–146**, 89–101 (1987)

[R] Risler, J.J.: Complexité et géomérie réelle (d'après A. Khovansky). Séminaire Bourbaki, N° 637 (1984); Astérisque **133–134**, 89–100 (1986)

[T] Tougeron, J.-C.: Sur certains algèbres de fonctions analytiques. Séminaire sur la géométrie algébrique réelle. Publ. Math. Univ. Paris VII **24**, 35–121 (1986)

[Y] Yoccoz, J.-C.: Non-accumulation de cycles limites. Séminaire Bourbaki, N° 690 (1987); Astérisque **161–162**, 87–103 (1989)

Invent. math. 105, 443 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Erratum

Binary forms and orders of algebraic number fields

J. Nakagawa

Joetsu University, Department of Mathematics, Joetsu, 943-Japan

Invent. Math. **97**, 219–235 (1989)

There is the following serious mistake in the proof of Theorem 3:

The estimate of Lemma 3.1 is applicable only when $X^{1/(2n-2)} \gg p$. Hence if $p \geqq X^{1/(2n-2)}$, the estimate (5.19) on p. 234 does not follow from Lemma 3.1.

The author could not recover the proof of it. Consequently, Theorems 3 and 4 are not proved. There is the following typographical mistake:

The inequality $\geqq$ on p. 229, l. 16, should be replaced by $\leqq$.

Invent. math. 105, 445–472 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

L-functions and Fourier-Jacobi coefficients for the unitary group U(3)*

Stephen S. Gelbart and Jonathan D. Rogawski*

Department of Theoretical Mathematics, Weizmann Institute of Science, Rehovot 76100 Israel

Oblatum 10-VII-1990

To the memory of Takuro Shintani

Table of contents

Introduction 445

1 Preliminaries 449
2 Fourier expansions 451
3 Global Weil representations 454
4 Shimura integral 461
5 Characterization of A-packets 465
6 Fourier-Jacobi coefficients 468

Introduction

Theta series were among the first known examples of modular forms. They are distinguished by the simplicity of their Fourier coefficients, which are, in many cases of number-theoretical interest. The representation theory underlying theta series was explained in Weil's well-known paper [We]. From this point of view, the Howe correspondence for dual reductive pairs provides a general setting for the construction of theta series, and it has proven to be an important tool for producing interesting examples of automorphic representations. On the other hand, for a general reductive group, the theory of endoscopy leads conjecturally [A, L] to a partition of the set of automorphic representations into packets of various types. This partition is essential to the understanding of Langlands functoriality. The question then arises: where do representations constructed by the Howe correspondence fit in to the classification in terms of packets?

In this paper, we investigate this problem for the dual reductive pairs $(U(m), U(3))$, $m = 1, 2$. One of our main results asserts that the discrete representations of $U(3)$ belonging to an endoscopic packet coincide with the theta series liftings from a unitary group in 1 or 2 variables (not necessarily quasi-split). We also consider some questions related to the non-vanishing of Fourier-Jacobi coefficients of automorphic forms on $U(3)$.Here and below, $U(n)$ denotes the quasi-split unitary group in n variables with respect to a quadratic extension E/F.

* *Permanent address:* University of California, Los Angeles, CA 90049, USA

In more detail, let $G = U(3)$ and assume that F is a number field. In [R], a certain class of enlarged L-packets ([A]), or A-packets, as we call them, have been defined. They are parametrized by automorphic characters ϱ of the unique elliptic endoscopic group $H = U(2) \times U(1)$ for G. The A-packet $\prod(\varrho)$ is the transfer of ϱ with respect to functoriality for an embedding $\xi_H : {}^L H \to {}^L G$. Of course, $\prod(\varrho)$ is a "tensor product" $\otimes \prod(\varrho_v)$. For all v, $\prod(\varrho_v)$ contains a certain non-tempered representation $\pi^n(\varrho_v)$, and it contains an additional representation $\pi^s(\varrho_v)$ precisely when v remains prime in E. Let $X = \{v : \pi_v = \pi^s(\varrho)\}$. By [$R_2$, Theorem 1.1], an element π of $\prod(\varrho)$ is discrete (i.e., occurs discretely in $L^2(G(F) \setminus G(\mathbf{A}))$ if and only if $(-1)^{|X|} = \varepsilon(\frac{1}{2}, \varphi)$ with φ a character of E determined by ϱ and ξ_H. The discrete representations which belong to an A-packet $\prod(\varrho)$ are of a very special type. They are almost everywhere non-tempered. The discrete representations which do not belong to any $\prod(\varrho)$ with $\dim(\varrho) = 1$ are partitioned into L-packets, and the conjectures of Arthur [A] predict that the infinite dimensional ones are all tempered everywhere.

Let Ω be the set of representations of G in the image of the Howe correspondence for $(U(1), U(3))$. They will be called Weil representations; all of them are discrete. We show that Ω coincides with the set of discrete representations which belong to $\prod(\varrho)$ for some ϱ with $\dim(\varrho) = 1$. It is not hard to check that if $\pi \in \Omega$, then for almost all v, there is a character ϱ_v of H_v such that $\pi_v = \pi^n(\varrho_v)$. We may then apply [R, Theorem 13.3.6(c)], according to which a discrete representation with local component of the form $\pi^n(\varrho_v)$ for a some place v which does not split in E belongs to $\prod(\varrho)$ for some ϱ.

To prove that a cuspidal member $\pi \in \prod(\varrho)$ belongs to Ω, we use an integral representation of "Shimura type" for the standard L-functions $L(s, \pi \otimes \xi)$ attached to π:

$$(*) \qquad L_S(s, \pi \otimes \xi) \times c(s) = \int_{G(F)\setminus G(\mathbf{A})} \varphi(g)\, \overline{\theta(g)}\, E(g, \xi, s)\, dg\,.$$

Here $L_S(s, \pi \otimes \xi)$ is the partial L-function with respect to a finite set of places S, φ belongs to the space of π, and $E(g, \xi, s)$ is an Eisenstein series attached to a Hecke character ξ of E. The function θ is a theta series belonging to the space of some $\omega \in \Omega$ and $c(s)$ is a product of local integrals at the places in S. As is to be expected, the proofs are based on an analysis of the possible pole of (*) at $s = 1$, where the Eisenstein series can have a non-zero residue which generates a one-dimensional representation.

To state the relation between Ω and the A-packets more precisely, we must take care in parametrizing the elements of Ω. Suppose that G is the unitary group of a skew-Hermitian form Φ on an E-vector space V of dimension 3. Then G embeds in the symplectic group $\mathrm{Sp}(W)$, where W is the space V viewed as an F-vector space with the alternating form $\mathrm{Tr}_{E/F}(\Phi)$. The metaplectic covering $\mathrm{Mp}_{\mathbf{A}}(W) \to \mathrm{Sp}_{\mathbf{A}}(W)$ splits over $G(\mathbf{A})$, and the metaplectic representation ω_ψ of $\mathrm{Mp}_{\mathbf{A}}(W)$ restricts to give a representation $\omega(s, \psi)$ of $G(\mathbf{A})$ which depends on the choice of splitting s. Here ψ is a character of $F \setminus \mathbf{A}_F$. The center Z of G is isomorphic to $U(1)$, and for each automorphic character χ of $U(1)$, the subrepresentation $\omega(s, \psi, \chi)$ of $\omega(s, \psi)$ on which $Z(\mathbf{A})$ acts via χ is irreducible and discrete. The action of the diagonal subgroup M of the standard Borel subgroup B of G in a Schrödinger model for $\omega(s, \psi)$

is described in terms of a certain Hecke character γ of the idele group I_E whose restriction to I_F is $\omega_{E/F}$. On the other hand, a pair (γ, ψ) consisting of a character of this type and an additive character ψ determine a unique splitting s (cf. §3.1). Hence we can write $\omega(\gamma, \psi, \chi)$ for $\omega(s, \psi, \chi)$. Let $\delta \in F^*$ act on the set of characters of $F \setminus \mathbf{A}$ by setting $\psi^\delta(x) = \psi(\delta x)$. We prove the following two results. First, the representations $\omega(\gamma, \psi, \chi)$ and $\omega(\gamma', \psi', \chi')$ are equivalent if and only if $(\gamma', \psi', \chi') = (\gamma, \psi^\delta, \chi)$ for some $\delta \in N_{E/F}(E^*)$. Secondly, there is a simple correspondence between the set of pairs (γ, χ) and the set of automorphic characters ϱ of H, and we show that the set of discrete members of $\prod(\varrho)$ consists precisely of the representations $\omega(\gamma, \psi, \chi)$, where ψ ranges over a set of representatives for the characters of $F\setminus\mathbf{A}$ modulo the action of $N_{E/F}(E^*)$. As a corollary, we obtain results about the Howe correspondence and the theta series lifting for the simplest of all dual pairs $(U(1), U(1))$ (§5.2). In this case, the Howe correspondence is the identity whenever it is defined (recall that the global Howe correspondence is defined by piecing together the local correspondences). However, by Cor. 5.2.2., if an automorphic χ is in the domain of the Howe correspondence, then its theta lift to $U(1)$ is non-zero if and only if $L(\frac{1}{2}, \gamma\chi_E^{-1}) \neq 0$, where χ_E is the "base change" of χ to $\mathrm{GL}(1)_{/E}$.

The integral representation (*) is closely related to the Fourier-Jacobi expansion of φ along the center U of the unipotent radical N of B. Let R be the centralizer of U in B. The Fourier-Jacobi expansion can be written as a sum $\varphi(g) = \sum \varphi_\tau(g)$ where τ ranges over the set $R^\wedge$ of infinite-dimensional, irreducible representations of R which occur as subrepresentations of $L^2(R(F)\setminus R(\mathbf{A}))$. The representations τ are also indexed by triples (γ, ψ, χ) and the correspondence $\tau(\gamma, \psi, \chi) \to \omega(\gamma, \psi, \chi)$ gives a bijection between the set of $B(F)$-orbits in $R^\wedge$ and Ω. The integral (*) is non-zero for some φ if and only if φ_τ is non-zero for some τ in a certain $B(F)$-orbit which is simply related to the $B(F)$-orbit corresponding to ω. This enables us to investigate the non-vanishing of the Fourier-Jacobi coefficients φ_τ in some cases. Here our motivation comes from the theory of modular forms of half-integral weight. Shimura ([S_2]) pointed out the problem of relating the Fourier coefficients $a(n)$ of a half-integral weight Hecke eigenform $\sum a(n)e^{2\pi inz}$ to its Hecke eigenvalues. The difficulty of the question lies in the fact that recursion relations provided by the Hecke operators only relate coefficients in a fixed orbit modulo squares $\{a(tm^2)$ $m = 1, 2, \ldots\}$. However, for some forms, called exceptional, the coefficients vanish off of a single orbit. Examples are given by the classical theta series $\sum \chi(n)n^\nu e^{2\pi inz}$, where χ is a Dirichlet character and ν is a positive integer such that $\chi(-1) = (-1)^\nu$. In fact, these are the only examples ([G_1, V]). In this paper, we prove a similar result for $G = U(3)$. Following Piatetski-Shapiro ([P_2]), we call an discrete representation of G exceptional if the coefficients φ_τ are all identically zero for τ not belonging to a single $B(F)$ orbit in $R^\wedge$ and if, in addition, the constant terms φ_U and φ_N coincide. Our result states that the Weil representations, which are easily seen to be exceptional, are the only exceptional representations of G.

Now let $\varrho = \varrho_2 \otimes \varrho_1$ be a *cuspidal* L-packet on H. The functional transfer of ϱ to an L-packet $\prod(\varrho)$ on G with respect to the embedding $\xi_H : {}^LH \to {}^LG$ exists by [R] and such L-packets are called endoscopic. Let $\pi \in \prod(\varrho)$ and $\tau = \tau(\gamma, \psi, \chi) \in R^\wedge$. Theorem 6.1.1 states that if the Fourier-Jacobi coefficient

φ_τ is not identically zero for some φ is the space of π, then $L(\frac{1}{2}, \gamma\varrho_{1E}^{-1}) \neq 0$ and π is a theta series lifting from a unitary group $U(\Phi')$ of the Hermitian form $\Phi' = \mathrm{diag}(\delta, -1)$, where δ is determined as follows. By our results for $U(1)$, the condition $L(\frac{1}{2}, \gamma\varrho_{1E}^{-1}) \neq 0$ implies that ϱ_1 occurs in the ψ^δ-oscillator representation of $U(1)$ relative to the splitting determined by γ for a unique $\delta \in F^*$ modulo NE^*; this is the δ which describes Φ'. This result shows, in particular, that all cuspidal members of $\prod(\varrho)$ are theta series lifts. Conversely, if σ is a cuspidal representation of $U(\Phi')$, then all irreducible subrepresentations of the theta-series lift of σ to G belong to $\prod(\varrho)$ for some ϱ. The relation between Fourier coefficients and the non-vanishing of an L-function is similar to the well-known result of Waldspurger on Fourier coefficients of modular forms of half-integral weight.

For unitary groups in general, we expect the following. For each integer $m < n$, the group $U(m) \times U(n-m)$ is an endoscopic group for $U(n)$, and there is an embedding of L-groups ${}^L(U(m) \times U(n-m)) -> {}^L(U(n)$. Assuming that the notion of L-packets is defined, let σ be a cuspidal L-packet on $U(m)$ and let η be a one-dimensional automorphic representation of $U(n-m)$. Then Arthur's conjecture would predict the existence of an A-packet $\prod(\sigma \times \eta)$ on $U(n)$ associated to σ, η and the L-group embedding. On the other hand, if $U(\Phi')$ is a unitary group in m variables, which is not necessarily quasi-split, then $(G, U(\Phi'))$ is a dual reductive pair and we can consider the theta liftings to $U(n)$ of cuspidal representations of $U(\Phi')$. It seems plausible that every discrete member of $\prod(\sigma \times \eta)$ is a theta lift from a cuspidal representation of $U(\Phi')$ for some Φ' and, conversely, that every such theta lift belongs to an A-packet $\prod(\sigma \times \eta)$, where σ is cuspidal and η is one-dimensional.

A discrete representation π of G which belongs neither to an endoscopic L-packet nor to an A-packet is called stable. It would be interesting to determine when the Fourier-Jacobi coefficients φ_τ are non-zero for a stable π. A necessary condition is that $\mathrm{Hom}_{R_v}(\pi_v, \tau_v) \neq 0$ for all v. Is this sufficient? We note that the coefficients φ_τ for holomorphic Eisenstein series were calculated explicitly by Hickey ([Hi]) and found to equal a ratio of special values of Hecke L-series.

The Weil representations arising from the pair $(U(1), U)$, where U is a unitary group in n variables, have been studied by Kazhdan [K_2] and Borel-Wallach [BW] in the anisotropic case and by Shimura [S_1] in the general case. These authors were interested in producing discrete subgroups Γ of finite co-volume in $U(\mathbf{R})$ such that $H^1(\Gamma, \mathbf{C}) \neq 0$. The pair $(U(1), U)$ can be used for this purpose because the Weil representations π are often such that $H^1(\mathrm{Lie}(U), K, \pi_\infty) \neq 0$. For $U = U(3)$, the results of this paper, together with the characterization of representations π with non-trivial H^1 in [R], show that every discrete π with $H^1(\mathrm{Lie}(U), K, \pi_\infty) \neq 0$ is a Weil representation. These Weil representations were also considered by Howe and Piatetski-Shapiro [HP], who observed that they yield examples of non-tempered cuspidal representations.

The first investigations of general automorphic forms on $U(3)$ from the point of view of representation theory and L-functions were undertaken by I.I. Piatetski-Shapiro and T. Shintani in the late seventies. In his 1977 course at Yale [P_1] and 1978 lecture at the Helsinki ICM [P_2], Piatetski-Shapiro obtained the factorization of the Shimura integral based on his theory of Heisenberg models without computing the local factors. Independently and at around the same time, Shintani [Sh_1] investigated the same integral in classical language for holomorphic automorphic forms of $U(3)$ of level one. He obtained an

expression for (*) in terms of L-functions in that case, by means of an ingenious analysis of the Fourier-Jacobi coefficients of such a form. Subsequently, Gelbart and Piatetski-Shapiro [GP_2] studied L-functions and base change for $U(3)$, dealing primarily with the case of generic representations, and Kudla [Ku] used Shintani's approach to study the Howe pairing for $(U(2), U(3))$. The approach of the present work is a synthesis of the methods introduced in [P_1, P_2, GP_2, Sh_1] with the results of [R], which are obtained by trace formula techniques.

Shintani's innovative paper was unfortunately never published. Shortly before his untimely death, he remarked in a footnote to [Sh_2] that he had a found several gaps in [Sh_1]. We are unable to understand his proof of the assertion ([Sh_1, pp. 78–80]) that a certain series is in fact a θ-series, but since no account is taken there of the need to choose a splitting of the metaplectic cover over the unitary group, it appears to us that this may be the at least one of the gaps to which Shintani referred.

Acknowledgements. The first-named author is grateful to Roger Howe, Ilya Piatetski-Shapiro, and David Soudry for helpful conversations. Thanks are due to Michael Harris for pointing out an error in our original proof of Prop. 3.1.1. Both authors are supported by a grant from the U.S.-Isarel Bi-national Science Foundation. The second-named author was a Varon Visiting Professor at the Weizmann Institute of Science during the completion of this work and he is grateful to the Varon family and the Weizmann Institute for support. He also wishes to thank The Institute for Advanced Study and to acknowledge support from the NSF (DMS-8610730) and the Sloan foundation during the spring semester 1990.

1. Preliminaries

1.1 Notation

E/F = quadratic extension of number fields, conjugation denoted by a bar
$\mathbf{A}$ = adele ring of F
E^1 = norm-one elements in E^*
v = place of F.

We will say that v is split (resp., non-split) if v splits (resp., remains prime) in E.
ψ will denote a non-trivial additive character of $F \setminus \mathbf{A}$.
$U(n)$ will denote a quasi-split unitary group in n variables. We identify $U(1)$ with the algebraic group associated to E^1. For $n = 3$, we fix

$$\Phi = \begin{pmatrix} & & 1 \\ & \xi & \\ -1 & & \end{pmatrix}$$

where $\xi \in E^*$, $\mathrm{Tr}_{E/F}(\xi) = 0$ and denote by G the (quasi-split) unitary group of the skew-Hermitian form Φ.

Z = center of G (isomorphic to E^1)
B = standard Borel subgroup of upper triangular matrices in G
N = unipotent radical of B
M = diagonal subgroup of G
U = center of N
R = centralizer of U in B

We use the following notation for elements of N and M:

$$[w,t] = \begin{pmatrix} 1 & w\xi & \frac{\xi w\overline{w}}{2} + t \\ & 1 & w \\ & & 1 \end{pmatrix} \quad \text{where} \quad w \in E \quad \text{and} \quad t \in F,$$

$$\operatorname{diag}(\alpha, \beta, \overline{\alpha}^{-1}) = \begin{pmatrix} \alpha & & \\ & \beta & \\ & & \overline{\alpha}^{-1} \end{pmatrix} \quad \text{where} \quad \alpha \in E^*, \quad \beta \in E^1.$$

The group N is the Heisenberg group attached to the F-vector space E with symplectic form $\langle x, y\rangle = \operatorname{Tr}_{E/F}(\xi x \overline{y})$. Its center is $U = \{[0,t] : t \in F\}$.

We have $R = SN$, where $S = \{\operatorname{diag}(\alpha, \beta, \alpha), \alpha, \beta \in E^1\}$. Furthermore $N \setminus R$ is isomorphic to $(E^1)^2$.

1.2 Conventions. All representations will tacitly be assumed admissible. If π is a representation, V_π will denote the space on which π acts.

If ν is a character of E^1 (locally or globally), we define a character ν_E of E^* by $\nu_E(\alpha) = \nu(\alpha/\overline{\alpha})$. It is the *base change* of ν to $\mathrm{GL}_{1/E}$.

1.3 L-functions. Let π be a discrete representation of G. The *base change* π_E of π to $\mathrm{GL}_{3/E}$ is an automorphic representation of $\mathrm{GL}_{3/E}$ ([R]). For all Hecke characters ξ of E, the "standard" L-function $L(s, \pi \otimes \xi)$ is, by definition, the GL_3 L-function $L(s, \pi_E \otimes \xi)$. For a description at the unramified places, cf. [GR_1, § 1.2].

1.4 Endoscopy. Fix a character μ of the idele class group C_E whose restriction to C_F is the character $\omega_{E/F}$ associated to E/F by class field theory. The choice of μ fixes an embedding of L-groups $\xi_H : {}^L H \to {}^L G$, where $H = U(2) \times U(1)$ is the unique elliptic endoscopic group of G (cf. [R, pp. 51–52]). One ξ_H is fixed, we obtain a transfer from discrete L-packets on H to automorphic L-packets on G. As usual, an L-packet $\prod$ on G is a "tensor product" $\otimes \prod_v$ of local L-packets $\prod_v$ such that $\prod_v$ contains an unramified representation for almost all v. An L-packet ϱ on H is of the form $\varrho = \varrho_2 \times \varrho_1$, where ϱ_j is an L-packet on $U(j)$. Of course, an L-packet on $U(1)$ consists of a single automorphic character. If $\dim(\varrho) \neq 1$, we denote by $\prod(\varrho)$ the L-packet on G corresponding to ϱ with respect to ξ_H. For all $\pi \in \prod(\varrho)$, we have $L(s, \pi \otimes \xi) = L(s, \varrho_{2E} \otimes \mu\xi) L(s, \varrho_{1E} \otimes \xi)$, where ϱ_{jE} denotes the base change lift of ϱ_j to $\mathrm{GL}_{j/E}$. The base change lift ϱ_{2E} is an automorphic representation of $\mathrm{GL}_{2/E}$.

If $\dim(\varrho) = 1$, we let $\prod(\varrho)$ denote the A-packet associated to ϱ, as described in the Introduction. Suppose that ϱ is defined by

$$\varrho((h_2, h_1)) = \eta(\det(h_2))\eta'(\det(h_2)h_1)$$

where η, η' are automorphic characters of E^1 and $h_j \in U(j)(\mathbf{A})$. For all v, let $\pi^n(\varrho_v)$ be the Langlands quotient of the principal series representation induced from the character ν of B_v defined by

$$\nu(d(\alpha, \beta, \overline{\alpha}^{-1})) = \mu_v(\alpha)\eta_v(\alpha/\overline{\alpha})\,\eta'_v((\alpha/\overline{\alpha})\beta)\,\|\alpha\|_v^{1/2}.$$

If v is split, then $\prod(\varrho_v) = \{\pi^n(\varrho_v)\}$. If v remains prime, then there exists another irreducible representation $\pi^2(\varrho_v)$ such that $\prod(\varrho_v) = \{\pi^n(\varrho_v), \pi^s(\varrho_v)\}$. If v is finite, then $\pi^s(\varrho_v)$ is supercuspidal, and if v is archimedean then $\pi^s(\varrho_v)$ is either in the discrete series (holomorphic or anti-holomorphic) or limit of discrete series. If v is finite and the characters μ_v, η_v, and η_v' are unramified, then $\pi^n(\varrho_v)$ is unramified, and hence if $\pi = \otimes\pi_v \in \prod(\varrho)$, then $\pi_v = \pi^n(\varrho_v)$ for almost all v.

2. Fourier expansions and exceptional automorphic forms

2.1 Let φ be an automorphic form on $G(\mathbf{A})$. Recall that ψ denotes throughout a non-trivial character of $F \setminus \mathbf{A}$. We also view it as a character of $U(F) \setminus U(\mathbf{A})$, and define

$$\varphi_\psi(g) = \int_{U(F)\setminus U(\mathbf{A})} \varphi(ug)\,\psi(u)\,du \tag{2.1.1}$$

This is the ψ^{th} *Fourier-Jacobi coefficient* of φ along U. The constant term φ_U of φ along U is defined similarly by $\int \varphi(ug)\,du$. By harmonic analysis on the abelian group U,

$$\varphi(g) = \varphi_U(g) + \sum_\psi \varphi_\psi(g). \tag{2.1.2}$$

Observe that the function $r \to \varphi_\psi(rg)$ belongs to $L^2(R(F) \setminus R(\mathbf{A}))$ for all $g \in G(\mathbf{A})$. The constant term φ_N can be written in terms of φ_U as follows:

$$\varphi_N(g) = \int_{N(F)\setminus N(\mathbf{A})} \varphi(ng)\,dn = \int_{N(F)U(\mathbf{A})\setminus N(\mathbf{A})} \varphi_U(ng)\,dn.$$

2.2 Let $R^\wedge$ be the set of infinite-dimensional representations of $R(\mathbf{A})$ which occur as subrepresentations of $L^2(R(F) \setminus R(\mathbf{A}))$. The unipotent radical N is a Heisenberg group with center U, and hence for each non-trivial character ψ of $F \setminus \mathbf{A}$, there is a unique isomorphism class τ_ψ of irreducible admissible representations of $N(\mathbf{A})$ on which $U(\mathbf{A})$ acts via ψ (Stone-von Neumann theorem). Let $R^\wedge(\psi)$ be the set of elements of $R^\wedge$ on which $U(\mathbf{A})$ acts via ψ. Each element of $R^\wedge(\psi)$ is an extension of τ_ψ to $R(\mathbf{A})$, and if $\tau_1, \tau_2 \in R^\wedge(\psi)$, then τ_1 is isomorphic to $\tau_2 \otimes \nu$, where ν is a character of $N(\mathbf{A}) \setminus R(\mathbf{A})$. In fact, ν is trivial on $R(F)$, and hence can be viewed as an automorphic character of $(E^1 \setminus E^1_{\mathbf{A}})^2$ (cf. §2.5). The next theorem will be proved in §2.5.

Theorem 2.2.1. *Each irreducible constitutent of $L^2(R(F) \setminus R(\mathbf{A}))$ occurs with multiplicity one.*

2.3 For $\tau \in R^\wedge(\psi)$, let $\varphi_\tau(rg)$ be the projection of the function $r \to \varphi_\psi(rg)$ onto the τ-subspace of $L^2(R(F) \setminus R(\mathbf{A}))$ (uniquely defined by Theorem 2.2.1). Then (2.1.2) reads

$$\varphi(g) = \varphi_U(g) + \sum_{\tau \in R^\wedge} \varphi_\tau(g). \tag{2.3.1}$$

If π is a discrete representation, set

$$\Lambda(\pi) = \{\tau \in R^\wedge : \varphi_\tau \neq 0 \quad \text{for some} \quad \varphi \in V_\pi\}.$$

Observe that R is normal in B and $B(F)$ acts on $R^\wedge$. If $b = mn \in B(F)$, where $n \in N(F)$ and $m = d(\alpha, \beta, \bar{\alpha}^{-1})$, then b maps $R^\wedge(\psi)$ to $R^\wedge(\psi^{N\alpha})$, and $\varphi_{\tau^b}(g) = \varphi_\tau(bg)$. The set $\Lambda(\pi)$ is a union of $B(F)$-orbits.

Definition 2.3.1. *A discrete representation (π, V) is called exceptional if the following two conditions hold:*

(1) *$\Lambda(\pi)$ is a single $B(F)$-orbit.*
(2) *For each $\varphi \in V$, $\varphi_U = \varphi_N$.*

Locally, $R_v^\wedge$ will denote the set of irreducible, infinite-dimensional representations of R_v. Let π_v be an irreducible representation of G_v. A non-zero R_v map from π_v to τ_v, where $\tau_v \in R_v^\wedge$, is called a *Heisenberg model* for π_v. By a theorem of Piatetski-Shapiro ([P_1, GR_2]), a Heisenberg model exists for all infinite-dimensional π_v. Furthermore, Heisenberg models are unique in the following sense: $\dim(\mathrm{Hom}_{R_v}(\pi_v, \tau_v)) \leqq 1$ for all τ_v. Observe that the set of τ_v such that $\mathrm{Hom}_{R_v}(\pi_v, \tau_v) \neq \{0\}$ is a union of B_v-orbits in $R_v^\wedge$. If the set of $\tau_v \in R_v^\wedge$ such that $\mathrm{Hom}(\pi_v, \tau_v) \neq \{0\}$ is a single B_v-orbit, then π_v is called locally exceptional. We will see later (Theorem 2.4.1, 5.1.1) that an exceptional automorphic representation is locally exceptional at all finite places.

2.4 Recall the definition of the A-packet $\prod(\varrho)$ given in § 1.4.

Theorem 2.4.1. *If π is exceptional, then π belongs to $\prod(\varrho)$ for some automorphic character ϱ of H.*

Proof. By [R, Theorem 13.3.6], if there exists a non-split place v of F such that $\pi_v = \pi^n(\varrho_v)$ for some one-dimensional representation ϱ_v of H_v, then $\pi \in \prod(\varrho)$ for some automorphic character ϱ of H. On the other hand, by [GR_1, Proposition 5.1.6.], if v is non-split and unramified in E, then an unramified locally exceptional representation of G_v is of the form $\pi^n(\varrho_v)$. Hence it will suffice to show that π_v is locally exceptional at all places v which do not split in E.

Choose $\tau \in R^\wedge$ so that the subspace $T(\pi)$ of $L^2(R(F) \setminus R(\mathbf{A}))$ generated by $\{\varphi_\tau(r) : \varphi \in V_\pi\}$ is non-zero. Such a τ exists, by Lemma 2.4.2 below. By Theorem 2.2.1, $T(\pi)$ is irreducible, and hence is isomorphic to τ. We call $\tau_v \in R_v^\wedge$ unramified if the residual characteristic of v is different from 2, the central character ψ_v has conductor zero, and τ_v contains a vector fixed under $R(\mathcal{O}_v) = S(\mathcal{O}_v)N(\mathcal{O}_v)$, where $S(\mathcal{O}_v)$ is the maximal compact subgroup of S. By [GR_1, Lemma 8.1.1(a)], the space of $R(\mathcal{O}_v)$-fixed vectors in an unramified τ_v is one-dimensional. Let S be a finite set of places such that π_v and τ_v are unramified for $v \notin S$. We identify π with $\otimes\pi_v$ relative to a collection $\{x_v^\circ\}_{v \notin S}$, where x_v° is a K_v-fixed vector in π_v for $v \notin S$. Similarly, we identify τ with $\otimes\tau_v$ relative to a collection $\{y_v^\circ\}_{v \notin S}$ where y_v° is unit vector in τ_v fixed by $R(\mathcal{O}_v)$.

Define $L : V_\pi \to V_\tau$ by $L(\varphi) = \varphi_\tau$. Since L is non-zero, it follows that $\mathrm{Hom}_{R_v}(\pi_v, \tau_v) \neq \{0\}$ for all v. We may assume that $R(\mathcal{O}_v) = R_v \cap K_v$ for $v \notin S$. In this case, $L_v(x_v^\circ)$ is a non-zero multiple of y_v° for all $L_v \in \mathrm{Hom}_{R_v}(\pi_v, \tau_v)$, and we may assume that $L_v(x_v^\circ) = y_v^\circ$ for $v \notin S$. Now, by a standard argument involving uniqueness of Heisenberg models (cf. JL, §9]), we can choose $L_v \in \mathrm{Hom}_{R_v}(\pi_v, \tau_v)$ for $v \in S$ so that L is the tensor product $\otimes L_v$.

Suppose that v is non-split and that π_v is not locally exceptional. Then $\mathrm{Hom}_{R_v}(\pi_v, \tau_v') \neq \{0\}$ for some $\tau_v' \in R_v^\wedge$ which does not belong to the B_v-orbit of τ_v, and by [GR_1, Proposition 4.1.3.] there exists a vector $w_v^\circ \in V_{\pi_v}$ such that $L_v(\pi_v(b)w_v^\circ) = 0$ for all $b \in B_v$. Choose a non-zero decomposable vector $\varphi \in V_\pi$ with local component w_v° at v. Then $\varphi_\tau(b) = 0$ for all $b \in B(\mathbf{A})$, and by the density of $B(F) \setminus B(\mathbf{A})$ in $G(F) \setminus G(\mathbf{A})$, $\varphi_\tau \equiv 0$. Observe that $\varphi_\tau b(g) = \varphi_\tau(bg)$ for all $b \in B(F)$ and hence, by the exceptionality of π, $\varphi_\tau' \equiv 0$ for all $\tau' \in R^\wedge$. However, as already noted, this is not possible by the following lemma.

Lemma 2.4.2. *Let (π, V) be an infinite-dimensional discrete automorphic representation of G, and let φ be a non-zero element of V. Then $\varphi_\psi \not\equiv 0$ for some non-trivial ψ.*

Proof. Let S be the closed subgroup of $G(\mathbf{A})$ generated by $U(\mathbf{A})$ and $\mathrm{SU}_3(F)$. If $\varphi_\psi \equiv 0$ for all non-trivial ψ, then $\varphi = \varphi_U$. Since φ and φ_U are left-invariant under $\mathrm{SU}_3(F)$ and $U(\mathbf{A})$, respectively, this assumption implies φ is left-invariant under S. To obtain a contradiction (to the infinite-dimensionality of π), it suffices to show that $S = \mathrm{SU}_3(\mathbf{A})$ (since $S \setminus G(\mathbf{A})$ will then be compact abelian). Observe that for all v, S contains the closed subgroup S_v generated by $\{\gamma u \gamma^{-1} : \gamma \in \mathrm{SU}_3(F), u \in U_v\}$. Since $\mathrm{SU}_3(F)$ is dense in $\mathrm{SU}_3(F_v)$ (for example, by strong approximation), S_v is a closed normal subgroup of $\mathrm{SU}_3(F_v)$. Hence $S_v = \mathrm{SU}_3(F_v)$ and $S = \mathrm{SU}_3(\mathbf{A})$. □

2.5 Proof of Theorem 2.2.1. The constitutents on which $U(\mathbf{A})$ acts trivially are one-dimensional and the assertion is clear for these. So let $L_\psi^2(N(F) \setminus N(\mathbf{A}))$ (resp., $L_\psi^2(R(F) \setminus R(\mathbf{A}))$) be the L^2-space of measurable functions on $N(F) \setminus N(\mathbf{A})$ (resp., $R(F) \setminus R(\mathbf{A})$) which transform under $U(\mathbf{A})$ via ψ. By a theorem of Moore [M], $L_\psi^2(N(F) \setminus N(\mathbf{A}))$ is irreducible under the action of $N(\mathbf{A})$, and hence is isomorphic to τ_ψ.

Let $(\tau_\circ, V_\circ)$ be an irreducible subrepresentation of $L_\psi^2(R(F) \setminus R(\mathbf{A}))$. The map from $V_\circ$ to $L_\psi^2(N(F) \setminus N(\mathbf{A}))$ given by restriction is an isomorphism of $N(\mathbf{A})$-representations. It is also an isomorphism of $R(F)$-modules, where $r \in R(F)$ acts on $L_\psi^2(N(F) \setminus N(\mathbf{A}))$ by sending $\varphi(n)$ to $\varphi(r^{-1}nr)$. Hence any two subrepresentations of $L_\psi^2(R(F) \setminus R(\mathbf{A}))$ differ by a twist by a character of $R(F)N(\mathbf{A}) \setminus R(\mathbf{A})$.

We now show that if τ_0 is isomorphic to $\tau_0 \otimes \nu$, where ν is a character of $R_F N(\mathbf{A}) \setminus R(\mathbf{A})$, then ν is trivial. It will suffice to show that ν_v is trivial for almost all v. Since ν_v is unramified for almost all v, it is trivial for almost all non-split v. To handle the case when ν_v and τ_{0v} are unramified and v splits, we argue indirectly. We have shown in [GR_1] that there is a bijection between the set of orbits $B_v \setminus R_v^\wedge$ and the set Ω_v of Weil representations of G_v for all finite places v. Let $\omega(\tau_v)$ be the Weil representation corresponding to the orbit of τ_v. If ν_v is non-trivial, then $\omega(\tau_{\circ v})$ and $\omega(\tau_{\circ v} \otimes \nu_v)$ are seen to be Langlands quotients which are not isomorphic ([GR_1, § 2]). Hence $\tau_{\circ v}$ and $\tau_{\circ v} \otimes \nu_v$ do not lie in the same B_v-orbit, and in particular they are not isomorphic.

It will therefore suffice to show that if $(\tau_\circ, V_\circ)$ and (τ_1, V_1) are irreducible subrepresentations of $L_\psi^2(R(F) \setminus R(\mathbf{A}))$, in the obvious notation, then there is a character ν of $R_F N(\mathbf{A}) \setminus R(\mathbf{A})$ such that $V_1 = \{\varphi\nu : \varphi \in V_\circ\}$.

Let $r_\circ \in R(\mathbf{A})$ and consider the map $T_j(r_\circ)$ from V_j to $L_\psi^2(N(F) \setminus N(\mathbf{A}))$

which sends φ to the function $n \to \varphi(nr_\circ)$. Clearly T_j is non-zero since V_j is invariant under translation. Furthermore, T_j intertwines $\tau_j|_{N(A)}$ with the representation $\tau_\psi^{r_\circ}$ defined by $\tau_\psi^{r_\circ}(n) = \tau_\psi(r_\circ n r_\circ^{-1})$. Now $\tau_\psi^{r_\circ}$ is clearly irreducible, and hence T_j is an isomorphism. It follows that $T_1(r_0)^{-1}T_0(r_0)$ defines an $N(\mathbf{A})$-isomorphism from V_0 to V_1. Since the V_j are irreducible under the action of $N(\mathbf{A})$, $T_1(r_0)^{-1}T_0(r_0)$ differs from $T_1(1)^{-1}T_0(1)$ by a non-zero scalar $f(r_0)$. In other words, for all $\varphi_1 \in V_1$, there exists a unique $\varphi_0 \in V_0$ such that $\varphi_1(n) = \varphi_0(n)$ for $n \in N(\mathbf{A})$, and in this case $\varphi_1(r) = f(r)\varphi_0(r)$ for all $r \in R(\mathbf{A})$. It is easy to check that $f(r)$ is left-invariant under $R_F N(\mathbf{A})$ and a smooth function on the compact abelian group $R_F N(\mathbf{A}) \setminus R(\mathbf{A}) = (E^1 \setminus E_{\mathbf{A}}^1)^2$. Thus we can write f as a Fourier series with respect to characters v of $(E^1 \setminus E_{\mathbf{A}}^1)^2$: $f = \sum a_v \cdot v$, where $a_v \in \mathbf{C}$. But τ_1 is isomorphic to $\tau_0 \otimes v_0$ for some character v_0 of $(E^1 \setminus E_{\mathbf{A}}^1)^2$, and therefore $a_v = 0$ unless $v = v_0$.

3 Global Weil representations

3.1 Compatible splittings. To study theta series on the unitary group, we must embed it in the metaplectic group.

Let (W, ϕ) be a symplectic space over F and let $H(W) = W \oplus F$ be the associated Heisenberg group. As usual, the symplectic group $\mathrm{Sp}(W)$ acts as a group of automorphisms of $H(W)$, fixing the center elementwise. For each non-trivial character ψ, let ϱ_ψ be an irreducible unitary representation of $H_{\mathbf{A}}(W)$ with central character ψ (ϱ_ψ is unique up to isomorphism). We define the global metaplectic group $\mathrm{Mp}_{\mathbf{A}}(W)$ as the group of pairs (g, M_g) where $g \in \mathrm{Sp}_{\mathbf{A}}(W)$ and M_g is an operator on the space of ϱ_ψ such that $M_g \varrho_\psi(h) M_g^{-1} = \varrho_\psi(g(h))$ for all $h \in H(\mathbf{A})$. The map $(g, M_g) \to M_g$ defines a unitary representation of $\mathrm{Mp}_{\mathbf{A}}(W)$, which we denote by ω_ψ. The projection $\pi((g, M_g)) = g$ yields an exact sequence

$$0 \longrightarrow \mathbf{C}^* \longrightarrow \mathrm{Mp}_{\mathbf{A}}(W) \xrightarrow{\pi} \mathrm{Sp}_{\mathbf{A}}(W) \longrightarrow 0.$$

For each finite place v of F, the local metaplectic covering $\pi_v : \mathrm{Mp}_v(W) \to \mathrm{Sp}_v(W)$ is defined similarly. Let $\mathrm{Sp}(\mathcal{O}_v)$ be the standard maximal compact subgroup of integral matrices in $\mathrm{Sp}_v(W)$. For almost all v, there is a unique splitting $\alpha_v : \mathrm{Sp}(\mathcal{O}_v) \to \mathrm{Mp}_v(W)$. Let $\prod' \mathrm{Mp}_v(W)$ be the restricted direct product over all places v with respect to the collection of subgroups $\alpha_v(\mathrm{Sp}(\mathcal{O}_v))$. There is a natural map from $\prod' \mathrm{Mp}_v(W)$ to $\mathrm{Mp}_{\mathbf{A}}(W)$ with kernel $\{(t_v) \in \prod' \mathrm{Mp}_v(W) : t_v \in \mathbf{C}^*, \prod t_v = 1\}$. By means of this map, we view α_v as a splitting of π over $\mathrm{Sp}(\mathcal{O}_v)$. It is known ([We]) that π splits uniquely over the group of F-rational points $\mathrm{Sp}_F(W)$. We denote this splitting by i.

Let (V, Φ) be a skew Hermitian space, where V is an n-dimensional vector space over E. In this §, let G denote the unitary group attached to (V, Φ). For all v, let K_v be the subgroup of integral points in G_v. For almost all v, K_v is a maximal compact subgroup of G_v. Let (W, ϕ) be the associated symplectic space over F. That is, W coincides with V, but viewed as an F-vector space, and $\phi = \mathrm{Tr}_{E/F}(\Phi)$.

Proposition 3.1.1. *The covering π splits over $G(\mathbf{A})$. There exists a continuous section $s : G(\mathbf{A}) \to \mathrm{Mp}_{\mathbf{A}}(W)$ such that $s(G(F))$ is contained in $i(\mathrm{Sp}_F(W))$.*

Proof. Assume first that $\dim(V) = 2$ and that G is quasi-split. We may choose Φ so that $G = T \cdot \mathrm{SL}_2(F)$, where $T = \{\mathrm{diag}(a, \bar{a}^{-1}) : a \in E^*\}$. An explicit local splitting $s_v : G_v \to \mathrm{Mp}_v(W)$ for all places v is given in [MVW, pp. 56–57]. We shall reconstruct this s_v and check that if $g \in G(\mathbf{A})$, then the product $\prod s_v(g)$ belongs to $\prod' \mathrm{Mp}_v(W)$. It must be shown that $s_v(K_v) \subset \alpha_v(\mathrm{Sp}(\mathcal{O}_v))$ for almost all v. Identify $\mathrm{Mp}_v(W)$ with $\mathrm{Sp}_v(W) \times \mathbf{C}^*$ relative to the explicit choice of two cocycle $c(g, g')$ of [MVW, Theorem, p. 56]. The multiplication law in $\mathrm{Sp}_v(W) \times \mathbf{C}^*$ is $(g, z)(g', z') = (gg', c(g, g')zz')$. Let $g = HK$, where $K = \mathrm{diag}(k, \bar{k}^{-1})$ and $h = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(F)$. Similarly, let $g' = H'K'$. Assume that $E = F((-D)^{1/2})$ where $D \in F$. The restriction of c to G_v has the form:

$$c(g, g') = \begin{cases} (-1, -D)_v (D, N(k''))_v (N(k), N(k'))_v & \text{if } cc' \neq 0, c'' = 0 \\ (N(k), D)_v (N(k), N(k'))_v & \text{if } c'c'' \neq 0, c = 0 \\ (N(k'), D)_v (N(k), N(k'))_v & \text{if } cc'' \neq 0, c' = 0 \\ (N(k), N(k'))_v & \text{if } c = c' = c'' = 0 \\ (cc'c'', -D)_v (N(k), N(k'))_v & \text{if } cc'c'' \neq 0. \end{cases}$$

where $(\ ,\ )_v$ is the Hilbert symbol for F_v and N is the norm relative to E/F. Fix a global additive character ψ of $F \setminus \mathbf{A}$. If f is a quadratic form over F_v, then $\psi_v \circ f$ is a character of the second degree and the Weil factor $\gamma(\psi_v \circ f)$ is defined. For $a \in F_v^*$, set $\gamma_v(a) = \gamma(\psi_v \circ ax^2)/\gamma(\psi_v \circ x^2)$. By [We, p. 176], γ_v is trivial on $(F_v^*)^2$ and the relation

$$\frac{\gamma_v(ab)}{\gamma_v(a)\,\gamma_v(b)} = (a, b)_v$$

holds. Now choose a Hecke character λ of E whose restriction to F is $\omega_{E/F}$, and define

$$b_v(HK) = \begin{cases} (c, -D)_v (N(k), D)_v \gamma_v(N(k)) \lambda_v(k) & \text{if } c \neq 0 \\ (d, -D)_v \gamma_v(N(k)) \lambda_v(k) & \text{if } c = 0. \end{cases} \tag{3.1.2}$$

(Note that in [MVW, p. 57], a factor $\chi(N(k))$ appears instead of $\lambda_v(k)$, where χ is a square root of $\omega_{E/F}$; however such a square root need not exist). Observe that $b_v(hk)$ is well-defined: If $m = \mathrm{diag}(a, a^{-1}) \in \mathrm{SL}_2(F) \cap T$ and $c \neq 0$, then

$$\begin{aligned} b_v((Hm^{-1})(mK)) &= (ac, -D)_v (a^{-2}N(k), D)_v \gamma_v(a^{-2}N(k)) \lambda_v(a^{-1}k) \\ &= (c, -D)_v (N(k), D)_v \gamma_v(N(k)) \lambda_v(k) [(a, -D)_v \lambda_v(a^{-1})] \\ &= b_v(HK) \end{aligned}$$

since $(a, -D)_v = \omega_{E/F}(a)$. Similarly if $c = 0$. One can check by cases that c is trivialized by b_v, that is: $c(g, g') = b_v(gg') b_v(g)^{-1} b_v(g')^{-1}$. The splitting s_v is defined by $s_v(g) = (g, b_v(g))$.

Let e_1, e_2 be the standard basis vectors on which $\mathrm{diag}(a, \bar{a}^{-1})$ acts with eigenvalues a and $\bar{a}^{-1}$, respectively and let $W_j = Ee_j$. Following [Rao], we let

$\mathrm{Sp}_v(W)$ act on W on the right and realize the oscillator representation $\omega_{\psi v}$ on $L^2(W_1)$. Let M_v be the subgroup of $\mathrm{Sp}_v(W)$ which stabilizes the subspaces W_1 and W_2. If the restriction of $g \in M_v$ to W_1 is g_1, where $g_1 \in \mathrm{GL}(W_1)$, then:

$$\omega_{\psi v}((g,1))f(w) = \gamma_v(\det(g_1))^{-1}|\det(g_1)|^{1/2} f(wg_1) \tag{3.1.3}$$

for $f \in L^2(W_1)$, by [Rao, Theorem 5.3].

Let f_0 be the characteristic function of $\mathcal{O}_v e_1$. For $g \in \mathrm{Sp}(\mathcal{O}_v)$, $\alpha_v(g)$ is the unique element in $\pi_v^{-1}(g)$ such that $\omega_{\psi v}(\alpha_v(g))$ fixes f_0. The subgroup $\mathrm{SL}_2(\mathcal{O}_v) = K_v \cap \mathrm{SL}_2(F_v)$ coincides with its commutator subgroup for almost all v. Hence there is at most one splitting of π_v over $\mathrm{SL}_2(\mathcal{O}_v)$ and it is automatic that $s_v(\mathrm{SL}_2(\mathcal{O}_v)) \subset \alpha_v(\mathrm{Sp}(\mathcal{O}_v))$. If $K = \mathrm{diag}(k, \bar{k}^{-1})$, then $s_v(K) = (K, \lambda_v(k)\gamma_v(N(k)))$. We have $\det(g_1) = N(k)$, $|\det(g_1)| = 1$, and

$$\omega_{\psi v}(s_v(K))f(w) = \lambda_v(k)\omega_{\psi v}((K,1))f(w) = \lambda_v(k)f(wk).$$

For almost all v, λ_v is unramified, so that $\lambda_v(k) = 1$ and $\omega_{\psi v}(s_v(K))$ fixes f_0. This shows that $s_v(K) \in \alpha_v(\mathrm{Sp}(\mathcal{O}_v))$ for a.a.v.

We may now define $s(g)$ to be the image of $\prod s_v(g)$ in $\mathrm{Mp}_{\mathbf{A}}(W)$. The cocycle $c(g,g')$ is a product of Hilbert symbols and Hasse invariants. By the product formulas for both of these, $c(g,g') = 1$ if $g, g' \in \mathrm{Sp}_F(W)$. It follows that the (unique) splitting of π over $\mathrm{Sp}_W(F)$ is given by $i(g) = (g,1)$. On the other hand, the product formula also shows that $b(g) = \prod b_v(g) = 1$ if $g \in G(F)$. This shows that $s(g) = i(g)$.

Now let G be the unitary group of a skew Hermitian form Φ_n in n variables, and let G_0 be the associated special unitary group. Let $(V, \Phi) = (V_n \oplus V_2, \Phi_n \oplus \Phi_2)$, where (V_2, Φ_2) is a two-dimensional quasi-split skew Hermitian space, and let $H_0 = \mathrm{SU}(\Phi)$. We view $S(U(\Phi_2) \times G)$ as a subgroup of H_0 in the obvious way and define an embedding $q : G \to S(U(\Phi_2) \times G)$ as follows. Choose a homomorphism $p : U(1) \to U(\Phi_2)$ such that $\det(p(x)) = x$, and define $q(g) = (g, p(\det g^{-1}))$. (To obtain p, note $U(\Phi_2)$ contains a Cartan subgroup isomorphic to $U(1) \times U(1)$, since Φ_2 can be chosen diagonal; then p can be the map from $U(1)$ to the first of the components of $U(1) \times U(1)$.)

Let ϕ_2 and ϕ be the symplectic forms associated to Φ_2 and Φ', respectively, and let W, W_n, W_2 be the F-vector spaces associated to V, V_n, V_2. Let $s_2 : U(\Phi_2) \to \mathrm{Mp}_{\mathbf{A}}(W_2)$ be the splitting just constructed. According to [K_2], there exists a splitting $s : H_0(\mathbf{A}) \to \mathrm{Mp}_{\mathbf{A}}(W)$. We also have the exact sequence

$$0 \longrightarrow \mathbf{C}^* \xrightarrow{j} \mathrm{Mp}_{\mathbf{A}}(W_n) \times \mathrm{Mp}_{\mathbf{A}}(W_2) \xrightarrow{\beta} \mathrm{Mp}_{\mathbf{A}}(W)$$

where $j(z) = (z, z^{-1})$. For $g \in G(\mathbf{A})$, there exists a unique $s_n(g) \in \mathrm{Mp}_{\mathbf{A}}(W_n)$ lying above g such that $\beta(s_n(g) \times s_2(p(\det g^{-1})) = s(q(g))$.

Let i, i_n, i_2 be the splittings of the metaplectic covers above $\mathrm{Sp}_F(W)$, $\mathrm{Sp}_F(W_n)$, $\mathrm{Sp}_F(W_2)$. The following diagram commutes by the uniqueness of i, i_n, i_2:

$$\begin{array}{ccc} \mathrm{Sp}_F(W_n) \times \mathrm{Sp}_F(W_2) & \xrightarrow{i_n \times i_2} & \mathrm{Mp}_{\mathbf{A}}(W_n) \times \mathrm{Mp}_{\mathbf{A}}(W_2) \\ \downarrow & & \downarrow \\ \mathrm{Sp}_F(W) & \xrightarrow{i} & \mathrm{Mp}_{\mathbf{A}}(W) \end{array}$$

The elements $i_n \times i_2(g)$ and $s_n \times s_2(q(g))$ differ by an element of $j(\mathbf{C}^*)$. But the second components of these elements are equal since i_2 and s_2 have the same restriction to $U_F(\Phi_2)$. It follows that $i_n(g) = s_n(g)$. □

Remark. A choice of splitting s as in the proposition will be called *compatible.* It is determined by the function $b(HK)$ defined above. Thus, a choice of s is equivalent to a choice of Hecke character of E whose restriction to F is $\omega_{E/F}$, once ψ is fixed.

Now fix $G = U(3)$ and s a compatible splitting of G. We denote by ω_ψ^s the representation $\omega_\psi \circ s$ of $G(\mathbf{A})$. Such a representation will be called a *oscillator representation* of $G(\mathbf{A})$. If s^* is any other compatible splitting, then $s^* = s \otimes v'$, where v' is an automorphic character of E^1, regarded as a character of G with values in the central subgroup $\mathbf{C}^*$ of $\mathrm{Mp}_{\mathbf{A}}(W)$, and $\omega_\psi^{s^*} = \omega_\psi^s \otimes v$, where v is v', viewed as a complex character.

3.2 Mixed model. Let V_j be the subspace of column vectors ${}^t(w_1, w_2, w_3)$ in V such that $w_j = 0$, and let W_j denote the underlying F-vector space. Set $W_{13} = W_1 \oplus W_3$. This decomposition of W_{13} is a polarization with respect to the symplectic structure inherited from W. Let $(\varrho_\psi^1, \mathscr{F})$ be the space of a global ψ-representation of $H_{\mathbf{A}}(W_2)$. The mixed model with respect to the decomposition $W = W_{13} \oplus W_2$ is a realization of ω_ψ on the space $\mathscr{S}(W_3(\mathbf{A}), \mathscr{F})$ (cf. [MVW, GR$_1$] for the local case). We identify $W_3(\mathbf{A})$ with $E_{\mathbf{A}}$ and write $\mathscr{S}(E_{\mathbf{A}}, \mathscr{F})$. We also identify N with $H(W_2)$ and view ϱ_ψ as a representation of $N(\mathbf{A})$.

We have the following explicit formulas for the action of $G(\mathbf{A})$ on $\mathscr{S}(E_{\mathbf{A}}, \mathscr{F})$. For $u \in U_1(\mathbf{A})$,

$$\omega_\psi^s \begin{pmatrix} 1 & & \\ & u & \\ & & 1 \end{pmatrix} \Phi(w) = \omega_\psi^{s'}(u)(\Phi(w)). \tag{3.2.1}$$

Here $\omega_\psi^{s'}$ is an oscillator representation of $U_1(\mathbf{A})$ on $\mathscr{F}$. In other words, it is the restriction to $U_1(\mathbf{A})$ of the metaplectic representation of $\mathrm{Mp}_{\mathbf{A}}(W_2)$ on $\mathscr{F}$ with respect to some lifting s' of $U_1(\mathbf{A})$ to $\mathrm{Mp}_{\mathbf{A}}(W_2)$. The lifting s' is determined by s and ψ. For $\alpha \in E_{\mathbf{A}}^*$,

$$\omega_\psi^s \begin{pmatrix} \alpha & & \\ & 1 & \\ & & \bar{\alpha}^{-1} \end{pmatrix} \Phi(w) = \gamma(\alpha) \, \|\alpha\|^{1/2} \Phi(\bar{\alpha} w) \tag{3.2.2}$$

where γ is a unitary character of I_E. For $n = [c, t] \in N(\mathbf{A})$,

$$\omega_\psi^s([(c, t])\Phi(w) = \psi(tN_{E/F}(w))\varrho_\psi^1([cw, 0])\Phi(w). \tag{3.2.3}$$

These formulas follow from their local counterparts, discussed in [GR$_1$, § 2].

Remarks. (1) The character γ appearing above is a Hecke character of E whose restriction to F is $\omega_{E/F}$ (cf. the proof of Prop. 3.1.1). If γ^* is any other Hecke character with this property, then $\gamma^* = \gamma v_E$, where v is an automorphic

character of U_1 and $v_E(\alpha) = v(\alpha/\bar{\alpha})$. If $s^* = s \otimes v$, then γ^* appears in the analogue of formula (3.2.2) for $\omega_\psi^{s^*}$.

(2) By (3.2.2), we may also view γ as follows. Let W be a two-dimensional isotropic skew Hermitian space and choose a polarization $W = W' \oplus W''$, where W' and W'' are isotropic lines for the Hermitian form. Identify the torus T which stabilizes W' and W'' with E^* in such a way that $\alpha \in E^*$ acts as multiplication by α (resp., $\bar{\alpha}^{-1}$) on W' (resp., W''). If we fix a compatible splitting of $U(W)$ in $\mathrm{Mp}(W)$, then $\alpha \in T$ acts on the Schrödinger realization on $\mathscr{S}(W'')$ by the formula

$$\varphi(w) \to \gamma(\alpha)\, \|\alpha\|^{1/2} \varphi(\alpha w).$$

As above, γ is a Hecke character of E^* whose restriction to F^* is $\omega_{E/F}$. It depends only on the splitting data ψ and λ, and is independent of the choice of polarization.

More generally, suppose that V_n is an n-dimensional skew Hermitian space of the form $W_1 \oplus W_2 \oplus \cdots \oplus W_r \oplus W$, where $W_1, \ldots, W_r$ are two-dimensional isotropic skew-Hermitian spaces and W is also a skew Hermitian space of any dimension. We have a natural embedding of $U(2)^r$ in $U(V_n)$. Fix a splitting of $U(V_n)$ in $\mathrm{Mp}(V_n)$. Let (W_j', W_j'') be a polarizations of W_j and let $\mathscr{F}$ be model for the oscillator representation of $\mathrm{Mp}(W)$. Consider a mixed Schrödinger model for $\mathrm{Mp}(V_n)$ on the Schwartz space $\mathscr{S}(W_1'' \oplus \cdots W_r'', \mathscr{F})$. Let M_j be the torus isomorphic to E^* in the j^{th} copy of $U(2)$ determined by (W_j', W_j''). Then $\alpha \in M_j$ acts on $\mathscr{S}(W_1' \oplus \cdots \oplus W_r', \mathscr{F})$ by the formula

$$\varphi((w_1, \ldots, w_r)) \to \gamma_j(\alpha)\, \|\alpha\|^{1/2}\, \varphi((w_1, \ldots, \bar{\alpha} w_j, \ldots, w_r))$$

where γ_j is a character of E^* of the above type. We observe that in fact the γ_j all coincide because the r embeddings of $U(2)$ in $\mathrm{Mp}(V_n)$ are all conjugate and γ_j depends only on the conjugacy class of the embedding by the previous paragraph.

(3) By (1), we have a splitting $s(\psi, \gamma)$ for each choice of data ψ, γ as above. By $\omega(\psi, \gamma)$ we denote the representation $g \mapsto \omega_\psi(s(\psi, \gamma)(g))$ of $U(V)$. We observe that $\omega(\psi, \gamma)$ *depends only on* ψ *modulo norms from* E. To see this, choose a central unitary similitude S with multiplier a in $N(E^x)$. As a symplectic similitude, conjugation by S lifts to an automorphism $\widetilde{ad(S)}$ of $M_p(W)$ with the property that $\pi(\widetilde{ad(S)}(g)) = g$ for any splitting $s : U(V) \to M_p(W)$. Therefore

$$\widetilde{ad(S)}(s(g)) = \chi(\det g) s(g)$$

for some character χ. On the other hand, it is known (cf. p. 36 of [MVW]) that $\omega_\psi(\widetilde{ad(S)}(s(g))) = \omega_{\psi^a}(s(g))$. Therefore equation (3.1.3) above implies that $\chi(\det g) = \chi_a(\det g) = (a, d\bar{d})$ if $\det g = d/\bar{d}$. Since it is also clear from (3.1.2) that $s(\psi_a, \gamma)(g) = \chi_a(\det g) s(\psi, \gamma)(g)$, our observation follows.

3.3 Theta-series. Recall that if ϱ_ψ is a ψ-representation of a Heisenberg group $H(\mathbf{A})$, then there is a unique (up to multiples) $H(F)$-invariant functional on ϱ_ψ, called the theta functional. Let $\theta_1 : \mathscr{F} \to \mathbf{C}$ be a theta-functional on $\mathscr{F}$ and for

$\Phi \in \mathscr{S}(E_{\mathbf{A}}, \mathscr{F})$, set

$$\theta(\Phi) = \sum_{x \in E} \theta_1(\Phi(x)). \tag{3.3.1}$$

Since, up to their centers, $H_{\mathbf{A}}(W_{13})$ and $H_{\mathbf{A}}(W_2)$ act separately on $\mathscr{S}(E_{\mathbf{A}})$ and $\mathscr{F}$, the functional $f \otimes v \to \theta_2(f)\theta_1(v)$ is $H_F(W)$-invariant if θ_2 is a theta-functional on $\mathscr{S}(E_{\mathbf{A}})$. A theta-function on $\mathscr{S}(E_{\mathbf{A}})$ being given by $\sum_{x \in E} \Phi(x)$, it follows that (3.3.1) is a theta-functional on $\mathscr{S}(E_{\mathbf{A}}, \mathscr{F})$.

A theta series, in the sense of Weil, is a function of $m \in \mathrm{Mp}_{\mathbf{A}}(W)$ defined by

$$\theta_\Phi(m) = \theta(\omega_\psi(m)\Phi)$$

for $\Phi \in \mathscr{S}(E_{\mathbf{A}}, \mathscr{F})$. It is well-known that $\theta_\Phi(m)$ is a smooth function on $\mathrm{Sp}_F(W) \setminus \mathrm{Mp}_{\mathbf{A}}(W)$. If s is a compatible section as in Proposition 3.1.1, let $\theta_{\Phi,s}(g) = \theta_\Phi(s(g))$ for $g \in G(\mathbf{A})$. The resulting space of "theta series" on $G(\mathbf{A})$ is denoted by $\Theta(s, \psi)$ or $\Theta(\gamma, \psi)$ (cf. Remark (3), § 3.2).

3.4 Weil representations. If χ is an automorphic character of the center Z of G, let $\Theta(\gamma, \psi, \chi)$ denote the χ-eigenspace of $\Theta(\gamma, \psi)$, and let

$$\theta_{\chi,\gamma}(\Phi)(g) = \int_{Z(F)\setminus Z(\mathbf{A})} \theta_{\Phi,s}(zg)\chi^{-1}(z)\, dz$$

denote the composition of the map $\Phi \to \theta_{\Phi,s}$ with the projection onto the χ-eigenspace. We have $\mathscr{S}(E_{\mathbf{A}}, \mathscr{F}) = \otimes \mathscr{S}(E_v, \mathscr{F})$, where $\mathscr{S}(E_v, \mathscr{F})$ is the space of an oscillator representation $\omega_v(\gamma_v, \psi_v) = \omega_v(s_v, \psi_v)$ of G_v. Let $\mathscr{S}(\chi_v, \psi_v)$ be the maximal quotient of $\mathscr{S}(E_v, \mathscr{F}_v)$ on which Z_v acts according to χ_v. Then there exists a map $\otimes \mathscr{S}(\chi_v, \psi_v) \to \Theta(\gamma, \psi, \chi)$ such that the diagram

$$\begin{array}{ccc} \mathscr{S}(E_{\mathbf{A}}, \mathscr{F}) & \longrightarrow & \Theta(\gamma, \psi, \chi) \\ & \searrow \quad \nearrow & \\ & \otimes \mathscr{S}(\chi_v, \psi_v) & \end{array}$$

commutes.

Proposition 3.4.1. (1) *$\Theta(\gamma, \psi)$ is contained in $L^2(G(F) \setminus G(\mathbf{A}))$.*
(2) *For any χ, the closure of $\Theta(\gamma, \psi, \chi)$ in $L^2(G(F) \setminus G(\mathbf{A}))$ generates a non-zero irreducible automorphic representation.*

Proof. The square-integrability of the theta functions $\theta_{\Phi,s}$ on $G(F) \setminus G(\mathbf{A})$ follows immediately from the argument of [H_1, Theorem 2.5] , adapted to the case at hand. By [GR_1, § 2], each G_v-module $\mathscr{S}(\chi_v, \psi_v)$ is irreducible. Therefore, it remains to show that $\Theta(\gamma, \psi, \chi) \neq \{0\}$. For this purpose, we shall compute the Fourier- Jacobi coefficients of a theta series.

A typical element in $\Theta(\gamma, \psi)$ looks like

$$\theta_{\Phi,s}(g) = \sum_{x \in E} \theta_1((\omega_\psi^s(g)\Phi)(x)).$$

By (3.2.3), $(\omega_\psi^s([0,t])\Phi)(x) = \psi(tN_{E/F}(x))\Phi(x)$ for $t \in \mathbf{A}$, and therefore for $\delta \in F^*$,

$$(\theta_{\Phi,s})_\psi \delta(g) = \sum_{x \in E} \theta_1((\omega_\psi^s(g)\Phi)(x)) \int_{F\backslash \mathbf{A}} \psi(tN_{E/F}(x))\psi^\delta(-t)\,dt,$$

or,

$$(3.4.1) \qquad (\theta_{\Phi,s})_\psi \delta(g) = \begin{cases} \sum_{\substack{x \in E \\ N_{E/F}(x)=\delta}} \theta_1((\omega_\psi^s(g)\Phi)(x)) & \text{if } \delta \in N_{E/F}(F^*) \\ 0 & \text{otherwise.} \end{cases}$$

Now take $\Phi = \Phi_0 \otimes v_0$, where $v_0 \in \mathscr{F}$ and $\Phi_0 \in \mathscr{S}(E_\mathbf{A})$. For $z' = \mathrm{diag}\,(z,z,z) \in Z(\mathbf{A})$,

$$(\omega_\psi^s(z')\Phi)(x) = \gamma(z)\Phi_0(z^{-1}x)\omega_\psi^{s'}(z)(v_0).$$

Hence

$$(\theta_{\Phi,s})\psi(z') = \sum_{x \in E^1} \theta_1((\omega_\psi^s(z')\Phi)(x) = \sum_{x \in E^1} \gamma(z)\Phi_0(z^{-1}x)\theta_1(\omega_\psi^{s'}(z)(v_0))$$

and therefore

$$\begin{aligned} \theta_{\chi,s}(\Phi)\psi(1) &= \int_{E^1\backslash E^1_\mathbf{A}} \sum_{x \in E^1} \gamma(z)\Phi_0(z^{-1}x)\theta_1(\omega_\psi^{s'}(z)(v_0))\chi^{-1}(z)\,dz \\ &= \int \Phi_0(z^{-1})\theta_1(\omega_\psi^{s'}(z)(v_0))\gamma(z)\chi^{-1}(z)\,dz\,. \end{aligned}$$

We can choose v_0 so that $\theta_1(\omega_\psi^{s'}(z)(v_0))$ is not identically zero as a function of $E^1_\mathbf{A}$. It is then clear that the last integral above is non-zero form some choice of Φ_0. □

We denote the representation on $\Theta(\gamma,\psi,\chi)$ by $\omega(\gamma,\psi,\chi)$ and call it a Weil representation. Let Ω denote the set of all Weil representations.

We now introduce a parametrization of $R^\wedge$ by the data (γ,ψ,χ). First, observe that elements of $R^\wedge$ may be constructed as follows. Recall (§1.1) that N is the Heisenberg group attached to E, and let V be the space of a ψ-representation of $N(\mathbf{A})$. The metaplectic group $\mathrm{Mp}_\mathbf{A}(E)$ also acts on V, and we denote by ω the joint action of $N(\mathbf{A})\rtimes \mathrm{Mp}_\mathbf{A}(E)$ on V. Observe that $R = ZS'N$, where $S' = \{d(1,\beta,1) : \beta \in E^1\}$. The action of $S'(\mathbf{A})$ on $N(\mathbf{A})$ by conjugation yields an embedding of $S'(\mathbf{A})$ in $\mathrm{Sp}_\mathbf{A}(E)$. Fixing a compatible splitting $s' : S'(\mathbf{A}) \to \mathrm{Mp}_\mathbf{A}(E)$ (which exists by Proposition 3.1.1), we obtain an action of $S'(\mathbf{A})N(\mathbf{A})$ on V. Let θ_1 be a theta functional on V, and for all $v \in V$ define $\varphi_v(zsn) = \chi(z)\theta^1(\omega(sn)v)$ for $z \in Z(\mathbf{A})$, $s \in S'(\mathbf{A})n \in N(\mathbf{A})$. Then φ_v belongs to $L^2(R(F)\setminus R(\mathbf{A}))$, and the map $v \to \varphi_v$ is a embedding of $R(\mathbf{A})$-representations. The representation on the space of functions $\{\varphi_v\}$ will be denoted by $\tau(s',\psi,\chi)$. If v is a Hecke character of $R(F)N(\mathbf{A})\setminus R(\mathbf{A})$ of the form $v(zs'n) = v_1(z)v_2(s')$, then $\tau(s',\psi,\chi)\otimes v = \tau(s'\otimes v_2,\psi,\chi v_2)$, and hence as we vary (s',ψ,χ) we obtain all elements of $R^\wedge$ (cf. §2.5). Next, observe that the data

(s', ψ) determines $\omega_\psi^{s'}$, hence a mixed model for $G = U(3)$ as in § 3.2, and hence a Hecke character γ of E. The choices (s', ψ) or (γ, ψ) mutually determine each other and hence we can denote $\tau(s', \psi, \chi)$ by $\tau(\gamma, \psi, \chi)$.

Theorem 3.4. *Let $\pi = \omega(\gamma, \psi, \chi) \in \Omega$. Then*
(a) *$\pi \in \prod(\varrho)$, where $\varrho((h_2, h_1)) = \eta(\det(h_2))\,\eta'(\det(h_2)h_1)$ is determined by the formulas:*

$$\gamma = \mu\eta_E\eta'_E, \quad \chi(\gamma^1)^{-1} = \eta' \tag{3.4.2}$$

(b) *π is exceptional.*

Proof. For Part (a), let $\pi = \omega(\gamma, \psi, \chi)$. If v is non-split and π_v is unramified, then π_v is isomorphic to $\pi^n(\varrho_v)$ for some character of H_v by [GR_1], §8.1, and hence $\pi \in \prod(\varrho)$ for some automorphic character ϱ as in the proof of Theorem 2.4.1 above. This means that for almost all v, π_v is the Langlands quotient of the principal series representation of G_v induced from $\nu_v(d(\alpha, \beta, \bar{\alpha}^{-1})) = \mu_v(\alpha)\eta_v(\alpha/\bar{\alpha})\eta'_v((\alpha/\bar{\alpha})\beta)||\alpha||_v^{1/2}$ for some automorphic characters η, η', of E^1 (cf. 1.4). On the other hand, if $\chi_v(\gamma_v^1)^{-1}$ occurs in $(\omega_\psi^{S'})_v$, where γ^1 is the restriction of γ to E^1, then by [GR_1], Proposition 5.2.3, the Weil representation of π_v is the Langlands quotient of the principal series representation induced from $\xi_v(d(\alpha, \beta, \bar{\alpha}^{-1})) = \gamma_v(\alpha)\chi_v\gamma_v^{-1}(\beta)\,||\alpha||_v^{1/2}$. For such v, $\gamma_v = \mu_v\eta_{Ev}\eta'_{Ev}$. However, for any fixed v, we can choose a global χ so that $\chi_v(\gamma_v^1)^{-1}$ occurs in $(\omega_\psi^{S'})_v$. Thus $\gamma = \mu\eta_E\eta'_E$, and $\eta' = \chi(\gamma^1)^{-1}$.

To prove (b), suppose that $\tau_1, \tau_2 \in \Lambda(\pi) = \{\tau \in R^\wedge : \varphi_\tau \neq 0 \text{ for some } \varphi \in V_\pi\}$. By (3.4.1), the characters by which $U(\mathbf{A})$ acts on τ_1 and τ_2 lie in the same $B(F)$-orbit and hence we may assume they coincide. We then have $\tau_1 = \tau(\gamma_1, \psi, \chi)$ and $\tau_1 = \tau(\gamma_2, \psi, \chi)$ for some ψ, where χ is the central character of π. Now recall that $\pi_v = \omega(\gamma_v, \psi_v, \chi_v)$ is locally exceptional for each finite v. (In [Gr_1], we showed that π_v is locally exceptional for all finite non-split v; however the argument easily adapts to the split case.) Since τ_{1v} and τ_{2v} belong to $\Lambda(\pi_v)$, this implies $\gamma_{1v} = \gamma_{2v}$. But the γ_j are Hecke characters, and hence coincide.

It remains to check that $\varphi_U(g) = \varphi_N(g)$ for all $\varphi \in V_\pi$ and $g \in G(\mathbf{A})$. For this we may assume that $g = 1$. But by (3.2.3), $\omega_\psi^s([c, t])\Phi)(0) = \Phi(0)$, and by (3.4.1),

$$\varphi_U([c, t]) = \theta_1((\omega_\psi^s([c, t])\Phi)(0)) = \theta_1(\Phi)(0)).$$

Thus $\varphi_U(n)$ is independent of n, and

$$\varphi_N(1) = \int_{N(F)U(\mathbf{A})\backslash N(\mathbf{A})} \varphi_U(n)\,dn = \varphi_U(1)$$

provided that measures are suitably normalized. □

Remark. For $\tau = \tau(\gamma, \psi, \chi) \in R^\wedge$, set $\omega(\tau) = \omega(\gamma, \psi, \chi)$. Then the map $\tau \to \omega(\tau)$ induces a bijection between $B(F) \setminus R^\wedge$ and Ω. Indeed, if $\tau = \tau(\gamma, \psi, \chi)$, then $\tau^\delta = \tau(\gamma, \psi^\delta, \chi)$, and so we need only show that $\omega(\gamma, \psi, \chi) = \omega(\gamma', \psi', \chi')$ if and only if $\gamma' = \gamma, \chi' = \chi$, and $\psi' = \psi^\delta$ for some $\delta \in N_{E/F}(E^*)$. This is clear from the local bijection ([GR_1, Prop. 5.1.1]), but with the statement $\psi' = \psi^\delta$ for

some $\delta \in N_{E/F}(I_E)$. However, both ψ' and ψ are trivial on F, and hence δ belongs to F^*; since δ is also a norm locally everywhere, $\delta \in N_{E/F}(F^*)$.

4 Shimura integral

4.1 Eisenstein series. Let ξ be a unitary Hecke character of E. We regard ξ as a character of $B(F) \setminus B(\mathbf{A})$ by setting $\xi(d(\alpha, \beta, \bar{\alpha}^{-1})n) = \xi(\alpha)$ for $n \in N(\mathbf{A})$. Set $\alpha(d(\alpha, \beta, \bar{\alpha}^{-1})n) = \|\alpha\|$. Let $F = \prod F_v$ be a smooth function on K such that $F(bk) = \xi(b)F(k)$ for $b \in K \cap B(\mathbf{A})$ and for $s \in \mathbf{C}$, set $F(g, s, \xi) = \alpha(b)^{s+1}\xi(b)F(k)$. The Eisenstein series

$$E(g, s, \xi) = \sum_{\gamma \in B(F)\setminus G(F)} F(\gamma g, s, \xi)$$

converges for $\mathrm{Re}(s) \gg 0$ and has a meromorphic continuation to the complex s-plane. The possible poles of $E(g, s, \xi)$ are well- known. Set

$$R(s) = \frac{L(2s, \xi_F \omega_{E/F})L(s, \xi)}{L(2s+1, \xi_F \omega_{E/F})L(s+1, \xi)}$$

where ξ_F denotes the restriction of ξ to F. The set of poles of $E(g, s, \xi)$ is contained in the set of poles of $R(s)$ ([HC, Thm. 7, p. 105]). In particular, for $\mathrm{Re}(s) > 0$, the possibilities are

(1) $s = 1$, ξ trivial,

(2) $s = \frac{1}{2}$, $\xi_F = \omega_{E/F}$, $L(\frac{1}{2}, \xi) \neq 0$.

4.2 Let ω be a cuspidal representation of G. Let $\pi \in \Omega$. If θ is a theta series in V_ω and $\varphi \in V_\pi$, we define the Shimura integral

$$\mathrm{SH}(\varphi, \theta, \xi, s) = \int_{G(F)\setminus G(\mathbf{A})} \varphi(g)\,\overline{\theta(g)}\,E(g, s, \xi)\,dg\,.$$

The absolute and uniform convergence in compact subsets of this integral is clear from the rapid decrease of φ and the slow increase of $\theta(g)E(g, s, \xi)$.

Suppose that $\tau \in \Lambda(\pi)$, where we recall that $\Lambda(\pi)$ is the set of τ such that φ_τ is non-zero for some $\varphi \in V_\pi$. Factor the $R(\mathbf{A})$-map $L : \pi \to \tau$, $\varphi \to \varphi_\tau | R(\mathbf{A})$, as a product $\prod L_v$ as in the proof of Theorem 2.4.1, where $L_v : \pi_v \to \tau_v$. Set $\tau^* = \tau \otimes \xi_R$, where ξ_R is the restriction of ξ to $R(\mathbf{A})$, and assume that $\omega = \omega(\tau^*)$. Assume that φ and θ correspond to decomposible elements $x = \otimes x_v$ and $w = \otimes w_v$ with respect to decompositions $\pi = \otimes \pi_v$ and $\omega = \otimes \omega_v$, and define the local Fourier coefficients:

$$\varphi_v(g) = L_v(\pi_v(g)x_v), \quad \theta_v(g) = L_v^*(\omega_v(g)w_v),$$

where $L^* = \prod L_v^*$ is the $R(\mathbf{A})$-map $\omega(\tau^*) \to \tau^*$. let $\langle , \rangle_v$ be a non-zero Hermitian inner product on τ_v, which we view as a Hermitian ξ_{Rv}-pairing between τ_v and $\tau_v^* : \langle \tau(r)w, \tau^*(r)w^* \rangle_v = \xi_R(r)^{-1}\langle w, w^* \rangle_v$. Then $\langle , \rangle_v$ is unique up to multiples.

Let S be a finite set of places of F, including the archimedean places, such that $\pi_v, \omega_v, \tau_v, \xi_v$ and ψ_v are unramified for $v \notin S$. We also assume that F_v is the

characteristic function of K_v for $v \notin S$. As in the proof of Theorem 2.4.1, let $\{y_v^0\}_{v \notin S}$ be a collection of vectors in τ_v and assume that $\langle y_v^0, y_v^0 \rangle_v = 1$. We can then assume $L_v(x_v) = L_v^*(w_v) = y_v^0$ for almost all v. Note also that the integral on the left below,

$$\int_{R(F)\backslash R(\mathbf{A})} \varphi_\tau(r)\, \overline{\theta_\tau^*(r)}\, \xi_R(r)\, dr = \prod_v \langle L_v(x_v), L_v^*(w_v) \rangle_v,$$

defines a ξ_R-pairing between τ and τ^*. Since such a pairing is unique up to multiples, we can assume the local pairings $\langle , \rangle_v$ for $v \in S$ chosen so that the above equality holds (with almost all terms in the product on the right equal to one).

Proposition 4.2.1. *For* $\mathrm{Re}(s) \gg 0$, $\mathrm{SH}(\varphi, \theta, \xi, s)$ *is equal to the product* $\prod \mathrm{SH}(\varphi_v, \theta_v, \xi_v, s)$ *where*

$$\mathrm{SH}(\varphi_v, \theta_v, \xi_v, s) = \int_{R_v \backslash G_v} \langle \varphi_v(g), \theta_v(g) \rangle_v\, F_v(g, s, \xi)\, dg\,.$$

Proof. We follow [PS], part III (cf. also [GP_1]). From the definition of $E(g, s, \xi)$, we have

$$\mathrm{SH}(\varphi, \theta, \xi, s) = \int_{B(F)\backslash G(\mathbf{A})} \varphi(g)\, \overline{\theta(g)}\, F(g, s, \xi)\, dg\,.$$

Since θ is exceptional

$$\theta(g) = \theta_N(g) + \sum_{b \in R(F)\backslash B(F)} \theta_{\tau^*}(bg)\,.$$

The constant term $\theta_N(g)$ contributes nothing to the Shimura integral, since

$$\int_{B(F)\backslash G(\mathbf{A})} \varphi(g)\, \overline{\theta_N(g)}\, F(g, s, \xi)\, dg = \int_{B(F)N(\mathbf{A})\backslash G(\mathbf{A})} \overline{\theta_N(g)}\, F(g, s, \xi) \left[\int_{N(F)\backslash N(\mathbf{A})} \varphi(ng)\, dn \right] dg$$

and the inner integral vanishes for cuspidal φ. Therefore

$$\mathrm{SH}(\varphi, \theta, \xi, s) = \int_{R(F)\backslash G(\mathbf{A})} \varphi(g)\, \overline{\theta_{\tau^*}(g)}\, F(g, s, \xi)\, dg\,.$$

Since $F(rg, s, \xi) = \xi_R(r) F(g, s, \xi)$,

$$\mathrm{SH}(\varphi, \theta, \xi, s) = \int_{R(\mathbf{A})\backslash G(\mathbf{A})} \langle \varphi_\tau(g), \theta_{\tau^*}(g) \rangle F(g, s, \xi)\, dg$$

where we set

$$\begin{aligned} \langle \varphi_\tau(g), \theta_{\tau^*}(g) \rangle &= \int_{R(F)\backslash R(\mathbf{A})} \varphi(rg)\, \overline{\theta_{\tau^*}(rg)}\, \xi_R(r)\, dr \\ &= \int_{R(F)\backslash R(\mathbf{A})} \varphi_\tau(rg)\, \overline{\theta_{\tau^*}(rg)}\, \xi_R(r)\, dr\,. \end{aligned}$$

The second equality follows from the Fourier-Jacobi expansion of φ and the fact that θ_{τ^*} lies in the τ^*-subspace of $L^2(R(F)\backslash R(\mathbf{A}))$. By the remarks preceding the statement of the proposition, $\langle \varphi_\tau(g), \theta^*_\tau(g)\rangle = \prod \langle L_v(\pi(g_v)x_v), L^*_v(\pi(g_v)y_v)\rangle_v$. □

4.3 We continue with the notation of the previous section. By γ we denote the Hecke character of E such that $\omega = \omega(\gamma, \psi, \chi)$.

Theorem 4.3.1. *Assume that* $\tau \in \Lambda(\pi)$. *Then for* $\mathrm{Re}(s) \gg 0$,

$$\mathrm{SH}(\varphi, \theta, \xi\gamma, s) = c_S(s)\frac{L_S(s+\tfrac{1}{2}, \pi \otimes \xi)}{L_S(2s+1, \xi_F)L_S(s+1, \xi\gamma)}$$

where $c_S(s) = c\prod_{v\in S} \mathrm{SH}(\varphi_v, \theta_v, \xi_v\gamma_v, s)$ *and* c *is a non-zero constant.*

Proof. We have $dg = \alpha(m)^{-2}dmdndk$. For $v \notin S$, the integral over $R_v \backslash G_v$ reduces to

$$\mathrm{SH}(\varphi_v, \theta_v, \xi_v\gamma_v, s) = \mathrm{meas}\,(K_v)\int_{S_v\backslash M_v} \langle \varphi_v(m), \theta_v(m)\rangle_v\, \alpha_v(m)^{s-1}\xi_v(m)\gamma_v(m)\, d_v m.$$

The integral on the right here has been calculated in [GR_1, Theorems 8.2.6 and 8.7.1]. Since we have the normalization $\langle L_v(x_v), L^*_v(y_v)\rangle v = 1$, the result is

$$\mathrm{SH}(\varphi_v, \theta_v, \xi_v\gamma_v, s) = c_v\frac{L_v(s+\tfrac{1}{2}, \pi \otimes \xi)}{L_v(2s+1, \xi_F)L_v(s+1, \xi\gamma)}$$

where c_v depends on the normalizations of dk_v and $d_v m$. □

Proposition 4.3.2. *Given* π *and* ξ, *we may choose* φ_v, θ_v *and* F_v *for* $v \in S$ *so that* $c_S(s)$ *is non-zero for all* s.

Proof. $\mathrm{SH}(\varphi_v, \theta_v, \xi_v\gamma_v, s)$ is equal to

$$\int_{K_v}\left[\int_{S_v\backslash M_v} \langle \varphi_v(mk), \theta_v(mk)\rangle_v \alpha_v(m)^{s-1}\xi_v(m)\gamma_v(m)\, dm\right] F_v(k)\, dk. \tag{4.3.1}$$

We will first show that there exists data φ_v, θ_v and F_v so that $\langle \varphi_v(m), \theta_v(m)\rangle_v$ is compactly supported modulo S_v and the integral

$$\int_{S_v\backslash M_v} \langle \varphi_v(m), \theta_v(m)\rangle_v \alpha_v(m)^{s-1}\xi_v(m)\gamma_v(m)\, dm \tag{4.3.2}$$

(which consequently converges absolutely for all s) is non-zero for all s. By translation, we may assume that $\theta_v(1) \neq 0$. Furthermore, since τ_v is irreducible, we may, replacing φ_v by a linear combination of its translates if necessary, choose φ_v so that $\langle \varphi_v(1), \theta_v(1)\rangle_v \neq 0$. Next, let $\phi(u)$ be a Schwartz-Bruhat function on F_v, which we identify with U_v, and for x in V_{π_v}, set

$$x^* = \int_{U_v} \phi(u)\pi([0,u])x\, du.$$

For archimedean v, we take x to be a smooth vector, by [JS, Prop. 2.4] x^* is again smooth. In general,

$$\varphi_v^*(m) = L_v(\pi(m)x^*) = \phi^\wedge(\alpha\alpha)L_v(\pi(m)x)$$

where $\phi^\wedge(t) = \int \phi(x)\psi(tx)\,dx$. It is thus clear that we may replace $\varphi_v(m)$ by $\phi^\wedge(\alpha(m)\varphi_v(m))$, and for $\hat{\phi}$ with support in a small neighborhood of 1, (4.3.2) is non-zero.
Now $F_v(k)$ may be an arbitrary smooth function on K_v subject to the condition $F_v(bk) = \xi(b)F_v(k)$ for $b \in B_v \cap K_v$. Therefore (4.3.1) is non-zero for a suitable choice of F_v. □

5 Characterization of A-packets

5.1 Suppose that ϱ is an automorphic character of H given by

$$\varrho((h_2, h_1)) = \eta(\det(h_2))\,\eta'(\det(h_2)h_1). \tag{5.1.1}$$

If $\pi \in \prod(\varrho)$, then

$$L\left(s+\frac{1}{2}, \pi \otimes \xi\right) = L_E(s+1, \mu\eta_E\eta'_E\xi)L_E\left(s+\frac{1}{2}, \eta'_E\xi\right)L_E(s, \mu\eta_E\eta'_E\xi).$$

Theorem 5.1.1. *Let π be a discrete representation of G. Then the following are equivalent:*
(a) *π is a Weil representation $\omega(\gamma, \psi, \chi)$.*
(b) *π is exceptional*
(c) *$\pi \in \prod(\varrho)$ for some automorphic character ϱ of H.*
If (a)–(c) *hold, the relation between (γ, χ) and ϱ is:*

$$\gamma = \mu\eta_E\eta'_E \quad \text{and} \quad \chi = \gamma^1\eta', \tag{5.1.1}$$

where γ^1 is the restriction of γ to the norm one elements. In particular, $\prod(\varrho) = \{\omega(\gamma, \psi, \chi)\}$, where ψ runs through a set of representatives for the characters of $F \setminus \mathbf{A}$ modulo $N_{E/F}(E^*)$.

Remark. In terms of Langlands functoriality, this result says that for fixed γ, the set of theta-liftings of χ, relative to the splittings determined by (γ, ψ), coincides with the A-packet obtained as the endoscopic transfer of the following one-dimensional representation ϱ of H. Let ϱ_2 be the character of E^1 such that $\varrho_2(z/\bar{z}) = \mu^{-1}\gamma(z)$ for $z \in E^*$ and let $\varrho_1(z) = \chi(\gamma^1)^{-1}(z)$. Then $\varrho((h_2, h_1)) = \varrho_2(\det(h_2))\varrho_1(h_1)$.

Proof. We have already shown (*a*) ⇒ (*b*) ⇒ (*c*), and shall now show (*c*) ⇒ (*a*). Suppose first that $\pi \in \prod(\varrho)$ is cuspidal. Choosing data as in §4, we have for $\mathrm{Re}(s) \gg 0$

$$\mathrm{SH}(\varphi, \theta, \xi\gamma, s) = c_S(s)\frac{L_S(s+\frac{1}{2}, \pi \otimes \xi)}{L_S(2s+1, \xi_F)L_S(s+1, \xi\gamma)} \tag{5.1.2}$$

where $c_S(s)$ is a nowhere vanishing analytic function of s. The equality therefore holds for all s by meromorphic continuation of both sides. If ϱ is given by (5.1.1), then $L_S(s+\frac{1}{2}, \pi \otimes \xi)$ has a simple pole at $s=1$ if $\xi \mu \eta_E \eta'_E$ is trivial (note that $L_S(s+1, \mu\eta_E\eta'_E\xi)L_S(s+\frac{1}{2}, \eta'_E\xi)$ is finite and non-zero at $s=1$). But in this case, the left-hand side of (5.1.2) has a simple pole at $s=1$, which implies that $\xi\gamma$ is trivial, i.e., $\gamma = \mu\eta_E\eta'_E$. Furthermore, the residue of $E(g,1,s)$ at $s=1$ is a (non-zero) constant, and we may conclude that

$$\int_{G(F)\backslash G(\mathbf{A})} \varphi(g)\theta(g)\,dg \neq 0.$$

We have now shown that $(c) \Rightarrow (a)$ for cuspidal π. To prove it for all discrete π, we need a local analogue:

Lemma 5.1.2. *For all v, $\prod(\varrho_v) = \{\omega(\gamma_v, \psi_v, \chi_v)\}$, where ψ_v ranges over a set of representatives for additive characters modulo $N_{E/F}(E_v^*)$.*

Proof. If v is split, $\prod(\varrho_v)$ is a singleton and the Lemma is obvious. For non-split v, let (γ, ψ, χ) be global data which specializes to $(\gamma_v, \psi_v, \chi_v)$ at v, and define ϱ in terms of (γ, χ). We know that $\omega(\gamma, \psi, \chi) \in \prod(\varrho)$ and hence locally $\omega(\gamma_v, \psi_v, \chi_v) \in \prod(\varrho_v)$. On the other hand, for any finite set X of non-split places v, let π_X be the representation $\otimes \pi_{X_v}$ in $\prod(\varrho)$ such that $\pi_{X_v} = \pi^s(\varrho_v)$ if and only if $v \in X$. By the multiplicity rule of [R_2] (see our Introduction), there exists an ε equal to ± 1 such that π_X is discrete if and only if $(-1)^{|X|} = \varepsilon$. If π_X is discrete and $|X| \neq 0$, then π_X is cuspidal, and hence of the form $\omega(\gamma, \psi, \chi)$ for some ψ by what we have shown so far. Thus it is clear from the multiplicity rule that for all $\pi' \in \prod(\varrho_v)$, there exists a cuspidal $\pi \in \prod(\varrho)$ such that $\pi_v = \pi'$. However, $\pi = \omega(\gamma, \psi, \chi)$ for some ψ, and hence $\pi' = \omega(\gamma_v, \psi_v, \chi_v)$ for some ψ_v.

Suppose finally that $\pi \in \prod(\varrho)$ is discrete but not cuspidal. Then $\pi_v = \pi^n(\varrho_v)$ for all v. Let X be any non-empty set of non-split places of even cardinality, so that π_X is cuspidal and hence of the form $\omega(\gamma, \psi, \chi)$. Let ψ' be an additive character of $F \setminus \mathbf{A}$ whose class modulo $N_{E/F}(E^*)$ differs from that of ψ precisely at the places of X. Then $\omega(\gamma_v, \psi'_v, \chi_v)$ differs from $\omega(\gamma_v, \psi_v, \chi_v)$ at precisely the places in X. But then $\omega(\gamma_v, \psi'_v, \chi_v)$ is isomorphic to π_v for all places v, and so by multiplicity one for $G([R]), \pi = \omega(\gamma, \psi', \chi)$. □

Remark. We note $(c) \Rightarrow (a)$ in Theorem 5.1.1 could be proved without recourse to the Shimura integral, if we had an independent prove of Lemma 5.1.2 (which itself reduces to a statement about the Howe correspondence pair $(U(1), U(1))$ in the local case, namely, that for given γ and χ, there exists a unique class of ψ such that χ occurs in $\omega^1(\gamma, \psi)$, cf. Corollary 5.2.2).

5.2 θ-series liftings. We recall the notion of theta-series lifting. Suppose that G and H are unitary groups in n and m variables, respectively. Then there exists an embedding s of $G \times H$ in the metaplectic group on $2nm$ variables (the obvious analogue of Prop. 3.1.1 holds for $G \times H$) and we may consider the restrictions to $G \times H$ of metaplectic theta series relative to s and ψ. Let $\vartheta(g,h)$ be such a theta function. Let π and τ be irreducible cuspidal representations of G and H, respectively. We say that π is a theta-series lifting of τ if V_τ is

contained in the space $\Theta(\psi, s, \tau)$ of functions of the form

$$g \longrightarrow \int_{H(F)\backslash H(\mathbf{A})} \vartheta(g,h) f(h)\, dh$$

for some $\vartheta(g,h)$ and some $f \in V_\tau$. It is known that if $\Theta(\psi, s, \tau)$ is cuspidal, then it is irreducible (cf. [Ra]).

We have seen that the choice of data (γ, ψ) simultaneously fixes embeddings of $U(1)$ and $U(3)$ in their corresponding metaplectic groups. We denote the corresponding oscillator representations by $\omega^1(\gamma,\psi)$ and $\omega^3(\gamma,\psi)$. For any automorphic character χ of $U(1)$, let $\theta^1(\gamma,\psi,\chi)$ and $\theta^3(\gamma,\psi,\chi)$ be the space of associated theta functions on $U(1)$ and $U(3)$, respectively.

Proposition 5.2.1. *The following three conditions are equivalent:*

(1) $\theta^3(\gamma,\psi,\chi)$ *is cuspidal*

(2) $\theta^1(\gamma,\psi,\chi(\gamma^1)^{-1} = \{0\}$

(3) *Either there exists a place* v *such that* $\chi_v(\gamma_v^1)^{-1}$ *does not occur in* $\omega^1(\gamma_v,\psi_v)$, *or*

$$L\left(\frac{1}{2}, \gamma^3 \chi_E^{-1}\right) = 0.$$

Proof. By the calculation of § 3.4,

$$\theta_{\chi,s}(\Phi)_U(1) = \left[\int_{E^1\backslash E^1_{\mathbf{A}}} \theta_1(\omega_\psi^{s'}(z)(v_0))\gamma(z)\chi^{-1}(z)\, dz\right]\Phi_0(0),$$

where $\Phi(x) = \Phi_0(x)v_0$, $\Phi_0 \in \mathscr{S}(E_A)$, $v_0 \in \mathscr{F}$. The integral is, by definition, the theta-series lifting of $\gamma^1\chi^{-1}$ from $U(1)$ to $U(1)$. Since Φ_0 is arbitrary, the equivalence of (1) and (2) is immediate.

To establish the equivalence of (1) and (3), recall from [GR_1, Prop. 5.2.2], that $\omega^3(\chi_v,\psi_v,\chi_{\bar{v}})$ is supercuspidal if and only if $\chi_v(\gamma_v^1)^{-1}$ does not occur in $\omega^1(\gamma_v,\psi_v)$.

Now let $\pi = \theta^3(\gamma,\psi,\chi)$ and let ϱ be as in Theorem 5.1.1(c). Assume π is cuspidal. If $\chi_v(\gamma_v^1)^{-1}$ occurs in $\omega^1(\gamma,\psi)$ for all v, then $\pi_v = \pi^n(\varrho_v)$ for all v, and then π is cuspidal if and only if $L(\frac{1}{2}, \mu\eta_E) = 0$ by [R, Theorem 13.3.6(b)]. (This last condition is equivalent to $L(\frac{1}{2}, \gamma^3\chi_E^{-1}) = 0$, since $(\gamma^1)_E = \gamma^2$.) Finally, if $\chi_v(\gamma_v^1)^{-1}$ does not occur in $\omega^1(\gamma_v,\psi_v)$ for some v, then $\pi_v = \pi^s(\varrho_v)$ for some v and π is cuspidal. □

For any dual reductive pair, there is the notion of the Howe lifting, locally and globally [MVW]. In the context of $(U(1), U(1))$, the Howe lift of χ relative to $\omega^1(\gamma,\psi)$ is defined only if χ occurs in $\omega^1(\gamma,\psi)$, in which case the Howe lift is χ itself. In particular, a Howe lift is defined globally if and only if it is defined locally everywhere.

Corollary 5.2.2. *Let* χ *be an automorphic character of* $U(1)$. *For all* γ, *there exists a* ψ *such that the corresponding theta lift* $\theta^1(\gamma,\psi,\chi)$ *is non-zero if and only if* $L(\frac{1}{2}, \gamma\chi_E^{-1}) \neq 0$. *(When such a* ψ *exists, it is unique modulo norms.)*

Remarks. (1) The following additional result about $(U(1,), U(1))$ liftings is shown in [R_2]: Given χ and γ as in Corollary 5.2.2, there exists a ψ such that χ_v "occurs" in $\omega^1(\gamma_v, \psi_v)$ for all v if and only if $\varepsilon(\frac{1}{2}, \gamma\chi_E^{-1}) = +1$.
(2) Let ξ be a Hecke character of E whose restriction to F is $\omega_{E/F}$. If $E(g, s, \xi)$ has a non-trivial residue at $s = 1/2$, than the residue generates a representation $\pi(\xi)$ of the form $\omega(\gamma, \psi, \chi)$ with $(\gamma, \chi) = (\xi, \xi^1)$. By the multiplicity one theorem for $U(3)$, we obtain the following Siegel-Weil relation: $\pi(\xi) = \omega(\xi, \xi^1, \psi)$.
(3) In [GP, p. 167], it is stated $\omega(\gamma, \chi, \psi)$ cannot be cuspidal if it is of the form $\otimes\pi^n(\varrho_v)$. This is obviously incorrect, in view of the above results.

6 Fourier-Jacobi coefficients of endoscopic forms

6.1 In this section, let $\varrho = \varrho_2 \times \varrho_1$ be a cuspidal L-packet on H.

Theorem 6.1.1. *Assume that $\pi \in \prod(\varrho)$ is cuspidal. Let $\tau = \tau(\gamma, \psi, \chi)$ and suppose that $\varphi_\tau \neq 0$ for some $\varphi \in V_\pi$. Then*
(1) $L(\frac{1}{2}, \gamma(\varrho_{1E})^{-1} \neq 0$.
(2) *π is a theta-series lifting of a cuspidal representation on a unitary group $U(\Phi')$ in two variables. The Hermitian form Φ' belongs to the class of* $\operatorname{diag}(-1, \delta)$, *where δ belongs to the unique class in $F^*/N_{E/F}(E^*))$ such that ϱ_1 occurs in $\omega^1(\gamma, \psi_\delta)$.*

Proof. It follows from the definition of the embedding ξ_H that

$$L_S\left(s + \frac{1}{2}, \pi \otimes \xi\right) = L_S\left(s + \frac{1}{2}, \varrho_{2E} \otimes \mu\xi\right) L_S\left(s + \frac{1}{2}, \varrho_{1E}\xi\right).$$

If ϱ_{2E} is cuspidal, then $L_S(s + \frac{1}{2}, \varrho_{2E} \otimes \mu\xi)$ does not vanish at $s = \frac{1}{2}$ ([JS]). Similarly, if ϱ_{2E} is not cuspidal, then there are unitary Hecke characters θ and θ' such that $L_S(s+\frac{1}{2}, \varrho_{2E}\otimes\mu\xi)$ is equal to the product $L_S(s+\frac{1}{2}, \theta\otimes\xi)L_S(s+\frac{1}{2}, \theta'\otimes\xi)$ ([R]), which again does not vanish at $s = \frac{1}{2}$. In either case, $L_S(s + \frac{1}{2}, \pi \otimes \xi)$ has a pole at $s = \frac{1}{2}$ if $\xi\varrho_{1E}$ is trivial.
Set $\eta = \gamma(\varrho_{1E})^{-1}$. Then Theorem 4.3.1 and Proposition 4.3.2 imply that $\mathrm{SH}(\varphi, \theta, \eta, s)$ also has a pole at $s = \frac{1}{2}$ for a suitable choice of data, and hence so does $E(g, \eta, s)$. By the result quoted in §4.1, this implies that $L(\frac{1}{2}, \eta) \neq 0$. Moreover, the pole of $E(g, \eta, s)$ is simple and its residue is a theta series belonging to $\omega(\eta, \eta^1, \psi^\delta)$, where δ belongs to the unique coset in $F^*/N_{E/F}(E^*)$ such that η^1 occurs in $\omega^1(\eta, \psi^\delta)$ (cf. remark of §5.2). Therefore

$$\int_{G(F)\backslash G(\mathbf{A})} \varphi(g)\,\overline{\theta(g)}\,\theta^*(g)\,dg \neq 0.$$

To prove (2), we note that $\overline{\theta(g)}\,\theta^*(g)$ is equal to $\vartheta(g, 1)$, where $\vartheta(g, h)$ is a theta series on $G \times U$ relative to some metaplectic embedding s of $G \times U$ and additive character ψ, and where U is a unitary group in two variables. In fact, it is

easy to check that $U = U(\Phi')$, where Φ' is the Hermitian form $\mathrm{diag}(-1, \delta)$. This shows that the space of theta-series $\Theta(\psi, s, \pi)$ on $U(\Phi')$ is non-zero. Using the Schrödinger model, one can see that $\Theta(\psi, s, \pi)$ generates an irreducible cuspidal representation σ of $U(\Phi')$, which in this case theta-series lifts back to an irreducible *cuspidal* representation of G[1]. Using "multiplicity one" for G, it follows that π itself must be the theta-series lift of σ. □

Remark. In general, one cannot expect that all cuspidal members π of $\prod(\varrho)$ will be theta-series lifts from the quasi-split group $U(2)$. However, if π has a Whittaker model, then π is in fact a theta-series lift from the quasi-split group $U(2)$ ([GP$_2$]).

6.2 By the above theorem, every cuspidal element in $\prod(\varrho)$ is a theta-series lift from some $U(\Phi')$. We now check that the converse holds.

Proposition 6.2.1. *Let σ be a cuspidal representation of $U(\Phi')$, and s a splitting as above. Then all irreducible cuspidal subrepresentations π of $\Theta(\psi, s, \sigma)$ belong to $\prod(\varrho)$ for some ϱ.*

Together with Theorem 5.1.1, this shows that the set of discrete endoscopic representations of $U(3)$ coincides with the set of theta series lifts from unitary groups $U(\Phi')$ where Φ' is a Hermitian form in 1 or 2 variables.

Proof. For almost all v, π_v is the Howe lift of σ_v and hence, by the strong multiplicity one theorem for L-packets ([R]), it suffices to prove the following proposition.

Lemma 6.2.2. *Let σ be a cuspidal representation of $U(\Phi')$, where Φ' is a Hermitian form in two variables. Then there exists a cuspidal representation ϱ such that for almost all v, the Howe lift of σ_v belongs to $\prod(\varrho_v)$. More precisely, in the proof below, we define Hecke characters γ_2 of E^* and γ_3 of E^1, where the restriction of γ_2 to F^* is $\omega_{E/F}$. Let ξ be the character of E^1 defined by the relation $\mu^{-1}\gamma_2^{-1}(\alpha) = \xi(\alpha/\bar{\alpha})$. Let σ' be the cuspidal L-packet on $U(2)$ corresponding to σ by the Jacquet- Langlands correspondence for two variable unitary groups. Then $\varrho = \varrho_2 \times \varrho_1$ where $\varrho_2 = \sigma' \otimes (\xi\gamma_3 \circ \det)$ and $\varrho_1 = \gamma_3$.*

Proof. We recall that the Howe correspondence is known for almost all v. To check the proposition, we need only calculate the Howe lift locally of unramified representations from $U(2)$ to $U(3)$ (and GL_2 to GL_3 for the split places). We carry out the calculation for the pair $(U(2), U(3))$ by considering the following linear functional:

$$\ell(\phi) = L(v)^{-1} \int_{E^*} \phi(t(w_{-1} \otimes e_1))v(t)\, d^*t$$

Here ϕ lies in the Schwartz space $\mathcal{S}(w_{-1} \otimes V)$, where $V = E^3$ in the space of $U(3)$. By W, we denote the space of $U(2)$, by w_{-1} an α-eigenvector of $\mathrm{diag}(\alpha, \bar{\alpha}^{-1})$ in W, and by $\{e_1, e_2, e_3\}$ be the standard basis of V so that e_1 is an α-eigenvector of $\mathrm{diag}(\alpha, \beta, \bar{\alpha}^{-1})$. Finally, v is a character of E^* such that

[1] That the theta-lift of σ back to G is again cuspidal follows from Theorem 13.3.2 of [R] *plus* a computation expressing the Jacquet module of the Howe lift of almost every σ_v (to G_v) in terms of its (Howe) lift down to $U(1)$; this argument was kindly pointed out to us by D. Soudry.

$\mathrm{Re}(v) > 0$ and $L(v)$ is the standard local L-factor associated to v. As is well-known, the above integral extends to a holomorphic function of v and ℓ defines a non-zero linear functional for all v by analytic continuation.

We view $\mathscr{S}(w_{-1} \otimes V)$ as a Schrödinger model of the oscillator representation $\omega(h, g)$ of $U(2) \times U(3)$ relative to the polarization $(w_{-1} \otimes V, w_1 \otimes V)$ of $W \otimes V$, where w_1 is an isotropic vector fixed by the lower triangular unipotent subgroup of $U(2)$. Implicit here is a choice of embedding s of $U(2) \times U(3)$ in $\mathrm{Mp}(W \otimes V)$. To obtain s, we view $U(2) \times U(3)$ as a subgroup of $U(W \otimes V)$ and embed $U(W \otimes V)$ in $\mathrm{Mp}(W \otimes V)$ (cf. Proposition 3.1.1).

The above set-up also makes sense globally. Let T be the torus in $U(V \otimes W)$ fixing the vectors $w_i \otimes e_j$ and let

$$d = \mathrm{diag}\,(\alpha_1, \alpha_2, \alpha_3, \bar{\alpha}_1^{-1}, \bar{\alpha}_2^{-1}, \bar{\alpha}_3^{-3})$$

relative to the basis $w_{-1} \otimes e_1$, $w_{-1} \otimes e_2$, $w_{-1} \otimes e_3$, $w_1 \otimes e_1$, $w_1 \otimes e_2$, $w_1 \otimes e_3$. Then

$$\omega(d)\phi(w_{-1} \otimes v) = \gamma(\alpha_1\alpha_2\alpha_3)\, \|\alpha_1\alpha_2\alpha_3\|^{1/2}\phi(w_{-1} \otimes dv)$$

where γ is a character of the type considered in § 3.2, namely, γ is a Hecke character of E^* whose restriction to F^* is $\omega_{E/F}$. We also obtain the formulas

$$\omega(\mathrm{diag}\,(\alpha, \bar{\alpha}^{-1}), 1)\phi(w_{-1} \otimes e_1) = \gamma_2(\alpha)\, \|\alpha\|^{-3/2}\phi(\alpha^{-1}w_{-1} \otimes e_1)$$

and

$$\omega(1, \mathrm{diag}\,(\alpha, \beta, \bar{\alpha}^{-1}))\phi(w_{-1} \otimes e_1) = \gamma_3((\alpha/\bar{\alpha})\beta)\phi(\alpha^{-1}w_{-1} \otimes e_1)$$

where $\gamma_2 = \gamma$ and $\gamma_3 = \gamma^1$. A different choice of s changes γ_2 by a Hecke character of the form $\chi(\alpha/\bar{\alpha})$ and γ_3 by a Hecke character of E^1.

Returning to the local case, it is straightforward to check that $\ell(\omega(h, g)\phi)$ belongs to the representation $\mathrm{Ind}_{U(2)}(v\gamma_2) \otimes \mathrm{Ind}_G(v \otimes \gamma_3 \circ \det)$ of $U(2) \times G$, where the first factor is induced from the lower triangular Borel subgroup of $U(2)$ and the second factor is induced from the upper triangular Borel subgroup of G. Hence we obtain a non-zero element in $\mathrm{Hom}_{G \times H}(\omega, \mathrm{Ind}_{U(2)}(v\gamma_2) \otimes \mathrm{Ind}_G(v \otimes \gamma_3 \circ \det))$. Assume that v, γ_2, and γ_3 are unramified, and that the residual characteristic of v is different from 2. (In this case, the Howe correspondence is known (cf. [Wd$_2$]). Let $\pi_{U(2)}(v\gamma_2)$ and $\pi_G(v \otimes \gamma_3 \circ \det)$ denote the unramified constituents of $\mathrm{Ind}_{U(2)}(v\gamma_2)$ and $\mathrm{Ind}_G(v \otimes \gamma_3 \circ \det)$, respectively. We want to show that $\pi_G(v \otimes \gamma_3 \circ \det)$ is the Howe lift of $\pi_{U(2)}(v\gamma_2)$. When $\mathrm{Ind}_{U(2)}(v\gamma_2)$ and $\mathrm{Ind}_G(v \otimes_3 \circ \det)$ are irreducible, this is clear. To handle the general case, we argue as follows. We may assume the v is chosen within its Weyl group orbit so that $\pi_G(v \otimes \gamma_3 \circ \det)$ is a quotient of $\mathrm{Ind}_G(v \otimes \gamma_3 \circ \det)$. Let ϕ_0 be the characteristic function of $\{w_{-1} \otimes v : v \in \mathcal{O}^3\}$ and let K_v be the stabilizer of $\mathcal{O}^3$ in G_v. Then $\ell(\omega(h, g)\phi_0)$ is non-zero and invariant under $g \to gk$ for $k \in K_v$, and hence ℓ yields a non-trivial element of

$$\mathrm{Hom}_{G \times H}(\omega, \mathrm{Ind}_{U(2)}(v\gamma_2) \otimes \pi_G(v \otimes \gamma_3 \circ \det)).$$

It follows $\pi_G(v \otimes \gamma_3 \circ \det)$ is paired with a constituent of $\mathrm{Ind}_{(2)}(v\gamma_2)$ by the Howe correspondence and this constituent must also be unramified ([H$_2$, Theorem 7.1]). This shows that the Howe lift of $\pi_{U(2)}(v\gamma_2)$ is $\pi_G(v \otimes \gamma_3 \circ \det)$, or equivalently, the Howe lift of $\pi_{U(2)}(v)$ is $\pi_G(v\gamma_2^{-1}) \otimes \gamma_3 \circ \det$.

On the other hand, the endoscopic transfer of the representation $\pi_{U(2)}(v) \otimes \chi$ of $H = U(2) \times U(1)$ relative to the L-map ${}^L H \to {}^L G$ defined by μ is $\pi_G(\chi')$, where $\chi'(\mathrm{diag}(\alpha, \beta, \alpha)) = v\mu(\alpha)\chi(\beta)$. Define $\varrho = \varrho_2 \otimes \varrho_1$ by $\varrho_2 = \sigma' \otimes (\xi\gamma_3 \circ \det)$, where $\mu^{-1}\gamma_2^{-1}(\alpha) = \xi(\alpha/\bar{\alpha})$ and let $\varrho_1 = \gamma_3$. By the above computation, the Howe lift of σ_v belongs to $\prod(\varrho_v)$ if v remains prime and σ_v is a principal series representation.

One can argue similarly that this holds if v splits. In this case, identify H and G with $\mathrm{GL}_2(F_v)$ and $\mathrm{GL}_3(F_v)$, respectively, by choosing a place w of E above v, and view γ_{2w} and γ_{3w} as characters of F_v^*. If suffices to calculate the Howe lifting determined by the oscillator representation of $\mathrm{GL}_2 \times \mathrm{GL}_3$ on $\mathscr{S}(M_{2\times 3}(F))$ defined by $\omega(h,g)\phi(X) = |\det(h)^{3/2}|\det(g)|\phi({}^t hXg)$. If σ'' is an irreducible representation of GL_2, define a map from $\mathscr{S}(M_{2\times 3}(F))$ to σ'' by:

$$F(g,\phi,v) = \int_{\mathrm{GL}_2(F)} \omega(h,g)\phi[I_2,0])\sigma(h)v\,dh$$

for $v \in V^{\sigma''}$. This is an integral of Godement-Jacquet type in the range of absolute convergence and for all $\phi, v, F(g,\phi,v)$ belongs to the space of the representation π induced from $\sigma^* \otimes 1$ on the lower (2,1)-parabolic subgroup P of GL_3. Here σ^* is the contragredient of σ''. It follows that $\mathrm{Ind}_P(\sigma \otimes 1)$ is the Howe lift of σ'' (cf. [MVW]). □

References

[A] Arthur, J.: Unipotent representations: conjectures. In: Orbites unipotentes et répresentations, II. (Asterisque 171–172, pp. 13–71) Paris: 1989

[BW] Borel, A. and Wallach, N.: Continuous cohomology, discrete subgroups and representations of reductive groups. (Annals of Math. Studies, vol. 94) Princeton: Princeton University Press 1980

[GP_1] Gelbart, S., Piatetski-Shapiro, I.: Distinguished representations and modular forms of half-integral weight. Invent. Math. **59,** 145–188, (1980)

[GP_2] Gelbhart, S., Piatetski-Shapiro, I.: Automorphic forms and L-functions for the unitary group. In: Lie group representations II. (Lect. Notes Math., vol. 1041, pp. 141–184) Berlin Heidelberg New York: Springer 1984

[GR_1] Gelbart, S., Rogawski, J.: Expectional representations and Shimura's intgral for the local unitary group $U(3)$. In: Festschrift in honor of Piatetski-Shapiro. (Israel Math. Conf. Proc., vol. 2, pp. 19–75) 1990

[GR_2] Gelbart, S., Rogawski, J.: Heisenberg models for $U(3)$. (in preparation)

[GIR] Glauberman, G., Rogawski, J.: On theta functions with complex multiplication. J. Reine Angew. Math. **395,** 68–101 (1989)

[HC] Harish-Chandra: Automorphic forms on semisimple Lie groups. (Lect. Notes Math., vol. **62**) Heidelberg Berlin New York: Springer 1968

[Hi] Hickey, T.: Thesis. University of Chicago (1986)

[H_1] Howe, R.: Automorphic forms of low rank. In: Carmona, J., Vergne M. (eds.) Non-commutative harmonic analysis and Lie groups. (Lect. Notes Math. vol. **880,** pp. 211–248) Berlin Heidelberg New York: Springer 1981

[H_2] How, R.: θ-series and invariant theory. In: Automorphic forms, representations, and L-functions, (Proc. Symp. Pure Math., vol. 33, Part I) Providence: Am. Math. Soc. 1979

[HP] Howe, R. and Piatetski-Shapiro, I.: A counterexample to the "generalized Ramanujan conjecture". In: Automorphic forms, representations and L-functions. (Proc. Symp. Pure Math., vol. 33, Part I) Providence: Am. Math. Soc. 1979

[JL] Jacquet, H., Langlands, R.P.: Automorphic forms of GL(2). (Lect. Notes Math., vol. **114**) Berlin Heidelberg New York: Springer 1970

[JS] Jacquet, H., Shalika, J.: On Euler products and the classification of automorphic forms, II. Am. J. Math, **103,** 777–815 (1981)

[K_1] Kazhdan, D.: Connection of the dual space of a group with the structure of its closed subgroups. Funct. Analysis Appl. **1,** 63–65 (1967)

[K_2] Kazhdan, D.: Some applications of the Weil representation. J. Anal. Math. **32,** 235–248 (1977)

[Ku] Kudla, S.: On certain Euler products for SU(2,1). Compos. Math. **42,** 321–344 (1981)

[L] Langlands, R.: Les débuts d'une formule des traces stable. Publ. Math. Univ. Paris VII **13** (1982)

[M] Moore, C.: Decomposition of unitary representations defined by discrete subgroups of nilpotent groups. Ann. Math. **82,** 146–182 (1965)

[MVW] Moeglin, C., Vigneras, M.-F., Waldspurger, J.-L.: Correspondances de Howe sur un corps *p*-adique. (Lect. Notes Math., vol **1291**) Berlin Heidelberg New York: Springer 1988

[P_1] Piatetski-Shapiro, I.I.: Yale University Lectures Notes 1977

[P_2] Piatetski-Shapiro, I.I.: Tate theory for reductive groups and distinguished representations. Proc. Int. Cong. Math, Helsinki 1978, pp. 585–590. Helsinki: Acad. Sci. Fennica 1980

[R] Rogawski, J.: Automorphic representations of unitary groups in three variables. (Annals of Math. Studies vol. 123) Princeton: Princeton Univertsity Press 1990

[R_2] Rogawski, J.: "The multiplicity formula for A-packets: an erratum to [R]". In: *The Zeta-Functions of Picard Modular Surfaces* Langlands, R.P., Ramakrishnan, D. (eds) Les Publications CRM, Université de Montreal (in press)

[Ra] Rallis, S.: On the Howe duality conjecture. Comp. Math **51,** 333–399 (1984)

[Rao] Rao, R.: On some explicit formulas in the theory of Weil representation. (unpublished)

[S_1] Shimura, G.: Automorphic forms and the periods of abelian varieties, J. Math. Soc. Japan **31,** 561–592 (1979)

[S_2] Shimura, G.: On modular forms of half-integral weight. Ann. Math. **97,** 440–481 (1973)

[Sh_1] Shintani, T.: On automorphic forms on unitary groups of order 3. (unpublished manuscript)

[Sh_2] Shintani, T.: A remark on the zeta-functions of algebraic number fields. In: Automorphic forms, representations theory and arithmetic. Bombay Colloquium 1979. Tata Institute of Fundamental Research Studies in Math. No. 10, Bombay 1981, pp. 255–260

[V] Vigneras, M.-F.: Facteurs gamma et équations fonctionnelles. In: Serve J.-P., Zagier D.B. (eds.) Modular Functions of one Var. VI. Proceedings, Bombay 1976. (Lect. Notes Math., vol. **627,** pp. 227–239 Berlin Heidelberg New York: Springer 1977

[Wd_1] Waldspurger, J.-L.: Correspondance de Shimura. J. Math. Pures Appl., **59,** 1–133 (1980)

[Wd_2] Waldspurger, J.L.: Demonstration d'une conjecture de dualité de Howe dans le cas *p*-adique, $p \neq 2$. In: Festschrift in honor of Piatetski-Shapiro. (Israel Math. Conf. Proc., vol. **2,** pp. 267–324 (1990)

[W] Weil, A.: Sur certains groupes d'operateurs unitaires. Acta, (1964) Math. **111** 143–211, (1964)

Invent. math. 105, 473–545 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

The 3-manifold invariants of Witten and Reshetikhin-Turaev for sl(2, C)

Robion Kirby[1] **and Paul Melvin**[2]

[1] Department of Mathematics, University of California, Berkeley, CA-94720, USA
[2] Department of Mathematics, Bryn Mawr College, Bryn Mawr, PA-19010, USA

Oblatum 24-I-1991

0 Introduction

In the fall of 1988, Witten [W] gave the first intrinsically 3-dimensional interpretation of the Jones polynomial [J1, J2] of a link in the 3-sphere, using a quantum field theory with Chern-Simons action. In the process, he uncovered a family of new invariants for arbitrary closed framed 3-manifolds and for links in 3-manifolds.

Shortly afterwards, Reshetikhin and Turaev defined closely related invariants using the theory of quantum groups. In particular, starting with a simple Lie algebra $\mathfrak{g}$, they defined invariants for a framed link L in the 3-sphere using representations of associated quantum groups (which are Hopf algebra deformations of the universal enveloping algebra of $\mathfrak{g}$) [RT1]. For $\mathfrak{g}=\mathrm{sl}(2, \mathbf{C})$, these invariants $J_{L,\mathbf{k}}$ (which depend on a *coloring* $\mathbf{k}$ of L by representations) generalize the Jones polynomial. Their values at a fixed r^{th} root of unity $q=e^{2\pi i/r}$ can be combined to produce a complex valued invariant of the oriented 3-manifold obtained by surgery on L [RT2], shown independent of L using [K1]. (Every 3-manifold can be obtained in this way [L1, Wa].) Presumably these 3-manifold invariants can be defined for any simple Lie algebra. What is needed is that the associated quantum groups have the structure of a *modular* Hopf algebra [RT2].

This paper gives a self-contained proof of the existence of the 3-manifold invariants for $\mathfrak{g}=\mathrm{sl}(2, \mathbf{C})$ and $q=e^{2\pi i/r}$ (§§ 1–5). It is similar in spirit to [RT2], but relies more on the elementary representation theory of the relevant Hopf algebra $\mathscr{A}$ and the topology of framed links, and less on the abstract theory of modular Hopf algebras.

There are several new features in the present treatment (described below). These yield manageable formulas for the 3-manifold invariants which can be interpreted in terms of familiar topological invariants for some small values of r (§§ 6–7). Perhaps more importantly, they lead to a family of new invariants underlying those of Witten and Reshetikhin-Turaev. These new invariants and some applications are discussed in the last section of the paper.

First, the 3-manifold invariants of Reshetikhin and Turaev are modified by splitting off a term involving the first Betti number. The resulting invariants

$\tau_r(M)$ conjugate under orientation reversal (that is $\tau_r(-M)=\overline{\tau_r(M)}$) and are thus useful in answering questions of amphicheirality. They also appear to be *exactly* the same as Witten's invariants of M with the canonical 2-framing of Atiyah [A] (normalized at 1 for the 3-sphere). This has been experimentally verified in many cases by Freed and Gompf [FG], and can in fact be proved for plumbed 3-manifolds using the formulas for Witten's invariants in [FG] together with the formulas for $\tau_r(M)$ in this paper (see Remark 1.9).

Second, a cabling formula (4.15) which reduces the generalized Jones polynomials $J_{L,\mathbf{k}}$ to the classical Jones polynomial for cables of L is derived. (Similar formulas have been obtained independently by Morton and Strickland [MS].) From this one obtains a formula for $\tau_r(M)$ in terms of classical Jones polynomials (4.17) (first announced in [KM1] with a proof sketched in [KM2]). Recently, Lickorish [L3, L4, L5] has found an elegant new proof of the invariance of such a formula, using the Temperley-Lieb algebra and linear skein theory and thereby avoiding the explicit use of quantum groups. It is clear however that the algebra $\mathscr{A}$ should not be deemphasized, for it encodes deep combinatorial properties of the Jones polynomial and appears to make calculations more accessible.

Using this formula, $\tau_r(M)$ is expressed in terms of familiar topological invariants for $r=3$ and 4, as the Jones polynomial has topological meaning at the corresponding roots of unity. (Note that $\tau_2(M)=1$ for all M. The Jones polynomial is also understood at the sixth root of unity (Appendix B), and so one would expect a similar formula for $\tau_6(M)$; see [KM2] for partial results.) For example,

$$\tau_3(M)=\exp(-2\pi i/8)^{\mu(M)+(IH,MI^2-I)/2}$$

if M is a $\mathbf{Z}/2\mathbf{Z}$-homology sphere with μ-invariant $\mu(M)$ (6.5), and

$$\tau_4(M)=\sum_{\Theta}\exp(-6\pi i/16)^{\mu(M_\Theta)}$$

for general M, where $\mu(M_\Theta)$ is the μ-invariant of M with spin structure Θ, and the sum is over all spin structures (7.1). The general formula (6.3) for $\tau_3(M)$ depends on the (mod 2) first betti number and the Brown invariant [Br] of M, and is therefore a homotopy invariant, whereas $\tau_4(M)$ can distinguish homotopy equivalent manifolds. (The derivation of the formula for $\tau_4(M)$ suggested a new, elementary combinatorial proof of Rohlin's theorem, which appears in Appendix C.)

Finally, a Symmetry Principle (4.20) is proved which reduces greatly the number of steps required in calculating $\tau_r(M)$, and (coupled with the cabling formula) leads to an elementary proof of its invariance (see § 5). In particular, this proof avoids the difficult analysis of the structure of tensor products of irrreducible representations of $\mathscr{A}$, which is central to the treatment of [RT2]. Note also the application to Jones polynomials of cables of a framed link at a root of unity (4.25).

The Symmetry Principle has many other interesting consequences. Some of these are discussed in § 8:

(1) For r odd,

$$\tau_r(M)=\begin{cases}\tau_3(M)\,\tau_r'(M) & \text{if } r\equiv 3 \pmod 4\\ \overline{\tau_3(M)}\,\tau_r'(M) & \text{if } r\equiv 1 \pmod 4\end{cases}$$

where $\tau_r'(M)$ is an invariant of M (see 8.10). In particular, $\tau_3(M)=0$ implies that $\tau_r(M)=0$ for all odd r. Note: it is shown in (6.3) that $\tau_3(M)=0$ if and only if there exists α in $H^1(M; \mathbf{Z}/2\mathbf{Z})$ with $\alpha\smile\alpha\smile\alpha\neq 0$.

(2) For r divisible by 4,

$$\tau_r(M)=\sum_{\Theta}\tau_r(M, \Theta)$$

where $\tau_r(M, \Theta)$ is an invariant of the manifold M with spin structure Θ, and the sum is over all spin structures on M (see 8.27). A similar statement holds for $r\equiv 2 (\mathrm{mod}\ 4)$, with the spin structure replaced by an element of $H^1(M; \mathbf{Z}/2\mathbf{Z})$ (8.32). This result has been observed independently by Turaev [T3], also using the Symmetry Principle.

(3) The Casson invariant of a homology sphere M obtained by Dehn surgery on a knot is determined (mod 5) by $\tau_5(M)$ (8.20).

(4) The Jones polynomial of a knot K at the fifth root of unity is an invariant for integral surgery on K (8.14).

(5) If $K_{1/n}$ denotes the homology sphere obtained by $1/n$ surgery on the knot K, then $\tau_r(K_{1/n})$ is periodic in n with period r for odd r (8.15) and period $r/2$ for even r (8.26).

The paper is organized as follows: In §1 there is a general discussion of framed links, the K-move and 3-manifolds, the colored framed link invariants $J_{L,\mathbf{k}}$, and the definition of $\tau_r(M)$. The algebra $\mathscr{A}$ (2.7) is described in §2, along with its finite dimensional representations V^k (2.8) and W_k^r (2.16). Explicit formulas for the R-matrix (2.18 and 2.32) are derived. In §3, the $\mathscr{A}$-linear tangle operators F_T are defined (3.6), which specialize to $J_{L,\mathbf{k}}$ (3.25). Their behavior under direct sums, extensions and tensor products of colors (3.10) and under changes in orientation (3.18) is explored. The cabling formula (4.15) and symmetry principle (4.20) for $J_{L,\mathbf{k}}$ are established in §4, and §5 contains the proof of the invariance of τ_r under K-moves, and of its behavior under connected sums and orientation reversal. The evaluations of $\tau_r(M)$ for $r=3$ and 4 are found in §6 and §7. In §8, applications of the Symmetry Principle to the study of τ_r for odd r (8.7–21) and even r (8.23–33) are given, and the new invariants mentioned above are defined. Appendix A contains a combinatorial proof of the deepest identity in the algebra needed to define τ_r. Appendix B has a treatment of the Jones polynomial at $q=e^{2\pi i/6}$, and Appendix C deals with μ-invariants.

In a future paper [KM4] we will calculate $\tau_r(M)$ for lens spaces and Seifert fibered 3-manifolds, and give a Dehn surgery formula. The calculation of $\tau_r(M)$ for lens spaces led to a new definition of the Dedekind sum in terms of signatures, and new formulae for signature defects and the signature cocycle defining a central extension of $\mathrm{SL}(2, \mathbf{Z})$ by $\mathbf{Z}$ [KM3].

Acknowledgements. We are especially grateful to Nicolai Reshetikhin for his lectures and conversations on his work, and to Vaughan Jones, Greg Kuperberg and Antony Wasserman for valuable insights into quantum groups. We also thank Andrew Casson, Peter Gilkey, Raymond Lickorish, Jim Milgram, Peter Sarnak, Frank Schmidt and Larry Taylor for helpful conversations. Part of this work was carried out while the second author was visiting Stanford University and the University of California at Berkeley. We thank both institutions for their hospitality.

1 The invariants of Reshetikhin and Turaev

Fix an integer $r>1$. In this section we describe in general terms the 3-*manifold invariant* τ_r of Reshetikhin and Turaev [RT2], which assigns a complex number $\tau_r(M)$ to each oriented, closed connected 3-manifold M. It satisfies the following properties (see (5.9)):

(1) (multiplicativity) $\tau_r(M \# N) = \tau_r(M)\,\tau_r(N)$
(2) (orientation) $\tau_r(-M) = \overline{\tau_r(M)}$
(3) (normalization) $\tau_r(S^3) = 1$.

In fact, τ_r is a slight modification of the invariant that appears in [RT2, § 3.3.2], which does not satisfy (2) (see (1.4) below).

$\tau_r(M)$ is defined as a linear combination of certain *colored framed link invariants* $J_{L,\mathbf{k}}$ (defined in [RT1, § 5]) of a framed link L associated with M. The $J_{L,\mathbf{k}}$ are generalizations of the Jones polynomial of L, and are described in more detail below and in § 3.

First we fix some notation to be used throughout the paper. Writing $e(a)$ for $\exp(2\pi i a)$, set

$$q = e\left(\frac{1}{r}\right) \qquad s = e\left(\frac{1}{2r}\right) \qquad t = e\left(\frac{1}{4r}\right)$$

so $q = s^2 = t^4$. (Reshetikhin and Turaev consider other roots of unity $e\left(\frac{m}{r}\right)$, but we restrict to $m=1$ for simplicity.) For any integer k, define

$$[k] = \frac{s^k - \bar{s}^k}{s - \bar{s}} = \frac{\sin\frac{\pi k}{r}}{\sin\frac{\pi}{r}} \tag{1.1}$$

(cf. $[k]_q$ in (2.28) below). Observe that $[k]$ depends on r and has symmetries $[k] = [r-k] = -[k+r]$. Finally, set

$$b = \sqrt{\frac{2}{r}}\sin\frac{\pi}{r} \tag{1.2}$$

$$c = e\left(-\frac{3(r-2)}{8r}\right).$$

Framed links and 3-manifolds

Let L be a framed link in S^3. Recall that L determines a smooth oriented 4-manifold W_L obtained by adding 2-handles to the 4-ball B^4, oriented as the unit ball in $\mathbf{C}^2$, along the components L_i of L in $S^3 = \partial B^4$ [K2]. If L is oriented then each L_i is identified with an element of $H = H_2(W_L; \mathbf{Z})$, also called L_i, formed from an oriented Seifert surface for L_i in S^3 and the core of the associated 2-handle. The L_i form a basis for H, and with respect to this basis the intersection

form on H, denoted by $\cdot$, coincides with the linking matrix of L. That is, $L_i \cdot L_j = lk(L_i \cdot L_j)$ for $i \neq j$ and $L_i \cdot L_i$ is the framing on L_i. We write σ_L (or σ) for the *signature* of the linking matrix of L, or equivalently the index of W_L.

The 3-manifold $M_L = \partial W_L$, oriented using the "outward first" convention for boundaries, is the result of surgery on L in S^3. Any oriented 3-manifold M may be obtained in this way [L1, Wa], and if $M = M_L = M_{L'}$, then one can pass from L to L' by isotopy in S^3 and a combination of the following two moves [K1]:

Move 1 (blow up). *Add (or delete) a disjoint unknotted component with framing* ± 1.

Move 2 (handle slide). *For some* $i \neq j$, *replace* L_i *with* $L_i' = L_i \# L_j$, *a band connected sum of* L_i *with a push off of* L_j *(along the first vector in the framing), with framing* $L_i' \cdot L_i' = (L_i + L_j) \cdot (L_i + L_j)$.

In Move 1, *disjoint* means separated by a 2-sphere from the rest of the link. Move 2 corresponds to sliding the 2-handle for L_i over the 2-handle for L_j. These two moves can be combined into one [FR] which is more convenient for the work of Reshetikhin and Turaev [RT2]:

m-strand **K-move** (of type $\varepsilon = \pm 1$). *Locally, the following are interchangeable*

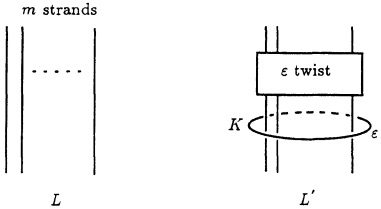


Fig. 1.3

where the framings on corresponding components J *and* J' *of* L *and* L' are related by $J' \cdot J' = J \cdot J + \varepsilon (K \cdot J')^2$.

The colored framed link invariants

At the heart of the 3-manifold invariant $\tau_r(M)$ are the colored framed link invariants $J_{L,\mathbf{k}}$. Here L is a framed link in S^3 with $M = M_L$ and $\mathbf{k}$ is a *coloring* of L, i.e. the assignment of an $\mathscr{A}$-module (or *color*) to each component of L, where $\mathscr{A}$ is a certain Hopf algebra over $\mathbf{C}$ (that depends on the fixed integer r) arising in the theory of *quantum groups.*

The algebra $\mathscr{A}$ will be described in detail in §2. The reason for using a *Hopf* algebra is that the set of representations (i.e. $\mathscr{A}$-modules) is closed under taking tensor products and duals over the ground field $\mathbf{C}$. This is important for the construction of $J_{L,\mathbf{k}}$, given in §3.

Here we will give a heuristic description of $J_{L,\mathbf{k}}$ in the language of topological quantum field theory (see e.g. [AHLS]). Orient L and represent it by a planar diagram D. After removing the extreme points (maxima and minima) of D,

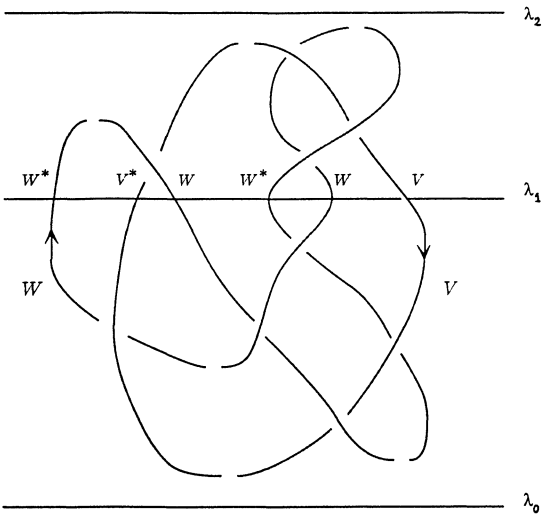


Fig. 1.4

assign the module V, or its dual V^*, to each V-colored strand of L (according to whether the strand is oriented down or up, see Fig. 1.4).

Any horizontal line λ which avoids crossings and extreme points hits D in a collection of points labeled by the colors and their duals. Associate to this line λ the module V_λ which is the tensor product of the labels in order. In Fig. 1.4, $V_{\lambda_0} = \mathbf{C} = V_{\lambda_2}$ and $V_{\lambda_1} = W^* \otimes V^* \otimes W \otimes W^* \otimes W \otimes V$. To each extreme point and each crossing, assign an $\mathscr{A}$-linear operator from the module just below to the one just above. The composition of these operators maps $\mathbf{C}$ to $\mathbf{C}$. Hence it is multiplication by a scalar which (after adjusting for framings) is defined to be $J_{L,\mathbf{k}}$.

This construction should be independent of the orientation on L and the chosen diagram D, so as to give an invariant of (unoriented) colored framed links. Suitable operators are provided in [RT1] using additional structure that exists on the algebra $\mathscr{A}$.

The 3-manifold invariant

Recall that r is a fixed integer >1. To define the 3-manifold invariant in terms of colored framed link invariants $J_{L,\mathbf{k}}$, it is necessary to restrict the colorings $\mathbf{k}$ to lie in a distinguished family $\mathscr{M}$ of $\mathscr{A}$-modules, consisting of one irreducible module V^k in each dimension $0<k<r$ (see §2). We call these $\mathscr{M}$-*colorings* or write $\mathbf{k} \subset \mathscr{M}$, and denote the dimension of the color assigned to the component L_i of L by k_i, also called the *color* of L_i.

Now consider the following linear combination of colored framed link invariants.

(1.5) **Definition.** For any framed link L, define

$$\tau_L = \alpha_L \sum_{\mathbf{k} \subset \mathcal{M}} [\mathbf{k}] J_{L,\mathbf{k}}$$

where $\alpha_L = b^{n_L} c^{\sigma_L}$ and $[\mathbf{k}] = \prod_{i=1}^{n} [k_i]$. (Recall (1.2) that $b = \sqrt{\frac{2}{r}} \sin\frac{\pi}{r}$, $c = e\left(-\frac{3(r-2)}{8r}\right)$; $n = n_L$ is the number of components in L and σ_L is the signature of the linking matrix of L.) Note that since $\mathcal{M} = \{V^1, \ldots, V^{r-1}\}$, the sum $\sum_{\mathbf{k} \subset \mathcal{M}}$ may be written in multi-index notation as $\sum_{\mathbf{0} < \mathbf{k} < \mathbf{r}}$ (see § 4).

(1.6) **Theorem** [RT2] *τ_L is invariant under K-moves on L.*

Theorem 1.6 can be proved for 0 or 1-strand K-moves using three local properties of $J_{L,\mathbf{k}}$ and a standard Gauss sum (see 3.27, 5.1 and 5.4). Then the proof for m-strand K-moves is an easy inductive argument using the Symmetry Principle 4.20 (see (5.6)).

It follows from Theorem 1.6 and [K1, FR] that there is a well defined invariant for closed, oriented 3-manifolds:

(1.7) **Definition.** $\tau_r(M) = \tau_L$, where L is any framed link with $M = M_L$.

(1.8) *Remarks.* (1) The invariant that actually appears in [RT2] differs from $\tau_r(M)$ by a factor of c^ν, where ν is the nullity of the linking matrix of L (= first Betti number of M). The advantage of $\tau_r(M)$ is that it behaves nicely under orientation reversal.

As a convenience to the reader, here is a dictionary relating the notation of [RT2] with the notation in this paper: The algebra U_t of [RT2, § 8] (where $t = e(1/4r) = \exp(2\pi i/4r)$ as above) is our $\mathcal{A} = \mathcal{A}_r$, and the U_t-module V_i is our V^{i+1}. (Note that in U_t, the element uv^{-1} is just K^2 and $\dim V_i = [i+1]$). The link invariant $F(\Gamma(L, \omega, \lambda))$ of [RT2, § 3] (where ω is the orientation on L and λ is the coloring) corresponds to our $J_{L,\lambda}$. The constant C defined in [RT2, § 3.2 and 8.3.8] is our c^2, the coefficient d_i is equal to our $bc[i+1]$, and so the invariant $\{L\} = (bc)^n \sum_{\mathbf{k} \subset \mathcal{M}} [\mathbf{k}] J_{L,\mathbf{k}}$. Thus the 3-manifold invariant defined in [RT2, § 3.3.2] is

$$\mathcal{F}(M, L) = b^n c^{\sigma - \nu} \sum_{\mathbf{k} \subset \mathcal{M}} [\mathbf{k}] J_{L,\mathbf{k}} = c^{-\nu} \tau_r(M).$$

(2) It is often useful for calculations to write $\tau_r(M)$ (where $M = M_L$) in other ways. For example, let $a_i = L_i \cdot L_i$ be the framing on L_i, and write L_0 for the link L with all framings changed to zero. It is shown in Lemma 3.27b below that $J_{L,\mathbf{k}}$ changes by $t^{\pm(k_i^2 - 1)}$ if the framing of L_i is changed by ± 1. Thus $J_{L,\mathbf{k}} = t^{\sum a_i (k_i^2 - 1)} J_{L_0,\mathbf{k}}$. Now, since $c = t^3 e(-\frac{3}{8})$, (1.5) becomes

$$\tau_r(M) = b^n e(-\tfrac{3}{8})^\sigma t^{3\sigma - \sum a_i} \sum_{\mathbf{k} \subset \mathcal{M}} [\mathbf{k}] t^{\sum a_i k_i^2} J_{L_0,\mathbf{k}}. \tag{1.9}$$

The term $\varphi(L) = 3\sigma - \sum a_i$ has been identified in [FG] as the difference between the 2-framing on M determined naturally by L and the canonical 2-framing on M [A].

Furthermore, the Gauss sum $G=\sum_{k=1}^{4r} t^{k^2}$ (see [La] and (5.1)) equals $\sqrt{8r}e(\frac{1}{8})$, and so $b=\frac{G}{4r}(s-\bar{s})\,e(-\frac{3}{8})$ since $s-\bar{s}=2i\sin\left(\frac{\pi}{r}\right)$. Formula (1.9) becomes

$$\tau_r(M)=\left(G\frac{(s-\bar{s})}{4r}\right)^n e(-\tfrac{3}{8})^{n+\sigma}\, t^{3\sigma-\Sigma a_i}\sum_{\mathbf{k}\subset\mathscr{M}}[\mathbf{k}]\,t^{\Sigma a_i k_i^2}J_{L_0,\mathbf{k}}. \tag{1.10}$$

It follows that $\tau_r(M)$ belongs to the ring $\mathbf{Z}[t^{\pm\frac{1}{2}}, r^{-1}]$ since the $\frac{1}{4^n}$ disappears using results from §8. It is necessary to use $t^{\pm\frac{1}{2}}$ rather than t, since for example (see (5.11) and (6.3)) $\tau_3(S^1\times S^2)=\frac{1}{b}=\sqrt{2}=t^{3/2}-t^{3/2}$. (Integer powers of t suffice if the nullity is put back into τ, see (1.8)). Question: is $\tau_r(M)$ always an element of $\mathbf{Z}[t^{\pm\frac{1}{2}}]$?

2 The quasitriangular Hopf algebra $\mathscr{A}$

In this section we shall define the Hopf algebra $\mathscr{A}$ and produce an R-matrix R in $\mathscr{A}\otimes\mathscr{A}$ making $\mathscr{A}$ a quasitriangular Hopf algebra. We also show that the associated operators $\bar{R}$ in representations of $\mathscr{A}$ satisfy the Yang-Baxter equation.

Throughout this section, r will be a fixed integer greater than 1.

The algebra $\mathscr{A}$

As motivation, we first recall the definition of the Lie algebra sl(2, **C**) and its representations: sl(2, **C**) is a 3-dimensional complex vector space with preferred basis X, Y, H and Lie bracket given by $[H, X]=2X$, $[H, Y]=-2Y$ and $[X, Y]=H$. It has a unique (up to isomorphism) k-dimensional irreducible representation V^k for each positive integer k. Explicitly, sl(2, **C**) acts on V^k (with preferred ordered basis $e_m, e_{m-1}, \ldots, e_{-m}$ where $k=2m+1$) by

$$\begin{aligned} Xe_j&=(m+j+1)\,e_{j+1}\\ Ye_j&=(m-j+1)\,e_{j-1}\\ He_j&=2je_j. \end{aligned} \tag{2.1}$$

Note that the subscripts are integers if k is odd and half integers if k is even.

For example, the 1, 2 and 3-dimensional representations of sl(2, **C**) are

(2.2) (1) $X=Y=H=0$

(2) $X=\begin{pmatrix}0&1\\0&0\end{pmatrix}$, $Y=\begin{pmatrix}0&0\\1&0\end{pmatrix}$ and $H=\begin{pmatrix}1&0\\0&-1\end{pmatrix}$

(3) $X=\begin{pmatrix}0&2&0\\0&0&1\\0&0&0\end{pmatrix}$, $Y=\begin{pmatrix}0&0&0\\1&0&0\\0&2&0\end{pmatrix}$ and $H=\begin{pmatrix}2&0&0\\0&0&0\\0&0&-2\end{pmatrix}$

respectively.

We extend this discussion to the *universal enveloping algebra* $\mathbf{U}=U(\mathrm{sl}(2,\mathbf{C}))$, which is just the associative algebra over $\mathbf{C}$ with the same generators and relations as $\mathrm{sl}(2,\mathbf{C})$. (The bracket is interpreted as for matrices. Thus $HX-XH=2X$, or equivalently $HX=X(H+2)$. Similarly $HY=Y(H-2)$ and $XY-YX=H$.) The representations above evidently extend to algebra representations of $\mathbf{U}$ and so there are unique irreducible $\mathbf{U}$-modules V^k in each dimension.

Note that $\mathbf{U}$ has a Hopf algebra structure (with comultiplication $\Delta\colon \mathbf{U}\to\mathbf{U}\otimes\mathbf{U}$, antipode $S\colon \mathbf{U}\to\mathbf{U}$ and counit $\varepsilon\colon \mathbf{U}\to\mathbf{C}$ given by $\Delta(\alpha)=\alpha\otimes 1+1\otimes\alpha$, $S(\alpha)=-\alpha$ and $\varepsilon(\alpha)=0$ for all α in $\mathrm{sl}(2,\mathbf{C})$). This allows one to define $\mathbf{U}$-module structures on the duals $V^*=\mathrm{Hom}_{\mathbf{C}}(V,\mathbf{C})$ and tensor products $V\otimes W=V\otimes_{\mathbf{C}}W$ of $\mathbf{U}$-modules V and W. In particular, $(\alpha f)(v)=f(S(\alpha)v)$ and $\alpha(v\otimes w)=\Delta\alpha\cdot(v\otimes w)$ (where $\mathbf{U}\otimes\mathbf{U}$ acts diagonally on $V\otimes W$), for $\alpha\in\mathbf{U}$, $f\in V^*$, $v\in V$ and $w\in W$. (This is the reason for using *Hopf* algebras in the construction of the colored framed link invariants.)

Next we consider the *quantized universal enveloping algebra* $\mathbf{U}_h=U_h(\mathrm{sl}(2,\mathbf{C}))$, found by Kulish and Reshetikhin [KuR]. It can be defined as the algebra $\mathbf{U}[[h]]$ of *formal* power series in h (=Planck's constant for the physicists) with coefficients in $\mathbf{U}$, with the same relations as in $\mathbf{U}$ except that $[X,Y]=H$ is replaced with

$$[X,Y]=\frac{\sinh(hH/2)}{\sinh(h/2)}=H+\frac{H^3-H}{24}h^2+\dots.$$

Setting $q=e^h$ and (in analogy with the usual notation) $s=e^{h/2}$, $t=e^{h/4}$, $\bar{s}=s^{-1}=e^{-h/2}$ and

$$[H]=\frac{s^H-\bar{s}^H}{s-\bar{s}},$$

these relations may be written

$$\begin{aligned} HX&=X(H+2)\\ HY&=Y(H-2)\\ [X,Y]&=[H]. \end{aligned}\tag{2.3}$$

If we introduce the element

$$K=t^H=e^{\frac{hH}{4}}=1+\frac{H}{4}h+\frac{H^2}{2!\,4^2}h^2+\dots$$

(which will play an important role in the sequel), then we obtain the associated relations

$$\begin{aligned} KX&=sXK\\ KY&=\bar{s}YK\\ [X,Y]&=[H]=\frac{K^2-\bar{K}^2}{s-\bar{s}} \end{aligned}\tag{2.4}$$

where $\bar{K}=K^{-1}=\bar{t}^H$.

There is a Hopf algebra structure on $\mathbf{U}_h$ as a module over the ring $\mathbf{C}[[h]]$ of formal power series, discovered by Sklyanin [Sk], with comultiplication Δ, antipode S and counit ε given by

(2.5)
$$\begin{aligned}
\Delta(X) &= X\otimes K + \bar{K}\otimes X\\
\Delta(Y) &= Y\otimes K + \bar{K}\otimes Y\\
\Delta(H) &= H\otimes 1 + 1\otimes H\\
S(X) &= -sX\\
S(Y) &= -\bar{s}Y\\
S(H) &= -H\\
\varepsilon(X) &= \varepsilon(Y) = \varepsilon(H) = 0.
\end{aligned}$$

One may readily compute

(2.6)
$$\begin{aligned}
\Delta(K) &= K\otimes K\\
S(K) &= \bar{K}\\
\varepsilon(K) &= 1.
\end{aligned}$$

We would like to specialize $\mathbf{U}_h$ at particular values of h, namely $h=\dfrac{2\pi i}{r}$ $\left(\text{so } q=e^h=e\left(\dfrac{1}{r}\right)\right)$, and then look for complex representations. This cannot be done using the full algebra $\mathbf{U}_h$, because of the presence of divergent series, and so we first restrict to the subalgebra, over the ring of *convergent* power series in h (i.e. entire functions), generated by X, Y, K and $\bar{K}$. Now, following Reshetikhin and Turaev, we define

$$\mathscr{A} = \mathscr{A}_r$$

(denoted $\mathbf{U}_t$ in [RT2]) to be the quotient of this subalgebra obtained by setting $h=\dfrac{2\pi i}{r}$, $X^r=0$, $Y^r=0$ and $K^{4r}=1$. (Omitting the last three relations yields the infinite dimensional algebra $\mathbf{U}_q$ of [Ji, RT1] known as the *q-analogue* of $\mathbf{U}$.) Thus $\mathscr{A}$ is a finite dimensional algebra over $\mathbf{C}$ with generators X, Y, K, $\bar{K}$ and relations

(2.7)
$$\begin{aligned}
\bar{K} &= K^{-1}\\
KX &= sXK\\
KY &= \bar{s}YK\\
[X, Y] &= \frac{K^2-\bar{K}^2}{s-\bar{s}}\\
X^r &= Y^r = 0\\
K^{4r} &= 1
\end{aligned}$$

where $s=e\left(\dfrac{1}{2r}\right)$, as usual. (We will retain the notation

$$[H]=\frac{s^H-\bar{s}^H}{s-\bar{s}}=\frac{K^2-\bar{K}^2}{s-\bar{s}}$$

even though H is no longer in our algebra.) $\mathscr{A}$ acquires a complex Hopf algebra structure from $\mathbf{U}_h$, and so tensor products and duals (over $\mathbf{C}$) of $\mathscr{A}$-modules are still $\mathscr{A}$-modules.

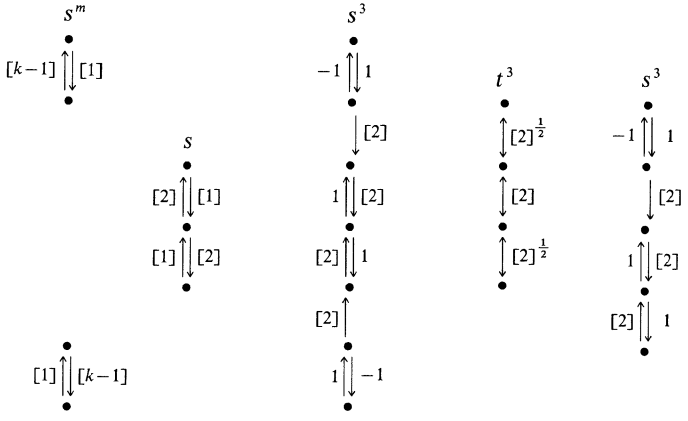


Fig. 2.11

Representations of $\mathscr{A}$

As with **U**, there are $\mathscr{A}$-modules V^k in each dimension $k>0$. In particular, $\mathscr{A}$ acts on V^k (with basis $e_m, \ldots, e_{-m}$, for $k=2m+1$) by

$$
\begin{aligned}
Xe_j &= [m+j+1]\, e_{j+1} \\
Ye_j &= [m-j+1]\, e_{j-1} \\
Ke_j &= s^j e_j = t^{2j} e_j
\end{aligned} \tag{2.8}
$$

(cf. (2.1), but note the brackets). The relation $[X, Y]=[H]$ follows from the identity $[a][b]-[a+1][b-1]=[a-b+1]$.

For example, the 1, 2 and 3 dimensional representations of $\mathscr{A}$ are

(2.9) (1) $X=0$, $Y=0$ and $K=1$

(2) $X=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$, $Y=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$ and $K=\begin{pmatrix}t & 0\\ 0 & \bar{t}\end{pmatrix}$

(3) $X=\begin{pmatrix}0 & [2] & 0\\ 0 & 0 & [1]\\ 0 & 0 & 0\end{pmatrix}$, $Y=\begin{pmatrix}0 & 0 & 0\\ [1] & 0 & 0\\ 0 & [2] & 0\end{pmatrix}$ and $K=\begin{pmatrix}s & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & \bar{s}\end{pmatrix}$

respectively.

(2.10) *Remark* It is useful to represent V^k by a graph in the plane with one vertex at height j for each basis vector e_j, and with oriented edges from e_j to $e_{j\pm1}$ labeled by $[m\pm j+1]$ if $[m\pm j+1]\neq 0$ (recall $k=2m+1$), indicating the actions of X and Y. This graph, with the top vertex labeled by its *weight* (i.e. eigenvalue for K) s^m, is called the *diagram* of V^k with respect to $e_m, \ldots, e_{-m}$. (See Fig. 2.11a, with the special cases $k=3$ and 7 for $r=5$ shown in 2.11b

and c, using the identities $[j]=[r-j]=-[r+j]$.) Similar diagrams can be used to describe other finite dimensional $\mathscr{A}$-modules with respect to bases of *weight vectors* (i.e. eigenvectors for K).

The diagram of a module is often simplified by rescaling the basis (i.e. changing the lengths of the basis vectors). In particular, if v is changed to av, then the label on each edge which starts (or ends) at the corresponding vertex is multiplied (or divided) by a.

For example for $k \leqq r$, the basis $e_m, \ldots, e_{-m}$ for V^k can be rescaled to a unique basis $b_m, \ldots, b_{-m}$ (up to a multiple), called a *balanced* basis, with

$$\begin{aligned} Xb_j &= ([m+j+1][m-j])^{1/2} b_{j+1} \\ Yb_j &= ([m-j+1][m+j])^{1/2} b_{j-1} \\ Kb_j &= s^j b_j. \end{aligned}$$

(Indeed $e_j = \left[\frac{2m}{m-j}\right]^{1/2} b_j$, where $\left[\frac{n}{k}\right] = \frac{[n]!}{[k]![n-k]!}$ is the quantized binomial coefficient, cf. (2.29).) Observe that in the corresponding diagram, any two oppositely oriented edges with the same end points have equal labels of the form $([i][k-i])^{1/2}$, and so can be combined into one doubly oriented edge. The case $r=5$, $k=4$ is shown in Fig. 2.11 d.

It turns out that the representations V^k are irreducible if and only if $k \leqq r$ (see below), and so it is natural to define the distinguished family

$$\mathscr{M} = \{V^1, \ldots, V^{r-1}\} \tag{2.12}$$

of $\mathscr{A}$-modules to be used in constructing the 3-manifold invariant τ_r. (V^r is excluded for technical reasons; see Lemma 3.29 below.)

The structure of the $\mathscr{A}$-modules V^k for $k \leqq r$, and their tensor products $V^k \otimes V^{k'}$ for $k+k' \leqq r+1$, is parallel to the classical case, and is summarized in the following well known result (see e.g. [Ji, Lu, RT2]).

(2.13) **Theorem** *If $k \leq r$, then the modules V^k are irreducible and self dual. In particular, the map $D\colon (V^k)^* \to V^k$ given by $D(b^j) = (-s)^j b_{-j}$, where b_j is a balanced basis with dual basis b^j, is an $\mathscr{A}$-linear isomorphism.* (*Equivalently,* $D(e^j) = \left[\frac{2m}{m-j}\right]^{-1} (-s)^j e_{-j}$.)

Furthermore, if $k+k' \leqq r+1$, then

$$V^k \otimes V^{k'} = \bigoplus_{p \in k \otimes k'} V^p$$

where $k \otimes k' = \{k+k'-1, k+k'-3, \ldots, |k-k'|+1\}$.

Proof. It is evident from its diagram that for $k \leqq r$, V^k contains no proper submodules generated by weight vectors e_j. But for $k < 2r$, *every* submodule of V^k is generated by weight vectors, since the weights of the e_j are distinct. Indeed, for any nonzero vector $v = \sum a_j e_j$, each e_j is a multiple of v by a suitable polynomial in K, and so the submodules generated by v and by the e_j with $a_j \neq 0$ coincide. It follows that V^k is irreducible for $k \leqq r$.

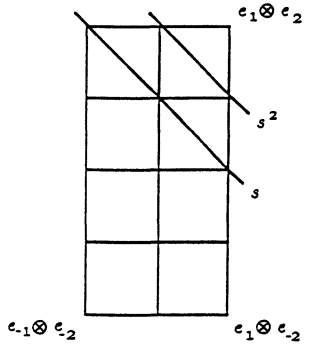


Fig. 2.14

For the second statement, note that a **C**-linear isomorphism D is $\mathscr{A}$-linear provided $D(\alpha b^j) = \alpha D(b^j)$ for $\alpha = K$, X and Y. For D of the form $D(b^j) = c_j b_{-j}$, this imposes no restriction on the c_j for $\alpha = K$, and the sole restriction $c_j = (-s)\, c_{j-1}$ for $\alpha = X$ or Y. To see this, one may compute the action of $\mathscr{A}$ on V^{k*}:

$$Xe^j = -s[m+j]\, e^{j-1}$$
$$Ye^j = -\bar{s}[m-j]\, e^{j+1}$$
$$Ke^j = \bar{s}^j e^j,$$

or in the balanced basis,

$$Xb^j = -s([m-j+1][m+j])^{1/2}\, b^{j-1}$$
$$Yb^j = -\bar{s}([m+j+1][m-j])^{1/2}\, b^{j+1}$$
$$Kb^j = \bar{s}^j b^j.$$

$\left(\text{Note that } b^j = \begin{bmatrix} 2m \\ m-j \end{bmatrix}^{1/2} e^j.\right)$ The self duality result follows.

Finally, to verify the decomposition of $V^k \otimes V^{k'}$ observe that there are weight vectors v_p in $V^k \otimes V^{k'}$ of weight t^{p-1} with $Xv_p = 0$ for each p in $k \otimes k'$, since the dimensions of the corresponding weight spaces decrease as p increases. (To see this it is useful to use the diagram for $V^k \otimes V^{k'}$ with respect to the weight basis $e_i \otimes e_j$ with vertices at (i, j), see Fig. 2.14 for the case $V^3 \otimes V^5$.) It follows that the submodule generated by v_p is isomorphic to V^p. These subspaces are independent since any collection of equal weight vectors lying in distinct V^p are annihilated by distinct powers of X. A dimension count completes the argument. □

As a consequence of Theorem 2.13, each of the irreducible modules V^k can be expressed as a linear combination of powers of V^2 in the representation ring of $\mathscr{A}$ (where for example $2V + UW$ means $V \oplus V \oplus (U \otimes W)$, $(V)^2$ means $V \otimes V$ (not V^2), and $U = V - W$ means $U \oplus W = V$):

(2.15) **Corollary** *For $0 \leqq n < r$, the equality*

$$V^{n+1} = \sum_{j=0}^{n/2} (-1)^j \binom{n-j}{j} (V^2)^{n-2j}$$

holds in the representation ring, where the sum is over all integers j with $0 \leqq 2j \leqq n$.

Proof. Theorem 2.13 implies that $V^{n+1} = V^2 V^n - V^{n-1}$. Thus by induction (which starts trivially at $n=0$),

$$\begin{aligned}
V^{n+1} &= V^2 \sum_j (-1)^j \binom{n-1-j}{j} (V^2)^{n-1-2j} - \sum_j (-1)^j \binom{n-2-j}{j} (V^2)^{n-2-2j} \\
&= \binom{n-1}{0} (V^2)^n - \binom{n-2}{1} (V^2)^{n-2} + \binom{n-3}{2} (V^2)^{n-4} - + \dots \\
&\quad - \binom{n-2}{0} (V^2)^{n-2} + \binom{n-3}{1} (V^2)^{n-4} - + \dots \\
&= \binom{n}{0} (V^2)^n - \binom{n-1}{1} (V^2)^{n-2} + \binom{n-2}{2} (V^2)^{n-4} - + \dots \\
&= \sum_j (-1)^j \binom{n-j}{j} (V^2)^{n-2j}. \quad \square
\end{aligned}$$

It is amusing to note that the same identity holds with the bracket $[n]$ replacing V^n. The same proof works.

For small values of n, we have

$$\begin{aligned}
V^3 &= V^2 \otimes V^2 - V^1 \\
V^4 &= V^2 \otimes V^2 \otimes V^2 - 2V^2 \\
V^5 &= V^2 \otimes V^2 \otimes V^2 \otimes V^2 - 3V^2 \otimes V^2 + V^1.
\end{aligned}$$

Remark. The structure of the tensor product $V^k \otimes V^{k'}$ for $k + k' > r+1$ is more complicated. This has been analyzed by Reshetikhin and Turaev [RT2] (and independently by A. Wasserman and J. Frölich-G. Keller) and is central to their proof of the invariance of the 3-manifold invariant. We will give a different proof of the invariance in §5 which depends on the Symmetry Principle 4.20. This in turn is based on the structure of certain r-dimensional $\mathscr{A}$-modules W_p^r discussed below (which arise as well in the general discussion of tensor products).

In contrast with the case of **U**, the $\mathscr{A}$-modules V^k are reducible for $k > r$. In particular, the subspace V_k^r generated by e_j for $j > m-r$ is an r-dimensional submodule, since $Ye_{m-r+1} = [r]\, e_{m-r} = 0$. These modules are called *Verma modules.*

Observe that $V_r^r = V^r$, and (as is easily seen using Remark 2.10) V_k^r and V_{k+4r}^r are isomorphic. If $r \leqq k < 2r$, then V_k^r contains V^p, where $p = 2r-k$, as its *unique* proper submodule, which motivates the notation adopted in [RT2]

$$W_p^r = V_{2r-p}^r$$

for $0<p<r$. (See Fig. 2.11e for the diagram of $W_3^5 = V_7^5$.) Indeed, it is clear from the diagram of W_p^r that V^p is the only submodule generated by weight vectors, but every submodule is of this form by the argument in the proof of Theorem 2.13.

In fact, the Verma modules W_p^r may be described as extensions of V^p by a twisted version $V^{r-p}(i)$ of V^{r-p}. In particular, observe that there are exactly four 1-dimensional $\mathscr{A}$-modules $V^1(\alpha)$, where $\alpha^4 = 1$, given by $K = \alpha$ and $X = Y = 0$. (The value of K follows from the relation $[X, Y] = \dfrac{K^2 - \bar{K}^2}{s - \bar{s}}$, which gives $K^4 = 1$ since $[X, Y] = 0$ in $\mathbf{C}$, and the values of X and Y are immediate from the relations $KX = sXK$ and $KY = \bar{s}YK$.) For any $\mathscr{A}$-module V, put

$$V(\alpha) = V \otimes V^1(\alpha).$$

(A diagram for $V(\alpha)$ is obtained from one for V by multiplying the vertex weights by α.) It is now easy, using Remark 2.10, to establish the following result.

(2.16) **Lemma** [RT2, § 8.4] *There is a short exact sequence*

$$0 \to V^p \to W_p^r \to V^{r-p}(i) \to 0$$

for $0<p<r$, where V^p is the unique proper submodule of W_p^r.

Similar considerations apply to the general Verma modules V_k^r, since V_{k+nr}^r is isomorphic to $V_k^r(i^n)$.

The R-matrix

The algebra $\mathscr{A}$ has the additional structure of a *quasi-triangular* Hopf algebra (see Drinfeld [D2]). That is, there exists an invertible element R in $\mathscr{A} \otimes \mathscr{A}$ satisfying the following properties:

(2.17) (a) $R\Delta(\alpha)R^{-1} = \check{\Delta}(\alpha)$ for all α in $\mathscr{A}$

(b) $(\Delta \otimes \mathrm{id})(R) = R_{13}R_{23}$

$(\mathrm{id} \otimes \Delta)(R) = R_{13}R_{12}$

where $\check{\Delta}(\alpha) = P(\Delta(\alpha))$, P is the permutation endomorphism of $\mathscr{A}$ given by $P(\alpha \otimes \beta) = \beta \otimes \alpha$, $R_{12} = R \otimes 1$, $R_{23} = 1 \otimes R$ and $R_{13} = (P \otimes \mathrm{id})(R_{23})$. (Explicitly, if $R = \sum \alpha_i \otimes \beta_i$, then $R_{12} = \sum \alpha_i \otimes \beta_i \otimes 1$, $R_{23} = \sum 1 \otimes \alpha_i \otimes \beta_i$ and $R_{13} = \sum \alpha_i \otimes 1 \otimes \beta_i$.) Such an element R is called a *universal R-matrix* for $\mathscr{A}$, and is the central ingredient in the definition of the colored framed link invariants.

Historically, an R-matrix was first discovered in the algebra $\mathbf{U}_h$ by Drinfeld [D1] and independently by Jimbo [J] in the algebra $\mathbf{U}_q$. R-matrices in $\mathscr{A}$ have been written down by several authors, including Reshetikhin and Turaev [RT2] and A. Wasserman. We give a formula for R of the form $R = \sum c_{nab} X^n K^a \otimes Y^n K^b$, which is derived by recursively solving for the constants c_{nab} using the defining relation (2.17a). (This approach to finding R was suggested to us by A. Wasserman, who has previously carried out a similar calculation.)

(2.18) **Theorem** *The element*

$$R=\frac{1}{4r}\sum_{n,a,b}\frac{(s-\bar{s})^n}{[n]!}\,t^{ab+(b-a)n+n}X^nK^a\otimes Y^nK^b,$$

where the sum is over all $0\leqq n<r$ *and* $0\leqq a,\ b<4r$, *is a universal R-matrix for* $\mathscr{A}$.

Proof. As mentioned above, we assume that there exists an R-matrix of the form

$$R=\sum_{n,a,b}c_{nab}X^nK^a\otimes Y^nK^b \qquad 0\leqq n<r,\ 0\leqq a,\ b<4r, \tag{2.19}$$

and then find c_{nab} by solving $R\Delta(X)=\check{\Delta}(X)R$ and $R\Delta(Y)=\check{\Delta}(Y)R$.

First note the following commutation relations in $\mathscr{A}$ (see 2.7):

$$\begin{aligned} KX&=sXK \quad\text{and}\quad [H]X=X[H+2]\\ KY&=\bar{s}YK \quad\text{and}\quad [H]Y=Y[H-2]\\ YX&=XY-[H] \end{aligned} \tag{2.20}$$

where $[H+n]=\dfrac{s^{H+n}-\bar{s}^{H+n}}{s-\bar{s}}=\dfrac{s^nK^2-\bar{s}^n\bar{K}^2}{s-\bar{s}}$. By induction, using the identity

$$[a][H+c+b]+[b][H+c-a]=[a+b][H+c],$$

we can generalize the last relation in (2.20) to

$$\begin{aligned} Y^nX&=XY^n-[n][H+n-1]\,Y^{n-1}\\ YX^n&=X^nY-[n][H-n+1]\,X^{n-1}. \end{aligned}$$

Now we have (recalling from (2.5) that $\Delta(X)=X\otimes K+\bar{K}\otimes X$)

$$\begin{aligned} 0&=R\Delta(X)-\check{\Delta}(X)R\\ &=\sum_{n,a,b}c_{nab}(X^nK^aX\otimes Y^nK^{b+1}+X^nK^{a-1}\otimes Y^nK^bX\\ &\quad -KX^nK^a\otimes XY^nK^b-X^{n+1}K^a\otimes\bar{K}Y^nK^b)\\ &=\sum_{n,a,b}c_{nab}(s^aX^{n+1}K^a\otimes Y^nK^{b+1}+s^bX^nK^{a-1}\otimes XY^nK^b\\ &\quad -\frac{s^b[n]}{s-\bar{s}}(\bar{s}^{n-1}X^nK^{a-1}\otimes Y^{n-1}K^{b+2}-s^{n-1}X^nK^{a-1}\otimes Y^{n-1}K^{b-2})\\ &\quad -s^nX^nK^{a+1}\otimes XY^nK^b-s^nX^{n+1}K^a\otimes Y^nK^{b-1}). \end{aligned}$$

Thus the coefficient of $X^nK^a\otimes XY^nK^b$ is $-s^nc_{n,a-1,b}+s^bc_{n,a+1,b}=0$ which implies that

$$c_{n,a+2,b}=s^{n-b}c_{nab}=\bar{s}^{b-n}c_{nab}. \tag{2.21}$$

Also, the coefficient of $X^n K^a \otimes Y^{n-1} K^b$ is

$$s^a c_{n-1,a,b-1} - \frac{s^{b-n-1}[n]}{s-\bar{s}} c_{n,a+1,b-2} + \frac{s^{b+n+1}[n]}{s-\bar{s}} c_{n,a+1,b+2} - s^{n-1} c_{n-1,a,b+1} = 0. \tag{2.22}$$

Similarly, using $R\Delta(Y) - \check{\Delta}(Y)R = 0$, we get

$$\begin{aligned} 0 = \sum_{n,a,b} c_{nab} & (\bar{s}^a X^n Y K^a \otimes Y^n K^{b+1} + \bar{s}^b X^n K^{a-1} \otimes Y^{n+1} K^b \\ & - s^n X^a K^{a+1} \otimes Y^{n+1} K^b - s^n X^n Y K^a \otimes Y^n K^{b-1} \\ & + \frac{s^n[n]}{s-\bar{s}} (s^{n-1} X^{n-1} K^{a+2} \otimes Y^n K^{b-1} - \bar{s}^{n-1} X^{n-1} K^{a-2} \otimes Y^n K^{b-1}), \end{aligned}$$

and so the coefficient of $X^n Y K^a \otimes Y^n K^b$ is $\bar{s}^a c_{n,a,b-1} - s^n c_{n,a,b+1} = 0$ which implies

$$c_{n,a,b+2} = \bar{s}^{a+n} c_{n,a,b}. \tag{2.23}$$

Also, the coefficient of $X^{n-1} K^a \otimes Y^n K^b$ is

(2.24)

$$\bar{s}^b c_{n-1,a+1,b} - s^{n-1} c_{n-1,a-1,b} + \frac{[n]}{s-\bar{s}} s^{2n-1} c_{n,a-2,b+1} - \frac{[n]}{s-\bar{s}} s c_{n,a+2,b+1} = 0.$$

Using (2.22) and (2.23) we obtain

$$c_{n,a,b} = \frac{s-\bar{s}}{[n]\, s^b} c_{n-1,a-1,b-1} \tag{2.25}$$

which can also be obtained from (2.21) and (2.24).

If we choose $c_{0,0,0} = 1$, then from (2.21) and (2.23) $c_{0,2a,2b} = \bar{s}^{2ab} = \bar{t}^{4ab}$. Thus it is a natural choice to let $c_{0ab} = \bar{t}^{ab}$ (which is consistent with (2.21) and (2.23)). It follows from (2.25) that for the values $c_{nab} = \frac{(s-\bar{s})^n}{[n]!} \bar{t}^{ab+(b-a)n+n}$, the element R of (2.19) satisfies the first defining relation (2.17a) of the R-matrix. In fact, in order to satisfy the second defining relation as well, it is necessary to normalize by multiplying by $\frac{1}{4r}$. Thus, we put

$$c_{nab} = \frac{1}{4r} \frac{(s-\bar{s})^n}{[n]!} \bar{t}^{ab+(b-a)n+n} \tag{2.26}$$

which gives the desired

$$R = \frac{1}{4r} \sum_{n,a,b} \frac{(s-\bar{s})^n}{[n]!} \bar{t}^{ab+(b-a)n+n} X^n K^a \otimes Y^n K^b$$

satisfying (2.17a), and it remains to verify that R is invertible and satisfies (2.17b). We will check the axiom

$$(\Delta \otimes \mathrm{id})\, R = R_{13}\, R_{23} \tag{2.27}$$

and leave the rest as an exercise.

We need a quantized version of the binomial coefficients (this goes back to Gauss). If $BA = AB$, then $(A+B)^n = \sum_k \binom{n}{k} A^k B^{n-k}$ defines the binomial coefficient $\binom{n}{k}$. Similarly, if $BA = qAB$ (q arbitrary), then

$$(A+B)^n = \sum_k \begin{bmatrix} n \\ k \end{bmatrix}_q A^k B^{n-k}$$

defines the binomial q-coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q$. It can be verified by induction that

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{[n]_q!}{[k]_q![n-k]_q!}$$

where

$$[n]_q = \frac{q^n - 1}{q-1}. \tag{2.28}$$

This is just an unbalanced version of the $[n]$ used in this paper, i.e. $[n]_q = s^{n-1}[n]$. It follows that

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = s^{k(n-k)} \begin{bmatrix} n \\ k \end{bmatrix} \quad \text{where} \quad \begin{bmatrix} n \\ k \end{bmatrix} = \frac{[n]!}{[k]![n-k]!}. \tag{2.29}$$

Returning to the axiom (2.27), the left hand side $(\Delta \otimes \mathrm{id})\, R$ is

$$\frac{1}{4r} \sum_{n,a,b} \frac{(s-\bar{s})^n}{[n]!}\, \bar{t}^{ab+(b-a)n+n} (X \otimes K + \bar{K} \otimes X)^n (K \otimes K)^a \otimes Y^n K^b \tag{2.30}$$

$$= \frac{1}{4r} \sum_{n,a,b} c_{nab} \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_{\bar{q}} X^k K^{k-n+a} \otimes s^{k(n-k)} X^{n-k} K^{k+a} \otimes Y^n K^b,$$

since $(\bar{K} \otimes X)(X \otimes K) = \bar{q}(X \otimes K)(\bar{K} \otimes X)$, and the right hand side $R_{13} R_{23}$ is

$$\frac{1}{(4r)^2} \sum_{\substack{n',a',b' \\ n'',a'',b''}} c_{n'a'b'}\, c_{n''a''b''}\, X^{n'} K^{a'} \otimes X^{n''} K^{a''} \otimes \bar{s}^{b'n''} Y^{n'+n''} K^{b'+b''}. \tag{2.31}$$

We need to show equal the coefficients of

$$X^k K^{a'} \otimes X^{n-k} K^{a''} \otimes Y^n K^b$$

where $a'=k-n+a$ and $a''=k+a$ in (2.30), and $k=n'$, $n''=n-k$ and $b=b'+b''$ in (2.31). Furthermore, note that $a'=a''-n$ in (2.30), but there are terms in (2.31) for which $a'\neq a''-n$ so these terms must be shown zero.

Using (2.29) and (2.26), we compute this coefficient in (2.30) to be

$$\begin{aligned}&\frac{1}{4r}\,c_{n,a,b}\begin{bmatrix}n\\k\end{bmatrix}\bar{s}^{k(n-k)}\,s^{k(n-k)}\\&=\frac{1}{4r}\,\frac{(s-\bar{s})^n}{[n]!}\,\bar{t}^{ab+(b-a)n+n}\cdot\frac{[n]!}{[k]![n-k]!}\\&=\frac{1}{4r}\,\frac{(s-\bar{s})^n}{[k]![n-k]!}\,\bar{t}^{ab+(b-a)n+n}.\end{aligned}$$

Since b' and b'' can vary as long as $b'+b''\equiv b(\mathrm{mod}\,4r)$, the corresponding coefficient in (2.31) equals

$$\begin{aligned}&\frac{1}{16r^2}\sum_{b'=0}^{4r-1}c_{ka'b'}\,c_{n-k,a'',b-b'}\,\bar{s}^{b'(n-k)}\\&=\frac{1}{16r^2}\sum_{b'=0}^{4r-1}\frac{(s-\bar{s})^k}{[k]!}\,\bar{t}^{a'b'+(b'-a')k+k}\\&\quad\cdot\frac{(s-\bar{s})^{n-k}}{[n-k]!}\,\bar{t}^{a''(b-b')+(b-b'-a'')(n-k)+n-k}\cdot\bar{t}^{2b'(n-k)}\\&=\frac{1}{16r^2}\sum_{b'=0}^{4r-1}\frac{(s-\bar{s})^n}{[k]![n-k]!}\,\bar{t}^{b'(a'-a''+n)+(-a'k+a''b+bn-bk-a''n+a''k+n)}\\&=\frac{1}{16r^2}\,\frac{(s-\bar{s})^n}{[k]![n-k]!}\,\bar{t}^{(-a'k+a''b+bn-bk-a''n+a''k+n)}\sum_{b'=0}^{4r-1}\bar{t}^{b'(a'-a''+n)}.\end{aligned}$$

When $a'-a''+n\neq 0$, the sum on the right is zero, as it should be for there is no corresponding term in (2.30). If $a'-a''+n=0$, then the sum equals $4r$, and substituting $a'=k-n+a$ and $a''=k+a$, we obtain

$$\frac{1}{4r}\,\frac{(s-\bar{s})^n}{[k]![n-k]!}\,\bar{t}^{ab+(b-a)n+n}$$

as the common value of this coefficient in (2.30) and (2.31). □

(2.32) **Corollary** (a) *R acts on the module $V^k\otimes V^{k'}$ by*

$$R(e_i\otimes e_j)=\sum_n\frac{(s-\bar{s})^n}{[n]!}\,\frac{[m+i+n]!}{[m+i]!}\,\frac{[m'-j+n]!}{[m'-j]!}\,t^{4ij-2n(i-j)-n(n+1)}\,e_{i+n}\otimes e_{j-n}$$

where $k=2m+1$, $k'=2m'+1$, $[n]!=[n][n-1]\ldots[1]$ $(=1$ for $n=0)$, and $\dfrac{[p]!}{[n]!}=[p][p-1]\ldots[n+1]$. The sum is over all $n\geqq 0$ with $i+n\leqq m$ and $j-n\geqq -m'$. In particular, if $i=m$ or $j=-m'$, then $R(e_i\otimes e_j)=q^{ij}e_i\otimes e_j$.

(b) *If V is an $\mathscr{A}$-module and v is a vector in V of weight t^p (i.e. $Kv=t^p v$), then R maps $v\otimes e$ in $V\otimes V^1(\alpha)$ to $\alpha^p v\otimes e$, and $e\otimes v$ in $V^1(\alpha)\otimes V$ to $e\otimes\alpha^p v$. (For the definition of $V^1(\alpha)$, look above Lemma 2.16.)*

Proof. For the k-dimensional representation V^k with basis $e_m, e_{m-1}, \ldots, e_{-m}$, recall that $Ke_i=s^i e_i$, $Xe_i=[m+i+1]e_{i+1}$ and $Ye_i=[m-i+1]e_{i-1}$. It follows from the theorem that R acts on $V^k\otimes V^{k'}$ by

$$R(e_i\otimes e_j)=\frac{1}{4r}\sum_{n\geqq 0}\frac{(s-\bar{s})^n}{[n]!}\frac{[m+i+n]!}{[m+i]!}\frac{[m'-j+n]!}{[m'-j]!}\sum_{a,b}\bar{t}^{ab+(b-a)n+n-2ai-2bj}e_{i+n}\otimes e_{j-n}. \tag{2.33}$$

Observe that the exponent of $\bar{t}$ can be written as $(a+n-2j)(b-n-2i)+(n-2j)(n+2i)+n$. It is an elementary fact that

$$4r=\sum_{0\leqq a,b<4r}\bar{t}^{ab}=\sum_{0\leqq a,b<4r}\bar{t}^{(a+n-2j)(b-n-2i)}$$

so it follows that

$$\sum_{0\leqq a,b<4r}\bar{t}^{ab+(b-a)n+n-2ai-2bj}=4r\cdot\bar{t}^{n(n+1)-4ij+2n(i-j)}$$

and substitution in (2.33) gives

$$R(e_i\otimes e_j)=\sum_n\frac{(s-\bar{s})^n}{[n]!}\frac{[m+i+n]!}{[m+i]!}\frac{[m'-j+n]!}{[m'-j]!}\bar{t}^{n(n+1)+2n(i-j)-4ij}e_{i+n}\otimes e_{j-n}$$

which proves (a).

For (b), note that e is an element of the (twisted) 1-dimensional module $V^1(\alpha)$ with $\alpha=i^m=t^{rm}$ for some m (so $Ke=\alpha e=t^{rm}e$), and so we have

$$\begin{aligned}R(v\otimes e)&=\frac{1}{4r}\sum_{a,b}\bar{t}^{ab-ap-brm}v\otimes e\\&=t^{rmp}\left(\frac{1}{4r}\sum_{a,b}\bar{t}^{(a-rm)(b-p)}\right)v\otimes e.\end{aligned}$$

Thus $R(v\otimes e)=\alpha^p v\otimes e$, since the sum equals $4r$ as above. The other case follows in the same way. □

(2.34) *Remark* The action of the R-matrix in the modules $V^k\otimes V^{k'}$ given in the previous corollary can also be derived using Drinfeld's R-matrix in $\mathbf{U}_h$ [D2, p. 816],

$$R=\sum_{n=0}^{\infty}h^n Q_n(h)\,t^{H\otimes H+n(H\otimes 1-1\otimes H)}(X^+)^n\otimes(X^-)^n.$$

Here $X^{\pm}$ and H are generators for $\mathbf{U}_h$ with relations [D2, p. 807]

$$[H, X^{\pm}] = \pm 2X^{\pm}$$

$$[X^+, X^-] = \frac{2}{h}\sinh\left(\frac{hH}{2}\right)$$

$$= \frac{s-\bar{s}}{h}[H],$$

$$Q_n(h) = \bar{s}^n \prod_{k=1}^{n} \frac{q-1}{q^k-1} = \prod_{k=1}^{n} \frac{s-\bar{s}}{(s^k-\bar{s}^k)\,s^k}$$

$$= \prod_{k=1}^{n} \frac{1}{[k]}\,\bar{s}^k = \frac{1}{[n]!}\,\bar{t}^{n(n+1)}.$$

The generators X and Y correspond to $\left(\frac{h}{s-\bar{s}}\right)^{\frac{1}{2}} X^+$ and $\left(\frac{h}{s-\bar{s}}\right)^{\frac{1}{2}} Y^+$ (giving the relation $[X, Y] = [H]$), and substituting these in Drinfeld's R-matrix gives

$$R = \sum_{n=0}^{\infty} \frac{(s-\bar{s})^n}{[n]!}\, t^{H\otimes H + n(H\otimes 1 - 1\otimes H) - n(n+1)}\, X^n \otimes Y^n.$$

(Since $(s-\bar{s})^n \bar{t}^{n(n+1)} = (1-\bar{q})^n \bar{t}^{n(n-1)}$, it follows that there is a missing $\bar{t}^{n(n-1)}$ in the formula for the R-matrix in [RT1, § 7.4], and a missing $t^{n(n-1)}$ in [KiR, § 1.7].)

To compute the action of R on $e_i \otimes e_j$ in $V^k \otimes V^{k'}$, observe that $t^{H\otimes H} = 1\otimes 1 + \frac{h}{4} H\otimes H + \left(\frac{h}{4}\right)^2 \frac{1}{2!} H^2 \otimes H^2 + \ldots$ and $t^{n(H\otimes 1 - 1\otimes H)} = (K^n \otimes 1)(1 \otimes \bar{K}^n)$ $= K^n \otimes \bar{K}^n$, which gives

$$t^{H\otimes H} e_i \otimes e_j = t^{4ij} e_i \otimes e_j$$

$$t^{n(H\otimes 1 - 1\otimes H)} e_i \otimes e_j = t^{2n(i-j)} e_i \otimes e_j.$$

Thus

$$R(e_i \otimes e_j) = \sum_n \frac{(s-\bar{s})^n}{[n]!}\, t^{4(i+n)(j-n) + 2n((i+n)-(j-n)) - n(n+1)}$$

$$\cdot \frac{[m+i+n]!}{[m+i]!} \frac{[m'-j+n]!}{[m'-j]!}\, e_{i+n} \otimes e_{j-n},$$

which readily yields the formula given in Corollary 2.32.

(2.35) **Definition** The R-matrix, viewed as an operator on $V \otimes W$ for $\mathcal{A}$-modules V and W, can be composed with the permutation operator P to give an operator

$$\check{R} = P \circ R\colon\ V \otimes W \to W \otimes V$$

which we call the *$\check{R}$-matrix* (read "R flip matrix") on $V \otimes W$.

These are the operators associated with crossings in the definition of the colored framed link invariants.

(2.36) **Lemma.** *The $\check{R}$-matrices are $\mathcal{A}$-linear and satisfy the Yang-Baxter equation*

$$\check{R}_{23}\check{R}_{12}\check{R}_{23}=\check{R}_{12}\check{R}_{23}\check{R}_{12}$$

(as operators $U\otimes V\otimes W\to W\otimes V\otimes U$ for $\mathcal{A}$-modules U, V and W), where $\check{R}_{12}=\check{R}\otimes\mathrm{id}$ and $\check{R}_{23}=\mathrm{id}\otimes\check{R}$.

Proof. The first statement follows readily from the first defining property (2.17a) of R. Indeed, for X in $V\otimes W$, we have

$$\begin{aligned}\check{R}(\alpha X)&=P(R\Delta\alpha\cdot X)\qquad(\cdot=\text{diagonal action})\\&=P(P(\Delta\alpha)R\cdot X)\qquad(\text{by (2.17a)})\\&=\Delta\alpha P(R\cdot X)\\&=\alpha\check{R}(X).\end{aligned}$$

The second also follows from the defining properties of R. First we derive the Yang-Baxter equation for the R-matrix, namely

$$R_{12}R_{13}R_{23}=R_{23}R_{13}R_{12},$$

as follows:

$$\begin{aligned}R_{12}R_{13}R_{23}&=R_{12}\cdot(\Delta\otimes\mathrm{id})(R)\qquad(\text{by (2.17b)})\\&=(\check{\Delta}\otimes\mathrm{id})(R)\cdot R_{12}\qquad(\text{by (2.17a)})\\&=(P\otimes\mathrm{id})(R_{13}R_{23})\cdot R_{12}\qquad(\text{by (2.17b)})\\&=R_{23}R_{13}R_{12}.\end{aligned}$$

Now view this as an equation of operators $U\otimes V\otimes W\to U\otimes V\otimes W$ and multiply on the left by the operator $P_{23}P_{12}P_{23}=P_{12}P_{23}P_{12}$, where $P_{12}=P\otimes\mathrm{id}$ and $P_{23}=\mathrm{id}\otimes P$. Observing that $P_{ij}R_{ik}=R_{jk}P_{ij}$ for $j\neq k$, we obtain the Yang-Baxter equation for $\check{R}$. □

(2.37) **Examples** If the preferred weight basis $e_i\otimes e_j$ for $V^k\otimes V^{k'}$ is put in decreasing lexicographic order with respect to $(i+j, i, j)$, then the R-matrices given in Theorem 2.18 decompose into block sums, with constant $i+j$ in each block. For example, the R-matrix in $V^2\otimes V^2$ is given by

$$(t)\oplus\begin{pmatrix}\bar{t} & \bar{t}(s-\bar{s})\\ 0 & \bar{t}\end{pmatrix}\oplus(t)$$

(with respect to the basis $e_{1/2}\otimes e_{1/2}$, $e_{1/2}\otimes e_{-1/2}$, $e_{-1/2}\otimes e_{1/2}$, $e_{-1/2}\otimes e_{-1/2}$), and the corresponding $\check{R}$-matrix is

$$(t)\oplus\begin{pmatrix}0 & \bar{t}\\ \bar{t} & \bar{t}(s-\bar{s})\end{pmatrix}\oplus(t).$$

Similarly the $\check{R}$-matrix in $V^3 \otimes V^3$ is

$$(q) \oplus \begin{pmatrix} 0 & 1 \\ 1 & q-\bar{q} \end{pmatrix} \oplus \begin{pmatrix} 0 & 0 & \bar{q} \\ 0 & 1 & 1-\bar{q} \\ \bar{q} & (q-\bar{q})(1+\bar{q}) & (q-\bar{q})(1-\bar{q}) \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & q-\bar{q} \end{pmatrix} \oplus (q),$$

and the $\check{R}$-matrix in $V^2 \otimes V^3$ is

$$(s) \oplus \begin{pmatrix} 0 & \bar{s} \\ 1 & \bar{t}(q-\bar{q}) \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ \bar{s} & \bar{t}(q-\bar{q})(s+\bar{s}) \end{pmatrix} \oplus (s).$$

3 Tangle operators and link invariants

The link invariants $J_{L,\mathbf{k}}$ (see Definition 3.25) are special cases of the more general *tangle operators* which we define first (in Theorem 3.6).

Tangles

Recall that a *tangle* T is a 1-manifold properly embedded (up to isotopy) in the unit cube I^3 in $\mathbf{R}^3 = S^3 - \infty$, with $\partial T \subset \frac{1}{2} \times I \times \partial I$. Define $\partial_- T = T \cap (I^2 \times 0)$ and $\partial_+ T = T \cap (I^2 \times 1)$, and call T an (m, n)-*tangle* if $m = |\partial_- T|$ and $n = |\partial_+ T|$. Thus a link is a (0, 0)-tangle, and a general tangle consists of a link together with a collection of proper arcs. All tangles will be assumed oriented and perpendicular to $I^2 \times \partial I$.

A *framed tangle* is a tangle T equipped with a framing of its normal bundle (up to isotopy rel ∂T) which is standard $(\mathbf{i}, \pm\mathbf{j})$ on ∂T (where the sign is chosen so that the frame followed by the oriented tangent to T is the standard frame on $\mathbf{R}^3$). Since we are working in S^3, there is a natural 0-*framing* on each component of T, and so the framings may be specified by integers in the usual way. Alternatively, they may be specified by thickening the embeddings to ribbons in the direction of the second vector of the framing, as in the *homogeneous ribbon tangles* of [RT1].

As for links, one often studies tangles by their *diagrams* D in the square I^2 (obtained by regular projection onto $0 \times I^2$) with $\partial D \subset I \times \partial I$. Note that the 0-framing of a tangle T is in general different from the *blackboard framing* coming from a diagram D of T, in which the second vector is always parallel to $0 \times I^2$ (see Fig. 3.1).

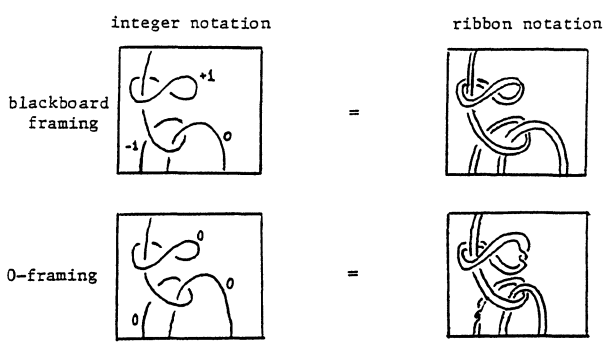


Fig. 3.1

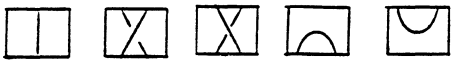


Fig. 3.2

Fig. 3.3

If the framing on T and the blackboard framing coincide, we call D a *good* diagram of T. Any diagram of T may be made into a good diagram by adding kinks. It is well known that every tangle diagram can be factored into the *elementary diagrams* I, R, L, $\cap$ and $\cup$ shown in Fig. 3.2 (with all possible orientations) using the *composition* $\circ$ (when defined) and the *tensor product* $\otimes$ of diagrams (see Fig. 3.3). Of course, distinct *factored diagrams* may represent the same tangle. For example $L \circ R = I \otimes I$ (with appropriate orientations), which may also be written as $\asymp = \|$. In fact, the following result follows easily from the work of Reidemeister [R].

(3.4) **Theorem** ([Ye, FY, T2, RT1]). *Any two factored good diagrams of a given framed tangle are related by a sequence of the following moves (with all possible orientations)*

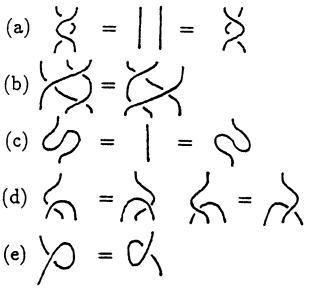


Fig. 3.4

together with the implicit associativity and identity relations and $(S \circ T) \otimes (S' \circ T') = (S \otimes S') \circ (T \otimes T')$ *(i.e. tangles are morphisms of a strict monoidal category, see e.g.* [FY]).

Remark. Moves (a)–(d) generate regular isotopy of tangle diagrams [Kf, Tr].

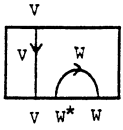


Fig. 3.5

Colored framed tangle operators F_T

Now fix a quasitriangular Hopf algebra (A, R) and define a *coloring* of a tangle T (or one of its diagrams) to be an assignment of an A-module to each component of T. This induces a coloring of ∂T as follows: if S is an arc of color V, then assign V to each endpoint of S where S is oriented down, and the dual module V^* to each endpoint where S is oriented up (see Fig. 3.5). Tensoring from left to right, this gives *boundary A-modules* $T_\pm$ assigned to $\partial_\pm T$ ($T_+ = V$ and $T_- = V \otimes W^* \otimes W$ in Fig. 3.5). By convention, the empty tensor product is $\mathbf{C}$, and so $T_\pm = \mathbf{C}$ if T is a link.

In the next result we show how to obtain tangle operators $T_- \to T_+$ for colored framed tangles T which behave well with respect to compositions and tensor products. This construction depends on some additional structure on the quasitriangular Hopf algebra (A, R) (with even more structure one obtains the *ribbon Hopf algebras* of Reshetikhin-Turaev, cf. Theorem 5.1 in [RT1] and Remark 3.16 below).

(3.6) **Theorem** *Let μ be an invertible element of a quasitriangular Hopf algebra $(A, R = \sum \alpha_i \otimes \beta_i)$ satisfying*
(a) $\mu \alpha \bar{\mu} = S^2(\alpha)$ *for all* α *in* A
(b) $\sum \alpha_i \bar{\mu} \beta_i = \sum \beta_i \mu \alpha_i$
where $\bar{\mu}$ denotes the inverse of μ. Then there exist unique A-linear operators

$$F_T = F_T^{A,R,\mu}\colon T_- \to T_+$$

assigned to each colored framed tangle T which satisfy $F_{T \circ T'} = F_T \circ F_{T'}$, $F_{T \otimes T'} = F_T \otimes F_{T'}$, and for the tangles given by the elementary diagrams (of Fig. 3.2) with the blackboard framing,

$$F_{|} = \mathrm{id}$$
$$F_{\mathbf{R}} = \check{R} \quad \text{and} \quad F_L = \check{R}^{-1}$$
$$F_{\curvearrowright} = E \quad \text{and} \quad F_{\curvearrowleft} = E_\mu$$
$$F_{\curvearrowright} = N \quad \text{and} \quad F_{\curvearrowleft} = N_{\bar{\mu}}$$

where $E(f \otimes x) = f(x)$, $E_\mu(x \otimes f) = f(\mu x)$, $N(1) = \sum e_i \otimes e^i$ and $N_{\bar{\mu}}(1) = \sum e^i \otimes (\bar{\mu} e_i)$ (for any basis e_i). Note that for a link L (i.e. $(0, 0)$-tangle), $F_L\colon \mathbf{C} \to \mathbf{C}$ is just a scalar.

Proof. First assign operators $F_D\colon D_- \to D_+$ to each elementary diagram D, as in the theorem. (Observe that orientations are implicit in the definitions of the first three operators. For example $\diagdown\!\!\diagup$ and $\diagdown\!\!\diagup$ are assigned $\check{R}$-matrices on
$U\ V$ $U\ V$

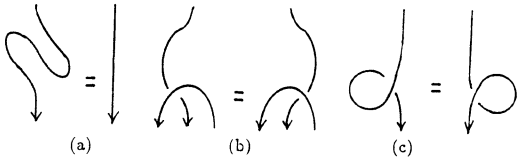


Fig. 3.7

$U \otimes V$ and $U^* \otimes V$, respectively.) A-linearity follows from Lemma 2.36 (for the crossings), and from the Hopf algebra axioms and property (3.6a) of μ (for the extrema). Indeed, for E: $V^* \otimes V \to \mathbf{C}$ we have, for all α in A

$$\begin{aligned} E(\alpha(f \otimes x)) &= f((m(S \otimes \mathrm{id})\, \Delta\alpha)\, x) \qquad \text{where } m \text{ is multiplication} \\ &= f(\varepsilon(\alpha)\, x) \qquad \text{by the antipode axiom} \\ &= \varepsilon(\alpha)\, f(x) \\ &= \alpha E(f \otimes x). \end{aligned}$$

Similarly N: $\mathbf{C} \to V \otimes V^*$ is A-linear. For E_μ: $V \otimes V^* \to \mathbf{C}$ we have

$$\begin{aligned} E_\mu(\alpha(x \otimes f)) &= E_\mu(\Delta\alpha(x \otimes f)) \\ &= E_\mu(\textstyle\sum (a_i\, x) \otimes (b_i\, f)) \qquad \text{where } \Delta\alpha = \sum a_i \otimes b_i \\ &= f(\textstyle\sum S(b_i)\, \mu a_i\, x) \\ &= f(\textstyle\sum S(b_i) S^2(a_i)\, \mu x) \qquad \text{by (3.6a)} \\ &= f(S(m(S \otimes \mathrm{id})\, \Delta\alpha)\, \mu x) \\ &= f(S(\varepsilon(\alpha)\, 1)\, \mu x) \qquad \text{by the antipode axiom} \\ &= f(\varepsilon(\alpha)\, \mu x) \\ &= \alpha E_\mu(x \otimes f). \end{aligned}$$

A similar argument shows that $N_{\bar{\mu}}$ is A-linear. Note that the role of μ (in the operators for the backward extrema) is essential, since the maps $x \otimes f \mapsto f(x)$ and $1 \mapsto \sum e^i \otimes e_i$ are not in general A-linear, due to the fact that the permutation map P is not in general A-linear.

Now extend the definition of F_D to arbitrary factored good diagrams D of colored framed tangles T by the rules $F_{D \circ D'} = F_D \circ F_{D'}$ and $F_{D \otimes D'} = F_D \otimes F_{D'}$. To show that these induce well defined operators F_T on tangles, it remains to show that F_D is invariant under the moves (3.4a–e).

Move (3.4a) follows from $\check{R}\check{R}^{-1} = I = \check{R}^{-1}\check{R}$, and (3.4b) follows from the Yang-Baxter equation of Lemma 2.36 (which holds, by the same proof, in any quasitriangular Hopf algebra). For moves (3.4c–e), we establish the cases shown in Fig. 3.7 and leave the rest as exercises.

For move (3.7a), which says $(E_\mu \otimes \mathrm{id})(\mathrm{id} \otimes N_{\bar\mu}) = \mathrm{id}$, we have

$$\begin{aligned}(E_\mu \otimes \mathrm{id})(\mathrm{id} \otimes N_{\bar\mu})(x) &= (E_\mu \otimes \mathrm{id})(\textstyle\sum x \otimes e^i \otimes \bar\mu e_i)\\ &= \textstyle\sum e^i(\mu x)\, \bar\mu e_i\\ &= \bar\mu \textstyle\sum e^i(\mu x)\, e_i\\ &= \bar\mu \mu x\\ &= x.\end{aligned}$$

For move (3.7b), which says $(\mathrm{id} \otimes E_\mu)(\check R \otimes \mathrm{id}) = (E_\mu \otimes \mathrm{id})(\mathrm{id} \otimes \check R^{-1})$, first note that the inverse of the R-matrix $R = \sum \alpha_i \otimes \beta_i$ can be computed easily using the antipode axiom as

$$R^{-1} = (S \otimes \mathrm{id})(R) = \textstyle\sum S(\alpha_i) \otimes \beta_i \tag{3.8}$$

(see e.g. [RT1, § 3.1.6]), and so

$$\begin{aligned}(\mathrm{id} \otimes E_\mu)(\check R \otimes \mathrm{id})(x \otimes y \otimes f) &= (\mathrm{id} \otimes E_\mu)(\textstyle\sum \beta_i\, y \otimes \alpha_i\, x \otimes f)\\ &= \textstyle\sum \beta_i\, yf(\mu \alpha_i\, x)\\ &= \textstyle\sum \beta_i\, yf(S^2(\alpha_i)\, \mu x) \quad \text{by (3.6a)}\\ &= (E_\mu \otimes \mathrm{id})(\textstyle\sum x \otimes S(\alpha_i)\, f \otimes \beta_i\, y)\\ &= (E_\mu \otimes \mathrm{id})(\mathrm{id} \otimes \check R^{-1})(x \otimes y \otimes f).\end{aligned}$$

Finally, move (3.7c) says

$$(E \otimes \mathrm{id})(\mathrm{id} \otimes \check R)(N_{\bar\mu} \otimes \mathrm{id}) = (\mathrm{id} \otimes E_\mu)(\check R \otimes \mathrm{id})(\mathrm{id} \otimes N).$$

The value of the left and right hand sides on an element x are readily computed as $(\sum \alpha_i\, \bar\mu \beta_i)\, x$ and $(\sum \beta_i\, \mu \alpha_i)\, x$, respectively, and these are equal by (3.6b). □

Properties of tangle operators

Now we establish various properties of tangle operators. We shall always assume that we are in the setting of Theorem 3.6, so

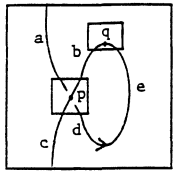
$$F_T = F_T^{A, R, \mu}$$

for some fixed quasitriangular Hopf algebra (A, R) and unit μ in A satisfying (3.6ab).

We begin with an elementary but useful fact about operators of (1, 1)-tangles.

(3.9) **Lemma** *Let T be a colored framed* (1, 1)*-tangle and V be the color of its (unique) arc component.*
(a) *If V is irreducible, then F_T is a scalar operator (i.e. a multiple of the identity).*
(b) *If V is reducible with a unique proper submodule, then F_T is the sum of a scalar operator and a nilpotent operator.*

Proof. Observe that any eigenspace for the operator F_T is a submodule of V (or V^* depending upon the orientation of the arc of T), since F_T is A-linear. The Lemma follows immediately by considering the Jordan canonical form of F_T. □



$a, b, c, d \in B_V$, $e \in B_V^*$
$\sigma_+(p) = a \otimes b$, $\sigma_-(p) = c \otimes d$
$\sigma_+(q) = 1$, $\sigma_-(q) = b \otimes e$
$\sigma_+ = a$, $\sigma_- = c$

Fig. 3.12

Next we consider how tangle operators behave under direct sums, extensions and tensor products of colors. Parts of this result were stated previously in §6.4 of [RT1].

(3.10) **Lemma** *Let T be a colored framed tangle and K be a component of color V. Write TX for the tangle obtained by changing the color on K to X.*
(a) *If $V = X \oplus Y$, or more generally V is an extension of Y by X (i.e. there is a short exact sequence $0 \to X \to V \to Y \to 0$ of A-modules), and K is closed, then*

$$F_T = F_{TX} + F_{TY}.$$

(see [RT1, §6.4.1] *for the case $V = X \oplus Y$).*
(b) *If $V = X \oplus Y$ and K is an arc between the bottom and top of the tangle, then*

$$F_T = F_{TX} \oplus F_{TY}$$

where the modules $T_\pm$ are naturally identified with $TX_\pm \oplus TY_\pm$.
(c) *If $V = X \otimes Y$, then*

$$F_T = F_{TXY}$$

where TXY is the tangle obtained by replacing K by two parallel pushoffs of itself (using the framing) colored X and Y, respectively. (See [RT1, §6.4.2].)

(3.11) *Remarks.* (1) Statements similar to (a) and (b) hold for arcs K joining the bottom or top of the tangle to itself.

(2) We will prove (a) using the following *states model* for F_T. Fix a factored good diagram D for T and preferred bases B_V for each color V. Let P be the set of critical points (i.e. extrema and double points) of D, and denote the elementary factor of D corresponding to a point p in P by Dp (i.e. Dp is the diagram in a small box about p).

A *state* σ of D is the assignment of a label $\sigma(S)$ to each component S of $D - P$ as follows: If S is V-colored, then $\sigma(S)$ is an element of B_V or B_V^* (the dual basis) according to whether S is oriented downward or upward. By taking tensor products, a state yields elements $\sigma_\pm(E)$ in the modules $E_\pm$ for any factor E of D. In particular, set $\sigma_\pm(p) = \sigma_\pm(Dp)$ and $\sigma_\pm = \sigma_\pm(D)$. (See Fig. 3.12).

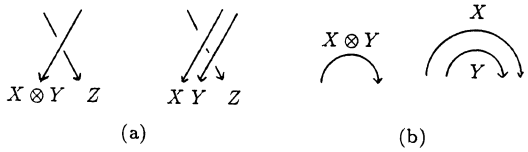


Fig. 3.14

Define the *weight* of a state σ to be the product

$$w(\sigma)=\prod_{p\in P} w_p(\sigma)$$

where $w_p(\sigma)$ is defined as the coefficient of $\sigma_+(p)$ in $F_{D_P}(\sigma_-(p))$.

Now consider *basic* elements of $T_\pm$, i.e. tensor products of preferred basis elements. The operator F_T is determined by the coefficients $(F_T)^{x_+}_{x_-}$ of x_+ in $F_T(x_-)$, for all basic $x_\pm$ in $T_\pm$, and these evidently have the following *states formula*

$$(F_T)^{x_+}_{x_-}=\sum w(\sigma) \tag{3.13}$$

where the sum is over all states σ with $\sigma_\pm = x_\pm$.

Proof of 3.10 For (a) we adopt the notation of the previous remark, choosing preferred bases B_V, B_X and B_Y so that $B_X \subset B_V$ (viewing X as a subspace of V) and B_Y is the projection of $\tilde{B}_Y = B_V - B_X$. All state labels from B_X or B_X^* will be called *X-labels*, and those from $\tilde{B}_Y$ or $\tilde{B}_Y^*$ will be called *Y-labels*.

Observe that if σ is a state of D with non-zero weight, then the corresponding labels on the arcs of K must be either all X-labels (written $\sigma \mid K \subset X$) or all Y-labels (written $\sigma \mid K \subset Y$). This follows from the A-invariance of $X \subset V$ (and dually of $Y^* \subset V^*$), which shows that one cannot move from an X-label to a Y-label while traversing K in the direction opposite to its orientation.

Now for basic $x_\pm$ in $T_\pm$, the states formula (3.13) gives

$$\begin{aligned}(F_T)^{x_+}_{x_-} &= \sum w(\sigma)\\ &= \sum_{\sigma \mid K \subset X} w(\sigma) + \sum_{\sigma \mid K \subset Y} w(\sigma)\\ &= (F_{TX})^{x_+}_{x_-} + (F_{TY})^{x_+}_{x_-}\end{aligned}$$

and so $F_T = F_{TX} + F_{TY}$.

The proof of (b) is similar but easier, and is left to the reader.

Finally, (c) follows from definitions, including the second defining property (2.17b) of the R-matrix. In particular, the two operators corresponding to a crossing involving K in a diagram D for T, and the corresponding crossings in the associated diagram of TXY, are equal. We illustrate this with the right crossing shown in Fig. 3.14a, where both operators map $X \otimes Y \otimes Z$ to $Z \otimes X \otimes Y$.

Indeed, with the obvious notation

$$\begin{aligned}\check{R}_{X\otimes Y,Z} &= P_{X\otimes Y,Z}\circ R_{X\otimes Y,Z}\\ &= P_{X\otimes Y,Z}\circ(\Delta\otimes \mathrm{id})(R)_{(X\otimes Y)\otimes Z}\\ &= P_{X\otimes Y,Z}\circ(R_{13}R_{23})_{X\otimes Y\otimes Z}\\ &= \sum_{i,j}\beta_i\beta_j\otimes\alpha_i\otimes\alpha_j\\ &= (\check{R}_{X,Z}\otimes \mathrm{id}_Y)\circ(\mathrm{id}_X\otimes\check{R}_{Y,Z})\end{aligned}$$

where $R=\sum\alpha_i\otimes\beta_i$. An easier argument shows that the operators corresponding to an extreme point of D, for example $E_{X\otimes Y}$ and $E_X\circ(\mathrm{id}\otimes E_Y\otimes\mathrm{id})$ for the case shown in Fig. 3.14b, are equal. The result follows by the definition of tangle operators. □

Orientations

Finally we consider orientation questions for the tangle operators $F_T=F_T^{A,R,\mu}$. It turns out that to get a reasonable theory, one must assume that the element μ and its antipode $S(\mu)$ are inverses. Such an element μ, i.e. a unit in A satisfying

(3.15) (a) $\mu\alpha\bar{\mu}=S^2(\alpha)$ for all α in A

(b) $\sum\alpha_i\bar{\mu}\beta_i=\sum\beta_i\mu\alpha_i$ where $R=\sum\alpha_i\otimes\beta_i$

(c) $S(\mu)=\bar{\mu}$,

is said to be *charmed*.

(3.16) *Remark* If (A, R, v) is a *ribbon Hopf algebra* in the sense of [RT1], then $\mu=u\bar{v}$ is charmed, where $u=\sum S(\beta_i)\alpha_i$. We do not know if a charmed element μ in an arbitrary quasitriangular Hopf algebra (A, R) gives rise to a ribbon structure (with $v=u\bar{\mu}$).

Observe that any μ satisfying (3.15a) induces an A-linear isomorphism

$$E_\mu\colon V\to V^{**} \tag{3.17}$$

for any A-module V, given by $E_\mu(x)=(\mu x)^{**}$ (=evaluation on μx), i.e. $E_\mu(x)(f)=f(\mu x)$. Indeed $E_\mu(\alpha x)=(\mu\alpha x)^{**}=(S^2(\alpha)\mu x)^{**}=\alpha(\mu x)^{**}=\alpha E_\mu(x)$. (Since this map is canonically identified with the map $E_\mu\colon V\otimes V^*\to\mathbf{C}$ in (3.6), we use the same notation.)

(3.18) **Lemma** *Let T be a colored framed tangle with a preferred component K of color V, and μ be charmed.*
(a) *If T^* is obtained from T by replacing K by $-K$ (opposite orientation) with color V^* (the dual module), then*

$$F_T=F_{T^*}$$

where (if K is not closed) $T_\pm$ and $T_\pm^*$ are identified by the isomorphism E_μ of (3.17) *between the V-colored endpoints of K and the V^{**}-colored endpoints of $-K$.*

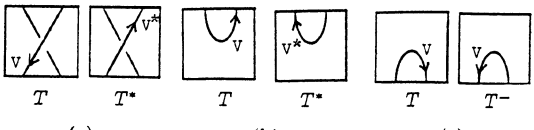
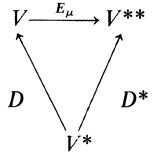


Fig. 3.19

(b) *If T^- is obtained from T by replacing K by $-K$ without changing the color, and if V is self dual by an isomorphism $D: V^* \to V$ for which $E_\mu = \pm D^* D^{-1}$, i.e.*



commutes up to sign, then

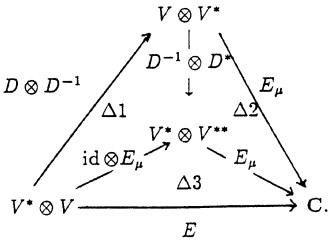
$$F_T = \varepsilon F_{T^-}$$

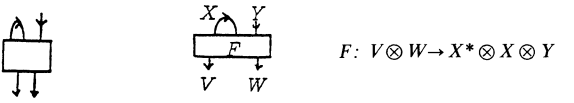
where (if K is not closed) $T_\pm$ and $T_\pm^-$ are identified by D between corresponding endpoints of K, and the sign $\varepsilon = \pm 1$ is -1 if and only if $E_\mu = -D^* D^{-1}$ and K is an arc joining one end of the tangle to itself. In particular $\varepsilon = +1$ if T is a link (i.e. (0, 0)-tangle).

Proof. It suffices to prove the lemma for elementary tangles, and for these it is straightfoward from definitions.

We illustrate the proof of (a) for the two (hardest) cases shown in Fig. 3.19ab. For (3.19a), we have $F_T(x \otimes y) = \sum \beta_i y \otimes \alpha_i x$ and $F_{T^*}(x \otimes y) = \sum \beta_i y \otimes E_\mu^{-1}(\alpha_i(\mu x)^{**})$. But these are equal by (3.15a), since $E_\mu(\alpha_i x) = (\mu \alpha_i x)^{**} = (S^2(\alpha_i)\mu x)^{**} = \alpha_i(\mu x)^{**}$. For (3.19b), we have $F_T(1) = \sum e_i \otimes e^i = \sum s(\mu) e_i \otimes \bar{\mu} e^i$ (note that $s(\mu) e_i$ and $\bar{\mu} e^i$ are dual bases) and $F_{T^*}(1) = \sum E_\mu^{-1}((e^i)^*) \otimes \bar{\mu} e^i$. These are equal, since $s(\mu) e_i = \bar{\mu} e_i$ and $E_\mu(\bar{\mu} e_i) = e_i^{**} = (e^i)^*$.

Note (b) follows from (a) by a diagram chase. For example, for the case shown in (3.19c) we must show $E = \varepsilon E_\mu (D \otimes D^{-1})$, where $E_\mu = \varepsilon D^* D^{-1}$, $\varepsilon = \pm 1$. Consider the diagram





(a) coupon (b) colored coupon

Fig. 3.21

The small triangle $\Delta 1$ ε-commutes by hypothesis, $\Delta 2$ commutes by the A-linearity of D^{-1}, and $\Delta 3$ commutes since $S(\mu)=\bar{\mu}$. Thus the outer triangle ε-commutes, as desired. The remaining cases are left to the reader. □

(3.20) *Remark* It is convenient to rephrase the previous lemma using *framed* (or *ribbon*) *graphs*, introduced in [RT1], which are formed from compositions and tensor products of framed tangles and *coupons*. A *coupon* can be thought of as an empty tangle (with diagram □) which is permitted to compose with arbitrary tangles, as shown in Fig. 3.21a. The coupons are thought of as the vertices of the framed graph. *Coloring* the framed graph then consists of coloring the edges (and loops) of the graph by A-modules and the vertices (i.e. coupons) by appropriate A-linear operators, as indicated in Fig. 3.21b. As with colored framed tangles, there is an operator F_G associated with any colored framed graph G.

Now in Lemma 3.18, the identification of $T_\pm$ with $T_\pm^*$ in (a) and $T_\pm^-$ in (b) can be accomplished by inserting coupons colored with E_μ and D (and their inverses). For example, the tangle operator equalities (after suitable identifications), illustrated in Fig. 3.19 become exact graph operator equalities, as shown in Fig. 3.22.

Note that the operator equality of Fig. 3.22c shows that pushing a $D^{\pm 1}$-colored coupon over a maximum changes the associated operator by a factor of ε, and the same remark holds for minima. As a consequence we have the following corollary of Lemma 3.18.

(3.23) **Corollary** *Let K be a closed V-colored component of a colored framed tangle T, and assume that μ is charmed that there is an isomorphism $D\colon V^*\to V$ of A-modules with $E_\mu=(-1)^m D^* D^{-1}$ for some integer m. If G is a colored framed graph formed by introducing two $D^{\pm 1}$-colored coupons on K (changing orientations appropriately) at points separated by p extreme points of K (in some good diagram of T), then $F_G=(-1)^{mp} F_T$.*

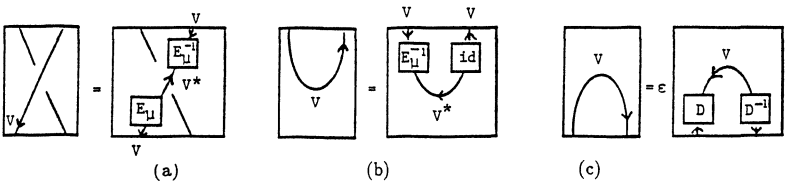


Fig. 3.22

Tangle operators for $\mathscr{A}$

We now specialize to the quasitriangular Hopf algebras $\mathscr{A}=\mathscr{A}_r$ (quantized sl(2, **C**)) discussed in § 2, with R given in (2.18).

(3.24) **Theorem** *The element $\mu=K^2$ in $\mathscr{A}$ is charmed, i.e. satisfies* (3.15a–c).

The proof will be given below. The associated tangle operators F_T will be denoted by J_T, in honor of V.F.R. Jones (see § 4):

(3.25) **Definition** For any integer $r>1$ and colored framed tangle T, with coloring **k**, define

$$J_T=J_{T,\mathbf{k}}=F_T^{\mathscr{A}_r,R,K^2}.$$

Note that the integer r, and sometimes the coloring **k**, will be suppressed in this notation. If T is a link L, then $J_{L,\mathbf{k}}$ is a scalar which will be called the *colored framed link invariant* of $(L, \mathbf{k})$.

(3.26) *Remark* The tangle operators $J_{T,\mathbf{k}}$ are independent of the orientation on any closed component K of T whose color is one of the irreducible modules V^k $(1\leqq k\leqq r)$. Indeed, the isomorphism $D\colon (V^k)^*\to V^k$ given in Theorem 2.13 satisfies

$$D^*D^{-1}=(-1)^{k-1}E_{K^2}$$

(since $D^*D^{-1}(b^j)=(-s)^{2j}b_j^{**}=(-1)^{k-1}q^jb_j^{**}=(-1)^{k-1}E_{K^2}(b^j)$), and so the remark follows from Lemma 3.18b since K has an even number of extreme points. In fact, from Corollary 3.23, we obtain the more refined result that

$$F_G=(-1)^{(k-1)p}F_T$$

if G is the colored framed graph obtained by introducing two $D^{\pm 1}$-colored coupons on K separated by p extreme points.

Proof of 3.24 Properties (3.15a) and (c) are immediate from the definition of $\mathscr{A}$. In particular $KX=sXK$, $KY=\bar{s}YK$, $S(K)=\bar{K}$, $S(X)=-sX$ and $S(Y)=-\bar{s}Y$ imply $K^2X\bar{K}^2=qX=S^2(X)$, $K^2Y\bar{K}^2=\bar{q}Y=S^2(Y)$ and $S(K^2)=\bar{K}^2$. Property (3.15b) is deeper, and is proved in Appendix A. □

We conclude this section with some computations of specific tangle operators associated with the algebra $\mathscr{A}=\mathscr{A}_r$, using as colors the irreducible modules V^k (often identified by their dimensions $k\leqq r$) and the associated Verma modules W_k^r (for $0<k<r$), defined in § 2.

We begin with a basic result about local modifications of tangles.

(3.27) **Lemma** *Fix a colored framed tangle T and a preferred component J of color j (i.e. V^j with $j\leqq r$). Let T_0 be T with a disjoint k-colored unknot adjoined, let T_1 be T with the framing on J increased by* 1, *and let T_2 be T with a k-colored unknotted meridian to J adjoined. Then*

$$J_{T_i}=c_iJ_T$$

where $c_0=[k]$, $c_1=t^{j^2-1}$ *and* $c_2=\dfrac{[jk]}{[j]}$ *(interpreted as* $(-1)^{k-1}k$ *when* $j=r$*). In pictures:*

(a) $J_{J\cup_k}=[k]J_J$ (c) $J_{J\,\phi_k}=\dfrac{[jk]}{[j]}J_J$

(b) $J_{J\circlearrowright_j}=t^{j^2-1}J_J$

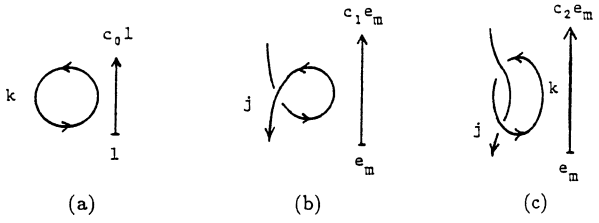


Fig. 3.28

Proof. It follows from Lemmas 3.9a and 3.18b that $J_{T_i} = c_i J_T$ for scalars c_i which are independent of orientations. Thus we may find the c_i by computing the values of the operators on 1 (for $i=0$) and e_m (for $i=1, 2$), where $m=\frac{j-1}{2}$, as indicated in Fig. 3.28.

For (a), $1 \mapsto \sum_{i=-n}^{n} e_i \otimes e^i \mapsto \sum e^i(K^2 e_i) = \sum q^i = [k]$ $\left(\text{where } n=\frac{k-1}{2}\right)$, and so $c_0 = [k]$.

For (b), $e_m \mapsto \sum_{i=-m}^{m} e_m \otimes e_i \otimes e^i \mapsto \sum q^{im} e_i \otimes e_m \otimes e^i$ (by Corollary 2.32) $\mapsto \sum q^{im} e^i(K^2 e_m)\, e_i = q^{m(m+1)} e_m = t^{j^2-1} e_m$, and so $c_1 = t^{j^2-1}$.

For (c), $e_m \mapsto \sum_{i=-n}^{n} e_m \otimes e_i \otimes e^i \mapsto \sum q^{mi} e_i \otimes e_m \otimes e^i \mapsto \sum q^{2mi} e_m \otimes e_i \otimes e^i$ (plus terms which will vanish at the next step) $\mapsto \sum q^{2mi} e^i(K^2 e_i)\, e_m = \sum q^{2mi+i} e_m$ $= \sum_{i=-n}^{n} q^{ji} e_m = \frac{[jk]}{[j]} e_m$ for $j<r$ $\left(\text{where } n=\frac{k-1}{2}\right)$, and so $c_2 = \frac{[jk]}{[j]}$ for $j<r$. For $j=r$, we have $q^{ji} = (-1)^{k-1}$, and so $c_2 = \sum_{i=-n}^{n} (-1)^{k-1} = (-1)^{k-1} k$. □

We conclude with a global result about links. Recall from Lemma 2.15 that W_k^r contains V^k as a unique proper submodule. In particular, using the standard basis $e_{r-m-1}, \ldots, e_{-m}$ coming from the inclusion $W_k^r \subset V^{2r-k}$ $\left(\text{where } m=\frac{k-1}{2}\right)$, V^k is spanned by $e_m, \ldots, e_{-m}$.

(3.29) **Lemma** *If a colored framed link* $(L, \mathbf{k})$ *has a component of color* V^r *or* W_k^r, *then* $J_{L,\mathbf{k}} = 0$.

Proof. Write L as the closure of a (1, 1)-tangle T as shown in Fig. 3.30 with $V = V^r$ or W_k^r.

If $V = V^r$, then by Lemma 3.9a, J_T is a scalar operator and so $J_{L,\mathbf{k}}$ is a scalar multiple of the invariant $J_{\bigcirc r}$ of the r-colored unknot. But this is just $[r] = 0$, by (3.27a).

If $V = W_k^r$, then J_T is still a scalar operator (cf. Lemma 3.9b). For, if λ is an eigenvalue for J_T, with eigenspace U, then U is a submodule and so $U \supset V^k$, i.e. $J_T(e_i) = \lambda e_i$ for $i \leqq m$. Thus we must show $U \neq V^k$, i.e. $J_T(e_i) = \lambda e_i$ for some

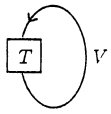


Fig. 3.30

$i>m$, for then $U=W_k^r$ and J_T is multiplication by λ. To see $J_T(e_{m+1})=\lambda e_{m+1}$, for example, linearity $J_T(\alpha e_{m+1})=\alpha J_T(e_{m+1})$ with $\alpha=Y$ shows that $J_T(e_{m+1})=\lambda e_{m+1}+\nu e_{-m}$ for some ν. But then linearity with $\alpha=K$, together with the fact that e_{m+1} and e_{-m} have distinct eigenvalues for K, shows $\nu=0$.

The proof is completed as in the case $V=V^r$ with the observation that

$$\begin{aligned} J_{\bigcirc} w_k^r &= \sum_{i=-m}^{r-m-1} q^i \\ &= \bar{q}^m \sum_{i=0}^{r-1} q^i \\ &= 0. \quad \square \end{aligned}$$

4 Skein theory, cabling and the symmetry principle

The computation of the $\mathcal{M}$-colored framed link invariant $J_{L,\mathbf{k}}$ (see § 1 and (2.12)) directly from the R-matrices defined in § 2 becomes impractical as the crossing number of L and the colors k_i increase. If all $k_i=2$, however, then $J_{L,\mathbf{k}}$ (as a function of $q=e\left(\frac{1}{r}\right)$) is just a variant of the Jones polynomial of L (Corollary 4.11) or the Kauffman bracket polynomial (Corollary 4.13) and can therefore be computed by the Conway skein calculus or (for $r=3$, 4 or 6) by topological means. Using this fact, we will give an expression for $J_{L,\mathbf{k}}$ as a linear combination of Jones polynomials of certain cables of L (Theorem 4.15). This will yield an alternative form for the 3-manifold invariant $\tau_r(M)$ (Theorem 4.17) which can be exploited for calculations.

At the end of this section we prove a symmetry principle (4.20) which describes the change in $J_{L,\mathbf{k}}$ when a color k is changed to $r-k$. This also leads to simplifications in the computation of $\tau_r(M)$, and appears to have interesting applications as well (including a simplified proof of the existence of $\tau_r(M)$; see § 5 below).

Skein theory, the Jones polynomial and the Kauffman bracket

(4.1) **Definition** Let L be a framed (unoriented) link. Define

$$J_L = J_{L,2}$$

where **2** denotes the constant 2-coloring (i.e. each component of L is colored with the untwisted 2-dimensional irreducible $\mathscr{A}$-module V^2).

The purpose of this subsection is to prove that J_L is equal to the value of the Jones polynomial at q, suitably normalized to account for the framing (Corollary 4.11), or equivalently the Kauffman bracket for a good diagram of L (Corollary 4.13).

Observe that for the zero framed m-component unlink $\bigcirc^m$, we have

$$J_{\bigcirc^m} = [2]^m = (s+\bar{s})^m \tag{4.2}$$

by m applications of Lemma 3.27a.

Furthermore, J_L satisfies the following skein relations.

(4.3) **Theorem** (1) *(oriented skein relations) Let L_+, L_- and L_0 be oriented framed links with good diagrams (i.e. the framings are the blackboard framings) which are identical except in a disc where they are as shown in Fig.* 4.4a. *Then*

(a) $$tJ_{L_+} - \bar{t}J_{L_-} = (s-\bar{s})J_{L_0}.$$

If the framings are adjusted so that $L_+ \cdot L_+ = L_- \cdot L_- = L_0 \cdot L_0$, then (a) *becomes*

(b) $$qJ_{L_+} - \bar{q}J_{L_-} = (s-\bar{s})J_{L_0}.$$

(2) *(unoriented skein relations) Let R, V and H denote unoriented framed links with identical good diagrams except as shown in Fig.* 4.4*b*. *Then*

(a) $$J_R = tJ_V + \bar{t}J_H$$

if the two strands in the crossing come from different components of R, and

(b) $$J_R = \varepsilon(tJ_V - \bar{t}J_H)$$

if the two strands come from the same component of R, producing a crossing of sign $\varepsilon = \pm 1$ (i.e. appearing as in L_ε of Fig. 4.4a *if R is oriented).*

Proof. Recall from Example 2.37 that action of $\check{R}$ on $V^2 \otimes V^2$ is given (in the preferred basis) by

$$\check{R} = (t) \oplus \begin{pmatrix} 0 & \bar{t} \\ \bar{t} & \bar{t}(s-\bar{s}) \end{pmatrix} \oplus (t).$$

We find (e.g. by computing the characteristic polynomial of $\check{R}$) that

$$t\check{R} - \bar{t}\check{R}^{-1} = (s-\bar{s})I, \tag{4.5}$$

and (1a) follows. To adjust the framings to become equal we may add a left kink to L_+ and a right kink to L_- (see e.g. Fig. 3.28b), which changes $J_{L_\pm}$ by $t^{\mp 3}$ by Lemma 3.27b, and so the coefficients of $J_{L_\pm}$ become $t^{\pm 1}t^{\pm 3} = q^{\pm 1}$. This gives (1b).

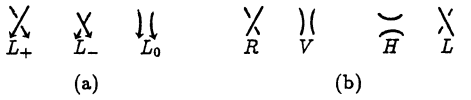


Fig. 4.4

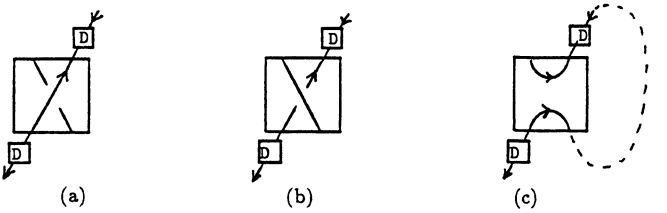


Fig. 4.8

For (2a), orient R so that the crossing looks like L_+, and then (1a) yields

$$tJ_R - \bar{t}J_L = (s-\bar{s})J_V \tag{4.6}$$

where L is as shown in Fig. 4.4b. Now reverse the orientation on one strand so that the same crossing looks like L_- when rotated by 90°, and so

$$-\bar{t}J_R + tJ_L = (s-\bar{s})J_H \tag{4.7}$$

by (1a) again. Multiplying (4.6) by t and (4.7) by $\bar{t}$ and adding gives (2a).

For (2b), first suppose $\varepsilon = 1$, so (4.6) follows as above by orienting R. Now we may locally reorient one strand by introducing two coupons on R, giving the framed graph (see Remark 3.20) R' shown in Fig. 4.8a, and similarly construct L' and H' (Figs. 4.8bc). As in (4.7) above, we get $-\bar{t}J_{R'} + tJ_{L'} = (s-\bar{s})J_{H'}$. But by Remark 3.26 we have $J_{R'} = J_R$ and $J_{L'} = J_L$ (since our color 2 is odd and there are no extreme points in Figs. 4.8ab) and $J_{H'} = -J_H$ (since there are an odd number of extreme points between the coupons in Fig. 4.8c), and so (4.7) is replaced by

$$-\bar{t}J_R + tJ_L = -(s-\bar{s})J_H. \tag{4.9}$$

Multiplying (4.6) by t and (4.9) by $\bar{t}$ and adding now gives (2b) for $\varepsilon = 1$.

If $\varepsilon = -1$, then the same argument establishes (4.7) and a revised (4.6), with the right hand side negated. This gives (2b) for $\varepsilon = -1$, as above. □

Remark. Theorem 4.3(2) can also be proved directly by computing the appropriate local tangle operators as in our proof of (1).

The skein relations in Theorem 4.3(1b) lead to a variant $\tilde{V}_L$ of the original Jones polynomial V_L of an *oriented* link L (in the variable q), characterized by

(1) $$\tilde{V}_{\bigcirc} = 1$$

(2) $$q\tilde{V}_{L_+} - \bar{q}\tilde{V}_{L_-} = (s-\bar{s})\tilde{V}_{L_0}$$

where $\bigcirc$ denotes the unknot and L_+, L_- and L_0 are as in Fig. 4.2. (Note that for V_L one swaps q and $\bar{q}$ on the left side of (2).) In fact

$$\tilde{V}_L = V_L(\bar{q}) \tag{4.10}$$

where $\bar{q}^{1/2}$ must be chosen to be $-\bar{s}$ on the right (i.e., recalling that $V_L(q)$ is a polynomial in $q^{\pm 1/2}$, we substitute $\bar{q}$ for q, $-\bar{s}$ for $q^{1/2}$ and $-s$ for $q^{-1/2}$). Equivalently $\tilde{V}_L$ may be defined as the specialization of the Homfly polynomial at $(iq, i(\bar{s}-s))$ (see e.g. [L2]).

(4.11) **Corollary** *If L is a framed link, then*

$$J_L = [2]\, t^{3L\cdot L}\, \tilde{V}_L$$

for any orientation on L.

Proof. The right hand side is characterized by the same skein relations, (4.2) and (4.3(1a)), as J_L. □

(4.12) *Remarks* (1) The values of $\tilde{V}_L$ at certain roots of unity have topological significance, as they do for V_L [LM1, Lp, Mu]. In particular, the values at $q = e\left(\frac{1}{r}\right)$, for $r = 1, 2, 3, 4$ and 6, are as follows:

r	$\tilde{V}_L$	V_L
1	2^{n-1}	$(-2)^{n-1}$
2	$\det L$	$\det L$
3	1	$(-1)^{n-1}$
4	$a\sqrt{2}^{n-1}$	$a(-\sqrt{2})^{n-1}$
6	$\sqrt{3}^{d}(-i)^{\omega}$	$(-\sqrt{3})^{d}(-i)^{\omega}$

where n is the number of components of L, $\det L$ is the value at -1 of the (normalized) Alexander polynomial of L, a is $(-1)^{\mathrm{Arf}(L)}$ when L is proper (so the Arf invariant is defined) and 0 otherwise, d is the nullity of $Q \pmod 3$ where Q is the quadratic form of L (represented by $S+S^t$ for any Seifert matrix S of L), and ω is the Witt class of $Q \pmod 3$ in $W(\mathbf{Z}/3\mathbf{Z}) = \mathbf{Z}/4\mathbf{Z}$ (see Appendix B). It is well known that $|\det L| = |H_1(M)|$, where M is the 2-fold branched cover of S^3 along L, and $d = \dim H_1(M; \mathbf{Z}/3\mathbf{Z})$ (since any matrix representing Q is a presentation matrix for $H_1(M)$). Our expression for the value of V_L at $e(\frac{1}{6})$ may appear unfamiliar, although it can be shown to be equivalent to Lipson's.

(2) It is often simpler to use J_L as the basic ingredient rather than $\tilde{V}_L$. For example, $J_\emptyset = 1$ is a better normalization than $\tilde{V}_\bigcirc = V_\bigcirc = 1$ since formulas are simpler (e.g. [2] disappears). Of course J_L does require a framing on L, but if one chooses a framing for which $L\cdot L = 0$, then $J_L = [2]\, \tilde{V}_L$ (cf. 4.3(1b)).

The skein relations in Theorem 4.3(2) remind one of Kauffman's bracket polynomial [Kf] (also see [L2]) in the variable t defined for a link diagram D. Our version $B_D(t)$ is normalized differently and is characterized by

(1) $$B_{\bigcirc^m} = (-[2])^m = (-s-\bar{s})^m \quad (s = t^2)$$

(2) $$B_R = tB_V + \bar{t}B_H$$

where $\bigcirc^m$ is the standard diagram of the unlink of m components, and R, V and H are diagrams which are identical except as shown in Fig. 4.4b.

(4.13) **Corollary** *If L is a framed link, then*

$$J_L=(-i)^{L\cdot L}B_D(it)$$

for any good diagram D of L.

Proof. First note that $L\cdot L$ is only defined for oriented links, but $L\cdot L(\mathrm{mod}\,4)$ is independent of the choice of orientation (i.e. $(A+B)(A+B)\equiv(A-B)\cdot(A-B)$ (mod 4)).

We prove the corollary by showing that the right side satisfies the same characterizing skein relations as J_L, namely (4.2) (obviously) and (4.3(2)). Observe that $R\cdot R\equiv V\cdot V+1\equiv H\cdot H-1(\mathrm{mod}\,4)$ if the two strands in the crossing belong to different components of R, whereas $R\cdot R\equiv V\cdot V+\varepsilon\equiv H\cdot H+\varepsilon$ if they belong to the same component. In the former case we compute

$$\begin{aligned}(-i)^{R\cdot R}B_R(it)&=(-i)^{V\cdot V+1}(it)B_V(it)+(-i)^{H\cdot H-1}(-i\bar{t})B_H(it)\\&=t((-i)^{V\cdot V}B_V(it))+\bar{t}((-i)^{H\cdot H}B_H(it))\end{aligned}$$

and the latter case

$$\begin{aligned}(-i)^{R\cdot R}B_R(it)&=(-i)^{V\cdot V+\varepsilon}(it)B_V(it)+(-i)^{H\cdot H+\varepsilon}(-i\bar{t})B_H(it)\\&=\varepsilon(t(-i)^{V\cdot V}B_V(it)-\bar{t}(-i)^{H\cdot H}B_H(it)).\end{aligned}$$ □

Remark. Alternatively, (4.13) can be proved directly from (4.10) and Corollary 4.11 by using the well known relation between the Jones polynomial and the bracket

$$[2]\ V_L(\bar{q})=(-(it))^{-3L\cdot L}B_D(it)$$

(see e.g. [L2], and note that the [2] is there because of our normalization of the bracket, and the i's are there because of the choice of $q^{1/2}$ in (4.10)).

Cabling

The next result gives a formula for the general $\mathscr{M}$-colored framed link invariant $J_{L,\mathbf{k}}$ in terms of Jones polynomials $\left(\text{at } q=e\left(\frac{1}{r}\right)\right)$ of certain cables of L. (Recall that $\mathscr{M}=\{V^1,\ldots,V^{r-1}\}$.)

We will need the following lemma (which will correspond to zero cabling).

(4.14) **Lemma** *Let $(L,\mathbf{k})$ be a colored, framed link. If S is a sublink of L obtained by removing some 1-colored components, then*

$$J_{L,\mathbf{k}}=J_{S,\mathbf{k}|S}.$$

Proof. It was observed in (2.9) that K^2 acts by the identity on V^1, as does $\check{R}$ on $V^1\otimes V^k$ and $V^k\otimes V^1$. Thus we may ignore 1-colored components of L when computing $J_{L,\mathbf{k}}$. □

Now define a *cabling* $\mathbf{c}$ of a framed link L to be an assignment of nonnegative integers c_i to the components L_i of L. The associated *cable* of L, denoted $L^{\mathbf{c}}$, is obtained by replacing each L_i with c_i parallel pushoffs (using the framing). If $c_i=0$, simply delete L_i.

If L is oriented, then there is a natural choice of orientation on $L^{\mathbf{c}}$: for each component L_i of L, orient the pushoffs so that their sum is homologous in a tubular neighborhood of L_i to L_i or to 0 (depending upon whether c_i is odd or even). With this choice we say L and $L^{\mathbf{c}}$ are *compatibly oriented.*

We will use the multi-index notation $f(\mathbf{k})=\prod f(k_i)$, $\mathbf{k}<\mathbf{n}$ if $k_i<n_i$ for all i, etc. For example, $(-1)^{\mathbf{k}}=\prod(-1)^{k_i}=(-1)^{\Sigma k_i}$, $\binom{\mathbf{n}}{\mathbf{k}}=\prod\binom{n_i}{k_i}$, and $\sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{n}}$ is the sum over all $\mathbf{k}$ with $1\leqq k_i\leqq n_i$.

(4.15) **Theorem** *Let L be a framed link and $\mathbf{k}$ be an $\mathscr{M}$-coloring of L. Then setting $\mathbf{n}=\mathbf{k}-\mathbf{1}$,*

$$J_{L,\mathbf{k}}=\sum_{\mathbf{j}=\mathbf{0}}^{\mathbf{n}/2}(-1)^{\mathbf{j}}\binom{\mathbf{n}-\mathbf{j}}{\mathbf{j}}J_{L^{\mathbf{n}-2\mathbf{j}}}$$

$$=[2]\sum_{\mathbf{j}=\mathbf{0}}^{\mathbf{n}/2}(-1)^{\mathbf{j}}\binom{\mathbf{n}-\mathbf{j}}{\mathbf{j}}t^{3L^{\mathbf{n}-2\mathbf{j}}\cdot L^{\mathbf{n}-2\mathbf{j}}}\tilde{V}_{L^{\mathbf{n}-2\mathbf{j}}}$$

for any orientation on $L^{\mathbf{n}-2\mathbf{j}}$. In particular, if L and $L^{\mathbf{n}-2\mathbf{j}}$ are compatibly oriented for all $\mathbf{j}$, then

$$J_{L,\mathbf{k}}=[2]\,t^{3S\cdot S}\sum_{\mathbf{j}=\mathbf{0}}^{\mathbf{n}/2}(-1)^{\mathbf{j}}\binom{\mathbf{n}-\mathbf{j}}{\mathbf{j}}\tilde{V}_{L^{\mathbf{n}-2\mathbf{j}}}$$

where S is the even colored sublink of L, consisting of all L_i with k_i even. (By convention $J_{L^0,2}=1$, $\tilde{V}_{L^0}=\dfrac{1}{[2]}$ and $L^0\cdot L^0=0$.)

Proof. The first equality is an immediate consequence of Corollary 2.15 and Lemmas 3.10 and 4.14. The second equality uses Corollary 4.11 and the last equality follows from the definition of compatibly oriented.

Remark. There is an analogous statement if $\mathbf{k}$ is only an $\mathscr{M}$-coloring on a sublink S of L,

$$J_{L,\mathbf{k}}=\sum_{\mathbf{j}=\mathbf{0}}^{\mathbf{n}/2}(-1)^{\mathbf{j}}\binom{\mathbf{n}-\mathbf{j}}{\mathbf{j}}J_{S^{\mathbf{n}-2\mathbf{j}}\cup(L-S),\,\mathbf{2}\cup\mathbf{k}|(L-S)}$$

where $\mathbf{n}=\mathbf{k}|S-\mathbf{1}$.

(4.16) *Examples* If K is a framed knot, then

$$J_{K,k}=\sum_{j=0}^{n/2}(-1)^j\binom{n-j}{j}J_{K^{n-2j}}$$

where $n=k-1$. In particular

$$J_{K,3}=J_{K^2}-1$$
$$J_{K,4}=J_{K^3}-2J_K$$
$$J_{K,5}=J_{K^4}-3J_{K^2}+1.$$

As a consequence of Theorem 4.15, we obtain a formula for $\tau_r(M_L)$ in terms of Jones polynomials $\left(\text{at } q=e\left(\frac{1}{r}\right)\right)$ of cables of L. Recall from §1 that

$$\tau_r(M_L)=\alpha_L \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{r}-\mathbf{1}} [\mathbf{k}]\, J_{L,\mathbf{k}}$$

where $\alpha_L=\left(\sqrt{\frac{2}{r}}\sin\frac{\pi}{r}\right)^{n_L}\left(e\left(\frac{-3(r-2)}{8r}\right)\right)^{\sigma_L}$.

(4.17) **Cabling Theorem** *For any framed link L,*

$$\tau_r(M_L)=\alpha_L \sum_{\mathbf{c}=0}^{\mathbf{r}-2} \langle \mathbf{c}\rangle\, J_{L^{\mathbf{c}}}$$

(i.e. sum over all cables $\mathbf{c}=(c_1, \ldots, c_n)$ *with* $0\leqq c_i\leqq r-2$*), where*

$$\langle \mathbf{c}\rangle = \sum_{\mathbf{j}=\mathbf{0}}^{(\mathbf{r}-\mathbf{c}-\mathbf{1})/2} (-1)^{\mathbf{j}}[\mathbf{c}+2\mathbf{j}+\mathbf{1}]\binom{\mathbf{c}+\mathbf{j}}{\mathbf{j}}$$

(i.e. sum over all $\mathbf{j}\geqq\mathbf{0}$ *with* $\mathbf{c}+2\mathbf{j}+\mathbf{1}<\mathbf{r}$*).*

The formula 4.17 can be rewritten in terms of the Jones polynomial variant $\tilde{V}$ by using the third equation in Theorem 4.15 and orienting cables compatibly:

(4.18) $$\tau_r(M_L)=\alpha_L[2]\sum_{\mathbf{c}} t^{3L_{\mathbf{c}}\cdot L_{\mathbf{c}}}\langle \mathbf{c}\rangle\, \tilde{V}_{L^{\mathbf{c}}}$$

where $L_{\mathbf{c}}$ is the sublink of L consisting of components L_i with c_i odd, (thus $L^{\mathbf{c}-2\mathbf{j}}\cdot L^{\mathbf{c}-2\mathbf{j}}=L^{\mathbf{c}}\cdot L^{\mathbf{c}}=L_{\mathbf{c}}\cdot L_{\mathbf{c}}$).

(4.19) *Remark* It is often easier to calculate with cables by first changing all framings of L to zero (and adjusting by the appropriate power of t) and then taking cables using the 0-framing.

Symmetry Principle

Finally we state the Symmetry Principle, which allows us to switch a color k to $r-k$. This cuts the number of terms in $\tau_r(M_L)$ from the order of $(r-1)^n$ to $\left(\frac{r}{2}\right)^n$, and makes possible an elementary proof of the invariance of τ under the m-strand k-move for $m>1$ (see § 5).

It is convenient to adopt the notation $L\cup K$ for a framed link with a distinguished component K. Colorings of $L\cup K$ will be written $\mathbf{l}\cup k$, where $\mathbf{l}$ is a coloring of L and k is the color of K. If the colors are selected from the modules V^1, V^2, ... defined in (2.8), then as above we identify these modules by their dimensions (also called colors) and so $\mathbf{l}$ is just a list of positive integers (i.e. l_i, or V^{l_i}, is the color of the component L_i).

(4.20) **Symmetry Principle** *Let* $L\cup K$ *be a framed link where* K *has framing* a*, and* $\mathbf{l}=(l_1, \ldots, l_n)$ *be a coloring of* $L=L_1\cup\ldots\cup L_n$ *by the modules defined in* (2.8)*. If* $0<k<r$*, then*

$$J_{L\cup K,\mathbf{l}\cup(r-k)}=i^{(r-2k)a+2\lambda}\, J_{L\cup K,\mathbf{l}\cup k}$$

where $\lambda = \sum_{\text{even } l_i} K \cdot L_i$. *(Note that the exponent of* i *can also be written as* $(r-2)K \cdot K + 2K \cdot E$ *where* E *is the even colored sublink of* $L \cup K$ *for the coloring* $\mathbf{l} \cup k$.*)*

Before proving the Symmetry Principle, we illustrate its use in the context of 3-manifolds given by surgery on a knot. A more systematic study of its applications appears in §8, including a form of the Symmetry Principle for τ_r (Theorem 8.5). Also see the end of this section for an application to Jones polynomials of cables.

(4.21) *Example* For $r=5$ and K a knot with framing a, we have

$$\begin{aligned}\tau_5(M_K) &= \alpha_K \sum_{k=1}^{4} [k]\, J_{K,k} \\ &= \alpha_K([1]+[2]\, J_K + [3]\, i^a\, J_K + [4]\, i^{3a}) \\ &= \alpha_K((1+i^{-a}) + [2](1+i^a)\, J_K).\end{aligned}$$

Thus if $a \equiv 2 \pmod 4$, then $\tau_5(M_K)=0$. If $a \not\equiv 2 \pmod 4$, then this shows that J_K (and so also the Jones polynomial of K at the fifth root of unity) is determined by $\tau_5(M_K)$, and thus is an invariant of M_K (cf. Theorem 8.22).

The proof of the Symmetry Principle needs

(4.22) **Lemma** *Let* $L \cup K$ *be a framed link, where* K *has framing zero, with colorings* $\mathbf{l}$ *on* L *and* $V^1(i)$ *on* K *(see above Lemma* 2.16*). Then*

$$J_{L \cup K, \mathbf{l} \cup V^1(i)} = (-1)^{1+\lambda} J_{L,\mathbf{l}}$$

where $\lambda = \sum_{\text{even } l_i} K \cdot L_i$.

Proof. Orient $L \cup K$ and draw it as a counterclockwise braid with the blackboard framing. We need to inspect three kinds of crossings as in Fig. 4.23.

In the first two cases, according to Corollary 2.32b, $\check{R}(e_j \otimes e) = i^{2j} e \otimes e_j$ and $\check{R}(e \otimes e_j) = i^{2j} e_j \otimes e$ since e_j has weight t^{2j} and $\alpha = i$. Similarly in the third case $\check{R}(e \otimes e) = i^r e \otimes e$ since e has weight $i = t^r$. On negative crossings we get the inverses $\check{R}^{-1}(e \otimes e_j) = i^{-2j} e_j \otimes e$, $\check{R}^{-1}(e_j \otimes e) = i^{-2j} e \otimes e_j$ and $\check{R}^{-1}(e \otimes e) = i^{-r} e \otimes e$ respectively.

Since K is 0-framed, K has an equal number of positive and negative crossings in the braid diagram and an odd number of maxima. Thus the self crossings of K together contribute nothing to $J_{L \cup K, \mathbf{l} \cup V^1(i)}$, and (since $K^2 = -1$ for $V^1(i)$) the maxima together contribute -1. That is, K contributes -1 (whether knotted or not). Hence we may change the self crossings of K so that K is unknotted, and then $L \cup K$ can be drawn as in Fig. 4.24. The crossings of K with L_i occur

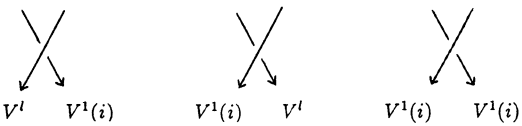


Fig. 4.23

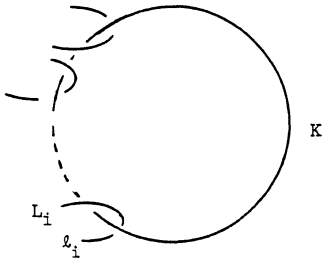


Fig. 4.24

in right or left handed pairs which algebraically sum to $K \cdot L_i$. So the contribution of pairs of crossings between K and L_i is $(-1)^{2jK \cdot L_i}$. When l_i is odd, then j is an integer so $(-1)^{2j}=1$ and there is no contribution. When l_i is even, then j is a half integer and $(-1)^{2j}=-1$, so we get a multiplicative contribution of $(-1)^{K \cdot L_i}$. □

Proof of the Symmetry Principle. Since we have the short exact sequence (2.16)

$$0 \to V^k \to W_k^r \to V^{r-k} \otimes V^1(i) \to 0$$

and since $J_{L,\mathbf{k}}$ is additive under extensions (3.10a) and $J_{L,W_k}=0$ (3.29), it follows that

$$0 = J_{L \cup K, \mathbf{1} \cup W_k} = J_{L \cup K, \mathbf{1} \cup k} + J_{L \cup K, \mathbf{1} \cup V^{r-k} \otimes V^1(i)}.$$

If the framing on K is zero, then replacing K with color $V^{r-k} \otimes V^1(i)$ by the 2-cable (using framing zero) of K with colors V^{r-k} and $V^1(i)$, it follows by Lemma 4.22 that we can eliminate the copy of K with color $V^1(i)$ by multiplying by $(-1)^{1+\lambda} = -i^{2\lambda}$ where $\lambda = \sum_{\text{even } l_i} K \cdot L_i$, and the result follows.

Finally, if the framing on K is a rather than zero, then we can change to framing zero by multiplying by $\bar{t}^{a(k^2-1)}$. Now apply the just proved zero framed case to switch the color on K from k to $r-k$, and then shift the framing back to a by multiplying by $t^{a((r-k)^2-1)}$. So the net change gives

$$\begin{aligned} J_{L \cup K, \mathbf{1} \cup (r-k)} &= t^{a((r-k)^2-1)} \cdot i^{2\lambda} \cdot \bar{t}^{a(k^2-1)} J_{L \cup K, \mathbf{1} \cup k} \\ &= i^{(r-2k)a+2\lambda} J_{L \cup K, \mathbf{1} \cup k}. \quad \square \end{aligned}$$

We conclude this section with an application of the Symmetry Principle to the study of the values $J_{L^{\mathbf{c}}}$ of Jones polynomials for cables $L^{\mathbf{c}}$ of a framed link L at a *fixed* root of unity $q=e\left(\frac{1}{r}\right)$.

A cabling $\mathbf{m}$ of L will be called *minimal* if $\mathbf{m} \leqq \frac{r}{2}-1$ $\left(\text{i.e. each component of } L \text{ is replaced by at most } \frac{r}{2}-1 \text{ parallel copies}\right)$. Now an easy inductive argument using the Symmetry Principle and the cabling formula (Theorem 4.15 and following remark) shows that for *any* cabling $\mathbf{c}$, $J_{L^{\mathbf{c}}}$ is an integer linear combination of the values $J_{L^{\mathbf{m}}}$ for minimal cablings $\mathbf{m}$ where the coefficients of the linear combination depend only on the linking matrix of L (and of course on $\mathbf{c}$ and r). That is, writing $J^{\mathbf{c}}(L)$ for $J_{L^{\mathbf{c}}}$:

(4.25) **Corollary** *For each framed link L and cablings* $\mathbf{c}$ *and* $\mathbf{m}$ *with* $\mathbf{m}$ *minimal, there exist integers* $a^{\mathbf{c}}_{\mathbf{m}}(L)$ *such that*

$$J^{\mathbf{c}}(L)=\sum_{\text{minimal cablings } \mathbf{m}} a^{\mathbf{c}}_{\mathbf{m}}(L)\, J^{\mathbf{m}}(L),$$

and such that $a^{\mathbf{c}}_{\mathbf{m}}(L)=a^{\mathbf{c}}_{\mathbf{m}}(L')$ *if L and L' have the same linking form.*

In practice, the integers $a^{\mathbf{c}}_{\mathbf{m}}(L)$ can be computed easily.

(4.26) *Example* Let $r=5$. Consider 0-framed knots K and let v^c denote the value J_{K^c} of the Jones polynomial of the c-cable of K at $e(\frac{1}{5})$. By the Symmetry Principle $J_{K,2}=J_{K,3}$, and so by (4.16), $v^1=v^2-1$. Induction shows that $v^c=f_c\, v^1+f_{c-1}$, where f_c is the c^{th} term in the Fibonacci sequence 1, 1, 2, 3, 5, ...

5 The 3-manifold invariant $\tau_r(M)$

Proof that τ_r is a 3-manifold invariant

If M is described by surgery on a framed link L, then we have defined (1.5 and 1.7)

$$\tau_r(M)=\tau_L=\alpha_L \sum_{\mathbf{k}=\mathbf{1}}^{\mathbf{r-1}} [\mathbf{k}]\, J_{L,\mathbf{k}}$$

where $\alpha_L=b^{n_L}\, c^{\sigma_L}$, $b=\sqrt{\frac{2}{r}}\sin\frac{\pi}{r}$, $c=e\left(\frac{-3(r-2)}{8r}\right)$, n_L is the number of components of L and σ_L is the signature of its linking matrix. For $\tau_r(M)$ to be well defined we must finally prove Theorem 1.6, that τ_L is invariant under K-moves on L, and hence $\tau_r(M)$ is independent of the choice of framed link used to describe M.

The proof of Theorem 1.6 depends on an elementary identity for Gauss sums. $\left(\text{Recall that } t=e\left(\frac{1}{4r}\right) \text{ and } s=e\left(\frac{1}{2r}\right).\right)$

(5.1) **Lemma** $\displaystyle\sum_{k=1}^{r-1} [j\,k]\,[k\,l]\, t^{j^2+k^2+l^2}=\frac{[j\,l]}{b}\, e\left(\frac{3}{8}\right).$

Proof. Note that

$$\begin{aligned}
&(s-\bar{s})^2 \sum_{k=1}^{4r} [j\,k]\,[k\,l]\, t^{j^2+k^2+l^2}\\
&=\sum_{k=1}^{4r} (t^{2jk}-\bar{t}^{2jk})(t^{2kl}-\bar{t}^{2kl})\, t^{j^2+k^2+l^2}\\
&=\sum_{k=1}^{4r} (\bar{t}^{2jl}(t^{(k+(j+l))^2}+t^{(k-(j+l))^2})-t^{2jl}(t^{(k+(j-l))^2}+t^{(k-(j-l))^2}))\\
&=2(\bar{t}^{2jl}-t^{2jl})\sum_{k=1}^{4r} t^{k^2}\\
&=2(\bar{s}^{jl}-s^{jl})\, 2\sqrt{2r}\, e(\tfrac{1}{8})
\end{aligned}$$

where the last equality is a standard Gauss sum (see [La]). Now $\sum_{k=1}^{r-1} = \frac{1}{4}\sum_{k=1}^{4r}$ by the symmetries of the bracket [ ], so

$$\sum_{k=1}^{r-1} [jk][kl]\, t^{j^2+k^2+l^2} = \frac{[jl]}{\bar{s}-s}\sqrt{2r}\, e\left(\frac{1}{8}\right)$$

$$= \frac{[jl]}{b}\, e\left(\frac{3}{8}\right)$$

since $\bar{s}-s = 2\sin\frac{\pi}{r}\, e(-\frac{1}{4}) = \sqrt{2r}\, b\, e(-\frac{1}{4})$. □

Now to prove Theorem 1.6 consider an m-strand K-move $L \leftrightarrow L^\varepsilon$ of type $\varepsilon = \pm 1$. Choose diagrams for L and L^ε which agree everywhere except for the tangles shown in Fig. 5.2.

For any $\mathcal{M}$-coloring $\mathbf{l}$ of L, let $\mathbf{l}\cup k$ denote the induced coloring of L^ε with the new component K colored $k<r$. Then $[\mathbf{l}\cup k] = [\mathbf{l}][k]$. Since $n_{L^\varepsilon} = n_L + 1$ and $\sigma_{L^\varepsilon} = \sigma_L + \varepsilon$, and so $\alpha_{L^\varepsilon} = b\, c^\varepsilon \alpha_L$, we have

$$\tau_L = \alpha_L \sum_{\mathbf{l}=1}^{r-1} [\mathbf{l}]\, J_{L,\mathbf{l}}$$

$$\tau_{L^\varepsilon} = \alpha_{L^\varepsilon} \sum_{\mathbf{l}=1}^{r-1} \left(\sum_{k=1}^{r-1} [\mathbf{l}\cup k]\, J_{L^\varepsilon,\mathbf{l}\cup k} \right)$$

$$= \alpha_L \sum_{\mathbf{l}=1}^{r-1} [\mathbf{l}] \left(b\, c^\varepsilon \sum_{k=1}^{r-1} [k]\, J_{L^\varepsilon,\mathbf{l}\cup k} \right).$$

Thus to prove $\tau_{L^\varepsilon} = \tau_L$ it suffices to establish the identity

$$b\, c^\varepsilon \sum_{k=1}^{r-1} [k]\, J_{L^\varepsilon,\mathbf{l}\cup k} = J_{L,\mathbf{l}} \tag{5.3}$$

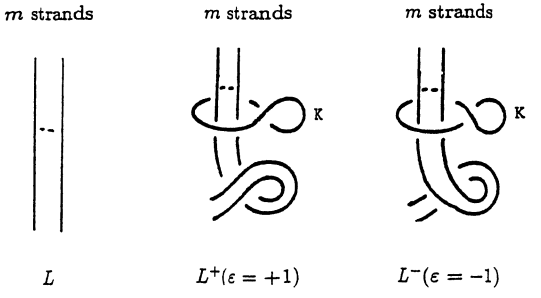


Fig. 5.2

for any fixed $\mathscr{M}$-coloring $\mathbf{l}$ on L. We will prove (5.3) by induction on the number of strands m, starting the induction for $m \leqq 1$ in the next result.

(5.4) **Lemma** *Identity* (5.3) *holds for* $m=0$ *and* 1, *and therefore* τ_L *is invariant under m-strand K-moves for* $m=0$ *and* 1.

Proof. The proof for $m=0$ is a special case of the proof for $m=1$ when the color j on the strand of L passing through K is 1, by Lemma 4.14. So we assume $m=1$. Then $J_{L^\varepsilon,\mathbf{l}\cup k}=\dfrac{[jk]}{[j]}\,t^{\varepsilon(j^2-1+k^2-1)}J_{L,\mathbf{l}}$ by Lemma 3.27, and so (5.3) reduces to

$$b\,c^\varepsilon\sum_{k=1}^{r-1}[k]\,\frac{[jk]}{[j]}\,t^{\varepsilon(j^2-1+k^2-1)}=1.$$

Since these two identities (for $\varepsilon=\pm 1$) are conjugate, we need only consider the case $\varepsilon=+1$. But this identity follows from Lemma 5.1, with $l=1$, since $c=e(-\frac{3}{8})t^3$. □

(5.5) *Remark* S^3 can be obtained by $+1$ surgery on the unknot, so

$$\tau_r(S^3)=b\,c\sum_{k=1}^{r-1}[k]^2\,t^{k^2-1}=1$$

by Lemma 3.27 and the identity established in the proof of (5.4) with $j=1$. Furthermore, it is not hard to show that τ_L is the only invariant of framed links under 1-strand K-moves of the form $a^\sigma\sum_{\mathbf{k}}\left(\prod_{i=1}^{n}d_{k_i}\right)J_{L,\mathbf{k}}$ with value 1 on the 1-framed unknot. (This is essentially how Reshetikhin and Turaev arrived at their formula.) Indeed, one readily shows as in the proof above that for any such invariant

$$a^\varepsilon\sum_{k=1}^{r-1}\left(\frac{[jk]}{[j]}\,t^{\varepsilon(j^2-1+k^2-1)}\right)d_k=1$$

for all $0<j<r$ and $\varepsilon=\pm 1$. Solving for the d_k, using Lemma 5.1 and the fact that the matrix $(b[jk])$ is its own inverse (this is the well known orthogonality relation $b^2\sum_{j=1}^{r-1}[ij][jk]=\delta_{ik}$), gives $d_k=b\,c^\varepsilon a^{-\varepsilon}[k]$. Equating the values for $\varepsilon=\pm 1$ shows $a=\pm c$ and the case $a=-c$ is eliminated by the normalization on the unknot. Thus $a=c$ and $d_k=b[k]$, so the invariant is just τ_L.

Finally we prove the inductive step, completing the proof of Theorem 1.6.

(5.6) **Lemma** *Identity* (5.3) *holds for m-strand K-moves for* $m>1$ *provided it holds for n-strand K-moves for all* $n<m$. *Thus* τ_L *is invariant under K-moves.*

Proof. First suppose that L and L^ε are trivial outside of the tangles shown in Fig. 5.2, as shown in Fig. 5.7 (with blackboard framings).

Using the Symmetry Principle 4.20, we may assume that all colors j of components J of L satisfy $j\leqq\frac{r}{2}$. Indeed, if $j>\frac{r}{2}$, then change j to $r-j\leqq\frac{r}{2}$ on J (and

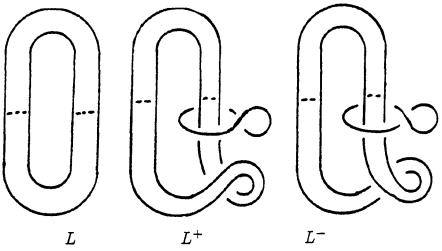


Fig. 5.7

on the corresponding component J^ε of L^ε). This changes the left side of (5.3) by

$$i^{\varepsilon(r-2j)+2(J^\varepsilon \cdot S+k-1)}$$

where S is the even-colored sublink of $L^\varepsilon-(J^\varepsilon \cup K)$, and leaves the right side unchanged. Next change k to $r-k$ on K. Then, using $[k]=[r-k]$, the left side of (5.3) changes by

$$i^{\varepsilon(r-2k)+2(J\cdot S+(r-j)-1)}$$

while the right side remains unchanged. Noting that $K\cdot S=J^\varepsilon\cdot S=|S|$, we see that the net change on the left side is

$$i^{2\varepsilon(r-j-k)+4(K\cdot S-1)+2(r-j+k)}=1$$

as it is on the right side.

Now by Lemma 3.10c, we may replace two components L_1 and L_2 of L, with colors l_1 and l_2, by a single component colored by $V^{l_1}\otimes V^{l_2}$ $=V^{l_1+l_2-1}\oplus\ldots\oplus V^{|l_1-l_2|+1}$ (using Theorem 2.13). Thus by Lemma 3.10a and distributivity, it is enough to establish (5.3) when L_1 and L_2 are replaced by a single j-colored component for $j<r$. But this is covered by the induction hypothesis.

Now consider the general case shown in Fig. 5.8, where T is an arbitrary tangle. We will reduce to the special case above (Fig. 5.7) using cabling and skein theory.

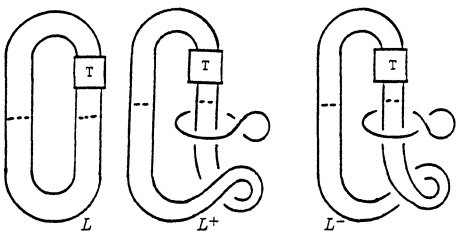


Fig. 5.8

First suppose that $\mathbf{l}=2$, the constant 2-coloring. Then we prove (5.3) by induction on the number of crossings in T. The induction begins with zero crossings, which is covered by the special case. (Note that T may have maxima and minima, as well as vertical strands, which simply pull through K, reducing m.) In general, we may smooth a crossing of T in two ways in both L and L^ε, and (5.3) follows by induction for each smoothing, using Theorem 4.3(2).

Finally, for general $\mathbf{l}$, $J_{L,\mathbf{l}}$ and $J_{L^\varepsilon,\mathbf{l}\cup k}$ can be computed using the cabling formula of Theorem 4.15 (and the subsequent remark) applied to L and $L^\varepsilon - K$, respectively. This reduces the proof of (5.3) to the case $\mathbf{l}=\mathbf{2}$ proved above. □

(5.9) **Theorem** *τ_r satisfies the following three properties*
(1) $\tau_r(M \# N)=\tau_r(M)\,\tau_r(N)$
(2) $\tau_r(-M)=\tau_r(M)$
(3) $\tau_r(S^3)=1$.

Proof. For (1), choose framed links L and L' with $M_L=M$, $M_{L'}=N$, and so $M_{L\cup L'}=M \# N$ where $L\cup L'$ denotes disjoint, (L and L' are separated by a 2-sphere). Note that

$$J_{L\cup L',\mathbf{k}\cup\mathbf{k}'}=J_{L,\mathbf{k}}\,J_{L',\mathbf{k}'}$$

(this is immediate from the definition of the colored framed link invariants, see the proof of Theorem 3.6), and so (1) follows from the definition of τ_r and distributivity.

For (2), observe that $(-M_L)=M_{\bar{L}}$, while $\bar{L}$ is the mirror image (obverse) of L. Now (2) follows from the Cabling Theorem 4.17 since $J_{\bar{L}}=\bar{J}_L$ and $\sigma_{\bar{L}}=-\sigma_L$.

Finally, (3) was shown in Remark 5.5. (Alternatively (3) follows from (1) once it is known that τ_r is nontrivial, i.e. $\tau_r(M)\neq 0$ for some M.) □

(5.10) *Remark* Observe that $\tau_2(M)=1$ for all M. Indeed $\tau_2(M_L)=J_{L,1}=1$. In the subsequent sections we will give formulas for $\tau_r(M)$ for small values of $r>2$.

Examples

Computations similar to the one made for S^3 in Remark 5.5 can be made for $S^2\times S^1$ and the lens spaces $L(p,1)$, obtained by surgeries on the unknot with framings 0 and p, respectively. For example

$$\tau_r(S^2\times S^1)=b\sum_{k=1}^{r-1}[k]^2=\frac{1}{b}=\sqrt{\frac{r}{2}}\csc\left(\frac{\pi}{r}\right) \tag{5.11}$$

(which approaches ∞ like $r^{3/2}$ as $r\to\infty$), and for *even* r

$$\tau_r(\mathbf{R}P^3)=b\,c\sum_{k=1}^{r-1}[k]^2\,t^{2(k^2-1)}=\frac{1}{\sqrt{2}}\sec\left(\frac{\pi}{2r}\right) \tag{5.12}$$

(which approaches $1/\sqrt{2}$ as $r\to\infty$)) where the last equality is derived as in the proof of Lemma 5.1 by expanding, completing the square, and using a Gauss sum. (Note that $\mathbf{R}P^3=L(2,1)$.) A similar argument shows that for r odd, $\tau_r(\mathbf{R}P^3)=0$ (cf. (8.9)). In [KM4], we will give general formulas for τ_r (lens space).

6 The case $r=3$

In this section we give two formulas for $\tau_3(M)$. The first (6.1) depends only on the linking matrix of L, where $M=M_L$. It follows that $\tau_3(M)$ (unlike $\tau_4(M)$, see §7) is a homotopy invariant (Remark 6.2). The second (Theorem 6.3) expresses $\tau_3(M)$ in terms of "classical" invariants of M.

Let L be a framed link with n components and signature σ. If $\mathbf{k}$ is a coloring of L with all colors 1 or 2, and S is the 2-colored sublink of L, then by Corollary 4.11 and Lemma 4.14, $J_{L,\mathbf{k}}=J_{S,\mathbf{2}}=[2]\,t^{3S\cdot S}\,\tilde{V}_S$ (for any orientation on S). In particular for $r=3$

$$J_{L,\mathbf{k}}=i^{S\cdot S}$$

since $\tilde{V}_S=1$ (Remark 4.12), $t^3=i$ and $[2]=1$. Thus the 3-manifold invariant (1.7) reduces to

$$\tau_3(M_L)=\frac{1}{\sqrt{2^n}}\,c^{\sigma}\sum_{S<L} i^{S\cdot S} \tag{6.1}$$

where $c=e\left(-\frac{1}{8}\right)=\frac{1-i}{\sqrt{2}}$. Here $<$ denotes sublink and $\emptyset\cdot\emptyset=0$ by convention. (Alternatively, (6.1) follows from the cabling formulas in Theorem 4.15 or from the Symmetry Principle 4.20.)

(6.2) *Remark* Evidently Formula (6.1) depends only on the linking matrix A of L. Since it is a 3-manifold invariant, it must be invariant under change of orientation on L, and under blowups and handle slides, that is under *stable equivalence* of A. It follows that $\tau_3(M_L)$ is a homotopy invariant, determined in fact by the first Betti number of M_L and the linking pairing on $\operatorname{Tor} H_1(M_L)$ (for it is known that these determine the stable equivalence class of A [KP, Du, Wk]).

Note that there is an easy direct proof of the invariance of Formula (6.1) under stable equivalence of A (giving an elementary proof, using the two moves in [K1], that τ_3 is a 3-manifold invariant). First observe that the formula is multiplicative under block sums of matrices (since σ is additive). Invariance under blowing up (summing with (± 1)) is now evident since $\frac{1}{\sqrt{2}}\,c^{\pm 1}(1+i^{\pm 1})=1$. Reversing the orientation on a component of L (multiplying a row and corresponding column by -1) leaves σ unchanged and alters $S\cdot S$ by a multiple of 4 for all sublinks S, and thus leaves the formula invariant. Finally consider handle slides. Let L' be obtained from L by sliding component L_i over L_j and then, for convenience, reversing the orientation on L_j (i.e. replace L_i by $L'_i=L_i+L_j$ and L_j by $L'_j=-L_j$). Each sublink S of L corresponds to a sublink S' of L' with $S\cdot S=S'\cdot S'$, namely

$$S'=\begin{cases} S & \text{if } S \text{ does not contain } L_i \\ S-(L_i+L_j)+L'_i & \text{if } S \text{ contains } L_i \text{ and } L_j \\ S-L_i+(L'_i+L'_j) & \text{if } S \text{ contains } L_i \text{ but not } L_j. \end{cases}$$

(In fact $S=S'$ as homology classes in W_L.) This correspondence is one-to-one, and so Formula (6.1) is invariant under handle slides.

The (cumbersome) sum in Formula (6.1) can be eliminated by using Ed Brown's $\mathbf{Z}/8\mathbf{Z}$ invariant λ of the linking matrix A of L, defined as follows ([Br, Ma]): View A as the matrix of a $\mathbf{Z}/4\mathbf{Z}$-valued quadratic form on a $\mathbf{Z}/2\mathbf{Z}$-vector space by reducing mod 4 along the diagonal and mod 2 off the diagonal. Two matrices are *Witt equivalent* if they represent the same form after possibly block summing with copies of $(1)\oplus(-1)$. It is easy to show that A is Witt equivalent to a diagonal matrix (mod 2). Let n_j denote the number of diagonal entries congruent to j (mod 4). Assume $n_2=0$. Then the Brown invariant is defined by

$$\lambda = n_1 - n_3 \,(\mathrm{mod}\ 8).$$

(If $n_2 \neq 0$, then the form is classified up to Witt equivalence by its nullity over $\mathbf{Z}/2\mathbf{Z}$, and the Brown invariant is not defined.)

Observe that if $n_2=0$ (which is equivalent to the topological statement that there exists α in $H^1(M;\mathbf{Z}/2\mathbf{Z})$ with $\alpha\smile\alpha\smile\alpha\neq 0$, see Theorem 6.3) then

$$\beta(M)=\sigma-\lambda \,(\mathrm{mod}\ 8)$$

is an invariant of the 3-manifold $M=M_L$ by [K1], since σ and λ change equally under blowing up and remain unchanged under handle slides. This will be called the *Brown invariant* of M. We can now state:

(6.3) **Theorem** *Let M be a closed, oriented 3-manifold. Then $\tau_3(M)=0$ if and only if any one of the following equivalent conditions holds:*
(1) *M has two spin structures with distinct μ-invariants* mod 4 *(see Appendix C)*
(2) *M contains an embedded closed surface of odd euler characteristic*
(3) *there exists α in $H^1(M;\mathbf{Z}/2\mathbf{Z})$ with $\alpha\smile\alpha\smile\alpha\neq 0$.*
Otherwise,

$$\tau_3(M)=\sqrt{2^{b(M)}}\,c^{\beta(M)}$$

where $b(M)=\mathrm{rk}\,H^1(M;\mathbf{Z}/2\mathbf{Z})$, $c=e(-1/8)$ and $\beta(M)$ is the Brown invariant (defined above).

Proof. First we show the equivalence of the three conditions. Let Θ_1 and Θ_2 be two spin structures on M. Following [KT, Theorem 4.11] or [T1], we have $\mu(\Theta_1)-\mu(\Theta_2)=2\beta(F)$ where F is a surface which is Poincaré dual to the class in $H^1(M;\mathbf{Z}/2\mathbf{Z})$ which measures the difference between Θ_2 and Θ_1; F gets a Pin^- structure from Θ_1 and $\beta(F)$ is its Pin^- bordism class in $\Omega_2^{\mathrm{Pin}^-}=\mathbf{Z}/8\mathbf{Z}$. Now the odd classes in $\Omega_2^{\mathrm{Pin}^-}$ are represented by odd multiples of $\mathbf{R}P^2$, and the even by even, and so the equivalence of (1) and (2) follows. The equivalence of (2) and (3) is well known, and follows from an elementary geometric argument (see for example [KT]).

Now choose a framed link L of n components and signature σ with $M=M_L$. Orient L and let A be the associated linking matrix. As above, we may assume that A is diagonal (mod 2) with n_j diagonal entries congruent to j (mod 4).

Observe that

$$b(M)\equiv n_0+n_2 \,(\mathrm{mod}\ 4)$$

since A is a presentation matrix for $H_1(M)$. By Formula (6.1),

$$\tau_3(M) = c^\sigma w$$

where

$$w = \frac{1}{\sqrt{2^n}} \sum_{s<L} i^{s \cdot s}.$$

Now, since w (as a function of A) is multiplicative under block sum and depends only on A mod 4 on the diagonal and mod 2 off the diagonal, we have

$$w = \prod_{j=0}^{3} w_j^{n_j}$$

where $w_0 = \sqrt{2}$, $w_1 = \bar{c}$, $w_2 = 0$ and $w_3 = c$. Thus

$$w = \begin{cases} \sqrt{2}^{n_0} c^{n_3 - n_1} = \sqrt{2}^{b(M)} c^{-\lambda} & \text{if } n_2 = 0 \\ 0 & \text{if } n_2 > 0 \end{cases}$$

and so

$$\tau_3(M) = \begin{cases} \sqrt{2}^{b(M)} c^{\beta(M)} & \text{if } n_2 = 0 \\ 0 & \text{if } n_2 > 0. \end{cases}$$

It remains to show that $n_2 = 0$ if and only if all the μ-invariants of M are congruent (mod 4). It is known that the spin structures on M are in one-to-one correspondence with the *characteristic* sublinks C of L (i.e. $C \cdot L_i \equiv L_i \cdot L_i \pmod 2$ for all components L_i of L), and their μ-invariants are given by

$$\mu_C = \sigma - C \cdot C + 8 \operatorname{Arf}(C) \pmod{16}$$

(see Appendix C).

It is evident that C is characteristic if and only if it contains all L_i with $L_i \cdot L_i$ odd (since $L_i \cdot L_j$ is even for $i \neq j$ by assumption). Now if $n_2 = 0$, then (working mod 4) $\mu_C \equiv \sigma - \lambda = \beta(M)$ for all characteristic C (since $C \cdot C \equiv \lambda$). If $n_2 > 0$, however, then $\mu_C \equiv \beta(M)$ if C contains an even number of L_i with $L_i \cdot L_i \equiv 2$, and $\mu_C = \beta(M) + 2$ otherwise. □

(6.4) **Corollary** *If M is a $\mathbf{Z}/2\mathbf{Z}$-homology sphere, then*

$$\tau_3(M) = \pm c^{\mu(M)}$$

where $c = e(-\frac{1}{8})$, $\mu(M)$ is the μ-invariant of M, and the sign is chosen according to whether $|H_1(M)| \equiv \pm 1$ or $\pm 3 \pmod 8$.

Proof. Since $b(M) = \operatorname{rk} H_1(M; \mathbf{Z}/2\mathbf{Z}) = 0$, we must show

$$\beta(M) \equiv \mu(M) + \delta(M) \pmod 8$$

where $\delta(M) = 0$ if $|H_1(M)| \equiv \pm 1 \pmod 8$, and $\delta(M) = 4$ otherwise.

First note that $n_0 = n_2 = 0$ (in the notation of the proof of Theorem 6.3). In addition, after a change of basis we may assume that A is diagonal (mod 4) with m_j diagonal entries congruent to $j \pmod 8$.

Working mod 8 we have

$$\lambda \equiv m_1 - m_3 + m_5 - m_7$$

$$|H_1(M)| \equiv 3^{m_3}\, 5^{m_5}\, 7^{m_7} \equiv \begin{cases} \pm 1 & \text{if } m_3 + m_5 \text{ is even} \\ \pm 3 & \text{otherwise.} \end{cases}$$

$$C \cdot C \equiv m_1 + 3m_3 + 5m_5 + 7m_7$$

where C is the (unique) characteristic sublink of L. Thus

$$\begin{aligned} \beta(M) &\equiv \sigma - \lambda \equiv \mu(M) + C \cdot C - \lambda \\ &\equiv \mu(M) + 4(m_3 + m_5) \equiv \mu(M) + \delta(M). \quad \square \end{aligned}$$

(6.5) *Remarks* (1) $\tau_3(M)$ is not in general determined by $H_1(M)$ and the μ-invariants of M (whereas $\tau_4(M)$ is, see § 7). For example, for $M = L(4, 1) \# L(8, 1)$ one readily computes $\beta(\pm M) = \pm 2$ whence $\tau_3(\pm M) = \pm 2i$. Yet M and $-M$ have the same homology and μ-invariants.

(2) Let $\nu(M) = \operatorname{rk} H_1(M)$. Then the modified invariant $c^{\nu(M)}\tau_3(M)$ (see Remark 1.8) is always a Gaussian integer. This follows from Theorem 6.3 and the elementary observation that $b(M) \equiv \beta(M) + \nu(M) \pmod 2$.

7 The case $r = 4$

In this section we give a formula for $\tau_4(M)$ in terms of the μ-invariants of spin structures on M (Theorem 7.1). It is derived using Rohlin's Theorem on the signature of spin 4-manifolds from a related formula (7.2) which involves the Arf invariants of sublinks of a framed link L with $M_L = M$. It turns out that Formula (7.2) can be shown directly to be an invariant of M using only elementary properties of the Arf invariant and [K1], and this in turn yields a new short proof of Rohlin's theorem (see Appendix C).

(7.1) **Theorem** *Let M be a closed, oriented 3-manifold. Then*

$$\tau_4(M) = \sum_{\Theta} c^{\mu(M_\Theta)}$$

where $c = e(-\frac{3}{16})$ and $\mu(M_\Theta)$ is the μ-invariant of the spin structure Θ on M (the sum is taken over all spin structures).

Proof. Choose a framed link L of n components and signature σ with $M_L = M$. By Theorem 4.17,

$$\tau_4(M) = \sqrt{2}^{1-2n} c^{\sigma} \sum_{\mathbf{c}=\mathbf{0}}^{\mathbf{2}} \langle \mathbf{c} \rangle J_{L^{\mathbf{c}}}$$

where the doubly cabled components of $L^{\mathbf{c}}$ are oppositely oriented, and

$$\langle \mathbf{c} \rangle = \bar{c}^{S \cdot S} \sum [\mathbf{c} + 2\mathbf{j} + \mathbf{1}](-1)^{\mathbf{j}} \binom{\mathbf{c}+\mathbf{j}}{\mathbf{j}}.$$

Here S is the sublink of L which is cabled once (S depends on $\mathbf{c}$), and the sum is over all $\mathbf{j} \geqq \mathbf{0}$ with $\mathbf{c}+2\mathbf{j}+\mathbf{1}<\mathbf{4}$.

First we show that $\langle \mathbf{c} \rangle = 0$ for any $\mathbf{c}$ in which some $c_p = 0$. Indeed, each $\mathbf{j}$ with $j_p = 0$ in the sum can be paired with $\mathbf{j}'$, identical to $\mathbf{j}$ except that $j'_p = 1$. The corresponding terms in the sum differ only in the p^{th} position, where we have $[0+0+1](-1)^0 \binom{0}{0} = 1$ for $\mathbf{j}$ and $[0+2+1](-1)^1 \binom{1}{1} = -1$ for $\mathbf{j}'$, and therefore cancel.

Now if $\mathbf{c}$ has all $c_i = 1$ or 2, then

$$\langle \mathbf{c} \rangle = \bar{c}^{S \cdot S} \sqrt{2}^{n_S}$$

where S (as above) is the sublink of all L_i with $c_i = 1$, and n_S is the number of components of S. Furthermore, recall from Remark 4.12 that $J_{L^c} = a\sqrt{2}^{2n-n_S-1}$, where $a = (-1)^{\text{Arf}(L^c)}$ or 0, depending upon whether L^c is proper or not (see Appendix C). But L^c is proper if and only if S is characteristic, since the components of $L-S$ are doubled. Hence

$$J_{L^c} = (-1)^{\text{Arf}(S)} \sqrt{2}^{2n-n_S-1}$$

if S is characteristic, and 0 otherwise.

Putting these calculations together gives

$$\tau_4(M) = \sqrt{2}^{1-2n} c^{\sigma} \sum_C \bar{c}^{C \cdot C} \sqrt{2}^{n_C} (-1)^{\text{Arf}(C)} \sqrt{2}^{2n-n_C-1} \tag{7.2}$$

$$= \sum_C c^{\sigma - C \cdot C + 8\,\text{Arf}(C)}$$

where the sum is over all characteristic sublinks C of L. It is shown in Appendix C that characteristic sublinks C of L naturally correspond to spin structures Θ on M, and the associated μ-invariants $\mu(M_\Theta)$ are given by $\sigma - C \cdot C + 8\,\text{Arf}(C)$ (mod 16) (see Eq. (C.3)). Thus Formula (7.2) may be written as in the statement of the theorem.

(7.3) *Remark* There is, of course, a quicker proof of Theorem 7.1 using the Symmetry Principle 4.20, which we leave to the reader.

8 Applications of the Symmetry Principle

In this section, the Symmetry Principle is used to simplify the formulas for $\tau_r(M)$, and to split $\tau_r(M)$ into finer invariants. Several applications are given. We begin by reformulating the Symmetry Principle.

The function $\phi_{\mathbf{k}}$ and the sum $m_{L,\mathbf{k}}$

Let L be a framed link and $\mathbf{k}$ be an $\mathcal{M}$-coloring of L. Denote the corresponding even colored sublink of L by $E_{\mathbf{k}}$. The Symmetry Principle (4.20) describes how

the invariant $J_{L,\mathbf{k}}$ changes if the color k on some component K is switched to $r-k$: it is multiplied by

$$i^{(r-2)K\cdot K+2E_{\mathbf{k}}\cdot K}.$$

Now we give a formula when several colors are switched in this way, and show how to apply this to the study of the 3-manifold invariant τ_r.

For each sublink S of L, let $\mathbf{k}_S$ denote the coloring of L obtained from $\mathbf{k}$ by switching the color as above on *each* component of S. Then applying the Symmetry Principle repeatedly, we have

$$J_{L,\mathbf{k}_S}=i^{\phi_{\mathbf{k}}(S)}J_{L,\mathbf{k}} \tag{8.1}$$

for some $\mathbf{Z}/4\mathbf{Z}$-valued function $\phi_{\mathbf{k}}$ on the sublinks of L, where

$$\phi_{\mathbf{k}}(K)=(r-2)K\cdot K+2E_{\mathbf{k}}\cdot K \pmod 4 \tag{8.2}$$

for a single component K of color k. Noting that there is a one to one correspondence between the sublinks of L and the elements of $H_2(W_L;\mathbf{Z}/2\mathbf{Z})$, where W_L is the 4-manifold defined by L (see §1), $\phi_{\mathbf{k}}$ may be viewed as a function

$$\phi_{\mathbf{k}}\colon H_2(W_L;\mathbf{Z}/2\mathbf{Z})\to\mathbf{Z}/4\mathbf{Z}.$$

(8.3) **Lemma** *If r is odd, then $\phi_{\mathbf{k}}$ is a quadratic enhancement of the $\mathbf{Z}/2\mathbf{Z}$ intersection form $\cdot$ on $H_2(W_L;\mathbf{Z}/2\mathbf{Z})$. If r is even, then $\phi_{\mathbf{k}}$ is linear.*

Proof. For r odd, we must show

$$\phi_{\mathbf{k}}(S+T)=\phi_{\mathbf{k}}(S)+\phi_{\mathbf{k}}(T)+2(S\cdot T)\pmod 4$$

for any (sublinks) S and T in $H_2(W_L;\mathbf{Z}/2\mathbf{Z})$. Note that by (8.2), $\phi_{\mathbf{k}}(K)+K\cdot K$ is even (since r is odd), and so this is immediate for $S=T=K$. The general case now reduces easily to the case when $T=K$ and S is an arbitrary sublink of L not containing K.

In this case, consider $S_0=E_{\mathbf{k}}\cap S$ (the even colored sublink of S) and $S_1=S-S_0$ (the odd colored sublink of S). Now compute $\phi_{\mathbf{k}}(S+K)$ by first switching colors on the components of S (this gives $\phi_{\mathbf{k}}(S)$), and then switching the color on K (which adds $(r-2)\ K\cdot K+2(E_{\mathbf{k}}+S_1-S_0)\cdot K\equiv\phi_{\mathbf{k}}(K)+2(S_1-S_0)\cdot K \pmod 4$, since the parities of all the colors on the components of S have changed). Hence,

$$\phi_{\mathbf{k}}(S+K)=\phi_{\mathbf{k}}(S)+\phi_{\mathbf{k}}(K)+2(S\cdot K)\pmod 4$$

since $S\cdot K=(S_1+S_0)\cdot K\equiv(S_1-S_0)\cdot K \pmod 2$.

When r is even, $\phi_{\mathbf{k}}(K)$ is even by (8.2). The argument now proceeds as above except that the parities of the colors on S remain unchanged, and so switching the color on K simply adds $\phi_{\mathbf{k}}(K)$. Thus $\phi_{\mathbf{k}}(S+K)=\phi_{\mathbf{k}}(S)+\phi_{\mathbf{k}}(K)\pmod 4$. □

For any function $\phi\colon V\to\mathbf{Z}/4\mathbf{Z}$ on an n-dimensional $\mathbf{Z}/2\mathbf{Z}$-vector space V, consider the *Monsky sum* [Br]

$$m_\phi=\sum_{v\in V}i^{\phi(v)}.$$

If ϕ is linear, then m_ϕ vanishes unless ϕ is identically zero, in which case $m_\phi = 2^n$ [Br, Lemma 3.1]. If ϕ is quadratic (i.e. $\phi(v+w)=\phi(v)+\phi(w)+2v\cdot w$ for some symmetric bilinear form $\cdot$ on V) then m_ϕ is either 0 or of the form $\sqrt{2^d}\, e(1/8)^\beta$, where $d = n + \text{nullity}_{\mathbf{Z}/2\mathbf{Z}}\phi$ and β is the Brown invariant of ϕ (see §6, or [Br, Theorem 1.20] for the nonsingular case).

Now, for any coloring $\mathbf{k}$ of L, let

$$m_{L,\mathbf{k}} = m_{\phi_\mathbf{k}|T_\mathbf{k}} = \sum_{S<L} i^{\phi_\mathbf{k}(S)} \tag{8.4}$$

where $T_\mathbf{k}$ is the sublink of L consisting of all components with colors unequal to $r/2$ (and of course $m_{L,\mathbf{k}}=1$ if $T_\mathbf{k}$ is empty). Note that $m_{L,\mathbf{k}} = m_{\phi_\mathbf{k}}$ for r odd, since $T_\mathbf{k} = L$ in this case.

In view of (8.1), these sums may be used to compute the 3-manifold invariants $\tau_r(M)$, where $M = M_L$. In particular, define two colorings $\mathbf{k}$ and $\mathbf{k}'$ of L to be *equivalent* if on each component, the corresponding colors are either equal or add up to r. Note that each equivalence class contains $2^{|T_\mathbf{k}|}$ elements, and exactly one of these is minimal (where $\mathbf{k}$ is called *minimal* if no color exceeds $r/2$, also written $\mathbf{k} \leqq \mathbf{r}/2$ in multi-index notation). Now we may group the colorings into equivalence classes and rewrite $\tau_r(M) = \alpha_L \sum_{0<\mathbf{k}\leqq\mathbf{r}} [\mathbf{k}]\, J_{L,\mathbf{k}}$ from Definition 1.5 (using $[k]=[r-k]$):

(8.5) **Theorem** $\tau_r(M) = \alpha_L \sum_{0<\mathbf{k}\leqq\mathbf{r}/2} m_{L,\mathbf{k}}[\mathbf{k}]\, J_{L,\mathbf{k}}$.

This reduces the number of terms by roughly a factor of 2^n where n is the number of components in L. Thus it is of interest to evaluate the sums $m_{L,\mathbf{k}}$.

(8.6) **Lemma** *Let L be a framed link and $\mathbf{k}$ be a coloring of L. If r is even, then*

$$m_{L,\mathbf{k}} = \begin{cases} 2^{|T_\mathbf{k}|} & \text{if } \phi_\mathbf{k}|T_\mathbf{k} \text{ is identically zero} \\ 0 & \text{otherwise} \end{cases}$$

where $T_\mathbf{k}$ is the sublink of L consisting of components with colors strictly less than $r/2$ in the coloring $\mathbf{k}$.

For r odd,

$$m_{L,\mathbf{k}} = i^{rE_\mathbf{k}\cdot E_\mathbf{k}} m_{L,\mathbf{1}},$$

where $\mathbf{1}$ is the constant 1-coloring and $E_\mathbf{k}$ is the even colored sublink of L for the coloring $\mathbf{k}$.

Proof. The formula for even r follows immediately from the fact that $\phi_\mathbf{k}$ is linear (Lemma 8.3) and the remarks above. For r odd, it suffices to establish the following:

Assertion. *If $\mathbf{k}$ and $\mathbf{k}'$ are colorings of L which differ on only one component K, with colors k and k' respectively, then*

$$m_{L,\mathbf{k}} = \begin{cases} m_{L,\mathbf{k}'} & \text{if } k \equiv k' \pmod 2 \\ i^{\phi_\mathbf{k}(K)} m_{L,\mathbf{k}'} & \text{if } k \not\equiv k' \pmod 2. \end{cases}$$

The formula in the Lemma then follows by induction on the number of components in $E_\mathbf{k}$ (since for k even, $E_\mathbf{k} = E_{\mathbf{k}'} \cup K$, and so as r is odd, $rE_\mathbf{k}\cdot E_\mathbf{k}$

$= rE_{\mathbf{k}'} \cdot E_{\mathbf{k}'} + 2rE_{\mathbf{k}'} \cdot K + rK \cdot K \equiv rE_{\mathbf{k}'} \cdot E_{\mathbf{k}'} + 2E_{\mathbf{k}} \cdot K + (r-2)K \cdot K \equiv rE_{\mathbf{k}'} \cdot E_{\mathbf{k}'} + \phi_{\mathbf{k}}(K)$ (mod 4)).

The assertion is obvious for $k \equiv k' \pmod 2$, since $\phi_{\mathbf{k}} = \phi_{\mathbf{k}'}$ by (8.2), and so we assume that k and k' have opposite parities.

First some notation. We shall write ϕ for $\phi_{\mathbf{k}}$ and ϕ' for $\phi_{\mathbf{k}'}$; these are both quadratic forms on $V = H_2(W_L; \mathbf{Z}/2\mathbf{Z})$ by Lemma 8.3. Observe that $E_{\mathbf{k}'} = E_{\mathbf{k}} + K$ (in V) and so

$$\phi'(S) \equiv \phi(S) + 2(S \cdot K) \pmod 4$$

for any S in V (e.g. $\phi'(K) \equiv -\phi(K) \pmod 4$). Also, let V_1 be the 1-dimensional $\mathbf{Z}/2\mathbf{Z}$-vector space generated by N with $N \cdot N = 1$, and $\phi^{\pm}$ be the quadratic forms on V_1 given by $\phi^{\pm}(N) = \pm 1 \pmod 4$. Note that

$$m_{\phi^{\pm}} = 1 \pm i.$$

Now to prove the assertion, it remains to show that $m_{\phi} = i^{\phi(K)} m_{\phi'}$. There are three cases.

Case 1 $\phi(K) \equiv 0 \pmod 4$. Then ϕ and ϕ' are equivalent, i.e. there is an isometry T on $(V, \cdot)$ with $\phi' = \phi \circ T$. Indeed, define $T(S) = S + (S \cdot K)K$ for any S in V. (In the language of the calculus of framed links, we slide over K each component of $L - K$ which links K oddly.) It follows that $m_{\phi} = m_{\phi'} = i^{\phi(K)} m_{\phi'}$.

Case 2 $\phi(K) \equiv \pm 1 \pmod 4$. Then $\phi \oplus \phi^{+}$ and $\phi' \oplus \phi^{\pm}$ are equivalent (as forms on $V \oplus V_1$). Indeed, an explicit isometry T is given by $T(S) = S + (S \cdot K)(K + N)$ for S in V (e.g. $T(K) = N$) and $T(N) = K$. (As framed links, we slide off K and over N each component of $L - K$ which links K oddly.) It follows that $(1 + i)m_{\phi} = (1 \pm i)m_{\phi'}$, since Monsky sums multiply under direct sums, and so $m_{\phi} = i^{\pm 1} m_{\phi'} = i^{\phi(K)} m_{\phi'}$.

Case 3 $\phi(K) \equiv 2 \pmod 4$. Then $\phi \oplus \phi^{+} \oplus \phi^{+}$ and $\phi' \oplus \phi^{-} \oplus \phi^{-}$ are equivalent (as forms on $V \oplus V_1 \oplus V_1$). To see this, consider $\psi = \phi \oplus \phi^{+}$ and $\psi' = \phi' \oplus \phi^{-}$. Note that $\psi(K + N) = -1$ and $\psi'(K + N) = 1$, and so we are in the situation of case 2 (with ϕ replaced by ψ and K replaced by $K + N$). Thus $\psi \oplus \phi^{+}$ and $\psi' \oplus \phi^{-}$ are equivalent. It follows that $(1 + i)^2 m_{\phi} = (1 - i)^2 m_{\phi'}$, and so $m_{\phi} = -m_{\phi'} = i^{\phi(K)} m_{\phi'}$. □

The case of odd r

Combining Theorem 8.5 and Lemma 8.6 we obtain the following formula for $\tau_r(M)$ when r is odd (where $M = M_L$ as usual).

(8.7) **Theorem** *If r is odd, then*

$$\tau_r(M) = m_{L,\mathbf{1}}\, \alpha_L \sum_{0 < \mathbf{k} < \mathbf{r}/2} i^{rE_{\mathbf{k}} \cdot E_{\mathbf{k}}} [\mathbf{k}]\, J_{L,\mathbf{k}},$$

where $E_{\mathbf{k}}$ *is the even colored sublink of L for the coloring* $\mathbf{k}$.

We now derive some consequences of this formula.

Splitting for odd r

Observe that there is a natural quadratic form ϕ_L on $(H_2(W_L; \mathbf{Z}/2\mathbf{Z}), \cdot)$ given by

$$\phi_L(S) = S \cdot S \pmod 4$$

with associated Monsky sum

$$m_L = m_{\phi_L} = \sum_{S<L} i^{S\cdot S} = \sqrt{2}^n \, e\left(\frac{\sigma}{8}\right) \tau_3(M), \tag{8.8}$$

where the last equality follows from the formula for $\tau_3(M)$ in (6.1). It is readily verified that

$$m_{L,\mathbf{l}} = \begin{cases} m_L & \text{if } r \equiv 3 \pmod 4 \\ \overline{m_L} & \text{if } r \equiv 1 \pmod 4 \end{cases}$$

and so we deduce from Theorem 8.7 that τ_r splits as a product for odd r:

(8.9) **Corollary** *If r is odd, set*

$$\tau'_r(L) = \sqrt{2}^n \, e\left(\frac{\pm\sigma}{8}\right) \alpha_L \sum_{0<\mathbf{k}<\mathbf{r}/2} i^{rE_\mathbf{k}\cdot E_\mathbf{k}} [\mathbf{k}] J_{L,\mathbf{k}}$$

with the + or − sign chosen according to whether $r \equiv 3$ *or* $r \equiv 1 \pmod 4$. *(Here n is the number of components in L and σ is the signature of the linking matrix of L.) Then*

$$\tau_r(M) = \begin{cases} \tau_3(M)\, \tau'_r(L) & \text{if } r \equiv 3 \pmod 4 \\ \overline{\tau_3(M)}\, \tau'_r(L) & \text{if } r \equiv 1 \pmod 4. \end{cases}$$

In particular $\tau_r(M) = 0$ *whenever* $\tau_3(M) = 0$.

The Corollary suggests that $\tau'_r(L)$, if invariant under the moves of the calculus of framed links (K-moves), would be a more useful invariant of $M = M_L$ than $\tau_r(M)$ because it would not vanish for "trivial" reasons. It is evident that it is invariant when $\tau_3(M) \neq 0$, since $\tau_3(M)$ and $\tau_r(M)$ are, and in fact it is invariant in general:

(8.10) **Theorem** *If L is a framed link and r is odd, then* $\tau'_r(L)$ *(defined in* (8.9)*) is invariant under K-moves on L, and hence defines an invariant* $\tau'_r(M)$ *of the associated 3-manifold* $M = M_L$.

Proof. We adopt the notation of the proof in § 5 of Theorem 1.6 (which established the invariance of $\tau_r(M)$). In particular we have a K-move $L \leftrightarrow L^\varepsilon$ of type $\varepsilon = \pm 1$, a fixed coloring $\mathbf{l}$ of L, and induced colorings $\mathbf{l} \cup k$ of L^ε for each k (the color of the new component K). Note that $K \cdot K = \varepsilon$.

Let E denote the even colored sublink of $L^\varepsilon - K$ for the coloring $\mathbf{l} \cup k$ (this is independent of k), and as above $E_\mathbf{l}$ and $E^\varepsilon_{\mathbf{l}\cup k}$ denote the even colored sublinks of L and L^ε for the colorings $\mathbf{l}$ and $\mathbf{l} \cup k$, respectively. Observe that

$$E^\varepsilon_{\mathbf{l}\cup k} = \begin{cases} E \cup K & \text{if } k \text{ is even} \\ E & \text{if } k \text{ is odd.} \end{cases} \tag{8.11}$$

It follows that

$$E^{\varepsilon}_{\mathbf{1}\cup k}\cdot E^{\varepsilon}_{\mathbf{1}\cup k}\equiv E_{\mathbf{1}}\cdot E_{\mathbf{1}}+\varepsilon\,\delta_k \pmod 4$$

where

$$\delta_k=\begin{cases}1 & \text{if } k\equiv E\cdot K \pmod 2\\ 0 & \text{if } k\not\equiv E\cdot K \pmod 2,\end{cases}$$

since $E\cdot E=E_{\mathbf{1}}\cdot E_{\mathbf{1}}+\varepsilon(E\cdot K)^2$ (by the way framings change under K-moves, see §1). Using this, the proof of the Theorem reduces (as in the proof (1.6)) to the identity

$$(1-i^{\varepsilon r})\,b\,c^{\varepsilon}\sum_{0<k<r/2} i^{\varepsilon\delta_k r}[k]\,J_{L^{\varepsilon},\mathbf{1}\cup k}=J_{L,\mathbf{1}}, \tag{8.12}$$

which is the analogue of (5.3).

To prove (8.12), consider the contribution $s_k=(1-i^{\varepsilon r})\,b\,c^{\varepsilon}\,i^{\varepsilon\delta_k r}[k]\,J_{L^{\varepsilon},\mathbf{1}\cup k}$ of each color k to left hand side. Using the Symmetry Principle (4.20) we have

$$s_k=b\,c^{\varepsilon}([k]\,J_{L^{\varepsilon},\mathbf{1}\cup k}+[r-k]\,J_{L^{\varepsilon},\mathbf{1}\cup(r-k)}). \tag{8.13}$$

Indeed, $[r-k]\,J_{L^{\varepsilon},\mathbf{1}\cup(r-k)}=i^{\varepsilon(r-2)+2(Ef_{\cup k}\cdot K)}[k]\,J_{L^{\varepsilon},\mathbf{1}\cup k}=i^{\varepsilon(r-2)+2\delta_k}[k]\,J_{L^{\varepsilon},\mathbf{1}\cup k}$ (since $E^{\varepsilon}_{\mathbf{1}\cup k}\cdot K\equiv\delta_k \pmod 2$ by (8.11)). Thus to prove (8.13), it suffices to show $(1-i^{\varepsilon r})\,i^{\varepsilon\delta_k r}=1+i^{\varepsilon(r-2)+2\delta_k}$, which is readily verified.

Now the left hand side of (8.12) can be rewritten using (8.13) as

$$b\,c^{\varepsilon}\sum_{0<k<r}[k]\,J_{L^{\varepsilon},\mathbf{1}\cup k}.$$

But this is just the left hand side of (5.3), and so (8.12) follows. □

As an application, we illustrate the use of Theorem 8.10 in studying the manifolds K_a obtained by surgery on K with integer framing a. Recall from Example 4.21 that the value of the Jones polynomial of K at the fifth root of unity is an invariant of K_a provided the framing $a\not\equiv 2 \pmod 4$. This is in fact true for all a. In particular, by Theorem 8.10

$$\tau'_5(K_a)=\sqrt{2}\,e\left(\frac{-\sigma}{8}\right)\alpha_K(1+q^{2a}[2]^2\,\overline{V_K(q)})$$

is an invariant of K_a (where $q=e(1/5)$, and the right hand side is obtained using (4.10) and Corollary 4.11). It follows that $V_K(q)$ is as well. We have proved:

(8.14) **Theorem** *Let K and K' be knots in S^3 whose Jones polynomials have distinct values at the fifth root of unity $e(1/5)$. Then the 3-manifolds K_a and K'_a (obtained by surgery with framing a) are distinct for each integer a.*

Homology spheres and the Casson invariant

For another application of Theorem 8.7, consider the 3-manifold $K_{p/q}$ obtained by p/q Dehn surgery on a knot K in S^3. Then we have:

(8.15) **Corollary** (periodicity for homology spheres for odd r) *If r is odd, then $\tau_r(K_{1/n})=\tau_r(K_{1/(n+r)})$ for every integer n. The same statement holds for the invariants τ'_r.*

Proof. $K_{1/n}$ may be obtained by surgery on a two component link L consisting of K with the zero framing together with a meridian J (unknotted) of K with framing $-n$.

Observe that the linking matrix of L has zero signature. (Note that it follows immediately that $\tau_r(K_{1/n})=\tau_r(K_{1/(n+4r)})$, by the Definition 1.5 of τ_r and the way that the invariants $J_{L,\mathbf{k}}$ change under change of framing (3.27b).) One readily computes $m_{L,\mathbf{1}}=2$ and so by Theorem 8.7

$$\tau_r(K_{1/n})=2b^2 \sum_{0<j,k<r/2} [j][k]\, i^{rE_{J\cup K}\cdot E_{J\cup K}} J_{L,j\cup k}$$

where j and k are the colors on J and K respectively, and $E_{j\cup k}$ is the associated even colored sublink of L as usual.

Evidently $E_{j\cup k}\cdot E_{j\cup k}$ is 0 if j is odd, and is $-n$ or $2-n$ if j is even (depending upon whether k is odd or even), that is

$$E_{j\cup k}\cdot E_{j\cup k}\equiv 2(j-1)(k-1)+n(j^2-1)\,(\mathrm{mod}\,4).$$

Also we have

$$J_{L,j\cup k}=\frac{[jk]}{[k]}\, t^{-n(j^2-1)}$$

by Lemma 3.27c. Thus

$$\begin{aligned}\tau_r(K_{1/n}) &= 2b^2 \sum_{0<j,k<r/2} [j][jk]\, i^{r(2(j-1)(k-1)+n(j^2-1))}\, t^{-n(j^2-1)} J_{K,k}\\ &= 2b^2 \sum_{0<j,k<r/2} [j][jk](-1)^{(j-1)(k-1)}\, q^{n(j^2-1)(r^2-1)/4} J_{K,k}\end{aligned}$$

since $i=t^r$ and $q=t^4$. (Note that $(r^2-1)/4$ is an integer since r is odd.) It is now evident that changing n to $n+r$ does not change τ_r, since $q^r=1$.

The analogous result for τ'_r follows immediately from Corollary 8.9 since $\tau_3(K_{1/n})=1$ by Corollary 6.5, whence $\tau'_r(K_{1/n})=\tau_r(K_{1/n})$. □

(8.16) *Example* Let $r=5$. Then $b^2=\frac{2}{5}\sin^2\left(\frac{\pi}{5}\right)=\frac{1}{10}(2-q-q^4)$, and a straightforward calculation from the last formula in the proof of the previous proposition yields

$$\tau_5(K_{1/n})=\tfrac{1}{5}(2-q-q^4)((1+[2]^2 q^{3n})+[2](1-q^{3n})J_K)),$$

where $J_K=J_{K,2}$ as usual.

It is illuminating to write this formula in terms of the *reduced Jones polynomial* W_K of K, defined by $V_K(x)=1-P(x)\,W_K(x)$, where $P(x)=(1-x)(1-x^3)$. (That V_K can be so written follows from the evaluations (4.12) $V_K(1)=V_K(\omega)=1$, where $\omega=e(1/3)$, and $V'_K(1)=0$. Note that $P(1)=P'(1)=P(\omega)=0$. The polynomials W_K are tabulated in Jones' original papers [J1, J2].) Now, since $J_K=[2](1-P(\bar{q})\,W(\bar{q}))$ by (4.10) and Corollary 4.11, we compute

(8.17) $$\tau_5(K_{1/n}) = 1-(1+q)(1-q^{3n})\, W_K(\bar{q}).$$

$\left(\text{Recall that } \sum_{j=0}^{4} q^j = 0.\right)$

There is an interesting consequence of this formula, relating the Casson invariant $\lambda = \lambda(K_{1/n})$ (see [AM]) with $\tau = \tau_5(K_{1/n})$. (This relationship was first observed experimentally using data generated in Mathematica [Wo].)

First recall from Casson's surgery formula that $\lambda = \frac{n}{2} \Delta_K''(1)$, where Δ_K denotes the normalized Alexander polynomial of K. But $V_K''(1) = -3\Delta_K''(1)$ by a well-known skein computation [J2, p. 369], and so

(8.18) $$\lambda = n\, W_K(1)$$

since $P(1) = P'(1) = 0$ and $P''(1) = 6$.

Next observe from (8.17) that τ is an element of the ring $\mathbf{Z}[q]$ of cyclotomic integers in the cyclotomic field $\mathbf{Q}[q]$ (where $q = e(1/5)$). Consider the map

$$T\colon\ \mathbf{Z}[q] \to \mathbf{Z}/5\mathbf{Z}$$

given by $T(\sum \lambda_j q^j) = \sum \lambda_j \pmod 5$ $(= -\operatorname{tr}(\sum \lambda_j q^j) \pmod 5)$. Observe that T is both an additive and a multiplicative homomorphism (i.e. $T(\alpha+\beta) = T(\alpha) + T(\beta)$ and $T(\alpha\beta) = T(\alpha)\, T(\beta)$). Evidently $T(\tau - 1) = 0$ (or equivalently $\operatorname{tr}(\tau - 1) \equiv 0 \pmod 5$).

Noting that $T(W_K(\bar{q})) = W_K(1) \pmod 5$, it follows from (8.17) and (8.18) that

$$\lambda \equiv n\, T\left(\frac{1-\tau}{(1+q)(1-q^{3n})}\right) \pmod 5$$

for $n \not\equiv 0 \pmod 5$. (If $n \equiv 0 \pmod 5$, then $\lambda \equiv 0 \pmod 5$.)

The last expression is in fact independent of n. To see this, write

$$\frac{1-\tau}{(1+q)(1-q^{3n})} = \frac{1-\tau}{1-q}\, u_2 u_{3n}$$

where $u_j = \dfrac{1-q}{1-q^j}$. An easy computation $\left(\text{using the fact that } \prod_{j=1}^{4}(1-q^j) = 5\right)$ shows that $T(u_j) \equiv j \pmod 5$ for $j \equiv \pm 1 \pmod 5$ and $T(u_j) \equiv -j \pmod 5$ for $j \equiv \pm 2 \pmod 5$. It follows readily that

(8.19) $$\lambda \equiv T\left(\frac{1-\tau}{1-q}\right) \pmod 5.$$

(Note that this holds even if $n \equiv 0 \pmod 5$, for then $\tau = 1$ by (8.17), and so the right hand side is 0 as expected.) Thus the mod 5 Casson invariant is determined by τ. In summary, we have

(8.20) **Theorem** *Let M be a homology sphere obtained by Dehn surgery on a knot in S^3, and let q be the fifth root of unity $e(1/5)$. Then $\tau_5(M)$ is a cyclotomic integer (i.e. an element of the ring $\mathbf{Z}[q]$). Furthermore, the element $\alpha(M)$*

$= \tau_5(M) - 1 = \tau_5(M) - \tau_5(S^3)$ *has zero trace, is divisible by* $1-q$, *and satisfies*

$$\lambda(M) \equiv \operatorname{tr}\left(\frac{\alpha(M)}{1-q}\right) (\text{mod } 5)$$

where λ *denotes the Casson invariant.*

(8.21) *Remark* The theorem holds equally well for connected sums of homology spheres, each of which is obtained by Dehn surgery on a knot in S^3.

The case of even r

From Theorem 8.5 and Lemma 8.6, we obtain the formula

$$\tau_r(M) = \alpha_L \sum_{\substack{0 < \mathbf{k} \leq r/2 \\ \phi_{\mathbf{k}} | T_{\mathbf{k}} = 0}} 2^{|T_{\mathbf{k}}|} [\mathbf{k}] J_{L,\mathbf{k}} \tag{8.23}$$

for even r, where as usual $M = M_L$ and $T_{\mathbf{k}}$ is the sublink of components whose colors are less than $r/2$ for the coloring $\mathbf{k}$.

Observe that the condition $\phi_{\mathbf{k}} | T_{\mathbf{k}} = 0$ can be replaced by the more restrictive condition $\phi_{\mathbf{k}} = 0$. For if $\phi_{\mathbf{k}} \neq 0$ on some $r/2$-colored component K, then the Symmetry Principle applied to K yields $J_{L,\mathbf{k}} = -J_{L,\mathbf{k}}$ (since $\phi_{\mathbf{k}}$ is even valued for r even), and so $J_{L,\mathbf{k}} = 0$.

Furthermore, the condition $\phi_{\mathbf{k}} = 0$ holds if and only if

$$\begin{aligned} E_{\mathbf{k}} \cdot K &\equiv K \cdot K (\text{mod } 2) \quad \text{if } r \equiv 0 (\text{mod } 4) \\ E_{\mathbf{k}} \cdot K &\equiv 0 \quad (\text{mod } 2) \quad \text{if } r \equiv 2 (\text{mod } 4) \end{aligned} \tag{8.24}$$

for each component K of L, where $E_{\mathbf{k}}$ is the even colored sublink for $\mathbf{k}$. A coloring $\mathbf{k}$ satisfying (8.24) for all K will be called a *characteristic* coloring (since for r divisible by 4 this is just the condition that $E_{\mathbf{k}}$ be a characteristic sublink of L, see Appendix C). Thus we have

(8.25) **Theorem** *If r is even, then*

$$\tau_r(M) = \alpha_L \sum_{0 < \text{characteristic} \mathbf{k} \leq r/2} 2^{|T_{\mathbf{k}}|} [\mathbf{k}] J_{L,\mathbf{k}}$$

where $T_{\mathbf{k}}$ *is the sublink of* L *consisting of components with colors strictly less than* $r/2$ *in the coloring* $\mathbf{k}$.

Homology spheres

For an application, consider once again the homology spheres $K_{1/n}$ obtained by $1/n$ Dehn surgery on a knot K in S^3 (cf. Proposition 8.15).

(8.26) **Corollary** (periodicity for homology spheres for even r) *If r is even, then* $\tau_r(K_{1/n}) = \tau_r(K_{1/(n+(r/2))})$ *for every integer n.*

Proof. The argument is analogous to the proof of Corollary 8.15, and we adopt the notation used there. Using Theorem 8.25 in place of Theorem 8.7, we obtain

$$\tau_r(K_{1/n}) = b^2 \sum_{0 < \text{characteristic } j \cup k \leq r/2} 2^{|T_{j\cup k}|}[j][jk]\, t^{-n(j^2-1)} J_{K,k}.$$

Now if $r \equiv 0 (\mathrm{mod}\, 4)$, then $j \cup k$ characteristic means that j and $k+n$ are odd. If $r \equiv 2 (\mathrm{mod}\, 4)$, then it means that j and k are odd. In either case, j is always odd and so j^2-1 is divisible by 8. Thus the term $t^{n(j^2-1)}$ (which is the only term that depends on n) can be rewritten as $(q^2)^{mn}$ for some integer m. Since q^2 has order $r/2$, the Corollary follows. □

Splittings for even r

Recall from Corollary 8.9 and Theorem 8.10 that for odd r, the invariant $\tau_r(M)$ can be written as a *product* of two other invariants of M. It turns out that for even r, it can be written as a *sum* of invariants.

For r divisible by 4, these are invariants of spin structures on M. That this should be so is suggested by the fact that the only terms which contribute to $\tau_r(M)$ come from colorings whose even colored sublinks are *characteristic*, and characteristic sublinks correspond to spin structures.

(8.27) **Theorem** *Let M be a 3-manifold and Θ be a spin structure on M. Choose a framed link L for which $M = M_L$, and let C be the characteristic sublink corresponding to Θ (see Lemma C.1). If $r \equiv 0 (\mathrm{mod}\, 4)$, then*

$$\tau_r(M, \Theta) = \alpha_L \sum_{0 < \mathbf{k} \leq r/2 \text{ with } E_{\mathbf{k}} = C} 2^{|T_{\mathbf{k}}|}[\mathbf{k}]\, J_{L,\mathbf{k}}$$

is an invariant of the spin manifold M_Θ. Furthermore,

$$\tau_r(M) = \sum_{\Theta} \tau_r(M, \Theta)$$

where the sum is over all spin structures on M.

Remark. $\tau_r(M, \Theta)$ can equally well be written as

$$\tau_r(M, \Theta) = \alpha_L \sum_{0 < \mathbf{k} < r \text{ with } E_{\mathbf{k}} = C} [\mathbf{k}]\, J_{L,\mathbf{k}} \tag{8.28}$$

using Lemma 8.6 and the fact that k and $r-k$ have the same parity (for even r).

Proof of Theorem 8.27 We must show that the right hand side of (8.28), denoted $\tau_r(L, C)$, is invariant under K-moves $(L, C) \leftrightarrow (L^\varepsilon, C^\varepsilon)$ of characteristic pairs (see Appendix C). Here

$$C^\varepsilon = \begin{cases} C + K & \text{if } C \cdot K \text{ is even} \\ C & \text{if } C \cdot K \text{ is odd} \end{cases}$$

is the characteristic sublink of L^ε (Remark C.2), where C denotes both the characteristic sublink of L and the corresponding sublink of L^ε and K is the new component of L^ε (as usual).

Proceeding as in the proof of Theorem 1.6 in §5, or Theorem 8.10 above, the proof of the invariance of $\tau_r(L, C)$ reduces to the identity

$$b\,c^\varepsilon \sum_{\substack{0<k<r \\ k\equiv C\cdot K(\mathrm{mod}\,2)}} [k]\,J_{L^\varepsilon,\mathbf{l}\cup k} = J_{L,\mathbf{l}} \tag{8.29}$$

for any coloring $\mathbf{l}$ of L with $E_{\mathbf{l}} = C$.

To prove this identity, first assume that $C\cdot K$ is even. Then for *odd* k we have $\phi_{\mathbf{l}\cup k}(K) \equiv (r-2)\,\varepsilon + 2E_{\mathbf{l}\cup k}\cdot K \equiv 2 + 2C\cdot K \equiv 2 (\mathrm{mod}\,4)$, and so by the Symmetry Principle (4.20),

$$[r-k]\,J_{L^\varepsilon,\mathbf{l}\cup(r-k)} = -[k]\,J_{L^\varepsilon,\mathbf{l}\cup k}. \tag{8.30}$$

It follows that the condition $k\equiv C\cdot K(\mathrm{mod}\,2)$ may be omitted in the sum in (8.29), as the additional terms cancel in pairs, and so (8.29) reduces to the identity (5.3).

If $C\cdot K$ is odd, then for *even* k we have $\phi_{\mathbf{l}\cup k}(K) \equiv 2 + 2(C+K)\cdot K \equiv 2 (\mathrm{mod}\,4)$, and (8.30) follows. Thus (8.29) holds in this case as well, and so the first statement in the Theorem is proved.

The last statement in the Theorem follows immediately from Theorem 8.25. □

(8.31) *Example* If $r=4$, then $\tau_4(M, \Theta) = c^{\mu(M_\Theta)}$, where $c = e(-3/16)$. Indeed, since $b = 1/2$ and $[2] = \sqrt{2}$, we compute $\tau_4(M, \Theta) = c^\sigma J_{C,2}/\sqrt{2}^{|C|}$, where σ is the signature of a framed link L with $M = M_L$ and C is the characteristic sublink corresponding to Θ. Using Corollary 4.11 and Remark 4.12, it follows that $\tau_4(M, \Theta) = c^{\sigma - C\cdot C + \mathrm{Arf}(C)} = c^{\mu(M_\Theta)}$. (Note that this yields a proof of Theorem 7.1 without cabling, cf. Remark 7.3.)

For $r \equiv 2 (\mathrm{mod}\,4)$, the invariant $\tau_r(M)$ splits as a sum of invariants, one for each element in $H^1(M; \mathbf{Z}/2\mathbf{Z})$. Indeed, the only terms which contribute to the computation of $\tau_r(M)$ come from colorings whose even colored sublinks E intersect *each* component of L evenly (where $M = M_L$). Such sublinks E are in one-to-one correspondence with elements α of $H^1(M; \mathbf{Z}/2\mathbf{Z})$. In particular, α is the unique class which is one on meridians of E and zero on meridians of $L - E$.

(8.32) **Theorem** *Let M be a 3-manifold and α be an element of $H^1(M; \mathbf{Z}/2\mathbf{Z})$. Choose a framed link L for which $M = M_L$, and let E be the sublink corresponding to α (see above). If $r \equiv 2 (\mathrm{mod}\,4)$, then*

$$\tau_r(M, \alpha) = \alpha_L \sum_{0<\mathbf{k}\le \mathbf{r}/2 \text{ with } E_{\mathbf{k}} = E} 2^{|T_{\mathbf{k}}|}\,[\mathbf{k}]\,J_{L,\mathbf{k}}$$

is an invariant of (M, α). Furthermore,

$$\tau_r(M) = \sum_\alpha \tau_r(M, \alpha)$$

where the sum is over all elements α in $H^1(M; \mathbf{Z}/2\mathbf{Z})$.

The proof is similar to the proof of Theorem 8.27 and is left to the reader. Note that one uses a calculus for pairs (L, E), with $E \cdot K$ even for all K, where the K-move replaces E by E' given by

$$E' = \begin{cases} E & \text{if } E \cdot K \text{ is even} \\ E + K & \text{if } E \cdot K \text{ is odd.} \end{cases}$$

(8.33) *Example* If $\alpha = 0$, then the corresponding sublink E is empty, and so to compute $\tau_r(M, 0)$ we only consider odd colorings of L (where $M = M_L$). It follows readily that

$$\tau_6(M, 0) = \frac{i^{-\sigma}}{\sqrt{3^n}} \sum_{S < L} J_{S,3}$$

since $b = 1/\sqrt{12}$, $c = i^{-1}$ and $[3] = 2$.

Appendix A. Identities in $\mathscr{A}$

As noted after (2.20), the relation $YX = XY - [H]$ may be generalized to $Y^n X = XY^n - [n][H+n-1]Y^{n-1}$. A more general formula is

$$\text{(A.1)} \quad Y^n X^k = \sum_{0 \leq i \leq \min(n,k)} (-1)^i \begin{bmatrix} n \\ i \end{bmatrix} [i]! \begin{bmatrix} k \\ i \end{bmatrix} [i]! \begin{bmatrix} H+n-k-1+i \\ i \end{bmatrix} X^{k-i} Y^{n-i}$$

$\left(\text{where } \begin{bmatrix} H+m \\ j \end{bmatrix} = [H+m] \ldots [H+m-j+1]/[j]!\right)$, which follows by induction on k using the identity $[a][H+c+b] + [b][H+c-a] = [a+b][H+c]$.

In particular, we will need (A.1) when $k = n$:

$$\text{(A.2)} \quad Y^n X^n = \sum_{0 \leq i \leq n} (-1)^i \begin{bmatrix} n \\ i \end{bmatrix}^2 [i]!^2 \begin{bmatrix} H-1+i \\ i \end{bmatrix} X^{n-i} Y^{n-i}.$$

For our purposes, it is convenient to expand the term

$$\text{(A.3)} \quad \begin{bmatrix} H-1+i \\ i \end{bmatrix} = \frac{1}{[i]!(s-\bar{s})^i} (K^2 - \bar{K}^2)(sK^2 - \bar{s}K^2) \ldots (s^{i-1}K^2 - \bar{s}^{i-1}\bar{K}^2)$$

in powers of K:

(A.4) **Lemma**

$$(K^2 - \bar{K}^2)(sK^2 - \bar{s}\bar{K}^2) \ldots (s^{i-1}K^2 - \bar{s}^{i-1}\bar{K}^2) = \sum_{0 \leq j \leq i} (-1)^{i-j} c_{ij} K^{4j-2i}$$

where $c_{ij} = t^{(2j-i)(i-1)} \begin{bmatrix} i \\ j \end{bmatrix}$. $\left(\textit{The exponent of } K \textit{ arises by choosing } j \textit{ } K^2\textit{'s and } (i-j) \textit{ } \bar{K}^2\textit{'s.}\right)$

Proof. We use double induction starting with the cases $j=i$ (where $c_{ii}=s^{0+\ldots+(i-1)}=t^{i(i-1)}$ is the coefficient of K^{2i}) and $j=0$ (where $c_{i0}=\bar{s}^{0+\ldots+(i-1)}=\bar{t}^{i(i-1)}$ is $(-1)^i$ times the coefficient of $\bar{K}^{2i}$).

Recursively, $c_{ij}=s^{i-1}c_{i-1,j-1}-\bar{s}^{i-1}c_{i-1,j}$ for $j<i$, so c_{ij} is determined by c_{kk} and c_{k0} for $k\leqq j$. But $t^{(2j-i)(i-1)}\begin{bmatrix} i \\ j \end{bmatrix}$ satisfies the same recursive formula, using the quantized Pascal relation

$$s^j\begin{bmatrix} i \\ j \end{bmatrix}=s^i\begin{bmatrix} i-1 \\ j-1 \end{bmatrix}+\begin{bmatrix} i-1 \\ j \end{bmatrix}. \quad \square \tag{A.5}$$

Using (A.3–4) we may rewrite (A.2) as

$$Y^nX^n=\sum_{0\leqq j\leqq i\leqq n}(-1)^j\begin{bmatrix} n \\ i \end{bmatrix}^2[i]!\,\frac{t^{(2j-i)(i-1)}}{(s-\bar{s})^i}\begin{bmatrix} i \\ j \end{bmatrix}X^{n-i}Y^{n-i}K^{4j-2i}. \tag{A.6}$$

We are now in a position to complete the proof of Theorem 3.20.

(A.7) **Theorem** *The identity*

$$\sum\alpha_i\bar{K}^2\beta_i=\sum\beta_iK^2\alpha_i$$

is satisfied in $\mathscr{A}$, *where* $R=\sum\alpha_i\otimes\beta_i$ *is the R-matrix given in Theorem* 2.18.

Proof. Using Theorem 2.18, the left hand side is

$$\sum c_{nab}X^nK^a\bar{K}^2Y^nK^b=\sum c_{nab}\bar{s}^{n(a-2)}X^nY^nK^{a+b-2}$$

and the right hand side is

$$\sum c_{nab}Y^nK^bK^2X^nK^a=\sum c_{nab}s^{n(b+2)}Y^nX^nK^{a+b+2},$$

where the sums are over all $0\leqq n<r$ and $0\leqq a,\ b<4r$. Multiplying by $4r\bar{K}^2$ and substituting for c_{nab}, the left hand side becomes

$$\sum\frac{(s-\bar{s})^n}{[n]!}\bar{t}^{ab+(b+a)n-3n}X^nY^nK^{a+b-4}=\sum\lambda_{np}X^nY^nK^p \tag{A.8}$$

and the right hand side becomes

$$\sum\frac{(s-\bar{s})^n}{[n]!}\bar{t}^{ab-(b+a)n-3n}X^nY^nK^{a+b}=\sum\rho_{np}X^nY^nK^p \tag{A.9}$$

summed over all $0\leqq n<r$ and $0\leqq p<4r$. It remains to show that coefficients λ_{np} and ρ_{np} of $X^nY^nK^p$ are equal for all n and p.

From (A.8) we compute

$$\lambda_{np} = \sum_{a+b-4 \equiv p (\mathrm{mod}\, 4r)} \frac{(s-\bar{s})^n}{[n]!}\, t^{\bar{a}b+(b+a)n-3n}$$

$$= \sum_{0 \leqq a < 4r} \frac{(s-\bar{s})^n}{[n]!}\, t^{a^2-(p+4)a-(p+1)n}$$

$$= \sum_{0 \leqq a < 4r} \frac{(s-\bar{s})^n}{[n]!}\, t^{a^2-pa-2p-4-(p+1)n}$$

where the last equality is derived by replacing a by $a+2$, and noting that the sums $\sum_{0 \leqq a < 4r}$ and $\sum_{-2 \leqq a < 4r-2}$ are equal.

To compute ρ_{np}, first use (A.6) to move Y^n past X^n on the left side of (A.9)

$$\sum_{\substack{0 \leqq n < r \\ 0 \leqq a,b < 4r \\ 0 \leqq j \leqq i \leqq n}} \frac{(s-\bar{s})^n}{[n]!}\, t^{\bar{a}b-(b+a)n-3n}(-1)^j \begin{bmatrix} n \\ i \end{bmatrix}^2 [i]!\, \frac{t^{(2j-i)(i-1)}}{(s-\bar{s})^i} \begin{bmatrix} i \\ j \end{bmatrix} X^{n-i} Y^{n-i} K^{a+b+4j-2i}.$$

It follows as above that ρ_{np} is

$$\sum_{\substack{0 \leqq a < 4r \\ 0 \leqq j \leqq i < r-n}} \frac{(s-\bar{s})^{n+i}}{[n+i]!}\, t^{a^2-a(p-4j+2i)+(n+i)(p-4j+2i+3)+(2j-i)(i-1)}$$
$$(-1)^j \begin{bmatrix} n+i \\ i \end{bmatrix}^2 \frac{[i]!}{(s-\bar{s})^i} \begin{bmatrix} i \\ j \end{bmatrix}$$

$$= \sum_{\substack{0 \leqq a < 4r \\ 0 \leqq j \leqq i < r-n}} \frac{(s-\bar{s})^n}{[n]!}\, t^{a^2-pa+(pn+3n+2jp-4j^2-4jn-2j)+i(2j+2n+4)} \begin{bmatrix} n+i \\ i \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix}$$

where the second equality follows by cancellation and substitution of $a-2j+i$ for a.

Since n and p are fixed, λ_{np} and ρ_{np} have a common factor

$$C = \frac{(s-\bar{s})^n}{[n]!} \sum_{0 \leqq a < 4r} t^{a^2-pa}.$$

If p is odd, then $C=0$. Indeed, $\frac{p-1}{2}$ is an integer in this case, and so we can replace a by $a+\frac{p-1}{2}$ in the sum:

$$\sum_{0 \leqq a < 4r} t^{a^2-pa} = t^{\frac{1-p^2}{4}} \sum_{0 \leqq a < 4r} t^{a^2-a}.$$

But $t^{(2r-a+1)^2-(2r-a+1)} = t^{-2r+a^2-a} = -t^{a^2-a}$, since $t^{4r}=1$, and so the terms in the sum on the right cancel in pairs. Thus $\lambda_{np}=\rho_{np}=0$.

For p even, we divide λ_{np} and ρ_{np} by C, and it remains to show that

$$\text{(A.10)} \qquad t^{-2p-4-(p+1)n} = \sum_{0 \leqq j \leqq i < r-n} (-1)^j t^{m+i(2j+2n+4)} \begin{bmatrix} n+i \\ i \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix}$$

where $m = pn+3n+2jp-4j^2-4jn-2j$ is independent of i.

It is straightforward to check that the term on the right hand side of (A.10) corresponding to $i=j=r-n-1$ is equal to the left hand side. (Note that p is even, $(-1)^i = t^{2r(r-n-1)}$, and $\begin{bmatrix} n+i \\ i \end{bmatrix} = \begin{bmatrix} i \\ j \end{bmatrix} = 1$ in this case.)

We now finish by showing that for fixed $j < r-n-1$, the sum over i on the right of (A.10) is zero. Factoring out $(-1)^j t^m$, it suffices to show

$$\text{(A.11)} \qquad \sum_{j \leqq i < r-n} t^{i(2j+2n+4)} \begin{bmatrix} n+i \\ i \end{bmatrix} \begin{bmatrix} i \\ j \end{bmatrix} = 0.$$

It is convenient at this point to use the binomial q-coefficients $\begin{bmatrix} n \\ k \end{bmatrix}_q = s^{k(n-k)} \begin{bmatrix} n \\ k \end{bmatrix}$ (see (2.29)). The left hand side of (A.11) then becomes

$$\sum_{j \leqq i < r-n} t^{2j^2+4i} \begin{bmatrix} n+i \\ i \end{bmatrix}_q \begin{bmatrix} i \\ j \end{bmatrix}_q = \frac{t^{2j^2}}{[n]_q! [j]_q!} \sum_{j \leqq i < r-n} q^i \frac{[i+n]_q!}{[i-j]_q!}.$$

The last sum is just $\sum_{j \leqq i < r-n} q^i [i+n]_q \dots [i-j+1]_q$. To see that this is zero, note that the product of brackets ranges from a high of $[r-1]_q$ when $i=r-n-1$ to a low of $[1]$ when $i=j$. Thus we may extend the range of i to $0 \leqq i < r$, as that only adds terms containing a factor of either $[r]_q = 0$ or $[0]_q = 0$. Hence the left hand side of (A.11) is a multiple of

$$\begin{aligned}
&\sum_{i=0}^{r-1} q^i [i+n]_q \dots [i-j+1]_q \\
&= \frac{1}{(q-1)^{n+j}} \sum_i q^i (q^{i+n}-1) \dots (q^{i-j+1}-1) \\
&= \frac{1}{(q-1)^{n+j}} \sum_i q^i \left(\sum_{k=1}^{2^{n+j}} \pm q^{a_k i + b_k} \right) \quad \text{(for suitable } a_k, b_k) \\
&= \frac{1}{(q-1)^{n+j}} \sum_k \pm \left(\sum_i q^{(a_k+1)i+b_k} \right) \\
&= \frac{1}{(q-1)^{n+j}} \sum_k \pm \left(\sum_{i=0}^{r-1} q^i \right) \\
&= 0
\end{aligned}$$

since the sum of the r^{th}-roots of unity is zero. □

Appendix B. The Jones polynomial at the sixth root of unity

In this appendix we derive an expression for the value of the variant $\tilde{V}_L$ of the Jones polynomial at $q=e(\frac{1}{6})$ (see Remark 4.12). There are of course similarities between our derivation and that of Lipson [Lp] for the Jones polynomial V_L.

Let A be a symmetric matrix over $\mathbf{Z}/3\mathbf{Z}$. The corresponding quadratic form is classified up to Witt equivalence by its *nullity* d_A and *Witt class* ω_A in the Witt group $W(\mathbf{Z}/3\mathbf{Z})=\mathbf{Z}/4\mathbf{Z}$ (see e.g. [MH]). These invariants may be computed as follows: diagonalize A (over $\mathbf{Z}/3\mathbf{Z}$) writing all entries as 0 or ± 1. Then d_A is the number of diagonal 0's, and ω_A is the trace (or signature) viewed as an integer mod 4.

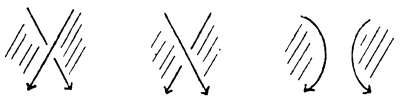
Set

$$\lambda_A=\sqrt{3}^{d_A}(-i)^{\omega_A}.$$

Note that λ multiplies under block sum. If A is the mod 3 reduction of a matrix representing the quadratic form of a link L, write $\lambda_L=\lambda_A$.

(B.1) **Theorem** $\tilde{V}_L=\lambda_L$ *at* $q=e(\frac{1}{6})$.

Proof. Let L_ε ($\varepsilon=+,-,0$) be as in Fig. 4.4a. Choose corresponding connected Seifert surfaces F_ε which locally appear as in Fig. B.2 and coincide otherwise.



F_+ F_- F_0

Fig. B.2

With respect to suitable bases of $H_1(F_\varepsilon)$, the associated symmetrized Seifert matrices A_ε satisfy

$$A_+=\begin{pmatrix} a & * \\ * & A_0 \end{pmatrix} \quad \text{and} \quad A_-=\begin{pmatrix} a+2 & * \\ * & A_0 \end{pmatrix}.$$

By a change of basis (first diagonalizing A_0 mod 3) we may arrange that either

$$A_+\equiv(a)\oplus B, \qquad A_-\equiv(a-1)\oplus B, \qquad A_0\equiv B \,(\text{mod } 3)$$

(for A_0 nonsingular), or

$$A_+\equiv\begin{pmatrix} a & b \\ b & 0 \end{pmatrix}\oplus B, \qquad A_-\equiv\begin{pmatrix} a-1 & b \\ b & 0 \end{pmatrix}\oplus B, \qquad A_0\equiv(0)\oplus B \,(\text{mod } 3)$$

for some matrix B. Here $\oplus$ denotes block sum.

Now set

$$d_\varepsilon=d_{A_\varepsilon}-d_B$$

$$\omega_\varepsilon=\omega_{A_\varepsilon}-\omega_B$$

$$\lambda_\varepsilon=\lambda_{A_\varepsilon}/\lambda_B=\sqrt{3}^{d_\varepsilon}(-i)^{\omega_\varepsilon}.$$

If A_0 is nonsingular, or A_0 is singular with $b \neq 0$, we have

		d_+	d_-	d_0	ω_+	ω_-	ω_0	λ_+	λ_-	λ_0
$a \equiv 0$		1	0	0	0	-1	0	$\sqrt{3}$	i	1
$a \equiv 1$	(mod 3)	0	1	0	1	0	0	$-i$	$\sqrt{3}$	1
$a \equiv 2$		0	0	0	-1	1	0	i	$-i$	1
$b \neq 0$		0	0	1	0	0	0	1	1	$\sqrt{3}$

For A_0 singular with $b=0$, the d_ε are 1 more than the corresponding d_ε in the nonsingular case, and the ω_ε remain the same, and so the λ_ε go up by a factor of $\sqrt{3}$. Now for $q = e\left(\frac{1}{6}\right) = \frac{1}{2} + \frac{\sqrt{3}}{2} i$, $s - \bar{s} = e\left(\frac{1}{12}\right) - e\left(-\frac{1}{12}\right) = i$, and one readily verifies that

$$q\lambda_+ - \bar{q}\lambda_- = (s - \bar{s})\lambda_0$$

in all cases. Multiplying by λ_B gives

$$q\lambda_{L_+} - \bar{q}\lambda_{L_-} = (s - \bar{s})\lambda_{L_0}$$

since $\lambda_{L_\varepsilon} = \lambda_{A_\varepsilon}$ by definition. It is evident that $\lambda_{\text{unknot}} = 1$, and so $\tilde{V}_L = \lambda_L$ as desired. □

Appendix C. μ-invariants

Let M be an oriented, closed connected 3-manifold. A *spin structure* Θ on M can be viewed as a homotopy class of trivializations of the tangent bundle τ_M over M-point [K2]. It is well known that M has a spin structure, since $w_2(\tau_M) = 0$. It follows by obstruction theory that the number of distinct spin structures is equal to the number of elements in $H^1(M; \mathbf{Z}/2\mathbf{Z}) = H_1(M; \mathbf{Z}/2\mathbf{Z})$.

Recall that the *μ-invariant* of M_Θ is defined to be the signature (mod 16) of any smooth, compact spin 4-manifold W with spin boundary M_Θ,

$$\mu(M_\Theta) = \sigma(W) \,(\text{mod } 16).$$

This is well defined by Rohlin's theorem, which states that the signature of a smooth, closed spin 4-manifold is divisible by 16.

Now suppose that M is described by a framed link L, so $M = M_L = \partial W_L$ (see § 1). A sublink C of L is *characteristic* if $C \cdot L_i \equiv L_i \cdot L_i (\text{mod } 2)$ for all components L_i of L, and the pair (L, C) is then called a *characteristic pair.*

(C.1) **Lemma** *There is a natural one-to-one correspondence between the spin structures on M_L and characteristic sublinks of L.*

Proof. Assign to any spin structure Θ on M_L the sublink C of L consisting of all components L_i such that Θ does not extend across the 2-handle in W_L attached to L_i. An elementary geometric argument shows that C must be characteristic (see e.g. § 3 of [MK]).

The map carrying Θ to C is one-to-one. Indeed, if C is assigned to some other spin structure Θ', then Θ and Θ' agree on the link L in M_L which is

the core of the surgery on L. Since $H=H_1(M_L; \mathbf{Z}/2\mathbf{Z})$ is carried by L', we have $\Theta=\Theta'$ by obstruction theory.

It remains to show that there are exactly $|H|$ characteristic sublinks (as this is the number of spin structures). To see this, recall that the mod 2 linking matrix A of L is a presentation matrix for H, and so $|H|=|\ker A|$. But C (viewed as a column vector of 0's and 1's) is characteristic if and only if $AC=D$, where D is the main diagonal of A, and so the number of characteristic sublinks is $|\ker A|$ as well. □

If θ is a spin structure on M_L and C is the corresponding characteristic sublink of L, then the spin manifold $(M_L)_\theta$ will be denoted by $M_{L,C}$.

(C.2) *Remark* The argument in [K1] on the calculus of framed links for oriented 3-manifolds yields a calculus of characteristic pairs for spin 3-manifolds. In particular, $M_{L,C}=M_{L',C'}$ as spin manifolds (i.e. there is a diffeomorphism between them which preserves spin structures) if and only if one can pass from (L, C) to (L', C') by isotopy in S^3 and a combination of the following two *moves of characteristic pairs* (cf. §1):

Move 1 (blow up) *Add (or delete) a disjoint unknotted component with framing* ± 1 *and replace* C *by* $C'=C+K$.

Move 2 (handle slide) *For some* $i\neq j$, *slide* L_i *over* L_j *to get* $L_i'=L_i+L_j$ *and replace* C *by*

$$C'=\begin{cases} C & \text{if } C \text{ does not contain } L_i \\ C-(L_i+L_j)+L_i' & \text{if } C \text{ contains } L_i \text{ and } L_j \\ C-L_i+(L_j+L_i') & \text{if } C \text{ contains } L_i \text{ but not } L_j \end{cases}$$

(cf. Remark 6.2). As in [FR], these moves may be combined into one, the *K-move*, defined as in §1 (Fig. 1.3) with

$$C'=\begin{cases} C+K & \text{if } C\cdot K \text{ is even} \\ C & \text{if } C\cdot K \text{ is odd.} \end{cases}$$

Here, C denotes both the characteristic sublink of L and the corresponding sublink of L′. Note that it can happen that $M_{L,C}=M_{L,C'}$ for some $C\neq C'$.

Now let (L, C) be a characteristic pair. Note that C is a *proper* link (i.e. characteristic as a sublink of itself, or equivalently $L_i\cdot(C-L_i)$ is even for all components L_i of C), and therefore has a well defined *Arf invariant* $\mathrm{Arf}(C)$. If L is oriented, then define the *μ-invariant* of (L, C) to be

$$\mu(L, C)=\sigma-C\cdot C+8\,\mathrm{Arf}(C)\ (\mathrm{mod}\ 16) \tag{C.3}$$

where $\sigma=\sigma(W_L)$ is the signature of the linking matrix of L.

(C.4) **Theorem** *$\mu(L, C)$ is an invariant of the spin manifold $M_{L,C}$.*

Proof. By the previous remark, it suffices to show that $\mu(L, C)$ is independent of the orientation on L and is invariant under moves 1 and 2 of characteristic pairs.

First suppose that the orientation on a component K of L is reversed, giving (L', C'). Evidently σ remains unchanged. If C does not contain K, then $C \cdot C$ and $\mathrm{Arf}(C)$ are unchanged as well, and so μ is unchanged. If C contains K, then homologically $C' = C - 2K$, and so $C' \cdot C' = C \cdot C - 4K \cdot (C - K)$. Thus the invariance of μ follows from the fact that $\mathrm{Arf}(C') \equiv \mathrm{Arf}(C) + \frac{1}{2} K \cdot (C - K) (\mathrm{mod}\ 2)$, or equivalently

$$(-1)^{\mathrm{Arf}(C') - \mathrm{Arf}(C)} = (-1)^{\frac{1}{2} K \cdot (C - K)}.$$

(Note that $K \cdot (C - K)$ is even since C is characteristic.) This fact can be proved by elementary methods using Seifert surfaces, but it is quicker to use the formula for the Jones polynomial at i (Remark 4.12), which gives $(-1)^{\mathrm{Arf}(C') - \mathrm{Arf}(C)} = V_{C'}(i)(V_C(i))^{-1}$, and the Jones reversing result (see Proposition 4.3 of [L2]), which gives $V_{C'}(i) = (-1)^{\frac{1}{2} K \cdot (C - K)} V_C(i)$. (Note: Jones' reversing result can be proved using Remark 3.26 and Corollary 4.11.)

If L is changed by Move 1, then σ and $C \cdot C$ change equally (by ± 1) and $\mathrm{Arf}(C)$ remains unchanged, and so μ is unchanged.

If L is changed by Move 2, sliding L_i over L_j to get $L_i' = L_i + L_j$, then σ is unchanged, and $C \cdot C$ and $\mathrm{Arf}(C)$ change only if C contains L_i. In that case, since μ is independent of orientation, we can change the orientation of L_j after the handle slide ($L_j' = -L_j$). Then homologically $C' = C$, and so $C' \cdot C' = C \cdot C$. Also $\mathrm{Arf}(C') = \mathrm{Arf}(C)$, since the Arf invariant does not change under orientation preserving band connected sum, and evidently we can get C from C' by summing L_i and L_j' appropriately. Thus μ is unchanged. □

(C.5) **Corollary** $\mu(M_{L,C}) = \mu(L, C)$.

Proof. It is known that (L, C) may be changed by moves of characteristic pairs to (L', C') with $C' = \emptyset$, and so $W_{L'}$ is a spin 4-manifold bounded by $M_{L,C}$ [Ka]. Thus $\mu(M_{L,C}) = \sigma(W_{L'}) (\mathrm{mod}\ 16) = \mu(L', C') = \mu(L, C)$. □

Remark. Normally Rohlin's theorem is used to show that the μ-invariant is a $\mathbf{Z}/16\mathbf{Z}$ invariant (rather than a $\mathbf{Z}/8\mathbf{Z}$ invariant as the algebra of intersection forms predicts). One may, however, reverse the order of things by showing the invariance of μ from the calculus of framed links (Lemma C.4), and then deduce Rohlin's theorem:

(C.6) **Corollary** (Rohlin's Theorem) *If W is a smooth, closed spin 4-manifold, then $\sigma(W) \equiv 0 (\mathrm{mod}\ 16)$.*

Proof. Using smoothness, decompose W as a handlebody with one 0 and one 4-handle. We may assume that W has no 1 or 3-handles, since they may be changed into 2-handles by surgery (preserving the spin structure and signature σ). Now $W_0 = W - (\text{4-handle}) = W_L$, with $M_L = S^3$, for some framed link L. Since W_L is spin, the characteristic sublink corresponding to the (unique) spin structure on S^3 is empty, and so $\mu(L, \emptyset) = \sigma(W_L) (\mathrm{mod}\ 16)$. But $S^3 = M_\emptyset$ as well, and so $\mu(L, \emptyset) = \mu(\emptyset, \emptyset) = 0$ by (C.4). Thus $\sigma(W) = \sigma(W_L) \equiv 0 (\mathrm{mod}\ 16)$. □

Whether this proof is really easier depends on ones view of the calculus of framed links whose proof ([K1]) uses Cerf theory.

References

[A] Atiyah, M.: On framings of 3-manifolds. Topology **29**, 1–7 (1990)

[AHLS] Atiyah, M.F., Hitchin, N., Lawrence, R., Segal, G.: Oxford seminar on Jones-Witten theory. Michaelmas term 1988

[Br] Brown, E.H.: Generalization of the Kervaire invariant. Ann. Math. **95**, 368–383 (1972)

[D1] Drinfel'd, V.G.: Hopf algebras and the quantum Yang-Baxter equation. Dokl. Akad. Nauk SSSR **283**, 1060–1064 (1985)

[D2] Drinfel'd, V.G.: Quantum groups. Proc. Int. Cong. Math. Berkeley, Calif. 1986, pp. 798–820. Providence, RI: Am. Math. Soc. 1987

[Du] Durfee, A.: Bilinear and quadratic forms on torsion modules. Adv. Math. **25**, 133–164 (1977)

[FR] Fenn, R., Rourke, C.: On Kirby's calculus of links. Topology **18**, 1–15 (1979)

[FG] Freed, D.S., Gompf, R.E.: Computer calculation of Witten's 3-manifold invariant. (Preprint)

[FY] Freyd, P., Yetter, D.: Braided, compact, closed categories with applications to low dimensional topology. Adv. Math. **77**, 156–182 (1989)

[GJ] Goldschmidt, D., Jones, V.: Metaplectic links invariants. (to appear)

[Ji] Jimbo, M.: A q-difference analogue of $U(\mathfrak{G})$ and the Yang-Baxter equation. Lett. Math. Phys. **10**, 63–69 (1985)

[J1] Jones, V.F.R.: A polynomial invariant for knots via von Neumann algebras. Bull. Am. Math. Soc. **12**, 103–111 (1985)

[J2] Jones, V.F.R.: Hecke algebra representations of braid groups and link polynomials. Ann. Math. **126**, 335–388 (1987)

[Ka] Kaplan, S.: Constructing 4-manifolds with given almost framed boundaries. Trans. Am. Math. Soc. **254**, 237–263 (1979)

[Kf] Kauffman, L.H.: On Knots. (Ann. Math. Stud., vol. 115) Princeton: Princeton University Press 1987

[K1] Kirby, R.C.: A calculus for framed links in S^3. Invent. Math. **45**, 35–56 (1978)

[K2] Kirby, R.C.: The Topology of 4-Manifolds. (Lect. Notes Math., vol. 1374) Berlin Heidelberg New York: Springer 1989

[KM1] Kirby, R., Melvin, P.: Evaluations of new 3-manifold invariants. Notices Am. Math. Soc. **10**, Abstract 89T-57-254, 491 (1989)

[KM2] Kirby, R., Melvin, P.: Evaluations of the 3-manifold invariants of Witten and Reshetikhin-Turaev for sl(2, **C**). In: Geometry of Low-dimensional Manifolds. Durham 1989. (Lond. Math. Soc. Lect. Note Ser., vol. 151) Cambridge: Cambridge University Press 1990

[KM3] Kirby, R., Melvin, P.: Dedekind sums, signature defects and the signature cocycle. (Preprint)

[KM4] Kirby, R., Melvin, P.: Quantum invariants of lens spaces and a Dehn surgery formula. (in preparation)

[KT] Kirby, R.C., Taylor, L.R.: Pin structures on low dimensional manifolds. In: Geometry of Low-dimensional Manifolds. Durham 1989. (Lond. Math. Soc. Lect. Note Ser., vol. 151) Cambridge: Cambridge University Press 1990

[KiR] Kirillov, A.N., Reshetikhin, N.Yu.: Representations of the algebra $U_q(\mathrm{sl}_2)$, q-orthogonal polynomials and invariants of links. Prepr. LOMI, E **9** (1988)

[KP] Kneser, M., Puppe, D.: Quadratische Formen und Verschlingungsinvarianten von Knoten. Math. Z. **58**, 376–384 (1953)

[KoS] Ko, K.H., Smolinsky, L.: A combinatorial matrix in 3-manifold theory. Pac. J. Math. (to appear)

[KuR] Kulish, P.P., Reshetikhin, N.Yu.: J. Sov. Math. **23**, 2435 (1983)

[La] Lang, S.: Algebraic Number Theory. Berlin Heidelberg New York: Springer 1986

[L1] Lickorish, W.B.R.: A representation of orientable, combinatorial 3-manifolds. Ann. Math. **76**, 531–540 (1962)

[L2] Lickorish, W.B.R.: Polynomials for links. Bull. Lond. Math. Soc. **20**, 558–588 (1988)

[L3] Lickorish, W.B.R.: Invariants for 3-manifolds derived from the combinatorics of the Jones polynomial. Pac. J. Math. (to appear)

[L4] Lickorish, W.B.R.: 3-manifolds and the Temperley-Lieb algebra. (to appear)
[L5] Lickorish, W.B.R.: Calculations with the Temperley-Lieb algebra. (Preprint)
[LM1] Lickorish, W.B.R., Millett, K.C.: Some evaluations of link polynomials. Comment. Math. Helv. **61**, 349–359 (1986)
[Lp] Lipson, A.S.: An evaluation of a link polynomial. Math. Proc. Camb. Phil. Soc. **100**, 361–364 (1986)
[Lu] Lustig, G.: Quantum deformations of certain simple modules over enveloping algebras. Adv. Math. **70**, 237–249 (1988)
[Ma] Matsumoto, Y.: An elementary proof of Rochlin's signature theorem and its extension by Guillou and Marin. In: Guillou, L., Marin, A. (eds.) A la Recherche de la Topologie Perdue (pp. 119–140). Boston: Birkhauser 1986
[MK] Melvin, P., Kazez, W.: 3-dimensional bordism. Mich. Math. J. **36**, 251–260 (1989)
[MH] Milnor, J., Husemoller, D.: Symmetric Bilinear Forms. Berlin Heidelberg New York: Springer 1973
[MS] Morton, H.R., Strickland, P.: Jones polynomial invariants for knots and satellites. (Preprint)
[Mu] Murakami, H.: A recursive calculation of the Arf invariant of a link. J. Math. Soc. Japan **38**, 335–338 (1986)
[RT1] Reshetikhin, N.Yu., Turaev, V.G.: Ribbon graphs and their invariants derived from quantum groups. Commun. Math. Phys. **127**, 1–26 (1990)
[RT2] Reshetikhin, N.Yu., Turaev, V.G.: Invariants of 3-manifolds via link polynomials and quantum groups. Invent. Math. **103**, 547–597 (1991)
[Sk] Sklyanin, E.K.: On an algebra generated by quadratic relations. Usp. Mat. Nauk **40**, 214 (1985)
[Tr] Trace, B.: On the Riedemeister moves of a classical knot. Proc. Am. Math. Soc. **89**, 722–724 (1983)
[T1] Turaev, V.G.: Spin structures on three-dimensional manifolds. Math. USSR, Sb. **48**, 65–79 (1984)
[T2] Turaev, V.G.: The Conway and Kauffman modules of the solid torus with an appendix on the operator invariants of tangles. Prepr. LOMI (1988)
[T3] Turaev, V.G.: State sum models in low-dimensional topology. (Preprint 1990)
[Wa] Wallace, A.H.: Modifications and cobounding manifolds. Can. J. Math. **12**, 503–528 (1960)
[Wk] Wilkens, D.L.: Closed $(s-1)$-connected $(2s+1)$-manifolds, $s=3, 7$. Bull. Lond. Math. Soc. **4**, 27–31 (1972)
[Wt] Witten, E.: Quantum field theory and the Jones polynomial. Commun. Math. Phys. **121**, 351–399 (1989)
[Wo] Wolfram, S.: Mathematica: A System for Doing Mathematics by Computer. Reading, Mass. Addison-Wesley 1988
[Ye] Yetter, D.N.: Markov algebras. In: Braids. (Contemp. Math., vol. 78) Providence, RI: Am. Math. Soc. 1988

Invent. math. 105, 547–569 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Interior estimates for hypersurfaces moving by mean curvature

Klaus Ecker[1] **and Gerhard Huisken**[2]

[1] Department of Mathematics, The University of Melbourne, Parkville VIC 3052, Australia
[2] Centre for Mathematical Analysis, Australian National University, GPO Box 4, Canberra ACT 2601, Australia

Oblatum 19-X-1990

Let $\mathbf{F}_t: M^n \to R^{n+1}$ be a hypersurface moving by its mean curvature in Euclidean space. That is, $\mathbf{F}_t = \mathbf{F}(\cdot, t)$ is a one-parameter family of smooth immersions, with images $M_t = \mathbf{F}_t(M^n)$, satisfying the evolution equation

$$\frac{d}{dt}\mathbf{F}(p,t) = \mathbf{H}(p,t), \qquad p \in M^n,\ t > 0. \tag{1}$$

Here $\mathbf{H}(p,t) = -H(p,t)\cdot\nu(p,t)$ is the mean curvature vector of M_t at $\mathbf{F}(p,t)$ and ν denotes a choice of unit normal for M_t.

Mean curvature evolution of smooth hypersurfaces was studied previously under various global assumptions: It was shown in [8, 6] that compact convex surfaces in R^{n+1}, $n \geqq 2$ and embedded curves in the plane contract smoothly to a point. In [4] the authors characterized the longterm behaviour of entire graphs of controlled growth.

It is the aim of this paper to study the local properties of mean curvature flow and obtain regularity estimates which are *interior* both in space and time. Working in a region sufficiently small such that M_t can be written as the graph of a function w over some hyperplane, we use the gradient function $v = \sqrt{1+|Dw|^2}$ to obtain interior estimates for the second fundamental form A and all its derivatives, compare Theorem 3.4. Under additional curvature assumptions on M_t we furthermore obtain improved estimates with respect to time dependence for the derivatives of A without having to assume that M_t can be locally written as a graph. It is worth noting that all the interior estimates also apply to the case of minimal hypersurfaces as these are stationary solutions of (1), compare Remark 3.6.

An immediate consequence of the interior estimates is a shorttime existence result for the mean curvature flow of complete surfaces, assuming only a uniform local Lipschitz condition for the initial surface. In Sect. 4 we extend a method of Liao and Tam [16] to prove a general maximum principle for parabolic equations on such complete Riemannian manifolds with time-dependent metric. As an application of this maximum principle we obtain that nonnegative mean curvature is preserved by (1) under reasonably weak conditions.

Finally, in Sect. 5 we use the interior estimates to deduce the somewhat surprising result that if M_0 is a locally Lipschitz continuous entire graph over R^n, then the mean curvature flow Eq. (1) has a smooth solution with initial data M_0 which exists for all times $t>0$. In particular, it is not necessary to make any assumption on the growth and the curvature of M_0 for $|x|\to\infty$.

1 Preliminaries

We will use the same notation as in [10] and [4], in particular we will often write $\mathbf{x}=\mathbf{F}(p,t)$ to denote the position of a point in M_t and write $A=\{h_{ij}\}$ for the second fundamental form of M_t.

Furthermore, if M_t is locally given as a graph over some hyperplane in R^{n+1} such that $\langle \nu,\omega\rangle>0$ for some fixed vector $\omega\in R^{n+1}$, $|\omega|=1$, we will consider the gradient function v defined by

$$v=\langle \nu,\omega\rangle^{-1}.$$

With this notation we recall the following evolution equations from [10] and [4].

1.1 Lemma. *If M_t satisfies (1) we have*

(i) $\left(\frac{d}{dt}-\Delta\right)|\mathbf{x}|^2=-2n,$

(ii) $\left(\frac{d}{dt}-\Delta\right)u=0,$ *where* $u=\langle \mathbf{x},\omega\rangle,$

(iii) $\left(\frac{d}{dt}-\Delta\right)v=-|A|^2v-2v^{-1}|\nabla v|^2,$

(iv) $\left(\frac{d}{dt}-\Delta\right)H=H|A|^2,$

(v) $\left(\frac{d}{dt}-\Delta\right)|A|^2=-2|\nabla A|^2+2|A|^4,$

(vi) $\left(\frac{d}{dt}-\Delta\right)|\nabla^m A|^2\leqq -2|\nabla^{m+1}A|^2+C(m,n)\sum_{i+j+k=m}|\nabla^m A||\nabla^i A||\nabla^j A||\nabla^k A|.$

In the case $M_t=\text{graph } w_t$, the parabolic system (1) is up to tangential diffeomorphisms equivalent to the quasilinear scalar equation

$$\frac{d}{dt}w(y,t)=\sqrt{1+|Dw|^2}\,\operatorname{div}_y\left(\frac{Dw}{\sqrt{1+|Dw|^2}}\right)(y,t). \tag{2}$$

Indeed, suppose that $w(y,t)$ is a smooth solution of Eq. (2) on $\Omega\times[0,T)$, where Ω is some open subset of R^n. Then we define in the obvious way

$$\tilde{\mathbf{F}}(y,t)=(y,w(y,t))\in R^{n+1},\qquad y\in\Omega,\ t\in[0,T)$$

and the normal component of $\frac{\partial\tilde{\mathbf{F}}}{\partial t}$ will yield the flow in (1). To be precise let

$$\varphi\colon\ \Omega\times[0,T)\to R^n$$

be the solution of the following system of ordinary differential equations:

$$\frac{\partial}{\partial t}\varphi(y,t) = -\left(\frac{\partial \tilde{\mathbf{F}}}{\partial y}(\varphi(y,t),t)\right)^{-1} \circ \left(\frac{\partial \tilde{\mathbf{F}}}{\partial t}(\varphi(y,t),t)\right)^{T} \tag{3}$$

$$\varphi(y,0) = y$$

where "T" denotes the tangential component of a vector. Now observe that $v = \sqrt{1+|Dw|^2}$, and the right hand side of (2) equals $-vH$. We then define

$$\mathbf{F}(y,t) = \tilde{\mathbf{F}}(\varphi(y,t),t)$$

and obtain from (2) and (3) that

$$\begin{aligned}\frac{d}{dt}\mathbf{F}(y,t) &= \frac{\partial}{\partial t}\tilde{\mathbf{F}}(\varphi(y,t),t) + \frac{\partial \tilde{\mathbf{F}}}{\partial y}(\varphi(y,t),t) \circ \frac{d}{dt}\varphi(y,t) \\ &= \left[\frac{\partial}{\partial t}\tilde{\mathbf{F}}(\varphi(y,t),t)\right]^{\perp} \\ &= \frac{d}{dt} w \cdot v^{-1} \cdot \nu \\ &= -H \cdot \nu = \mathbf{H}(y,t).\end{aligned}$$

Notice that the solution of (3) exists and is a diffeomorphism as long as $\varphi(y,t)$ remains in Ω since $\tilde{\mathbf{F}}$ is a smooth diffeomorphism. The quantity $\left(\frac{d}{dt}\right)\varphi = -Hv + Hv^{-1}e_{n+1}$ is the projection of $\left(-\left(\frac{d}{dt}\right)\tilde{\mathbf{F}}\right)^{T}$ onto the hyperplane R^n.

Conversely, given a smooth solution $\mathbf{F}_t$ of the mean curvature flow equation which can be written as a graph locally, it is easy to compute that any function $w(y,t)$ with

$$w(\mathbf{F}(p,t) - \langle \mathbf{F}(p,t), \mathbf{e}_{n+1}\rangle \mathbf{e}_{n+1}, t) = \langle \mathbf{F}(p,t), \mathbf{e}_{n+1}\rangle$$

satisfies

$$\begin{aligned}-Hv^{-1} &= \frac{d}{dt} w(\mathbf{F}(p,t) - \langle \mathbf{F}(p,t), \mathbf{e}_{n+1}\rangle \mathbf{e}_{n+1}, t) \\ &= \frac{\partial}{\partial t} w(\mathbf{F}(p,t) - \langle \mathbf{F}(p,t), \mathbf{e}_{n+1}\rangle \mathbf{e}_{n+1}, t) + Hv^{-1}|Dw|^2.\end{aligned}$$

Therefore

$$\frac{d}{dt} w(y,t) = -Hv$$

which is Eq. (2).

2 Local gradient estimates

Let us assume that M_0 is locally given as a graph over some hyperplane determined by $\omega \in R^{n+1}$, $|\omega| = 1$. Then we have the following result.

2.1 Theorem. *Let $R>0$ and $\mathbf{x}_0 \in R^{n+1}$ be arbitrary and define $\varphi(\mathbf{x}, t)=R^2-|\mathbf{x}-\mathbf{x}_0|^2-2nt$. If φ_+ denotes the positive part of φ we have the estimate*

$$v(\mathbf{x}, t)\,\varphi_+(\mathbf{x}, t) \leqq \sup_{M_0} v\varphi_+$$

as long as $v(\mathbf{x}, t)$ is defined everywhere on the support of φ_+.

2.2 Remark. By letting $R \to \infty$ we obtain the global gradient estimate for entire graphs which was derived in [4] by use of a monotonicity formula, see also [13].

Proof of Theorem 2.1. We may assume without loss of generality that $\mathbf{x}_0=0$. For $R>0$ we define $\eta(r)=(R^2-r)^2$. Note that η satisfies

$$\eta^{-1}(\eta')^2=4 \quad \text{and} \quad \eta''=2 \tag{4}$$

where $'$ denotes differentiation with respect to r. If $r=|\mathbf{x}|^2+2nt$ we derive from Lemma 1.1 (i) that

$$\left(\frac{d}{dt}-\Delta\right)\eta=-2|\nabla|\mathbf{x}|^2|^2. \tag{5}$$

Combining (5) with Lemma 1.1 (iii) we obtain

$$\left(\frac{d}{dt}-\Delta\right)v^2\eta=-2|A|^2v^2\eta-6|\nabla v|^2\eta-2|\nabla|\mathbf{x}|^2|^2v^2-2\nabla v^2\cdot\nabla\eta. \tag{6}$$

Observe that

$$-2\nabla v^2\cdot\nabla\eta=-6v\nabla v\nabla\eta+\eta^{-1}\nabla\eta\nabla(v^2\eta)-\eta^{-1}|\nabla\eta|^2v^2$$

as well as by (4)

$$\eta^{-1}|\nabla\eta|^2v^2=\eta^{-1}(\eta')^2|\nabla|\mathbf{x}|^2|^2v^2=4|\nabla|\mathbf{x}|^2|^2v^2.$$

Substituting these identities into (6) implies the estimate

$$\left(\frac{d}{dt}-\Delta\right)v^2\eta\leqq-6|\nabla v|^2\eta-6|\nabla|\mathbf{x}|^2|^2v^2-6v\nabla v\cdot\nabla\eta+\eta^{-1}\nabla\eta\cdot\nabla(v^2\eta).$$

Young's inequality and (4) yield

$$\begin{aligned}6v\nabla v\cdot\nabla\eta&\leqq 6|\nabla v|^2\eta+\tfrac{3}{2}v^2\eta^{-1}(\eta')^2|\nabla|\mathbf{x}|^2|^2\\&=6|\nabla v|^2\eta+6|\nabla|\mathbf{x}|^2|^2v^2.\end{aligned}$$

If we replace η by φ_+^2 this computation remains valid on the support of φ_+ as long as v is defined. The weak parabolic maximum principle then implies the result.

Next we shall establish a bound for the gradient in terms of the height over a fixed set in our reference hyperplane. As in [4] we define the height

of M_t with respect to the hyperplane orthogonal to ω by $u=\langle \mathbf{x}, \omega\rangle$, see Lemma 1.1 (ii).

In contrast, we define the height function $w(y,t)$ to be the height at time t over a fixed point y in this hyperplane.

2.3 Theorem. *The gradient of the height function w satisfies the estimate*

$$\sqrt{1+|Dw(y_0,t)|^2} \leqq C_1(n) \sup_{B_R(y_0)} \sqrt{1+|Dw_0|^2}$$
$$\cdot \exp[C_2(n) R^{-2} \sup_{[0,T]} (\sup_{B_R(y_0)\times[0,T]} w - w(y_0,t))^2]$$

where $0 \leqq t \leqq T$, $B_R(y_0)$ *is a ball in the hyperplane orthogonal to* ω *and* w_0 *denotes the initial height function over this hyperplane.*

2.4 Remark. This estimate should be compared with the corresponding interior gradient estimate for the equation

$$\frac{d}{dt} w = \operatorname{div}\left(\frac{Dw}{\sqrt{1+|Dw|^2}}\right)$$

on R^n obtained in [3].

Proof of Theorem 2.3. We are going to adapt the techniques employed in [14] to the parabolic setting and will work with the functions v and u rather than the height function w.

Let us first assume that $u<0$ in the set $\bar{M}_t=\{x\in M_t/|\mathbf{x}|^2-u^2\leqq 1\}$ for $t\in[0,T]$ and let $\eta=\eta(x,t)\geqq 0$ be a smooth function which for each $t\in[0,T]$ vanishes outside some compact subset of R^{n+1}. Suppose then that $\max_{M_t} v\eta$ reaches some value larger than $\max_{M_{t_0}} v\eta$ for the first time at $0<t_0\leqq T$. At a maximum point on M_{t_0} we have

$$\nabla(v\eta)=0 \tag{7}$$

$$\left(\frac{d}{dt}-\Delta\right)(v\eta)\geqq 0. \tag{8}$$

In view of the evolution equation for v in Lemma 1.1 (iii) we obtain from (7) and (8) the inequality

$$0\leqq v\left(\frac{d}{dt}-\Delta\right)\eta-|A|^2 v\eta$$

where we used the relation

$$-2\nabla\eta\cdot\nabla v=-2v^{-1}\nabla v\cdot\nabla(v\eta)+2v^{-1}|\nabla v|^2\eta$$
$$=2v^{-1}|\nabla v|^2\eta.$$

Since v and η are nonnegative we therefore conclude

$$\left(\frac{d}{dt}-\Delta\right)\eta \geqq 0 \tag{9}$$

at a maximum point of $v\eta$ on M_{t_0}.

Let now $\eta = -1+\exp \lambda\varphi$ where the function φ will be chosen later and where $\lambda > 0$. From (9) we infer

$$\left(\frac{d}{dt}-\Delta\right)\varphi \geqq \lambda |\nabla \varphi|^2. \tag{10}$$

As in [14] set

$$\varphi = \begin{cases} \left(\frac{1}{2\beta}u+1-(|\mathbf{x}|^2-u^2)\right)_+, & \mathbf{x}\in \bar{M}_t,\ \beta>0 \text{ to be chosen} \\ 0, & \text{else.} \end{cases}$$

Since $u<0$ in $\bar{M}_t$, φ is a Lipschitz function vanishing outside $\bar{M}_t$. On the set where φ is positive we compute

$$\begin{aligned} \nabla\varphi &= \frac{1}{2\beta}\nabla u - \nabla(|\mathbf{x}|^2-u^2) \\ |\nabla\varphi|^2 &= \frac{1}{4\beta^2}|\nabla u|^2 + |\nabla(|\mathbf{x}|^2-u^2)|^2 - \frac{1}{\beta}\nabla u\cdot\nabla(|\mathbf{x}|^2-u^2) \\ \left(\frac{d}{dt}-\Delta\right)\varphi &= \frac{1}{2\beta}\left(\frac{d}{dt}-\Delta\right)u - \left(\frac{d}{dt}-\Delta\right)(|\mathbf{x}|^2-u^2). \end{aligned} \tag{11}$$

From Lemma 1.1 (i) and (ii) we obtain finally

$$\left(\frac{d}{dt}-\Delta\right)\varphi = 2(n-|\nabla u|^2) \leqq 2n. \tag{12}$$

Substituting (11) and (12) into (10) we arrive at the inequality

$$\lambda\left(\frac{1}{4\beta^2}|\nabla u|^2 - \frac{1}{\beta}\nabla u\cdot\nabla(|\mathbf{x}|^2-u^2)\right) \leqq 2n. \tag{13}$$

Now note that $\nabla u = \omega - \langle \nu, \omega\rangle$ and $\nabla(|\mathbf{x}|^2-u^2) = 2(\mathbf{x}-\langle \mathbf{x}, \nu\rangle\, \nu - u\nabla u)$ such that

$$\nabla u\cdot\nabla(|\mathbf{x}|^2-u^2) = 2(u-\langle \mathbf{x}, \nu\rangle\langle \nu, \omega\rangle - u|\nabla u|^2).$$

Using also that $|\nabla u|^2 = 1-v^{-2}$ we therefore derive

$$\begin{aligned} \nabla u\cdot\nabla(|\mathbf{x}|^2-u^2) &= 2v^{-1}(uv^{-1}-\langle \mathbf{x}, \nu\rangle) = 2v^{-1}(u\langle \nu, \omega\rangle - \langle \mathbf{x}, \nu\rangle) \\ &= -2v^{-1}\langle \nu, \mathbf{x}-u\omega\rangle \leqq 2v^{-1}|\nu|\,|\mathbf{x}-u\omega| \leqq 2v^{-1}, \end{aligned}$$

since $|\mathbf{x}-u\omega|^2=|\mathbf{x}|^2-u^2\leqq 1$ in the set under consideration. Thus, again using $|\nabla u|^2=1-v^{-2}$ we deduce from (13) that

$$\lambda\left(\frac{1}{4\beta^2}(1-v^{-2})-\frac{2}{\beta}v^{-1}\right)\leqq 2n. \tag{14}$$

Let now $\lambda=64n\beta^2$. Then at the maximum point of $v\eta$ on M_{t_0} we have

$$v\leqq 4+16\beta.$$

Hence

$$\begin{aligned}\max_{M_{t_0}} v\eta &\leqq (4+16\beta)\sup_{M_{t_0}}\eta\\ &\leqq (4+16\beta)\,e^{\lambda}=(4+16\beta)\,e^{64n\beta^2},\end{aligned}$$

where we used the fact that $\varphi\leqq 1$. This proves that for all $t\in[0,T]$ we have inside $\bar{M}_t$ the estimate

$$v\eta\leqq\sup_{M_0} v\eta+(4+16\beta)\,e^{64n\beta^2}. \tag{15}$$

As described in Sect. 1 this estimate implies an equivalent bound for the height function w. At the point $y=0$ in the hyperplane orthogonal to ω we obtain in equivalence to (15) for $t\in[0,T]$ and arbitrary $\beta>0$

$$\begin{aligned}&[e^{64n\beta^2((2\beta)^{-1}w(0,t)+1)_+}-1]\sqrt{1+|Dw(0,t)|^2}\\ &\quad\leqq e^{64n\beta^2}(\sup_{B_1(0)}\sqrt{1+|Dw_0|^2}+4+16\beta).\end{aligned} \tag{16}$$

Now choose $\beta=\sup_{t\in[0,T]}-w(0,t)$. We then infer from (16)

$$\sqrt{1+|Dw(0,t)|^2}\leqq C_1(n)\sup_{B_1(0)}\sqrt{1+|Dw_0|^2}\exp[C_2(n)\sup_{[0,T]}(-w(0,t))^2].$$

To achieve the condition $w<0$ on $B_1(0)$ we replace w by $w-\sup_{B_1(0)\times[0,T]} w(y,t)$ and the estimate on $B_R(y_0)$ is then obtained by scaling and translating.

3 Local curvature and higher order estimates

In this section we prove that as long as M_t can be written as a graph with bounded gradient, the curvature and all its derivatives remain bounded as well.

Let $r=r(\mathbf{x},t)$ be a smooth nonnegative function such that

$$\left|\left(\frac{d}{dt}-\Delta\right)r\right|\leqq C(n)\quad\text{and}\quad|\nabla r|^2\leqq C(n)\,r \tag{17}$$

holds on any surface M_t moving by its mean curvature. We begin with a curvature estimate.

3.1 Theorem. *Let $R>0$ be such that $\{\mathbf{x}\in M_t / r(\mathbf{x},t)\leqq R^2\}$ is compact and can be written as a graph over some hyperplane for $t\in[0,T]$. Then for any $t\in[0,T]$ and $0\leqq\theta<1$ we have the estimate*

$$\sup_{\{(\mathbf{x}\in M_t/r(\mathbf{x},t)\leqq\theta R^2)} |A|^2\leqq c(n)(1-\theta)^{-2}(t^{-1}+R^{-2}) \sup_{\{\mathbf{x}\in M_s/r(\mathbf{x},s)\leqq R^2,\, s\in[0,t]\}} v^4.$$

3.2 Corollary. (i) *Let $R>0$ and $0\leqq\theta<1$ be as above. Then for $\mathbf{x}_0\in R^{n+1}$ and $t\in[0,T]$ we have the estimate*

$$\sup_{K(\mathbf{x}_0,t,\theta R^2)} |A|^2\leqq c(n)(1-\theta)^{-2}t^{-1} \sup_{0\leqq s\leqq t} \sup_{K(\mathbf{x}_0,s,R^2)} v^4$$

where $K(\mathbf{x}_0,t,\theta R^2)=\{x\in M_t/|\mathbf{x}-\mathbf{x}_0|^2+2nt\leqq\theta R^2\}$.

(ii) *Let $R>0$ and $0\leqq\theta<1$ be as above. Then we have for $t\in[0,T]$ the estimate*

$$\sup_{B_{\theta R}(y_0)} |A|^2(t)\leqq c(n)(1-\theta^2)^{-2}\left(\frac{1}{R^2}+\frac{1}{t}\right) \sup_{B_R(y_0)\times[0,t]} v^4$$

where $B_R(y_0)$ denotes a ball in the hyperplane orthogonal to ω.

Proof of Corollary 3.2. We may assume w.l.o.g. that $\mathbf{x}_0=0$, $y_0=0$. To prove (i) we note that $r=|\mathbf{x}|^2+2nt$ satisfies condition (17) in view of Lemma 1.1 (i) and the inequality $|\nabla|\mathbf{x}|^2|^2\leqq 4|\mathbf{x}|^2$. The estimate then follows in view of the fact $t\leqq\frac{R^2}{2n}$ when $r(\mathbf{x},t)\leqq R^2$. To prove the second estimate observe that $r(\mathbf{x},t)=|\mathbf{x}|^2-u^2$ satisfies condition (17) in view of Lemma 1.1 (i) and (ii), the inequality $|\nabla u|\leqq 1$ and the fact that $|\nabla(|\mathbf{x}|^2-u^2)|^2\leqq 4(|\mathbf{x}|^2-u^2)$.

3.3 Remark. (i) By combining Theorem 2.1 with Corollary 3.2 (i) and letting $R\to\infty$ we obtain the global curvature estimate derived in [4].

(ii) By combining Theorem 2.3 with Corollary 3.2 we can estimate the curvature in terms of a height bound and the initial gradient.

(iii) Corollary 3.2 (ii) includes the estimate in [2] for $H\equiv 0$ as a special case, see also Remark 3.6 (ii).

Proof of Theorem 3.1. From Lemma 1.1 (v) and Schwarz' inequality we infer that

$$\left(\frac{d}{dt}-\Delta\right)|A|^2\leqq 2|A|^4-2|\nabla|A||^2. \tag{18}$$

Utilizing an idea of Caffarelli, Nirenberg and Spruck in [2] we combine (18) with the evolution equation for the gradient in Lemma 1.1 (iii) to deduce that

$$\left(\frac{d}{dt}-\Delta\right)|A|^2\varphi(v^2)$$
$$\leqq 2(\varphi-\varphi' v^2)|A|^4-2|\nabla|A||^2\varphi-(6\varphi'+4\varphi'' v^2)|A|^2|\nabla v|^2-2\nabla|A|^2\cdot\nabla\varphi,$$

where $\varphi=\varphi(v^2)$ is a smooth nonnegative function of v^2 to be determined later and $'$ denotes derivatives with respect to v^2. The last term can be estimated by

$$\begin{aligned}-2\nabla|A|^2\cdot\nabla\varphi &= -\varphi^{-1}\nabla\varphi\cdot\nabla(|A|^2\varphi)+\varphi^{-1}|\nabla\varphi|^2|A|^2-4|A|\,v\varphi'\nabla|A|\cdot\nabla v\\ &\leqq -\varphi^{-1}\nabla\varphi\cdot\nabla(|A|^2\varphi)+4\varphi^{-1}(\varphi')^2v^2|\nabla v|^2|A|^2+2|\nabla|A||^2\varphi\\ &\quad +2|A|^2v^2\varphi^{-1}(\varphi')^2|\nabla v|^2\\ &= -\varphi^{-1}\nabla\varphi\cdot\nabla(|A|^2\varphi)+2|\nabla|A||^2\varphi+6\varphi^{-1}(\varphi')^2v^2|A|^2|\nabla v|^2.\end{aligned}$$

Hence we obtain

$$\begin{aligned}(19)\qquad \left(\frac{d}{dt}-\Delta\right)|A|^2\varphi &\leqq 2(\varphi-\varphi'v^2)|A|^4\\ &\quad -\varphi^{-1}\nabla\varphi\cdot\nabla(|A|^2\varphi)-(6\varphi'(1-\varphi^{-1}\varphi'v^2)+4\varphi''v^2)|A|^2|\nabla v|^2.\end{aligned}$$

As in [2] we let $\varphi(v^2)=\dfrac{v^2}{1-kv^2}$, $k>0$ to be chosen. Observe that then

$$\varphi-\varphi'v^2=-k\varphi^2,$$

$$6\varphi'(1-\varphi^{-1}\varphi'v^2)+4\varphi''v^2=\frac{2k}{(1-kv^2)^2}\varphi,$$

and

$$\varphi^{-1}\nabla\varphi=2\varphi v^{-3}\nabla v.$$

Substituting these identities into (19) we derive for $g=|A|^2\varphi(v^2)$ the inequality

$$(20)\qquad \left(\frac{d}{dt}-\Delta\right)g\leqq -2kg^2-\frac{2k}{(1-kv^2)^2}|\nabla v|^2g-2\varphi v^{-3}\nabla v\cdot\nabla g.$$

Now let $\eta=\eta(r)=(R^2-r)^2$ where $r=r(\mathbf{x},t)$ satisfies (17). Since by (4) and (17)

$$\begin{aligned}\left(\frac{d}{dt}-\Delta\right)\eta &= \eta'\left(\frac{d}{dt}-\Delta\right)r-\eta''|\nabla r|^2\\ &\leqq 2c(n)R^2-2|\nabla r|^2\end{aligned}$$

we compute from (20)

$$\begin{aligned}(21)\qquad \left(\frac{d}{dt}-\Delta\right)g\eta &\leqq -2kg^2\eta-\frac{2k}{(1-kv^2)^2}|\nabla v|^2g\eta-2\varphi v^{-3}\nabla v\cdot\nabla g\cdot\eta\\ &\quad +c(n)R^2g-2g|\nabla r|^2-2\nabla g\cdot\nabla\eta\end{aligned}$$

where $c(n)$ denotes constants depending on n only. The last term can be written as

$$(22)\qquad -2\nabla g\cdot\nabla\eta=-2\eta^{-1}\nabla\eta\cdot\nabla(g\eta)+8g|\nabla r|^2$$

in view of (4). We also estimate

$$
\begin{aligned}
(23)\quad & -2\varphi v^{-3}\nabla v\cdot\nabla g\eta \\
&= -2\varphi v^{-3}\nabla v\cdot\nabla(g\eta)+2\varphi v^{-3}g\eta'\nabla v\cdot\nabla r \\
&\leqq -2\varphi v^{-3}\nabla v\cdot\nabla(g\eta)+\frac{2k}{(1-kv^2)^2}|\nabla v|^2 g\eta+2\varphi^2 v^{-6}\frac{(1-kv^2)^2}{k}g|\nabla r|^2 \\
&= -2\varphi v^{-3}\nabla v\cdot\nabla(g\eta)+\frac{2k}{(1-kv^2)^2}|\nabla v|^2 g\eta+\frac{2}{kv^2}g|\nabla r|^2
\end{aligned}
$$

where we again used (4) and the fact that

$$\varphi^2 v^{-6}(1-kv^2)^2 = v^{-2}.$$

Substituting (22) and (23) into (21) and using (17) we derive

$$\left(\frac{d}{dt}-\Delta\right)g\eta \leqq -2kg^2\eta-2(\varphi v^{-3}\nabla v+\eta^{-1}\nabla\eta)\cdot\nabla(g\eta)+c(n)\left(\left(1+\frac{1}{kv^2}\right)r+R^2\right)g$$

and hence

$$
\begin{aligned}
\left(\frac{d}{dt}-\Delta\right)tg\eta \leqq & -2kg^2\eta t-2(\varphi v^{-3}\nabla v+\eta^{-1}\nabla\eta)\cdot\nabla(g\eta t) \\
& +c(n)\left(\left(1+\frac{1}{kv^2}\right)r+R^2\right)tg+g\eta.
\end{aligned}
$$

At a point where $m(T)=\sup\limits_{0\leqq t\leqq T}\ \sup\limits_{\{\mathbf{x}\in M_t/r(\mathbf{x},t)\leqq R^2\}} tg\eta$ is attained for $t_0>0$ we compute

$$2kg^2\eta t_0 \leqq c(n)\left(1+\frac{1}{kv^2}\right)R^2 g t_0+g\eta.$$

Multiplying by $\eta\dfrac{t_0}{2k}$ yields since $t_0\leqq T$

$$m(T)^2 \leqq \frac{c(n)}{2k}\left(\left(1+\frac{1}{kv^2}\right)R^2 T+\eta\right)m(T).$$

Since $\eta\leqq R^4$ we arrive at

$$(24)\qquad g\eta T\leqq\frac{c(n)}{2k}\left(\left(1+\frac{1}{kv^2}\right)R^2 T+R^4\right)$$

in the set $\{\mathbf{x}\in M_T/r(\mathbf{x},T)\leqq R^2\}$. Let now

$$k=\tfrac{1}{2}\inf_{\{\mathbf{x}\in M_t/r(\mathbf{x},t)\leqq R^2,\, t\in[0,T]\}} v^{-2}.$$

Since $\varphi(v^2) \geqq 1$ and $\eta \geqq (1-\theta)^2 R^4$ in $\{\mathbf{x} \in M_T / r(\mathbf{x}, T) \leqq \theta R^2\}$ we finally obtain from (24) replacing T by any $t \in [0, T]$

$$\sup_{\{\mathbf{x} \in M_t / r(\mathbf{x}, t) \leqq R^2\}} |A|^2 \leqq c(n)(1-\theta)^{-2}\left(\frac{1}{R^2}+\frac{1}{t}\right) \sup_{0 \leqq s \leqq t} \sup_{\{\mathbf{x} \in M_s / r(\mathbf{x}, s) \leqq R^2\}} v^4$$

as desired.

We will now extend these estimates to all derivatives of A.

3.4 Theorem. *Let $r = r(\mathbf{x}, t) \geqq 0$ be as in (17) and let $R > 0$ be such that $\{\mathbf{x} \in M_t / r(\mathbf{x}, t) \leqq R^2\}$ is compact and can be written as a graph over some hyperplane for $t \in [0, T]$. Then for any $m \geqq 0$, $0 \leqq \theta < 1$ and $t \in [0, T]$ we have the estimate*

$$\sup_{\{\mathbf{x} \in M_t / r(\mathbf{x}, t) \leqq R^2\}} |\nabla^m A|^2 \leqq c_m \left(\frac{1}{R^2}+\frac{1}{t}\right)^{m+1}$$

where $c_m = c_m(\theta, n, m, \sup_{\{\mathbf{x} \in M_s / r(\mathbf{x}, s) \leqq R^2, s \in [0, t]\}} v)$

As in Corollary 2.2 setting $r = |\mathbf{x} - \mathbf{x}_0|^2 + 2nt$ or $r = |\mathbf{x}|^2 - u^2$ implies

3.5 Corollary. (i) *Let $R > 0$ and $0 \leqq \theta < 1$, $m > 0$ be as above. Then we have for $\mathbf{x}_0 \in R^{n+1}$, $t \in [0, T]$ the estimate*

$$\sup_{K(\mathbf{x}_0, t, \theta R^2)} t^{m+1} |\nabla^m A|^2 \leqq c_m$$

where $K(\mathbf{x}_0, t, \theta R^2) = \{\mathbf{x} \in M_t / |\mathbf{x} - \mathbf{x}_0|^2 + 2nt \leqq R^2\}$ and c_m is as in Theorem 3.4.

(ii) *Let $R > 0$, $0 \leqq \theta < 1$ and $m \geqq 0$ be as above. Then we have for $t \in [0, T]$ the estimate*

$$\sup_{B_{\theta R}(y_0)} |\nabla^m A|^2(t) \leqq c_m \left(\frac{1}{R^2}+\frac{1}{t}\right)^{m+1}$$

where $B_R(y_0)$ denotes a ball in the hyperplane orthogonal to ω and c_m is as above.

3.6 Remark. (i) By combining Theorem 2.1 and Corollary 3.5 (i) and letting $R \to \infty$ we obtain the global estimates derived in [4].

(ii) From the proof of Theorem 3.4 we infer that the constant c_m is of the form $c_m = c(n, \theta, m)\, c_0^{\alpha(m)}$ where $\alpha(m)$ is a positive integer (bounded by 2^m) and c_0 is s.t.

$$\sup_{\{\mathbf{x} \in M_s / r(\mathbf{x}, s) \leqq R^2\}} |A|^2 \leqq c_0 \left(\frac{1}{R^2}+\frac{1}{s}\right)$$

for any $s \in [0, t]$. The graph property of M_t is only used to bound $|A|^2$ as in Theorem 3.1. One determines an optimal exponent $\alpha(m)$ by estimating the last term in the evolution equation for $|\nabla^m A|^2$ (see Lemma 1.1 (vi)) similarly as in Proposition 4.4 of [4].

(iii) Since minimal hypersurfaces ($H \equiv 0$) are stationary solutions of (1) the estimates 3.5 (i), (ii) for $m \geqq 0$ with t proportional to R^2 provide an interior regularity theory for solutions of the minimal surface equation obtained entirely from

geometric computations. To minimize the exponent of v in the $|\nabla^m A|$-estimate one uses the curvature bound in [5] and then finds the optimal exponent $\alpha(m)$ as in (ii) above.

Proof of Theorem 3.4. We proceed by induction on m in a similar way to that in [4]. The case $m=0$ was proved in Theorem 3.1. For fixed $R>0$ let $\psi(t) = \dfrac{R^2 t}{(R^2+t)}$. Suppose

$$\sup_{\{\mathbf{x}\in M_t / r(\mathbf{x},t)\leqq R^2,\, t\in[0,T]\}} \psi^k |\nabla^{k-1} A|^2 \leqq c_{k-1} \tag{25}$$

has been established for $1\leqq k\leqq m$. We then want to estimate $\psi^{m+1}|\nabla^m A|^2$ in $\{\mathbf{x}\in M_t / r(\mathbf{x},t)\leqq \theta R^2\}$. From Lemma 1.1 (vi) we obtain for $m\geqq 1$

$$\left(\frac{d}{dt}-\Delta\right)\psi^{m+1}|\nabla^m A|^2 \leqq -2\psi^{m+1}|\nabla^{m+1}A|^2$$
$$+c(m,n)\,\psi^{m+1}\sum_{i+j+k=m}|\nabla^i A||\nabla^j A||\nabla^k A||\nabla^m A| \quad +\left(\frac{d}{dt}\psi^{m+1}\right)|\nabla^m A|^2.$$

In $\{\mathbf{x}\in M_t / r(\mathbf{x},t)\leqq \theta R^2\}$ we have by (25)

$$\psi^{m+1}\sum_{i+j+k=m}|\nabla^i A||\nabla^j A||\nabla^k A||\nabla^m A| \leqq c\sum_{k\leqq m}\psi^k|\nabla^k A|^2$$

similarly as in [4], where $c=c(n,m,c_0,\ldots,c_{m-1})$. Since

$$\frac{d}{dt}\psi^{m+1} = (m+1)\,\psi^m \frac{R^4}{(R^2+t)^2} \leqq (m+1)\,\psi^m$$

we infer in $\{\mathbf{x}\in M_t / r(\mathbf{x},t)\leqq R^2\}$ the inequality

$$\left(\frac{d}{dt}-\Delta\right)\psi^{m+1}|\nabla^m A|^2 \leqq -2\psi^{m+1}|\nabla^{m+1}A|^2 + C\sum_{k\leqq m}\psi^k|\nabla^k A|^2,$$

where $c=c(n,m,c_0,\ldots,c_{m-1})$. Following the work of Shi in [17] we now define

$$f=\psi^{m+1}|\nabla^m A|^2\,(\Lambda+\psi^m|\nabla^{m-1}A|^2),$$

where $\Lambda>0$ will be chosen appropriately. We then compute, using again (25)

$$\left(\frac{d}{dt}-\Delta\right)f \leqq -2\psi^{m+1}|\nabla^{m+1}A|^2(\Lambda+\psi^m|\nabla^{m-1}A|^2)$$
$$+C\sum_{k\leqq m}\psi^k|\nabla^k A|^2\,(\Lambda+\psi^m|\nabla^{m-1}A|^2)$$
$$-2\psi^{2m+1}|\nabla^m A|^4 + \sum_{k\leqq m-1}\psi^k|\nabla^k A|^2\,\psi^{m+1}|\nabla^m A|^2$$
$$-2\psi^{2m+1}\nabla|\nabla^m A|^2\cdot\nabla|\nabla^{m-1}A|^2.$$

The last term we estimate by

$$2\psi^{2m+1}\nabla|\nabla^m A|^2\cdot\nabla|\nabla^{m-1}A|^2 \leqq 2\psi^{m+1}|\nabla^m A|^2(\Lambda+\psi^m|\nabla^{m-1}A|^2) + 8\psi^{2m+1}\frac{\psi^m|\nabla^{m-1}A|^2}{\Lambda+\psi^m|\nabla^{m-1}A|^2}|\nabla^m A|^4.$$

Noting that

$$\frac{\psi^m|\nabla^{m-1}A|^2}{\Lambda+\psi^m|\nabla^{m-1}A|^2} \leqq \frac{c_{m-1}}{\Lambda+c_{m-1}}$$

we obtain in $\{\mathbf{x}\in M_t/r(\mathbf{x},t)\leqq R^2\}$ that

$$\left(\frac{d}{dt}-\Delta\right)f \leqq -2\psi^{-1}\left(1-\frac{4c_{m-1}}{\Lambda+c_{m-1}}\right)(\psi^{m+1}|\nabla^m A|^2)^2 + c\psi^{-1}\left(\sum_{k\leqq m}\psi^{k+1}|\nabla^k A|^2(\Lambda+\psi^m|\nabla^{m-1}A|^2) + \sum_{k\leqq m-1}\psi^{k+1}|\nabla^k A|^2\,\psi^{m+1}|\nabla^m A|^2\right).$$

Setting $\Lambda=7c_{m-1}+1$ and collecting terms we obtain in view of (25) in $\{\mathbf{x}\in M_t/r(\mathbf{x},t)\leqq R^2\}$

$$\left(\frac{d}{dt}-\Delta\right)f \leqq -\psi^{-1}(\Lambda+\psi^m|\nabla^{m-1}A|^2)^{-2}f^2+c\psi^{-1}(1+f),$$

where c depends on n, m, $c_0,\ldots,c_{m-1}$. Applying now (25) to the f^2-term and using Young's inequality on the second term we arrive at

$$\left(\frac{d}{dt}-\Delta\right)f \leqq -\psi^{-1}(\delta f^2-c),$$

where δ and c depend on n, m, $c_0,\ldots,c_{m-1}$. Similarly as in the proof of Theorem 3.1 we compute for $\eta=(R^2-r)^2$ using (17) that

$$\left(\frac{d}{dt}-\Delta\right)f\eta \leqq \psi^{-1}(-\delta f^2+c)\eta-2\eta^{-1}\nabla\eta\cdot\nabla(f\eta)+c(n)R^2 f.$$

Since $\psi(0)=0$ we have $f\eta\equiv 0$ at $t=0$. At the point where $m(T)=\sup_{0\leqq t\leqq T}\sup_{\{\mathbf{x}\in M_t/r(\mathbf{x},t)\leqq R^2\}} f\eta$ is attained for $t_0>0$ we compute

$$f^2\eta \leqq \frac{1}{\delta}(c\eta+c(n)R^2 f\psi).$$

Multiplying by η and using $\eta\leqq R^4$ and $\psi\leqq R^2$ we obtain since $\delta=\delta(n,m_0,c_0,\ldots,c_{m-1})$ the estimate

$$f^2\eta^2 \leqq c(R^8+R^4 f\eta),$$

where $c=c(n,m,c_0,\ldots,c_{m-1})$. Hence, by Young's inequality we derive

$$m(T)^2 \leqq c\cdot R^8$$

and therefore

$$f\eta \leqq cR^4$$

in $\{\mathbf{x}\in M_T/r(\mathbf{x}, T)\leqq R^2\}$ with $c=c(n, m, c_0, \ldots, c_{m-1})$. Thus we finally arrive at

$$\sup_{\{\mathbf{x}\in M_T/r(\mathbf{x}, T)\leqq R^2\}} \psi^{m+1}|\nabla^m A|^2 \leqq c(1-\theta)^{-2}$$

where $c=c(n, m, c_0, \ldots, c_{m-1})$ and where we used the fact that $\eta\geqq(1-\theta)^2 R^4$ and $\Lambda\geqq 1$. Replacing T by any $t\in[0, T]$ yields the result.

If we assume additional smoothness of M_t we can improve the time dependence of the higher derivative estimates. In particular, it is not necessary to assume that M_t can be locally written as a graph.

3.7 Theorem. *Let $r=r(\mathbf{x}, t)\geqq 0$ be as in (17) and let $R>0$ be such that $\{\mathbf{x}\in M_t/r(\mathbf{x}, t)\leqq R^2\}$ is compact for $t\in[0, T]$. Then for $0\leqq\theta<1$, $t\in[0, T]$ and any integers $\ell\geqq 0$, $m\geqq 0$ we have the estimate*

$$\sup_{\{\mathbf{x}\in M_t/r(\mathbf{x}, t)\leqq R^2\}} |\nabla^{m+\ell} A|^2 \leqq c\left(1+\frac{1}{R^2}+\frac{1}{t}\right)^{\ell}$$

where $c=c\left(m, \ell, n, \theta, \sup_{\{\mathbf{x}\in M_s/r(\mathbf{x}, s)\leqq R^2, s\in[0, t]\}} \sum_{i=0}^{m} |\nabla^i A|^2\right)$.

In particular, we obtain the estimate

$$\sup_{K(\mathbf{x}_0, t, \theta R^2)} |\nabla^{m+\ell} A|^2 \leqq c t^{-\ell}.$$

3.8 Remark. For $m=0$ and $\ell=1$ we have

$$c=c(n, \theta)\cdot \sup_{\{\mathbf{x}\in M_s/r(\mathbf{x}, s)\leqq R^2, s\in[0, t]\}} |A|^2.$$

For general m and ℓ the constant can be optimized as outlined in Remark 3.6 (ii).

Proof of Theorem 3.7. Observe that similar as in the proof of Theorem 3.4 the function $f=\psi^{\ell}|\nabla^{m+\ell} A|^2 (\Lambda+\psi^{\ell-1}|\nabla^{m+\ell-1} A|^2)$ with $\psi=\left(1+\frac{1}{R^2}+\frac{1}{t}\right)^{-1}$ satisfies an inequality of the form

$$\left(\frac{d}{dt}-\Delta\right) f \leqq \psi^{-1}(\delta f^2 - C).$$

4 Flow of complete surfaces

In this section we want to establish a number of results for complete, noncompact surfaces moving by mean curvature. In particular, we will use the interior estimates from the previous section to obtain a short time existence result for arbitrary immersed, complete initial surfaces assuming only a uniform local Lipschitz condition. A similar approach to a short time existence result is also

known to R. Hamilton [9]. Furthermore, we will derive a general maximum principle for heat equations on Riemannian manifolds with a time dependent metric.

We begin with a result for the Dirichlet problem. Let M^n be a smooth manifold with boundary ∂M^n a compact $(n-1)$ dimensional smooth submanifold such that $\bar{M}^n = M^n \cup \partial M^n$ is compact. Then the following shorttime existence result holds.

4.1 Proposition. *Suppose that* $\mathbf{F}_0 \colon \bar{M}^n \to R^{n+1}$ *is a smooth isometric immersion. Then there is* $T_0 > 0$ *such that the initial value problem*

$$
\begin{aligned}
&\frac{d}{dt}\mathbf{F}(pt) = \mathbf{H}(p,t), && p \in M^n, \quad 0 < t < T_0 \\
&\mathbf{F}(p,0) = \mathbf{F}_0(p), && p \in M^n \\
&\mathbf{F}(p,t) = \mathbf{F}_0(p), && p \in \partial M^n, \quad 0 \leqq t \leqq T_0
\end{aligned} \tag{26}
$$

has a unique smooth solution on $0 \leqq t < T_0$. *The constant* $T_0 > 0$ *only depends on a bound for the second fundamental form of* M_0 *and* ∂M_0.

Proof. Since $\mathbf{F}_0$ is smooth there is a tubular neighbourhood of M_0 with Gaussian coordinates $\hat{x}(p) = (x_1, \ldots, x_n)$, y such that at a point q in this neighbourhood $y(q)$ is the length of the geodesic arc perpendicular to M_0 through q and $\hat{x}(q)$ denotes the coordinate vector at the base point of the geodesic on M_0. Let $\omega(\hat{x})$ be the unit normal to M_0 at $\hat{x}$. In these coordinates consider the scalar initial value problem

$$
\begin{aligned}
&\frac{d}{dt} w(\hat{x},t) = \langle v(\hat{x},t), \omega(\hat{x}) \rangle^{-1} H(\hat{x},t), && \hat{x} \in M_0, \quad 0 < t < T_0 \\
&w(\hat{x},0) = 0, && \hat{x} \in M_0 \\
&w(\hat{x},t) = 0, && \hat{x} \in \partial M_0, \quad 0 \leqq t < T_0
\end{aligned} \tag{27}
$$

where v and H are the unit normal and the mean curvature respectively of the hypersurface given by w. This is a quasilinear second order equation which is uniformly parabolic as long as $\langle v, \omega \rangle^{-1}$ remains uniformly bounded. At time $t = 0$ we have $\langle v, \omega \rangle \equiv 1$ and standard arguments yield an a priori estimate for $\langle v, \omega \rangle^{-1}$ on a time interval $[0, T_0)$ which only depends on a bound for the curvature of M_0 and ∂M_0. Note that an interior estimate for the gradient function $\langle v, \omega \rangle^{-1}$ also follows from Theorem 2.3. Then the results of Ladyzhenskaya et al. [15] apply and yield a smooth solution of (27) on $[0, T_0)$. Using an appropriate tangential diffeomorphism exactly as in Sect. 1 we then construct a smooth solution of the original problem (26).

Now let $\mathbf{F}_0 \colon M^n \to R^{n+1}$ be the smooth isometric immersion of a complete hypersurface with bounded curvature in R^{n+1}. Fix a point $p_0 \in M^n$ and as in [17] choose a family of domains $\{D_k | k = 1, 2, 3, \ldots\}$ on M^n such that for each k the boundary ∂D_k is a compact C_∞, $(n-1)$-dimensional submanifold of M^n and

(i) $\bar{D}_k = D_k \cup \partial D_k$ is a compact subset of M^n,

(ii) the geodesic ball $B(p,k)$ is contained in D_k.

For each $k = 1, 2, 3, \ldots$ let $\mathbf{F}_t^k$ be the solution of the Dirichlet problem (26) with initial data $\mathbf{F}_0$ on $\bar{D}_k$. Proposition 4.1 guarantees that such a solution exists for each k on a time interval $[0, T_0)$ with $0 < T_0$ independent of k. Since M_0 has bounded curvature our interior estimates in Theorem 3.4 apply and we

may take the limit $k \to \infty$ to obtain a smooth solution of the mean curvature flow (1) with initial data M_0.

Furthermore, in view of the interior gradient estimate in Theorem 2.3 a simple approximation argument shows that it is sufficient for the short time existence result to assume the following uniform local Lipschitz condition on M_0.

For $q \in M^n$ let $U_{r,q}$ be the q-component of $\mathbf{F}_0^{-1}(B_r^{n+1}(\mathbf{F}_0(q)))$. We say that M_0 satisfies a *uniform local Lipschitz condition* if there are fixed numbers $r_0 > 0$, $C_0 < \infty$ such that for each $q \in M^n$, $\mathbf{F}_0(U_{r_0,q})$ can be written as the graph of a Lipschitz function over a hyperplane in R^{n+1} through $\mathbf{F}_0(q)$ with Lipschitz constant less than C_0.

Using Theorem 2.3 and the special time dependence of the curvature estimates in Corollary 3.5 (ii) we obtain after approximation:

4.2 Theorem. *Let $\mathbf{F}_0 : M^n \to R^{n+1}$ satisfy a uniform local Lipschitz condition. Then the mean curvature flow problem*

$$\frac{d}{dt}\mathbf{F}(p,t) = \mathbf{H}(p,t), \qquad p \in M^n,\ t > 0$$
$$\mathbf{F}(p,0) = \mathbf{F}_0(p)$$

has a solution $\mathbf{F}_t$ on some time interval $[0, T_0)$. $\mathbf{F}_t$ is smooth for $t > 0$ and at $t = 0$ it is Hölder continuous in time with Hölder exponent $\alpha = \frac{1}{2}$.

We will now adapt a method of Liao and Tam [14] to establishing a maximum principle for heat equations which not only applies to our noncompact hypersurfaces, but to general Riemannian manifolds with a time dependent metric. Notice that a maximum principle for the mean curvature flow was obtained in [4] assuming more restrictive growth properties of the solution. Let M^n be a complete, non-compact Riemannian manifold with time dependent metric $g(t) = \{g_{ij}(t)\}$ for $0 \leqq t < T$. We denote by $B_r^t(p)$ the geodesic ball of radius r centred at p at time t and will often use a superscript or subscript t to denote other time dependent quantities.

4.3 Theorem. *Suppose that the manifold M^n with Riemannian metrics $g(t)$ satisfies a uniform volume growth restriction, namely*

$$\mathrm{vol}^t(B_r^t(p)) \leqq \exp(k(1 + r^2))$$

holds for some point $p \in M^n$ and a uniform constant $k > 0$ for all $t \in [0, T]$.

Let be a function on $M^n \times [0, T]$ which is smooth on $M^n \times (0, T]$ and continuous on $M^n \times [0, T]$. Assume that f and $g(t)$ satisfy

(i) $\dfrac{d}{dt} f \leqq \Delta^t f + \mathbf{a} \cdot \nabla f + bf$ *where the function b satisfies* $\sup_{M^n \times [0,T]} |b| \leqq \alpha_0$ *for some* $\alpha_0 < \infty$ *and the vector field* $\mathbf{a}$ *satisfies* $\sup_{M^n \times [0,T]} |\mathbf{a}| \leqq \alpha_1$ *for some* $\alpha_1 < \infty$,

(ii) $f(p, 0) \leqq 0$ *for all* $p \in M^n$,

(iii) $\int_0^T \left(\int_M \exp(-\alpha_2^2 r^t(p, y)^2) |\nabla f|^2 (y)\, d\mu_t \right) dt < \infty$, *for some* $\alpha_2 > 0$

(iv) $\sup_{M^n \times [0,T]} \left| \dfrac{d}{dt} g_{ij} \right| \leqq \alpha_3$, *for some* $\alpha_3 < \infty$.

Then we have $f \leqq 0$ on $M^n \times [0, T]$.

Proof. First notice that in view of (iv) and ([8, Lemma 14]) all metrics $g(t)$ are uniformly equivalent, i.e. there is a constant α_4 depending only on α_3 and T such that for any vector $v(q)$, $q \in M^n$ we have for $t \in [0, T]$

$$\alpha_4^{-1}\, g_{ij}(q,0)\, v^i(q)\, v^j(q) \leqq g_{ij}(q,t)\, v^i(q)\, v^j(q) \leqq \alpha_4\, g_{ij}(q,0)\, v^i(q)\, v^j(q). \tag{28}$$

We then proceed as in [16] and let $0<\eta<\min\left(T, \frac{1}{64K}, \frac{1}{32\alpha_2}\right)$ be a fixed constant. Define for $\theta>0$ to be chosen

$$h(y,s) = -\frac{\theta r_s^2(p,y)}{4(2\eta - s)}, \qquad 0<s<\eta$$

where $r_s(p,y)$ is the distance between p and y at time s. Then h is Lipschitz continuous and

$$\begin{aligned}\frac{d}{ds}h &= -\frac{\theta r_s^2(p,y)}{4(2\eta-s)^2} - \frac{\theta r_s}{2(2\eta-s)}\frac{d}{ds}r_s \\ &= -\theta^{-1}|\nabla h|^2 - \frac{\theta r_s}{2(2\eta-s)}\frac{d}{ds}r_s.\end{aligned}$$

Now observe that for any curve γ with length $l(s)$ we have

$$\left|\frac{d}{ds}l(s)\right| = \left|\frac{d}{ds}\int_\gamma \sqrt{g(\dot\gamma,\dot\gamma)}\, d\tau\right| \leqq \tfrac{1}{2}\alpha_3\, l(s)$$

such that in particular

$$\left|\frac{d}{ds}r_s\right| \leqq \tfrac{1}{2}\alpha_3\, r_s.$$

We therefore conclude that

$$\frac{d}{ds}h \leqq -\theta^{-1}|\nabla h|^2 + \theta^{-1}\alpha_3|\nabla h|^2\cdot(2\eta - s).$$

Using then $\eta \leqq \frac{1}{2\alpha_3}$ we obtain for $\theta = \frac{1}{4}$

$$\frac{d}{ds}h \leqq -2|\nabla h|^2. \tag{29}$$

Proceeding similarly as in [16] we obtain for $f_K = \max\{\min(f, K), 0\}$ and some smooth time independent compactly supported function φ on M^n that for $0<\varepsilon<\eta$

$$\begin{aligned}&\int_\varepsilon^\eta e^{-\beta s}\left(\int_M \varphi^2 e^h f_K\left(\Delta f - \frac{\partial f}{\partial s}\right)d\mu_s\right)ds \\ &\quad\geqq -\alpha_1\int_\varepsilon^\eta e^{-\beta s}\left(\int_M \varphi^2 e^h f_K|\nabla f|\, d\mu_s\right)ds - \alpha_0\int_\varepsilon^\eta e^{-\beta s}\left(\int_M \varphi^2 e^h f_K f\, d\mu_s\right)ds\end{aligned}$$

where $\beta > 0$ will be chosen later and where we used that $f_K f \geqq 0$. Performing an integration by parts yields

$$
\begin{aligned}
0 \leqq & -\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K \langle \nabla f_K, \nabla f\rangle \, d\mu_s\Big) ds \\
& -\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K \langle \nabla h, \nabla f\rangle \, d\mu_s\Big) ds \\
& -2\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi e^h f_K \langle \nabla h, \nabla f\rangle \, d\mu_s\Big) ds \\
& -\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K \frac{\partial f}{\partial s} \, d\mu_s\Big) ds + \alpha_0 \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K f \, d\mu_s\Big) ds \\
& +\alpha_1 \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K |\nabla f| \, d\mu_s\Big) ds \\
= & \, \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + \mathrm{V} + \mathrm{VI}.
\end{aligned}
$$

We compute

$$
\mathrm{I} = -\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h |\nabla f_K|^2 \, d\mu\Big) ds.
$$

Using Schwarz' inequality we obtain

$$
\mathrm{II} \leqq \tfrac{1}{4}\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h |\nabla f|^2 \, d\mu_s\Big) ds + \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h |\nabla h|^2 f_K^2 \, d\mu_s\Big) ds,
$$

$$
\mathrm{III} \leqq \tfrac{1}{2}\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h |\nabla f|^2 \, d\mu_s\Big) ds + 2\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M e^h |\nabla h|^2 f_K^2 \, d\mu_s\Big) ds.
$$

To estimate the terms IV + V observe that as in [16]

$$
-e^h f_K \frac{\partial f}{\partial s} \leqq -e^h f_K \frac{\partial f_K}{\partial s} + \frac{\partial}{\partial s}\{e^h f_K (f_K - f)\},
$$

whenever $\dfrac{\partial f_K}{\partial s}$ exists. Since f_K is uniformly Lipschitz continuous on compact subsets of $M^n \times [0, T]$ we derive

$$
\begin{aligned}
\mathrm{IV} + \mathrm{V} \leqq & -\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K \frac{\partial f_K}{\partial s} \, d\mu_s\Big) ds + \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 \frac{\partial}{\partial s}\{e^h f_K (f_K - f)\} \, d\mu_s\Big) ds \\
& +\alpha_0 \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K f \, d\mu_s\Big) ds.
\end{aligned}
$$

Now observe that

$$\left|\frac{d}{ds}(d\mu_s)\right| \leqq n\alpha_3 \cdot d\mu_s$$

in view of condition (iv).

Therefore,

$$\begin{aligned} \mathrm{IV} \leqq -\frac{1}{2}\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h \frac{\partial}{\partial s}(f_K^2)\, d\mu_s\Big)\, ds + \int_\varepsilon^\eta \frac{\partial}{\partial s}\Big[e^{-\beta s}\int_M \varphi^2 e^h f_K (f_K - f)\, d\mu_s\Big]\, ds \\ + \beta\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K (f_K - f)\, d\mu_s\Big)\, ds - n\alpha_3 \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K (f_K - f)\, d\mu_s\Big)\, ds \\ + \alpha_0 \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K f\, d\mu_s\Big)\, ds. \end{aligned}$$

Choosing now $\beta \geqq 2n\alpha_3 + 4\alpha_0$ and observing that $f_K(f_K - f) \leqq 0$ we derive

$$\begin{aligned} \mathrm{IV} = -\tfrac{1}{2} e^{-\beta s}\int_M \varphi^2 e^h f_k^2\, d\mu_s|_{s=\eta} + \tfrac{1}{2} e^{-\beta s}\int_M \varphi^2 e^h f_k^2\, d\mu_s|_{s=\varepsilon} \\ + \tfrac{1}{2}\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K^2 \frac{\partial h}{\partial s}\, d\mu_s\Big)\, ds - \tfrac{1}{4}\beta\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K^2\, d\mu_s\Big)\, ds \\ + e^{-\beta s}\int_M \varphi^2 e^h f_K (f_K - f)\, d\mu_s|_{s=\eta} - e^{-\beta s}\int_M \varphi^2 e^h f_K (f_K - f)\, d\mu_s|_{s=\varepsilon}. \end{aligned}$$

The integral VI can be estimated by

$$\mathrm{VI} \leqq \tfrac{1}{4}\int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h |\nabla f|^2\, d\mu_s\Big)\, ds + \alpha_1^2 \int_\varepsilon^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h f_K^2\, d\mu_s\Big)\, ds.$$

Combining the estimates for I–VI, using (29) and choosing $\beta \geqq \max(2n\alpha_3 + 4\alpha_0, \alpha_1^2)$ we conclude for $\varepsilon \to 0$ that

$$\begin{aligned} 0 \leqq -\int_0^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h |\nabla f_K|^2\, d\mu_s\Big)\, ds + \int_0^\eta e^{-\beta s}\Big(\int_M \varphi^2 e^h |\nabla f|^2\, d\mu_s\Big)\, ds \\ + 2\int_0^\eta e^{-\beta s}\Big(\int_M e^h |\nabla \varphi|^2 f_K^2\, d\mu_s\Big)\, ds - \tfrac{1}{2} e^{-\beta s}\int_M \varphi^2 e^h f_K^2\, d\mu_s|_{s=\eta}. \end{aligned}$$

Here we also used that $f_K(f_K - f) \leqq 0$ and $f_K \equiv 0$ at $s = 0$. For $R > 0$ let $0 \leqq \varphi \leqq 1$ satisfy $\varphi \equiv 1$ on $B_R^0(p)$, $\varphi \equiv 0$ outside $B_{R+1}^0(p)$ and $|\nabla \varphi|^0 \leqq 2$. We then conclude in view of (28)

$$\begin{aligned} \tfrac{1}{2}\int_{B_R^0(p)} e^h f_K^2\, d\mu_s|_{s=\eta} \cdot e^{-\beta\eta} \leqq \int_0^\eta e^{-\beta s}\Big(\int_{B_{R+1}^0(p)} e^h (|\nabla f|^2 - |\nabla f_K|^2)\, d\mu_s\Big)\, ds \\ + 2\alpha_4 \int_0^\eta e^{-\beta s}\Big(\int_{B_{R+1}^0(p)\setminus B_R^0(p)} e^h f_K^2\, d\mu_s\Big)\, ds. \end{aligned}$$

Since $0<\eta<\min\left(\frac{1}{64K}, \frac{1}{32\alpha_2}\right)$ we infer $h(y,s)\leqq -2kr_s^2(p,y)$ and $h(y,s)\leqq -\alpha_2 r_s^2(p,y)$ for all $0<s<\eta$. Since $f_K^2\leqq K^2$ we obtain from the volume growth assumptions on M^n that the second term on the RHS tends to zero as $R\to\infty$. Finally, since $0\leqq|\nabla f|^2-|\nabla f_K|^2\leqq|\nabla f|^2$ we derive for $R\to\infty$

$$\tfrac{1}{2}\int_M e^h f_K^2\, d\mu_s|_{s=\eta}\cdot e^{-\beta\eta}\leqq\int_0^\eta e^{-\beta s}\Big(\int_M e^h(|\nabla f|^2-|\nabla f_K|^2)\, d\mu_s\Big)\, ds$$

in view of assumption (iii).

Now let $K\to\infty$. Then $f_K^2\to(\max(f,0))^2$ and for all s, $|\nabla f_K|^2\to|\nabla f|^2$. Using again assumption (iii) we finally conclude

$$\tfrac{1}{2}\int_M e^h(\max(f,0))^2\, d\mu_s|_{s=\eta}\leqq 0,$$

where $0<\eta<\min\left(T, \frac{1}{64K}, \frac{1}{32\alpha_2}\right)$. By means of an inductive argument we arrive at $f\leqq 0$ everywhere in $M^n\times[0,T]$.

4.4 Corollary. *Let $\mathbf{F}_0=M^n\to R^{n+1}$ be a smooth isometric immersion of a complete hypersurface with bounded $C^{2,\alpha}$-norm. Suppose $M_0=\mathbf{F}_0(M^n)$ has nonnegative mean curvature. Then the smooth solution of* (1), *$M_t=\mathbf{F}_t(M^n)$, has nonnegative mean curvature for $t\in[0,T]$ where T depends on n and the initial curvature bound.*

Proof. Since M_0 has bounded curvature, $\sup_{M_0}|A|\leqq c_0$, it satisfies a uniform local Lipschitz condition as in Sect. 4. Using the gradient estimates in Theorem 2.1 we infer that v stays bounded on some small time interval $[0,T]$ depending on c_0. By a simple modification of the arguments in the proof of Theorem 3.1 we can bound $\sup_{M_t}|A|$ on $[0,T]$ in terms of c_0 again. Since $\mathrm{Ric}_{M_t}\geqq -2|A|^2$ the uniform volume growth condition of Theorem 4.3 holds for $t\in[0,T]$ as we have the uniform bound on $|A|$ on this time interval. In view of the $C^{2,\alpha}$ initial data and parabolic regularity theory we obtain the estimate

$$\sup_{M_t} t^{1-\alpha}|\nabla H|^2\leqq c$$

where c depends on n, T, c_0 and the $C^{2,\alpha}$-norm of M_0.

This and the volume growth condition then imply the validity of condition (iii) in Theorem 4.3. Moreover (iv) holds in view of the equation

$$\frac{d}{dt}g_{ij}=-2Hh_{ij}$$

derived in [8].

In view of Lemma 1.1 (iv) and the curvature bound the mean curvature H satisfies the conditions of Theorem 4.3 with $b=|A|^2$ for $t\in[0, T]$. This implies the result.

4.5 Remark. In a forthcoming paper Theorem 4.3 will be applied to spacelike slices moving by mean curvature in asymptotically flat spacetimes.

5 Flow for entire graphs of arbitrary growth

In [4] the authors established longtime existence for the initial value problem corresponding to (1) in the class of linearly growing graphs. In this section we use the interior estimates of Sect. 3 in conjunction with a spherical barrier argument to show that (1) admits a smooth solution for all $t>0$ in the class of entire graphs. In particular, no smoothness or growth assumptions have to imposed on the initial graph.

5.1 Theorem. *Let $M_0\equiv\mathbf{F}_0(R^n)$ be a locally Lipschitz continuous entire graph over R^n. Then the initial value problem* (1) *has a smooth solution $M_t=\mathbf{F}_t(R^n)$ for all $t>0$. Moreover, each M_t is an entire graph over R^n.*

5.2 Remark. (i) In view of the equivalence of solutions of (1) and (2) discussed in Sect. 1. Theorem 5.1 also establishes the existence of a smooth solution of the initial value problem

$$\frac{d}{dt}w=\sqrt{1+|Dw|^2}\,\operatorname{div}\left(\frac{Dw}{\sqrt{1+|Dw|^2}}\right) \quad \text{in } R^n\times(0,\infty)$$
$$w(\cdot,0))=w_0 \quad \text{in } R^n$$

assuming that w_0: $R^n\to R$ is merely locally Lipschitz continuous. Moreover, the special form of the curvature estimate in Corollary 3.2 (ii) in conjunction with the gradient estimate in Theorem 2.3 yields that at $t=0$ the solution w is Hölder continuous in time with Hölder exponent $\alpha=\frac{1}{2}$.

(ii) Note that in contrast to the above result a solution of the ordinary heat equation becomes unbounded in finite time, unless certain growth conditions for $|y|\to\infty$ are imposed on the initial data.

Proof of Theorem 5.1. Assume first $M_0=\mathbf{F}_0(R^n)=$ graph w_0 for $w_0\in C^{2,\alpha}(R^n)$, $\alpha>0$.

Using Theorem 2.1 of [12] as well as the tangential diffeomorphisms defined in Sect. 1 we infer that for any $R>0$ the boundary value problem

$$\frac{d}{dt}\mathbf{F}(y,t)=\mathbf{H}(y,t), \qquad (y,t)\in B_R(0)\times(0,\infty)$$
$$\mathbf{F}(y,0)=\mathbf{F}(y), \qquad y\in B_R(0)$$
$$\mathbf{F}(y,t)=\mathbf{F}(y), \qquad (y,t)\in\partial B_R(0)\times[0,\infty)$$

admits a smooth solution $M_t^R=\mathbf{F}_t^R(\bar{B}_R(0))$ which for each $t>0$ can be written as a graph over $\bar{B}_R(0)\subset R^n$.

Let now $T>0$ be fixed and $\Omega \subset R^n$ be compact. Choose $R_0>0$ s.t. $\Omega \subset B_{R_0}(0)$ and define $R_1=\sqrt{4R_0^2+2nT}$.

We also define $a^+=(0, \sup_{B_{R_1}(0)} w_0+R_1+1)\in R^{n+1}$. Then we have for $R>R_1$

$$M_0^R \cap B_{R_1}^{n+1}(a^+)=\emptyset$$

where $B_{R_1}^{n+1}$ denotes a ball in R^{n+1}. Furthermore

$$\partial M_t^R \cap B_{R_1}^{n+1}(a^+)=\emptyset$$

for $t\geqq 0$. Using Theorem 3.7 of [1], or Lemma 3.2 of [11], we infer

$$M_t^R \cap B_{2R_0}^{n+1}(a^+)=\emptyset$$

for $t\in[0, T]$.

Using a similar argument for $B_{R_1}^{n+1}(a^-)$, where $a^-=(0, \inf_{B_{R_1}(0)} w_0-R_1-1)$ $\in R^{n+1}$ we conclude that the height function $w_R(\cdot, t)$ of M_t^R satisfies

$$\sup_{B_{2R_0}\times[0, T]} |w_R|\leqq c_0$$

where $c_0=c_0(n, R_0, T, \sup_{B_{R_1}(0)} w_0)$ is independent of R.

Applying the gradient estimate of Theorem 2.3 we obtain

$$\sup_{B_{3/2R_0}(0)\times[0, T]} |Dw_R|\leqq c_1$$

where $c_1=c_1(n, R_0, c_0, \sup_{B_{2R_0}(0)} |Dw_0|)$.

From Corollary 3.5 (ii) we then conclude for any integer $m\geqq 0$

$$\sup_{B_{R_0}(0)\times[0, T]} |D^m w_R|\leqq c_m$$

where $c_m=c_m(m, n, R_0, c_0, c_1)$.

We can therefore select a sequence of solutions (w_{R_k}) for $R_k\to\infty$ $(R_k>R_1$ for any $k\geqq 2)$ s.t. $w_{R_k}\to w$ in C^∞ uniformly on $\Omega\times[0, T]$. Since Ω and $T>0$ were arbitrary this establishes the existence of a family of entire graphs M_t $=\text{graph}\, w(\cdot, t)$ solving (1) where $w\in C^\infty(R^n\times(0, \infty))$. As the second and higher order derivative estimates for w on each compact subset of R^n depend only on the initial height and gradient on a slightly larger subset, an approximation argument yields a smooth solution of (1) also for locally Lipschitz initial data.

References

1. Brakke, K.A.: The Motion of a Surface by its Mean Curvature. (Math. Notes) Princeton, NJ: Princeton University Press 1978
2. Caffarelli, L., Nirenberg, L., Spruck, J.: On a form of Bernstein's Theorem. (Preprint)
3. Ecker, K.: Estimates for evolutionary surfaces of prescribed mean curvature. Math. Z. **180**, 179–192 (1982)

4. Ecker, K., Huisken, G.: Mean curvature evolution of entire graphs. Ann. Math. **130**, 453–471 (1989)
5. Ecker, K., Huisken, G.: A Bernstein result for minimal graphs of controlled growth. J. Differ. Geom. **31**, 397–400 (1990)
6. Gerhardt, C.: Evolutionary surfaces of prescribed mean curvature. J Differ. Equations **36**, 139–172 (1980)
7. Grayson, M.: The heat equation shrinks embedded plane curves to round points. J. Differ. Geom. **26**, 285–314 (1987)
8. Hamilton, R.S.: Three-manifolds with positive Ricci curvature. J. Differ. Geom. **17**, 255–306 (1982)
9. Hamilton, R.S.: Heat Equations in Geometry. Honolulu, Hawaii (Lecture Notes) 1989
10. Huisken, G.: Flow by mean curvature of convex surfaces into spheres. J. Differ. Geom. **20**, 237–266 (1984)
11. Huisken, G.: Contracting convex hypersurfaces in Riemannian manifolds by their mean curvature. Invent. Math. **84**, 463–480 (1986)
12. Huisken, G.: Non-parametric mean curvature evolution with boundary conditions. J. Differ. Equations **77**, 369–378 (1989)
13. Huisken, G.: Asymptotic behaviour for singularities of the mean curvature flow. J. Differ. Geom. **31**, 285–299 (1990)
14. Korevaar, N.: An easy proof of the interior gradient bound for solutions to the prescribed mean curvature equation. Proc. Symp. Pure Math. **45**, Part 2, 81–89 (1986)
15. Ladyzhenskaya, O.A., Solonnikov, V.A., Uraltseva, N.N.: Linear and Quasilinear Equations of Parabolic Type. Transl., II. Ser., Am. Math. Soc. **23** (1968)
16. Liao, G.G., Tam, L.-F.: On the heat equation for harmonic maps from non-compact manifolds. (Preprint)
17. Shi, W.-X.: Deforming the metric on complete Riemannian manifolds. J. Differ. Geom. **30**, 223–301 (1989)
18. Angenent, S.: Parabolic equations for curves on surfaces (I). Curves with p-integrale curvature, preprint Univ. of Wisconsin, 1988

Invent. math. 105, 571–588 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

A Rankin-Selberg integral for the adjoint representation of GL_3

David Ginzburg

Yale University, Department of Mathematics, New Haven, CT 06520, USA

Oblatum 3-VII-1990

Introduction

In this paper we construct steps 1, 4 and part of 2 as formulated at [G-Sh, p. 19] for the function $L(s, \pi, r)$. Here π is a cusp form on GL_3 and r is the eight dimensional adjoint representation of GL_3, the L-group of GL_3. To do so we construct a Rankin-Selberg integral, using the embedding of SL_3 in G_2. Let us mention that the Langland's conjucture for this L-function can not be obtained by the Langlands-Shahidi method.

The paper is organized as follows. In the first section we collect some information on the structure of G_2 and compute the space of double cosets $P\backslash G_2/SL_3$, where P is a maximal parabolic subgroup of G_2. In Sect. two, the global integral is constructed and is shown to be Eulerian with Whittaker model. Finally, in Sect. three, we carry out the unramified computation using the local computation of the integral for $GL_3 \times GL_3$ as constructed in [J-S]. For simplicity we shall assume that π has a trivial central character.

Acknowledgements. I am very grateful to I. Piatetski-Shapiro and S. Rallis for helpful conversations. A special thanks to D. Bump for making useful suggestions for the unramified computation.

1 On the structure of G_2

Since we use the group G_2 in our integral, we collect some data about this group and the group SL_3 embedded inside it. For matrix representations and some other relations among the matrices of G_2 see [Hu] or [Ch].

G_2 has six positive roots: α, β, $\alpha+\beta$, $2\alpha+\beta$, $3\alpha+\beta$, $3\alpha+2\beta$. We will use the notation $x_\delta(t)$ to denote the one parameter unipotent subgroup generated by the root x_δ. SL_3 is embedded in G_2 as $SL_3 \simeq \langle \pm\beta, \pm(3\alpha+\beta), \pm(3\alpha+2\beta)\rangle$. We shall denote by $h(z_1, z_2, z_3)$, with $z_1 z_2 z_3 = 1$, the torus of G_2, which is also embedded in SL_3 in the obvious way. We also denote

$$N_2 = \begin{pmatrix} 1 & -x & z \\ & 1 & x \\ & & 1 \end{pmatrix}$$

Denote by w_α and w_β, the two simple reflections of the Weyl group of G_2. We have:

$$W_{G_2}=\{e, w_\alpha, w_\beta, w_\alpha w_\beta, w_\beta w_\alpha, w_\alpha w_\beta w_\alpha, w_\beta w_\alpha w_\beta, w_\alpha w_\beta w_\alpha w_\beta, w_\beta w_\alpha w_\beta w_\alpha,$$
$$w_\alpha w_\beta w_\alpha w_\beta w_\alpha, w_\beta w_\alpha w_\beta w_\alpha w_\beta, w_\alpha w_\beta w_\alpha w_\beta w_\alpha w_\beta\}$$

and

$$W_{SL_3}=\{e, w_\beta, w_\alpha w_\beta w_\alpha, w_\beta w_\alpha w_\beta w_\alpha, w_\alpha w_\beta w_\alpha w_\beta, w_\beta w_\alpha w_\beta w_\alpha w_\beta\}.$$

Let P be the parabolic subgroup of G_2 with Levi part GL_2, constructed from the root α. Then $P=GL_2\cdot U$ where $U=\langle x_\beta, x_{\alpha+\beta}, x_{2\alpha+\beta}, x_{3\alpha+\beta}, x_{3\alpha+2\beta}\rangle$. We denote by U_{G_2} the maximal unipotent subgroup of G_2.

For further calculations we shall collect some commutation relations of matrices in G_2. We use the notation $[g_1, g_2]=g_1^{-1} g_2^{-1} g_1 g_2$.

(1.1) $$w_\beta x_{\alpha+\beta}(1)=x_\alpha(1) w_\beta$$

(1.2) $$[x_\beta(s), x_{-(\alpha+\beta)}(t)]=x_{-\alpha}(ts) x_{-(2\alpha+\beta)}(-t^2 s) x_{-(3\alpha+\beta)}(t^3 s^2) x_{-(3\alpha+2\beta)}(-t^3 s)$$

(1.3) $$[x_{-(\alpha+\beta)}(r), x_{-\alpha}(t)]=x_{-(2\alpha+\beta)}(2rt) x_{-(3\alpha+\beta)}(3r^2 t) x_{-(3\alpha+2\beta)}(3rt^2)$$

(1.4) $$w_\alpha w_\beta w_\alpha x_\alpha(1)=x_{-(2\alpha+\beta)}(1) w_\alpha w_\beta w_\alpha$$

(1.5) $$w_\beta w_\alpha w_\beta w_\alpha x_\alpha(1)=x_{-(2\alpha+\beta)}(1) w_\beta w_\alpha w_\beta w_\alpha$$

(1.6) $$w_\beta w_\alpha w_\beta w_\alpha x_{2\alpha+\beta}(1)=x_{-(\alpha+\beta)}(1) w_\beta w_\alpha w_\beta w_\alpha$$

(1.7) $$w_\beta h(z_1, z_2, z_3)=h(z_2, z_1, z_3) w_\beta$$

(1.8) $$x_{-\beta}(r) x_{\alpha+\beta}(t)=u x_{-\beta}(r) \quad \text{where} \quad u\in U_{G_2}$$

For all other pairs of unipotent matrices g_1, g_2 used in this paper, we have $[g_1, g_2]=1$.

For our Rankin-Selberg integral, we shall need the following:

Lemma. $|P\backslash G_2/SL_3|=2$. *As representatives, one can choose* e *and* $x_{-(\alpha+\beta)}(1) w_\beta$.

Proof. Denote by B_3 the Borel subgroup of SL_3. Using the Bruhat decomposition in G_2 and the fact that $x_\alpha\in P$ we conclude that a set of representatives for $P\backslash G_2/B_3$ is: e, w_β, $w_\beta w_\alpha$, $w_\beta w_\alpha x_\alpha(1)$, $w_\beta w_\alpha w_\beta$, $w_\beta w_\alpha w_\beta x_{\alpha+\beta}(1)$, $w_\beta w_\alpha w_\beta w_\alpha$, $w_\beta w_\alpha w_\beta w_\alpha x_\alpha(1)$, $w_\beta w_\alpha w_\beta w_\alpha x_{2\alpha+\beta}(1)$, $w_\beta w_\alpha w_\beta w_\alpha x_\alpha(1) x_{2\alpha+\beta}(1)$, $w_\beta w_\alpha w_\beta w_\alpha w_\beta$, $w_\beta w_\alpha w_\beta w_\alpha w_\beta x_{\alpha+\beta}(1)$, $w_\beta w_\alpha w_\beta w_\alpha w_\beta x_{2\alpha+\beta}(1)$, $w_\beta w_\alpha w_\beta w_\alpha w_\beta x_{\alpha+\beta}(1) x_{2\alpha+\beta}(1)$.

Write $g_1 \sim g_2$ if g_1 and g_2 are in the same double coset for $P\backslash G_2/SL_3$. Then:

$$w_\beta \sim e;\ w_\beta w_\alpha \sim w_\alpha w_\beta w_\alpha \sim e;\ w_\beta w_\alpha w_\beta \sim w_\beta w_\alpha \sim e;$$
$$w_\beta w_\alpha w_\beta x_{(\alpha+\beta)}(1) \underset{(1.1)}{=} w_\beta w_\alpha x_\alpha(1) w_\beta \sim w_\beta w_\alpha x_\alpha(1)$$
$$\sim w_\alpha w_\beta w_\alpha x_\alpha(1) \underset{(1.4)}{=} x_{-(2\alpha+\beta)}(1) w_\alpha w_\beta w_\alpha \sim x_{-(2\alpha+\beta)}(1);$$
$$w_\beta w_\alpha w_\beta w_\alpha \sim e;\ w_\beta w_\alpha w_\beta w_\alpha x_\alpha(1) \underset{(1.5)}{=} x_{-(2\alpha+\beta)}(1) w_\beta w_\alpha w_\beta w_\alpha \sim x_{-(2\alpha+\beta)}(1);$$
$$w_\beta w_\alpha w_\beta w_\alpha x_{2\alpha+\beta}(1) \underset{(1.6)}{=} x_{-(\alpha+\beta)}(1) w_\beta w_\alpha w_\beta w_\alpha \sim x_{-(\alpha+\beta)}(1);$$
$$w_\beta w_\alpha w_\beta w_\alpha x_\alpha(1) x_{2\alpha+\beta}(1) \underset{(1.5)(1.6)}{=} x_{-(\alpha+\beta)}(1) x_{-(2\alpha+\beta)}(1)$$
$$w_\beta w_\alpha w_\beta w_\alpha \sim x_{-(\alpha+\beta)}(1) x_{-(2\alpha+\beta)}(1);\ w_\beta w_\alpha w_\beta w_\alpha w_\beta \sim w_\beta w_\alpha w_\beta w_\alpha w_\beta;$$
$$w_\beta w_\alpha w_\beta w_\alpha w_\beta x_{(\alpha+\beta)}(1) \sim w_\beta w_\alpha w_\beta w_\alpha x_\alpha(1);\ w_\beta w_\alpha w_\beta w_\alpha w_\beta x_{2\alpha+\beta}(1)$$
$$\sim w_\beta w_\alpha w_\beta w_\alpha x_{2\alpha+\beta}(1);$$
$$w_\beta w_\alpha w_\beta w_\alpha w_\beta x_{\alpha+\beta}(1) x_{2\alpha+\beta}(1) \sim w_\beta w_\alpha w_\beta w_\alpha x_\alpha(1) x_{2\alpha+\beta}(1).$$

We have:

$$(*) \qquad x_{-(\alpha+\beta)}(1) \sim x_{-(\alpha+\beta)}(1)\ x_{-(2\alpha+\beta)}(1) \sim x_{-(2\alpha+\beta)}(1)$$

Indeed:

$$x_{-(\alpha+\beta)}(1) \sim x_\beta(1) x_{-(\alpha+\beta)}(1) \underset{(1.2)}{=}$$
$$= x_{-(\alpha+\beta)}(1) x_\beta(1) x_{-\alpha}(1) x_{-(2\alpha+\beta)}(-1) x_{-(3\alpha+\beta)}(t_1) x_{-(3\alpha+2\beta)}(t_2)$$
$$\sim x_{-(\alpha+\beta)}(1) x_{-\alpha}(1) x_\beta(1) x_{-(2\alpha+\beta)}(-1) \underset{(1.3)}{=}$$
$$= x_{-\alpha}(1) x_{-(\alpha+\beta)}(1) x_{-(2\alpha+\beta)}(2) x_{-(3\alpha+\beta)}(r_1) x_{-(3\alpha+2\beta)}(r_2) x_{-(2\alpha+\beta)}(-1) x_\beta(1)$$
$$\sim x_{-(\alpha+\beta)}(1) x_{-(2\alpha+\beta)}(1).$$

In the same way we prove the right relation in (*).

Since $x_{-(\alpha+\beta)}(1) \sim x_{-(\alpha+\beta)}(1) w_\beta$ the lemma follows.

2 The global integral

Let F be a global field and $\mathbf{A}$ its ring of adeles. Let φ be a cusp form on GL_3. We know that φ is generic, i.e. if

$$W(g) = \int_{(F\backslash\mathbf{A})^3} \varphi\left[\begin{pmatrix} 1 & y & z \\ & 1 & x \\ & & 1 \end{pmatrix} g\right] \psi(x+y)\, dx\ dy\ dz \qquad g \in GL_3(\mathbf{A})$$

then $W(g) \not\equiv 0$.

Let δ_P be the modular character of P. Then:

$$\delta_P(h(z_1, z_2, z_3)) = |z_1 z_3^{-1}|^3.$$

If $f \in \mathrm{Ind}_{P_A}^{G_2(\mathbf{A})} \delta_P^s$, where s is a complex number, then we define the Eisenstein series on $G_2(\mathbf{A})$, as follows:

$$E^f(g, s) = \zeta(3s)\,\zeta(6s-2)\,\zeta(9s-3) \sum_{\gamma \in P_F \backslash G_2(F)} f(\gamma g, s) \qquad g \in G_2(\mathbf{A}).$$

Here $\zeta(s)$ is the Dedekind zeta function of F.

This Eisenstein series was studied by I. Piatetski-Shapiro and S. Rallis.

From the general theory of Eisenstein series, we know that $E^f(g, s)$ converges absolutely in some right half plane $\mathrm{Re}(s) \geqq s_0$, has a meromorphic continuation to the whole complex plane with finite number of poles and satisfies a functional equation of the form $E^f(g, s) = E^{M(s)f}(g, 1-s)$. Here $M(s)$ is a suitable, normalized intertwining operator.

We define our global integral:

$$I(\varphi, f, s) = \int_{\mathrm{SL}_3(F)\backslash \mathrm{SL}_3(\mathbf{A})} \varphi(g)\, E^f(g, s)\, dg.$$

Our main theorem is:

Theorem 1 (a) *$I(\varphi, f, s)$ converges for all s, except possibly for a finite number of values of s where the Eisenstein series has poles, and satisfies a functional equation*

$$I(\varphi, f, s) = I(\varphi, M(s)f, 1-s).$$

(b) *If φ is a cusp form such that $W(g) = 0$ then*

$$I(\varphi, f, s) = \zeta(3s)\,\zeta(6s-2)\,\zeta(9s-3) \int_{N_2(\mathbf{A})\backslash \mathrm{SL}_3(\mathbf{A})} W(g) f(\gamma g, s)\, dg$$

where $\gamma = x_{-(\alpha+\beta)}(1)\, w_\beta$.

Proof. (a) follows from the properties of cusp forms and Eisenstein series.

(b) Unfolding the Eisenstein series we get:

$$I(\varphi, f, s) = \zeta(3s)\,\zeta(6s-2)\,\zeta(9s-3) \int_{\mathrm{SL}_3(F)\backslash \mathrm{SL}_3(\mathbf{A})} \varphi(g) \sum_{\delta \in P_F \backslash G_2(F)} f(\delta g, s)\, dg$$

Changing the order of integration and summation gives us:

$$I(\varphi, f, s) = \zeta(3s)\,\zeta(6s-2)\,\zeta(9s-3) \sum_{\gamma \in P_F \backslash G_2(F)/\mathrm{SL}_3(F)} \int_{\mathrm{SL}_3^\gamma(F)\backslash \mathrm{SL}_3(\mathbf{A})} \varphi(g) f(\gamma g, s)\, dg$$

where $\mathrm{SL}_3^\gamma = \mathrm{SL}_3 \cap \gamma P \gamma^{-1}$.

By the lemma in Sect. 1, we know that there are two representatives for $P \backslash G_2 / \mathrm{SL}_3$: e and $x_{-(\alpha+\beta)}(1)\, w_\beta$.

It is clear that $SL_3^e = SL_3 \cap P \supset N$, where N is the maximal unipotent subgroup of SL_3; hence by cuspidality of φ, the integral for $\gamma = e$ vanishes. As for $\gamma = x_{-(\alpha+\beta)}(1)\, w_\beta$, a simple matrix multiplication shows that

$$SL_3^\gamma = SL_3 \cap \gamma P \gamma^{-1} = N_2 \cdot GL_1 = \begin{pmatrix} 1 & -x & z \\ & 1 & x \\ & & 1 \end{pmatrix} \begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1} \end{pmatrix}.$$

Hence:

$$I(\varphi, f, s) = \zeta(3s)\,\zeta(6s-2)\,\zeta u(9s-3) \int\limits_{N_2(F)GL_1(F)\backslash SL_3(\mathbf{A})} \varphi(g) f(\gamma g, s)\, dg$$

$$= \zeta(3s)\,\zeta(6s-2)\,\zeta(9s-3) \int\limits_{N_2(\mathbf{A})GL_1(F)\backslash SL_3(\mathbf{A})}$$

$$\cdot \int\limits_{N_2(F)\backslash N_2(\mathbf{A})} \varphi\left[\begin{pmatrix} 1 & -x & z \\ & 1 & x \\ & & 1 \end{pmatrix} g\right] dx\, dz\, f(\gamma g, s)\, dg.$$

It is not hard to obtain the following expansion:

$$\sum_{a \in F^*} W\left[\begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1} \end{pmatrix} g\right] = \int\limits_{N_2(F)\backslash N_2(\mathbf{A})} \varphi\left[\begin{pmatrix} 1 & -x & z \\ & 1 & x \\ & & 1 \end{pmatrix} g\right] dx\, dz.$$

Indeed, by definition

$$W(g) = \int\limits_{(F\backslash A)^3} \varphi\left[\begin{pmatrix} 1 & y & z \\ & 1 & x \\ & & 1 \end{pmatrix} g\right] \psi(x+y)\, dx\, dy\, dz.$$

Replacing g by $\begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1} \end{pmatrix} g$ for $a \in F^*$ and summing over all $a \in F^*$ gives,

$$\sum_{a \in F^*} W\left[\begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1} \end{pmatrix} g\right]$$

$$= \sum_{a \in F^*} \int\limits_{(F\backslash A)^3} \varphi\left[\begin{pmatrix} 1 & y & z \\ & 1 & x \\ & & 1 \end{pmatrix} \begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1} \end{pmatrix} g\right] \psi(x+y)\, dx\, dy\, dz.$$

Conjugating $\begin{pmatrix} 1 & y & z \\ & 1 & z \\ & & 1 \end{pmatrix}$ by $\begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1} \end{pmatrix}$, and using the fact that φ is invariant

under $GL_3(F)$ we get,

$$\sum_{s\in F^*} W\left[\begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1}\end{pmatrix} g\right]$$
$$= \sum_{a\in F^*} \int_{(F\backslash A)^3} \varphi\left[\begin{pmatrix} 1 & a^{-1}y & a^{-2}z \\ & 1 & a^{-1}x \\ & & 1\end{pmatrix} g\right] \psi(x+y)\, dx\, dy\, dz.$$

Next we change variables $a^{-1}y \to y$ and, $a^{-1}x \to x$ and $a^{-2}z \to z$; hence

$$\sum_{a\in F^*} W\left[\begin{pmatrix} a & & \\ & 1 & \\ & & a^{-1}\end{pmatrix} g\right]$$
$$= \sum_{a\in F^*} \int_{(F\backslash A)^3} \varphi\left[\begin{pmatrix} 1 & x & z \\ & 1 & y \\ & & 1\end{pmatrix} g\right] \psi(a(x+y))\, dx\, dy\, dz.$$

By cuspidality, we may sum over all $a\in F$ in the R.H.S. and using the Fourier expansion formula we get the desired identity.

Plugging this into the integral gives (b).

3 The unramified computations

In this section we carry out the unramified computations. Let F be a local field and assume that π is unramified. For simplicity, let us assume that π has a trivial central character. Denote $\zeta(s)=(1-q^{-s})^{-1}$, where $q^{-1}=|p|$ and p is a generator of the maximal ideal in the ring of integers of F. We prove the following theorem:

Theorem 2 *If $W(g)$ is the class one Whittaker function for the local unramified representation π, and $f(g, s)$ is the class one function in $\mathrm{Ind}_{P^2}^{G_2} \delta_P^s$ then*

$$\int_{N_2(F)\backslash SL_3(F)} W(g) f(\gamma g, s)\, dg = \frac{L(3s-1, \pi, \mathrm{Ad})}{\zeta(3s)\zeta(6s-2)\zeta(9s-3)} \tag{3.1}$$

Proof. Completing N_2 to N, the maximal unipotent subgroup of SL_3, we get that the L.H.S. of (3.1) equals

$$\int_{N(F)\backslash SL_3(F)} W(g) \int_F f(\gamma x_\beta(r) g, s) \psi(r)\, dr\, dg.$$

Using the Iwasawa decomposition for SL_3 we obtain

$$\int_{(F^*)^2} W(a, b) \int_F f(\gamma x_\beta(r) h(ab, b, a^{-1}b^{-2}), s) \psi(r)\, dr\, |a^4 b^6|^{-1} d^*a\, d^*b. \tag{3.2}$$

Here $W(a, b)=W(h(ab, b, a^{-1}b^{-2}))$. Now we use the definition of γ and some commutation relations in G_2 to obtain:

$$\begin{aligned}\gamma x_\beta(r)\, h(ab, b, a^{-1}b^{-2}) &= x_{-(\alpha+\beta)}(1)\, w_\beta\, x_\beta(r)\, h(ab, b, a^{-1}b^{-2}) \\ &= x_{-(\alpha+\beta)}(1)\, x_{-\beta}(r)\, w_\beta\, h(ab, b, a^{-1}b^{-2}) \\ &\underset{(1.7)}{=} x_{-(\alpha+\beta)}(1)\, x_{-\beta}(1)\, h(b, ab, a^{-1}b^{-2})\, w_\beta \\ &= h(b, ab, a^{-1}b^{-2})\, x_{-(\alpha+\beta)}(b)\, x_{-\beta}(a^{-1}r)\, w_\beta.\end{aligned}$$

Using this and the invariant properties of $f(g, s)$, (3.2) equals

$$\begin{aligned}(3.3)\quad &\int_{(F^*)^2} W(a, b) \int_F f(x_{-(\alpha+\beta)}(b)\, x_\beta(a^{-1}r), s)\, \psi(r)\, dr\, |a|^{3s-4} \cdot |b|^{9s-6} d^*a\, d^*b \\ &= \int_{(F^*)^2} W(a, b) \int_F f(x_{-\beta}(r)\, x_{-(\alpha+\beta)}(b), s)\, \psi(ar)\, dr\, |a|^{3s-3} |b|^{9s-6} d^*a\, d^*b.\end{aligned}$$

The last step was obtained by changing variables $a^{-1}r \to r$ and using the fact that $x_{-\beta}$ and $x_{-(\alpha+\beta)}$ commute.

Let K be the maximal compact subgroup of G_2. We will need the following matrix identities:

$$\begin{aligned}(3.4)\quad &(1)\ \textit{If } |b|>1 \textit{ then } x_{-(\alpha+\beta)}(b)=h(b^{-2}, b, b)\, x_{\alpha+\beta}(b)k \\ &(2)\ \textit{If } |r|>1 \textit{ then } x_{-\beta}(r)=h(r^{-1}, r, 1)\, x_\beta(r)k.\end{aligned}$$

Here $k \in K$.

Clearly $W(a, b)=0$ if $|a|>1$ or $|ab^3|>1$. We break the domain of integration with respect to b into two parts $|b| \leqq 1$ and $|b|>1$.

If $|b| \leqq 1$, we have

$$\int_{|a|, |b| \leqq 1} W(a, b) \int_F f(x_{-\beta}(r), s)\, \psi(ar)\, dr\, |a|^{3s-3} |b|^{9s-6} d^*a\, d^*b.$$

Define

$$G(a, s)=\int_F f(x_{-\beta}(r), s)\, \psi(ar)\, dr.$$

Then our integral on $|b| \leqq 1$ is

$$\int_{|a|, |b| \leqq 1} W(a, b)\, G(a, s) |a|^{3s-3} |b|^{9s-6} d^*a\, d^*b.$$

On $|b|>1$ we have

$$\int_{\substack{|a| \leqq 1, |b|>1 \\ |ab^3| \leqq 1}} W(a, b) \int_F f(x_{-\beta}(r)\, x_{-(\alpha+\beta)}(b), s)\, \psi(ar)\, dr\, |a|^{3s-3} |b|^{9s-6} d^*a\, d^*b.$$

Using (3.4) (1) and (1.8) we get

$$\begin{aligned}&\int_{\substack{|a| \leqq 1, |b|>1 \\ |ab^3| \leqq 1}} W(a, b) \int_F f(x_{-\beta}(r)\, h(b^{-2}, b, b)\, x_{\alpha+\beta}(b), s) \\ &\quad \cdot \psi(ar)\, dr\, |a|^{3s-3} |b|^{9s-6} d^*a\, d^*b\end{aligned}$$

$$= \int\limits_{\substack{|a| \leqq 1, |b|>1 \\ |ab^3| \leqq 1}} W(a,b) \int\limits_F f(x_{-\beta}(b^{-3}r), s)\, \psi(ar)\, dr\, |a|^{3s-3} |b|^{-6} d^*a d^*b.$$

Changing variables $b^{-3}r \to r$, we obtain

$$\int\limits_{\substack{|a| \leqq 1, |b|>1 \\ |ab^3| \leqq 1}} W(a,b)\, G(ab^3, s)\, |a|^{3s-3} |b|^{-3} \mathrm{d^*a\, d^*b}.$$

Next we compute $G(t, s)$. Since for $|\varepsilon|=1$, $G(t\varepsilon, s)=G(t, s)$ we shall compute $G(p^m, s)$ for $m \geqq 0$.

$$G(p^m, s) = \int\limits_F f(x_{-\beta}(r), s)\, \psi(p^m r)\, dr = \int\limits_{|r| \leqq 1} + \int\limits_{|r|>1}$$
$$= \int\limits_{|r| \leqq 1} f(e, s)\, dr + \int\limits_{|r|>1} f(x_{-\beta}(r), s)\, \psi(p^m r)\, dr.$$

Here we used the assumption that $m \geqq 0$; hence if $|r| \leqq 1$ so is $|p^m r|$. Also we assumed that $\psi(t) \equiv 1$ for $|t| \leqq 1$. Using relation (3.4) (2) we obtain:

$$G(p^m, s) = 1 + \int\limits_{|r|>1} f(h(r^{-1}, r, 1)\, x_\beta(r), s)\, \psi(p^m r)\, dr$$

$$= 1 + \int\limits_{|r|>1} |r|^{-3s} \psi(p^m r)\, dr = 1 + \sum_{k=1}^{\infty} q^{-3ks} \int\limits_{|r|=q^k} \psi(p^m r)\, dr$$

$$= 1 + \sum_{k=1}^{\infty} q^{-3ks+m} \int\limits_{|r|=q^{k-m}} \psi(r)\, dr = 1 - q^{-3s(m+1)+m}$$

$$+ (1-q^{-1}) \sum_{k=1}^{m} q^{(-3s+1)k}.$$

The last equality from

$$\int\limits_{|r|=q^i} \psi(r)\, dr = \begin{cases} (1-q^{-1})\, q^{+i} & i \leqq 0 \\ -1, & i=1 \\ 0, & i>1. \end{cases}$$

Using the formula for geometric sums, we get

$$G(p^m, s) = \left(\frac{1-q^{-3s}}{1-q^{-3s+1}}\right)(1-q^{(-3s+1)(m+1)}).$$

Next, using the [Ka, C-S] formula, our integral reduces to

$$(3.5)\qquad \frac{1-q^{-3s}}{1-q^{-3s+1}}\Bigg[\sum_{n,k=0}^{\infty}\mathrm{tr}(2n+3k,n+3k)\,q^{(-3s+1)(n+3k)} - \sum_{n,k=0}^{\infty}\mathrm{tr}(2n+3k,n+3k)\,q^{(-3s+1)(2n+3k+1)} + \sum_{\substack{n=0\\k=1\\n-3k\geqq 0}}^{\infty}\mathrm{tr}(2n-3k,n-3k)\,q^{(-3s+1)n} - \sum_{\substack{n=0\\k=1\\n-3k\geqq 0}}^{\infty}\mathrm{tr}(2n-3k,n-3k)\,q^{(-3s+1)(2n-3k+1)}\Bigg].$$

Here if m, n are integers with $m\geqq n\geqq 0$, then (m, n) denotes the irreducible representation of $SL(3, \mathbf{C})$ whose highest weight is the character

$$\begin{pmatrix} a & & \\ & b & \\ & & c \end{pmatrix}\mapsto a^m b^n$$

of the torus, and $\mathrm{tr}(m, n)$ denotes the trace of this irreducible representation applied to a representative of the semi-simple conjugacy class in the L-group $GL_3(\mathbf{C})$ associated with π. (This conjugacy class lies in $SL(3, \mathbf{C})$ by our assumption that the central character of π is trivial.)

Next we compute the R.H.S. of (3.1).

For any square matrix A, we have the known identity,

$$\frac{1}{\det[I-Ax]}=\sum_{r=0}^{\infty}\mathrm{tr}(\mathrm{sym}^r(A))\,x^r.$$

Hence,

$$\frac{L(3s-1,\pi,\mathrm{Ad})}{\zeta(3s)\,\zeta(6s-2)\,\zeta(9s-3)}=\frac{\sum_{r=0}^{\infty}\mathrm{tr}(\mathrm{sym}^r(A_p))\,q^{-3s+1}}{\zeta(3s)\,\zeta(6s-2)\,\zeta(9s-3)}.$$

Here A_p is the 8×8 matrix obtained by applying the adjoint representation to a representative of the semi-simple conjugacy class of $GL_3(\mathbf{C})$ corresponding to π. Comparing this to (3.5) we see that $\zeta(3s)$ is cancelled and denoting $x=q^{-3s+1}$, (3.1) reduces to proving

$$(3.6)\qquad \sum_{n,k=0}^{\infty}\mathrm{tr}(2n+3k,n+3k)\,x^{n+3k}-\sum_{n,k=0}^{\infty}\mathrm{tr}(2n+3k,n-3k)\,x^{2n+3k+1} + \sum_{\substack{n=0\\k=1\\n-3k\geqq 0}}^{\infty}\mathrm{tr}(2n-3k,n-3k)\,x^{n}-\sum_{\substack{n=0\\k=1\\n-3k\geqq 0}}^{\infty}\mathrm{tr}(2n-3k,n-3k)\,x^{2n-3k+1}$$
$$=(1-x)(1-x^2)(1-x^3)\sum_{r=0}^{\infty}\mathrm{tr}(\mathrm{sym}^r(A_p))\,x^r.$$

Next we will use the local computations for the L-functions of $GL_3 \times GL_3$ constructed by [J-S]. They prove the identity,

$$\sum_{n,k=0}^{\infty} \mathrm{tr}[(n+k,k)\otimes(n+k,n)]\, x^{n+2k} = (1-x^3) \sum_{r=0}^{\infty} \mathrm{tr}(\mathrm{sym}^r(A'_p))\, x^r$$

where A'_p is the 9×9 matrix which is the direct sum of A_p with the 1×1 identity matrix. The above formula can also be found in [Mac], as formula (4.3) on p. 33. Using the identity

$$\mathrm{sym}^r(\tau_1+\tau_2) = \sum_{k_1+k_2=r} \mathrm{sym}^{k_1}(\tau_1)\otimes \mathrm{sym}^{k_2}(\tau_2)$$

for $\tau_1=(2,1)$ $\tau_2=(0,0)$ we get

$$\mathrm{sym}^r(A'_p) = \sum_{m=0}^{r} \mathrm{sym}^m(A_p).$$

Summing over all r gives,

$$\sum_{r=0}^{\infty} \mathrm{tr}(\mathrm{sym}^r(A'_p))\, x^r = \frac{1}{1-x} \sum_{m=0}^{\infty} \mathrm{tr}(\mathrm{sym}^m(A_p))\, x^m.$$

Hence,

$$\frac{1}{1-x^3} \sum_{n,k=0}^{\infty} \mathrm{tr}[(n+k,k)\otimes(n+k,n)]\, x^{n+2k}$$

$$= \sum_{r=0}^{\infty} \mathrm{tr}(\mathrm{sym}^r(A'_p))\, x^r = \frac{1}{1-x} \sum_{m=0}^{\infty} \mathrm{tr}(\mathrm{sym}^m(A_p))\, x^m.$$

Plugging this into the R.H.S. of (3.6) then (3.1) becomes,

$$\text{(3.7)} \quad \text{L.H.S. of (3.6)} = (1-x)^2(1-x^2) \sum_{n,k=0}^{\infty} \mathrm{tr}[(n+k,k)\otimes(n+k,n)]\, x^{n+2k}.$$

Joining the two summations with minus sign on the L.H.S. of (3.6) to $\sum_{\substack{n=0 \\ n+3k\geq 0}}^{\infty}$, and comparing the r-th power in both sides of (3.7), then (3.1) is proved if we show that,

$$\text{(3.8)} \quad \sum_{\substack{n,k\geq 0 \\ n+3k=r}} \mathrm{tr}(2n+3k, n+3k) + \sum_{\substack{k\geq 1 \\ r-3k\geq 0}} \mathrm{tr}(2r-3k, r-3k)$$

$$- \sum_{\substack{n\geq 0 \\ n+3k\geq 0 \\ 2n+3k=r-1}} \mathrm{tr}(2n+3k, n+3k) = \sum_{n+2k=r} \mathrm{tr}[(n+k,k)\otimes(n+k,n)]$$

$$-2 \sum_{n+2k=r-1} \mathrm{tr}[(n+k,k)\otimes(n+k,n)]$$

$$+2 \sum_{n+2k=r-3} \mathrm{tr}[(n+k,k)\otimes(n+k,n)] - \sum_{n+2k=r-4} \mathrm{tr}[(n+k,k)\otimes(n+k,n)].$$

After changing indices in the L.H.S. of (3.8), our equality reduces to,

$$(3.9)\quad \bigoplus_{n+2k=r} (n+k,k)\otimes(n+k,n) - 2\bigoplus_{n+2k=r-1}(n+k,k)\otimes(n+k,n) + 2\bigoplus_{n+2k=r-3}(n+k,k)\otimes(n+k,n) - \bigoplus_{n+2k=r-4}(n+k,k)\otimes(n+k,n)$$
$$\cong \bigoplus_{k=1}^{\lfloor \frac{r}{3}\rfloor}(2r-3k,r) + \bigoplus_{k=0}^{\lfloor \frac{r}{3}\rfloor}(2r-3k,r-3k) - \bigoplus_{\substack{n=0\\ n+r-1\equiv 0 \bmod 3}}^{r-1}(r-1,r-n-1).$$

Here $\lfloor m\rfloor$ denotes the greatest integer function.

The proof of (3.9)

Before going into details, let us give a brief description of the way we carry out the proof.

The basic idea is to show that every highest weight (x, y) appears on both sides of (3.9) with the same multiplicity. To do so, we first compute the multiplicity of (x, y) in $(n+k, k)\otimes(n+k, n)$.

Next define $S(r, x, y)$ to be the multiplicity of (x, y) in $\bigoplus_{n+2k=r}(n+k, k)\otimes(n+k, n)$. Since x, y are fixed, we write $S(r)$ for $S(r, x, y)$.

Using this notation, (3.9) will be proved once we prove:

$$S(r)-2S(r-1)+2S(r-3)-S(r-4)=\varepsilon$$

where ε is the multiplicity of (x, y) on the R.H.S. of (3.9). (It is not hard to see that ε obtains one of the values $-1, 0, 1$.)

Denote $d(r)=S(r)-S(r-2)$, then the above equality reads,

$$d(r)-2d(r-1)+d(r-2)=\varepsilon.$$

In order to compute $d(r)$, we have to distinguish a number of cases. In some of the cases, we compute $d(r)$ by first computing $S(r)$. In other cases, we will compute $d(r)$ directly (although it is possible to compute $S(r)$ in these cases too).

We proceed to the proof of (3.9).

Fix a pair $(x, y)\, x\geqq y$. For fixed n, k with $n\leqq k$, we denote by M the multiplicity of (x, y) in $(n+k, k)\otimes(n+k, n)$. We shall give a formula for M.

Let us call a pair (x, y) admissible if it satisfies the conditions: $x+y\equiv 0 \bmod 3$; $-n\leqq\frac{2y-x}{3}\leqq n$; $\frac{x+y}{3}\leqq n+k$ and $x-y\leqq n+2k$. Notice that $x+y\equiv 0 \bmod 3$ implies $2y-x\equiv 0 \bmod 3$ and $y-2x\equiv 0 \bmod 3$.

Applying the Littlewood-Richardson rule ([Mac], p. 68), we conclude that if (x, y) is admissible, then:

For $2y\leqq x$

$$(3.10)\qquad M=\begin{cases} k+n-\frac{2x-y}{3}+1, & k<\frac{x+y}{2} \quad (1)\\ n-\frac{x-2y}{3}+1, & n\leqq\frac{x+y}{3}\leqq k \quad (2)\\ y+1, & \frac{x+y}{3}<n \quad (3). \end{cases}$$

For $x<2y$

$$M=\begin{cases} k+n-\dfrac{x+y}{3}+1, & k<\dfrac{2x-y}{3} \\ n-\dfrac{2y-x}{3}+1, & n\leqq\dfrac{2x-y}{3}\leqq k \\ x-y+1, & \dfrac{2x-y}{3}<n. \end{cases}$$

If (x, y) is not admissible, then $M=0$. If $n>k$, then by the symmetry of the tensor product we get the same expressions for M except that n and k are interchanged.

It is easy to see that in (3.9) (x, y) and $(x, x-y)$ appear with the same multiplicity. Indeed, the formulas for M are invariant under the transformation $y\to x-y$ and so are the pairs on the R.H.S. of (3.9). Hence we may assume that $2y\leqq x$. Doing so, we may ignore, on the R.H.S. of (3.9), the pairs with $y=r$, since for them $2y>x$. The same is true with "half" of the pairs with $x=r-1$.

Fix r. We shall assume that $r\geqq 5$. For $r<5$, (3.9) can be checked directly. Denote by $S(r)$ the multiplicity of (x, y) in $\bigoplus_{n+2k=r}(n+k, k)\otimes(n+k, n)$. In applying the Littlewood-Richardson rule, we assumed $n\leqq k$. Hence it is convenient to index the pairs $[n, k]$ contributing to $S(r)$ such that $n\leqq k$. More precisely, by the symmetry of $(n+k, k)\otimes(n+k, n)$ with respect to n and k we may write

$$\bigoplus_{n+2k=r}(n+k, k)\otimes(n+k, n)=\bigoplus_{\substack{n+2k=r\\ n\leqq k}}(n+k, k)\otimes(n+k, n)+\bigoplus_{\substack{k+2n=r\\ n<k}}(n+k, k)\otimes(n+k, n).$$

Now, index the points $[n_i, k_i]$ occurring in the above sums by increasing values of n_i in the first sum and then decreasing n_i in the second. For example if $r\equiv O(6)$ then $r=6u$ and:

$$r_i=\begin{cases} 2i, & 0\leqq i\leqq u \\ \dfrac{r}{2}-i, & u<i\leqq\dfrac{r}{2} \end{cases}$$

$$k_i=\begin{cases} \dfrac{r}{2}-i, & 0\leqq i\leqq u \\ 2i, & u<i\leqq\dfrac{r}{2}. \end{cases}$$

If we graph n_i and k_i as functions of i we find:

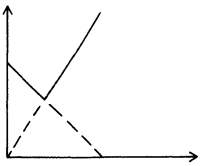


Fig. 1

In this graph the solid graph is the graph of k i.e. the line that corresponds to values of k. The dotted graph is the graph of n. The graphs of all other cases depending on r mod 6, have a similar shape. I will refer to the intersection point between the graphs of n and k as the peak point. Notice that on the right of the peak point the values of n are changed by one, and those of k by two. On the left of the peak k is changed by one and n by two.

Since the formula for M depends on the relation between $\frac{x+y}{3}$, n and k, we have to distinguish between a few cases. To do so, let us draw in the above graph a horizontal line at the point $\frac{x+y}{3}$. We have four cases:

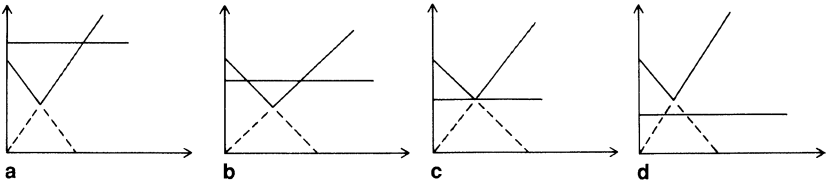


Fig. 2a (1)←(2); **b** (2)←(1)←(2); **c** (2); **d** (2)←(3)←(2)

If, for example, the line $\frac{x+y}{3}$ intersects the k-line as in (a) we will compute $S(r)$, going from right to left, starting with formula (2) in (3.10) and then as soon as $\frac{x+y}{3}$ intersects the k-line, proceed with (1). We denote this case by (1)←(2). The other graphs represent all other possibilities.

Let (x, y) be an admissible pair. Since $2y \leqq x$, then $0 \leqq \frac{x-2y}{3} \leqq n$ and since $n+2k=r$ (or $2n+k=r$) we have $x-y \leqq r$. We shall now compute $S(r)$.

(a) $S(r)$ is computed by (1)←(2).

Going from right to left, since n increases by one in this point, the first point such that M may be positive is when $n=\frac{x-2y}{3}$ (by admissibility). Notice that on the right part of the peak $2n+k=r$. Hence, for this point, since $r \geqq x-y$

$$k=r-2n=r-\frac{2}{3}(x-2y) \geqq x-y-\frac{2}{3}(x-2y)=\frac{x+y}{3}.$$

Using (2) in (3.10) we get for this n, k pair:

$$M=n-\frac{x+2y}{3}+1=\frac{x-2y}{3}-\frac{x-2y}{3}+1=1.$$

At the next pair we have (still using (2))

$$M=\frac{x-2y}{3}+1-\frac{x-2y}{3}+1=2.$$

As long as we use (2), M increases by one since n does. Let A be the point where the line $\frac{x+y}{3}$ intersects the graph of k. Assume first that the point A corresponds to a pair n_A, k_A such that $2n_A+k_A=r$, then $k_A=\frac{x+y}{3}$. For this pair $M_A=n_A+\frac{2y-x}{3}+1$. Here M_A is the value of M at the point n_A, k_A. At the next point n_A+1, k_A-2, we have to use (1) in (3.10), and the value of M at this point is

$$n_A+1+k_A-2-\frac{2x-y}{3}+1=n_A+\frac{x+y}{3}-\frac{2x-y}{3}=M_A-1.$$

Since the value of $n+k$ decreases by one as we go from right to left, then the value of M of the next point (using (1)) is M_A-2, and so on. By admissibility, the last possible value of n is $\frac{x-2y}{3}$ or $\frac{x-2y}{3}+1$, but using (1), we get that M for these values of n, k is $M=k+n-\frac{2x-y}{3}+1\leqq k+\frac{x-2y}{3}+1-\frac{2x-y}{3}+1$ $=k-\frac{x+y}{3}+2\leqq 1$ $\left(\text{since for (1) } k<\frac{x+y}{3}\right)$. Overall, we conclude that the positive values of M in this case are

$$1, 2, \ldots, M_A-1, M_A, M_A-1, \ldots, 2, 1.$$

Hence $S(r)=M_A^2$. To find M_A we recall that $2n_A+k_A=r$ and $k_A=\frac{x+y}{3}$. Hence, using (2), we get $M_A=\frac{1}{2}(r-x+y)+1$.

Notice that A corresponds to a point with $2n_A+k_A=r$ which implies that $r-x+y$ is even. If A does not correspond to a point such that $2n+k=r$, let A_1 be the first point from the right of A such that $2n+k=r$. Using the same arguments, we conclude that the positive values of M in this case are,

$$1, 2, \ldots, M_{A_1}-1, M_{A_1}, M_{A_1}, M_{A_1}-1, \ldots, 2, 1.$$

Similar computations show that $S(r)=(1+M_{A_1})\,M_{A_1}$ and that $M_{A_1}=\lfloor\frac{1}{2}(r-x+y)\rfloor+1$.

(b) $S(r)$ computed by (2) $\leftarrow$ (1) $\rightarrow$ (2).

Going from right to left, then up to the point where (1) is changed to (2), we proceed as in (a). Denote this point by B. Since, in this part, k is changed by one, there is a pair n_B, k_B with $n_B+2k_B=r$, and $k_B=\frac{x+y}{3}$. At the previous point, say B_1, $n=n_B+2$, $k=k_B-1$ and using (1)

$$M_{B_1}=k_B-1+n_B+2-\frac{2x-y}{3}+1=k_B+n_B-\frac{2x-y}{3}+2.$$

The value of M_B, using (2), is $M_B=n_B+\frac{2y-x}{3}+1$. Hence $M_{B_1}-M_B=k_B-\frac{2x-y}{3}+2-\frac{2y-x}{3}-1=\frac{x+y}{3}-\frac{x+y}{3}+1=1$. Since $n_B+2k_B=r$ then $n_B=r-2k_B=r-2\frac{x+y}{3}$ and $M_B=r-\frac{2(x+y)}{3}+\frac{2y-x}{3}+1=r-x+1$. From this point, as we go to the left, using (2), M decreases by two (since n does). Using the same argument as in (a) we see that the last positive value of M will be one or two depending whether $r-x+1$ is odd or even. Hence, if $r-x$ is even, the positive values of M are:

$$1, 3, \ldots, r-x+1, \ldots (\textit{continues as in} \text{ (a)})$$

and if $r-x$ is odd we have:

$$2, 4, \ldots, r-x+1, \ldots (\textit{continues as in} \text{ (a)}).$$

Hence, in order to compute $S(r)$ in this case, we have to subtract from the formulas in (a) the number $2+4+\ldots+(r-x)$ if $r-x$ is even, and $1+3+\ldots+r-x$ if $r-x$ is odd. We summarize:

(a) $S(r)$ computed by $(1)\leftarrow(2)$

$$S(r)=\begin{cases}(\frac{1}{2}(r-x+y)+1)^2, & r-x+y \text{ even}\\ (\lfloor\frac{1}{2}(r-x+y)\rfloor+2)(\lfloor\frac{1}{2}(r-x+y)\rfloor+1), & r-x+y \text{ odd}\end{cases}$$

(b) $S(r)$ computed by $(2)\leftarrow(1)\rightarrow(2)$

$$S(r)=(\textit{the expression in} \text{ (a)})-\begin{cases}\frac{1}{4}(r-x+2)(r-x), & r-x \text{ even}\\ \frac{1}{4}(r-x+1)^2, & r-x \text{ odd.}\end{cases}$$

It is possible to compute $S(r)$ for the remaining cases, however, it is more convenient to proceed as follows.

Denote $d(r)=S(r)-S(r-2)$.

We shall compute this number.

There are several cases. Suppose $S(r)$ and $S(r-2)$ are computed by $(1)\leftarrow(2)$. If $r-x+y$ is even, so is $r-2-x+y$ and we have $d(r)=r-x+y+1$. The same value is obtained if $r-x+y$ is odd. Suppose $S(r)$ is computed by $(2)\leftarrow(1)\leftarrow(2)$ and $S(r-2)$ by $(1)\leftarrow(2)$. If $r-x$ is even, then we have the previous difference minus $\frac{1}{4}(r-x+2)(r-x)$, i.e.

$$d(r)=r-x+y+1-\tfrac{1}{4}(r-x+2)(r-x).$$

If $r-x$ is odd, then $d(r)=r-x+y+1-\frac{1}{4}(r-x+1)^2$.

Before we proceed to the other cases, let us examine the possible values of n with $n+2k=r$. In the case of $r=3u$, n has the values

$$\ldots, u-4, u-2, u, u-1, u-2, \ldots, 2, 1, 0.$$

The values of n with $n+2\ k=r-2$ and $r=3u$ are

$$\ldots, u-4, u-2, u-1, u-2, \ldots, 2, 1, 0.$$

We see that in the case $r=3u$, the only difference between the sequences of the values of n as we pass from r to $r-2$ is an extra value with $n=u$. It is not difficult to see that this is always the case, i.e. whenever we pass from r to $r-2$, the only difference between the values of n is an extra value with $n=u$. Since, by admissibility, the first pair from the left and from the right which contribute a positive M, satisfies $n=\frac{1}{3}(x-2y)$ or $n=\frac{1}{3}(x-2y)+1$, then by the above argument, the values of n for these two first points will be exactly the same for r and $r-2$.

We now continue with the computations of $d(r)$. Assume that each $S(r)$ and $S(r-2)$ is computed by (2) or by (2)$\leftarrow$(3)$\leftarrow$(2). Since the decision whether to use (2) or (3) depends only on n, the M values for all pairs of r and $r-2$ will be computed by the same formula. Hence, except for $n=u$, the contribution of all other pairs to $d(r)$ is zero. Assume $n=u$. If, for this pair, M is computed by (3), then $M=y+1$, and hence $d(r)=y+1$. If by (2), then $n\leqq\frac{1}{3}(x+y)\leqq k$ and $M=u-\dfrac{x-2y}{3}+1$. For $r=3u$, we have $n=k=u$ and hence $\frac{1}{3}(x+y)=u$ and we get $M=\dfrac{x+y}{3}-\dfrac{x-2y}{3}+1=y+1$. Hence again, $d(r)=y+1$. A similar argument shows that for $r=3u+1$ and $r=3u+2$, $d(r)=y+1$. The last case is when $S(r)$ is computed by (2) and $S(r-2)$ by (2)$\leftarrow$(1)$\leftarrow$(2). It is not hard to see that we get, as above, $d(r)=y+1$. We summarize:

$S(r)$ and $S(r-2)$ computed by (1)$\leftarrow$(2)

$$d(r)=r-x+y+1$$

$S(r)$ computed by (2)$\leftarrow$(1)$\leftarrow$(2) and $S(r-2)$ by (1)$\leftarrow$2

$$d(r)=r-x+y+1-\begin{cases}\frac{1}{4}(r-x+2)(r-x), & r-x \text{ even}\\ \frac{1}{4}(r-x+1)^2, & r-x \text{ odd.}\end{cases}$$

In all other cases,

$$d(r)=y+1.$$

Using $d(r)$ the L.H.S. of (3.9) is

$$\Delta(r)\underset{\text{def}}{=}d(r)-2d(r-1)+d(r-2).$$

Let us compute $\Delta(r)$ for all pairs (x, y). Assume first that all $S(r-i)$ are computed by (1)$\leftarrow$(2). If $d(r-1)\neq 0$, then $d(r)\neq 0$ and we get the contribution $(r-x+y+1)-2(r-1-x+y+1)+d(r-2)$. Since $d(r-1)\neq 0$, then $r-x+y\neq 0$. If $r-x+y=1$, then $d(r-2)=0$ and we get zero. If $r-x+y>1$, then $d(r-2)\neq 0$, and we get zero again, since $d(r-2)=(r-2)-x+y+1$. If $d(r-1)=0$, then $r-x+y=0$ and $d(r)=1$. Going back to the computations of $S(r)$ in the case (1)$\leftarrow$(2), we see that $n=\frac{1}{3}(x-2y)$ and $k=\frac{1}{3}(x+y)$ with $2n+k=r$. Hence

$$y=\tfrac{1}{3}(x+y)-\tfrac{1}{3}(x-2y)=k-n=r-2n-n=r-3n$$
$$x=y+r=2r-3n.$$

So a pair (x, y) contributes 1 to $\Delta(r)$ if $(x, y)=(2r-3n, r-3n)$ with $n \leqq \left[\frac{r}{3}\right]$. By admissibility, each such pair appears.

To summarize, the contribution to $\Delta(r)$ from this case is one if $(x, y)=(2r-3i, r-3i)$ with $i \leqq \left[\frac{r}{3}\right]$ and zero otherwise.

Next assume that $S(r)$ is computed by $(2) \leftarrow (1) \leftarrow (2)$ and the others by $(1) \leftarrow (2)$. If this is the case, then $k=\frac{x+y}{3}$ and $n=0$, 1 (see the graphs). If $n=1$, then r is odd. But then the highest value of k which can be obtained with $n+2k=r$ or $n+2k=r-1$, is the same, which means that $S(r-1)$ is computed by $(2) \leftarrow (1) \leftarrow (2)$. A contradiction. Hence $n=0$. By admissibility, $\frac{x-2y}{3} \leqq 0$ which means $x=2y$. Hence $k=y$ and $r=2k+n=x$. Since $r-x$ is even, we get: $d(r)=r-x+y+1-\frac{(r-x+2)(r-x)}{4}=y+1$; $d(r-1)=r-1-x+y+1=y$; $d(r-2)=r-2-x+y+1=y-1$. Hence the contribution, in this case, to $\Delta(r)$ is zero.

We turn to the case where $S(r)$ and $S(r-1)$ are computed by $(2) \leftarrow (1) \leftarrow (2)$ and the others by $(1) \leftarrow (2)$. Similar considerations as above show that $x=r-1$ (now n need not be zero!). Since $r-x$ is odd, we get: $d(r)=r-x+y+1-\left(\frac{r-x+1}{2}\right)=y+1$, $d(r-1)=r-1-x+y+1-\frac{1}{4}(r-1-x+2)(r-1-x)=y+1$; $d(r-2)=r-2-x+y+1=y$. Hence the contribution to $\Delta(r)$ in this case is -1. Let us show that these pairs are of the form $(r-1, r-i-1)$ with i, $i+3j \geqq 0$, $2i+3j=r-1$. This is the same as $0 \leqq i \leqq r-1$ with $n+i-1 \equiv 0 \bmod 3$ which is the third sum on the R.H.S. of (3.9). We have $x=r-1$ and $\frac{1}{3}(x+y)=k$; hence $y=3k-x=3k-r+1=k+r-n-r+1=k-n+1$. Denote $i=r-k+n-2$ and $j=1-n$. Then $i=k+2n-2 \geqq 0$ for $r \geqq 3$. Also $i+3j=r-k+n-2+3-3n=k-n+1 \geqq 0$ since $n \leqq k$. Finally, $2i+3j=2r-2k+2n-4+3-3n=r-1$.

Since each such pair is admissible, we conclude that all of them appear in $\Delta(r)$ with multiplicity -1.

If $S(r-1)$ is computed by $(2) \leftarrow (1) \leftarrow (2)$ and the rest by $(1) \leftarrow (2)$, then as above $x=r-2$, and it is easy to check that the contributions to $\Delta(r)$ are zero. Finally, if all $S(r-i)$ are computed by $(2) \leftarrow (1) \leftarrow (2)$, then $d(r-i)=y+1$, and we get zero contribution.

We summarize: The only pairs which contribute to $\Delta(r)$, a nonzero multiplicity, are $(x, y)=(2r-3i, r-3i)$ with $0 \leqq i \leqq \left[\frac{r}{3}\right]$, in this case the multiplicity is one, and the pairs $(r-1, r-i-1)$ with i, $i+3j \geqq 0$, $2i+3j=r-1$ with multiplicity -1. But these are exactly the pairs which appear on the R.H.S. of (3.9). We are done.

References

[C-S] Casselman, W., Shalika, J.: The unramified principal series of p-adic groups II: The Whittaker function. Comput. Math. **41**, Fasc. 2, 207–231 (1980)

[Ch] Chang, B.: The conjugate classes of Chevalley groups of type G_2. J. Algebra, **9**, 190–211 (1968)

[G-Sh] Gelbart, S., Shahidi, F.: Analytic Properties of Automorphic L-functions. Perspect. Math. **6**, 1988

[Hu] Humphreys, J.E.: Introduction to Lie Algebras and Representation Theory. (Grad. Texts Math., vol. 9) Berlin Heidelberg New York: Springer

[J-S] Jacquet, H., Shalika, J.A.: On Euler products and the classification of automorphic representations I and II. Am. J. Math. **103**, 499–558 and 777–815 (1981)

[Ka] Kato, S.: On an explicit formula for class one Whittaker functions on split reductive groups over p-adic fields. University of Tokyo (Preprint 1978)

[Mac] Macdonald, I.G.: Symmetric Functions and Hall Polynomials. Oxford: Clarendon Press 1979

Invent. math. 105, 589–607 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

The uniqueness of groups of type J_4*

Michael Aschbacher and Yoav Segev

[1] California Institute of Technology, Department of Mathematics, Pasadena, CA 91125, USA
[2] Ben Gurion University, Beer Shiva 84105, Israel

Oblatum VIII-1990 & 31-I-1991

We give the first computer free proof of the uniqueness of groups of type J_4. In addition we supply simplified proofs of some properties of such groups, such as the structure of certain subgroups.

A *group of type* J_4 is a finite group G possessing an involution z such that $H = C_G(z)$ satisfies $F^*(H) = Q$ is extraspecial of order 2^{1+3}, H/Q is isomorphic to $\mathbf{Z}_3$ extended by $\mathrm{Aut}(M_{22})$, and $z^G \cap Q \neq \{z\}$. We prove:

Main Theorem. *Up to isomorphism there exists at most one group of type* J_4.

Janko was the first to consider groups of type J_4 in [6], where he established various properties of such groups. For example Janko showed that each group G of type J_4 is simple, he determined the order of G, and he described the normalizers of all subgroups of G of prime order. However Janko left open the question of whether there exist groups of type J_4 and whether all groups of type J_4 are isomorphic. Thus Janko is said to have *discovered* J_4 and indeed J_4 was the last of the 26 sporadic simple groups to be discovered.

Around 1980, Norton, Parker, and Thackray proved the existence and uniqueness of groups of type J_4 using extensive machine computation. This work is discussed briefly in [7]. Norton et al. construct J_4 as a linear group in 112 dimensions over the field of order 2. While the notion of a 2-*local geometry* did not exist at that time, this geometry plays an implicit role in [7]. For example Table 1 in [7] lists the orbits of the stabilizers of certain flags in this geometry on the points of the geometry.

Our proof is based on the theory of *uniqueness systems* developed in [3]. We apply this theory to the 2-local geometry Γ of a group G of type J_4 and its collinearity graph Δ. Given the theory in [3], the only difficult step in the uniqueness proof is to show that the graph Δ is *simply connected* in the language of [3]. To do so requires various facts about groups of type J_4. Many of those facts already appear in [6] or [7], but others do not. For completeness and

* This work was partially supported by BSF 88-00164. The first author is partially supported by NSF DMS-8721480 and NSA MDA 90-88-H-2032.

clarity we supply our own proofs of all necessary facts about J_4, and thus make no appeal to [6] or [7]. One by-product of this approach is better proofs of some properties of J_4.

1 Two lemmas on graphs

In this section Δ is an undirected graph. For $S \subseteq \Delta$ let $\Delta(S) = \bigcap_{s \in S} \Delta(s)$; thus for example $\Delta(x, y) = \Delta(x) \cap \Delta(y)$. For $x, y \in \Delta$, $X, Y \subseteq \Delta$, write $d_\Delta(x, y) = d(x, y)$ for the distance from x to y in Δ and define

$$d_\Delta(X, Y) = d(X, Y) = \min \{d(x, y): x \in X, y \in Y\}.$$

Given a symmetric relation R on Δ, write $R(x)$ for the set of y such that $(x, y) \in R$. We also write R for the graph with vertex set Δ and edge set R. Thus d_R denotes the distance function in this graph and a subset of Δ is *R-connected* if it is connected as a subgraph of R.

Let $\Delta^n(x) = \{y \in \Delta: d(x, y) = n\}$ and $\Delta^{\leqq n}(x) = \{y \in \Delta: d(x, y) \leqq n\}$. In addition we adopt the notation and terminology from Sects. 2 and 3 of [3]; in particular the notions of *triangulable graph* and 4-*generated graph* appear there.

Consider the following hypothesis:

Hypothesis A *The following two conditions hold:*

(A1) $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_N$ *is a chain of symmetric relations on* Δ *such that* $R_0 = \Delta^{\leqq 1}$ *and* $R_N = \Delta^{\leqq 2}$.

(A2) *For each* $0 \leqq k < N$, $x \in \Delta$, $y \in R_{k+1}(x) \cap \Delta^2(x)$, *and* $z \in \Delta^2(x)$,

$$d_{R_m}(R_k(y) \cap \Delta(x), \Delta(x, z)) \leqq 1,$$

where $m = \min \{k, 1\}$.

In the remainder of this section let P be the set of paths of Δ and $\sim$ be the equivalence relation determined by some closed subset S of P containing all triangles of Δ. (cf. [3]) We will be assuming hypothesis (A1). Given $(x, y) \in R_N$, let $\text{type}(x, y) = \min \{i: (x, y) \in R_i\}$ and given $p = x_0 \ldots x_n \in P$ let $\text{type}(p) = \min \{\text{type}(x_i, x_{i+2}): 0 \leqq i \leqq n-2\}$.

We often use the following elementary observation implicitly without comment.

(1.1) *If* $p = x_0 \ldots x_n \in P$ *such that* $p \not\sim q$ *for each* $q \in P$ *of length less than* n, *then* $x_{i+2} \in \Delta^2(x_i)$ *for each* $i \leqq n-2$.

Proof. If not, $p \sim q$ where q is the path obtained by deleting x_{i+1} from p.

(1.2) *If Hypothesis* A *holds then* Δ *is* 4-*generated.*

Proof. Let S be the closure of the set of all triangles and squares of Δ. Pick $c = x_0 \ldots x_n$ to be a S-nontrivial cycle of minimal length n, and subject to this constraint with $\text{type}(c) = t$ minimal. Then $n \geqq 5$ and $t \geqq 1$. Relabeling the vertices of c if necessary, we may assume $x_2 \in R_t(x_0)$. Let $p = x_0 \ldots x_4$. Then by (A2),

$$d_{R_m}(R_k(x_0) \cap \Delta(x_2), \Delta(x_2, x_4)) \leqq 1$$

where $k=t-1$ and $m=\min\{k, 1\}$. So there exists $u\in R_k(x_0)\cap\Delta(x_2)$ and $v\in\Delta(x_2, x_4)$ such that $u\in R_m(v)$.

Now if $k=0$ then $u\in x_0^\perp\cap v^\perp$ so $q=x_0uvx_4\sim p$ by (3.2) in [3]. Indeed this argument shows

(*) If $r=y_0\dots y_4\in P$ with $y_2\in R_1(y_0)$ then $r\sim r'$ for some $r'\in P$ of length 3.

Further in our case $c\sim c'=q\cdot(x_4\dots x_n)$, and by minimality of n, $c'\sim 1$, contradicting c nontrivial.

So $k>0$ and there exists $y\in x_0^\perp\cap u^\perp$. Next by (*) applied to $r=ux_2vx_4x_5$, $r\sim r'$ for some $r'\in P$ of length 3. Now $r\sim ux_2x_3x_4x_5$ by (3.2) in [3], and as $d=x_0x_1x_2uyx_0$ has type at most type $(x_0, u)\leqq k$, $d\sim 1$ by minimality of t. Then by (3.2) in [3], $c\sim c'=x_0yur'x_5\dots x_n$ of length n. But type $(x_0, u)\leqq k<t$, so $c'\sim 1$, completing the proof.

Before stating our next result we need more notation and terminology. Given a path xuy in Δ write $C\Delta(xuy)$ for the connected component of u in $\Delta(x, y)$. A ternary relation $\hat{R}$ on Δ is *symmetric* if $(x, u, y)\in\hat{R}$ implies $(y, u, x)\in\hat{R}$. Given a symmetric ternary relation $\hat{R}$ write R for the symmetric binary relation $(x, y)\in R$ if and only if $(x, u, y)\in\hat{R}$ for some $u\in\Delta$.

Hypothesis B *The following three conditions hold:*

(B1) $\hat{R}_0\subseteq\dots\subseteq\hat{R}_N$ *is a chain of symmetric ternary relations on* Δ *such that* $\hat{R}_0$ *is the set of all triangles of* Δ *and* $\hat{R}_N$ *is the set of all paths of* Δ *of length* 2.

(B2) *For all* $0\leqq k<N$, $(x, u, y)\in\hat{R}_{k+1}-\hat{R}_0$, *and* $(x, v, z)\in\hat{R}_N-\hat{R}_0$,

$$d_{R_m}(D_k(xuy), C\Delta(xvz))\leqq 1$$

where $m=\min\{k, 1\}$ *and* $D_k(xuy)$ *consists of* $C\Delta(xuy)$ *together with all* $w\in\Delta(x)$ *such that* $(y, c, w)\in\hat{R}_k$ *for some* $c\in C\Delta(xuy)$.

(B3) *If* $y\in R_1(x)\cap\Delta^2(x)$ *and* $w\in\Delta(y)\cap\Delta^2(x)$ *then* $d_\Delta(\Delta(x, y), w)\leqq 1$.

(1.3) *Hypothesis* B *implies Hypothesis* A *with respect to the binary relations* R_i *induced by* $\hat{R}_i$.

Proof. Evidently (B1) implies (A1). Further $D_k(xuy)\subseteq R_k(y)\cap\Delta(x)$, so by (B2), $d_{R_m}(R_k(y)\cap\Delta(x), \Delta(x, z))\leqq d_{R_m}(D_k(xuy), C\Delta(xvz))\leqq 1$, and hence (A2) holds.

(1.4) *If Hypothesis* B *holds then* Δ *is triangulable.*

Proof. Let S be the closure of the triangles. By (1.2) and (1.3) it suffices to show each square is trivial. Given a square $p=abcda$ let $t(p)$ be the minimum t such that $uvw\in\hat{R}_t$ for some subpath uvw of some cyclic permutation of p. Pick $p\notin S$ with $t=t(p)$ minimal. Without loss $abc\in\hat{R}_t$.

Observe first that if $y\in\Delta^2(x)\cap R_1(x)$ then $\Delta(x, y)$ is connected. This follows from (B2) with $k=0$ and $y=z$. Namely as $k=0$, $D_k(xuy)=C\Delta(xuy)$, so by (B2), $d_\Delta(C\Delta(xuy), C\Delta(xvy))\leqq 1$, and hence $C\Delta(xuy)=C\Delta(xvy)$.

In particular this observation together with (3.4) in [3] shows $t>1$. Similarly if $x\in\Delta(a, b, c)$ then $abcxa\sim 1$ by (3.2.1) in [3]. Thus we may replace b by any $b'\in C\Delta(abc)$. Therefore if $d_{R_1}(C\Delta(abc), C\Delta(adc))\leqq 1$ we may assume $d\in R_1(b)$. Then as $\Delta(b, d)$ is connected, p is trivial by (3.4) in [3].

So by (B2) there exists $w \in \Delta(a)$ such that $cbw \in \hat{R}_{t-1}$ and $d \in R_1(w)$. Then p is the product of the triangle $abwa$, and the pentagon $q = dawbcd$. Thus it remains to show $q \sim 1$.

But by (B3) there is $x \in \Delta(w, d, c)$, so q is the product of the squares $q_1 = wxdaw$, $q_2 = wbcxw$, and the triangle $xcdx$. Also $t(q_1) \leqq 1$ and $t(q_2) \leqq t-1$, so $q_1 \sim q_2 \sim 1$, and hence $q \sim 1$, completing the proof.

2 4-dimensional binary symplectic space

In this section F is the field of order 2, V is a 4-dimensional vector space over F, and $(\ ,\ )$ is a nondegenerate symmetric bilinear form on V. That is $(V, (\ ,\ ))$ is a 4-dimensional symplectic space over F. Let G be the isometry group of $(\ ,\ )$; thus $G \cong Sp_4(2) \cong S_6$.

Points $u, v \in V^\#$ are *adjacent* if u is orthogonal to v. An *oval* is a set of five pairwise nonadjacent points of V. Given $x \in V^\#$ define the *disc* $\mathrm{Disc}(x)$ at x to be the set of points adjacent to x but distinct from x; that is $\mathrm{Disc}(x) = x^\perp - \langle x \rangle$.

We record the following well known facts which follow from easy counting arguments:

(2.1) (1) *Each point is in exactly 2 ovals and each pair of ovals intersects in a unique point.*
(2) *Each oval intersects each singular line in a unique point.*

(2.2) *Let $\varnothing \neq S \subseteq V^\#$ such that for each $s \in S$ and each oval $\mathcal{O}$ in the set $\mathcal{O}(S)$ of ovals meeting S, $|\mathcal{O} \cap S| = 3$. Then either*
(1) *S is a disc, $|S| = 6$, and $|\mathcal{O}(S)| = 4$.*
(2) *$|S| = 9$.*

Proof. Let $n = |S|$ and $m = |\mathcal{O}(S)|$. Then by (2.1), counting the set of pairs $(s, \mathcal{O})$ with $s \in S$ and $s \in \mathcal{O} \in \mathcal{O}(S)$, we conclude $2n = 3m$. Pick $s \in S$ and let A, B be the ovals through s. Then $A \cup B$ contains 4 points of S distinct from s and these are the points in $S - s^\perp$. So $n = 5 + k$ where $k = |S \cap \mathrm{Disc}(s)|$. As $n \equiv 0 \bmod 3$, $k \equiv 1 \bmod 3$, so $k = 1$ or 4.

Now for $s \neq a \in S \cap A$, $s + a$ is the unique member of B distinct from s not adjacent to a. Thus if $k = 1$, $a + s \in S$, so $B \cap S = \{s, a+s, a'+s\}$ where $A \cap S = \{s, a, a'\}$. Therefore $S = \{s, a, a', a+s, a'+s, a+a'\} = \mathrm{Disc}(s+a+a')$ when $n = 6$.

Thus we may take $k = 4$, so $n = 9$.

Remark 2.3 Define singular lines of V to be adjacent if their intersection is nontrivial. There is a duality of the geometry of V induced by the outer automorphism of G interchanging points and singular lines and preserving adjacency and inclusion. Thus all the notions above dualize, and it is these dual statements which we use in later sections.

For example an *oval* of singular lines is a set $\mathcal{O}$ of 5 singular lines no two of which are adjacent; equivalently $\mathcal{O}$ is a partition of V by singular lines. Similarly if l is a singular line then the *disc* $\mathrm{Disc}(l)$ at l is the set of 6 singular lines adjacent to l; i.e. the 6 singular lines intersecting l nontrivially. Of course there are dual versions of (2.1) and (2.2).

3 The geometry of M_{24}

In this section G is the Mathieu group M_{24} and Γ is the 2-local geometry of G. Thus Γ is a geometry on $I=\{1, 2, 3\}$. We view the objects of type i as the *octads*, *trios*, *sextets* permuted by G, for $i=1, 2, 3$, respectively. To correspond with terminology used in later sections we also refer to trios as *lines* and the sextets as *planes*.

Let $\hat{V}$ be the Golay code module for G. Thus $\hat{V}$ is a 12-dimensional module over the field F of order 2 and we view $\hat{V}$ as a subspace of the power set of the set X of 24 points permuted by G with addition equal to symmetric difference. There is a triple of forms P, C, f defined on $\hat{V}$ via

$$P(x)=|x|/4 \bmod 2, \qquad C(x, y)=|x\cap y|/2 \bmod 2, \qquad f(x, y, z)=|x\cap y\cap z| \bmod 2.$$

Moreover f is a symmetric trilinear form with $f(x, x, y)=0$ for all $x, y=\hat{V}$. (cf. Sect. 1 in [2].) The radical of $\hat{V}$ is X, so we have the induced forms on $V=\hat{V}/\langle X\rangle$ which we also write as P, C, f. Thus V is the *11-dimensional Golay code module* or *dual Todd module*.

For $x\in V$, f_x is the bilinear form $f_x(y, z)=f(x, y, z)$, C_x is the quadratic form $C_x(y)=C(x, y)$ (which has f_x as its associated bilinear form) and $R(x)=\mathrm{Rad}(f_x)$. We say a subspace U of V is *singular* if P is trivial on U and U is singular with respect to f_x for all $x\in V$.

Given $S\subseteq V$, let $S\theta$ consist of those $v\in V$ such that $f_v=0$ on S. Thus $S\theta$ is a subspace of V. We say a subspace U of V is *hyperbolic* if $U\theta$ is a hyperplane of V and P is trivial on U.

We identify each trio and sextet with the set of octads incident with the object. Subject to this convention the members of Γ are subspaces of V. The octads are the singular points of V and the trios are the singular lines of V, (so the trios are indeed "lines") and the sextets are 4-dimensional and the maximal hyperbolic subspaces of V. Incidence in Γ becomes inclusion of subspaces under this representation. Moreover

(3.1) *Let U be a sextet and $f_U=f_v$ for $v\in V-U\theta$. Then (U, f_U) is a 4-dimensional symplectic space over F and the stabilizer G_U of U is isomorphic to $S_6/Z_3/E_{2^6}$ and induces the symplectic group $O(U, f_U)\cong Sp_4(2)$ on U. For $A\leqq U$, $R(A)\cap U$ is the subspace of U orthogonal to A under f_U, where $R(A)=\bigcap_{a\in A} R(a)$.*

(3.2) *If x is an octad then*
(1) $\dim(R(x))=5$, *G_x is the stabilizer of x in* $\mathrm{GL}(R(x))$, *and $R(x)$ is the union of the lines through x.*
(2) *$(V/R(x), C_x)$ is a 6-dimensional orthogonal space of* sign $+1$, *G_x acts as $\Omega_6^+(2)$ on $V/R(x)$, and for $U\in\Gamma_3(x)$, $U+R(x)/R(x)$ is a singular point of $V/R(x)$ and $U\theta/R(x)$ is the subspace orthogonal to that point.*
(3) *The map $U\mapsto R(x)\cap U$ is a bijection between $\Gamma_3(x)$ and the set of 3-subspaces of $R(x)$ containing x.*
(4) *Each pair of distinct coplanar lines is contained in a unique plane.*
(5) *If $U\in\Gamma_3(x)$ then $G_{x,U}$ has three orbits on $\Gamma_3(x)$: $\{U\}$, the 18 planes W in $U\theta$ with $U\cap W\in\Gamma_2$, and the 16 planes Z not contained in $U\theta$ with $Z\cap U=\langle x\rangle$.*
(6) *G is rank 4 on octads with orbitals $(x, x_i)G$, where $x_1=x$, $\langle x, x_2\rangle$ is a singular line, $\langle x, x_3\rangle$ is a hyperbolic line, and $x+x_4$ is nonsingular.*

(7) *Each hyperbolic line h is contained in a unique plane U, and $U = h \oplus R(h)$.*

Proof. Parts (1) and (6) are easy and well known. Let $U \in \Gamma_3(x)$. Then $R(x) \cap U$ is the hyperplane of U orthogonal to x under f_U, so $\bar{U}$ is a point in $\bar{V} = V/R(x)$. Further $(\bar{V}, C_x)$ is a nondegenerate orthogonal space of dimension 6 and G_x induces $\Omega_6^+(2)$ on $\bar{V}$, so the space is of sign $+1$. By definition of C_x, $\bar{U}$ is a singular point in $\bar{V}$ and $U\theta/R(x) = \bar{U}^\perp$. That is (2) holds. Also as $G_{x,U}$ is the parabolic of G_x stabilizing $R(x) \cap U$ and $\bar{U}$, (3) and (5) hold.

If l and m are lines on U then either $l + m = U$ or $l \cap m$ is a an octad which we take to be x. In the first case clearly U is the unique plane containing l and m. In the second $l + m = R(x) \cap U$, so U is unique by (3). Thus (4) is established.

It remains to prove (7). As each plane contains a hyperbolic line h and (6) says G is transitive on hyperbolic lines, we may take $h \leqq U \in \Gamma_3$. Then $U\theta \leqq h\theta$, so as $U\theta$ is a hyperplane, $U\theta = h\theta$. Then as G is transitive on planes and $N_G(U) = N_G(U\theta)$, U is the unique plane containing h. By (3.1), $U \cap R(h) = k$ is the line orthogonal to h in U, so $U = h \oplus k$. Further if $x \in R(h)$ then $x + h \leqq W \in \Gamma_3$ by (3), so by uniqueness of U, $W = U$ and hence $x \in k$.

Given a trio l let $\zeta(l) = \langle R(x) : x \in l \rangle$. Then

(3.3) $\dim(\zeta(l)) = 8$ *and* $\zeta(l) = R(x) + R(y)$ *with* $R(x) \cap R(y) = l$ *for each pair of distinct octads* $x, y \in l$, *and each octad in* $\zeta(l) - l$ *is in a unique* $R(u)$, $u \in l^\#$.

Proof. From (3.2), $G_{x,y}$ is irreducible on $R(x)/l$, so $R(x) \cap R(y) = l$. Thus $\dim(R(x) + R(l)) = 8$, so as G_l has a composition series on V with degrees 2, 6, 3, it follows that $\zeta(l) = R(x) + R(y)$ is of dimension 8, and each octad in $\zeta(l) - l$ is in at most one $R(u)$, $u \in l^\#$. As $\zeta(l)/l$ is the tensor product of natural modules for $L_2(2)$ and $L_3(2)$ as a G_l-module, G_l has two orbits on vectors in $\zeta(l) - l$, and as one of these contains nonsinular points, each such octad is in at least one $R(u)$.

(3.4) *The stabilizer G_l of l has 5 orbits $\delta_i(l)$, $0 \leqq i \leqq 4$ on Γ_2 as follows:*
(1) $\delta_0(l) = \{l\}$ *and* $G_l \cong (L_3(2) \times S_3)/E_{2^6}$.
(2) $\delta_1(l) = \{k : 0 \neq k \cap l\}$ *is of order* 42.
(3) $\delta_2(l) = \{k : k + l \in \Gamma_3\}$ *is of order* 56.
(4) $\delta_3(l) = \{k : \dim(k \cap \zeta(l)) = 1\}$ *is of order* 1,008.
(5) $\delta_4(l) = \{k : k \cap \zeta(l) = 0\}$ *is of order* 2,688.

Proof. First G_l is transitive on the 3 octads on l and taking $x \in l$, by (3.2), $R(x)$ contains 15 lines through x and G_x acts 2-transitively as $L_4(2)$ on these lines, so there are $3 \cdot 14 = 42$ lines in $\delta_1(l)$ and they form an orbit under G_l.

Second G_l is transitive on the 7 sextets incident with l and if U is such a sextet, then from (3.1), $G_{U,l}$ is transitive on the 8 singular lines of U not adjacent to l, so there are $7 \cdot 8 = 56$ lines in $\delta_2(l)$ and they form an orbit under G_l.

Third by (3.2), $G_{l,x}$ is transitive on the 28 octads z in $R(x) - l$. By (3.2.3), $z + l$ is contained in a unique sextet U and by (3.2.3) and (3.2.7), $\zeta(l) \cap R(z) = U \cap R(z)$ is a 3-subspace of $R(z)$ with $G_{z,U} = G_{z,U \cap R(z)}$. Let $x = l \cap R(z)$. Then $G_{z,x,U} = G_{z,l} O_2(G_z)$, so $G_{l,z}$ is transitive on the 12 lines in $R(z) - \zeta(l)$. Thus the $3 \cdot 28 \cdot 12 = 1{,}008$ lines in $\delta_3(l)$ form an orbit. This leaves 2,688 lines k with $k \cap \zeta(l) = 0$. Checking the permutation character of G on G/G_l, we find G is rank 5 on lines, and hence conclude that $\delta_4(l)$ forms the fifth orbit.

(3.5) *The stabilizer G_U of a sextet U has orbits $\Xi_i(U)$, $0 \leqq i \leqq 3$, on sextets, where:*
(1) $\Xi_0(U) = \{U\}$.
(2) $\Xi_1(U) = \{W: W \cap U \in \Gamma_2\}$ *of order* 90.
(3) $\Xi_2(U) = \{W: W \cap U \in \Gamma_1\}$ *of order* 240.
(4) $\Xi_3(U) = \{W: U \cap W = 0\}$ *of order* 1,440.
Moreover members of $\Xi_i(U)$ are not contained in $U\theta$ for $i = 2, 3$.

Proof. As G_U is transitive on the 15 lines $l \in \Gamma_2(U)$ and G_l is 2-transitive on the 7 planes in $\Gamma_3(l)$, $\Xi_1(U)$ is an orbit of length $15 \cdot 6 = 90$. As G_U is transitive on the 15 octads $x \in \Gamma_1(U)$ and by (3.2), $G_{U,x}$ is transitive on the 16 planes $W \in \Gamma_3(x)$ with $\langle x \rangle = U \cap W$, $\Xi_2(U)$ is an orbit of length $15 \cdot 16 = 240$. Further by (3.2), $W \not\leqq U\theta$. This leaves 1,440 sextets W with $W \cap U = 0$. From the character table of G, G is rank 4 on G/G_U, so $\Xi_3(U)$ is the fourth orbit of G_U on planes.

(3.6) *Let $U \in \Gamma_3$, $W \in \Xi_1(U)$, and $l \in \Gamma_2(W)$ with $l \cap U = 0$. Then $\zeta(l) \cap U = W \cap U \in \Gamma_2$.*

Proof. If $x \in U \cap \zeta(l)^\#$ then by (3.3), $x \in R(y)$ for some $y \in l^\#$. Further if $\langle x, x' \rangle$ is a line in U then as $U = (R(x) \cap U) + (R(x') \cap U)$, $(R(x) + U) \cap (R(x') + U) = U + (R(x) \cap R(x')) = U$, so $R(y) \cap U = \langle x \rangle$. Thus $\zeta(l) \cap U = (R(y) \cap U) + (R(y') \cap U) = W \cap U$, where $l = \langle y, y' \rangle$.

(3.7) *For each sextet U, G_U is transtive on the* 384 *octads and* 640 *nonsingular points in $V - U\theta$, and if x is such an octad then (U, C_x) is an orthogonal space of* sign -1 *and $U_{U,x} \cong \mathbf{Z}_3 \times S_5$ induces the isometry group of C_x on U with kernel of order* 3.

Proof. Let $x \in V - U\theta$ be an octad. Now $O_2(G_U) = E$ is the group of transvections with axis $U\theta/U$ on V/U, so E is regular on $V/U - U\theta/U$. Thus $H = G_{U, U+x}$ is a complement to E in G_U, so $H \cong S_6/\mathbf{Z}_3$. Now H has two orbits $x_i H$ on $x + U$ of order 6, 10. So as 27 does not divide the order of G_x, we conclude $x = x_1$ and G_U is transitive on the $64 \cdot 6$ octads in $V - U\theta$, and $G_{U,x}$ is as claimed.

(3.8) *The stabilizer of a sextet U has orbits $\kappa_i(U)$, $0 \leqq i \leqq 3$, on Γ_2, where:*
(1) $\kappa_0(U) = \Gamma_2(U)$ *of order* 15.
(2) $\kappa_1(U) = \{l: l \cap U \in \Gamma_1\}$ *of order* 180.
(3) $\kappa_2(U) = \{l: 0 = l \cap U$ *and* $l \leqq W \in \Xi_1(U)\}$ *of order* 720.
(4) $\kappa_3(U) = \{l: l \not\leqq U\theta\}$ *of order* 2,880.

Proof. Of course $\kappa_0(U)$ is an orbit of length 15. Next G_U is transitive on the 15 octads $x \in \Gamma_1(U)$ and by (3.2), $G_{x,U}$ is transitive on the 12 lines l with $U \cap l = \langle x \rangle$, so $\kappa_1(U)$ is an orbit of length $15 \cdot 12 = 180$.

Now $W \in \Xi_1(U)$ contains 8 lines l with $l \cap U = 0$ and by (3.6), $W \cap U = \zeta(l) \cap U$. Then $W = (\zeta(l) \cap U) + l$ is uniquely determined, so $\kappa_2(U)$ is of order $90 \cdot 8 = 720$. This leaves 2,880 lines. From (3.7), G_U is transitive on the 384 octads $x \in V - U\theta$, and $G_{x,U}$ is transitive on the 15 lines through x, so G_U is transitive on the $384 \cdot 15/2 = 2{,}880$ lines in $\kappa_3(U)$.

For $l \in \Gamma_2$ write $\mathrm{Co}(l)$ for the set of lines coplanar with l and let Co be the graph on Γ_2 in which lines are adjacent if and only if they are coplanar.

(3.9) *Lines l, m are coplanar if and only if $m \in \delta_i(l)$ for $i \leqq 2$.*

Proof. This follows as G is flag transtive on Γ and if l is a line a plane U then the orbits of $G_{l,U}$ on $\Gamma_2(U)$ are $\delta_i(l) \cap U$ for $i \leqq 2$.

(3.10) *Let l, m be lines with $m \in \delta_3(l)$. Then*
(1) *There exists a unique plane $\pi(l, m) \in \Gamma_3(l)$ with $m \in \kappa_1(\pi(l, m))$.*
(2) $\mathrm{Co}(l, m) \subseteq \Gamma_2(\pi(l,m)) \cup \Gamma_2(\pi(m, l))$.
(3) *Let $\langle y \rangle = m \cap \pi(l, m)$ and $\langle x \rangle = l \cap \pi(m, l)$. Then*

$$\mathrm{Co}(l, m) \cap \Gamma_2(\pi(l, m)) = \Gamma_2(\pi(l, m), y) \subseteq \delta_1(m)$$

with $x + y \in \delta_1(l)$ and the remaining two lines in $\Gamma_2(\pi(l, m), y)$ contained in $\delta_2(l)$.

Proof. As $m \in \delta_3(l)$, $m \cap \zeta(l) = \langle y \rangle$ for some octad y. Then by (3.3) there exists an octad $x \in l \cap R(y)$ and by (3.2) there is a unique plane U containing $\langle l, y \rangle$. As $m \not\leqq \zeta(l)$, $m \not\leqq U$, so $m \in \kappa_1(U)$. Conversely if $U' \in \Gamma_3(l)$ with $m \in \kappa_1(U')$ then $m \cap U' \leqq m \cap \zeta(l)$, so $m \cap U' = \langle y \rangle$. Then as U is the unique plane containing $\langle l, y \rangle$. $U = U'$. That is (1) holds with $U = \pi(l, m)$.

Suppose $k \in \mathrm{Co}(l, m)$. Then there is a plane W containing $l + k$ and $k \in \delta_{i(s)}(s)$ for $s = l, m$ and $i(s) = 1, 2$. Also $k \leqq \zeta(l) \cap \zeta(m)$, so if z is an octad in $k \cap m$ then $z \in m \cap \zeta(l) = \langle y \rangle$ and hence $z = y$. In particular if $k \in \delta_1(l, m)$ then by symmetry, $k = x + y$. So assume $k \in \delta_2(l)$. Then $\langle y, l \rangle \leqq k + l = W$, so $W = U$ by uniqueness of U, and (2) and (3) hold in this case.

Thus we may assume $k \in \delta_2(l, m)$. Then $k \leqq W \cap k + m$, so $k + m \in \Xi_1(W)$ and $k = k + m \cap W$. Hence by (3.6), $k = \zeta(m) \cap W$, impossible as $x \in \zeta(m) \cap W$.

(3.11) *Let U, π be a planes and D a disc or oval of lines in U. Then*
(1) *For each line l, $(\mathrm{Co}(l) \cup \delta_3(l)) \cap D \neq \emptyset$.*
(2) $d_{\mathrm{Co}}(D, \Gamma_2(\pi)) \leqq 1$.

Proof. As $\dim(U) > \mathrm{codim}(\zeta(l))$, there exists an octad $x \in \zeta(l) \cap U$. As D is a disc or oval in U, $x \in d$ *for some* $d \in D$. Thus $d \cap \zeta(l) \neq \emptyset$, so $d \notin \delta_4(l)$ and (1) holds.

Similarly $U \cap \pi\theta$ is of codimension at most 1 in U, so as D is a disc or oval, there exists $d \in D$ with $d \leqq \pi\theta$. Thus $d \in \kappa_i(\pi)$ for $i = 0, 1, 2$. If $i = 0, 1$ then there is an octad $y \in d \cap \pi$ and then by (3.2), d is coplanar with any $m \in \Gamma_2(y, \pi)$.

Thus we may assume $i = 2$. Hence $d \leqq W \in \Xi_1(\pi)$, so $k = \pi \cap W \in \delta_2(d)$, completing the proof of (2).

(3.12) *Let l be a line, $g \in G$, and $R = O_2(G_l)$. Then*
(1) *If $R \cap R^g \neq 1$ then $lg \in \mathrm{Co}(l)$.*
(2) *If $|R : R_{lg}| \leqq 4$ then $lg \in \mathrm{Co}(l)$.*

Proof. $R \cong E_{64}$ and $m_2(G_l/R) = 3$, so if $|R : R_{lg}| \leqq 4$ then there is $1 \neq r \in R \cap R^g$. As R is weakly closed in $N_G(R) = G_l$ with respect to G, G_l controls fusion in R and hence $C_G(r)$ is transitive on those $S \in R^G$ with $r \in S$. Thus R^g is conjugate to R under $C_G(r)$.

Now G has two classes of involutions. If r is 2-central then $C_G(r)$ fixes an octad $x \in l$ so $x \in lg$ and hence $lg \in \mathrm{Co}(l)$ by (3.9). On the other hand if r is not 2-central then $C_G(r)$ fixes a plane U with $l \leqq U$, so $lg \leqq U$, so again $lg \in \mathrm{Co}(l)$.

(3.13) *Let l be a line. Then*
(1) *If U is a plane with $l \in \kappa_1(U)$ then $l \notin \delta_4(m)$ for any $m \in \Gamma_1(U)$.*
(2) *If $m \in \delta_4(l)$ then $\delta_2(l) \cap \delta_2(m) \neq \emptyset$.*

Proof. If $l \in \kappa_1(U)$ there is an octad $x \in l \cap U$. Then $R(x) \cap m \neq 0$, so $m \notin \delta_4(l)$. Thus (1) holds. Next let π be a plane with $l \in \kappa_2(\pi)$. By (3.6) there is a line

$k \leqq \pi$ with $l \in \delta_2(k)$. Pick $m \in \delta_2(k)$ on π. By (3.4) and (3.6), $m \notin \delta_1(l) \cup \delta_2(l)$. By (3.10), $m \notin \delta_3(l)$. Thus $m \in \delta_4(l)$ and $k \in \delta_2(l) \cap \delta_2(m)$, so (2) holds by transitivity of G_l on $\delta_4(l)$.

4 Groups of type J_4

In this section we assume the following hypothesis:

Hypothesis J_4 *G is a finite group, z an involution in G, $H = C_G(z)$, $Q = F^*(H)$ is extraspecial of order 2^{13}, $H/Q \cong \mathrm{Aut}(M_{22})/\mathbf{Z}_3$, and z is not weakly closed in Q with respect to G.*

A group satisfying Hypothesis J_4 will be said to be of *type* J_4. We write $\tilde{H}$ for $H/\langle z \rangle$, and view $\tilde{Q}$ as an orthogonal space over GF(2) and H/Q as a subgroup of the orthogoal group on $\tilde{Q}$-

(4.1) *Let $X \in \mathrm{Syl}_3(O_{2,3}(H))$. Then $H = QN_H(X)$ with $\langle z \rangle = C_Q(X)$, $C_{\tilde{H}}(X)$ quasisimple, and $\tilde{Q}$ the unique 6-dimensional irreducible* GF(4)*-module for $C_{\tilde{H}}(X)$.*

Proof. By a Frattini argument, $H = QN_H(X)$. Next $\tilde{Q}$ is a faithful GF(2) module for $N_H(X)Q/Q$ and a GF(4)-module for $C_H(X)Q/Q$. Hence the lemma follows from the list of 2-modular irreducibles for coverings of M_{22} contained in James [5].

(4.2) *H has two orbits on involutions in $Q - \langle z \rangle$ with representatives r_1 and r_2, where $C_H(r_1)Q/Q \cong S_5/E_{32}$ and $C_H(r_2)Q/Q \cong S_5/E_{16}$.*

Proof. This follows from the structure of the 6-dimensional irreducible for H/Q and the fact that $t \in Q - \langle z \rangle$ is an involution if and only if $\tilde{t}$ is singular in the ortogonal space $\tilde{Q}$.

(4.3) (1) $z^G \cap Q = \{z\} \cup r_1^H$.
(2) *$m(Q \cap Q^g) = 7$ and $Z = Q \cap Q^g \cap Q^k$, where $r_1 = z^g$ and $Z = \langle z, [r_1, X] \rangle = \langle z, z^g, z^k \rangle$.*

Proof. By Hypothesis J_4 there is $r = z^g \in Q - \langle z \rangle$ and by (4.2) we may take $r = r_1$ or r_2. As Q has width 6, the elementary theory of large extraspecial subgroups says $m(Q \cap Q^g) = m \leqq 7$ and $(Q^g \cap H)Q/Q$ is normal in $C_H(\tilde{r})/Q$ and of 2-rank $12 - m$. Hence $m(O_2(C_H(\tilde{r}))/Q \geqq 12 - m \geqq 5$, so we conclude from (4.2) that $r = r_1$ and $m = 7$. Further $I = Q \cap Q^g \trianglelefteq C_H(\tilde{r})$, while from the structure of the GF(4) H/Q-module $\tilde{Q}$ this uniquely determines $\tilde{I}$ and forces the 1-dimensional GF(4)-subspace $\tilde{Z}$ to be contained in $\tilde{I}$ with $\tilde{I} \cap \tilde{I}^s = \tilde{Z}$ for $s \in N_H(Z)$ inverting X. So (2) holds.

(4.4) *Let $K = \langle Q^x : z^x \in Z \rangle$ and $S = C_G(Z)$. Then $N_G(Z) = KS$ with $K \cap S = R = O_2(K) = O_2(S)$, $K/R \cong L_3(2)$, $S/R \cong S_5$, R is special with $Z(R) = Z$ and $m(R/Z) = 12$, and R/Z is the tensor product of natural modules for K/R and $E(S/R) \cong L_2(4)$.*

Proof. Lemma 4.3.2 and the elementary theory of large extraspecial subgroups implies $R = O_2(K)$ is special of order 2^{15} with center Z and with R/Z the sum of natural modules for $K/Z \cong L_3(2)$ dual to Z. In particular as K induces $\mathrm{GL}(Z)$ on Z, $N_G(Z) = KS$. Further $S = C_H(Z)$ with $|O_2(C_H(Z))| = 2^{15}$ and $S/O_2(S) \cong S_5$,

so $R = O_2(S)$ and $KS/R = K/R \times S/R \cong L_3(2) \times S_5$, with R/Z the tensor product of 3 and 4 dimensional irreducibles for the factors. Finally the irreducible $E(S/R)$-submodules on R/Z are $L_2(4)$-modules as RQ/Q is such a module.

From the structure of H/Q there exists a subgroup $G_{1,3}$ of H such that $QX \leqq G_{1,3}$ and $G_{1,3}/QX \cong S_6/E_{16}$, while from the structure of the GF(4)-module $\tilde{Q}$, $C_{\tilde{Q}}(O_2(G_{1,3})) = \tilde{E}_1$ is a 6-dimensional irreducible for $E(G_{1,3}/O_2(G_{1,3})) \cong S_6/\mathbf{Z}_3$. Let $T \in \mathrm{Syl}_2(G_{1,3})$, so that $T \in \mathrm{Syl}_2(H)$. As $\langle z \rangle = Z(T)$, we have

(4.5) $T \in \mathrm{Syl}_2(G)$.

Further $|C_{\tilde{Q}}(T)| = 2$, so from (4.2), we may choose T so that $\langle \tilde{r}_1 \rangle = C_{\tilde{Q}}(T)$, and hence $\tilde{Z} = [\tilde{r}_1, X] \trianglelefteq \tilde{T}$ and $Z \leqq E_1$. Let $G_{1,2,3} = G_{1,3} \cap N_G(Z)$, so that $|G_{1,3} : G_{1,2,3}| = 15$. Let $E = (O_2(G_{1,3}) \cap C_G(X))E_1$, $M = N_G(E)$, and $M^* = M/E$.

(4.6) (1) $E = J(T) \cong E_{2^{11}}$.

(2) $G^*_{1,3}$ *is the split extension of* $Q^* \cong E_{64}$ *by* $(G_{1,3} \cap N_G(X))^* \cong S_6/\mathbf{Z}_3$, *and in particular is determined up to isomorphism.*

Proof. Let $I = G_{1,3} \cap C_G(X)$. Then I is transitive on $O_2(I/\langle z \rangle)^\#$ of order 15, so $O_2(I) \cong E_{32}$. Notice $O_2(I) = C_E(X)$. Next $E_1 \cong E_{128}$ with $E_1 = \langle z \rangle \times [E_1, X]$ so as $O_2(I)$ acts on $[E_1, X]$ and centralizes $\tilde{E}_1$, we have $E \cong E_{2^{11}}$. By construction $G_{1,3} \cap N_G(X)$ acts on E and as $E_1 \trianglelefteq Q$ with $[\tilde{Q}, E] = \tilde{E}_1$, Q acts on E. Hence $E \trianglelefteq Q(G_{1,3} \cap N_G(X)) = G_{1,3}$. Moreover (2) is immediate.

To prove (1), it suffices to show that if $A^* \leqq T^*$ with $A^* \cong E_{2^k}$, then $k < m = m(E/C_E(A))$. Now $m \leqq m_1 + m_2$, where $m_1 = m(\tilde{E}_1/C_{\tilde{E}_1}(A))$ and $m_2 = m(E/E_2)$, where $E_2/E_1 = C_{E/E_1}(A)$. Now $\tilde{E}_1$ and E/E_1 are the 6 and 4 dimensional modules for $G_{1,3}$ acting as $S_6/\mathbf{Z}_3$ and S_6, respectively, so $m(AQE/QE) \leqq 3$, $m_1 + m_2 \geqq 4$, and if $m(AQE/QE) = 3$ then $m_1 \geqq 4$ and $m_2 \geqq 4$ and $m_2 \geqq 1$. It follows that $m(A^* \cap Q^*) \geqq 2$ and $m_1 \geqq 4$ in case of equality. To complete the proof we show that if $B^* \leqq Q^*$ then $m(E/C_E(B)) \geqq m(B^*) + 4$ when $m(B^*) > 2$ and $m(E/C_E(B)E_1) \geqq 3$ when $m(B^*) = 2$.

First from the structure of the H/Q-module $\tilde{Q}$, if $E_1 \leqq D \leqq E$, then $\tilde{E}_1 = C_{\tilde{Q}}(D)$ when $|D : E_1| > 2$, while $m(C_{\tilde{Q}}(D)) = 8$ when $|D : E_1| = 2$. It follows that if $E_1 \leqq B \leqq Q$ and $m(B^*) > 2$ then $C_E(B) \leqq C_E(\tilde{B}) = E_1$, so that $m(E/C_E(B)) = 4 + m(E_1/C_{E_1}(B)) = 4 + m(B^*)$, as E_1 is maximal abelian in Q. On the otherhand if $m(B^*) = 2$ then $|C_E(\tilde{B}) : E_1| \leqq 2$.

Define $G_{1,2} = N_G(Z) \cap N_G(E)$.

(4.7) $K \leqq G_{1,2}$ with $G^*_{1,2}$ *the split extension of* E_{64} *by* $L_3(2) \times S_3$ *and* $R^* = O_2(G^*_{1,2})$ *the tensor product of natural modules for the factors.*

Proof. As $E \leqq S$ and $E = J(T)$, $E \trianglelefteq KT$. Indeed $ER/R \in \mathrm{Syl}_2(E(S/R))$, so $N_S(E)/R \cong S_4$ and $G_{1,2}/RE \cong S_3 \times L_3(2)$. Finally as $R/Z \cong E_{2^{12}}$, $RE/E \cong E_{64}$ is the sum of two natural modules for KE/E, so the lemma holds.

(4.8) M *is the split extension of* E *by* M_{24} *with* E *the Todd module for* M^*.

Proof. Let $\langle t \rangle = Z(T^*)$. We show $C_{M^*}(t) \cong L_3(2)/2^{1+6}$. Then the existence of $G^*_{1,2}$ and $G^*_{1,3}$ show $M^* \cong M_{24}$. (cf. Held [4]) Now by James [5], there are just two faithful 11-dimensional GF(2) M_{24}-modules, the Todd module and its dual, so as $G_{1,3}$ is the stabilizer of z, E is the Todd module for M^*. To show M splits over E we construct a complement to E in $G_{1,3}$ and appeal to Gaschutz' Theorem.

Namely $C_{\tilde{H}}(X) \cong M_{22}/\mathbf{Z}_3$ is quasisimple, so from the structure of Aut (M_{22}), $\tilde{G}_{1,3} \cap N(X)$ has a complement $\tilde{D}$ to $C_{\tilde{E}}(X)$. Now as M_{22} has one class of involutions, involutions in $E(\tilde{D})$ lift to involutions in D, so $E(\tilde{D})$ splits over $\langle z \rangle$. Thus $E(D) \cong S_6/\mathbf{Z}_3$. Further there is a D-complement $\tilde{U}_1$ to $\tilde{E}_1$ in $\tilde{Q}$ and $U_1 = \langle z \rangle \oplus U$, where $U = [U_1, X]$ is D-invariant. Finally every involution in $M^* \cong M_{24}$ is fused into U_1^*, so if s^* is an involution inverting X^* with $m([C_{\tilde{E}_1}(X), s]) = 2$ then we may choose s to be an involution inverting X by the Baer-Suzuki Theorem. (cf. 39.6 in [1]) If $\tilde{r}$ is an involution in $\tilde{s}C_{\tilde{E}_1}(X)$, r is conjugate to s or sz, *so* r is an involution. Therefore there is a complement D_1 to $\langle z \rangle$ in D and $D_1 U$ is our complement to E in $G_{1,3}$.

Thus it remains only to show $C = C_{M^*}(t) \cong L_3(2)/2^{1+6}$. Let $C_i = C \cap G_{1,i}^*$ and $A_i = O_2(G_{1,i}^*)$ for $i = 2, 3$. By (4.6) and (4.7), $G_{1,i}^*$ are determined up to isomorphism with T^* Sylow in each of these groups and in M^* by (4.5). So as M_{24} possess such subgroups, it follows from the structure of M_{24} that there is $P \trianglelefteq C_i$ with $P \cong 2^{1+6}$, $C_i/P \cong S_4$, and $C_0 = \langle C_1, C_2 \rangle$ induces $L_3(2)$ on $P/\langle t \rangle$.

Next $I = [E, t] \leqq E_1$ with $B = C_3 \cap C(I) = A_3 \cap P$. Then as $C_3 = C \cap C(z)$, $B = C \cap C(I) \trianglelefteq C$ and $C/B = C_0/B \cong L_3(2)/E_8$ or $L_4(2)$. The latter case is out as $B \trianglelefteq C$.

(4.9) $C_H(X) = E(C_H(X))$ *and* $N_H(X)$ *splits over* $C_H(X)$.

Proof. As E is the Todd module for $M^* \cong M_{24}$, $z \in [C_{E_1}(X), C_M(X)]$, so $C_H(X) = E(C_H(X))$. We saw during the proof of the previous lemma that X in inverted by an involution in H, so H, so $N_H(X)$ splits over $C_H(X)$.

5 The geometry of a group of type J_4

In this section we continue the hypothesis and notation established in the previous section. In addition let $G_1 = M$, $G_2 = N_G(Z)$, and $G_3 = H$. We have already defined $G_{1,2}$ and $G_{1,3}$. Let $G_{2,3} = N_H(Z)$. Thus by construction, $G_{i,j} = G_i \cap G_j$. Notice also that $G_{1,2,3} = G_1 \cap G_2 \cap G_3$, where $G_{1,2,3} = G_{1,3} \cap N_G(Z)$ as defined in the previous section. Observe $G_i = \langle G_{i,j}, G_{i,k} \rangle$ for all distinct i, j, k and $G_2 = G_{1,2} G_{2,3}$.

Let Γ be the coset geometry on $\{G_1, G_2, G_3\}$. Thus $\Gamma_i = G/G_i$ is the set of objects of Γ of type i. We call the members of Γ_i *points*, *lines*, and *planes*, for $i = 1, 2, 3$, respectively, and write (x, l, π) for the flag (G_1, G_2, G_3). For $\gamma \in \Gamma$, write $G_\gamma^{*\gamma} = G_\gamma^*$ for the induced action of G_γ on its residue $\Gamma(\gamma)$ and let $Q(\gamma)$ denote the kernel of the action of G_γ on $\Gamma(\gamma)$. Thus $Q(x) = E$, $Q(l) = R$, and $Q(\pi) = O_{2,3}(H) = QX$. Write $\Delta^d(x)$ for the points of distance d from x in the collinearity graph Δ of Γ, let $\Delta(x) = \Delta^1(x)$, and $x^\perp = \{x\} \cup \Delta(x)$.

(5.1) (1) $G_{\Gamma_3(l)} = K$ *induces* S_5 *on the* 5 *points in* $\Gamma_1(l)$ *and* $G_{\Gamma_1(l)} = S$ *induces* $L_3(2)$ *on the* 7 *planes in* $\Gamma_3(l)$.
(2) *For* $y \in \Gamma_1(l) - \{x\}$, $M_y = KN_{S \cap M}(Y)$ *for some* $Y \in \mathrm{Syl}_3(S \cap M)$, (*so that* M_y *is determined up to conjugation in* M) $O_2(M_y) = R$, $Q(y) \cap E = Z$, *and* $|Q(y) : Q(y)_x| = 4$.

Proof. From (4.4) and (4.7), $|G_2 : G_{1,2}| = 5$ with $G_2/K \cong S_5$, so the first statement of (1) holds. A similar argument establishes the second.

Now (1) implies the first two statements of (2). As we observed during the proof of (4.7), $ER/R \in \mathrm{Syl}_2(E(S/R))$, so $|E: E_y| = 4$, which is the last statement of (2) as $Q(x) = E$. Finally from (4.4), R/Z is the sum of 3 natural modules for $E(S/R)$, so $\dim(C_{R/Z}(ER/R)) = 6 = \dim((E \cap R)/Z)$ and $C_{R/Z}(ER/R) \cap C_{R/Z}(Q(y)R/R) = 0$. Thus $(E \cap R)/Z = C_{R/Z}(ER/R)$ and $Q(y) \cap E = Z$.

(5.2) (1) *Γ is a residually connected string geometry and G is flag transitive on Γ.*
(2) *Each pair a, b of distinct collinear points is incident with a unique line $a+b$.*
(3) *If $a, b \in \Gamma_1(\pi)$ are collinear then $a+b \in \Gamma_2(\pi)$.*

Proof. Part (1) follows as $G_i = \langle G_{i,j}, G_{i,k} \rangle$ for all distinct i, j, k, $G_2 = G_{1,2} G_{2,3}$, but $G_1 \neq H_{1,2} G_{1,3}$ and $G_3 \neq G_{1,3} G_{2,3}$. (cf. [3]) Actually we also need $G = \langle H, M \rangle$, but we postpone that verification until (6.5). For the moment (1) at least holds for $G_0 = \langle H, M \rangle$.

Next $G_{1,i}$ is maximal in G_1 for $i = 2, 3$ and by (5.1.2), $Q(x)$ is transitive on $\Gamma_1(l) - \{x\}$, so (2) follows from (7.4) in [3]. Finally Q is weakly closed in H, so (3) follows from (7.7) in [3].

(5.3) (1) *H is rank 3 on $\Gamma_1(\pi)$ and $G_{1,3}$ has orbits $\{x\}$, $\Delta(x) \cap \Gamma(\pi)$, $\Gamma_1(\pi) - x^\perp$ of order 1, 60, 16, respectively.*
(2) *For $u \in \Gamma_1(\pi) - x^\perp$, π is the unique plane incident with x and u, so $M_u = M_{\pi,u}$ with $Q = O_2(M_u)$, $M_u E = G_{1,3}$, and $Q(u)_x = Q(u) \cap Q$ is of order 2^7 with $\tilde{Q}(u)_x$ a complement to $\tilde{E}_1$ in $\tilde{Q}$.*
(3) *H has three orbits on $\Gamma_1(\pi) \times \Gamma_2(\pi)$ with representatives (x, l_i), $1 \leqq i \leqq 3$, where $x \in l = l_1$, $l_2 \subseteq \Delta(x)$, and $|l_3 \cap \Delta(x)| = 3$.*

Proof. The action of H on $\Gamma_1(\pi)$ is equivalent to the action of H on $H/G_{1,3}$, which is a rank 3 action in which $G_{1,3}$ has orbits $\{x\}$, X_1, and X_2, where X_1 consists of the points $y \in \Gamma_1(\pi)$ such that $x+y$ is a conjugate of l under G_3, $|X_1| = 60$, and $|X_2| = 16$. So $X_1 \subseteq \Delta(x)$. Also for $u \in X_2$, $H^{*\pi}_{x,u}$ is a complement to $O_2(H^{*\pi}_x) = E^{*\pi}$ in $H^{*\pi}_x$, so $H_{x,u} E = G_{1,3}$. Then as $G_{1,3}$ is maximal in M, $M_u \leqq G_{1,3}$. Finally $\tilde{E}_1$ is the unique nontrival irreducible for $G_{1,3}$ is maximal in M, $M_u \leqq G_{1,3}$. Finally $\tilde{E}_1$ is the unique nontrival irreducible for $G_{1,3}$ on $\tilde{Q}$ and M_u is irreducible on $\tilde{Q}/\tilde{E}_1$, so as $|Q(u)_x| = 2^7$, (2) follows. Then (2) and (5.1.2) imply $u \notin \Delta(x)$, completing the proof of (1).

Part (3) is just a property of the 2-local geometry for $\mathrm{Aut}(M_{22})$.

Write $\Delta_1^2(x)$ for the set of $u \in \Delta^2(x)$ such that $\Gamma_3(x, u)$ is nonempty. By (5.3), $\Delta_1^2(x)$ is an orbit of M.

(5.4) *Let $y \in \Gamma_1(l) - \{x\}$. Then*
(1) *$G_{1,2}$ has 5 orbits $\delta_i(x, l)$, $0 \leqq i \leqq 4$, on $\Gamma_2(x)$.*
(2) *$\delta_0(x, l) = \{l\}$ and $\bigcup_{i=0}^{2} \delta_i(x, l)$ is the set of lines in $\Gamma_2(x)$ coplanar with l.*
(3) *$|\delta_1(x, l)| = 42$ and for $k \in \delta_1(x, l)$, $|k \cap \Delta(y)| = 3$.*
(4) *$|\delta_2(x, l)| = 56$ and $K \subseteq \Delta(y)$ for $k \in \delta_2(x, l)$.*
(5) *$|\delta_i(x, l)| = 1{,}008$, $2{,}688$, for $i = 3, 4$, respectively.*
(6) *For $i > 1$ and $k \in \delta_i(x, l)$, $Q(x)_y$ is transitive on $\Gamma_1(k) - \{x\}$.*

Proof. Part (1) follows from (3.4), which also supplies the length of the orbits. Part 2 is (3.9). Lemma 5.3.3 completes the proof of (3) and (4). It remains to establish that E_y is transitive on $k - \{x\}$ in (6). From (4.8), E is the Todd module

for $M^* \cong M_{24}$ and by (7.4.1) in [3], $G_{1,2} = EG_{x,y}$. Hence by (5.1.2), E_y is a $G_{1,2}$-invariant subspace of E of codimension 2. Thus from the structure of the Todd module, E_y is the *unique* such subspace of E and $E = E_y E_u$ for all $u \in \Delta(x)$ with $x + u \in \delta_i(x, l)$ for $i > 1$, completing the proof of (5). (This is easiest to see in the dual E^* of E where $G_{1,2}$ is the stabilizer of the line E_y^* and $E_y^* \cap E_u^* = 0$ unless $i \leqq 1$; cf. the discussion in Sect. 3.)

(5.5) *Each triangle in Δ is incident with a plane.*

Proof. By (7.8) in [3], (4.4), and (5.1.2), it suffices to show $|R^* : R_k^*| > 4$ for $k \in \delta_i(x, l)$ with $i > 2$. This is (3.12).

(5.6) *Let abc be a triangle in Δ. Then*
(1) *$a + c \in \delta_i(a, a+b)$ for $i \leqq 2$.*
(2) *$N_G(\{a, b, c\})$ induces S_3 on $\{a, b, c\}$.*
(3) *If $a + c \in \delta_i(a, a+b)$ then $b + c \in \delta_i(b, a+b)$.*

Proof. Notice (2) and (5.2.2) imply (3). Further to prove (2) it suffices to show each transposition on $\{a, b, c\}$ is induced in G.

By (5.5) we may assume $a, b, c \in \Gamma_1(\pi)$, and then by (5.2.3), $a+b$, $a+c$, $b+c \in \Gamma_2(\pi)$. If $c \in a+b$ then (2) holds by (5.1.1), so we assume $c \notin a+b$. Hence by (5.4.2), $a + c \in \delta_i(a, a+b)$ for $i = 1$ or 2. In particular (1) is established.

As δ_i is a self paired orbital of G_a on $\Gamma_2(a)$, there is $g \in G_a$ with cycle $(a+b, a+c)$. By (5.1.2), E is transitive on $a + c - \{a\}$, so replacing g by a suitable member of gE we may assume $bg = c$. Similarly $bg^2 \in a + b - \{a\}$ and if $i = 2$ then by (5.4.6), E_c is transitive on $a + b - \{a\}$, so replacing g by a suitable member of gE_c we may take $cg = b$. If $i = 1$ then as $bg = c \in \Delta(b)$, $cg \in \Delta(b)$. Further $E_c \neq E_b$ and E/E_b is regular on $a + b - \{a\}$ by (5.1), while by (5.4.3), $|\Delta(c) \cap a + b| = 3$, so E_c is transitive on $\Delta(c) \cap a + b - \{a\}$. Thus again adjusting in E_c we may take $cg = b$, completing the proof.

(5.7) *Let $y \in \Delta_1^2(x) \cap \Gamma_1(\pi)$. Then $\Delta(x, y) \subseteq \Gamma_1(\pi)$.*

Proof. Let $u \in \Delta(x, y)$; then $y = xg$ for some $g \in G_u$. By (5.3.2), $z \in E \cap E^g$. But E fixes each line through x, so $E \cap E^g$ acts on $x + u \cap y + u = \{u\}$, and hence $z \in G_u$. If $z \in Q(u)$ then $u \in \Gamma_1(\pi)$, so assume $1 \neq z^{*u}$. Then $1 \neq E_u^{*u} \cap (E^g)_u^{*u}$, so by (3.12.1), $u + x$ is coplanar with $u + y$ in $\Gamma(u)$, and hence $u \in \Gamma_1(\pi)$, since $\pi = \Gamma_3(x, y)$ by (5.3.2).

6 Generation of a group of type J_4

In this section we continue the hypothesis and notation of Sects. 4 and 5.

(6.1) *$z^g \in E$ if and only if $\pi g \in \Gamma_3(x)$.*

Proof. As $E = J(T)$, M is transitive on $z^G \cap E$.

(6.2) *Assume $E \cap E^g \cap z^G$ is nonempty. Then*
(1) *$xg \in x^\perp \cup \Delta_1^2(x)$.*
(2) *$E \cap E^g \in Z^G$ if $xg \in \Delta(x)$ while $E \cap E^g = \langle z \rangle$ if $xg \in \Delta_1^2(x)$.*
(3) *If $E \neq E^g$ then $E^\# \cap E^g \subseteq z^G$.*

Proof. We may assume $z \in E \cap E^g$; equivalently by (6.1), x, $xg \in \Gamma_1(\pi)$. Hence (5.3.1) implies (1), (5.3.2) implies the second statement in (2), and (5.1.2) implies the first statement of (2). Finally (2) implies (3).

(6.3) (1) *M has two orbits on $E^\#$ with representatives z and t.*
(2) $C_M(t)^{*x} \cong \mathrm{Aut}(M_{22})$.
(3) *E is the unique member of E^G containing t, so $C_G(t) = C_M(t)$.*

Proof. As E is the Todd module for M^{*x}, (1) and (2) hold. Let $I = C_G(t)$. As $E = J(T)$, $I \leqq M$ if and only if E is the unique member of E^G containing t, so it remains to show $I \leqq M$. We need the following facts about M_{24} and its action on the Todd module:
(a) M^* has two classes of involutions with representatives s_1 and s_2, where s_1 is 2-central, fixes 8 of the 24 points permuted by M_{24}, and $m(C_E(s_1)) = 7$, while s_2 has no fixed points on the 24 points and $m(C_E(s_2)) = 6$.
(b) If $t \in U \leqq E$ with $U^\# \subseteq t^M$ and $m(U) > 1$ then $m(U) = 2$ and $N_{(I \cap M)^*}(U) \cong \mathbf{Z}_2/L_3(4)$ has 2-rank 4.
(c) $(I \cap M)^*$ has 2-rank 5 and each involution in $E((I \cap M)^*)$ is in s_1^G.

We conclude from (6.2.3) and (b) that:
(d) For all $g \in I - M$, $E^\# \cap E^g \subseteq t^G$ and $m(E \cap E^g) \leqq 2$.

Pick $U \leqq E$ as in (b). By (d), $N_I(V) \leqq I \cap M$ for all $V \leqq E$ properly containing U. Thus as $2 + m(N_{I \cap M^*}(U) < 7 = m(C_E(s))$ for each involution s in the subgroup $E(N_{(I \cap M)^*}(U))$ of index 2 in $N_{(I \cap M)^*}(U)$, a standard argument shows $N_{(I \cap M)}(U)$ is the unique fixed point of V on $N_I(U)/N_{I \cap M}(U)$. As $N_{I \cap M}(U)$ is not transitive on $(E/U)^\#$ it follows that $N_I(U) \leqq M$. Now applying the same argument to the action of I on $I/(I \cap M)$, we conclude $I \leqq M$, as desired.

(6.4) *G has two classes of involutions with representatives t and z.*

Proof. First each involution in $M - E$ is conjugate under M to an element of $s_i E$, $i = 1, 2$ where s_i^* are the elements of (a) from the proof of (6.3). Let $I(sE)$ be the set of involutions in sE. Then $I(sE) = sC_E(s)$ and $s^M \cap sE = s[E, s]$. Now $|C_E(s_2)\colon [E, s_2]| = 2$ while $|C_E(s_1)\colon [E, s_1]| = 8$ and $C_L(s_1)$ is transitive on $C_E(s_1)/[E, s_1]$, so representatives for orbits of M on involutions are s_i, $s_i e_i$, $i = 1, 2$, $e_i \in C_E(s_i) - [E, s_i]$.

Next $s_i E$ contains an involution in Q, which we may take to be s_i. Further $[E, s_i]$ is properly contained in $C_{E \cap Q}(s_i)$, so all involutions in M are fused into Q by the previous paragraph. Hence by (4.2), G has at most two classes of involutions, establishing the lemma.

(6.5) (1) *G is simple.*
(2) *$G = \langle G_i, G_j \rangle$ for all distinct $i, j \in \{1, 2, 3\}$.*

Proof. By the theory of large extraspecial subgroups, $F^*(G) = \langle z^G \rangle$ is simple. Then $M = \langle z^G \cap M \rangle \leqq F^*(G)$ by the discussion in the proof of (6.4), so by a Frattini argument, $G = F^*(G) N_G(E) = F^*(G)$.

Let $K = \langle G_1, G_2, G_3 \rangle$ and observe $K = \langle G_i, G_j \rangle$ for distict $i, j \in \{1, 2, 3\}$. Further $M = \langle C_M(z), C_M(t) \rangle$, while $H = C_G(z)$ and by (6.3), $C_M(t) = C_G(t)$, so $K = \langle H, M \rangle = \langle C_G(z), C_G(t) \rangle$.

Notice K is of type J_4, so all results proved for G apply equally well to K. So by (6.4), $s^G \cap K = s^K$ and $C_G(s) = C_K(s)$ for all involutions s in K. It follows that either $K = G$ or K is strongly embedded in G, and the latter is impossible as G has two classes of involutions.

7 The J_4 graph is simply connected

Continue the hypothesis and notation of Sect. 5. We view $\Gamma(x)$ as the geometry of lines and planes of M_{24} discussed in Sect. 3. In particular two lines $l, m \in \Gamma_2(x)$ are adjacent in the sense of Sect. 3 when $m \in \delta_1(x, l)$ and l is perpendicular to m when $m \in \delta_2(x, l)$. From Sect. 3, if $\pi \in \Gamma_3(x)$ then $\Gamma_2(x, \pi)$ can be viewed as vectors in 4-dimensional binary symplectic space as in Sect. 2. In particular we have notions of ovals, discs, and singular lines in that space.

For $x, y \in \Delta$, $l \in \Gamma_2$, let $L(x, y) = \{x+u: u \in \Delta(x, y)\}$ and

$$L(x, l) = \{x+u: u \in \Delta(x) \cap \Gamma_1(l)\}.$$

We will use Lemma 1.4 to prove that Δ is triangulable. Thus we need to produce a chain $(\hat{R}_i: 1 \leqq i \leqq N)$ of symmetric ternary relations satisfying Hypotheses B. We adopt the notational coventions of Sect. 1. Of course $\hat{R}_0$ is the set of triangles of Δ. Define $\hat{R}_1$ to consist of all triples (x, u, y) such that $u+x$ and $u+y$ are coplanar. Then by (5.4.2), $\hat{R}_1$ consists of the triples (x, u, y) such that $u+y \in \bigcup_{i=0}^{2} \delta_i(u, u+x)$ and $R_1(x) = R_0(x) \cup \Delta_1^2(x)$.

Let $\hat{R}_2$ consist of $\hat{R}_1$ together with the triples (x, u, y) such that $y+u \in \delta_3(u, u+x)$. Finally let $\hat{R}_3$ consist of all paths xuy in Δ.

(7.1) *Let $l \in \Gamma_2(x)$ and $y \in \Gamma_1(l) - \{x\}$. Then*
(1) *If $m \in \delta_1(x, l)$ then $|\Gamma_1(m) \cap \Delta(y)| = 3$.*
(2) *If $m \in \delta_2(x, l)$ then $\Gamma_1(m) \subseteq \Delta(y)$.*

Proof. See (5.4).

(7.2) *For each triangle xyz in Δ, $N_G(\{x, y, z\})$ induces S_3 on $\{x, y, z\}$. In particular if $x+y \in \delta_i(x, x+z)$ then $y+z \in \delta_i(y, y+x)$ for $i = 1, 2$.*

Proof. This is (5.6).

(7.3) *Let $x \in \Gamma_1(\pi)$ and $m \in \Gamma_2(\pi) - \Gamma_1(x)$. Then*
(1) *If $|\Gamma_1(m) \cap \Delta(x)| = 3$ then $L(x, m)$ is a singular line in $\Gamma_2(x, \pi)$.*
(2) *If $\Gamma_1(m) \subseteq \Delta(x)$ then $L(x, m)$ is an oval in $\Gamma_2(x, \pi)$.*

Proof. If $|\Gamma_1(m) \cap \Delta(x)| = 3$ then by (7.2), $L(x, m)$ is a clique of size 3 in $\Gamma_1(x, \pi)$, so $L(x, m)$ is a singular line. Similarly if $\Gamma_1(m) \subseteq \Delta(x)$ then $L(x, m)$ is a coclique of order 5, and hence is an oval.

(7.4) *Let $(x, u, y) \in \hat{R}_2 - \hat{R}_1$. Then there exists $\pi \in \Gamma_3(x, u)$ such that*
(1) *$D = \Delta(x, y) \cap \Gamma_1(\pi)$ is a clique of order 6 and $L(x, y) \cap \Gamma_2(x, \pi)$ is a disc in $\Gamma_2(x, \pi)$.*
(2) *$D_1(xuy) \cap m \neq \emptyset$ for all $m \in \Gamma_2(x, \pi)$.*

Proof. We first show that if $v \in \Delta(x, u, y)$ with $u+v \in \delta_i(u, u+x) \cap \delta_1(u, u+y)$ for $i = 1$ or 2, then $(x, v, y) \in \hat{R}_2 - \hat{R}_1$. First by (5.6), $v+u \in \delta_i(v, v+x) \cap \delta_1(v, v+y)$. Let $\sigma \in \Gamma_3(x, v, u)$. As $v+u \in \delta_1(v, v+y)$, $v+y \in \kappa_1(u, \sigma)$, in the language of Sect. 3. Thus by (3.13.1), $v+x \notin \delta_4(v, v+y)$. Further if $x, y \in \Gamma_1(\pi)$ for some plane π then by (5.7), $\Delta(x, y) \subseteq \Gamma_1(\pi)$, contradicting $(x, u, y) \notin \hat{R}_1$. So the claim is established.

By (3.10) there exist $\pi \in \Gamma_3(x+u)$ and $l \in \Gamma_2(u, \pi)$ with $l \in \delta_1(u, u+x) \cap \delta_1(u, u+y)$. Further there are lines $m_i(x) \in \delta_1(u, u+y) \cap \delta_2(u, u+x)$, $i=1, 2$. Now by (7.1), $m_i(x)$ contains 3 points $u, x_{i,j}, j=1, 2$, of $\Delta(y, x)$. Claim

$$(*) \qquad |l \cap \Delta(x, y)| > 1.$$

For $|\Delta(x) \cap l| = 3$ and for $u \neq v \in \Delta(x) \cap l$, $\Delta(v) \cap y+u \neq \{u\}$. But $Q(u)_x$ is transitive on $y+u-\{u\}$ by 5.4, so $(*)$ follows.

Next as $m_i(x) \in \delta_2(u, u+x)$, $L(x, m_i(x))$, $i=1, 2$, are the two ovals in $\Gamma_2(x, \pi)$ through $x+u$ by (7.3) and (2.1). As this holds for each $u \in D = \Delta(x, y) \cap \Gamma_1(\pi)$, (2.2) implies that either (2) holds or $|D|=9$. But as $y \notin \Gamma_1(\pi)$, D is a clique by (5.7). Then by (3.10), $\{u+d: d \in D\}$ is a singular line of lines in $\Gamma_1(u, \pi)$. Thus if $a, b \in D$ with $u+a \neq u+b$, we have $u+b \in \delta_1(u, u+a)$, so $|D \cap u+a| \leqq |\Delta(b) \cap u+a| = 3$, and hence $|D| \leqq 1+2 \cdot 3 = 7$.

Thus (1) holds. Finally let $m \in \Gamma_2(x, \pi)$; we must show $D_1(xuy) \cap m \neq \emptyset$. By (7.3), $L(x, l)$ is a singular line in the symplectic space $\Gamma_2(x, \pi)$. We may assume m is not on this line as then it contains the point $l \cap m$ coplanar with y. Also we have just seen there are two points $u, u' \in \Delta(x, y) \cap l$, and as m is not on the line $L(x, l)$, m is adjacent to at most one of $x+u$, $x+u'$, so $m \in \delta_2(x, x+v)$ for $v \in \{u, u'\}$. Then by (7.3), $L(v, m)$ is an oval in $\Gamma_2(v, \pi)$. But by (3.10), $\delta_1(v, v+y) \cap \Gamma_2(v, \pi)$ is a singular line in $\Gamma_2(v, \pi)$, so as the oval meets each singular line there is $k \in L(v, m) \cap \Gamma_2(v, \pi) \cap \mathrm{Co}(v+y)$. Then $k \cap m \in D_1(xuy)$, completing the proof.

(7.5) *Let $y \in \Delta_1^2(x)$. Then*
(1) *There exists a unique $\pi \in \Gamma_3(x, y)$.*
(2) *$\Delta(x, y) \subseteq \Gamma_1(\pi)$ and $\Delta(x, y)$ is connected.*
(3) *$|\Delta(x, y) \cap m| \geqq 3$ for all $m \in \Gamma(x, \pi)$.*

Proof. Part (1) holds by (5.3.2), while part (2) follows from (5.7) and (3) from (7.1).

(7.6) *Let $\pi \in \Gamma_3(x)$ and $S \subseteq \Gamma_1(x, \pi)$. Then*
(1) *If $S \cap \Gamma_1(l) \neq \emptyset$ for all $l \in \Gamma_2(x, \pi)$ then $d_{R_1}(S, C\Delta(xuy)) \leqq 1$ for all $y \in \Delta^2(x)$ and $u \in \Delta(x, y)$.*
(2) *$d_\Delta(\Delta(x, y), C\Delta(xuv)) \leqq 1$ for all $y \in R_1(x) \cap \Delta^2(x)$, $v \in \Delta^2(x)$, and $u \in \Delta(x, v)$.*

Proof. Let $v \in \Delta^2(x)$ and $u \in \Delta(x, v)$. We first observe there exists $U \in \Gamma_3(x, u)$ such that

$$D = \{d \in \Gamma_2(x, U): d \cap u^\perp \cap \Delta(x, y) \neq \emptyset\}$$

contains a disc or oval. This follows from (7.4) and (7.5) unless $(x, u, v) \in \hat{R}_3 - \hat{R}_2$, where it follows from (3.13.2) and (7.3.2).

By the observation and (3.11.2), there exists $k \in \Gamma_2(x, \pi)$ and $d \in D$ with k and d coplanar. Let $w \in \Delta(x, v) \cap \Gamma_1(d) \cap u^\perp$. Then as $|\Delta(w) \cap k| \geqq 3$ and $|k|=5$, if $|S \cap k| \geqq 3$ then $S \cap k \cap \Delta(w) \neq \emptyset$. Therefore (2) follows from (7.5.3). Similarly if $s \in S \cap k$ then s is coplanar with w, so that (1) is established.

(7.7) *Let $(x, u, y) \in \hat{R}_3 - \hat{R}_2$. Then there exists $\pi \in \Gamma_3(x, u)$ such that $D_2(xuy) \cap \Gamma_1(m) \neq \emptyset$ for all $m \in \Gamma_2(x, \pi)$.*

Proof. By (3.13), there is $l \in \delta_2(u, u+x) \cap \delta_2(u, u+y)$. Let $\pi \in \Gamma_3(x, l)$. By (7.3), $L(x, l)$ is an oval in the symplectic space $\Gamma_2(x, \pi)$. Thus there exists $v \in l$ with $m \in \delta_2(x,$

$x+v$). Then by (7.2) and (7.3), $L(v, m)$ is an oval in $\Gamma_2(v, \pi)$. Hence by (3.11.1), m contains a member of $D_2(xuy)$.

(7.8) *Let $y \in \Delta(x)$. Then either*
(1) $d_{Co}(\Gamma_2(x, \pi), x+y)=1$ *and z fixes y, or*
(2) $d_{Co}(\Gamma_2(x, \pi), x+y)>1$ *and z moves y.*

Proof. Suppose $l \in \Gamma_2(x, \pi)$ is coplanar with $x+y$. Then there is $u \in \Gamma_1(l) \cap \Delta(y)$ distinct from x and as z fixes all lines through x and u, z fixes $x+y \cap u+y=\{y\}$.

Conversely assume $x+y$ is coplanar with no member of $\Gamma_2(x, \pi)$. By (3.8), $H \cap M$ is transitive on the set of lines m coplanar with no line in $\Gamma_2(x, \pi)$ and by (5.4.6), E is transitive on $m-\{x\}$, so if z fixes y then z fixes $\Delta(x)$ pointwise, which is not the case.

Theorem 7.9 *Δ is triangulable.*

Proof. By (1.4) it suffices to show Hypothesis B of Sect. 1 is satisfied. By construction the chain $\hat{R}_0 \subseteq \ldots \subseteq \hat{R}_3$ satisfies Hypothesis (B1). By (7.4.3) and (7.5.2), (7.7), and (7.6.1), (B2) is satisfied. Thus it remains only to verify (B3).

Let $y \in \Delta_1^2(x)$ and $w \in \Delta(y) \cap \Delta^2(x)$. We must show $d_\Delta(\Delta(x, y), w) \leqq 1$. By (5.3.2), $\Gamma_3(x, y)=\{\pi\}$. If $d_{Co}(\Gamma_2(y, \pi), y+w) \leqq 1$ then (7.5.3) and (7.1) complete the proof, so assume otherwise. Then by (7.8), z does not fix w.

By (7.6.2) and (7.5.2), there is $v \in \Delta(x, w)$ with $d_{Co}(\Gamma_2(x, \pi), x+v) \leqq 1$. Then by (7.8), z fixes v. Now as $Q(y)^{\Gamma_1(y+w)}$ is regular on $\Gamma_1(y+w)$ and z does not fix $\Gamma_1(y+w)$ pointwise, y is the unique fixed point of z on $\Gamma_1(y+w)$. Further $w, wz \in \Gamma_1(y+w) \cap \Delta(v)$, so $|\Gamma_1(y+w) \cap \Delta(v)|=3$ or 5, and hence $\Delta(v) \cap \Gamma_1(y+w) = \Gamma_1(y+w)$ or $\{y, w, wz\}$. In either case $v \in \Delta(x, y, w)$, and the theorem holds.

8 Proof of the Main Theorem

We continue the notation and hypothesis of Sect. 5. In addition let $I=H_x$, $\Delta_H=\Gamma_1(\pi)$, $x \neq y \in \Gamma_1(l)$, and $\mathscr{U}=(G, H, \Delta, \Delta_H)$.

We wish to prove that, up to isomorphism, G is the unique group of type J_4. We will do so by appealing to the Corollary to Theorem 1 in [3]. In particular the reader is referred to [3] for the definition of various concepts like *uniqueness system*, etc.

(8.1) *$\mathscr{U}$ is a uniqueness system.*

Proof. This follows from Theorem 7.3 in [3]. To see that Hypotheses (Γ0) through (Γ5) of Sect. 7 of [3] are satisfied, use (5.1.1), (5.2), and (5.5).

(8.2) (1) $|\mathrm{Aut}(H): \mathrm{Inn}(H)|=2$ *and* $\mathrm{Aut}(H)=\mathrm{Inn}(H)\langle\alpha\rangle$ *where* $O^2(H)=C_H(\alpha)$.
(2) $|\mathrm{Aut}(M): M|=2$ *and* $\mathrm{Aut}(M)=M\langle\beta\rangle$, *where* $E\langle\beta\rangle=O_2(\mathrm{Aut}(M)) \cong E_{2^{12}}$.
(3) *Hypothesis* V *of* [3] *is satisfied.*

Proof. Let α be the automorphism of H with $C_H(\alpha)=O^2(H)$ and $h\alpha=hz$ for $h \in H-O^2(H)$. As $[H, \alpha]=\langle z\rangle \leqq I$, α acts on I and the restriction α_I of α to I is in $\mathrm{Aut}(I)$.

By (4.8) there is a complement L to E in M and E is the Todd module for L. Thus there is a 12-dimensional indecomposible E_0 for L with $E=[E_0, L]$.

Pick $\beta \in E_0 - E$. Here $[\beta, M] = E \leqq I$, so β acts on I and the restriction β_I of β to I is in $\mathrm{Aut}(I)$. Indeed $[\beta, I] \not\leqq Q$.

We will show that

(*) $\mathrm{Aut}(I) = \mathrm{Inn}(I)\langle \alpha_I \rangle \langle \beta_I \rangle$, $\mathrm{Aut}(H) = \mathrm{Inn}(H) N_{\mathrm{Aut}(H)}(I)$, and $\mathrm{Aut}(M) = MN_{\mathrm{Aut}(M)}(I)$.

Notice (*) implies (1) and (2), since I contains a Sylow 2-subgroup of H and M, so by (1.1) in [3], $C_{\mathrm{Aut}(H)}(I) \leqq \mathrm{Inn}(H)$ and $C_{\mathrm{Aut}(M)}(I) \leqq M$. Of course (*) also shows $\mathrm{Aut}(I) = \mathrm{Aut}_{\mathrm{Aut}(H)}(I)\,\mathrm{Aut}_{\mathrm{Aut}(M)}(I)$, which is (V 1). As I is maximal in H, $I = N_H(I)$. Therefore (*) implies (V 3) of Hypothesis V as $[\alpha, H] \leqq I$.

Next $M_y = M_l^\infty N_{M_l}(P)$ for $P \in \mathrm{Syl}_3(C_{M_l}(M_l^\infty / O_2(M_l^\infty)))$, so we may choose β to act on M_y, and hence (*) implies (V 2). Finally I acts on the 15 lines in $\Gamma_2(x, \pi)$ with l the unique fixed point of I_y, so $N_I(I_y) \leqq I_l$. Similarly x, y are the fixed points of I_y on $\Gamma_1(l)$, so $I_y = N_I(I_y)$, establishing (V 4) and completing the verification of Hypothesis V.

So it remains to establish (*). As $L = \mathrm{Aut}(L)$, $\mathrm{Aut}(M) = MC_{\mathrm{Aut}(M)}(M/E)$ and thus as $E \leqq I$, we have $\mathrm{Aut}(M) = MN_{\mathrm{Aut}(M)}(I)$. Similarly $\mathrm{Aut}(H) = \mathrm{Inn}(H) N_{\mathrm{Aut}(H)}(I)$. Let $B = \mathrm{Aut}(I)$, $S = O_2(I)$, $X \in \mathrm{Syl}_3(O_{2,3}(I))$, and $A = C_B(I/Q) \cap N(X)$. It remains to show $B = \langle \alpha_I \rangle \langle \beta_I \rangle \mathrm{Inn}(I)$.

First Q is the unique subgroup of a Sylow 2-subgroup T of H of its isomorphism type, so Q char I. Also $S/Q \cong E_{16}$ is the natural module for $I/SX \cong Sp_4(2)$, so I/SX is self normalizing in $\mathrm{GL}(S/Q)$ and $\dim(H^1(I/SX, S/Q)) = 1$. Therefore $|B : \mathrm{Inn}(I) C_B(I/Q)| \leqq 2$, so as $[\beta, I] \not\leqq Q$, we conclude $B = \langle \beta_I \rangle \mathrm{Inn}(I) C_B(I/Q)$. So it suffices to show $C_B(I/Q) = \mathrm{Aut}_Q(I)\langle \alpha_I \rangle$. But $C_B(I/Q) = \mathrm{Aut}_Q(I) A$ by a Frattini argument. Thus we must show $A = \langle \alpha_I \rangle$.

Now a complement J/Q to S/Q in I/Q has two inequivalent irreducibles of degree 6 on $Q/\langle z \rangle$ which are absolutely irreducible over GF(2). Then as A commutes with J/Q, A centralizes each of these irreducibles, so $[A, Q] \leqq \langle z \rangle$. Then $[A, N_I(X)] \leqq N_I(X) \cap C(Q/\langle z \rangle) = \langle z \rangle$, so A centralizes $O^2(N_I(X))$. Further as Q is extraspecial, $A\,\mathrm{Aut}_Q(I) = A_0\,\mathrm{Aut}_Q(I)$, where $A_0 = C_{A\,\mathrm{Aut}_Q(I)}(Q)$. Also $[X, A_0] \leqq C_I(Q) = \langle z \rangle$, so as X char $X\langle z \rangle$, $[A_0, X] = 1$ and hence $A = A_0$. Thus A centralizes $O^2(N_I(X))Q = O^2(I)$. Then $[A, I] \leqq C_i(Q) = \langle z \rangle$ and $|I : O^2(I)| = 2$, so $|A| \leqq 2$. Finally as $\alpha_I \in A$ is of order 2, $A = \langle \alpha_I \rangle$, completing the proof of the lemma.

We now begin the proof of the Main Theorem. Assume $\bar{G}$ is a second group of type J_4 with subgroups $\bar{H}$, $\bar{M}$, etc. By (8.1) applied to $\bar{G}$, we have a uniqueness system $\bar{\mathscr{U}} = (\bar{G}, \bar{H}, \bar{\Delta}, \bar{\Delta}_{\bar{H}})$. Now

(8.3) *There exists an isomorphism* $\zeta: H \to \bar{H}$ *with* $I\zeta = \bar{I}$ *and* $H(\{x, y\})\zeta = \bar{H}(\{\bar{x}, \bar{y}\})$.

Namely by (4.9), $N_H(X)$ is the split extension of $C_H(X)$ by an involution such that $C_H(X)$ is quasisimple with center $X\langle z \rangle$ and $N_H(X)/X\langle z \rangle \cong \mathrm{Aut}(M_{22})$. So as the Schur multiplier of M_{22} is cyclic, $N_H(X)$ is determined up to isomorphism. Similarly the representation of $N_H(X)$ on $Q/\langle z \rangle$ is determined up to equivalence (cf. James [J]) so $H = H_0/\langle z z_0 \rangle$ is determined up to isomorphism, where H_0 is the semidirect product of a copy Q_0 of Q with center $\langle z_0 \rangle$ by $N_H(X)$ with the representations of $N_H(X)$ on Q and Q_0 equivalent.

Thus there is an isomorphism $\zeta: H \to \bar{H}$. Further as I is determined up to conjugacy in H by its isomorphism type, we can choose $I\zeta = \bar{I}$. Similarly H_l is determined up to conjugacy under I by its isomorphism type and the fact

that I_l contains a Sylow 2-subgroup of H, so we may take $H_l \zeta = \bar{H}_{\bar{l}}$. Thus ζ induces an equivalence of the actions of H on $\Gamma(\pi)$ and $\bar{H}$ on $\Gamma(\bar{\pi})$, so $H(\{x, y\})\zeta = \bar{H}(\{\bar{x}, \bar{y}\})$ for suitable $\bar{y} \in \Gamma(\bar{l})$.

A similar argument using 4.8 shows:

(8.4) *There exists an isomorphism* $\alpha: M \to \bar{M}$ *with* $I\alpha = \bar{I}$ *and* $M_y \alpha = \bar{M}_{\bar{y}}$.

To see that we may choose $M_y \alpha = \bar{M}_{\bar{y}}$ use the description of M_y in the proof of (8.2).

Now (8.2.3), (8.3), and (8.4) establish the hypotheses of Theorem 4 of [3], so that result shows $\mathscr{U}$ is similar to $\bar{\mathscr{U}}$. Then by (7.9.3) in [3], $\mathscr{U}$ is equivalent to $\bar{\mathscr{U}}$. This fact together with (7.9) and (5.5) establish the hypotheses of the Corollary to Theorem 1 of [3], and that result says $G \cong \bar{G}$, completing the proof of the Main Theorem.

References

1. Aschbacher, M.: Finite Group Theory. Cambridge: Cambridge University Press, 1986
2. Aschbacher, M.: The geometry of trilinear forms. In: W. Kantz (ed.) Finite Geometries, Buildings, and Related Topics, pp. 75–84. Oxford: Oxford University Press, 1990
3. Aschbacher, M., Segev, Y.: Extending morphisms of groups and graphs. Ann. Math. (to appear)
4. Held, D.: The simple groups related to M_{24}. J. Algebra **13**, 253–296 (1969)
5. James, G.: The modular characters of the Mathieu groups. J. Algebra **27**, 57–111 (1973)
6. Janko, Z.: A new finite simple group of order 86, 775, 571, 046, 077, 562, 880 which possesses M_{24} and the full cover of M_{22} as subgroups. J. Algebra **42**, 564–596 (1976)
7. Norton, S.: The construction of J_4. Proc. Sym. Pure Math. **37**, 271–278 (1980)

Invent. math. 105, 609–611 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

On the length of an extremal rational curve*

Yujiro Kawamata

Department of Mathematics, University of Tokyo, Hongo, Bunkyo, Tokyo 113, Japan

Oblatum 20-XI-1990

The purpose of this paper is to supplement the cone theorem ([Ka1]) and prove the following (we refer to [KMM] for the notation):

Theorem 1 *Let $g: X \to Y$ be a projective morphism of algebraic varieties over a field of characteristic zero, and Δ a $\mathbb{Q}$-divisor on X such that the pair (X, Δ) has only log-terminal singularities. Let E be an irreducible component of the degenerate locus $\mathrm{Exc}(g) := \{x \in X;\ g$ is not an isomorphism at $x\}$, and let $n = \dim E - \dim g(E)$. Assume that $-(K_X + \Delta)$ is g-ample. Then E is covered by a family of rational curves $\{L_\lambda\}_{\lambda \in \Lambda}$ such that the $g(L_\lambda)$ are points and that $(-(K_X + \Delta) \cdot L_\lambda) < 2n$ (resp. $\leqq 2n$) if $E \neq X$ (resp. $E = X$).*

Let $f: X \to S$ be a projective morphism of algebraic varieties such that the pair (X, Δ) is log-terminal for a $\mathbb{Q}$-divisor Δ, and $g: X \to Y$ a contraction morphism associated to an extremal ray R for $f: X \to S$ with respect to the log-canonical divisor $K_X + \Delta$. Then g is an S-morphism which satisfies the conditions of Theorem 1. If X is smooth, S is a point and if $\Delta = 0$, then by Mori [M] and Ionescu [I], E is covered by a family of rational curves $\{L_\lambda\}_{\lambda \in \Lambda}$ such that $(-K_X \cdot L_\lambda) \leqq \dim X + 1 - 2 \operatorname{codim}_X E$. The number $(-K_X \cdot L)$ is called the *length* of L. We note that our theorem is new even in the case where X has only terminal singularities and $\Delta = 0$. As a corollary we obtain an alternative proof of the discreteness of extremal rays:

Corollary ([Ko]). *For $f: X \to S$ as above, let H be an f-ample Cartier divisor and ε a positive rational number. Then the number of extremal rays R such that $((K_X + \Delta + \varepsilon H) \cdot R) < 0$ is finite.*

Proof. These extremal rays are generated by rational curves L such that $(H \cdot L) < (-(K_X + \Delta) \cdot L)/\varepsilon \leqq 2/\varepsilon \cdot \dim X$. Q.E.D.

The key to the proof of Theorem 1 is the following lemma on the adjunction which is proved similarly as [Ka2, Theorem A.1].

Lemma. *Let $f: X \to Y$ be a projective birational morphism of normal algebraic varieties, H an f-ample Cartier divisor on X, Δ an effective $\mathbb{Q}$-divisor on X, E*

* Dedicated to Professor Heisuke Hironaka on his sixtieth birthday

an irreducible component of the degenerate locus $\mathrm{Exc}(f)$ *of* f, $n=\dim E$, *and* $\nu\colon \bar{E}\to E$ *the normalization. Suppose that the pair* (X, Δ) *has only log-terminal singularities and* $f(E)$ *is a point. Then*

$$(H^{n-1}\cdot(K_X+\Delta)\cdot E)>((\nu^*H)^{n-1}\cdot K_{\bar{E}}).$$

Proof. We may assume that Y is affine and H is very ample. We may also assume that $n=1$. In fact, if $n>1$, then we replace X, E and $\bar{E}$ by a general member X_1 of $|H|$, $E_1=E\cap X_1$ and the normalization $\bar{E}_1$ of E_1, respectively. Since we have $K_{X_1}=(K_X+H)|_{X_1}$ and $K_{\bar{E}_1}=(K_{\bar{E}}+\nu^*H)|_{\bar{E}_1}$, the assertion for X follows from that for X_1.

Suppose that $n=1$ and $((K_X+\Delta)\cdot E)\leqq \deg K_{\bar{E}}$. Then there exists a Cartier divisor A_0 on E such that $\deg A_0\geqq((K_X+\Delta)\cdot E)\in\mathbb{Q}$ and $H^0(\bar{E}, K_{\bar{E}}-\nu^*A_0)\neq 0$. By the trace map, we have $H^0(E, \omega_E(-A_0))\neq 0$. We can extend A_0 to a Cartier divisor A on X if we replace X by a small analytic neighborhood of E. Since $A-(K_X+\Delta)$ is f-nef, we have $R^1f_*\mathcal{O}_X(A)=0$ by [KMM, 1.2.5] and [N, §3]. Hence $H^1(E, A_0)=0$, and $H^0(E, \omega_E(-A_0))=0$ by the Serre duality, a contradiction. Q.E.D.

Proof of Theorem 1. Let Y_1 be an affine open subset of Y which intersects $g(E)$ and Y_0 a generic linear space section of Y_1 whose codimension is equal to $\dim g(E)$. We set $X_0=g^{-1}(Y_0)$, $g_0=g|_{X_0}$, $\Delta_0=\Delta\cap X_0$, and $E_0=E\cap X_0$. Let H be a g_0-ample Cartier divisor on X_0, $\nu\colon \bar{E}_0\to E_0$ the normalization, and C the intersection curve of $n-1$ general elements of $|\nu^*H|$. We also let $M=-\nu^*(K_X+\Delta)$, which is an ample $\mathbb{Q}$-Cartier divisor on $\bar{E}_0$.

First, we assume that $E\neq X$. Then by Lemma, we have $\deg_C(K_{\bar{E}_0}|_C)<(-M\cdot C)<0$. In the case where $n=1$, since $R^1g_{0*}\mathcal{O}_{X_0}=0$ by [KMM, 1.2.5], we have $C=\bar{E}_0=E_0\simeq\mathbb{P}^1$, and $(-(K_X+\Delta)\cdot C)<2$. If $n>1$, then by replacing H by its multiple if necessary, we may assume that C is not a rational curve. By [MM], for an arbitrary point x on C, there exists a rational curve L on $\bar{E}_0$ passing through x such that $(M\cdot L)\leqq 2n(M\cdot C)/(-K_{\bar{E}_0}\cdot C)<2n$.

In the case where $E=X$, if $n=1$, then $X_0\simeq\mathbb{P}^1$, while if $n>1$, then there exists a rational curve L through a general point of X_0 such that $(M\cdot L)\leqq 2n(M\cdot C)/(-K_{X_0}\cdot C)\leqq 2n$, since $(\Delta_0\cdot C)\geqq 0$. Q.E.D.

We have another application of Lemma.

Theorem 2 *Let* $f\colon X\to Y$ *be a projective surjective morphism,* Δ *an effective* $\mathbb{Q}$*-divisor on* X, *and* E *an irreducible component of* $\{x\in X;\ \dim_x f^{-1}f(x)>\dim X-\dim Y\}$. *Suppose that the pair* (X, Δ) *has only log-terminal singularities and that* $-(K_X+\Delta)$ *is* f*-nef. Then* E *is covered by a family of rational curves.*

Proof. We may assume that Y is affine and $f(E)$ is a point. Let $\nu\colon \bar{E}\to E$ be the normalization, H a very ample divisor on X, X_1 the intersection of d general members of $|H|$ for $d=\dim X-\dim Y$, $\Delta_1=\Delta\cap X_1$, $E_1=E\cap X_1$, and $\bar{E}_1=\nu^{-1}(E_1)$. If $d=0$ and $\dim E=1$, then $E\simeq\mathbb{P}^1$ as before. Otherwise, by Lemma $\bar{E}_1$ is covered by a family of curves C such that $(K_{\bar{E}_1}\cdot C)<((K_{X_1}+\Delta_1)\cdot\nu_*C)\leqq d(\nu^*H\cdot C)$. Hence $(K_{\bar{E}}\cdot C)<0$ and E is covered by rational curves by [MM]. Q.E.D.

References

[I] Ionescu, P.: Generalized adjunction and applications. Math. Proc. Camb. Philos. Soc. **99**, 457–472 (1986)

[Ka1] Kawamata, Y.: The cone of curves of algebraic varieties. Ann. Math. **119**, 603–633 (1984)

[Ka2] Kawamata, Y.: Moderate degenerations of algebraic surfaces. Proc. Symp. Bayreuth 1990 (to appear)

[KMM] Kawamata, Y., Matsuda, K., Matsuki, K.: Introduction to the minimal model problem. In: Oda, T. (ed.) Algebraic Geometry. Proc. Symp., Sendai 1985. (Adv. Stud. Pure Math. vol. 10, pp. 283–360) Tokyo: Kinokuniyo 1987

[Ko] Kollár, J.: The cone theorem: Note to [Ka1]. Ann. Math. **120**, 1–5 (1984)

[MM] Miyaoka, Y., Mori, S.: A numerical criterion of uniruledness. Ann. Math. **124**, 65–69 (1986)

[M] Mori, S.: Threefolds whose canonical bundles are not numerically effective. Ann. Math. **116**, 133–176 (1982)

[N] Nakayama, N.: The lower semi-continuity of the plurigenera of complex varieties. In: Oda, T. (ed.) Algebraic Geometry. Proc. Symp., Sendai 1985. (Adv. Stud. Pure Math. vol. 10, pp. 551–590) Tokyo: Kinokuniya 1987

Invent. math. 105, 613–639 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

On the packing densities of superballs and other bodies

N.D. Elkies [1,*], A.M. Odlyzko [2] and J.A. Rush [3,]**

[1] Department of Mathematics, Harvard University, One Oxford Street, Cambridge, Massachusetts 02138, USA
[2] AT & T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA
[3] Department of Mathematics, University of Edinburgh, James Clerk Maxwell Building, Edinburgh EH9 3JZ, Scotland, UK

Oblatum 25-IX-1990

Summary. A method of obtaining improvements to the Minkowski-Hlawka bound on the lattice-packing density for many convex bodies symmetrical through the coordinate hyperplanes, described by Rush [18], is generalized so that centrally symmetric convex bodies can be treated as well. The lower bounds which arise are very good.

The technique is applied to various shapes, including the classical l_σ-ball,

$$\{x \in \mathbf{R}^n : |x_1|^\sigma + |x_2|^\sigma + \ldots + |x_n|^\sigma \leqq 1\},$$

for $\sigma \geqq 1$. This generalizes the earlier work of Rush and Sloane [17] in which σ was required to be an integer. The superball above can be lattice packed to a density of $(b/2)^{n+o(1)}$ for large n, where

$$b = \sup_{t>0} \frac{\int_{x=-\infty}^{\infty} e^{-|tx|^\sigma} dx}{\sum_{k=-\infty}^{\infty} e^{-|tk|^\sigma}}.$$

This is as good as the Minkowski-Hlawka bound for $1 \leqq \sigma \leqq 2$, and better for $\sigma > 2$.

An analogous density bound is established for superballs of the shape

$$\{x \in \mathbf{R}^n : f(x_1, \ldots, x_k)^\sigma + f(x_{k+1}, \ldots, x_{2k})^\sigma + \ldots + f(x_{n-k+1}, \ldots, x_n)^\sigma \leqq 1\},\ k \mid n,$$

where f is the Minkowski distance function associated with a bounded, convex, centrally symmetric, k-dimensional body.

* Supported by the Harvard Society of Fellows and NSF Grant DMS-87-18965.
** Supported by a Seggie-Brown research fellowship; *current address:* Department of Mathematics GN-50, University of Washington, Seattle, WA 98195, USA.

Finally, we consider generalized superballs for which the defining inequality need not even be homogeneous. For these bodies as well, it is often possible to improve on the Minkowski-Hlawka bound.

1 Introduction

For our purposes, *bodies* will always be closed, bounded, convex subsets of $\mathbf{R}^n$, centrally symmetric, and with positive volume.

A basic problem in the geometry of numbers is the estimation of the maximum lattice-packing density δ_L of a given body. This is the problem of maximizing the fraction of $\mathbf{R}^n$ which is occupied by a lattice Λ of disjoint translates of the body. Overlap on the boundaries is disregarded. For a body G,

$$\delta_L(G)=\frac{\mathrm{Vol}(G)}{\inf \det(\Lambda)},$$

the infimum being taken over packing lattices Λ of G. Under a nonsingular linear transformation, δ_L is preserved, since the volume of the body and the determinant of the packing lattice change proportionally.

The quantity δ_L is known exactly for only a few bodies; and the greater the dimension n, the fewer the bodies for which it is known. For spheres, δ_L is known in dimensions 1 through 8. Usually we can only hope for an estimate.

The first major lower bound on lattice-packing density, valid in any number of dimensions, was Minkowski's bound

$$\delta_L(G)\geqq 2^{1-n}\,\zeta(n),$$

which he proved by a tour de force in his own reduction theory of quadratic forms, in the case that G is an n-dimensional sphere.

Hlawka [7] proved that this holds not only for spheres but for all bodies (as we define them), thus verifying a previously unproven assertion of Minkowski. This is the famous *Minkowski-Hlawka bound* [2, p. 175; 6, p. 199; 10, p. 145; 16].

Rogers [16] proved that for spheres

$$\delta_L\geqq\frac{n\,2^{1-n}\,\zeta(n)}{e(1-e^{-n})},$$

improving on Minkowski's bound by a factor approaching n/e for large n.

Rush and Sloane [17] found that the l_σ-balls having $\sigma=3, 4, 5$, etc., have

$$\delta_L\geqq\gamma_\sigma^n,$$

in which each γ_σ exceeds 1/2. This was the first *essential* improvement, i.e., improvement by a factor exponential in n, for these bodies.

The article [18] mentioned in the abstract described a code-theoretic method, based on Construction A of Leech and Sloane [9], giving rise to dense lattice

packings of what were called *bodies of type H*. Such a body is compact, convex, symmetric through the coordinate hyperplanes, and has the unit vectors

$$e_1=(1, 0, \ldots, 0), \quad e_2, \ldots, e_n$$

on its surface. It was claimed in the introduction to that paper that often the method yields asymptotic densities much higher than

$$2^{-n+o(n)},$$

as assured by the Minkowski-Hlawka bound. However, there has been no published evidence for this alleged oftenness; and considering the scarcity, indeed the veritable absence, of such examples[1] in the literature, the shapes discussed hereinafter with very high packing densities will be of interest.

Here are some bodies and our estimates for their lattice-packing densities δ_L as $n\to\infty$. All the error factors $\exp o(1)$ can be removed by allowing nonlattice packings.

Bodies defined by Cartesian products. Suppose that J is a bounded, convex, centrally symmetric body with maximum lattice-packing density $\delta_L(J)$ in $\mathbf{R}^k$, centered at the origin, and let $f(x)=\inf_{\mu\geqq 0}\{\mu: x\in\mu J\}$. The n-dimensional body given by

$$\max(f(x_1, \ldots, x_k), f(x_{k+1}, \ldots, x_{2k}), \ldots, f(x_{n-k+1}, \ldots, x_n))\leqq 1, \quad k|n,$$

can be packed to a density $\delta_L(J)^{n/k+o(n)}$. The result itself is not surprising; indeed one gets density $\delta_L(J)^{n/k}$ by using the orthogonal sum of n/k copies of the closest-packing lattice for J. But it is interesting that our method gives essentially the same bound. It would be nice to know whether $\delta_L(J)^{n/k}$ is the best possible. (See the third open problem in Sect. 9.)

Cross polytopes, Euclidean spheres, and other l_σ-balls.[2] If $\sigma\geqq 1$ and

$$G=\{x=(x_1, \ldots, x_n)\in\mathbf{R}^n: |x_1|^\sigma+\ldots+|x_n|^\sigma\leqq 1\}$$

then

$$\delta_L(G)\geqq 2^{-c_\sigma n+o(1)} \quad \text{as} \quad n\to\infty$$

for certain constants c_σ. It will be found that c_σ is less than one whenever $\sigma>2$, resulting in an improvement over the Minkowski-Hlawka bound for all such l_σ-balls. As $\varepsilon\to 0+$,

$$\varepsilon^{-1}(\ln 1/\varepsilon)^{3/2}(1-c_{2+\varepsilon})\to\frac{\sqrt{\pi}\,\zeta(3)}{2\ln 2}$$

which shows the behavior of the bound near $\sigma=2$.

---

[1] Granted, examples can be improvised, say as gently jostled tilings, or by nestling a body snugly within each tile. But we exclude such artful contrivances, which violate the spirit of the packing problem: Given a *body* find a dense *packing* thereof, not vice versa.

[2] This example from [17, 18] is not appearing here for the first time, though the exponent in those papers had an error term $O(n/\ln n)$ that is improved here to $o(1)$ and eliminated entirely for nonlattice packings. The bodies called *superballs* there are here called l_σ-balls, and are subsumed as a special case. We shall use the term superball in a wider sense. (See Sect. 3.)

Some numerical values of c_σ are given in Table 1.

Table 1. Truncated decimal values of c_σ for some l_σ-balls

σ	c_σ
$1 \leqq \sigma \leqq 2$	1
2.001	0.9999641166
2.01	0.999438314
2.1	0.989213040
3	0.822600380
4	0.674242663
5	0.569240536
50	0.074523381
$\sigma \to \infty$	$c_\sigma \sim \frac{\ln \ln \sigma}{\sigma \ln 2} \to 0$

Minkowskian complex superballs. These bodies were of interest to Minkowski [14, I, p. 261] in connection with discriminants of algebraic number fields.

Let $G \subseteq \mathbf{R}^n = \mathbf{R}^{\rho+2s}$ be given by the inequality

$$|x_1|^\sigma + \ldots + |x_\rho|^\sigma + (y_1^2 + z_1^2)^{\sigma/2} + \ldots + (y_s^2 + z_s^2)^{\sigma/2} \leqq 1$$

where $\sigma \geqq 1$. (See [6, p. 411].) Let $\rho/n = a_1$ and $2s/n = a_2$, so that $a_1 + a_2 = 1$. Then $\delta_L(G) \geqq 2^{-c_\sigma n + o(1)}$, where

$$c_\sigma = 1 + a_1 \log_2 \left(\left(\ln \frac{1}{r_1} \right)^{1/\sigma} \frac{\theta_\sigma(r_1)}{2\Gamma(1+1/\sigma)} \right) + \frac{a_2}{2} \log_2 \left(\left(\ln \frac{1}{r_2} \right)^{2/\sigma} \frac{\sqrt{3}\,\psi_\sigma(r_2)}{2\pi\Gamma(1+2/\sigma)} \right) \tag{1}$$

in which

$$\Gamma(1+z) = z! = \int_0^\infty t^z e^{-t}\, dt \quad \text{for } z > 0,$$

$$\theta_\sigma(z) = \sum_{m=-\infty}^{\infty} z^{|m|^\sigma},$$

$$\psi_\sigma(z) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} z^{(k^2+kl+l^2)^{\sigma/2}},$$

and r_1, r_2 are positive numbers chosen to minimize c_σ in (1).

Bodies based on the diamond shape. Let $G \subseteq \mathbf{R}^n = \mathbf{R}^{\rho+2s}$ be given by the inequality

$$|x_1|^\sigma + \ldots + |x_\rho|^\sigma + (|y_1| + |z_1|)^\sigma + \ldots + (|y_s| + |z_s|)^\sigma \leqq 1$$

where $\sigma \geqq 1$. Let $\rho/n = a_1$ and $2s/n = a_2$, so that $a_1 + a_2 = 1$. Then $\delta_L(G) \geqq 2^{-c_\sigma n + o(1)}$, where

$$c_\sigma = 1 + a_1 \log_2 \left(\left(\ln \frac{1}{r_1} \right)^{1/\sigma} \frac{\theta_\sigma(r_1)}{2\Gamma(1+1/\sigma)} \right) + \frac{a_2}{2} \log_2 \left(\left(\ln \frac{1}{r_2} \right)^{2/\sigma} \frac{\psi_\sigma(r_2)}{4\Gamma(1+2/\sigma)} \right) \tag{2}$$

in which $\theta_\sigma(z)$ is as in the previous example,

$$\psi_\sigma(z) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} z^{\max(|k|,|l|)^\sigma} = 1 + \sum_{m=1}^{\infty} 8m z^{m^\sigma},$$

and r_1, r_2 are positive numbers chosen to minimize c_σ in (2).

It was mentioned in [18] that the case $a_1 = 0$, $a_2 = 1$ of the above body admitted an essential improvement to the Minkowski-Hlawka bound for $\sigma = 6$, 7, 8, etc. Here we do even better. By considering instead the body

$$\max(|x_1|, |x_2|)^\sigma + \ldots + \max(|x_{n-1}|, |x_n|)^\sigma \leqq 1,$$

which, being linearly equivalent, has the same lattice-packing density, we can improve on both the Minkowski-Hlawka bound and its cited improvement, for all σ greater than about 2.134485.

Bodies based on the cube. Let $G \subseteq \mathbf{R}^n = \mathbf{R}^{\rho+3s}$ be given by the inequality

$$|w_1|^\sigma + \ldots + |w_\rho|^\sigma + \max(|x_1|, |y_1|, |z_1|)^\sigma + \ldots + \max(|x_s|, |y_s|, |z_s|)^\sigma \leqq 1$$

where $\sigma \geqq 1$. Let $\rho/n = a_1$ and $3s/n = a_2$, so that $a_1 + a_2 = 1$. Then $\delta_L(G) \geqq 2^{-c_\sigma n + o(1)}$, where

$$c_\sigma = 1 + a_1 \log_2 \left(\left(\ln \frac{1}{r_1} \right)^{1/\sigma} \frac{\theta_\sigma(r_1)}{2\Gamma(1+1/\sigma)} \right) + \frac{a_2}{3} \log_2 \left(\left(\ln \frac{1}{r_2} \right)^{3/\sigma} \frac{\psi_\sigma(r_2)}{8\Gamma(1+3/\sigma)} \right) \tag{3}$$

in which $\theta_\sigma(z) = \sum_{k \in \mathbf{Z}} z^{|k|^\sigma}$,

$$\psi_\sigma(z) = \sum_{(j,k,l) \in \mathbf{Z}^3} z^{\max(|j|,|k|,|l|)^\sigma} = 1 + \sum_{m=1}^{\infty} ((2m+1)^3 - (2m-1)^3) z^{m^\sigma},$$

and r_1, r_2 are positive numbers chosen to minimize c_σ in (3).

Bodies based on the sphere. Let $G \subseteq \mathbf{R}^n = \mathbf{R}^{\rho+3s}$ be given by

$$|w_1|^\sigma + \ldots + |w_\rho|^\sigma + (x_1^2 + y_1^2 + z_1^2)^{\sigma/2} + \ldots + (x_s^2 + y_s^2 + z_s^2)^{\sigma/2} \leqq 1$$

where $\sigma \geqq 1$. Let $\rho/n = a_1$ and $3s/n = a_2$, so that $a_1 + a_2 = 1$. Then $\delta_L(G) \geqq 2^{-c_\sigma n + o(1)}$, where

$$c_\sigma = 1 + a_1 \log_2 \left(\left(\ln \frac{1}{r_1} \right)^{1/\sigma} \frac{\theta_\sigma(r_1)}{2\Gamma(1+1/\sigma)} \right) + \frac{a_2}{3} \log_2 \left(\left(\ln \frac{1}{r_2} \right)^{3/\sigma} \frac{3\sqrt{2}\,\psi_\sigma(r_2)}{8\pi\Gamma(1+3/\sigma)} \right) \tag{4}$$

in which $\theta_\sigma(z) = \sum_{k \in \mathbf{Z}} z^{|k|^\sigma}$,

$$\psi_\sigma(z) = \sum_{(j,k,l) \in \mathbf{Z}^3} z^{(j^2+k^2+l^2+jk+kl)^{\sigma/2}},$$

and r_1, r_2 are positive numbers chosen to minimize c_σ in (4).

Bodies based on the hypercube. Let $G \subseteq \mathbf{R}^n = \mathbf{R}^{\rho+4s}$ be given by

$$|v_1|^\sigma + \ldots + |v_\rho|^\sigma + \max(|w_1|, |x_1|, |y_1|, |z_1|)^\sigma + \ldots + \max(|w_s|, |x_s|, |y_s|, |z_s|)^\sigma \leqq 1$$

where $\sigma \geqq 1$. Let $\rho/n = a_1$ and $4s/n = a_2$, so that $a_1 + a_2 = 1$. Then $\delta_L(G) \geqq 2^{-c_\sigma n + o(1)}$, where

$$c_\sigma = 1 + a_1 \log_2 \left(\left(\ln \frac{1}{r_1} \right)^{1/\sigma} \frac{\theta_\sigma(r_1)}{2\Gamma(1+1/\sigma)} \right) + \frac{a_2}{4} \log_2 \left(\left(\ln \frac{1}{r_2} \right)^{4/\sigma} \frac{\psi_\sigma(r_2)}{16\Gamma(1+4/\sigma)} \right) \tag{5}$$

in which $\theta_\sigma(z) = \sum_{k \in \mathbf{Z}} z^{|k|^\sigma}$,

$$\psi_\sigma(z) = \sum_{(i,j,k,l) \in \mathbf{Z}^4} z^{\max(|i|,|j|,|k|,|l|)^\sigma} = 1 + \sum_{m=1}^{\infty} ((2m+1)^4 - (2m-1)^4)\, z^{m^\sigma},$$

and r_1, r_2 are positive numbers chosen to minimize c_σ in (5).

Table 2 lists some numerical values of these c_σ.

Table 2. Numerical values of c_σ for bodies based on various shapes. Decimals are truncated, and $a_2 = 1 - a_1$

Bodies based on the	σ	a_1	c_σ
Circle	$1 \leqq \sigma \leqq 2$	0	1
(Minkowskian complex	3	0	0.860949908
superballs)	3	1/2	0.843869375
	4	0	0.729110596
	4	1/2	0.703515560
	10	0	0.392842258
	∞	$0 \leqq a_1 \leqq 1$	$1 - a_1 + \frac{a_2}{2} \log_2 \left(\frac{\sqrt{3}}{2\pi}\right)$
Diamond	$1 \leqq \sigma \leqq 2.134485793$	0	1
	3	0	0.832284542
	4	0	0.685724360
	4	1/2	0.686104293
	5	0	0.581353081
	∞	$0 \leqq a_1 \leqq 1$	0
Cube	$1 \leqq \sigma \leqq 2.212882112$	0	1
	3	0	0.845323587
	3	1/2	0.864968538
	4	0	0.698698206
	5	0	0.594016546
	∞	$0 \leqq a_1 \leqq 1$	0
Sphere	$1 \leqq \sigma \leqq 2$	0	1
	3	0	0.895000891
	3	1/2	0.863338582
	4	0	0.780495356
	5	0	0.692044379
	6	0	0.624274637
	∞	$0 \leqq a_1 \leqq 1$	$1 - a_1 + \frac{a_2}{3} \log_2 \left(\frac{3\sqrt{2}}{8\pi}\right)$
Hypercube	$1 \leqq \sigma \leqq 2.285497723$	0	1
	3	0	0.858882563
	3	1/2	0.903475529
	4	0	0.711485684
	∞	$0 \leqq a_1 \leqq 1$	0

All the foregoing examples are special or limiting cases of Theorem 6, whose statement and proof we shall withhold until the sixth section, "The packing densities of superballs". In that section we also give an argument along the lines of [17] and [18] which gives slightly weaker bounds, with the same exponential growth but with error $\exp o(n)$ rather than $\exp o(1)$. Although the error term is worse, the method provides more information on how to find a dense lattice packing (Theorem 7), by specifying the parameter r in the $[n, k, r, p, G]$ codes to which Construction A is applied. Theorems 6 and 7 depend for their proofs upon the intermediate results of Sects. 2 through 5. These contain versions of our density bound valid for bodies generally, the definition of a superball, expressions for the volume of a superball, and a bound on the number of lattice points in a superball, respectively. Section 7 is devoted to the l_σ-ball, and Sect. 8 is concerned with the generalization of superballs to the inhomogeneous case

of mixed exponents. In Sect. 9 we state some questions left unanswered by this paper.

Besides the various references already cited throughout this introduction, see the articles [8, 11, 19–22] and the books [4, 13, 15] for related work.

2 Two general versions of the bound

A lattice Λ is said to be *admissible* for the body G, or to be G-admissible, if the body is centered at the origin of Λ, and has no further points of Λ in its interior $\mathrm{int}(G)$. Although the usage is nonstandard, it is sometimes syntactically convenient to interchange G and Λ. Thus, under the foregoing circumstances, we refer to G as a Λ-admissible body.

The results of this paper can be obtained from either of the two following theorems, which are asymptotically of the same strength. Indeed the proofs are based on the same fundamental idea, but with different emphases: as usual in the geometry of numbers one has the dual choices of fixing either a lattice (Theorem 1) or a body (Theorem 2).

Theorem 1 *Let p be an odd prime, and $0 \leqq r \leqq p$. A body $G \subseteq \mathbf{R}^n$, admissible for the lattice of integer points $\mathbf{Z}^n$, can be lattice packed with density*

$$\frac{2^{1-n} r^n \,\mathrm{Vol}(G)}{(p-1)\,\mathrm{card}(\mathbf{Z}^n \cap pQ \cap (p\mathbf{Z}^n + rG))}$$

where Q is the unit hypercube

$$\{x \in \mathbf{R}^n : \max(|x_1|, \ldots, |x_n|) \leqq 1/2\}.$$

Proof. We give only the merest sketch.

Provided

$$k < n + 1 - \log_p\left(\frac{p-1}{2}\,\mathrm{card}(\mathbf{Z}^n \cap pQ \cap (p\mathbf{Z}^n + rG))\right),$$

there exists an $[n, k, r, p, G]$ code C, say. That is, there is a k-dimensional subspace C of $GF(p)^n$, such that every nonzero codeword c in C has G-norm at least r. The *G-norm*,

$$\|\cdot\|_G : GF(p)^n \to \mathbf{R},$$

is defined as

$$\|x\|_G = \inf_{\mu \geqq 0} \{\mu : x \in p\mathbf{Z}^n + \mu G\},$$

where we identify $GF(p)^n$ in the obvious way with the set of points of $\mathbf{Z}^n$ lying within the hypercube pQ. If $r \leqq p$, then the Construction A lattice $p\mathbf{Z}^n + C$ provides a packing of $rG/2$, and the density of the packing is at least that asserted by Theorem 1. The reader can consult [18] for details of this proof, and [3, 9], regarding Construction A and its genesis.

Theorem 2 *If Λ is a lattice and G is a body centered at the origin of $\mathbf{R}^n$, then there are enough translates of G by vectors of 2Λ to pack $\mathbf{R}^n$ with density at least*

$$\frac{2^{-n}\,\mathrm{Vol}(G)}{\mathrm{card}(\Lambda \cap \mathrm{int}(G))\det(\Lambda)}.$$

Also, for any prime $p \geqq \mathrm{card}(\Lambda \cap \mathrm{int}(G)) - 1$, there exists a sublattice $\Lambda' \subset \Lambda$ of index $[\Lambda : \Lambda'] = p$, such that $2\Lambda'$ provides a lattice packing for G, with density at least $2^{-n}\,\mathrm{Vol}(G)\,p^{-1}\det(\Lambda)^{-1}$.

Proof. For the first assertion we can simply use any maximal collection of translates of G by lattice vectors of 2Λ with pairwise disjoint interiors. (That is, pack $\mathbf{R}^n$ arbitrarily with such translates until there is no room left.) Indeed, if we replace each translate $G+v$ in such a collection by the translate $2G+v$ of $2G$, every lattice point in 2Λ is contained in at least one of these copies of $2G$; but each copy contains only $\mathrm{card}(\Lambda \cap \mathrm{int}(G))$ such points, so the translation vectors v have density at least $1/\mathrm{card}(\Lambda \cap \mathrm{int}(G))$ in 2Λ, whence the translated copies of G have density at least $2^{-n}\,\mathrm{Vol}(G)/(\mathrm{card}(\Lambda \cap \mathrm{int}(G))\det(\Lambda))$ in $\mathbf{R}^n$.

For the second assertion, note first that none of the $(\mathrm{card}(\Lambda \cap \mathrm{int}(G)) - 1)$ nonzero lattice points v contained in G may be of the form pv' for some $v' \in \Lambda$, for then G would contain the $2p+1$ lattice points $0, \pm v', \pm 2v', \ldots, \pm pv'$, contradicting $p \geqq \mathrm{card}(\Lambda \cap \mathrm{int}(G)) - 1$. Thus, for each nonzero $v \in \mathrm{int}(G) \cap \Lambda$, only $(p^{n-1}-1)/(p-1)$ of the $(p^n-1)/(p-1)$ sublattices Λ' of index p in Λ contain v. (An index-p sublattice $\Lambda' \subset \Lambda$ corresponds naturally to a point in the projective space of dimension $(n-1)$ over $\mathbf{Z}/p\mathbf{Z}$ associated to the dual of $\Lambda/p\Lambda$, and the sublattices Λ' containing v correspond to the points in the codimension-1 subspace associated to those functionals $\Lambda/p\Lambda \to \mathbf{Z}/p\mathbf{Z}$ taking the image of v in $\Lambda/p\Lambda$ to zero.) Thus there are at most $(\mathrm{card}(\Lambda \cap \mathrm{int}(G)) - 1)((p^{n-1}-1)/(p-1))$ index-p sublattices containing some nonzero $v \in \mathrm{int}(G) \cap \Lambda$; since by our assumption on p this is less than the total number $(p^n-1)/(p-1)$ of these sublattices, we conclude that there exists some sublattice Λ' of index p in Λ that meets $\mathrm{int}(G)$ only at the origin, and therefore that the translates of G by $2\Lambda'$ have pairwise disjoint interiors, Q.E.D.

3 Distance functions and superballs

The *distance function*[3] of a body G is defined by $g(x) = \inf_{\mu \geqq 0}\{\mu : x \in \mu G\}$.

A distance function is continuous, nonnegative (zero only at the origin), and satisfies the triangle inequality $g(x+y) \leqq g(x) + g(y)$ and the homogeneity condition $g(tx) = |t|\,g(x)$ for all real t.

A *superball* is a body of the form

$$(6) \qquad \{x \in \mathbf{R}^n : f(x_1, \ldots, x_k)^\sigma + f(x_{k+1}, \ldots, x_{2k})^\sigma + \ldots + f(x_{n-k+1}, \ldots, x_n)^\sigma \leqq 1\}$$

where f is some k-dimensional body's distance function, and n is a multiple of k. In order to pack them successfully, we require that $\sigma \geqq 1$, lest the superball be nonconvex for $n > k$.

In Sect. 8 we shall introduce *generalized superballs*, in which a variety of σ's and f's may occur within one body.

[3] Also called the *Minkowski distance function* or *gauge function.*

4 Volume of a superball

In order to apply the density bounds of Sect. 2 to the superballs of Sect. 3, we need to determine their volume. The following theorem provides three important volume formulas.

Theorem 3 *Let* $\phi_j: \mathbf{R}^{m_j} \to \mathbf{R}$ *be distance functions,* $m_1 + \ldots + m_s = n$, $\sigma > 0$, *and*

$$V_j = \mathrm{Vol}\{x \in \mathbf{R}^{m_j}: \phi_j(x) \leqq 1\}, \quad j = 1, 2, \ldots, s.$$

Then the volume of the superball $f(x_1, \ldots, x_n) \leqq 1$, *where* f *is the distance function*

$$f = \sqrt[\tau]{\phi_1^\tau + \ldots + \phi_s^\tau},$$

is given by each of the following three expressions:

$$V_1 V_2 \ldots V_s \frac{\Gamma(1+m_1/\tau)\,\Gamma(1+m_2/\tau)\ldots\Gamma(1+m_s/\tau)}{\Gamma(1+n/\tau)}, \tag{7}$$

$$\Gamma(1+n/\sigma)^{-1} \iint\limits_{x \in \mathbf{R}^n} \ldots \int e^{-f(x)^\sigma}\, dx_1 \ldots dx_n, \tag{8}$$

and

$$\Gamma(1+n/\sigma)^{-1} \lim_{t \to 0+} t^{n/\sigma} \sum_{x \in \mathbf{Z}^n} e^{-t f(x)^\sigma}. \tag{9}$$

Proof. By means of the disk method from calculus, we can establish the first formula. Let V be the desired volume of $f \leqq 1$. We approximate the superball by a finite number of disks

$$x_1(j_1) \leqq \phi_1 \leqq x_1(j_1+1), \ldots, x_s(j_s) \leqq \phi_s \leqq x_s(j_s+1)$$

where

$$0 \leqq x_1 \leqq 1, \quad 0 \leqq x_2 \leqq \sqrt[\tau]{1-x_1^\tau}, \ldots, 0 \leqq x_s \leqq \sqrt[\tau]{1-x_1^\tau-\ldots-x_{s-1}^\tau},$$

of volumes

$$\prod_{i=1}^{s} (x_i(j_i+1)^{m_i} V_i - x_i(j_i)^{m_i} V_i).$$

For $i = 1, \ldots, s$ we let the largest of the $\Delta x_i(j_i) = x_i(j_i+1) - x_i(j_i)$ approach zero, and find that

$$\begin{aligned} V &= \int_{x_1=0}^{1} \int_{x_2=0}^{\sqrt[\tau]{1-x_1^\tau}} \cdots \int_{x_s=0}^{\sqrt[\tau]{1-x_1^\tau-\ldots-x_{s-1}^\tau}} d(V_1 x_1^{m_1}) \ldots d(V_s x_s^{m_s}) \\ &= V_1 \ldots V_s \iint\limits_{\substack{x_1,\ldots,x_s \geqq 0 \\ x_1^\tau + \ldots + x_s^\tau \leqq 1}} \ldots \int m_1 x_1^{m_1-1} \ldots m_s x_s^{m_s-1}\, dx_1 \ldots dx_s \\ &= V_1 \ldots V_s\, m_1 \ldots m_s\, \tau^{-s} \frac{\Gamma(m_1/\tau)\ldots\Gamma(m_s/\tau)}{\Gamma(1+m_1/\tau+\ldots+m_s/\tau)} \\ &= V_1 \ldots V_s \frac{\Gamma(1+m_1/\tau)\ldots\Gamma(1+m_s/\tau)}{\Gamma(1+n/\tau)}. \end{aligned}$$

(It may be helpful to consult [5, pp. 620–625] wherein multiple integrals of this and similar sorts are tabulated.) This proves the first expression for V.

The second expression actually represents the volume V of $f \leqq 1$ for *any* distance function f: $\mathbf{R}^n \to \mathbf{R}$, not just this particular one, as we can see in the following way.

$$\begin{aligned}
\int_{x \in \mathbf{R}^n} e^{-f(x)^\sigma}\, dx_1 \ldots dx_n &= \int_{x \in \mathbf{R}^n} \left(\int_{\substack{y \in \mathbf{R} \\ y \geqq f(x)^\sigma}} e^{-y}\, dy \right) dx_1 \ldots dx_n \\
&= \int_{y=0}^{\infty} e^{-y} \left(\int_{\substack{x \in \mathbf{R}^n \\ f(x)^\sigma \leqq y}} dx_1 \ldots dx_n \right) dy \\
&= \int_{y=0}^{\infty} e^{-y} y^{n/\sigma} \operatorname{Vol}(f \leqq 1)\, dy \\
&= V\Gamma(1 + n/\sigma).
\end{aligned}$$

Now for the third volume formula, which is really just the definition of the second volume formula as the limit of a Riemann sum. We prove it using some Poisson inversion machinery which will be crucial later.

The expression (9) is understood to mean that $t \to 0$ from above. Let $t > 0$ be fixed for the time being. We begin by writing the periodic function

$$\sum_{y \in \mathbf{Z}^k} e^{-t f(y+u)^\sigma},$$

for $u \in \mathbf{R}^k$, as a Fourier series

$$\sum_{y \in \mathbf{Z}^k} e^{2\pi i(u_1 y_1 + \ldots + u_k y_k)} \int_{x \in \mathbf{R}^k} e^{-t f(x)^\sigma} e^{-2\pi i(x_1 y_1 + \ldots + x_k y_k)}\, dx_1 \ldots dx_k.$$

This is justified since $\exp(-t f(y)^\sigma)$ and all its partial derivatives decay rapidly as y moves away from the origin. Setting $u = 0$ we get a version of the Poisson summation formula [23, I, p. 35],

$$\sum_{y \in \mathbf{Z}^k} e^{-t f(y)^\sigma} = \sum_{y \in \mathbf{Z}^k} \int_{x \in \mathbf{R}^k} e^{-t f(x)^\sigma - 2\pi i(x_1 y_1 + \ldots + x_k y_k)}\, dx_1 \ldots dx_k.$$

Replacing x with $t^{-1/\sigma} x$ we obtain

$$\sum_{y \in \mathbf{Z}^k} e^{-t f(y)^\sigma} = t^{-k/\sigma} \sum_{y \in \mathbf{Z}^k} \int_{x \in \mathbf{R}^k} e^{-f(x)^\sigma - 2\pi i t^{-1/\sigma}(x_1 y_1 + \ldots + x_k y_k)}\, dx_1 \ldots dx_k.$$

Multiplying both sides by $t^{k/\sigma}$, we get

$$\begin{aligned}
t^{k/\sigma} \sum_{y \in \mathbf{Z}^k} e^{-t f(y)^\sigma} &= \int_{x \in \mathbf{R}^k} e^{-f(x)^\sigma}\, dx_1 \ldots dx_k \\
&\quad + \sum_{\substack{y \in \mathbf{Z}^k \\ y \neq 0}} \int_{x \in \mathbf{R}^k} e^{-f(x)^\sigma - 2\pi i t^{-1/\sigma}(x_1 y_1 + \ldots + x_k y_k)}\, dx_1 \ldots dx_k.
\end{aligned}$$

Let $t \to 0$ along the positive real axis. The sum over nonzero lattice points vanishes by the Riemann-Lebesgue lemma, so

$$\lim_{t \to 0+} t^{k/\sigma} \sum_{y \in \mathbf{Z}^k} e^{-t f(y)^\sigma} = \int_{x \in \mathbf{R}^k} e^{-f(x)^\sigma} \, dx_1 \ldots dx_k .$$

By applying the second volume formula to the right-hand side, we complete the proof of the third expression, and so of Theorem 3.

5 The number of lattice points in a superball

The following theorem will enable us to get a useful upper bound on the denominators of the density bounds in Theorems 1 and 2. The simplicity of the proof belies the power of the result. Mazo and Odlyzko [12] discuss the method in the case of ordinary spheres.

Theorem 4 *Let $f: \mathbf{R}^k \to \mathbf{R}$ be a distance function, and s a positive number.*
(i) *The n-dimensional superball of positive radius r given by*

$$f(x_1, \ldots, x_k)^\sigma + f(x_{k+1}, \ldots, x_{2k})^\sigma + \ldots + f(x_{n-k+1}, \ldots, x_n)^\sigma \leqq r^\sigma, \qquad k \mid n,$$

contains no more than

$$e^{sr^\sigma} \Big(\sum_{x \in \mathbf{Z}^k} e^{-s f(x)^\sigma} \Big)^{n/k} \tag{10}$$

points of $\mathbf{Z}^n$.
(ii) *For some $r > 0$ the number of points of $\mathbf{Z}^n$ in this radius-r superball is at most*

$$\left(\Big(\sum_{x \in \mathbf{Z}^k} e^{-s f(x)^\sigma} \Big) \Big/ \Big(\int\int \ldots \int_{x \in \mathbf{R}^k} e^{-s f(x)^\sigma} \, dx_1 \ldots dx_k \Big) \right)^{n/k} \tag{11}$$

times its volume.

Proof. (i) Let N_y be the number of lattice points on the boundary of the superball of radius y, centered at the origin. Then

$$\sum_{\substack{y \geqq 0 \\ N_y \neq 0}} N_y \, e^{-s y^\sigma} = \left(\sum_{x \in \mathbf{Z}^k} e^{-s f(x)^\sigma} \right)^{n/k} . \tag{12}$$

But then

$$\begin{aligned} \sum_{y \leqq r} N_y &\leqq \sum_{\substack{y \geqq 0 \\ N_y \neq 0}} N_y \, e^{sr^\sigma - s y^\sigma} \\ &= e^{sr^\sigma} \left(\sum_{x \in \mathbf{Z}^k} e^{-s f(x)^\sigma} \right)^{n/k}, \end{aligned}$$

which yields the bound (10).
(ii) Let U_r be the number of lattice points in the interior of the superball of radius r, so

$$U_r = \sum_{\substack{y < r \\ N_r \neq 0}} N_y ,$$

and let V_r be its volume (of course $V_r = r^n V_1$ but we ignore this for the time being). Rewrite the first sum in (12) by partial summation, or equivalently by integration by parts of its representation as the Stieltjes integral $\int_{y=0}^{\infty} e^{-sy^\sigma}\, dU_y$, to obtain

$$\int_{y=0}^{\infty} U_r \cdot -d(e^{-sr^\sigma}) = \left(\sum_{x \in \mathbf{Z}^k} e^{-sf(x)^\sigma} \right)^{n/k};$$

likewise we have (as in Sect. 4)

$$\int_{y=0}^{\infty} V_r \cdot -d(e^{-sr^\sigma}) = \left(\iint \ldots \int_{x \in \mathbf{R}^k} e^{-sf(x)^\sigma}\, dx_1 \ldots dx_k \right)^{n/k}. \tag{13}$$

Thus we have

$$\int_{y=0}^{\infty} U_r \cdot -d(e^{-sr^\sigma}) = A \int_{y=0}^{\infty} V_r \cdot -d(e^{-sr^\sigma})$$

where A is the expression (11); since the measure $-d(e^{-sr^\sigma})$ is nonnegative there must thus be some r such that $U_r \leqq A\, V_r$, Q.E.D.

We can use part (ii) of this theorem directly to prove Theorem 6, and part (i) to give an explicit value of r that satisfies the estimate (11) up to an insignificant correction $\exp o(n)$ (Theorem 7). In practice we may have to apply an invertible linear transformation to $\mathbf{R}^k$ before invoking Theorem 4 so as to optimize the bounds (10) and (11). Equivalently we may replace $\mathbf{Z}^k$ by any lattice Λ in $\mathbf{R}^k$; the bound (11) for the ratio between the number of $\Lambda^{n/k}$-points in some radius-r superball and its volume then becomes

$$\left(\left(\sum_{x \in \Lambda} e^{-sf(x)^\sigma} \right) \Big/ \left(\iint \ldots \int_{x \in \mathbf{R}^k} e^{-sf(x)^\sigma}\, dx_1 \ldots dx_k \right) \right)^{n/k}. \tag{14}$$

In the remainder of this section we discuss the limitations of our basic method and some related problems. The critical values of r that lead to interesting results are asymptotically $r = c\, n^{1/\sigma}$ for a constant $c > 0$.

Mazo and Odlyzko [12] showed that for Euclidean spheres ($k = 1$ and $\sigma = 2$ in Theorem 4), the infimum of (10) over $s > 0$ gives the correct estimate (to within a factor of $\exp o(n)$) for the number of lattice points in the ball. Their proof can be easily extended to cover the cases of the other superballs. Thus Theorem 4 does not give away much, and for our method to give a result better than the Minkowski-Hlawka bound, we need to find an r for which the bound of part (i) of Theorem 4 is small. Now consider any superball as in Theorem 4, or more generally, any body. As we move the center of the superball around, the number of points of $\mathbf{Z}^n$ inside equals the volume of the superball, if we average over all possible positions of the center. Therefore a basic question is whether the number of points of $\mathbf{Z}^n$ inside a superball is minimized when the superball is centered at the origin or not. In the case of the Euclidean sphere ($k = 1$, $\sigma = 2$, and $r = c\, n^{1/\sigma}$ for a constant $c > 0$, as usual), Mazo and Odlyzko [12] showed that centering the ball at the origin maximizes the

number of lattice points inside (to within factors of $\exp o(n)$), no matter what c is. Therefore our method cannot improve on the Minkowski-Hlawka bound for Euclidean spheres. We now sketch a proof that this is also true for all l_σ-balls with $0<\sigma\leqq 2$. By the method of [12], the problem can be reduced to showing that if $s>0$ and

$$g(y)=\sum_{k=-\infty}^{\infty}\exp(-s|k-y|^\sigma),$$

then $g(y)$ is maximized for $y=0$. By the proof of the third volume formula of Theorem 3, this will follow if we show that the Fourier transform of $\exp(-s|u|^\sigma)$ is nonnegative for $0<\sigma\leqq 2$. There are proofs of this result in the literature, but they are quite cumbersome. We present here a very nice proof due to B.F. Logan, which comes from his unpublished work on completely monotonic functions. It follows easily from the following lemma (also due to Logan):

Lemma 5 *Let $s>0$, $0<\sigma\leqq 2$, and $h(u)=\exp(-s|u|^\sigma)$. Then*

$$h(u)=\int_{t=0}^{\infty}\exp(-tu^2)\,d\alpha(t),$$

where $\alpha(t)$ is bounded and non-decreasing and the integral converges for all u.

Proof. It suffices to prove this for $s=1$, since the general case will then follow by a change of variables. Let $\beta=\sigma/2$, so that $0<\beta\leqq 1$, and for $x\geqq 0$, let $H(x)=\exp(-x^\beta)$.

We first show that $H(x)$ is completely monotonic [24] on $(0,\infty)$; i.e., that $(-1)^k d^k H(x)/dx^k$ is nonnegative for all $k\geqq 0$ and all $x>0$. We use the induction hypothesis that

$$\frac{d^k H(x)}{dx^k}=(-1)^k\sum_{j=1}^{k}c(k,j)\,x^{j\beta-k}\,H(x),$$

where the $c(k,j)$ are $\geqq 0$. This is clearly true for $k=0$. If it's true for k, though, differentiating that expression term by term shows that it is also true for $k+1$, since all the exponents $j\beta-k$ are $\leqq 0$. This proves the complete monotonicity of $H(x)$, and so by a theorem of Bernstein [24, Theorem 12a], we find that

$$H(x)=\int_{t=0}^{\infty}e^{-xt}\,d\alpha(t),$$

where $\alpha(t)$ is bounded and non-decreasing and the integral converges for $0\leqq x<\infty$. We now make the substitution $x=u^2$, and obtain the claim of Lemma 5.

Since e^{-tx^2} has a nonnegative Fourier transform for all $t>0$, we see immediately that the Fourier transform of $h(u)$ is $\geqq 0$, which establishes our main claim. In fact, we can find out more from Lemma 5. Mazo and Odlyzko showed that the number of lattice points in a Euclidean sphere is minimized (within factors of $\exp o(n)$ again, and for r on the order of $n^{1/\sigma}$) when the center of

the ball is placed at the point $(1/2, \ldots, 1/2)$. Combining their method of proof with Lemma 5 we see that this also holds for l_σ-balls with $0<\sigma\leqq 2$.

When $\sigma>2$ (but $k=1$ still), the situation changes. As will be shown later, one can always choose a c so that for $r=cn^{1/\sigma}$ the number of lattice points inside the l_σ-ball is smaller than the volume by a factor exponential in n. However, this is not true for all c. It can be shown by the method of [12] that the positions for the center of the l_σ-ball where the number of lattice points is maximized or minimized (to within factors of $\exp o(n)$) are always of the form $(x, x, \ldots, x)$ for some fixed x, but the values of x vary depending on c. In many cases the extreme values of x are 0 and 1/2 (but in some cases 0 maximizes the number of lattice points, in other cases it minimizes this number). In other cases different behavior occurs. For example, for $\sigma=3$, there are at least 2 intervals of values of c for which the value of x that maximizes the number of lattice points inside the l_σ-ball is neither $x=0$ nor $x=1/2$.

6 The packing densities of superballs

Either of the next two theorems supplies a result promised in the abstract. They are related in the same way that Theorems 1 and 2 are related to each other.

Theorem 6 *Let $f\colon \mathbf{R}^k\to\mathbf{R}$ be the distance function of a body $\{x: f(x)\leqq 1\}$, and $\sigma\geqq 1$. The n-dimensional superball given by*

$$f(x_1, \ldots, x_k)^\sigma + \ldots + f(x_{n-k+1}, \ldots, x_n)^\sigma \leqq 1, \qquad k\,|\,n,$$

can be lattice packed with density at least

$$\left(\frac{1}{2}\sup_{\Lambda}\left(\frac{\int_{x\in\mathbf{R}^k} e^{-f(x)^\sigma}\,dx_1\ldots dx_k}{\det(\Lambda)\sum_{x\in\Lambda} e^{-f(x)^\sigma}}\right)^{1/k}\right)^{n+o(1)}$$

as $n\to\infty$. Here the supremum is taken over all lattices Λ in $\mathbf{R}^k$, and the $o(1)$ correction in the exponent can be removed by allowing nonlattice packings.

Proof. We apply the version (14) of the second part of Theorem 4 to an arbitrary lattice Λ in $\mathbf{R}^k$, setting $s=1$. (We could use any other s, but that would give the same result as taking $s=1$ and scaling Λ by a factor $s^{1/\sigma}$.) We obtain some r such that the number $M_n(r)$ of points of $\Lambda^{n/k}$ contained in the n-dimensional superball of radius r is bounded by $M_n(r)\leqq A\,V_r$, where again A is given by (14) and $V_r=r^n V_1$ is the volume of that superball. By Theorem 2 it follows that disjoint translates of that superball (and thus also of the radius-1 superball) pack $\mathbf{R}^n$ with density at least $(2\sqrt[k]{A\det\Lambda})^{-n}$. Now if $M_k(r)=1$, so Λ is admissible for the k-dimensional superball of radius r, clearly also $M_n(r)=1$ and our packing is already the lattice packing with lattice $2\Lambda^{n/k}$. Otherwise we have $M_n(r)\geqq 1+(n/k)(M_k(r)-1)\to\infty$ as $n\to\infty$, so there exists a prime $p\geqq M_n(r)-1$ with $p=(1+o(1))M_n(r)$; thus from Theorem 2 we obtain a lattice packing of the dimension-n superball with density at least $(1-o(1))(2\sqrt[k]{A\det\Lambda})^{-n}$. Q.E.D.

We next apply the methods of [17] and [18] to obtain a nearly identical bound (weaker only by a factor of $\exp o(n)$) while also obtaining a specific choice for r as a function of n by using the first part of Theorem 4. By a linear transformation of $\mathbf{R}^k$ we can make the lattice Λ of Theorem 6 homothetic to $\mathbf{Z}^k$, and will do this to conform with the practice of [17] and [18].

Theorem 7 *Let* $f\colon \mathbf{R}^k \to \mathbf{R}$ *be the distance function of a* $\mathbf{Z}^k$*-admissible body* $\{x : f(x) \leqq 1\}$, *and* $\sigma \geqq 1$. *The n-dimensional superball G given by*

$$f(x_1, \ldots, x_k)^\sigma + \ldots + f(x_{n-k+1}, \ldots, x_n)^\sigma \leqq 1, \qquad k \mid n,$$

can be lattice packed with density at least

$$\left(\frac{1}{2} \sup_{s>0} \left(\frac{\int_{x \in \mathbf{R}^k} e^{-s f(x)^\sigma} \, dx_1 \ldots dx_k}{\sum_{x \in \mathbf{Z}^k} e^{-s f(x)^\sigma}} \right)^{1/k} \right)^{n+o(n)}$$

as $n \to \infty$.

Proof. Our starting point is Theorem 1.

If there is any point of $\mathbf{Z}^n$ at which two different translates of rG by vectors of $p\mathbf{Z}^n$ overlap, then

$$\operatorname{card}(\mathbf{Z}^n \cap pQ \cap (p\mathbf{Z}^n + rG)) < \operatorname{card}(\mathbf{Z}^n \cap rG),$$

and if there is no such lattice point, then we instead have equality. In any event, the left-hand side never exceeds the right. Consequently

$$\delta_L(G) \geqq \frac{2^{1-n} r^n \operatorname{Vol}(G)}{(p-1) \operatorname{card}(\mathbf{Z}^n \cap rG)}.$$

Let p be the smallest prime greater than r, where

$$r = \left(\frac{n}{\sigma s} \right)^{1/\sigma},$$

and s is a positive constant at our disposal. Then

$$r < p \leqq 2r + 3, \qquad 2/(p-1) = e^{o(n)},$$

and it follows that

$$\frac{\ln \delta_L(G)}{n} \geqq \frac{1}{n} \ln \left(\frac{2^{-n} r^n \operatorname{Vol}(G)}{\operatorname{card}(\mathbf{Z}^n \cap rG)} \right) + o(1) \tag{15}$$

for large n.

Using (7) of Theorem 3, we get

$$\operatorname{Vol}(G) = \Gamma(1 + k/\sigma)^{n/k} \, \Gamma(1 + n/\sigma)^{-1} \operatorname{Vol}(f \leqq 1)^{n/k},$$

in which

$$\Gamma(1 + n/\sigma) = n^{n/\sigma} (\sigma e)^{-n/\sigma} e^{o(n)} \quad \text{as} \quad n \to \infty$$

by Stirling's formula. Hence

$$\ln(r^n \operatorname{Vol}(G)) = n \ln\left(\left(\frac{e}{s}\right)^{1/\sigma} \sqrt[k]{\Gamma(1+k/\sigma)\operatorname{Vol}(f\leqq 1)}\right) + o(n). \tag{16}$$

The first part of Theorem 4 gives

$$\operatorname{card}(\mathbf{Z}^n \cap rG) \leqq e^{n/\sigma}\left(\sum_{x\in\mathbf{Z}^k} e^{-sf(x)^\sigma}\right)^{n/k}. \tag{17}$$

Substituting the estimates (16) and (17) into the density bound (15), we find that

$$\frac{\ln \delta_L(G)}{n} \geqq \ln\left(\frac{1}{2}\, s^{-1/\sigma}\left(\frac{\Gamma(1+k/\sigma)\operatorname{Vol}(f\leqq 1)}{\sum_{x\in\mathbf{Z}^k} e^{-sf(x)^\sigma}}\right)^{1/k}\right) + o(1).$$

But by (8) from Theorem 3,

$$\begin{aligned} s^{-k/\sigma}\Gamma(1+k/\sigma)\operatorname{Vol}(f\leqq 1) &= s^{-k/\sigma}\int_{x\in\mathbf{R}^k} e^{-f(x)^\sigma}\, dx_1 \ldots dx_k \\ &= \int_{x\in\mathbf{R}^k} e^{-sf(x)^\sigma}\, dx_1 \ldots dx_k, \end{aligned}$$

and the proof of Theorem 7 is complete.

Let us now investigate how Theorem 7 behaves in the three limiting cases $s\to\infty$, $s\to 0+$, and $\sigma\to\infty$.

The expression

$$\left(\frac{\int_{x\in\mathbf{R}^k} e^{-sf(x)^\sigma}\, dx_1 \ldots dx_k}{\sum_{x\in\mathbf{Z}^k} e^{-sf(x)^\sigma}}\right)^{1/k} = h(s,\sigma),$$

say, whose supremum is taken, approaches 0 as $s\to\infty$, making the bound weak. But h approaches 1 as $s\to 0+$, so that the bound is of the same asymptotic strength as the Minkowski-Hlawka bound in that case. If there is some positive s which makes h greater than one, then we obtain an essential improvement on the Minkowski-Hlawka bound.

Next we ascertain the limiting form of Theorem 7 as $\sigma\to\infty$. The superball (6) becomes

$$G = \{x\in\mathbf{R}^n : \max(f(x_1, \ldots, x_k), \ldots, f(x_{n-k+1}, \ldots, x_n)) \leqq 1\},$$

a Cartesian product of k-dimensional bodies $f\leqq 1$.

Let $z = e^{-s}$,

$$\xi(z) = \sum_{x\in\mathbf{Z}^k} z^{f(x)^\sigma},$$

and $u(z)=h(s,\sigma)$. Instead of $s>0$, we seek a positive $z<1$ to optimize the density bound. Setting the logarithmic derivative of $u(z)$ equal to zero, we get

$$-\sigma=\frac{k\xi(z)}{z\xi'(z)\ln z}. \tag{18}$$

Let c be the number of points of $\mathbf{Z}^k$ on the surface $f=1$. Since

$$\xi(z)=1+cz+O(z^2)\to 1 \tag{19}$$

and

$$\xi'(z)=c+O(z)\to c,$$

Eq. (18) behaves like

$$\frac{c\sigma}{k}=\frac{1}{z\ln 1/z} \tag{20}$$

for small positive z. Note that

$$\lim_{z\to 0+}\frac{1}{z\ln 1/z}=+\infty.$$

So if σ is large, there is some small $\rho>0$ such that (18) has $z=\rho$ as its smallest positive root. Let us determine the asymptotic behavior of this ρ as $\sigma\to\infty$. Thus we seek the asymptotic behavior of the smaller solution ρ (there are two solutions as $\sigma\to\infty$ but the larger one is an artifact) of (20) or equivalently the larger solution x of

$$xe^{-x}=t \tag{21}$$

as $t\to 0$. Here we have set $x=\ln 1/\rho$ and $t=k/(c\sigma)$. Following the example of de Bruijn [1, p. 25] we write (21) as

$$x=\ln\frac{1}{t}+\ln x. \tag{22}$$

If $t>0$ is sufficiently small, then

$$\ln\frac{1}{t}<x<2\ln\frac{1}{t}$$

and so $\ln x=O(\ln\ln 1/t)$ as $t\to 0$. By (22),

$$x=\ln\frac{1}{t}+O\left(\ln\ln\frac{1}{t}\right)=\left(1+\frac{O(\ln\ln 1/t)}{\ln 1/t}\right)\ln\frac{1}{t}=\left(1+O\left(\frac{\ln\ln 1/t}{\ln 1/t}\right)\right)\ln\frac{1}{t}.$$

So

$$\ln x=\ln\ln\frac{1}{t}+\ln\left(1+O\left(\frac{\ln\ln 1/t}{\ln 1/t}\right)\right)=\ln\ln\frac{1}{t}+O\left(\frac{\ln\ln 1/t}{\ln 1/t}\right).$$

Using this in (22) we get

$$x = \ln \frac{1}{t} + \ln \ln \frac{1}{t} + O\left(\frac{\ln \ln 1/t}{\ln 1/t}\right),$$

or, going back to the original variables, we find that the optimal s is

$$s = \ln \frac{1}{\rho} = \ln(c\,\sigma/k) + \ln \ln(c\,\sigma/k) + O\left(\frac{\ln \ln(c\,\sigma/k)}{\ln(c\,\sigma/k)}\right) \tag{23}$$

as $\sigma \to \infty$. We could refine these estimates by successive substitution into (22), but (23) is already more accurate than we need.

Now we can obtain the limiting form of Theorem 7 as $\sigma \to \infty$. From (23) we see that $\rho \to 0$ and so from (19), $\xi(\rho) \to 1$. Clearly $\Gamma(1+k/\sigma) \to 1$. And again using (23) we find that $(\ln 1/\rho)^{1/\sigma} \to 1$. So $u(z) = h(s, \sigma)$ approaches $\mathrm{Vol}(f \leqq 1)^{1/k}$.

The limiting form of Theorem 7 is therefore as follows:

Theorem 8 *Let f: $\mathbf{R}^k \to \mathbf{R}$ be the distance function of a $\mathbf{Z}^k$-admissible body J. Suppose that k divides n, and that G is the Cartesian product of n/k copies of J, that is*

$$G = \{x \in \mathbf{R}^n : \max(f(x_1, \ldots, x_k), \ldots, f(x_{n-k+1}, \ldots, x_n)) \leqq 1\}.$$

Then the maximum lattice-packing density δ_L of G satisfies

$$\delta_L(G) \geqq \gamma^{n+o(n)}$$

as $n \to \infty$ where

$$\gamma = \tfrac{1}{2} \max(1, \sqrt[k]{\mathrm{Vol}(J)}).$$

If $\mathrm{Vol}(J) > 1$ then this is stronger than the Minkowski-Hlawka bound.

Suppose that $\mathbf{Z}^k$ is a closest-packing lattice for $J/2$. Then the bound $\delta_L(G) \geqq \delta_L(J)^{n/k+o(n)}$ given by Theorem 8 is essentially the same as one gets by using an orthogonal sum of closest-packing lattices for J. This justifies the example given in the introduction, concerning bodies defined by Cartesian products.

A careful justification of the passage to the limit which gave rise to Theorem 8 can be given as follows: Inside the body

$$G_\sigma = \{x \in \mathbf{R}^n : (f(x_1, \ldots, x_k)^\sigma + \ldots + f(x_{n-k+1}, \ldots, x_n)^\sigma)^{1/\sigma} \leqq 1\}$$

inscribe a ball μG,

$$G = G_\infty = \{x \in \mathbf{R}^n : \max(f(x_1, \ldots, x_k), \ldots, f(x_{n-k+1}, \ldots, x_n)) \leqq 1\},$$

with $\mu < 1$ as large as possible. Then apply Construction A to $[n, k, r, p, G_\sigma]$ codes. This gives packings of μG whose density is $\delta\,\mathrm{Vol}(\mu G)/\mathrm{Vol}(G_\sigma)$ where δ is the density of the associated packing of G_σ. It is obvious that if σ becomes

infinite quickly enough, as a function of n, then $\mathrm{Vol}(\mu G)/\mathrm{Vol}(G_\sigma) \to 1$ as $n \to \infty$. Thus we obtain packings of μG (and so of G, since the density is independent of scale) whose asymptotic densities are those of Theorem 8.

7 The case of the classical l_σ-ball

We assume, since we need convexity, that $\sigma \geqq 1$.

Theorem 9 *The l_σ-ball*

$$|x_1|^\sigma + |x_2|^\sigma + \ldots + |x_n|^\sigma \leqq 1$$

can be lattice packed with density at least $(b/2)^{n+o(1)}$ as $n \to \infty$, where

$$b = \sup_{t>0} \frac{\int_{x=-\infty}^{\infty} e^{-|tx|^\sigma}\, dx}{\sum_{k=-\infty}^{\infty} e^{-|tk|^\sigma}}. \tag{24}$$

When $1 \leqq \sigma \leqq 2$, this gives $b=1$, as in the Minkowski-Hlawka bound. But when $\sigma > 2$, we obtain $b > 1$, an essential improvement. As ε tends to $0+$, b is asymptotically

$$1 + \frac{\sqrt{\pi}}{2}\, \zeta(3)\, \varepsilon (\ln 1/\varepsilon)^{-3/2} (1 + o(1)) \tag{25}$$

for $\sigma = 2 + \varepsilon$. As $\sigma \to \infty$, $\ln(2/b) \sim (\ln \ln \sigma)/\sigma$.

Proof. The first statement is proved by setting $k=1$ in Theorem 6, and letting f be the one-variable distance function $f(x)=|x|$. For the second, rewrite the ratio in (24) as

$$\left(\int_{x=-\infty}^{\infty} e^{-|x|^\sigma}\, dx \right) \Big/ \left(t \sum_{k=-\infty}^{\infty} e^{-|tk|^\sigma} \right),$$

that is, as the ratio between the integral $\int_{-\infty}^{\infty} \exp(-|x|^\sigma)\, dx$ and a Riemann sum for the integral; in particular as the mesh size t approaches zero the Riemann sum tends to the integral, showing that $b \geqq 1$. By the Poisson inversion theorem the difference between the sum and the integral is $2 \sum_{m=1}^{\infty} g_\sigma(m/t)$, where $g_\sigma(w)$ is the Fourier transform

$$g_\sigma(w) = \int_{v=-\infty}^{\infty} \exp(-|v|^\sigma) e^{2\pi i w v}\, dv;$$

in particular, if $\sigma \leqq 2$ we have already seen that $g_\sigma(w) > 0$ for all w (Lemma 5), so the sum always exceeds the integral and $b=1$. For $\sigma > 2$ it is easy to see

that g_σ must take negative values for some w by borrowing a trick from analytic number theory: if $g_\sigma(w)$ were nonnegative we would have

$$3-4\exp(-|v|^\sigma)+\exp(-|2v|^\sigma)=8\int_{w=-\infty}^{\infty} g_\sigma(w)\sin^4(\pi w v)\,dw>0$$

for all v, but this inequality fails for sufficiently small positive v once $\sigma>2$. While this does not in itself show that $b>1$, it would follow from Poisson that $b>1$ if $g_\sigma(w)$ were known to be negative for all $|w|$ sufficiently large, for then the integral could be made to exceed its Riemann sum by taking t sufficiently small. We will obtain below an asymptotic formula for $g_\sigma(w)$ as $|w|\to\infty$ (first part of Lemma 10) that will show that $g_\sigma(w)<0$ for large $|w|$ provided σ lies in an open interval $(4m+2, 4m+4)$ for some $m=0, 1, 2, \ldots$; in particular this will give $b>1$ for $2<\sigma<4$. But for σ sufficiently large we can simply take $t=1$, when as a function of σ the Riemann sum $\sum_{k=-\infty}^{\infty}\exp(-|k|^\sigma)$ decreases to $1+2/e$ and $\int_{x=-\infty}^{\infty}\exp(-|x|^\sigma)=2\Gamma(1+1/\sigma)$ increases to $2>1+2/e$, and in fact numerical computation shows that (say) $\sigma>3$ is large enough to obtain $\sum_{k=-\infty}^{\infty}\exp(-|k|^\sigma)<\Gamma(1+1/\sigma)$. Combining these two estimates we conclude that indeed $b>1$ for all $\sigma>2$. It remains to estimate b for $\sigma=2+\varepsilon$ and $\sigma\to\infty$. For the former we use the second estimate in Lemma 10 below and apply the Poisson inversion theorem to obtain

$$t\left(\sum_{k=-\infty}^{\infty}\exp(-|tk|^{2+\varepsilon})\right)-2\Gamma\left(1+\frac{1}{2+\varepsilon}\right)$$
$$=2\sqrt{\pi}\left(\sum_{m=1}^{\infty}e^{-(\pi m/t)^2}\right)-\frac{\zeta(3)\,t^3}{2\pi^2}\varepsilon+O(\varepsilon^2t^2+\varepsilon t^4); \tag{26}$$

here the constant implied in $O(\varepsilon^2t^2+\varepsilon t^4)$ is absolute provided ε is small enough and t is bounded (and t clearly must be bounded, say by $2\Gamma(1+1/(2+\varepsilon))$, to be of any use to us). Next note that for fixed t the right-hand side of (26) is positive for ε small enough, so as $\varepsilon\to0$, the optimal t must tend to zero as well. Thus the error $O(\varepsilon t^4)$ is negligible and the sum $2\sqrt{\pi}\sum_{m=1}^{\infty}e^{-(\pi m/t)^2}$ is dominated by its first term $2e^{-(\pi/t)^2}$. It is then clear that the optimal t is $(1-o(1))\,\pi(\ln 1/\varepsilon)^{-1/2}$ (note that this decreases very slowly with ε – indeed computations show that the optimal t exceeds 1 for all $\sigma>2.011!$); thus (26) is asymptotic to

$$\frac{\pi}{2}\zeta(3)\,\varepsilon(\ln 1/\varepsilon)^{-3/2},$$

and dividing by $2\Gamma(1+1/(2+\varepsilon))=\sqrt{\pi}+O(\varepsilon)$ we obtain the estimate (25). Finally, for the behavior of b as $\sigma\to\infty$ (which was established in [17, Eq. (32)]), use Eq. (23) with $k=1$ and $c=2$.

Thus Theorem 9 is proved modulo the following lemma on the behavior of $g_\sigma(w)$:

Lemma 10 (i) *For fixed $\sigma > 0$, the Fourier transform*

$$g_\sigma(w) = \int_{v=-\infty}^{\infty} \exp(-|v|^\sigma) e^{2\pi i w v}\, dv$$

of $\exp(-|v|^\sigma)$ *has the asymptotic expansion*

$$g_\sigma(w) \sim 2 \sum_{m=0}^{\infty} \frac{(-1)^{m+1}}{m!} \sin\left(\frac{m\pi\sigma}{2}\right) \Gamma(m\sigma+1)(2\pi|w|)^{-m\sigma-1} \tag{27}$$

as $|w| \to \infty$.

(ii) *For* $\sigma = 2+\varepsilon$, *with* $0 < \varepsilon < 1$ *and* $|w| > 1$, *we have*

$$g_\sigma(w) = \sqrt{\pi}\, e^{-(\pi w)^2} - (1 + O(|w|^{-1})) \frac{\varepsilon}{4\pi^2 |w|^3} + O((\varepsilon/w)^2), \tag{28}$$

where the constants implied by the O-notation are absolute.

It is not crucial that $\varepsilon < 1$ and $|w| > 1$, only that ε be positive and bounded, and w bounded away from 0; changing the bounds on ε and w we only change the implicit O-constants in (28).

Proof of Lemma 10. (i) The basic idea is that $\exp(-|v|^\sigma)$ is smooth for all $v \neq 0$ so only the singularity at the origin can contribute polynomial terms to the asymptotic behavior of $g_\sigma(w)$ at infinity. We prove by induction that for each $M = 0, 1, 2, \ldots$

$$g_\sigma(w) = 2 \sum_{\substack{m \\ m\sigma < M}} \frac{(-1)^{m+1}}{m!} \sin\left(\frac{m\pi\sigma}{2}\right) \Gamma(m\sigma+1)(2\pi|w|)^{-m\sigma-1} + (-2\pi i w)^{-M} \int_{v=-\infty}^{\infty} E_M(v) e^{2\pi i w v}\, dv \tag{29}$$

where $E_M(v)$ is the M-th derivative of

$$e^{-|v|^\sigma} - \sum_{\substack{m \\ m\sigma < M}} \frac{(-1)^m}{m!} |v|^{-m\sigma}.$$

Since $E_M(v)$ is of bounded variation on $\mathbf{R}$ and tends to 0 for large $|v|$, its Fourier transform is $O(1/|w|)$ for large $|w|$, so (29) for all M will yield (27). Now for $M = 0$ the formula (29) is simply the definition of $g_\sigma(w)$. Having proved (29) for some M, we integrate $\int_{v=-\infty}^{\infty} E_M(v) e^{2\pi i w v}\, dv$ by parts, obtaining

$$\frac{1}{-2\pi i w} \int_{v=-\infty}^{\infty} \left(E_{M+1}(v) + \sum_{\substack{m \\ M \leq m\sigma < M+1}} \frac{(-1)^m}{m!} \frac{\Gamma(m\sigma+1)}{\Gamma(m\sigma-M)} \operatorname{sgn}(v)^M |v|^{m-M-1} \right) e^{2\pi i w v}\, dv. \tag{30}$$

(Actually if some $m\sigma$ exactly equals M the corresponding term must be treated specially. The coefficient $\Gamma(m\sigma+1)/\Gamma(m\sigma-M)$ vanishes due to the pole of the gamma function at the origin, reflecting the vanishing of the $(M+1)$-st derivative of $|v|^M$, but if M is odd the discontinuity of the M-th derivative at the origin forces us to split the integral $\int_{-\infty}^{\infty}$ into $\int_{-\infty}^{0}+\int_{0}^{\infty}$ before integrating by parts, giving rise to an extra term $(i/\pi w)((-1)^m/m!)\,M!$ in (30); however for both even and odd M this still produces the correct term

$$2\frac{(-1)^{m+1}}{m!}\sin\left(\frac{m\pi\sigma}{2}\right)\Gamma(m\sigma+1)(2\pi|w|)^{-m\sigma-1}$$

in the asymptotic expansion of $g_\sigma(w)$). But the sum in (30) can be integrated termwise via the well-known definite integral

$$\int_{v=0}^{\infty} x^{\mu-1}e^{iax}\,dx=\Gamma(\mu)\,e^{i\pi\mu/2}\,a^{-\mu}$$

valid for all $\mu\in(0,1)$ and $a>0$ (use contour integration to change the path of integration to the positive imaginary axis, or see the definite integrals #3.761 in [5]); it is easy to check that this completes the induction step in (29) and the proof of the asymptotic formula (27).

(ii) Expand $\exp(-|v|^\sigma)$ in powers of $\varepsilon=\sigma-2$, obtaining

$$e^{-|v|^\sigma}=e^{-v^2}-\varepsilon\,e^{-v^2}v^2\ln|v|+\varepsilon^2E(v,\varepsilon)$$

where $E(v,\varepsilon)=E_\varepsilon(v,\varepsilon)$ and its first derivative are both continuous and bounded by (say) $c\exp(-v^2/2)$ for some absolute constant c independent of v and ε, whence the Fourier transform of $E(v,\varepsilon)$ is $O(w^{-2})$ for $|w|>1$. (The exponent -2 can be improve but it suffices for our purposes.) Now the Fourier transform of e^{-v^2} is well known to be $\sqrt{\pi}\,e^{-(\pi w)^2}$; and the Fourier transform of $e^{-v^2}v^2\ln|v|$ can be estimated as before: replace $e^{2\pi iwv}$ by $\cos(2\pi wv)$ and integrate by parts three times to obtain

$$(1+O(|w|^{-1}))\frac{2}{(2\pi w)^3}\int_{v=0}^{\infty}\frac{2}{v}\sin(2\pi wv)\,dv=\frac{1+O(|w|^{-1})}{4\pi^2|w|^3}$$

$$\left(\text{recall that }\int_0^\infty \sin(wv)\,dv/v=(\pi/2)\operatorname{sgn}w\right),\quad \text{Q.E.D.}$$

Note. When σ is an even integer $\geqq 4$, $\sin(m\pi\sigma/2)=0$ for all integers m, so part (i) of the above result tells us only that the Fourier transform $g_\sigma(w)$ of $\exp(-|v|^\sigma)$ decreases faster than any negative power of $|w|$, which is clear anyway because for such σ the function $\exp(-|v|^\sigma)$ is an entire function of v. In this case it can be shown by using contour integration and the saddle-point method that $g_\sigma(w)$ has infinitely many real zeros and that its absolute magnitude decreases approximately as $\exp(-C|w|^{\sigma/(\sigma-1)})$ for a constant $C=C(\sigma)$ as $|w|\to\infty$.

8 Generalization to mixed exponents

Our definition of a superball, (6), did not include mixed-exponent bodies such as the one given by the inequality

$$|x_1|^{\sigma_1}+\ldots+|x_{\rho_1}|^{\sigma_1}+|x_{\rho_1+1}|^{\sigma_2}+\ldots+|x_{\rho_1+\rho_2}|^{\sigma_2}\leqq 1 \tag{31}$$

where $\rho_1+\rho_2=n$. But our methods can be easily adapted to give lower bounds on the packing densities of bodies defined by inequalities on sums of homogeneous functions with different exponents. For example, the body (31) can be packed with density at least $2^{-n}b_1^{\rho_1}b_2^{\rho_2}$ where $b_i/2=2^{c_{\sigma_i}}$ (c_σ as defined in the introduction); this improves on Minkowski-Hlawka if either σ_1 or σ_2 exceeds 2. In this section we give the most general such bound readily obtainable by these methods:

Let $\{k_i\}_{i\in I}$ be a finite collection of positive integers, and $n=\sum_{i\in I}k_i$ their sum. For each $i\in I$ let $\sigma_i\geqq 1$ be some real number and let $F_i=f_i^{\sigma_i}$ be the σ_i-th power of some distance function f_i on $\mathbf{R}^{k_i}$. For any $r=\{r_i\}$ with each $r_i>0$, define the *generalized superball of radius* r, $B(r)\subset\mathbf{R}^n=\bigoplus_i\mathbf{R}^{k_i}$, to be the body

$$B(r)=\{(\mathbf{x}_i)\in\mathbf{R}^n\colon \sum_{i\in I}F_i(\mathbf{x}_i/r_i)\leqq 1\}.$$

Note that for any r the generalized superball $B(r)$ may be obtained from the one of unit radius $B(1, 1, \ldots, 1)$ by a linear transformation, specifically scaling each coordinate $\mathbf{x}_i$ by the factor r_i. As in Sect. 4 we have

$$\prod_i \iint_{\mathbf{R}^{k_i}}\ldots\int e^{-F_i(\mathbf{x}_i)}\,d\mathbf{x}_i=\int_{y=0}^{\infty}e^{-y}V_y\,dy, \tag{32}$$

where V_y is the volume of $B(\{y^{1/\sigma_i}\})$. Note that $V_y=y^sV_1=y^s\,\mathrm{Vol}(B(1, 1, \ldots, 1))$ where $s=\sum_i k_i/\sigma_i$; this yields a formula for $\mathrm{Vol}(B(1, 1, \ldots, 1))$ generalizing the second formula of Theorem 3:

$$\mathrm{Vol}(B(1, 1, \ldots, 1))=\frac{\prod_i \iint_{\mathbf{R}^{k_i}}\ldots\int e^{-F_i(\mathbf{x}_i)}\,d\mathbf{x}_i}{\Gamma(1+\sum_i k_i/\sigma_i)}=\frac{\prod_i\left(\Gamma(1+k_i/\sigma_i)\,\mathrm{Vol}(B_i(1))\right)}{\Gamma(1+\sum_i k_i/\sigma_i)} \tag{33}$$

where $B_i(1)=\{\mathbf{x}_i\in\mathbf{R}^{k_i}\colon F_i(\mathbf{x}_i)\leqq 1\}$. Also, if for each i we take a lattice Λ_i in $\mathbf{R}^{k_i}$ and let U_y be the number of points of the lattice $\Lambda=\bigoplus_i\Lambda_i\subset\mathbf{R}^n$ in $B(\{y^{1/\sigma_i}\})$, we have as in Sect. 5

$$\prod_i\left(\sum_{\mathbf{x}_i\in\Lambda_i}e^{-F_i(\mathbf{x}_i)}\right)=\int_{y=0}^{\infty}e^{-y}U_y\,dy. \tag{34}$$

We can thus generalize Theorem 4 as follows:

Theorem 11 (i) *For all positive real numbers* y,

$$U_y < e^y \prod_i \Big(\sum_{\mathbf{x}_i \in \Lambda_i} e^{-F_i(\mathbf{x}_i)} \Big). \tag{35}$$

(ii) *There exists* $y > 0$ *such that*

$$U_y \leqq V_y \prod_i \frac{\sum_{\mathbf{x}_i \in \Lambda_i} e^{-F_i(\mathbf{x}_i)}}{\iint_{\mathbf{R}^{k_i}} \dots \int e^{-F_i(\mathbf{x}_i)}\, d\mathbf{x}_i}. \tag{36}$$

Proof. (i) Since U_y is positive and increasing, the right-hand side of (34) is greater than

$$\int_{t=y}^{\infty} e^{-t} U_t \, dt > U_y \int_{t=y}^{\infty} e^{-t}\, dt = e^{-y} U_y$$

whence (35) follows.
(ii) As in Theorem 3 this is evident upon comparison of the integrals (32) and (34).

We can now mimic the proof of Theorem 6. (We could also use (33) and (35) to generalize Theorem 7, but we would have to work harder only to obtain the same or slightly worse estimates.) Let b_i be the supremum over lattices $\Lambda_i \in \mathbf{R}^{k_i}$ of

$$\Big(\iint_{\mathbf{x}_i \in \mathbf{R}^{k_i}} \dots \int e^{-F_i(\mathbf{x}_i)}\, d\mathbf{x}_i \Big) \Big/ \Big(\det(\Lambda_i) \sum_{\mathbf{x}_i \in \Lambda_i} e^{-F_i(\mathbf{x}_i)} \Big). \tag{37}$$

Note that $b_i \geqq 1$. Then we have:

Theorem 12 *There exist enough translates of* $B(1, 1, \dots, 1)$ *by vectors of* $\mathbf{R}^n$ *to pack* $\mathbf{R}^n$ *with density at least* $2^{-n} \prod_i b_i$.

Proof. Let Λ_i be arbitrary lattices in $\mathbf{R}^{k_i}$, and Λ their direct sum in $\mathbf{R}^n$. Apply Theorem 2 to the lattice Λ and the body $B(\{y^{1/\sigma_i}\})$, and apply a linear transformation to obtain a packing of $B(1, 1, \dots, 1)$ of density at least $2^{-n} \prod_i b_i - \varepsilon$ for ε arbitrarily small. It is easy to see that this ε can be removed by partitioning $\mathbf{R}^n$ into ever thicker spherical shells and putting in the m-th shell a chunk of the density-$(2^{-n} \prod_i b_i - \varepsilon_m)$ packing of $B(1, 1, \dots, 1)$ for some sequence $\varepsilon_m \to 0$, giving the desired density $2^{-n} \prod_i b_i$.

Note. As usual with Theorem 2, this packing will in general not be a lattice packing, but by the second part of Theorem 2 we can also obtain a lattice packing whose density is at most smaller by a bounded factor, and even a factor $(1 - o(1))$ can be obtained if it can be established as in the proof of Theorem 9 that $\operatorname{card}(\Lambda \cap \operatorname{int} B(\{y^{1/\sigma_i}\})) \to \infty$.

9 Open problems

We conclude with three questions suggested by the above work:
(i) For any distance function f we can only improve on Minkowski-Hlawka once σ exceeds some critical value σ_0. In all cases we have computed, $\sigma_0 \geqq 2$, with equality only when f is equivalent to the Euclidean distance function. Is there any distance function for which $b > 1$, giving by our methods essential improvements on Minkowski-Hlawka, for some $\sigma < 2$?
(ii) If two or more of the σ_i of Sect. 8 are equal (as happens for the superballs of Sects. 2–7), say $\sigma_1 = \sigma_2 = \ldots = \sigma_j$, we can replace $F_1, \ldots, F_j$ by a single homogeneous function $F = \sum_{i=1}^{j} F_i$ on $\mathbf{R}^k$ with $k = \sum_{i=1}^{j} k_i$. The ratio (37) obtained from this F is certainly no worse than the product of the b_i of the component F_i's, as may be seen by using lattices $\bigoplus_{i=1}^{j} \Lambda_i$; but conceivably the ratio might be increased by using lattices that are not direct sums of this form. We have not found an example where this occurs, but neither can we prove that the optimal lattice must always be a direct sum.
(iii) Let g: $\mathbf{R}^k \to \mathbf{R}$ and h: $\mathbf{R}^m \to \mathbf{R}$ be the distance functions of bodies G and H respectively. Consider the Cartesian product body $G \times H \subseteq \mathbf{R}^n$, where $n = k + m$, with distance function $\max(g, h)$. It can be lattice packed with density $\delta_L(G)\,\delta_L(H)$ by using the orthogonal sum of closest-packing lattices for G and H. Are there bodies G, H for which a denser lattice packing of $G \times H$ is possible?

Similarly, it is obvious that $\delta(G \times H) \geqq \delta(G)\,\delta(H)$, where δ denotes the maximum density by a packing of translates. Does the strict inequality ever hold?

Acknowledgement. The authors thank B.F. Logan for permission to include his proof of Lemma 5, and P.M. Gruber for comments concerning the third problem in Sect. 9.

References

1. de Bruijn, N.G.: Asymptotic Methods in Analysis. New York: Dover 1981 (Reprint)
2. Cassels, J.W.S.: An Introduction to the Geometry of Numbers. Berlin Heidelberg New York: Springer 1959
3. Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices and Groups. Berlin Heidelberg New York: Springer 1987
4. Erdös, P., Gruber, P.M., Hammer, J.: Lattice Points. Pitman Monograph 39, Longman Scientific. New York: John Wiley & Sons, 1989
5. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series and Products. New York: Academic Press 1980 (Translated from Russian)
6. Gruber, P.M., Lekkerkerker, C.G.: Geometry of Numbers. Amsterdam: North-Holland Elsevier 1987 (This is an updated version of [10])
7. Hlawka, E.: Zur Geometrie der Zahlen. Math. Z. **49**, 285–312 (1943)
8. Kabatiansky, G.A., Levenshtein, V.I.: Bounds for packings on a sphere and in space (in Russian). Probl. Peredachy Inf. **14**, 3–25 (1978); (English translation in) Probl. Inf. Transm. **14**, 1–17 (1978)
9. Leech, J., Sloane, N.J.A.: Sphere packing and error-correcting codes. Can. J. Math. **23**, 718–745 (1971)
10. Lekkerkerker, C.G.: Geometry of Numbers. Groningen: Wolters-Noordhoff 1969

11. Litsin, S.N., Tsfasman, M.A.: Algebraic-geometric and number-theoretic packings of spheres (in Russian). Usp. Mat. Nauk **40**, 185–186 (1985)
12. Mazo, J.E., Odlyzko, A.M.: Lattice points in high-dimensional spheres. Monatsh. Math. **110**, 47–61 (1990)
13. Minkowski, H.: Geometrie der Zahlen, I. Leipzig: Teubner 1896
14. Minkowski, H.: Gesammelte Abhandlungen. New York: Chelsea 1969 (Reprint)
15. Rogers, C.A.: Packing and Covering. Cambridge: Cambridge University Press 1964
16. Rogers, C.A.: Existence Theorems in the Geometry of Numbers. Ann. Math. II. Ser. **48**, 994–1002 (1947)
17. Rush, J.A., Sloane, N.J.A.: An improvement to the Minkowski-Hlawka bound for packing superballs. Mathematika **34**, 8–18 (1987)
18. Rush, J.A.: A lower bound on packing density. Invent. Math. **98**, 499–509 (1989)
19. Rush, J.A.: Thin lattice coverings. (Submitted)
20. Rush, J.A.: Constructive packings of cross polytopes. (Submitted)
21. Sloane, N.J.A.: Self-dual codes and lattices. In: Relations Between Combinatorics and Other Parts of Mathematics. Proc. Symp. Pure Math **34**, 273–308 (1979)
22. Sloane, N.J.A.: Recent bounds for codes, sphere packings and related problems obtained by linear programming and other methods. Contemp. Math. **9**, 153–185 (1982)
23. Terras, A.: Harmonic Analysis on Symmetric Spaces and Applications. Berlin Heidelberg New York: Springer, Vol. I, 1985 and Vol. II, 1988
24. Widder, D.V.: The Laplace Transform. Princeton: Princeton University Press 1941

Invent. math. 105, 641–662 (1991)

Inventiones mathematicae
© Springer-Verlag 1991

Small cancellation theory and automatic groups: Part II

S.M. Gersten[1,*] **and H. Short**[2]

[1] Mathematics Department, University of Utah, Salt Lake City, Ut 84112, USA
[2] Mathematics Department, City College, C.U.N.Y., Convent Avenue at 138th Street, New York, 10031, USA

Oblatum X-1989 & 6-II-1991

In this article we continue the work of [GS] (which we shall henceforth refer to as Part I). There we showed that small cancellation groups satisfying additional conditions are automatic (we refer the reader to Part I for definitions and further references). Recall that a group is said to be automatic if there is

(1) a finite set of semigroup generators,
(2) a finite state automaton accepting a language L of words in these generators which contains at least one representative for each element of G,
(3) for each generator a there is a finite state automaton which accepts the set of pairs of words in L which represent elements of G which differ by multiplication on the right by a.

The main results obtained in Part I can be rephrased in the language of root systems and Euclidean planar tessellations as saying that finite $A_1 \times A_1$ and A_2 piecewise Euclidean (abbreviated PE) complexes of non-positive curvature have automatic fundamental groups. (A 2-dimensional complex is a $A_1 \times A_1$, respectively A_2, PE complex if every 2-cell can be identified with a unit square, resp., equilateral triangle in the Euclidean plane, in such a way that the induced metrics agree on intersections. Such a complex has non-positive curvature if the link of each vertex contains no circuits without backtracking of length less than 2π).

Here we prove using similar techniques that finite 2-complexes modelled on the other two root systems, B_2 and G_2, corresponding to tessellations by 45° right angle triangles and by 30°, 60° right angle triangles, also have automatic fundamental groups (Theorems 2.3 and 3.2). To be precise,

Theorem 1 *A finite 2-complex of non-positive curvature modelled on the $A_1 \times A_1$, A_2, B_2 or G_2 root system has automatic fundamental group.*

From this it follows that a torsion-free discrete group acting, with compact quotient, by isometries of an Euclidean building of rank 2 where all vertices

* Partially supported by N.S.F. grants No. 850-5550 and DMS 880-1376.

have finite links, is automatic (Theorem 4.1). This result provides a rich class of examples of automatic groups.

We also use the B_2 case to extend the result of Part I that groups of prime alternating links are automatic to finite sums of alternating links (Theorem 4.2).

In another direction, we show how to lift the extra conditions (that pieces have length 1, and that no relation be a proper power) we imposed on the small cancellation groups in Part I, using the solution given here of the B_2 and the G_2 cases (Sects. 5 and 6). This means that we prove:

Theorem 2 *A group possessing of finite presentation satisfying a* $C(6)$, $C(4)-T(4)$ *or* $C(3)-T(6)$ *small cancellation condition is automatic.*

In Sect. 7 we conclude with a definition of biautomatic structures (where, for each semigroup generator a there is an automaton whose language consists of accepted words which differ by multiplication on the left by a, as well as an automaton for multiplication on the right). All known examples of automatic groups have such biautomatic structures, and all the structures constructed here, in Part I and in [CEHPT] have this additional property. We prove:

Theorem 3 *A group with a biautomatic structure has a solvable conjugacy problem.*

The corresponding problem for automatic groups is open. The solution of the conjugacy problem for the $C(p)-T(q)$ small cancellation groups (for $(p, q)=(3, 6), (4, 4)$, or $(6, 3)$) was established by Schupp [LS] while for negatively curved (or word hyperbolic) groups the solution was given by Gromov [G, 7.4.B]. The proof offered here gives a unified treatment ot the problem.

Since this work was completed, we have further developed the notion of biautomatic group, and gone some way to understanding the subgroup structure of these groups. In particular we present in [GS 2] an alternative proof of Theorem 3.

1 Definitions and brief review of Part I

Corresponding to the two root systems (covering the plane by 45° right angle triangles and by 30°, 60° right angle triangles), we make the following two definitions of B_2 and G_2 complexes. Note that the G_2 case has a further restriction concerning orientations.

Definition. A B_2 *complex* is a PE complex, where all 2-cells are isometric to an isosceles right angle triangle $\mathbb{E}^2$ with short side of length one.

Let K be a finite PE complex where all 2-cells are oriented isometric to a $\frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}$ triangle with short side of length $\frac{1}{2}$, subject to further restrictions. The 1-cells in the boundaries of 2-cells thus fall into three equivalence classes, according to their lengths, which are $\frac{\sqrt{3}}{2}$, $\frac{1}{2}$ or 1 – call these type x, y and z edges. We assume in addition the vertices fall into equivalence classes X, Y, Z according to whether they lie opposite an x, y or z edge in some 2-cell. We orient the edges of a standard such triangle in $\mathbb{E}^2$, such that an edge is

oriented from the shorter to the longer of the edges that it meets. We say that K is a G_2 *complex* if it satisfies the following additional conditions:

(1) each 1-cell is oriented
(2) each two cell is oriented isometric to the standard triangle (by an isometry respecting the orientation on the 1-skeleton):

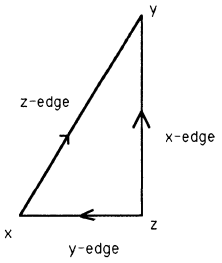


Fig. 1. Standard 2-cell for a G_2 complex

(3) if an x-edge e_1 and a y-edge e_2 have a common initial vertex, then there is a 2-cell containing these two edges.

In the non-positive curvature context in which we are principally interested, the 2-cell whose existence is asserted in condition (3) is necessarily unique.

The tessellations of the euclidean place $\mathbb{E}^2$ by isosceles right angle triangles and by (oriented) 30°, 60° right angle triangles we shall refer to as B_2 and G_2 tessellations.

A PE complex K has *non-positive curvature* (resp. *negative curvature*) if every nontrivial circuit without backtracking in the link of a vertex v subtends an angle of at least 2π (resp. greater than 2π) at v. To understand the subtended angle, we look at the *star graph* for the vertex v. This graph has a vertex for each oriented edge of $K^{(1)}$. Thus the vertices of the star graph occur in pairs $e, \bar{e}$, where $\bar{e}$ denotes the edge oppositely oriented to e. The attaching map of a 2-cell in K gives a circuit in $K^{(1)}$, defined up to inversion and starting point. If the syllable ab occurs in an attaching map of a 2-cell α of K (and thus $\partial_1(a) = \partial_0(b)$) then join the vertex labelled a to the vertex labelled b^{-1} in the star graph by an edge, and assign to this edge a length equal to the angle at the corner corresponding to ab in α measured in radians.

A PE complex has non-positive curvature if the sum of the lengths in a nontrivial circuit without backtracking in the star graph of any vertex is at least 2π. For instance, the 2-complex corresponding to the presentation $\langle a, b | abab^{-1}b^{-1} \rangle$ has non-positive curvature.

Some definitions recalled from Part I

We shall use the techniques of Part I. Let $\langle X; R \rangle$ be a presentation of the group G; as usual we take the set $\mathscr{A} = X \cup X^{-1}$ as a set of semigroup generators for G, and let $\mathscr{A}^*$ denote the free semigroup on $\mathscr{A}$.

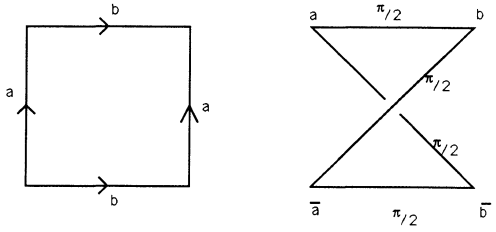


Fig. 2. Star graph for $\mathbb{Z} * \mathbb{Z}$

We denote the element of G represented by the word $w \in \mathscr{A}^*$ by $\bar{w}$. The main theorem used to assure the existence of an automatic structure is:

Theorem 1.1 ([CEHPT], see also [BGSS, 2.10] and Part I) *Suppose that there is a regular language of words in $\mathscr{A}^*$ which maps naturally on to the group G as geodesic (i.e. shortest) representatives. Suppose in addition that there is a constant k such that words in this language which begin and end at most distance 1 apart in the Cayley graph lie in a k-Hausdorff neighborhood of each other. Then G is automatic, and L is the language of an automatic structure for G.*

We describe the language in terms of turning angles, as in Part I. Let K be a PE complex, and let $v = a_1 \ldots a_n$ be a word in the labels on the edges of a closed path in $K^{(1)}$. If the word v represents the trivial word in $\pi_1(K)$, then there is a singular PE disc diagram whose boundary label is v (see Part I) and whose regions are isometric copies of the 2-cells of K (all diagrams are assumed reduced). If a segment of the boundary of a singular disc diagram is labelled by the word w, we say that the diagram *includes* w. Let D be a singular PE disc diagram which includes the word $w = a_1 \ldots a_n$. Then to each a_i, $1 < i < n$, D associates an ordered pair of turning angles (θ, θ') ocurring at the initial and final vertex of the edge of ∂D labelled a_i. If the edges corresponding to adjacent letters a_i and a_{i+1} do not lie on the topological boundary of an open 2-ball in D, then we assign the value $-\infty$ to the turning angle at the vertex between these two edges. In fact, for each $i = 1, \ldots, n-1$ the diagram D associates an angle θ_i to the vertex common to the boundary edges of D corresponding to a_i and a_{i+1}, and thus associates a sequence $\theta_1, \ldots, \theta_{n-1}$ of turning angles to the word w.

2 B_2 complexes of non-positive curvature

Let K be a finite B_2 complex of non-positive curvature.

We follow the steps of Part I, first giving explicitly a regular language for the fundamental group $\pi_1(K) = G$, and then establishing the various propositions required to show that this language satisfies the conditions of Theorem 1.1, and thus forms part of an automatic structure for G. To do this we shall consider reduced (singular) disc diagrams which inherit the PE structure from the B_2 complex, by pulling back the structure from K. We shall restrict our interest to the case when the boundary contains no hypotenuse edge: we call these

(singular) B_2 disc diagrams. Notice that a B_2 diagram becomes a singular disc diagram where all cells are rectangles by deleting all hypotenuse edges.

Let K be a B_2 complex, and let X be the set of oriented edges, and R the set of cyclic words in X corresponding to the attaching maps of the 2-cells. The fundamental group of K is a free factor of the group $\langle X; R\rangle$, which is the fundamental group $\pi_1(K)$ if K has just one vertex (we shall not specify the base point here). Let $\mathscr{A}$ denote the set of non hypotenuse edges; clearly these elements generate $\pi_1(K)$. Let $\Gamma_{\mathscr{A}}$ be the Cayley graph of $\pi_1(K)$ with respect to the generating set $\mathscr{A}$. Distance in the Cayley graph is measured by the metric induced by regarding each edge of $\Gamma_{\mathscr{A}}$ as an isometric copy of the unit interval.

Definition. A word $w = a_1 \dots a_n$ in $\mathscr{A}^*$ is called a B_2-*geodesic* if in all B_2 diagrams which include w, the sequence $\theta_1, \dots, \theta_{n-1}$ of associated turning angles never contains a non-negative turning angle followed by a positive turning angle.

We now follow the sequence of Part I to show that Theorem 1.1 applies.

Proposition 2.1 *Every element of* $\pi_1(K)$ *has a* B_2*-geodesic representative.*

Proof. Given an element $g \in G$, let w be a geodesic word w in $\mathscr{A}^*$ representing g.

Consider the various cases where a positive turning angle follows a non-negative turning angle in the boundary of a singular B_2 disc diagram which includes the word w. Notice that all turning angles associated are multiples of $\pi/2$ because of the restriction to words which contain no hypotenuse edges.

If a $\pi/2$ turning angle is followed a $\pi/2$ turning angle, then w could be shortened.

If a turning angle of $\pi/2$ follows a turning angle of zero, say at the letter a_i, then the following change is made, to give the word w'.

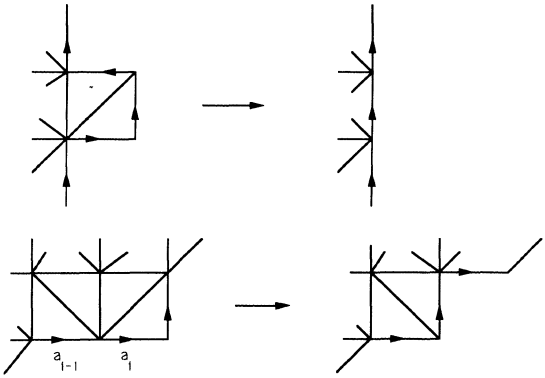


Fig. 3. Making a word into a B_2-geodesic

Notice that if the subword $a_{i-1}\, a_i$ occurs in another reduced disc diagram, then the turning angle between a_{i-1} and a_i cannot be $\pi/2$. If this were the case, then the non-positive curvature condition would be violated.

Let $H(i, w)$ be the number of values of j, $1<j\leqq i$, such that in some B_2 diagram which contains w, the turning angle θ_j is $\pi/2$; here $i\leqq n-1$ where n is the length of w.

Then, as in Proposition 5.4 of Part I,

$$\sum_2^n H(i, w)<\sum_2^n H(i, w')<n(n+1)/2,$$

where w' is the word obtained from w by the change illustrated in Fig. 3.

It follows that this process of 'pushing right angles to the front' terminates and gives a B_2-geodesic. □

To determine whether a word w is a B_2-geodesic or not, it suffices to check whether w contains one of a finite list of subwords. This remark proves:

Proposition 2.2 *The language of B_2-geodesics is regular.*

As all turning angles are multiples of $\pi/2$, deleting hypotenuse edges from a B_2 singular disc diagram gives a PE disc diagram where all cells are rectangles (i.e. an $A_1\times A_1$ diagram). The methods of Sect. 5 of Part I can now be used, and the proofs of Propositions 5.6 and 5.7 show that

Proposition 2.3 (i) *A B_2-geodesic is geodesic.*
(ii) *If two B_2-geodesics begin and end distance at most 1 apart in the Cayley graph $\Gamma_{\mathscr{A}}$, then they lie in a 2-Hausdorff neighborhood of each other.*

It thus follows that:

Theorem 2.4 *The fundamental group of a finite B_2 complex of non-positive curvature is automatic.*

3 G_2 Complexes of non-positive curvature

Recall that in a G_2 complex K, each 2-cell is oriented isometric to a standard triangle (Fig. 1).

Choose as base point for K some X-vertex. Let $\mathscr{A}$ be the set of y- and z-edges; clearly these elements generate $\pi_1(K)$. A G_2 disc diagram is defined analogous to a B_2 disc diagram, where here we do not have x edges on the boundary. As for B_2 complexes, we define the set of G_2 geodesics by conditions on associated turning angles, but now, instead of just considering adjacent turning angles, we consider triples of turning angles.

Definition. In a G_2 diagram which includes the word $w=a_1\ldots a_n\in\mathscr{A}^*$ let $\theta_1, \ldots, \theta_{n-1}$ be the associated turning angles. We say that w is a G_2 *geodesic* if for all G_2 diagrams which include w,

(1) if a_i is a z-label then we do not have $\theta_{i-1}\geqq 0$ and $\theta_i>0$;
(2) if a_j, a_{j+1} are y-labels, then we do not have $\theta_{j-1}\geqq 0$ and $\theta_j=0$ and $\theta_{j+1}>0$.

Proposition 3.1 *Every element of $\pi_1(K)$ has a G_2 geodesic representative.*

Proof. Let w be a word representing $g\in\pi_1(K)$, where w is chosen to minimise the G_2-length $\ell_G(w)=2\#\{z\text{-labels in } w\}+\#\{y\text{-labels in } w\}$. We show how to alter w to make it satisfy (1) and (2) above, without altering the G_2 length.

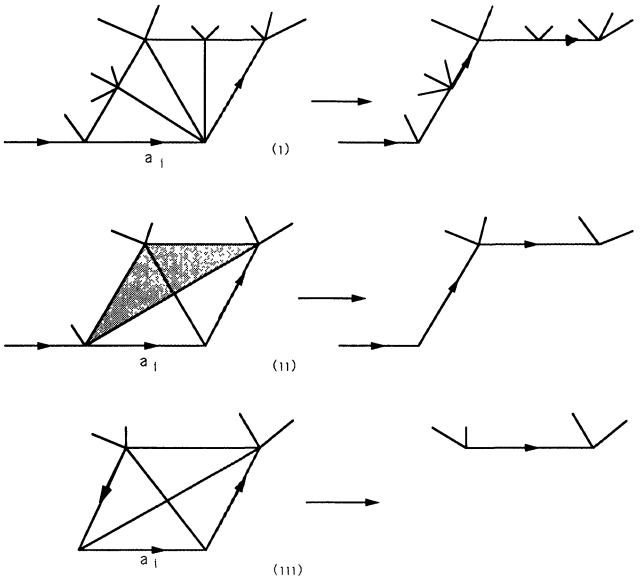


Fig. 4. Configurations prohibited by (1)

Case (1). *Suppose* (1) *is not satisfied*

The indicated changes are made. Notice that in case (ii) the condition (3) in the definition of G_2 complex is required to guarantee the existence of the shaded 2-cells. Case (iii) does not occur as it would lead to a reduction of G_2-length.

Case (2). *Suppose* (2) *is not satisfied*

The path continues past the edge labelled a_{j+1}. The path continues past the

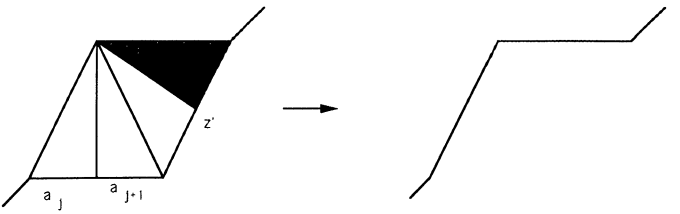


Fig. 5. Case (2)

Z-vertex z', as it must end at an X-vertex. The shaded triangle exists by condition (3) of the definition, so the indicated change can be made, without increasing the G_2 length.

Now set $H(i, w)$ to be the number of turning angles θ_j associated to w with $i \leqq j$ which can be given a value $< \pi$. As before, the above changes increase this value, so the process of replacement cannot continue indefinitely. □

Now everything reduces to the A_2 case, as a G_2 disc diagram which has no X-edges on the boundary, becomes an A_2 diagram when the X-edges are deleted from the interior. Thus § 5 of Part I shows that

Theorem 3.2 *The fundamental group of a G_2 complex of non-positive curvature is automatic.*

4 Applications

If we widen the concept of Euclidean building to allow the $A_1 \times A_1$ tesselation of the plane[1] we obtain:

Theorem 4.1 *Let X be a Euclidean building of rank 2, such that the link $\mathrm{Link}_X(v)$ of each vertex v is a finite complex. Let G be a discrete torsion-free group of isometries of X with compact quotient.*

Then G is automatic.

Proof. Let Γ denote the group of isometries of X; this is a locally compact group. The stabilizer Γ_v of the vertex v is profinite, i.e. the inverse limit of a sequence of finite groups, and thus compact: one sees this as follows. Let $B_v(n)$ be the ball of radius n about v, and let $\Gamma_v(n)$ be the subgroup of Γ which restricts to the identity on $B_v(n)$. Then we have projections

$$\Gamma_v \twoheadrightarrow \Gamma_v/\Gamma_v(n)$$

and

$$\bigcap_n \Gamma_v(n) = 1 \quad \text{so} \quad \Gamma_v \xrightarrow{1-1} \varprojlim \Gamma_v/\Gamma_v(n).$$

Now G is a discrete group, and so the stabilizer G_v of v is a discrete subgroup of a compact group, and so is finite. As G is assumed to be torsion-free, G_v is trivial, and the action is thus free. It follows that the quotient map $X \to X/G$ is a regular covering map, and so the quotient is a $A_1 \times A_1$, A_2, B_2 or G_2 complex. It remains to show that the non-positive curvature condition is satisfied in the quotient complex.

As the action is free, it is sufficient to show that the curvature condition is satisfied in X. Let C be a non-trivial reduced circuit in $\mathrm{Link}_X(v)$. If we cone the circuit C to the vertex v, we get a Euclidean cone; let θ denote the cone angle. If $\theta < 2\pi$, there are pairs of points in the cone which have two geodesic ares joining them, as can be seen as follows. If we cut the cone along a straight line from the cone point, and flatten this into a planar sector of internal angle θ, we see that there are two straight lines from a point on the angle bisector to points which were previously identified. In this way we see that geodesic arcs between points are not unique; this cannot happen in a Euclidean building (see [B, part (4) of the Theorem on p. 152]). □

[1] A building as defined in [B] is a simplicial complex. We understand buildings in the original sense of [BT] to be polysimplicial complexes subject to their axioms. In this latter sense the $A_1 \times A_1$ tesselation of $\mathbb{E}^2$ is an apartment.

Remark. We need G to be torsion-free in 4.1 since it is not known otherwise that G is linear, and hence whether (by Selberg's lemma) it contains a torsion-free subgroup of finite index.

In Part I, we used Weinbaum's description [W] of an alternating link group as a free factor of a $C(4)-T(4)$ (in fact the fundamental group of a finite $A_1 \times A_1$ complex) group, to show that such groups are automatic. We can now generalize this to:

Theorem 4.2 *The group of a connected sum of alternating links in automatic.*

Proof. Given a planar projection of a knot or link k, we can obtain a presentation (known as the Dehn presentation) of $\pi_1(S^3-k)*\mathbb{Z}$ where the generators correspond to the regions into which the projected curve divides plane, as follows. Enclose the projection in a large circle, and take a generator m_i for each compact region. We think of m_i as the homotopy class of a loop based at $+\infty$, the point from which projection takes place (or the viewer's eye), down through the region of the plane of projection corresponding to m_i, and on down to $-\infty$. As we are in S^3, this is in fact a loop. This provides a set of generators for the link group.

The relations for the presentation are of the form $m_1 m_2^{-1} m_3 m_4^{-1}$, when the regions labelled m_1, m_2, m_3, m_4 meet at a crossover.

When the link projection under consideration is alternating, the link is prime, and in addition there are no trivial crossings, there are four distinct regions at every crossing (see [W]). Moreover, the presentation obtained by the above method satisfies the $C(4)-T(4)$ small cancellation conditions. A minimal alternating projection of a prime knot or link satisfies all the above conditions.

For instance, the projection of Fig. 6 for a figure of eight knot gives a presentation of $\pi_1(S^3-k)*\mathbb{Z}$ with generators $m_1, m_2, m_3, m_4, m_5, m_6$ and relators

$$m_1 m_2^{-1} m_3 m_4^{-1}, m_1 m_6^{-1} m_3 m_2^{-1}, m_3 m_6^{-1} m_5 m_4^{-1}, m_5 m_6^{-1} m_1 m_4^{-1}.$$

Killing the generator corresponding to the region which has the extra added curve on its boundary gives the link group, but it is convenient to work with the free product as it is the above presentation which satisfies the small cancellation condition.

We deal first with the case of a sum of two prime links. In the above presentation, if m, m' are generators corresponding to neighboring regions, then mm'^{-1} is a meridian of the knot. The two presentations obtained give $A_1 \times A_1$ complexes K_1 and K_2.

Cutting each square into two triangles along a hypotenuse (in each square choose one of the two ways of doing this) turns K_i into a B_2 complex K_i'. Let m_i, m_i' be neighboring regions in the alternating projection of k_i, such that m_i, m_i' are neighboring non-hypotenuse edges in some triangle in K_i'. Then

$$\pi_1(S^3-(k_1 \# k_2)) = \pi_1(S^3-k_1)*_{m_1 m_1'^{-1} = m_2 m_2'^{-1}} \pi_1(S^3-k_2).$$

So we add to the disjoint union of the complexes K_i a cylinder triangulated as shown in Fig. 7, with the ends identified to the edges labelled m_i, m_i', and then identify all vertices.

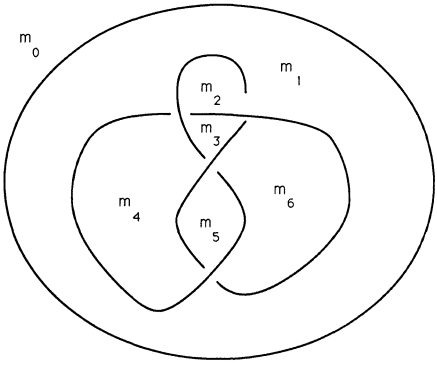


Fig. 6. A figure of eight knot

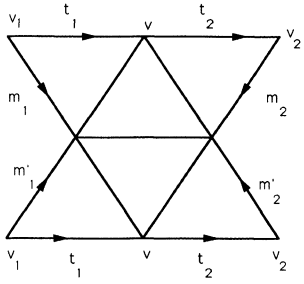


Fig. 7.

The fundamental group of the resulting B_2 complex is

$$\pi_1(S^3-(k_1\#k_2))*F_4$$

and the result follows for the sum of two prime alternating links.

Now to add on another alternating link to the sum already obtained, we follow the same procedure, again obtaining a B_2 complex. □

5 $C(4)-T(4)$ groups are automatic

We refer the reader to Part I for the relevant definitions of small cancellation conditions. For the rest of this section $\mathscr{P}=\langle X\,|\,R\rangle$ is a finite $C(4)-T(4)$ presentation of the group G, and K is the corresponding 2-complex. Recall that a non-trivial word $p\in F(X)$ is called a *piece* if there are two different words r_1, r_2 which are cyclic conjugates of a word (or the inverse of a word) in R such that $r_1=pr_1'$ and $r_2=pr_2'$. Each (cyclic) word $r\in R$ can be decomposed into no fewer than four pieces, possibly in several different ways. In Part I we avoided

the problem of different decompositions of a given relator by insisting that the pieces all be single letters.

The proof proceeds by first obtaining a new B_2-like 2-complex K' related to K where we can define turning angles, though lengths are not well-defined. In this new complex we choose a new set of generators for the fundamental group which we use to define representative words which have the same form as B_2-geodesics. The automatic structure is then constructed as in the B_2 case.

We build K' as follows: take a stellar decomposition of the 2-cell corresponding to $r=a_1 \dots a_n$. This introduces new 1-cells e_i from the barycenter to the initial vertex of the edge a_i, which we orient outwards from the barycenter. The 2-cell for r is thus replaced by n 2-cells, with attaching maps $e_i a_i e_{i+1}^{-1}$. For each piece $p=a_i \dots a_j$ of length greater than 1 occurring in the cyclic word r, add a new 2-cell to K with attaching map $e_i p e_{j+1}^{-1}$. Finally identify all the new barycenter vertices with the original vertex.

Let $E(r)=\{e_i | i=1, \dots, n\}$, and $E=\bigcup_{r \in R} E(r)$.

For instance, take the presentation $\langle x, y; x^3=y^2 \rangle$; the relator has three decompositions into pieces:

$$x \cdot x \cdot x = y \cdot y, \qquad (x^2) \cdot x = y \cdot y, \qquad x(x^2) = y \cdot y$$

which give the 2-cells shown in Fig. 8.

It is easy to see via Tietze transformations that:

Lemma 5.1 *$\pi_1(K')$ is the free product of G and a finitely generated free group.*

In K', we model each 2-cell on a 45° triangle, where the hypotenuse edge corresponds to the piece on the boundary. Notice that this does not impose a piecewise Euclidean structure on K', as different subdivisions of a relator would in general have to assign different lengths to the 1-cells common to different pieces.

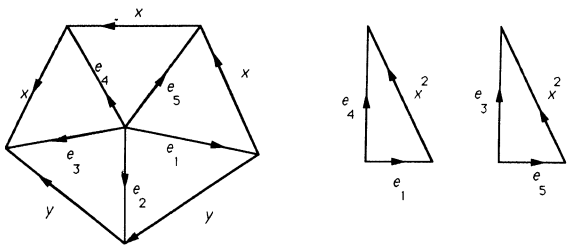


Fig. 8. Decomposition of a 2-cell

We now study maps of pairs $(D, \partial D) \to (K', K'^{(1)})$. After a homotopy, the cell structure of K' pulls back to a cell decomposition of a simply connected 2-complex which we call a singular disc diagram D. The orientations of the edges labelled in E divide the vertices in the interior of D into two classes. We refer to the vertices of D which correspond to barycenters, where the edges labelled

in E are all oriented outwards, as B-vertices. The others, where the edges labelled in E are all labelled inwards, we call V-vertices. If the boundary ∂D spells out the word w in $\pi_1(K'^{(1)})$, then we say that D is a disc diagram for $w=1$ in $\pi_1(K')$. Notice that the set E generates $\pi_1(K')$.

If two faces have an edge in common, and have the same label on their boundaries, then they can be cancelled, giving a diagram with fewer faces. A diagram is *reduced* if no faces can be cancelled. From now on, unless otherwise stated, all diagrams are assumed to be reduced.

Definition. We define an *admissible* disc diagram to be a singular disc diagram such that whenever two hypotenuse edges in D have a non-empty subsegment in common, then these entire edges are identified in D.

This means that in an admissible disc diagram D, the preimages of the vertex of the complex K' which lie in the interior of a hypotenuse edge meet no such other edge. It follows from the definition of admissible diagram that a hypotenuse edge in the interior is labelled by a piece, and that two hypothenuse edges either coincide or have no subsegment in common.

Definition. The regular language for $C(4)-T(4)$ presentations.

An *allowed* word is a cyclically reduced word in $F(E)$ corresponding to an edge path forming a circuit without backtracking in $K'^{(1)}$, satisfying the B_2 turning angle condition: in any admissible diagram which includes the word w, a non-negative turning angle is never followed by a positive turning angle (see Figs. 3, 4).

To obtain a diagram of non-positive curvature from a given diagram, we show how to remove interior vertices of positive curvature (i.e. where the sum of the corner angles is less than 2π) in a cell decomposition of a singular disc diagram.

Lemma 5.2 *If w is an allowed word representing the identity element of $\pi_1(K')$, then there is an admissible singular disc diagram of non-positive curvature at all interior vertices.*

Proof. Let D be a diagram for $w=1$ in $\pi_1(K')$. We first show that we can choose D to be admissible, and then we show how to remove interior vertices of positive curvature.

The important thing to note is that a non-trivial subword of a piece is also a piece, and so there is a 2-cell in K' with the hypotenuse edge labelled by that subword. Suppose that the hypotenuse edge h of a face f meets the hypotenuse edge h' of the face f' in a proper subsegment $h'' \subsetneq h$. The faces f, f' can be subdivided into faces corresponding to the subpieces labelling the segment h'' and the complement of h'' in h. In this way the diagram D can be made an admissible.

So from now on we assume that D is admissible.

Case 1: B-vertices of positive curvature

As all corner angles at a B-vertex are $\pi/2$, the only cases to consider are B-vertices meeting 2 or 3 edges.

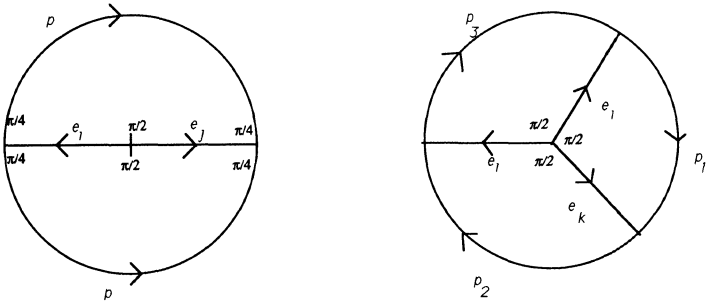


Fig. 9. *B*-vertices of positive curvature

In case (a), the two faces which have the vertex v on their boundary can be cancelled, as the diagram is not reduced (e_i and e_j define a unique pice p).

In case (b), there are neighboring faces in D, as the edges opposite the B-vertex cannot lie on the boundary of the diagram. The three pieces p_1, p_2, p_3 come from the same relator, but the $C(4)$ condition implies that three pieces do not cover the relator. Thus, for some choice of indices, $p_1 = p_3^{-1} p_2^{-1}$ and the diagram can be altered locally by subdivision and cancellation as shown in Fig. 10.

Although this change increases the curvature at two of the vertices, the number of faces and the number of vertices in the diagram is decreased.

Case 2: V-vertices of positive curvature

Interior V-vertices have even order, as E-edges alternate with hypotenuse edges.

In case (a), the two faces cancel, as before.

In case (b), the four faces can be replaced by two faces, as the two edges corresponding to pieces in $\pi_1(K)$ can be combined into a single piece, thus reducing the number of faces.

In case (c), the pieces p_j and p_{j+1} ($j = 1, 2, 3$, subindices mod 3) are subwords of a (cyclic) relator $r_j \in R$, as they lie on the boundary of faces f_j, f_j' which have a common edge labelled in E.

Either (i) we can write r_j as $p_j r_j' p_{j+1}^{-1}$, or (ii) the shorter, say p_j, of the words p_j, p_{j+1}, of the two words is an initial segment of the longer.

In subcase (ii), subdivide the faces f_2 and f_2' adjacent to the hypotenuse edge p_2 so that this common edge is divided in two, the first segment being labelled p_1. This creates two new faces, one of which can be cancelled with face f_1', producing a diagram with the same number of faces as before.

So suppose that subcase (ii) does not hold for any j. Then we can write the relators r_1, r_2, r_3 as:

$$r_1 = p_1 r_1' p_2^{-1}, \quad r_2 = p_2 r_2' p_3^{-1}, \quad r_3 = p_3 r_3 p_1^{-1}.$$

Then $r_j r_{j+1}$ is not cyclically reduced, and so by the $T(4)$ condition on G, two of these relators are identical, and two of the faces f_j', f_{j+1}' of D can be cancelled to give a diagram with fewer faces.

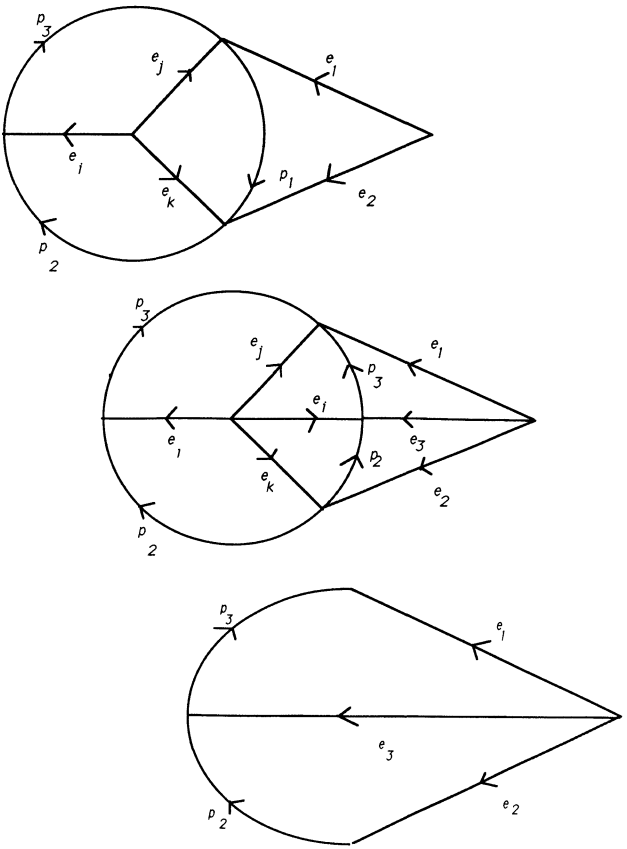


Fig. 10. Removal of a positive curvature B_2-vertex

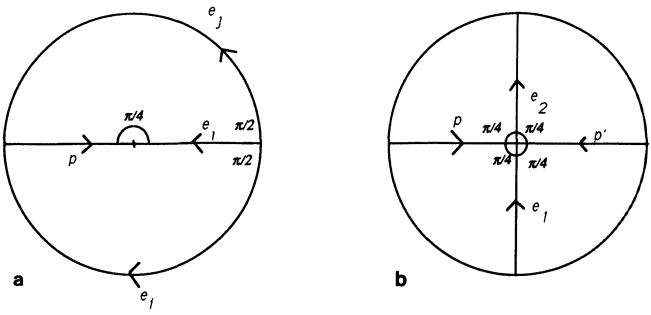


Fig. 11 a, b. V-vertices of order 2 and 4

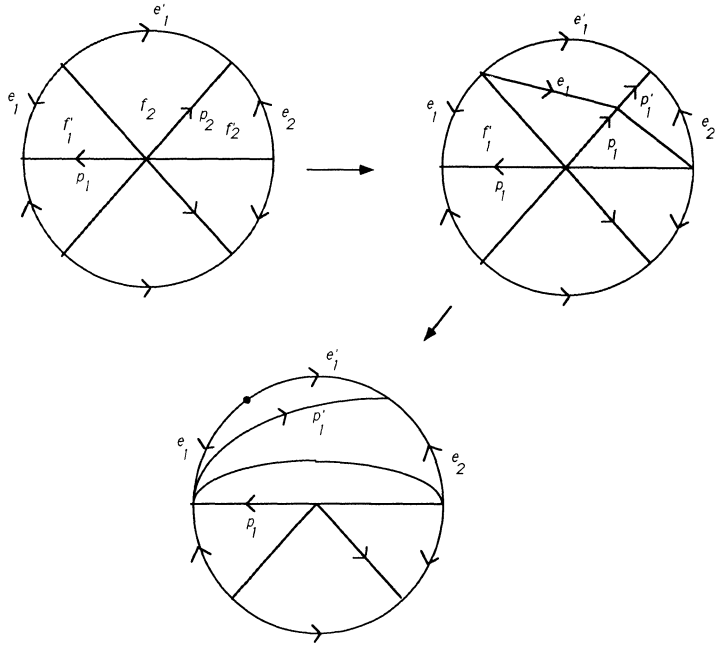


Fig. 12. Removing a V-vertex of order 6

It follows that a minimal admissible diagram for $w=1$ in $\pi_1(K')$ has non-positive curvature at all interior vertices, and the proof of Lemma 5.2 is complete. □

To establish the fact that the $C(4)-T(4)$ small cancellation groups are automatic, we now follow the argument for the B_2 case. We say that a word $w \in F(E)$ is an allowed representative of $g \in \pi_1(K')$ if it is shortest in $F(E)$ representing g, and satisfies the B_2 turning angle condition (i.e. this is just like an allowed word, but without the cyclically reduced condition, as we are not now considering cyclic words).

As in the B_2 case, each element of $\pi_1(K')$ has an allowed representative (see Proposition 2.1). For instance, given a geodesic representative w for the element $g \in G$, if a turning angle of $\frac{\pi}{2}$ follows a turning angle of 0 in a diagram D, then we can change w as in Fig. 3.

Only a finite number of configurations of subwords are used to define allowed representatives, so the language of allowed representatives is regular.

By considering admissible diagrams of non-positive curvature, the proof of Proposition 2.3 shows that two allowed representatives which begin and end distance at most 1 apart in the Cayley graph $\Gamma_E(\pi_1(K'))$ lie in a 2-Hausdorff neighborhood of each other.

Thus:

Proposition 5.3 *A group admitting a finite presentation satisfying a* $C(4)-T(4)$ *small cancellation condition is automatic.*

$C(6)-T(3)$ groups are automatic

The proof of this follows the $C(4)-T(4)$ case, though here we shall make use of the A_2 complexes.

Take the stellar subdivision of each 2-cell from the barycenter of each as before, adding a set E of 1-cells. For each relation r_i, we add a 2-cell for each piece, and identify the new barycenter vertices with the original vertex of K. Call the resulting complex K', where we give each 2-cell the structure of an equilateral triangle. As in the $C(4)-T(4)$ case, this does not define a piecewise Euclidean structure on K'. We now consider diagrams corresponding to maps of discs into the 2-complex K'.

Definition. An *admissible* diagram is a singular disc diagram such that if two edges in the interior of the diagram are labelled by pieces (i.e. when they are not E-edges) and intersect in a segment with non-empty interior, then the edges are identified. (An edge labelled by a piece of length greater than one consists of several segments, each of which we can think of as being labelled by a generator. If two pieces have a common subword, then in a diagram two cells could meet in the subsegment corresponding to the common subword.)

As before, the orientation of the E-edges divides the vertices of an admissible diagram into two classes, the B vertices corresponding to the barycentre vertices, where all edges are E-edges and oriented outwards, and V-vertices, where the edges are alternately E-edges (oriented inwards) and edges labelled by pieces.

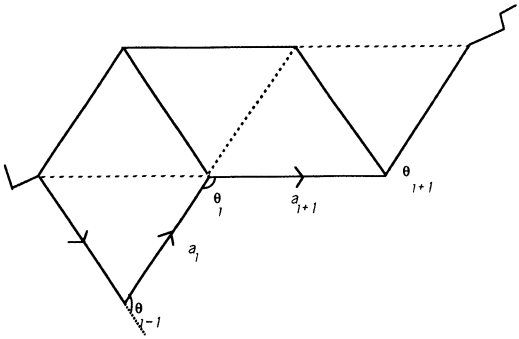


Fig. 13. Disallowed configuration of type (2) (dotted lines are non E edges)

An allowed word is a cyclically reduced word w in $F(E)$ corresponding to a circuit without backtracking, which satisfies the A_2 turning angle condition (1), and another more complicated condition (2). In any admissible A_2 diagram which includes $w=a_1 \dots a_n$, if $\theta_1 \dots \theta_{n-1}$ are the associated turning angles, then

(1) a positive turning angle does not follow a non-negative turning angle,

(2) we do not have $\theta_{i-1}=\frac{2\pi}{3}$, $\theta_i=-\frac{\pi}{3}$, $\theta_{i+1}>0$.

The reason for this more complicated second condition is to prohibit situations as in Fig. 14 below, where the paths grow arbitrarily far apart, though a positive turning angle never follows a non-negative one.

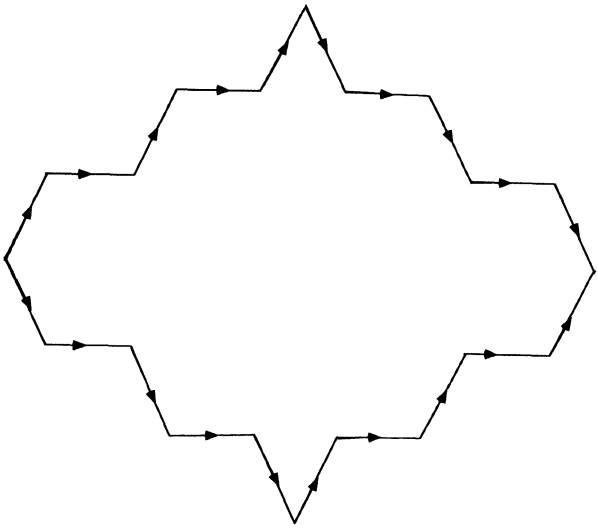


Fig. 14.

Lemma 6.1 *If w is an allowed word representing the identity element of $\pi_1(K')$, then there is an admissible disc diagram for w with non-positive curvature at all interior vertices.*

Proof. As in the proof of Lemma 5.2, a diagram can be made admissible by subdivision of edges labelled by pieces, so we can restrict our attention to admissible diagrams.

In the complex K', all corner angles are $\frac{\pi}{3}$, so that a vertex in the interior of a diagram has positive curvature if it meets fewer than 6 edges.

Case 1: B-vertices of positive curvature

These cases are treated in the same way as the case of B-vertices in Lemma 5.2.

A B-vertex of order 2 can be removed by cancellation of faces.

The $C(6)$ condition says that no relation is a product of fewer than 6 pieces. If $p_1, \ldots, p_n$, where $2<n<6$ are the pieces opposite a B-vertex, then as in subcase (b) of 5.2, for some index j, p_j^{-1} is an initial segment of the piece p_{j+1} or p_{j+1}^{-1}

is an initial segment of p_j (considering the indices cyclically). Subdivision of the corresponding face leads to cancellation and reduction of the number of vertices in the diagram.

Case 2: V-vertices of positive curvature

The fact that edges meeting a V-vertex are alternately E-edges and edges corresponding to pieces means that there an even number of edges, and so the only cases of interest are when there are 2 or 4 edges at a V-vertex. These two cases are dealt with as in the analogous subcases (a) and (b) of 5.2. □

We now continue as in Sect. 5. Define an allowed representative to be a reduced word in $F(E)$ satisfying the conditions for allowed words (without the restriction of being cyclically reduced), and shortest among all words representing the same element of $\pi_1(K')$.

A given word can be converted into an allowed word by operations, as before (specifically see Sect. 5 of Part I).

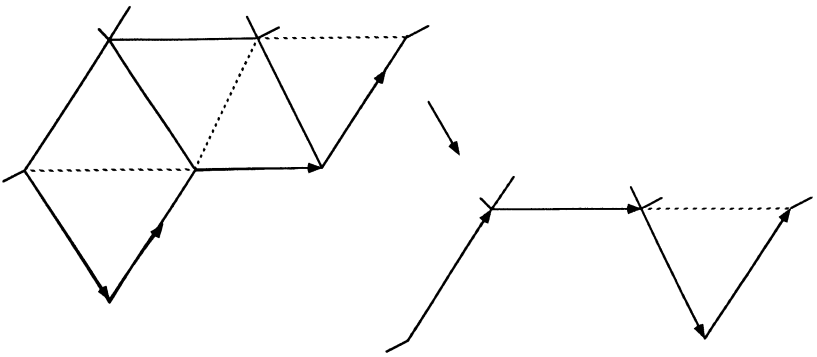


Fig. 15. Changing a disallowed configuration of type ('2)

To show that each element of $\pi_1(K')$ has an allowed representative, we must show that the process of changes cannot continue indefinitely. Here for the function $H(i, w)$ we use a count of the number of turning angles of $-\frac{\pi}{3}$ up to the i-th letter). When a change of type (2) occurs, as in Fig. 15, such an angle is interchanged with an earlier angle of $\frac{\pi}{3}$. Thus the sum of the $H(i, w)$ is definitely increased with each change.

It is clear that the language of allowed representatives is regular, as the conditions prohibit a finite number of ocnfigurations, which can be expressed in terms of finite state automata.

The results on A_2 diagrams of positive curvature contained in Part I, Sect. 6, show that allowed words which begin and end distance at most one apart lie in a 2-Hausdorff neighborhood of each other. Thus we have

Proposition 6.2 *A group which has a finite presentation satisfying the small cancellation conditions $C(6)-T(3)$ is automatic.*

7 $C(3)-T(6)$ groups are automatic

This case follows exactly the proof of Sect. 6. After subdivision of 2-cells etc., each triangle is given the structure of an isoceles $\frac{2\pi}{3}, \frac{\pi}{6}, \frac{\pi}{6}$ triangle with short sides of length one. We can think of this triangle as two copies of the standard triangle of the G_2 type, identified along the shortest side.

We define admissible diagram and allowed words as in § 6. In this case, the following configurations are disallowed:

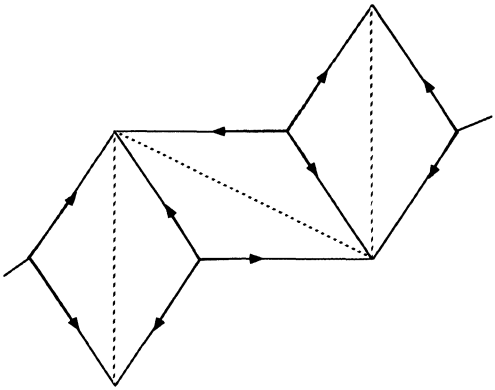


Fig. 16. Disallowed configuration

Lemma 7.1 *If an allowed word w represents the identity element of $\pi_1(K')$ then there is an admissible diagram for w with non-positive curvature at each interior vertex.*

Proof. As in 5.2, we can restrict our attention to admissible diagrams. Now all angles at a B-vertex are $\frac{2\pi}{3}$, and all angles at a V-vertex are $\pi/6$.

Case 1: B-vertices of non-positive curvature

As the angles involved are $\frac{2\pi}{3}$, the only case to consider is a B-vertex of order 2, which can be removed by cancellation of the neighboring faces as in 5.2.

Case 2: V-vertices of non-positive curvature

Here we must consider V-vertices of order less than 12. At a V-vertex, the E-edges and the edges labelled by pieces alternate. The proof is concluded by

treating such a vertex as in case 2(c) of 5.2, where the $T(6)$ condition is used instead of the $T(4)$ condition. □

The proof now concludes exactly as in § 5. We define an admissible representative as in § 6, and then we see that every element of $\pi_1(K')$ has such a representative, and that admissible words beginning and ending distance one apart lie in a 2-Hausdorff neighborhood of each other. Thus we have:

Proposition 7.2 *A group which has a finite presentation satisfying* $C(3)-T(6)$ *small cancellation condition is automatic.*

8 Remarks on the conjugacy problem

One of the interesting facts about automatic groups is that they have solvable word problem (see [CEHPT] or [BGSS, § 2.4]). At present it is unknown whether automatic groups have solvable conjugacy problem. The automatic structures constructed here and in Part I, have an additional property which gives a solution to the conjugacy problem using finite state automata. A way to think of this is that up until now, an automatic structure provides finite state automata which tell when two words differ by multiplication on the right. We can however equally well consider multiplication on the left, and multiplication on both sides. Let $\mathscr{A}$ be a finite set of semigroup generators for the automatic group G, and let L be the language of accepted words in $\mathscr{A}^*$ for an automatic structure on G. We say that L is the language of a *biautomatic structure* for G if in addition for each $a \in \mathscr{A}$, the language

$$\{(u, v)^{\$} \mid u, v \in L, \quad \text{and} \quad \bar{u} = \overline{av}\}$$

is a regular sublanguage in the padded language $v(\mathscr{A}^* \times \mathscr{A}^*)$. Recall that in the product language we pad the shorter of the two words u, v by adding a string of padding symbols \$ at the end of the shorter so that the two words have the same length, giving the set $v(\mathscr{A}^* \times \mathscr{A}^*) = \{(u, v)^{\$} \mid u, v \in L\}$ – see Part I, § 1.

Lemma 8.1 *If L is the language of a biautomatic structure on G, then for any words $x, y \in \mathscr{A}^*$, the language*

$$L(x, y) = \{(u, v)^{\$} \mid u, v \in L, \quad \textit{and} \quad \bar{u} = \overline{xvy}\}$$

is regular.

Proof. We prove by induction on the length of the words x, y, using the fact that the class of regular languages is closed under the operations of intersection, union, direct product and projection onto factors (see for instance [BGSS, Sect. 2][HU, Chapt. 3]).

We first show that the language

$$L(x) = \{(u, v)^{\$} \mid u, v \in L \quad \text{and} \quad \bar{u} = \overline{xv}\}$$

is regular (this is analogous to [BGSS, Prop. 2.7]). By definition this is true for words x of length 1.

Now suppose that the language $L(x)$ is regular for words of length $<n$. Let $x=ax'$ be a word of length n, where $a\in\mathscr{A}$, and $\ell(x')=n-1$. Then $L(x')$ and $L(a)$ are both regular by the induction hypothesis, and so the padded language

$$L(a)\times L(x')=\{(u, v, w, z)\,|\,\bar{u}=\overline{av} \quad \text{and} \quad \bar{w}=\overline{x'z}\}$$

is also regular. Now

$$L(M_0)=\{(v, w)\,|\,v, w\in L \quad \text{and} \quad \bar{v}=\bar{w}\}$$

is the regular language accepted by the equality checker M_0 of the automatic structure on G, so that

$$(\mathscr{A}^*\times L(M_0)\times\mathscr{A}^*)\cap(L(a)\times L(x'))$$

is the intersection of two regular padded languages and so is itself regular. Notice that this language consists of padded quadruples $(u, v, w, z)^\$$ where $\bar{u}=\overline{av}=\bar{a}\bar{w}=\overline{ax'z}$. We now project this language onto the product of the first and last factors, to give the regular language $L(x)$.

Similarly the language

$$L'(y)=\{(u, v)^\$\,|\,u, v\in L \quad \text{and} \quad \bar{u}=\overline{vy}\}$$

is regular, and the language $L(x, y)$ can be obtained by projecting the regular language

$$(L(x)\times L'(y))\cap(L\times L(M_0)\times L)$$

onto the product of the first and last factors. □

We have shown that all the languages that we have constructed in Part I and here have the essential property that geodesic words which begin and end distance at most one apart lie in a k-Hausdorff neighborhood of each other for some k. It follows that as in Part I, § 1, it is possible to build a standard automaton which accepts the languages $L(a)$, by simply altering the automaton given in the proof of 1.1 of Part I by making the start state $(s_0, s_0, \bar{a}^{-1})$, and the accept states to be $Y\times Y\times\{1\}$. Thus the proof of Theorem 1.1 of Part I shows that:

Lemma 8.2 *If G has an automatic structure in which the language L of accepted words consists of geodesics with the property that words which begin and end distance one apart lie in a k-Hausdorff neighborhood, then L is part of a biautomatic structure for G.*

Our constructions thus show that negatively curved groups have biautomatic structures, and that small cancellation groups are free factors of groups with biautomatic structures (and so themselves are biautomatic, following the argument of [BGSS, § 4]).

Theorem 8.3 *A group with a biautomatic structure has solvable conjugacy problem.*

Proof. Given two words $x, y\in\mathscr{A}^*$, then from the above result the language $L(x, y)$ is regular. But if $\bar{x}$ and $\bar{y}$ are conjugate in G, then there is some element

$g \in G$ such that $\bar{x} = g\bar{y}g^{-1}$. This means that there is some word $w \in L$ such that $\overline{xw} = \overline{wy} \Leftrightarrow \bar{w} = x^{-1}wy$, or alternatively, that the regular language

$$L(x^{-1}, y) \cap \{(w, w) | w \in L\}$$

is non-empty. But the problem of deciding whether a given regular language is empty is algorithmically solvable (see [HU, Chapt. 3]). □

Notice that given a small cancellation group G, we obtained languages of geodesics with the required properties which put biautomatic structures on a group $H = G * F$, where F is a finitely generated free group. But as the conjugacy problem is solvable in H, it is also solvable in G.

The conjugacy problem was originally solved for small cancellation groups by Schupp (see [LS, Chapt. V, Sect. 7)] and for negatively curved groups by Gromov [G].

References

[B] Brown, K.: Buildings. Berlin Heidelberg New York: Springer 1988

[BGSS] Baumslag, G., Gersten, S.M., Shapiro, M., Short, H.: Automatic groups and amalgams. J. Pure Appl. Algebra (to appear)

[BT] Bruhat, F., Tits, J.: Groupes réductifs sur un corps local. I. Données radicales valuées. Publ. Math., Inst. Hautes Étud. Sci. **41** 5–251 (1972)

[CEHPT] Cannon, J.W., Epstein, D.B.A., Holt, D.F., Patterson, M. S., Thurston, W.P.: Word processing and group theory. (University of Warwick preprint Jan. 1990)

[G] Gromov, M.: Hyperbolic groups. In: Essays in group theory Gersten, S.M. (ed.) M.S.R.I series, vol. 8, pp. 75–263, Berlin Heidelberg New York: Springer 1987

[GS] Gersten, S.M., Short, H.: Small cancellation theory and automatic groups. Invent. Math. **102**, 305–334 (1990)

[GS2] Gersten, S.M., Short, H.: Rational subgroups of biautomatic groups. (to appear) Ann. Math.

[HU] Hopcroft, J.E., Ullman, J.D.: Introduction to automata theory, languages and computation. Reading, MA: Addison-Wesley 1979

[LS] Lyndon, R.C., Schupp, P.E.: Combinatorial group theory. Berlin Heidelberg New York: Springer 1977

[W] Weinbaum, C.M.: The word and conjugacy problem for the knot group of any prime alternating knot. Proc. Am. Math. Soc. **22**, 22–26 (1971)