Inventiones mathematicae

Volume 72 1983

Springer-Verlag
Berlin · Heidelberg New York · Tokyo

Inventiones mathematicae

Springer-Verlag Berlin · Heidelberg · New York · Tokyo

Printers: Universitätsdruckerei H. Stürtz AG, Würzburg

Index

Covered by Zentralblatt für Mathematik and Current Mathematical Publications

Inventiones mathematicae

Vol. 72 Fasc. 1 1983

Covered by Zentralblatt für Mathematik
and Current Mathematical Publications

Springer-Verlag
Berlin Heidelberg New York Tokyo

222 Inventiones math. ISSN 0020-9910 INVMBH 72(1) 1–158 (1983)
April 1983

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.

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.

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

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
D-4540 Lengerich

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

Invent. math. 72, 1-25 (1983)

Inventiones mathematicae

Index for Subfactors

V.F.R. Jones

Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104, USA

§ 1. Introduction

One of the first things Murray and von Neumann did with their theory of continuous dimension for subspaces affiliated with a type II_1 factor was to define an invariant for its action on a Hilbert space. This invariant has come to be known as the coupling constant and it measures the relative mobility of the factor and its commutant. To be precise, if M is a II_1 factor on $\mathscr{H}$ and M' is its commutant, the coupling constant C_M is infinite if M' is an infinite factor and if M' is finite, one takes any non-zero vector ξ in $\mathscr{H}$ and one considers the closed subspaces $\overline{M\xi}$ and $\overline{M'\xi}$ affiliated with M' and M, respectively. It is a nontrivial result of Murray and von Neumann that the ratio $C_M = \dim_{M'}(\overline{M\xi})/\dim_M(\overline{M'\xi})$ does not depend on ξ. This real number between 0 and ∞ is called the coupling constant.

From a more modern point of view, the coupling constant measures the dimension of a Hilbert space on which M acts and may be defined in terms of intertwining maps in the category of normal M-modules. This leads to the notation $C_M = \dim_M(\mathscr{H})$. This notation is more natural and has been useful in the study of the von Neumann algebra of a foliation where the dimensions of certain geometric Hilbert spaces are measured by a von Neumann algebra (see [8]). We will use this more suggestive notation while keeping the term "coupling constant." Two normal representations of M are unitarily equivalent iff they have the same coupling constant.

It is an easy observation that the coupling constant can be used to define a conjugacy invariant for subfactors of II_1 factors. We call this invariant the index since if the subfactor comes from a subgroup in the group constructions of II_1 factors, the conjugacy invariant is the index of the subgroup. The index is defined in general as $\dim_N(L^2(M, \mathrm{tr}))$ where N is the subfactor and tr is the trace on M. This definition was probably noticed by Murray and von Neumann and appears more or less explicitly in works by Goldman [14], Suzuki

Supported in part by NSF Grant # MCS 79-03041

[26] and others. In fact Goldman proves that if a subfactor has index 2 then the whole factor may be expressed as the crossed product of the subfactor by a $\mathbb{Z}_2$ action. This is analogous to the fact that any subgroup of index 2 of a group is normal. It may be combined with Connes' classification of periodic automorphisms of the hyperfinite II_1 factor R ([6]) to yield the pleasing positive result that there is, up to conjugacy, only one subfactor of index 2 of R.

The question immediately arises: what possible values can the index take? The difficulty in answering this question is that there is very little to play with given an arbitrary subfactor of a II_1 factor. The only general result available is the existence of a conditional expectation onto the subfactor shown by Umegaki in [30]. In fact this tool will allow us to determine completely the possible values of the index for subfactors of R. Experience with dimension in II_1 factors suggests that the index will take on a continuum of values. This is indeed true but surprisingly one cannot turn on this effect until index 4 and even then it involves the fundamental group. It seems entirely plausible that for II_1 factors without full fundamental group, whose existence was shown by Connes in [9], the index may take only countably many values.

What, then, are the possible values of the index for subfactors of R? The answer is $\{4\cos^2\pi/n \mid n=3,4,\ldots\} \cup \{r\in\mathbb{R} \mid r\geqq 4\} \cup \{\infty\}$. The value ∞ and the real values $\geqq 4$ are easily obtained (2.1.19 and 2.2.5). The situation between 1 and 4 is more difficult. The basic idea of the analysis for index <4 is as follows: Let $N\subseteq M$ be II_1 factors. One represents M on $L^2(M,\mathrm{tr})$ and considers the extension e_N to $L^2(M,\mathrm{tr})$ of the conditional expectation onto N. One defines $\langle M, e_N\rangle$ to be the II_1 factor generated by M and e_N on $L^2(M,\mathrm{tr})$. (This construction appears also in [5], [24]).) The crucial observation at this point is that the index of M in $\langle M, e_N\rangle$ is the same as that of N in M. Thus one may iterate this extension process and one obtains a sequence of II_1 factors, each one obtained from the previous one by adding a projection. The inductive limit gives a II_1 factor and if the projections in the construction are numbered $e_1, e_2, \ldots,$ then they satisfy $e_i e_{i\pm 1} e_i = \tau e_i$, $e_i e_j = e_j e_i$ if $|i-j|\geqq 2$ and $\mathrm{tr}(e_{i_1} e_{i_2} \ldots e_{i_n}) = \tau^n$ if $|i_j - i_k| \geqq 2$ for $j \neq k$. Here τ is the reciprocal of the index of N in M and tr denotes the trace on the inductive limit. Analysis of the algebra generated by the e_i's yields that if $\tau > 1/4$ it can only be $\frac{1}{4}\sec^2\pi/n$ for $n=3,4,\ldots$. A large bonus of the analysis is that it shows how to construct subfactors with these values as τ.

It seems strange that the set of values contains a discrete and a continuous part, but this may yet be understood by the fact that, if the index is less than 4, the relative commutant is trivial while the constructions available for the continuous part all have nontrivial relative commutant. We have very little information on what happens to the values of the index if the relative commutant is required to be trivial, but note that a result of Popa in [23], together with Connes' result on injective factors [7] shows that any II_1 factor has a subfactor of infinite index with trivial relative commutant.

Before closing the introduction, I would like to pose 4 problems which I hope will lead to a more profound understanding of subfactors.

Problem 1 (due to Connes). What are the possible values of the index for subfactors of R with trivial relative commutant? Or, what is $\mathscr{C}_R$?

Problem 2. For each $n=3, 4, 5, \ldots$ are there only finitely many subfactors of R up to conjugacy with index $4\cos^2 \pi/n$?

Problem 3. If N is a subfactor of R, is N conjugate to $N \otimes R$ in a decomposition $R \cong R \otimes R$? ($[R:N] < \infty$)

Problem 4. If N is a subfactor of M which is regular and has trivial relative commutant, is M the crossed product of N by a group action? (True if $M=R$ by a result of Ocneanu [22], see [17].)

It would be impossible to thank everyone who helped me with this paper. I have tried to mention individual contributions in the text, but let me also thank especially B. Baker, A. Connes, F. Goodman, R. Powers, M. Takesaki, and A. Wassermann for many fruitful conversations. This paper is an extended version of the Comptes Rendus note [18].

Contents

§ 2. Generalities

§ 2.1. The Global Index

If M is a finite factor acting on a Hilbert space $\mathscr{H}$ with finite commutant M', the coupling constant $\dim_M(\mathscr{H})$ of M is defined as $\mathrm{tr}_M(E_\xi^{M'})/\mathrm{tr}_{M'}(E_\xi^M)$ where ξ is a non-zero vector in $\mathscr{H}$, tr_A denotes the normalized trace and E_ξ^A is the projection onto the closure of the subspace $A\xi$. This definition, due to Murray and von Neumann in [20], is independent of ξ.

We recall some rules of calculation associated with $\dim_M$. (See [10, p. 263].)

$$\dim_M(\mathscr{H})>0 \tag{2.1.1}$$

$$\dim_M(\mathscr{H})=(\dim_{M'}(\mathscr{H}))^{-1} \tag{2.1.2}$$

$$\text{If } e \text{ is a projection in } M',\ \dim_{M_e}(e\mathscr{H})=\mathrm{tr}_{M'}(e)\dim_M(\mathscr{H}) \tag{2.1.3}$$

$$\text{If } E \text{ is a projection in } M,\ \dim_{M_e}(e\mathscr{H})=(\mathrm{tr}_M(e))^{-1}\dim_M(\mathscr{H}) \tag{2.1.4}$$

If $M\otimes 1$ is the amplification of M on $\mathscr{H}\otimes\mathscr{K}$,

$$\dim_M(\mathscr{H}\otimes\mathscr{K})=\dim_{\mathbb{C}}(\mathscr{K})\dim_M(\mathscr{H}) \tag{2.1.5}$$

$\dim_M(\mathscr{H})=1$ iff M is standard on $\mathscr{H}$, i.e. there is a cyclic trace vector for M. (2.1.6)

Agree to put $\dim_M(\mathscr{H})=\infty$ if M' is infinite.

Proposition 2.1.7. *Let M be as above and N be a subfactor. The number $\dim_N(\mathscr{H})/\dim_M(\mathscr{H})$ is independent of $\mathscr{H}$ provided M' is finite.*

Proof. Any two such representations differ by an amplification and an induction. By (2.1.3) and (2.1.5), both $\dim_M$ and $\dim_N$ are multiplied by the same constants in this process. Q.E.D.

Definition. If N is a subfactor of M, the number $\dim_N(\mathscr{H})/\dim_M(\mathscr{H})$ defined in 2.1.7 is called the (global) *index* of N in M and written $[M:N]$. Note that $[M:N]=\infty$ means that N' is infinite for any normal representation of M.

By (2.1.7) and (2.1.6), $[M:N]=\dim_N(L^2(M,\mathrm{tr}))$. Thus $[M:N]$ is a conjugacy invariant for N as a subfactor of M.

The rules of calculation for $\dim_M$ give some rules for $[M:N]$.

Proposition 2.1.8. *If $P\subseteq Q\subseteq M$ are II_1 factors then*

$$[M:M]=1 \tag{2.1.9}$$

$$[M:P]\geqq 1 \tag{2.1.10}$$

$$[M:P]=[M:Q][Q:P] \tag{2.1.11}$$

$$[M:P]\geqq[M:Q] \tag{2.1.12}$$

$$[M:P]=[M:Q] \quad \textit{implies } P=Q \tag{2.1.13}$$

$$[M:P]=[P':M'] \quad \textit{if } P' \textit{ is finite.} \tag{2.1.14}$$

Proof. The only nontrivial property is (2.1.13). To prove it first note that by (2.1.11) and (2.1.9) we may suppose that $M=Q$ and that M acts on $L^2(M,\mathrm{tr})$ with cyclic trace vector ξ. Then $P\xi$ is dense in $\mathscr{H}$ by hypothesis. But then for any $a\in M$ there is a net b_n of elements of P with $b_n\xi\to a\xi$ in $\mathscr{H}$, i.e. $\|b_n-a_n\|_2\to 0$. By [10, Lemma 1, p. 270], this implies $a\in P$. Q.E.D.

We next examine how the index behaves under tensor products.

Proposition 2.1.15. *Let N_1 and N_2 be subfactors of the finite factors M_1 and M_2, respectively. Then $N_1\otimes N_2$ is a subfactor of $M_1\otimes M_2$ and $[M_1\otimes M_2:N_1\otimes N_2]=[M_1:N_1][M_2:N_2]$.*

Proof. Let M_1 and M_2 act with cyclic trace vectors $\xi_1 \in \mathcal{H}_1$ and $\xi_2 \in \mathcal{H}_2$ respectively. Then if e_1 and e_2 are the projections onto $\overline{N_1 \xi_1}$ and $\overline{N_2 \xi_2}$, $e_1 \otimes e_2$ is the projection onto $\overline{N_1 \otimes N_2(\xi_1 \otimes \xi_2)}$. Moreover $M_1 \otimes M_2$ is standard on $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $(N_1 \otimes N_2)' = N_1' \otimes N_2'$. Thus $\mathrm{tr}_{(N_1 \otimes N_2)'}(e_1 \otimes e_2) = \mathrm{tr}_{N_1'}(e_1)\,\mathrm{tr}_{N_2'}(e_2)$ which gives the desired result. Q.E.D.

Proposition 2.1.16. *Let N_i, M_i be as above and suppose $N_i' \cap M_i = \mathbb{C}$, $i=1,2$. Then $(N_1 \otimes N_2)' \cap (M_1 \otimes M_2) = \mathbb{C}$.*

Proof. $(N_1 \otimes N_2)' \cap (M_1 \otimes M_2) = (N_1' \otimes N_2') \cap M_1 \otimes M_2$ (see [28, p. 227]) and $(N_1' \otimes N_2') \cap (M_1 \otimes M_2) = (N_1' \cap M_1) \otimes (N_2' \cap M_2)$. Q.E.D.

We now introduce two isomorphism invariants for II_1 factors.

Definition. If M is a finite factor let

$$\mathcal{I}_M = \{r \in \mathbb{R} \cup \{\infty\} \mid \text{there is a } \mathrm{II}_1 \text{ subfactor } N \text{ of } M \text{ with } [M:N]=r\}$$

$$\mathcal{C}_M = \{r \in \mathbb{R} \cup \{\infty\} \mid \text{there is a } \mathrm{II}_1 \text{ subfactor } N \text{ of } M \text{ with } [M:N]=r \text{ and } N' \cap M = \mathbb{C}\}.$$

The determination of $\mathcal{I}_M$ and $\mathcal{C}_M$ is in general rather difficult. We shall gather some immediate results about them.

Proposition 2.1.17. *If $M \cong M \otimes M$ then both $\mathcal{I}_M$ and $\mathcal{C}_M$ are subsemigroups (with 1) of $\{r \in \mathbb{R} \mid r \geqq 1\} \cup \{\infty\}$ under multiplication.*

Proof. This follows from 2.1.15 and 2.1.16. Q.E.D.

Lemma 2.1.18. *If N is a hyperfinite subfactor of M then $[M:N] < \infty$ implies that M is hyperfinite.*

Proof. If M acts in such a way that N$'$ is finite, then N$'$ is hyperfinite (e.g. [29]) and by [7], M' and so M is hyperfinite. Q.E.D.

Corollary 2.1.19. *For any II_1 factor M, $\infty \in \mathcal{I}_M$.*

Proof. If $M \cong R$ then $R \otimes 1 \subseteq R \otimes R$ is of infinite index. Otherwise by [21] there is a hyperfinite subfactor of M which is of infinite index by 2.1.18. Q.E.D.

Proposition 2.1.20. *For any separable II_1 factor M, $\infty \in \mathcal{C}_M$.*

Proof. A result of Popa in [23] says that we can find a maximal abelian subalgebra A and a unitary u with $uAu^* = A$ such that u and A generate a subfactor N of M, isomorphic to R. Since N contains a maximal abelian subalgebra, $N' \cap M = \mathbb{C}$.

So if M is not hyperfinite, combining this result with 2.1.18 gives the required subfactor. If $M = R$, see [17] or §2.3. Q.E.D.

In this paper we will show that

$$\mathcal{I}_M \cap [1,4) = \mathcal{C}_M \cap [1,4) \subseteq \{4\cos^2 \pi/n \mid n=3,4,\ldots\}$$

and that

$$\mathcal{I}_R = \{4\cos^2 \pi/n \mid n=3,4,\ldots\} \cup \{r \in \mathbb{R} \mid r \geqq 4\} \cup \{\infty\}.$$

§ 2.2. The Local Index

If the global index of a subfactor is finite, one may define a finer invariant, which I call the local index, obtained by restricting the trace on N' to $N' \cap M$. I would like to thank A. Connes for suggesting this approach.

Definition. Let $N \subseteq M$ be II_1 factors and let $p \in N' \cap M$ be a projection. The *index of N at p* will be $[M_p : N_p] = [M:N]_p$.

Lemma 2.2.1. *The index at p and the global index are related by the formula*

$$[M:N]_p = [M:N]\,\mathrm{tr}_M(p)\,\mathrm{tr}_{N'}(p).$$

Proof. If M begins in standard form on $\mathscr{H}$ then by 2.1.4, $\dim_{M_p}(p\mathscr{H}) = \mathrm{tr}_M(p)^{-1}$. Also $\dim_N(\mathscr{H}) = [M:N]$ so by 2.1.3, $\dim_{N_p}(p\mathscr{H}) = [M:N]\,\mathrm{tr}_{N'}(p)$. Thus $[M:N]_p = \dim_{N_p}(p\mathscr{H})/\dim_{M_p}(p\mathscr{H}) = [M:N]\,\mathrm{tr}_M(p)\,\mathrm{tr}_{N'}(p)$. Q.E.D.

Lemma 2.2.2. *If $\{p_i\}$ is a partition of unity in $N' \cap M$ then*

$$[M:N] = \sum_i \mathrm{tr}_M(p_i)^{-1}\,[M:N]_{p_i}.$$

Proof. For each i, $[M:N]\,\mathrm{tr}_{N'}(p_i) = \mathrm{tr}_M(p_i)^{-1}[M:N]_{p_i}$ and summing over i gives the result.

Corollary 2.2.3. *If $[M:N] < \infty$ then $N' \cap M$ is finite dimensional.*

Proof. If $N' \cap M$ were infinite dimensional we could find arbitrarily large partitions of unity and by 2.2.2 $[M:N] = \infty$. Q.E.D.

In fact with a little more care one may obtain the bound $[M:N] \geqq \dim_{\mathbb{C}}(N' \cap M)$.

Corollary 2.2.4. *If $[M:N] < 4$, $N' \cap M = \mathbb{C}$.*

Proof. If $N' \cap M \neq \mathbb{C}$, then it contains two mutually orthogonal non-zero projections and since $[M:N]_{p_i} \geqq 1$, $[M:N] \geqq \mathrm{tr}(p_1)^{-1} + \mathrm{tr}(p_2)^{-1} \geqq 4$. Q.E.D.

Corollary 2.2.5. *If M has fundamental group $= \mathbb{R}$, $\mathscr{I}_M$ contains $\{r \in \mathbb{R} \mid r \geqq 4\}$.*

Proof. We must exhibit for any $r \geqq 4$ a subfactor of index r. Choose $d \in (0,1)$ with $1/d + 1/(1-d) = r$ and choose a projection p with $\mathrm{tr}_M(p) = d$. Then M_p and M_{1-p} are isomorphic so choose some isomorphism $\theta \colon M_p \to M_{1-p}$ and let N be the subfactor $\{x + \theta(x) \mid x \in M_p\}$. Then $N_p = M_p$ and $N_{1-p} = M_{1-p}$ so by 2.2.2, $[M:N] = 1/d + 1/(1-d) = r$. Q.E.D.

§ 2.3. Examples

We give three examples of subfactors and their indices.

Example 2.3.1. Suppose $M = N \otimes P$ where P is a type I_n factor. Then $[M : N \otimes 1] = n^2$. This follows from 2.1.5.

Thus for any II_1 factor M, $\mathscr{I}_M$ always contains $\{n^2 | n \in \mathbb{Z} - \{0\}\} \cup \{\infty\}$. It is not inconceivable that there are II_1 factors for which $\mathscr{I}_M$ is no larger than this set.

Example 2.3.2. Let A be a von Neumann algebra and G a countable discrete group of automorphisms for which the crossed product $A \rtimes G$ is a finite factor. If H is a subgroup of G such that $A \rtimes H$ is a finite factor then $[A \rtimes G : A \rtimes H] = [G:H]$.

Proof. Write G as a disjoint union of cosets, $G = \coprod_{i \in I} H g_i$. Then let $V_i = \overline{(A \rtimes H)\, u_{g_i}}$. The V_i are subspaces of $L^2(A \rtimes G, \mathrm{tr})$ affiliated with $(A \rtimes H)'$ which are mutually orthogonal, where tr is the extension to $A \rtimes G$ of a faithful normal G-invariant trace on A and the u_g's are the implementing unitaries of the crossed product. Moreover if p_i is the projection onto V_i then $\sum_{i \in I} p_i = 1$ and the p_i are mutually equivalent in $(A \rtimes H)'$ since $V_i = J u_{g_i} J V_0$ where $V_0 = \overline{A \rtimes H}$ and J is the involution on $L^2(A \rtimes G, \mathrm{tr})$. This is because $J u_{g_i} J \in (A \rtimes G)' \subseteq (A \rtimes H)'$.

Thus if $[G:H] = \infty$, $(A \rtimes H)'$ is infinite so $[A \rtimes G : A \rtimes H] = \infty$. If $[G:H] < \infty$, $(A \rtimes H)'$ is finite since reduction by p_0 puts $A \rtimes H$ in standard form. So $[G:H]\, \mathrm{tr}_{(A \rtimes H)'}(p_0) = 1$ which establishes that $[A \rtimes G : A \rtimes H] = [G:H]$. Q.E.D.

This example is the justification for the name "index".

Example 2.3.3. If M is a II_1 factor and G is a finite group of outer automorphisms of M with fixed point algebra M^G, $[M : M^G] = |G|$.

Proof. This result could be established using 2.3.2 but I shall give a different proof which brings in the basic construction of Chapter 3.

Let M act on $L^2(M, \mathrm{tr})$ and let u_g be the unitaries extending the action of G on M. Then the u_g's act also on M' and it is established in [1] that $(M^G)'$ is isomorphic in the obvious way to $M' \rtimes G$. The projection onto $\overline{M^G}$ is $|G|^{-1} \sum_{g \in G} u_g$ and by the isomorphism with the crossed product its trace is $|G|^{-1}$. Thus $[M : M^G] = |G|$. Q.E.D.

§ 3. The Basic Construction

§ 3.1. *Extending Finite von Neumann Algebras by Subalgebras*

Let M be a finite von Neumann algebra with faithful normal normalized trace tr and let N be a von Neumann subalgebra. By [30] there is a conditional expectation $E_N \colon M \to N$ defined by the relation $\mathrm{tr}(E_N(x) y) = \mathrm{tr}(x y)$ for $x \in M$, $y \in N$. The map E_N is normal and has the following properties:

$$E_N(axb) = a E_N(x) b \quad \text{for } x \in M,\ a, b \in N \text{ (the bimodule property)} \quad (3.1.1)$$

$$E_N(x^*) = E_N(x)^* \quad \text{for all } x \in M \tag{3.1.2}$$

$$E_N(x^*) E_N(x) \leqq E_N(x^* x) \quad \text{and} \quad E_N(x^* x) = 0 \quad \text{implies } x = 0. \tag{3.1.3}$$

Let ξ be the canonical cyclic trace vector in $L^2(M, \mathrm{tr})$. Identify M with the algebra of left multiplication operators on $L^2(M, \mathrm{tr})$. The conditional expectation E_N extends to a projection e_N on $\mathscr{H}$ via $e_N(x\xi) = E_N(x)\xi$. Let J be the involution $x\xi \mapsto x^*\xi$.

Proposition 3.1.4.
(i) *For* $x \in M$, $e_N x e_N = E_N(x) e_N$.
(ii) *If* $x \in M$ *then* $x \in N$ *iff* $e_N x = x e_N$.
(iii) $N' = \{M' \cup \{e_N\}\}''$.
(iv) J *commutes with* e_N.

Proof. (i) If y is an arbitrary element of M then $e_N x e_N(y\xi) = e_N x(E_N(y)\xi) = E_N(x) E_N(y)\xi$ by 3.1.1, and $E_N(x) e_N(y\xi) = E_N(x) E_N(y)\xi$. But the vectors $y\xi$ are dense in $L^2(M, \mathrm{tr})$.

(ii) Relation 3.1.1 shows that e_N commutes with N as in (i). Moreover if $x \in M$ and $e_N x = x e_N$ then $(e_N x)\xi = E_N(x)\xi = (xe)\xi = x\xi$. Since ξ is separating, $x = E_N(x)$.

(iii) It suffices to show that $\{M' \cup \{e_N\}\}' = N$. This follows from (i).

(iv) This follows from 3.1.2. Q.E.D.

These calculations lead to the following.

Definition. Let $\langle M, e_N \rangle$ be the von Neumann algebra on $L^2(M, \mathrm{tr})$ generated by M and e_N. This is the basic construction.

Proposition 3.1.5. (i) $\langle M, e_N \rangle = J N' J$.

(ii) *Operators of the form* $a_0 + \sum_{i=1}^{n} a_i e_N b_i$ *with* a_i, $b_i \in M$, *give a dense* $*$*-subalgebra of* $\langle M, e_N \rangle$.

(iii) $x \mapsto x e_N$ *is an isomorphism of* N *onto* $e_N \langle M, e_N \rangle e_N$.

(iv) *The central support of* e_N *in* $\langle M, e_N \rangle$ *is* 1.

(v) $\langle M, e_N \rangle$ *is a factor iff* N *is.*

(vi) $\langle M, e_N \rangle$ *is finite iff* N' *is.*

Proof. (i) and (ii) follow immediately from 3.1.4. For (iii), to show that $e_N \langle M, e_N \rangle e_N \subseteq N e_N$ it suffices by (ii) to show that $e_N(a e_N b) e_N \in N e_N$. This follows from (i) of 3.1.4. Moreover if $x e_N = 0$ then $x e_N \xi = x\xi = 0$ and ξ is separating. Thus $x \mapsto x e_N$ is injective. Affirmations (iv), (v) and (vi) are now easy. Q.E.D.

We want to consider special traces on $\langle M, e_N \rangle$.

Definition. If P is a subalgebra of $\langle M, e_N \rangle$, a trace Tr on $\langle M, e_N \rangle$ is called a (τ, P) trace if Tr extends tr and $\mathrm{Tr}(e_N x) = \tau\, \mathrm{tr}(x)$ for $x \in P$.

Lemma 3.1.6. *A* (τ, N) *trace is a* (τ, M) *trace.*

Proof. If $x \in M$, $\mathrm{Tr}(xe_N) = \mathrm{Tr}(e_N x e_N) = \mathrm{Tr}(E_N(x)e_N) = \tau\,\mathrm{tr}(E_N(x)) = \tau\,\mathrm{tr}(x)$. Q.E.D.

We shall now concentrate on the case where M and N are factors.

Proposition 3.1.7. *If M and N are factors then $[M:N] < \infty$ iff $\langle M, e_N\rangle$ is finite and in this case the canonical trace* Tr *on $\langle M, e_N\rangle$ is a (τ, M) trace where $\tau = [M:N]^{-1}$. In particular* $\mathrm{Tr}(e_N) = [M:N]^{-1}$. *Also* $[\langle M, e_N\rangle : N] = [M:N]$.

Proof. By 3.1.6 it suffices to show that Tr is a (τ, N) trace. But consider the map $y \to \mathrm{Tr}(e_N y)$ defined on N. This is a trace by (ii) of 3.1.4 so since N is a factor there is a constant K such that $\mathrm{Tr}(e_N y) = K\,\mathrm{tr}(y)$. Moreover the trace of e_N in N' is by definition τ so by (iv) of 3.1.4, (i) of 3.1.5 and uniqueness of the trace, this is the same as $\mathrm{Tr}(e_N)$. Thus $K = \tau$, and Tr is a (τ, N) trace.

To prove this last assertion note that

$$\begin{aligned}[\langle M, e_N\rangle : M] &= \dim_{M'}(L^2(M,\mathrm{tr}))/\dim_{N'}(L^2(M,\mathrm{tr})) = [\dim_{N'}(L^2(M,\mathrm{tr}))]^{-1} \\ &= \dim_N(L^2(M,\mathrm{tr})) = [M:N]. \quad \text{Q.E.D.}\end{aligned}$$

For some general results about projections onto finite subalgebras see Skau's paper [24].

Finally in this section I show that the basic construction is generic for subfactors of finite index in the sense that all such subfactors are of the form $M \subseteq \langle M, e_N\rangle$, although not canonically.

Lemma 3.1.8. *Let N be a II_1 factor acting on $L^2(N,\mathrm{tr})$ and let M be a II_1 factor containing N. Then $[M:N]$ is finite and there is a subfactor P of N such that $M = \langle N, e_P\rangle$.*

Proof. Since M' is a II_1 factor, $[M:N] < \infty$. Let $P = JM'J$. Then $[N:P] = [M:P]$ as in 3.1.7 and $\langle N, e_P\rangle$ is a subfactor of M with $[M:\langle N, e_P\rangle] = 1$. Thus by 2.1.13, $M = \langle N, e_P\rangle$. Q.E.D.

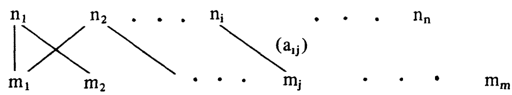
Corollary 3.1.9. *Let $N \subseteq M$ be II_1 factors with $[M:N] < \infty$ then there is a subfactor P of N and an isomorphism $\theta\colon M \to \langle N, e_P\rangle$ with $\theta|_N = \mathrm{id}$.*

Proof. Represent M on $L^2(M,\mathrm{tr})$. Choose a projection $p \in M'$ with $\mathrm{tr}(p) = [M:N]^{-1}$. Then by [20], on $pL^2(M,\mathrm{tr})$, M and N act with N in standard form. By 3.1.8 we are through. Q.E.D.

§ 3.2. Inclusions of Complex Semisimple Algebras

Let $N \subseteq M$ be finite dimensional complex semisimple algebras and $N = \bigoplus_{i=1}^{n} N_i$, $M = \bigoplus_{j=1}^{m} M_j$ be their canonical decompositions as direct sums of simple algebras, $N_i \cong M_{n_i}(C)$, $M_j \cong M_{m_j}(C)$. The inclusion of N in M is specified up to conjugacy by an $n \times m$ matrix $A_N^M = (a_{ij})$ where (a_{ij}) is the number of simple components of a simple M_j module viewed as an N_i module. This may be zero or any positive integer. If $\{p_i\}$ are the central idempotents for the N_i and q_j those for the M_j, $a_{ij} = 0$ iff $p_i q_j = 0$. We will call the matrix A_N^M the inclusion matrix.

The inclusion can also be described diagramatically as follows:

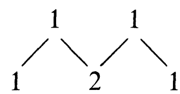


Here there are a_{ij} lines between n_i and m_j. This diagram will be called the Bratteli diagram after [4].

If the identity of M is the same as that of N we have the obvious relation $m_j = \sum_{i=1}^{n} a_{ij} n_i$ which we shall write as

$$\vec{m} = \vec{n} A_N^M. \tag{3.2.1}$$

A concrete example is



which is the diagram for the inclusion of $\mathbb{C}S_2$ in $\mathbb{C}S_3$.

If $N \subseteq M \subseteq P$, note the formula

$$A_N^P = A_N^M A_M^P. \tag{3.2.2}$$

If V is a faithful M-module, the centres of M and M'' are identical so that (if 3.2.1 holds) in the decompositions of M' and N' as simple algebras we may write $N' = \bigoplus_{i=1}^{n} N_i'$ and $M' = \bigoplus_{j=1}^{m} M_j'$. The following formula is in [3, §5, ex. 17].

$$A_{M'}^{N'} = (A_N^M)^T. \tag{3.2.3}$$

Since there is only one normalized trace on $M_n(\mathbb{C})$, a trace on $M = \bigoplus_j M_j$ may be specified by a column vector $\vec{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{pmatrix}$ where t_j is the trace of a minimal idempotent in M_j. The trace of the identity is the product $\vec{m} \cdot \vec{t}$, where $\vec{m} = (m_1, m_2, \ldots, m_m)$.

If $\vec{t}$ defines a trace on M whose restriction to N is defined by the vector $\vec{s}$ then the following relation is immediate.

$$\vec{s} = A_N^M \vec{t}. \tag{3.2.4}$$

Conversely if $\vec{s}$ and $\vec{t}$ define traces on N and M, respectively, then they agree on N if 3.2.4 holds.

§ 3.3. Finite Dimensional C^*-Algebras

A finite dimensional C^*-algebra is of course semisimple so the discussion and notation of §3.2 applies. The presence of the $*$-operation allows us to perform the basic construction of §3.1. The main question to be answered in this section is: given a faithful trace tr on the finite dimensional C^*-algebra M, when does there exist a (faithful) (τ, M) trace on $\langle M, e_N\rangle$? All traces in this section will be positive, i.e. $\vec{t}$ is a positive vector in the natural ordering of $\mathbb{R}^n$.

We begin with a lemma giving more information on the basic construction in finite dimensions.

Lemma 3.3.1. *Let $N \subseteq M$ be finite dimensional C^*-algebras and let* tr *be a faithful positive normalized trace on M. Let e_N and $\langle M, e_N\rangle$ be as in §3.1. Suppose $\{p_i \mid i=1,2,\dots,n\}$ are the minimal central projections of N. Then*

(i) *Jp_iJ are the minimal central projections of $\langle M, e_N\rangle$*

(ii) *$A_M^{\langle M, e_N\rangle} = (A_N^M)^T$ (with the obvious identification of the indices, $p_i \leftrightarrow Jp_iJ$)*

(iii) *$e_N Jp_iJ = e_N p_i$*

(iv) *$x \to e_N x Jp_iJ$ is an isomorphism from p_iN onto $(e_N Jp_iJ)\langle M, e_N\rangle(e_N Jp_iJ)$.*

Proof. (i) The p_i are the minimal central projections of N' and $\langle M, e_N\rangle = JN'J$.

(ii) This follows from 3.2.3 and (i).

(iii) If $x \in M$, $(e_N Jp_iJ)(x\xi) = e_N(xp_i\xi) = E_N(x)p_i\xi$, and

$$(e_N p_i)(x\xi) = e_N(p_i x\xi) = p_i E_N(x)\xi = E_N(x)p_i\xi.$$

(iv) Injectivity follows from (iii), and (iii) of 3.1.5. If $y \in p_i e_N\langle M, e_N\rangle p_i e_N$ then $y = e_N z$ for $z \in N$ and $p_i z = z$ by (iii) of 3.1.5 so $y = e_N p_i z$. Q.E.D.

Theorem 3.3.2. *Let M, N and* tr *be as above and let* tr *be given on M by the vector $\vec{t}$ and on N by the vector $\vec{s}$. Then there is a (τ, M) trace* Tr *on $\langle M, e_N\rangle$ iff*

(i) *$A^T A\vec{t} = (1/\tau)\vec{t}$*

(ii) *$AA^T\vec{s} = (1/\tau)\vec{s}$*

where $A = A_N^M$.

Proof. ($\Rightarrow$) Let Tr be given on $\langle M, e_N\rangle$ by $\vec{r}$, r_i being the trace of a minimal projection in $Jp_iJ\langle M, e_N\rangle$. By (iv) of 3.3.1, such a minimal projection may be chosen of the form $e_N q$, q being a minimal projection in p_iN. Since Tr is a (τ, M) trace, $r_i = \tau s_i$ so $\vec{r} = \tau\vec{s}$. But by (ii) of 3.3.1, 3.2.2 and 3.2.4, $\vec{s} = AA^T\vec{r} = \tau AA^T\vec{s}$, i.e. $AA^T\vec{s} = (1/\tau)\vec{s}$. Also $\vec{t} = A^T\vec{r} = \tau A^T AA^T\vec{r} = \tau A^T A\vec{t}$.

($\Leftarrow$) Define a trace Tr on $\langle M, e_N\rangle$ by the vector $\vec{r} = \tau\vec{s}$. It is a faithful positive trace since tr is. By 3.1.6 it suffices to show that it extends tr and that it is a (τ, N) trace. For the former, by 3.2.4 we need $A^T\vec{r} = \vec{t}$. But $A\vec{t} = \vec{s}$ and by (i), $A^T\vec{s} = (1/\tau)\vec{t}$ so $A^T\vec{r} = \vec{t}$. For the latter let q be a minimal projection in p_iN. Then $e_N q$ is a minimal projection in $Jp_iJ\langle M, e_N\rangle$ and by definition $\mathrm{Tr}(e_N q) = \tau\,\mathrm{tr}(q)$. The map $x \to \mathrm{Tr}(e_N x)$ is a trace on p_iN so by uniqueness and linearity, $\mathrm{Tr}(e_N x) = \tau\,\mathrm{tr}(x)$ for all $x \in N$. Q.E.D.

§3.4. Two Extensions; Goldman's Theorem

Let N be a proper von Neumann subalgebra of the finite von Neumann algebra M with faithful normal normalized trace tr. Suppose there is a faithful normal (τ, M) trace Tr on $\langle M, e_N\rangle$. Then we may form the extension $\langle\langle M, e_N\rangle, e_M\rangle$.

Proposition 3.4.1.

(i) $e_M e_N e_M = \tau e_M$

(ii) $e_N e_M e_N = \tau e_N$

(iii) $e_M \wedge e_N = e_M \wedge e_N^{\perp} = e_M^{\perp} \wedge e_N = 0$.

Proof. (i) By (i) of 3.1.4, $e_M e_N e_M = E_M(e_N) e_M$ and since Tr is a (τ, M) trace, $E_M(e_N) = \tau$.

(ii) By (ii) of 3.1.5 it suffices to verify the relation on vectors of the form $a\xi$ and $a e_N b \xi$ where $a, b \in M$. But

$$e_N e_M e_N(a\xi) = e_N(E_M(e_N a)\xi) = \tau e_N(a\xi),$$

and

$$e_N e_M e_N(a e_N b\xi) = e_N e_M(E_N(a) e_N b\xi) = e_N(\tau E_N(a) b\xi)$$

while $\tau e_N(a e_N b\xi) = \tau e_N E_N(\tau) b\xi$.

(iii) $e_N \wedge e_M = s - \lim\limits_{n\to\infty} (e_N e_M e_N)^n = 0$ since $\tau < 1$.

The other relations follow from $(1-\tau) < 1$. Q.E.D.

Now suppose there is a faithful $(\tau, \langle M, e_N\rangle)$ trace Tr on $\langle\langle M, e_N\rangle, e_M\rangle$.

Corollary 3.4.2. *If $\tau \neq 1/2$, the von Neumann algebra generated by e_N and e_M is isomorphic to $M_2(\mathbb{C}) \oplus \mathbb{C}$. If $\tau = 1/2$ it is isomorphic to $M_2(\mathbb{C})$ and we have the relation $e_M + e_N - e_M e_N - e_N e_M = 1/2$.*

Proof. The relations of 3.4.1 show immediately that the von Neumann algebra generated by e_N and e_M has dimension at most 5 and is not abelian. But $\mathrm{Tr}(e_M \vee e_N) = 2\tau - \mathrm{Tr}(e_M \wedge e_N) = 2\tau$. This is enough to prove the affirmations about the structure of the algebra. An easy calculation shows that $p = (1-\tau)^{-1}(e_M + e_N - e_M e_N - e_N e_M)$ is a projection with $p e_N = e_N$ and $p e_M = e_M$. Thus $p = e_1 \vee e_2$ and if $\tau = 1/2$, $p = 1$. Q.E.D.

Corollary 3.4.3. (Goldman's theorem [14]). *Let N be a subfactor of the* II_1 *factor M with $[M:N] = 2$. Then M decomposes as the crossed product of N by an outer action of $\mathbb{Z}_2$.*

Proof. By 3.1.9 we know that M is of the form $\langle N, e_P\rangle$ for a subfactor of index 2 of N. Thus N is generated by N and $u = 2e_P - 1$. Moreover since tr is a $(1/2, N)$ trace on M, $\mathrm{tr}(ux) = 0$ for $x \in N$, and the relationship of 3.4.2 implies that, if $v = 2e_N - 1$, $uv = -vu$. Thus for $x \in N$, $vuxu^* = uxu^*v$. So uxu^* commutes with e_N and by (ii) of 3.1.4, $uxu^* \in N$. Thus by [25], M is the crossed product of N by a $\mathbb{Z}_2$ action which is necessarily outer since M is a factor. Q.E.D.

§4. Possible Values of the Index

§4.1. Certain Algebras Generated by Projections

Chapter 4 will be largely devoted to proving the following result.

Theorem 4.1.1. *Let M be a von Neumann algebra with faithful normal normalized trace* tr. *Let $\{e_i \mid i = 1, 2, \ldots\}$ be projections in M satisfying*

a) $e_i e_{i\pm 1} e_i = \tau e_i$ *for some* $\tau \leqq 1$

b) $e_i e_j = e_j e_i$ *for* $|i-j| \geqq 2$

c) $\mathrm{tr}(w e_i) = \tau\, \mathrm{tr}(w)$ *if* w *is a word on* $1, e_1, e_2, \ldots, e_{i-1}$

Then if P *denotes the von Neumann algebra generated by the* e_i's,

(i) $P \cong R$ (*the hyperfinite* II_1 *factor*)

(ii) $P_\tau = \{e_2, e_3, \ldots\}''$ *is a subfactor of* P *with* $[P:P_\tau] = \tau^{-1}$.

(iii) $\tau \leqq 1/4$ *or* $\tau = \frac{1}{4}\sec^2 \pi/n$, $n = 3, 4, \ldots$.

In this section we will prove (i) and (ii). We begin the proof with some notation.

Definition. Let $A_{m,n}$ be the $*$-algebra generated by $1, e_m, e_{m+1}, \ldots, e_n$ for $1 \leqq m \leqq n \leqq \infty$. Let A_n be $A_{1,n}$, $A_0 = \mathbb{C}$. Thus $A_\infty = A_{1,\infty}$ is the $*$-algebra generated by 1 and the e_i's.

Next some combinatorial results. If w is an (associative) word on the e_i's, call it *reduced* if it is of minimal length for the grammatical rules $e_i e_{i\pm 1} e_i \leftrightarrow e_i$, $e_i e_j \leftrightarrow e_j e_i$ for $|i-j| \geqq 2$, $e_i^2 \leftrightarrow e_i$.

Lemma 4.1.2. *Let* $e_{i_1} e_{i_2} \ldots e_{i_k}$ *be a reduced word. Then if* $m = \max\{i_1, i_2, \ldots, i_k\}$, m *occurs only once in the list* $i_1, i_2, \ldots, i_k$.

Proof. By induction on the length of a reduced word. It is trivial for words of length $\leqq 1$. Suppose true for words of length $\leqq n$ and let w be a reduced word of length $n+1$. Suppose $w = w_1 e_m w_2 e_m w_3$ where m is the maximum index and w_2 does not contain e_m. Then there are 2 possibilities.

a) w_2 does not contain e_{m-1}. In this case e_m commutes with all the e_i's in w_2 so the length of w may be shortened using $e_m^2 \to e_m$.

b) w_2 contains e_{m-1}. Then since w is reduced, so is w_2 and by induction $w_2 = v_1 e_{m-1} v_2$ where v_1 and v_2 are words on $e_1, e_2, \ldots, e_{m-2}$. But then e_m commutes with v_1 and v_2 so that the length of w may be reduced using $e_m e_{m-1} e_m \to e_m$. Q.E.D.

It is clear that in the algebra A, any word on the e_i's is proportional to a reduced word.

Corollary 4.1.3. (i) A_n *is finite dimensional.*

(ii) *For* $x \in A_n$, $e_{n+1} x e_{n+1} = E_{A_{n-1}}(x) e_{n+1}$ *(here* $E_{A_{n-1}}\colon A_n \to A_{n-1}$ *is with respect to the restriction of* tr).

(iii) $x \mapsto x e_{n+1}$ *is an isomorphism of* A_{n-1} *onto* $e_{n+1} A_{n+1} e_{n+1}$.

Proof. (i) If there are only finitely many reduced words, A_n is finite dimensional. But this follows immediately by induction from 4.1.2.

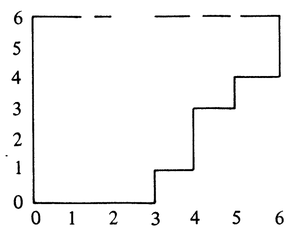
(ii) First of all $\mathrm{tr}(x e_{n+1}) = \tau\, \mathrm{tr}(x)$ for $x \in A_n$ follows immediately from 4.1.1(c) so that $E_{A_{n-1}}(e_n) = \tau$. By 4.1.2 and linearity it suffices to consider x of the form $w e_n w'$ with w and w' in A_{n-1}. But then $e_{n+1} x e_{n+1} = \tau w w' e_{n+1}$ by 4.1.1(a) and (b), and $E_{A_{n-1}}(x) e_{n+1} = \tau w w' e_{n+1}$ by the bimodule property of $E_{A_{n-1}}$.

(iii) If w is a reduced word on $e_1, e_2, \ldots, e_{n+1}$ write $w = x e_{n+1} y$ with $x, y \in A_n$. Then by (ii), $e_{n+1} w e_{n+1} = t e_{n+1}$ for $t \in A_{n-1}$. Thus $e_{n+1} A_{n+1} e_{n+1} \subseteq A_{n-1} e_{n+1}$. To show that the map is an isomorphism, suppose $x e_{n+1} = 0$, $x \in A_{n-1}$. Then $x x^* e_{n+1} = 0$ so $\mathrm{tr}(x x^*) = 0$, i.e. $x = 0$. Q.E.D.

Aside 4.1.4. In fact it is possible to uniquely order reduced words by pushing $e_{\max}$ to the right as far as possible. It is easy to show that such an ordered reduced word is of the form

$$(e_{j_1} e_{j_1-1} \dots e_{k_1})(e_{j_2} e_{j_2-1} \dots e_{k_2}) \dots (e_{j_p} e_{j_p-1} \dots e_{k_p})$$

where j_p is the maximum index, $j_i \geqq k_i$ and $j_{i+1} > j_i$, $k_{i+1} > k_i$. To each such word we may associate an increasing path on the integer lattice between $(0,0)$ and $(n+1, n+1)$, which does not cross the diagonal. For instance $(e_3 e_2 e_1)(e_4 e_3)(e_5 e_4)$ in A_5 would correspond to the path (I owe this observation to H. Wilf):



It is well known that such paths are counted by the Catalan numbers $1/(n+2)\binom{2(n+1)}{n+1}$ so we obtain $\dim A_n \leqq 1/(n+2)\binom{2(n+1)}{n+1}$. Uniqueness and linear independence of ordered reduced words would follow from $\dim A_n = 1/(n+2)\binom{2(n+1)}{n+1}$ which we will prove in §5.1 for $\tau \leqq 1/4$. See Aside 5.1.1.

We now want to study traces on A_∞. For this define a *totally reduced word* to be a reduced word on the e_i's where we also allow cyclic permutations. The following result is obvious.

Remark 4.1.5. Any trace on A_n is determined by its effect on totally reduced words.

Lemma 4.1.6. Any totally reduced word is of the form $w = e_{i_1} e_{i_2} \dots e_{i_n}$ with $|i_j - i_k| \geqq 2$, $j \neq k$, and $\operatorname{tr}(w) = \tau^n$.

Proof. The last assertion is immediate from 4.1.1(c). We prove the first assertion by induction on the length of a totally reduced word. It suffices to prove that if $m = \max\{i_1, i_2, \dots, i_n\}$ in a totally reduced word $e_{i_1} e_{i_2} \dots e_{i_n}$, then e_{m-1} does not occur. For this note that by a cyclic permutation we may suppose the word is of the form $e_m w$ and that e_{m-1} occurs at most once in w (since a totally reduced word is reduced). But then we may proceed to $e_m w e_m$ using $e_m^2 = e_m$ and then eliminate e_{m-1} from w using (a) and (b) of 4.1.1. Q.E.D.

We shall now show that any normal normalized trace on P is equal to tr. For this it suffices to show that it equals tr on completely reduced words. For each subset $I \subseteq N$ with $|i-j| \geqq 2$ whenever $i, j \in I$, $i \neq j$, define A_I to be the algebra generated by $\{e_i \mid i \in I\}$.

Lemma 4.1.7. *For any finite permutation σ of I, there is a unitary $u \in A_\infty$ such that $u e_k u^* = e_{\sigma(k)}$ for all $k \in I$.*

Proof. It suffices to show that any transposition $e_i \leftrightarrow e_j$ $(j \geqq i)$ can be effected by a unitary. We may even suppose that no k strictly between i and j is in I. If we can find a unitary $u \in A_{i,j}$ with $u e_i u^* = e_j$ and $u e_j u^* = e_i$ then this u will do since it commutes with all the other e_k's, $k \in I$. But for this it suffices to show that e_i and e_j are equivalent in $A_{i,j}$. And $e_i e_{i+1} \dots e_j$ is a multiple of a partial isometry v with $v v^* = e_i$, $v^* v = e_j$. Q.E.D.

Corollary 4.1.8. *Any normal normalized trace on P is equal to* tr *on A_I''.*

Proof. Since all the e_i's are independent for tr, A_I'' may be identified with an infinite tensor product Bernoulli shift algebra. By 4.1.7 the normalizer induces the obvious action of S_∞ on A_I''. This action is well known to be ergodic. Hence any invariant measure which is absolutely continuous with respect to tr is proportional to tr. Q.E.D.

Corollary 4.1.9. *P is a II_1 factor isomorphic to R.*

Proof. By 4.1.4, 4.1.5 and 4.1.8 there is only one normal normalized trace on P. Thus P is a factor. By [21] and (a) of 4.1.3, $P \cong R$. Q.E.D.

Note that our proof that P is a factor follows similar lines to the scheme laid out in [27].

Corollary 4.1.10. *P_τ is a subfactor of P.*

Proof. Writing $f_i = e_{i+1}$, the f_i satisfy the same relations as the e_i. Q.E.D.

Lemma 4.1.11. *For each n, the map $e_i \to e_{n-i}$ extends to a* tr*-preserving $*$-automorphism σ_n of A_n.*

Proof. The map obviously extends to a $*$-automorphism of the free involutive monoid on the self-adjoint e_i. It thus suffices to know that if w is a word on e_1, $e_2, \dots, e_n$ then $\mathrm{tr}(\sigma_n(w)) = \mathrm{tr}(w)$. For then if $x = \sum_i c_i w_i$ for $c_i \in \mathbb{C}$,

$$\mathrm{tr}(x x^*) = \mathrm{tr}(\sum_{i,j} c_i \bar{c}_j w_i w_j^*) = \mathrm{tr}(\sum_{i,j} c_i \bar{c}_j \sigma(w_i) \sigma(w_j^*))$$

so that $\sum_i c_i w_i \to \sum_i c_i \sigma(w_i)$ is a well defined isometry for the definite hermitian scalar product defined by tr. But the trace of w is determined by 4.1.1(a) and (b) and the formula of 4.1.6, all of which are invariant under the interchange $e_i \leftrightarrow e_{n-i}$. Q.E.D.

Corollary 4.1.12.

(i) $E_{A_{2,n}}(e_1) = \tau$

(ii) *for* $x \in A_{2,n}$, $e_1 x e_1 = E_{A_{3,n}}(x) e_1$

(iii) $E_{P_\tau}(e_1) = \tau$, $e_1 x e_1 = E_{A_{3,\infty}''}(x) e_1$ *for* $x \in P_\tau$.

Proof. (i) and (ii) follow from σ_n applied to (ii) of 4.1.3, and (iii) is just the limit as $n \to \infty$ of (i) and (ii). Q.E.D.

Proof of (ii) of 4.1.1 (calculation of $[P:P_\tau]$). Do the basic construction to obtain $\langle P, e_{P_\tau}\rangle$. Then $e_{P_\tau} e_1 e_{P_\tau} = \tau e_{P_\tau}$ follows from (i) of 4.1.12. We further claim that $e_1 e_{P_\tau} e_1 = \tau e_1$. By (iii) of 4.1.12, elements of the form $a_0 + \sum_{i=1}^{n} a_i e_1 b_i$ with a_i, $b_i \in P_\tau$ are dense in P so it suffices to verify $e_1 e_{P_\tau} e_1 = \tau e_1$ on $x\xi$ and $x e_1 y \xi$ with $x, y \in P_\tau$. But $e_1 e_{P_\tau} e_1(x\xi) = \tau e_1 x \xi = (\tau e_1)(x\xi)$, and

$$e_1 e_{P_\tau} e_1 (x e_1 y \xi) = e_1(E_{P_\tau}(e_1 E_{A'_{3,\infty}}(x) y)\xi) = \tau e_1 E_{A'_{3,\infty}}(x) y \xi = \tau e_1 (x e_1 y \xi).$$

These two relations imply that e_1 and e_{P_τ} are equivalent in $\langle P, e_{P_\tau}\rangle$. Now by (iii) of 3.1.5, e_{P_τ} is a finite projection and e_1 is in a II_1 factor so that $\langle P, e_{P_\tau}\rangle$ is necessarily a finite factor, i.e. $[P:P_\tau] < \infty$. But we know that $\operatorname{tr}(e_1) = \tau$ so $[P:P_\tau] = \operatorname{tr}(e_{P_\tau})^{-1} = \tau^{-1}$. Q.E.D.

§ 4.2. *Restrictions on* τ

We shall now prove (iii) of 4.1.1. We keep the notation of §4.1. Also define $s_n = e_1 \vee e_2 \vee \ldots \vee e_n$.

Lemma 4.2.1. *If* $1 - s_n \neq 0$ *then it is a minimal projection in* A_n *which belongs to* $Z(A_n)$.

Proof. If w is a word on $e_1, e_2, \ldots, e_n$ then

$$(1 - e_1 \vee e_2 \vee \ldots \vee e_n) w = 0. \quad \text{Q.E.D.}$$

Lemma 4.2.2. *If* $s_n \neq 1$ *then* $e_{n+1} \wedge s_n = e_{n+1} s_{n-1}$.

Proof. By 4.1.3(ii), $e_{n+1} s_n e_{n+1} = E_{A_{n-1}}(s_n) e_{n+1}$. But by 4.2.1 and the bimodule property of $E_{A_{n-1}}$, $E_{A_{n-1}}(s_n) \in Z(A_{n-1})$. Let $p_0, p_1, \ldots, p_k$ be the minimal projections in $Z(A_{n-1})$ with $p_0 = 1 - s_{n-1}$, and write $E_{A_{n-1}}(s_n) = \sum_{i=0}^{k} \lambda_i p_i$. Then since $e_{n+1} \wedge s_n \geqq e_{n+1} s_{n-1}$ and $\lim_{m\to\infty} (e_{n+1} s_n e_{n+1})^m = e_{n+1} \wedge s_n$, and by (iii) of 4.1.3, $\lambda_1 = \lambda_2 = \ldots = \lambda_k = 1$. Since $s_n \neq 1$, $E_{A_n}(s_n) \neq 1$ so by (iii) of 4.1.3, $\lambda_k < 1$. Thus $e_{n+1} \wedge s_n = e_{n+1} s_{n-1}$. Q.E.D.

Corollary 4.2.3. *If* $\operatorname{tr}(s_k) > 0$, $\operatorname{tr}(1 - s_{k+1}) = \operatorname{tr}(1 - s_k) - \tau \operatorname{tr}(1 - s_{k-1})$. *And* $\operatorname{tr}(1 - s_1) = 1 - \tau$, $\operatorname{tr}(1 - s_2) = 1 - 2\tau$.

Proof. The trace tr satisfies $\operatorname{tr}(p \vee q) = \operatorname{tr}(p) + \operatorname{tr}(q) - \operatorname{tr}(p \wedge q)$. So by 4.2.2 and (c) of 4.1.1, $\operatorname{tr}(1 - s_{k+1}) = \operatorname{tr}(1 - s_k) - \tau \operatorname{tr}(1 - s_{k-1})$. Q.E.D.

For this reason we define the polynomials $P_n(x)$ by $P_0(x) = 0$, $P_1(x) = 1$ and $P_{n+1} = P_n - x P_{n-1}$, so that if $P_{k+2}(\tau) > 0\ \forall k \leqq n$ then $P_{k+2}(\tau) = \operatorname{tr}(1 - s_k)$.

Lemma 4.2.4. *Let* $\sigma = \dfrac{1 + \sqrt{1 - 4x}}{2}$, $\tilde{\sigma} = \dfrac{1 - \sqrt{1 - 4x}}{2}$. *Then*

(i) $P_n(x) = \dfrac{\sigma^n - \tilde{\sigma}^n}{\sigma - \tilde{\sigma}}$

(ii) $P_n(\frac{1}{4}\sec^2\theta)=\sin n\theta/(2^{n-1}\cos^{n-1}\theta\sin\theta)$

(iii) $\deg P_n=\left[\frac{n-1}{2}\right]$.

Proof. (i) The general solution to the difference equation is $P_n=A\sigma^n+B\tilde{\sigma}^n$. The initial conditions give $A+B=0$, $A(\sigma-\tilde{\sigma})=1$.

(ii) Putting $\sigma=re^{i\theta}$, $\tilde{\sigma}=re^{-i\theta}$, $x=\frac{1}{4}\sec^2\theta$, $r=\frac{1}{2}\sec\theta$, $\sigma^n-\tilde{\sigma}^n=2ir^n\sin n\theta$, $\sigma-\tilde{\sigma}=2ir\sin\theta$.

(iii) Follows easily by induction from the difference equation. Q.E.D.

Corollary 4.2.5. (i) *The smallest root of* P_n *is* $\frac{1}{4}\sec^2\frac{\pi}{n}$.

(ii) $P_n(\tau)>0$ *for* $\tau<\frac{1}{4}\sec^2\frac{\pi}{n}$.

(iii) $P_{n+1}(\tau)<0$ *for* τ *between* $\frac{1}{4}\sec^2\frac{\pi}{n+1}$ *and* $\frac{1}{4}\sec^2\frac{\pi}{n}$.

Proof. (i) By counting the number of distinct values of $\frac{1}{4}\sec^2\frac{m\pi}{n}$ (which are roots of P_n by 4.2.4) we find that all roots of P_n are real and they are the numbers $\frac{1}{4}\sec^2\frac{m\pi}{n}$ with $\frac{m\pi}{n}<\frac{\pi}{2}$. The smallest is $\frac{1}{4}\sec^2\frac{\pi}{n}$.

(ii) By induction the coefficient of $x^{[(n-1)/2]}$ in $P_n(x)$ is positive when $\left[\frac{n-1}{2}\right]$ is even and negative when $\left[\frac{n-1}{2}\right]$ is odd.

(iii) $P_{n+1}(\tau)$ must be negative between its first and second real roots, and $\sec^2\pi/(n+1)<\sec^2\pi/n<\sec^2 2\pi/(n+1)$. Q.E.D.

Proof of (iii) of 4.1.1. Suppose $\tau>1/4$ and $\tau\neq\frac{1}{4}\sec^2\frac{\pi}{n}$, $n=3,4,5,\ldots$. Then there is a $k\geqq 3$ with $\frac{1}{4}\sec^2\pi/(k+1)<\tau<\frac{1}{4}\sec^2\pi/k$. But then $P_n(\tau)>0$ for all $n\leqq k$ so $P_{n+1}(\tau)=\operatorname{tr}(1-s_{n-1})$ by 4.2.3. But by (iii) of 4.2.5, $P_{n+1}(\tau)<0$ which is impossible since $1-s_{n-1}$ is a projection. Q.E.D.

§4.3. Values of the Index

Theorem 4.3.1. *If* N *is a subfactor of the* II_1 *factor* M *then either* $[M:N]\geqq 4$ *or* $[M:N]=4\cos^2\pi/n$ *for some* $n\geqq 3$.

Proof. If $[M:N]<\infty$, define the increasing sequence M_i, $i=0,1,2,\ldots$ of II_1 factors by the relations $M_0=M$, $M_1=\langle M,e_N\rangle$, $M_{i+1}=\langle M_i,e_{M_{i-1}}\rangle$ for $i\geqq 1$. The inductive limit becomes a II_1 factor with faithful normal normalized trace tr (by uniqueness of the trace - see also [19]). Moreover if $\tau=[M:N]^{-1}$ and $e_i=e_{M_i}$ then the e_i satisfy the conditions of 4.1.1 by 3.4.2 and 3.1.7. By Theorem 4.1.1 either $[M:N]\geqq 4$ or $[M:N]=4\cos^2\pi/n$ for some $n\in\mathbb{Z}$, $n\geqq 3$. Q.E.D.

Theorem 4.3.2. *For each $n=3,4,\ldots$ there is a subfactor P_τ of R with $[R:P_\tau]=4\cos^2\pi/n$ and $P'\cap R=\mathbb{C}$. For each $r\geqq 4$, $r\in\mathbb{R}$, there is a subfactor P of R with $[R:P]=r$.*

Proof. The existence of subfactors with index $r\geqq 4$ was shown in 2.2.5 and the assertion about the relative commutant when $[R:P]=4\cos^2\pi/n$ follows from 2.2.4. Thus we only need to construct subfactors with index $4\cos^2\pi/n$.

To do this note that the conditions of 3.3.2 for finite dimensional C^*-algebras $N\subseteq M$ together with (ii) of 3.3.1 show, by interchanging A and A^T, $\vec{s}$ and $\vec{t}$, that if there is a (τ, M) trace on $\langle M, e_N\rangle$ then there is a $(\tau, \langle M, e_N\rangle)$ trace on $\langle\langle M, e_N\rangle, e_M\rangle$ and so on. Thus we may iterate the basic construction once we have started it. Once the construction has been iterated, the inductive limit has a faithful normal normalized trace on it and the e_i's resulting from the iteration satisfy the conditions of 4.1.1 so by the result of 4.1.1 we may choose P_τ as the subfactor of R.

Thus it suffices to find $N\subseteq M$, finite dimensional C^*-algebras with inclusion matrix A and positive vectors $\vec{s}$ and $\vec{t}$ with $A^TA\vec{t}=(4\cos^2\pi/n)\vec{t}$ and $AA^T\vec{s}=(4\cos^2\pi/n)\vec{s}$. In fact the matrix A is enough since the subalgebra N can then be taken as the direct sum of as many copies of $\mathbb{C}$ as there are rows in the matrix.

Let A be the square $n\times n$ matrix (a_{ij}) with $a_{ij}=1$ if $|i-j|=1$ and 0 otherwise, e.g. $\begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix}$. We leave it to the reader to check the linear algebra. This choice of A was suggested by F. Goodman. Q.E.D.

Remark 4.3.3. If we had started with any $m\times n$ non-negative integer valued matrix we could have made the construction of 4.3.2. Thus 4.1.1 implies that for such a matrix, if $\|A\|\leqq 2$ then $\|A\|=2\cos\pi/n$, $n=3,4,\ldots$. This can also be proved using the well-known result of Kronecker which asserts that if z is an algebraic integer all of whose conjugates have absolute value equal to 1, then z is a root of unity.

§5. The Bratteli Diagram; the Relative Commutant

§5.1. The Bratteli Diagram when $\tau\leqq 1/4$

Let $\{e_1, e_2, \ldots\}$, M and tr be as in §4.1. Let $A_n=\{e_1, e_2, \ldots, e_n\}''$ and B_n be the algebra generated by $\{e_1, e_2, \ldots, e_n\}$ (without 1). We know that A_n and B_n are finite dimensional C^*-algebras. We want to determine the Bratteli diagram for them and the value of tr on minimal projections. In this section we suppose $\tau\leqq 1/4$. This means that

$$P_{n+2}(\tau)=\mathrm{tr}(1-e_1\vee e_2\vee\ldots\vee e_n)>0 \quad \text{for all } n\geqq 1$$

(where P_n is as in §4.2).

Let $\left\{ {n \atop b} \right\} = \binom{n}{b} - \binom{n}{b-1}$ (ordinary binomial symbols with the convention $\binom{n}{-1} = 0$). We shall show by induction that

(a) $A_n = \bigoplus_{k=0}^{[\frac{n+1}{2}]} Q_k^n$ where $Q_k^n \cong M_{\left\{ {n+1 \atop k} \right\}}(\mathbb{C})$.

(b) $Q_0^n = (1 - e_1 \vee e_2 \vee \ldots \vee e_n)\mathbb{C}$ so that

(c) $B_n = \bigoplus_{k=1}^{[\frac{n+1}{2}]} Q_k^n$

(d) The trace of a minimal projection in Q_k^n is $\tau^k P_{n+2-2k}(\tau)$ for $k = 0, 1, \ldots, \left[\frac{n+1}{2}\right]$.

(e) The inclusion matrix of A_{n-1} in A_n is

(i) When n is even

$$A = (a_{ij}) \text{ with } a_{ij} = \begin{cases} 1 & \text{if } j = i \text{ or } i+1 \\ 0 & \text{otherwise,} \end{cases}$$

$i, j = 0, 1, \ldots, \left[\frac{n+1}{2}\right]$

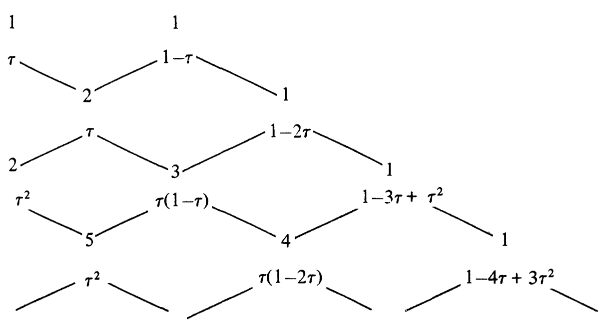
(here the indices i and j refer to the subscript of Q).

(ii) when n is odd

$$A = (a_{ij}) \; a_{ij} = \begin{cases} 1 & \text{if } j = i \text{ or } i+1 \\ 0 & \text{otherwise} \end{cases}$$

$i = 0, ., \ldots, (n+1)/2; \; j = 0, 1, \ldots, (n+3)/2.$

All this information is summed up by the following diagram (which also appears on p. 118 of [31]).



etc.

The numbers are the $\left\{ {n+1 \atop k} \right\}$ and the polynomials, which give the trace on minimal projections in the corresponding matrix algebra, change by multiplication by τ each step down a vertical column.

Proof. The proof will use the basic construction for the inclusion $A_{n-1} \subseteq A_n$ to obtain an almost faithful representation of A_{n+1}.

The truth of assertions (a)→(e) for $n=2$ follows immediately from §3.4. For the inductive step we shall treat only the case $n=2m$, the odd case being essentially identical. Suppose (a)→(e) are true for all $k \leqq n$ and apply the basic construction of §3.1 with $M=A_n$, $N=A_{n-1}$ with respect to tr. Let $E=E_{A_{n-1}}$, $e_{A_{n-1}}=e$. By induction and (ii) of 3.3.1 we know that we have the following Bratteli diagram:

$$\begin{tikzcd}[column sep=small]
A_{n-1} & \cdots & \left\{\begin{matrix} n \\ m \end{matrix}\right\} \arrow[dr, dash] & & \left\{\begin{matrix} n \\ m-1 \end{matrix}\right\} \arrow[dr, dash] & & \cdots & & \left\{\begin{matrix} n \\ 0 \end{matrix}\right\} \arrow[dl, dash] \arrow[dr, dash] & \\
\cap | & & & & & & & & & \\
A_n & \cdots & & \left\{\begin{matrix} n+1 \\ m \end{matrix}\right\} \arrow[dl, dash] \arrow[dr, dash] & & \left\{\begin{matrix} n+1 \\ m-1 \end{matrix}\right\} \arrow[dl, dash] \arrow[dr, dash] & \cdots & \left\{\begin{matrix} n+1 \\ 1 \end{matrix}\right\} \arrow[dl, dash] \arrow[dr, dash] & & \left\{\begin{matrix} n+1 \\ 0 \end{matrix}\right\} \arrow[dl, dash] \\
\cap | & & & & & & & & & \\
\langle A_n, e\rangle & \cdots & \left\{\begin{matrix} n+2 \\ m+1 \end{matrix}\right\} & & \left\{\begin{matrix} n+2 \\ m \end{matrix}\right\} & & \cdots & & \left\{\begin{matrix} n+2 \\ 1 \end{matrix}\right\} &
\end{tikzcd}$$

The numbers on the bottom line follow from 3.7.1, (ii) of 3.3.1 and the identities

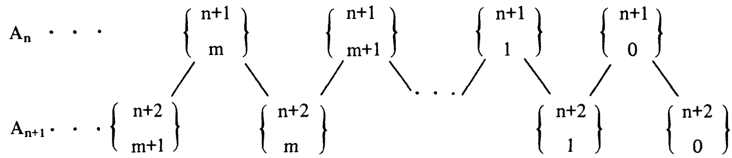
$$\left\{\begin{matrix} a \\ b \end{matrix}\right\} + \left\{\begin{matrix} a \\ b-1 \end{matrix}\right\} = \left\{\begin{matrix} a+1 \\ b \end{matrix}\right\}, \quad \left\{\begin{matrix} 2m+2 \\ m+1 \end{matrix}\right\} = \left\{\begin{matrix} 2m+1 \\ m \end{matrix}\right\}.$$

By assertion (c), the algebra generated by $\{e_1, e_2, \ldots, e_n\}$ is the direct sum of the first m terms on the A_n line so if we define the faithful (non-normalized) trace Tr on $\langle A_n, e\rangle$ by the rule $\mathrm{Tr}(ep_k)=\tau\,\mathrm{tr}(p_k)$ where p_k is a minimal projection in Q_{k-1}^{n-1} $(k=1,2,\ldots,m+1)$, the identity $P_j=P_{j+1}+\tau P_{j-1}$ ensures that Tr agrees with tr on B_n. Moreover as in 3.3.2, uniqueness of the trace and linearity show that $\mathrm{Tr}(ex)=\tau\,\mathrm{tr}(x)$ for all x in A_{n-1} and hence all x in A_n. Also Tr is a faithful positive trace since all the polynomials $P_n(\tau)$ are positive.

But now consider the algebras B_{n+1} and $\langle A_n, e\rangle$. B_{n+1} is generated by B_n and e_{n+1} so that any element can be written $a_0+\sum_{i=1}^{i} a_i e_{n+1} b_i$ with $a_0 \in B_n$, a_i, $b_i \in A_n$ for $i=1,2,\ldots,k$. Multiplication is defined by $e_{n+1} x e_{n+1}=E(x)e_{n+1}$ for $x \in A_n$ (see 4.1.3) and the faithful trace tr defined by $\mathrm{tr}(x e_{n+1})=\tau\,\mathrm{tr}(x)$ for $x \in A_n$. In $\langle A_n, e\rangle$, sums of the form $a_0+\sum_{i=1}^{j} a_i e b_i$ with $a_0 \in B_n$, a_i, $b_i \in A_n$ for $i \geqq 1$, form a 2-sided ideal. Since the central support of e is 1 ((iv) of 3.1.5), any element of $\langle A_n, e\rangle$ can be written in this form. Also $exe=E(x)e$ for $x \in A_n$ and the faithful trace Tr on $\langle A_n, e\rangle$ satisfies $\mathrm{Tr}(aeb)=\mathrm{tr}(aeb)$ for a, $b \in A_n$ and $\mathrm{Tr}(b)=\mathrm{tr}(b)$ for $b \in B_n$. Thus we may define a map from B_{n+1} to $\langle A_n, e\rangle$ by $a_0+\sum_i a_i e_{n+1} b_i \mapsto a_0+\sum_i a_i e b_i$ which is a surjective isometry for the definite hermitian scalar products defined by tr and Tr (and hence is well defined).

At this stage (†) we have obtained assertion (c) for $n+1$ and the values of tr on the minimal projections in Q_1^{n+1}, Q_2^{n+1}, ..., Q_{m+1}^{n+1}. But A_{n+1} is just $\{B_{n+1} \cup \{1\}\}''$ and since $\mathrm{tr}(1-e_1 \vee \ldots \vee e_{n+1})=P_{n+3}(\tau)>0$, A_{n+1} is $B_{n+1} \oplus (1-e_1 \vee e_2 \vee \ldots \vee e_{n+1})\mathbb{C}$. Moreover $x(1-e_1 \vee e_2 \vee \ldots \vee e_{n+1})=0$ for any $x \in B_n$

and $(1-e_1 \vee \ldots \vee e_{n+1})(1-e_1 \vee \ldots \vee e_n) \neq 0$ so the Bratteli diagram for $A_n \subseteq A_{n+1}$ is forced to be



This proves assertions (a) and (e) for $n+1$, and assertion (d) follows from $\operatorname{tr}(1-e_1 \vee \ldots \vee e_{n+1}) = P_{n+3}(\tau)$. This ends the proof. Q.E.D.

Aside 5.1.1. The binomial identity

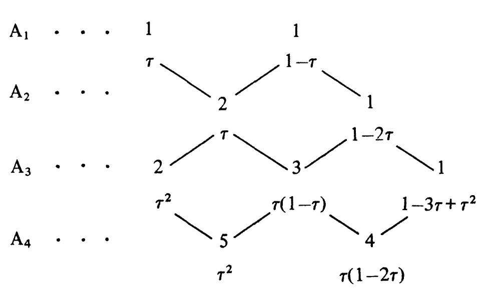
$$\sum_{i=0}^{\left[\frac{n+1}{2}\right]} \binom{n+1}{i}^2 = \frac{1}{n+2}\binom{2(n+1)}{n+1}$$

follows from [13, p. 63]. This shows that $\dim A_n$ is the same as the number of ordered reduced words which are thus linearly independent. See 4.1.4.

§5.2. The Bratteli Diagrams $\tau = \frac{1}{4}\sec^2\frac{\pi}{n}$

If $\tau = \frac{1}{4}\sec^2 \pi/(n+2)$ then $P_k(\tau) > 0$ for all $k \leqq n+1$ so the inductive argument of §5.1 goes through until the point marked (†) for assertions (a)→(e) up to step n. At this stage we find that $e_1 \vee e_2 \vee \ldots \vee e_n = 1$ so $B_n = A_n$. From this point on the basic construction for the pair $A_k \subseteq A_{k+1}$ will give an isometric (so faithful) surjective representation of A_{k+2} and the Bratteli diagram will not grow any wider.

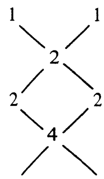
To convince the reader of these assertions without boring him with the details, we treat the case $n=4$. The argument of §5.1 shows that we have the following diagram



Now if $\vec{s} = \begin{pmatrix} \tau \\ 1-2\tau \end{pmatrix}$, $\vec{t} = \begin{pmatrix} \tau^2 \\ \tau(1-\tau) \\ 1-3\tau+\tau^2 \end{pmatrix}$, and $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$, $\vec{s}$ is an eigenvector for $A^T A$ with eigenvalue $3 = \tau^{-1} = 4\cos^2\frac{\pi}{6}$ and $\vec{t}$ is an eigenvector for AA^T with the same eigenvalue. They give normalized traces on A_2 and A_3 so by §3.3 the basic construction will continue to give faithful representations of the A_i by the same argument as §5.1, since $A_n = B_n$ for $n \geqq 4$.

The Bratteli diagrams will be:

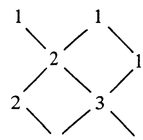
For $n=2$ $(\tau=1/2)$



etc.

Note that this is the same as the complex Clifford algebras (see [16, p. 148]).

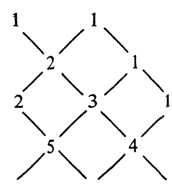
For $n=3$ $(\tau=1/\varphi^2,\ \varphi=\text{golden ratio})$



etc.

This diagram already appears in [4], [11].

For $n=4$



etc.

The pattern is now clear. The traces on minimal projections are the same as for the case $\tau < 1/4$ where this makes sense.

§5.3. *The Relative Commutant when* $\tau \leqq 1/4$

One of the main motivations of this paper was to decide whether there is a continuum of values of the index realized by subfactors with trivial relative commutant (see problem 1). It was originally thought that the subfactors P_τ, $\tau \leqq 1/4$, had trivial relative commutant. In this section we shall show that this is only true when $\tau = 1/4$. The calculations were first finished by A. Wassermann to whom the author is grateful. The difficulty is the absence of an orthonormal basis which makes it impossible to exhibit an element of the relative commutant.

The case $\tau < 1/4$. Let us adopt the notation of §4.1 so R is presented as the algebra generated by the e_i's. We intend to show that $E_{R \cap P'_\tau}(e_1)$ is not a scalar, i.e. $E_{R \cap P'_\tau}(e_1) \neq \tau$. For this we will show that $\|E_{R \cap P'_\tau}(e_1) - \tau\|_1 \neq 0$ (remember $\|x\|_1 = \mathrm{tr}(|x|)$). We shall need some lemmas.

Lemma 5.3.1. $E_{R\cap P'_\tau}(x)=s-\lim_{n\to\infty} E_{R\cap A'_{2,n}}(x)$ *for* $x\in R$.

Proof. Since $P_\tau=\left(\bigcup_{n=1}^{\infty} A_{2,n}\right)''$, the algebras $R\cap A'_{2,n}$ are a decreasing sequence of von Neumann algebras with intersection $P'_\tau\cap R$. The rest is well known (e.g. it is true in $L^2(R,\mathrm{tr})$). Q.E.D.

Lemma 5.3.2. $\|E_{A'_{2,n+1}\cap R}(e_1)-\tau\|_1=\|E_{A'_n\cap A_{n+1}}(e_{n+1})-\tau\|_1$.

Proof. Let du denote Haar measure on the unitary group of a finite dimensional C^*-algebra. Then

$$\|E_{A'_{2,n+1}\cap R}(e_1-\tau)\|_1=\|\int_{U(A_{2,n+1})}(ue_1u^*-\tau)\,du\|_1.$$

Applying the isomorphism σ_{n+1} of 4.1.11 we find that this expression equals

$$\|\int_{U(A_n)}(ue_{n+1}u^*-\tau)\,du\|_1$$

which equals $\|E_{A'_n\cap A_{n+1}}(e_{n+1})-\tau\|_1$. Q.E.D.

Thus to show that $P'_\tau\cap R\neq C$, it suffices to show that

$$\lim_{n\to\infty}\|E_{A'_n\cap A_{n+1}}(e_{n+1})-\tau\|_1\neq 0.$$

Since we know that the limit exists, it suffices to consider n odd; say $n=2m-1$. From §5.1 let $p_0,p_1,\dots,p_{m-1}$ be the minimal central projections in A_{n-1} corresponding to $Q_0^{n-1},Q_1^{n-1},\dots,Q_{m-1}^{n-1}$, similarly $q_0,q_1,\dots,q_m$ for A_n and $r_0,r_1,\dots,r_m$ for A_{n+1}. So $p_0=1-e_1\vee e_2\vee\dots\vee e_{n-1}$, $q_0=1-e_1\vee e_2\vee\dots\vee e_n$ and $r_0=1-e_1\vee e_2\vee\dots\vee e_{n+1}$. We also know from §5.1 and (iii) of 3.3.1 that

$$e_{n+1}r_i=e_{n+1}p_{i-1}\quad\text{for } i\geqq 1. \tag{5.3.3}$$

Since all the embeddings on the Bratteli diagram are of multiplicity at most one, the relative commutant of A_n in A_{n+1} is the abelian algebra generated by the mutually orthogonal projections $q_0r_0,q_0r_1,q_1r_0,\dots,q_mr_m$. Thus

$$E_{A'_n\cap A_{n+1}}(e_1)=\frac{\mathrm{tr}(q_0r_0e_{n+1})}{\mathrm{tr}(q_0r_0)}q_0r_0+\frac{\mathrm{tr}(q_0r_1e_{n+1})}{\mathrm{tr}(q_0r_1)}q_0r_0+\dots+\frac{\mathrm{tr}(q_mr_me_{n+1})}{\mathrm{tr}(q_mr_m)}q_mr_m$$

so that

$$\begin{aligned}\|E_{A'_n\cap A_{n+1}}(e_{n+1})-\tau\|_1&=\sum_{i=0}^{m}|\mathrm{tr}(q_ir_ie_{n+1})-\tau\,\mathrm{tr}(q_ir_i)|\\&\quad+\sum_{i=0}^{m-1}|\mathrm{tr}(q_ir_{i+1}e_{n+1}-\tau\,\mathrm{tr}(q_ir_i)|\\&\geqq\sum_{i=1}^{m}\tau|\mathrm{tr}(q_ip_{i-1})-\mathrm{tr}(q_ir_i)|\qquad\text{by 5.3.3}\\&\geqq\tau\left|\sum_{i=1}^{m}\mathrm{tr}(q_ip_{i-1})-\mathrm{tr}(q_ir_i)\right|.\end{aligned}$$

But from the Bratteli diagram, $q_i p_{i-1}$ is a projection of rank $\begin{Bmatrix} 2m-1 \\ i-1 \end{Bmatrix}$ in the matrix algebra Q_i^{2m-1} $(=q_i A_n)$ and $q_i r_i$ is of rank $\begin{Bmatrix} 2m \\ i \end{Bmatrix}$ in $Q_i^{2m}(=r_i A_{n+1})$. So the last sum may be written

$$L=\tau\left|\sum_{i=1}^{m}\left(\begin{Bmatrix} 2m-1 \\ i-1 \end{Bmatrix}\tau^i P_{2(m-i)+1}(\tau)-\begin{Bmatrix} 2m \\ i \end{Bmatrix}\tau^i P_{2(m-i)+2}(\tau)\right)\right|.$$

We saw in 4.2.4 that $P_n(\tau)=(\sigma^n-\tilde{\sigma}^n)/(\sigma-\tilde{\sigma})$ where $\sigma=(1+\sqrt{1-4\tau})/2$ and $\tilde{\sigma}=(1-\sqrt{1-4\tau})/2$. Note that for $\tau<1/4$, $\sigma>1/2$, $\tilde{\sigma}<1/2$ and $\tilde{\sigma}+(\tau/\tilde{\sigma})=1$, $\sigma+(\tau/\sigma)=1$. Let us now calculate the relevant limits.

Lemma 5.3.6. (a) $\lim\limits_{m\to\infty}\left(\sum\limits_{i=1}^{m}\begin{Bmatrix} 2m-1 \\ i-1 \end{Bmatrix}\tilde{\sigma}^{2(m-i)+1}\tau^i\right)=0$

(b) $\lim\limits_{m\to\infty}\left(\sum\limits_{i=1}^{m}\begin{Bmatrix} 2m \\ i \end{Bmatrix}\tilde{\sigma}^{2(m-i)+2}\tau^i\right)=0$

(c) $\lim\limits_{m\to\infty}\left(\sum\limits_{i=1}^{m}\begin{Bmatrix} 2m-1 \\ i-1 \end{Bmatrix}\sigma^{2(m-i)+1}\tau^i\right)=\tau(1-\tau/\sigma^2)$

(d) $\lim\limits_{m\to\infty}\left(\sum\limits_{i=1}^{m}\begin{Bmatrix} 2m \\ i \end{Bmatrix}\sigma^{2(m-i)+2}\tau^i\right)=\sigma^2(1-\tau/\sigma^2)$.

Proof. (a) and (c). Note that $\sigma^{2(m-i)+1}\tau^i=\tau(\tau/\sigma)^{i-1}\sigma^{2m-i}$ so if $x=\sigma$ or $\tilde{\sigma}$, in both cases we have to evaluate

$$\lim_{m\to\infty}\tau\left(\sum_{i=1}^{m}\binom{2m-1}{i-1}\left(\frac{\tau}{x}\right)^{i-1}x^{2m-i}-\frac{\tau}{x^2}\sum_{i=1}^{m}\binom{2m-1}{i-2}\left(\frac{\tau}{x}\right)^{i-2}x^{2m-i+1}\right).$$

Since $x+\tau/x=1$, we recognize the probability of $\geqq m$ successes in $2m-1$ Bernoulli trials. If $x=\sigma$, the probability of success is $>1/2$ so by the de Moivre-Laplace central limit theorem the limit is $\tau(1-\tau/\sigma^2)$ and if $x=\tilde{\sigma}$, the probability of success is $<1/2$ so the limit is 0.

(b) and (d) are proved in the same way with the substitution $x^{2(m-i)+2}\tau^i=x^2(\tau/x)^i x^{2m-i}$. Q.E.D.

Expanding L and using 4.3.5 we see that

$$(\sigma-\tilde{\sigma})/L=\tau|(1-\tau/\sigma^2)(\tau-\sigma^2)|\neq 0 \qquad \text{for } \tau\neq 1/4.$$

This shows that $P_\tau'\cap R\neq\mathbb{C}$ for $\tau<1/4$.

The case $\tau=1/4$. In this case we contend that $P_\tau'\cap R=\mathbb{C}$. Luckily we can use another model of $P_{1/4}$. Let R be realized as the closure of the Fermion algebra $\bigotimes\limits_{i=1}^{\infty}(M_2(\mathbb{C}))_i$ with respect to the trace tr. The fixed point algebra of the obvious infinite product action of $U(2)$ is generated by the representation of S_∞ coming from interchanging the tensor product components. The transpositions between successive components may be written $2e_i-1$ with $\mathrm{tr}(e_i)=1/4$ and it is a matter of calculation to show that the e_i's satisfy $e_i e_{i\pm1} e_i=\frac{1}{4}e_i$, $e_i e_j=e_j e_i$ for $|i-j|\geqq 2$. Thus the e_i algebra is $R^{U(2)}$ and the subfactor $P_{1/4}$ is $M_2(C)_1'\cap R^{U(2)}$.

But it is shown in [31] that $(R^{U(2)})' \cap R = \mathbb{C}$ so that $P'_{1/4} \cap R = M_2(\mathbb{C})_1$ and $M_2(\mathbb{C})_1 \cap R^{U(2)} = \mathbb{C}$. Thus $P'_{1/4} \cap R^{SU(2)} = \mathbb{C}$. For $R^{U(2)}$ see also [12], [15], [31], [2]. Q.E.D.

References

1. Aubert, P.L.: Théorie de Galois pour une W^*-algèbre. Comment. Math. Helv. **39**, 411–433 (1976)
2. Baker, B., Powers, R.: Product states and C^* dynamical systems of product type. Preprint, University of Pennsylvania
3. Bourbaki, N.: Algèbre, Ch. 8. Hermann, 1959
4. Bratteli, O.: Inductive limits of finite dimensional C^*-algebras. Trans. A.M.S. **171**, 195–234 (1972)
5. Christensen, E.: Subalgebras of a finite algebra. Math. Ann. **243**, 17–29 (1979)
6. Connes, A.: Periodic automorphisms of the hyperfinite factor of type II_1. Acta Sci. Math. **39**, 39–66 (1977)
7. Connes, A.: Classification of injective factors. Ann. Math. **104**, 73–115 (1976)
8. Connes, A.: Sur la théorie non-commutative de l'intégration. In: Algèbres d'opérateurs. Lecture Notes in Mathematics, Vol. 725. Berlin-Heidelberg-New York: Springer 1978
9. Connes, A.: A II_1 factor with countable fundamental group. J. Operator Theory **4**, 151–153 (1980)
10. Dixmier, J.: Les algèbres d'opérateurs dans l'espace Hilbertien. Deuxieme édition. Ganthier Villars, 1969
11. Effros, E., Shen, C.-L.: $A-F$ C^*-algebras and continued fractions. Indiana Math. J. **29**, 191–204 (1980)
12. Enomoto, M., Watatani, Y.: $C^*(\Pi)$ and Cuntz algebras from Young diagrams. Preprint
13. Feller, W.: Probability theory and its applications. I. Wiley and Sons, Inc. 1957
14. Goldman, M.: On subfactors of factors of type II_1. Mich. Math. J. **7**, 167–172 (1960)
15. Handelman, D.: Notes on ordered K-theory. Vancouver conference on operator algebras (1982)
16. Husemoller, D.: Fibre bundles. Grad. Texts in Math. Vol. 20. Berlin-Heidelberg-New York: Springer 1975
17. Jones, V.: Sur la conjugaison des sous-facteurs des facteur de type II_1. C.R. Ac. Sci. Paris t. **286**, 597–598 (1977)
18. Jones, V.: L'indice des sous-facteurs de facteurs de type II_1. C.R. Ac. Sci. Paris t. 286 (1977), 597–598
19. Kadison, R., Fugledge, B.: On a conjecture of Murray and von Neumann. Proc. Natl. Acad. Sci. U.S. **37**, 420–425 (1951)
20. Murray, F., von Neumann, J.: On rings of operators, II. Trans. A.M.S. **41**, 208–248 (1937)
21. Murray, F., von Neumann, J.: On rings of operators. IV. Ann. Math. **44**, 716–808 (1943)
22. Ocneanu, A.: Actions of discrete amenable groups on factors. To appear, Springer Lecture notes in Math. Berlin-Heidelberg-New York: Springer
23. Popa, S.: On a problem of R.V. Kadison on maximal abelian $*$-subalgebras in factors. Invent. Math. **65**, 269–281 (1981)
24. Skau, C.: Finite subalgebras of a von Neumann algebra. J. Funct. Anal. **25**, 211–235 (1977)
25. Suzuki, N.: Crossed products of rings of operators. Tohoku Math. J. **7**, 167–172 (1960)
26. Suzuki, N.: Extensions of rings of operators in Hilbert spaces. Tohoku Math. J. **14**, 217–232 (1962)
27. Stratila, S., Voiculescu, D.: Representations of AF algebras and of the group $U(\infty)$. Springer Lecture notes in Math.
28. Takesaki, M.: Theory of operator algebras I. Berlin-Heidelberg-New York: Springer 1979
29. Topping, D.: Lectures on von Neumann algebras. Nostrand Reinhold Math. Studies, 1969
30. Umegaki, H.: Conditional expectation in an operator algebra I. Tohoku Math. J. **6**, 358–362 (1954)
31. Wasserman, A.: Thesis, University of Pennsylvania, 1981

Oblatum 15-X-1982

Invent. math. 72, 27-55 (1983)

Inventiones mathematicae

The Unitary Spectrum for Real Rank one Groups

M.W. Baldoni Silva*[1] and D. Barbasch**[2]

[1] Institute for Advanced Study Princeton, NJ 08540, USA
[2] Rutgers University, New Brunswick, NJ 08903, USA

0. Introduction

Let G be a connected real semi-simple Lie group with finite center. One of the main problems in harmonic analysis is to determine the unitary spectrum of G. In this paper we treat this question in the case when real rank of G is 1.

Although the answer was known for G of classical type previously ([2], [5], [7], [15]), we have redone this work sometimes giving simpler arguments. With the arbitrary rank case in mind we have tried to deal with the problem of classifying unitary representations in as systematic a way as possible proceeding by induction on the dimension of G. We have concentrated on the case when G is linear and has a compact Cartan subgroup, the other cases being known already.

As an application we also give a list of the unitary representations that contribute to the L^2-index formula for the Dirac operator with coefficients. As is apparent from the calculations in [19], this is intimately connected with one of our main techniques for determining the unitary spectrum, the Dirac inequality. We will deal with some related problems in a future paper.

The paper is organized as follows:

In Sect. 1, we introduce notation and summarize the results necessary for the reduction to lower dimension groups.

Section 2 deals with the Dirac inequality and reducibility of principal series in terms of parabolic induction via the derived functor reviewed in section one. We show for the representations in question how to reduce the question of unitarity to lower (not necessarily strictly) dimensional groups. The results are summarized in Theorem 2.3.

In Sect. 3 we deal with the cases not considered in the previous section and we prove in particular that certain representations that are isolated are unitary.

* Supported in part by NSF grant MCS 81-08814
** Supported in part by an NSF grant

Theorem 3.2 provides a classification of the unitary irreducible representations in real rank one, modulo certain cases in $\mathrm{Sp}(n, 1)$, [2], as well as the spherical principal series case, [8].

Corollary 3.5 gives an affirmative answer to a conjecture of Zuckerman that a certain construction of irreducible representations via derived functors always leads to unitary representations.

In Sect. 4 we apply these results to determine the representations with $(\mathfrak{g}, K)$ cohomology and the representations contributing to the L^2-index formula alluded to earlier.

For a motivation for considering this problem we refer to the brief discussion at the beginning of Sect. 4 and to [19] and [13].

Our approach to classifying unitary representations owes a great deal to many conversations with D. Vogan. We would like to thank him for generously sharing his insight in the problems of reducibility of principal series and unitarity of representations.

We would also like to thank the referee for suggestions on improving the exposition of the paper and H. Kraljevic for valuable discussions on the composition series of $F_{4,1}$.

1. Preliminary Results

Let G be a semisimple real connected Lie group with finite center and Lie algebra $\mathfrak{g}_0$. We assume that $G \subseteq G_c$, where G_c is a linear complexification of G. We will denote by a subscript 0 a real vector space and we will drop the subscript for the complexification. For example, the Lie algebra of G_c will be denoted by $\mathfrak{g}$.

Extend θ, the Cartan involution for $\mathfrak{g}_0$, linearly to $\mathfrak{g}$ and let $\mathfrak{g}=\mathfrak{k}+\mathfrak{s}$ be the corresponding Cartan decomposition. We fix $\mathfrak{t}_0$ a θ stable fundamental Cartan of $\mathfrak{g}$ and we denote by K and T the analytic subgroups corresponding to $\mathfrak{k} \cap \mathfrak{g}$ and $\mathfrak{t}_0$.

Let $\mathfrak{p}_0=\mathfrak{m}_0+\mathfrak{a}_0+\mathfrak{n}_0$ be a minimal parabolic subalgebra of $\mathfrak{g}$ and $P_0 = M_0 A_0 N_0$ the corresponding subgroup.

We assume that $\dim \mathfrak{a}_0 = 1$ and $\mathfrak{t}_0 \subseteq \mathfrak{k}_0$. Let $\Delta=\Delta(\mathfrak{g}, \mathfrak{t})$ be the roots of $\mathfrak{g}$ relative to $\mathfrak{t}$ and $\Delta(\mathfrak{k})=\Delta(\mathfrak{k}, \mathfrak{t})$ the corresponding compact roots. If ψ is a positive system of roots for Δ, we denote by $\tilde{\psi}$ the corresponding closed Weyl chamber in $i\mathfrak{t}^*$ and we set $\rho(\psi)=\frac{1}{2}\sum_{\alpha\in\psi} \alpha$, $\rho_c(\psi)=\frac{1}{2} \sum_{\alpha \in \psi \cap \Delta(\mathfrak{k})} \alpha$ and $\rho_n(\psi)=\rho(\Psi) -\rho_c(\psi)$. In general for any $\mathfrak{t}$ stable vector space V we will write $\rho(V)=\frac{1}{2} \sum_{\alpha \in \Delta(V)} \alpha$, $\Delta(V)$ the roots of $\mathfrak{t}$ in V.

We denote by W the Weyl group of Δ and W_K the Weyl group of $\Delta(\mathfrak{k})$. If α is a root then s_α is the reflection by α.

We now fix a system of positive compact roots $\Delta^+(\mathfrak{k})$ and list all the positive systems for Δ containing $\Delta^+(\mathfrak{k})$ as follows:

a) $SO_e(2n,1)$.

$\Delta^+(\mathfrak{k})$: $\varepsilon_i \pm \varepsilon_j, \quad 1 \leqq i < j \leqq n$

$$\tilde{B}_1 = \left\{ \sum_{k=1}^{n} a_i \varepsilon_i,\ a_1 \geqq \ldots \geqq a_n \geqq 0 \right\}$$

$$\tilde{B}_2 = \left\{ \sum_{k=1}^{n} a_i \varepsilon_i,\ a_1 \geqq \ldots \geqq a_{n-1} \geqq -a_n \geqq 0 \right\}.$$

Then B_1 and B_2 contain $\Delta^+(\mathfrak{k})$ and

$$\rho_c = \rho(\Delta^+(\mathfrak{k})) = \sum_{k=1}^{n} (n-k)\, \varepsilon_k$$

$$\rho_n^1 = \rho_n(B_1) = \sum_{k=1}^{n} \varepsilon_k, \quad \rho_n^2 = \rho_n(B_2) = s_{\varepsilon_n} \rho_n^1.$$

b) $SU(n,1)$.

$\Delta^+(\mathfrak{k})$: $\varepsilon_i - \varepsilon_j, \quad 1 \leqq i < j \leqq n$

$$\tilde{A}_i = \left\{ \sum_{k=1}^{n+1} a_k \varepsilon_k,\ a_1 \geqq \ldots \geqq a_i \geqq a_{n+1} \geqq a_{i+1} \geqq \ldots \geqq a_n \right\}, \qquad 0 \leqq i \leqq n$$

$$\rho_c = \tfrac{1}{2} \sum_{k=1}^{n} (n+1-2k)\, \varepsilon_k$$

$$\rho_n^i = \rho_n(A^i) = \tfrac{1}{2} \left\{ \sum_{k=1}^{i} \varepsilon_k - \sum_{k=i+1}^{n} \varepsilon_k + (n-2i)\, \varepsilon_{n+1} \right\}.$$

c) $Sp(n,1)$.

$\Delta^+(\mathfrak{k})$: $\varepsilon_i \pm \varepsilon_j, \quad 1 \leqq i < j \leqq n; \qquad 2\varepsilon_i, \quad 1 \leqq i \leqq n+1.$

$$\tilde{C}_i = \left\{ \sum_{k=1}^{n+1} a_k \varepsilon_k :\ a_1 \geqq \ldots \geqq a_i \geqq a_{n+1} \geqq a_{i+1} \geqq \ldots \geqq a_n \geqq 0 \right\} \qquad 0 \leqq i \leqq n.$$

$$\rho_c = \sum_{k=1}^{n} (n-k+1)\, \varepsilon_k + \varepsilon_{n+1}$$

$$\rho_n^i = \rho_n(C_i) = \sum_{k=1}^{i} \varepsilon_k + (n-i)\, \varepsilon_{n+1}.$$

d) $F_{4,1}$.

$\Delta^+(\mathfrak{k})$: $\varepsilon_i \pm \varepsilon_j, \quad 1 \leqq i < j \leqq 4; \qquad \varepsilon_i, \quad 1 \leqq i \leqq 4.$

0——0 $\Longrightarrow$ 0——0

F_3: $\varepsilon_2 - \varepsilon_3\ \varepsilon_3 - \varepsilon_4\ \varepsilon_4\ \ 1/2(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4)$

$F_2 = s_{1/2(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4)} F_3$

$F_1 = s_{1/2(\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4)} F_2$

$\rho_c = 7/2\, \varepsilon_1 + 5/2\, \varepsilon_2 + 3/2\, \varepsilon_3 + 1/2\, \varepsilon_4$

$\rho_n^1 = \rho_n(F_1) = \varepsilon_1 + \varepsilon_2 + \varepsilon_3$

$\rho_n^2 = \rho_n(F_2) = 3/2\, \varepsilon_1 + 1/2(\varepsilon_2 + \varepsilon_3 + \varepsilon_4)$

$\rho_n^3 = \rho_n(F_3) = 2\, \varepsilon_1.$

If B is the Cartan Killing form on $\mathfrak{t}$ we let $\langle\ ,\ \rangle$ denote the bilinear form defined by B on $i\mathfrak{t}^*$.

By a θ-stable parabolic subalgebra of $\mathfrak{g}$ we will mean a subalgebra constructed from some element $\gamma \in i\mathfrak{t}^*$ in the following way. We set $\mathfrak{g}=\mathfrak{l}+\mathfrak{u}$ where $\mathfrak{t}\subseteq\mathfrak{l}$, $\Delta(\mathfrak{l})=\{\alpha\varepsilon\Delta:\langle\alpha,\gamma\rangle=0\}$ and $\Delta(\mathfrak{u})=\{\alpha\varepsilon\Delta:\langle\alpha,\gamma\rangle>0\}$. We will say that such a $\mathfrak{q}$ is determined by γ. Set $\mathfrak{l}_0=\mathfrak{l}\cap\mathfrak{g}_0$, then $\mathfrak{l}_0$ is reductive θ-stable subalgebra whose complexification is $\mathfrak{l}$. We fix, given such $\mathfrak{l}$, a positive system $\Delta^+(\mathfrak{l}\cap\mathfrak{k})$ of $\mathfrak{l}\cap\mathfrak{k}$ with respect to $\mathfrak{t}$ such that $\Delta^+(\mathfrak{l}\cap\mathfrak{k})\subseteq\Delta^+(\mathfrak{k})$. $W(\mathfrak{l})$, $W_{L\cap K}$ will denote the corresponding Weyl groups, and L the subgroup corresponding to $\mathfrak{l}_0$. Irreducible K-types (respectively $L\cap K$ types) are always parametrized by the highest weight with respect to $\Delta^+(\mathfrak{k})$ (respectively $\Delta^+(\mathfrak{l}\cap\mathfrak{k})$). Let now $P=MAN$ be a parabolic subgroup of G. If δ is a discrete series representation of M and v is a complex valued linear functional on the Lie algebra of A, then we define

$$\pi_P(\delta, v)=\operatorname{Ind}_P^G(\delta\otimes e^v\otimes 1),$$

where the induced representation is defined in such a way that G acts on the left and the representation is unitary of v is imaginary. In [16], Sect. 7, it is shown how to obtain the Langlands data for an irreducible $(\mathfrak{g}, K)$ module π from its lowest (in the sense of [16]) K-type μ. This is useful for calculations involving the Dirac operator. We review this procedure in the special case of our setting.

We fix μ a lowest K-type. Then we choose a positive system ψ in $\Delta(\mathfrak{g}, t)$ such that $\mu+2\rho_c$ is dominant with respect to it. Then $\mu+2\rho_c-\rho(\psi)$, according to Proposition 4.1 in [16], is either dominant with respect to ψ or else there is a simple noncompact root $\beta\varepsilon\psi$ such that

$$2\langle\mu+2\rho_c-\rho(\psi),\beta\rangle\langle\beta,\beta\rangle^{-1}=-1.$$

We define

$$\bar{\lambda}(\mu)=\begin{cases}\mu+2\rho_c-\rho(\psi) & \text{if } \mu+2\rho_c-\rho(\psi) \text{ is dominant}\\ \mu+2\rho_c-\rho(\psi)+\frac{1}{2}\beta & \text{if it is not.}\end{cases}$$

In the real rank one case $\bar{\lambda}(\mu)$ is dominant with respect to ψ and is precisely the parameter associated to μ by Proposition 4.1 of [16].

We denote by $\mathfrak{b}^0=\mathfrak{l}^0+\mathfrak{u}^0$ the θ-stable parabolic subalgebra determined by $\bar{\lambda}(\mu)$.

Then $\mathfrak{b}^0$ determines a cuspidal parabolic subgroup of $G, P=MAN$ and a discrete series representation δ of M. This is either G and a discrete series or a minimal parabolic subgroup $P_0=M_0A_0N_0$ and a finite dimensional unitary irreducible representation. Let $v\in\mathfrak{a}_0^*$ then $\pi_P(\delta, v)$ contains the K-type μ with multiplicity one. Moreover if $\pi_P(\delta, v, \mu)$ denotes the unique subquotient of $\pi_P(\delta, v)$ containing the K-type μ, then π is infinitesimally equivalent to it. If M is connected then $\delta=\bar{\lambda}(\mu)$. By abuse of notation we will write simply $\bar{\lambda}(\mu)$ for δ. (We also remind the reader that M_0 is disconnected only if $\mathfrak{g}=\mathfrak{sl}(2,\mathbf{R})$.)

In particular discrete series representations are parametrized by $\bar{\lambda}(\mu)$'s which are regular ($P=G$); and limit of discrete series by $(\bar{\lambda}(\mu), 0)$ and a positive system ψ such that $\bar{\lambda}(\mu)$ is dominant with respect to ψ and $\langle\bar{\lambda}(\mu), \alpha\rangle = 0$ for α simple implies α noncompact.

We also remark that if $P = P_0$ and $\operatorname{Re} \nu > 0$ then $\pi_{P_0}(\bar{\lambda}(\mu), \nu, \mu) \cong \bar{\pi}_{P_0}(\bar{\lambda}(\mu), \nu)$ where $\bar{\pi}_P(\bar{\lambda}(\mu), \nu)$ is the unique Langlands quotient of the corresponding induced representation. For a complete account of the quoted results compare [11, 10, 16].

It is well known that the problem of classifying the irreducible unitary representations of G amounts to describing precisely which representations $\bar{\pi}_{P_0}(\bar{\lambda}(\mu), \nu)$, $\nu > 0$, are infinitesimally unitary. We will consider only such representations and so from now on we will drop the subscript P_0.

Let $\mathfrak{u}(\mathfrak{g})$ be the universal enveloping algebra of $\mathfrak{g}$. Fix a K-type $\mu \in \Delta^+(\mathfrak{k})$ and let $\bar{\lambda}(\mu)$ be defined as above and $\mathfrak{b}^0$ the corresponding parabolic. Let $\mathfrak{q} \supseteq \mathfrak{b}^0$ be a θ-stable parabolic subalgebra and set $\mu_L = \mu - 2\rho(\mathfrak{u} \cap \mathfrak{s})$. Set $\lambda^0 = \bar{\lambda}(\mu)$ and $\lambda_L^0 = \bar{\lambda}(\mu_L) = \bar{\lambda}(\mu - 2\rho(\mathfrak{u} \cap \mathfrak{s}))$.

The following theorem is an immediate consequence of 4.17–4.21 in [14].

Theorem 1.1. *In the setting described, assume also that*

$$(*) \qquad \operatorname{Re}\langle(\lambda^0, \nu), \alpha\rangle \geqq 0 \quad \text{for each } \alpha \in \Delta(\mathfrak{u}).$$

Then

a) $\pi(\lambda^0, \nu)$ *has a number of composition factors which is less than or equal to the number of composition factors of* $\pi(\lambda_L^0, \nu)$. *In particular*

b) $\pi(\lambda^0, \nu)$ *is irreducible if* $\pi(\lambda_L^0, \nu)$ *is irreducible.*

We now need to recall the basic properties of the Harish-Chandra modules constructed by Vogan-Zuckerman in order to be able to compute various types of cohomology. We state the results we need without proof. A basic reference is [17].

Let $\mathfrak{q} = \mathfrak{l} + \mathfrak{u}$ be a θ-stable parabolic and L as before. Let W be an $(\mathfrak{l}, L \cap K)$ module. Then in [17] Sect. 6, a family of $(\mathfrak{g}, K)$ modules $\{\mathscr{R}^i W\}$ is associated to W with the following properties. Set $S = \dim(\mathfrak{u} \cap \mathfrak{k})$.

1. [17, cor. 6.3.21] $\mathscr{R}^i W = 0$ for $i < 0$ and $i > S$.

2. [17, th. 6.3.12] Let $\delta \in \Delta^+(\mathfrak{k})$. Set $W_K^1 = \{w \in W_k \colon w^{-1}\alpha < 0,\ \alpha \in \Delta^+(\mathfrak{k})$ then $\alpha \in \Delta(\mathfrak{u} \cap \mathfrak{k})\}$. Suppose δ occurs in $\mathscr{R}^{S-i}(W)$. Then there is $w \in W_k^1$ of length i, non-negative integers n_β, $\beta \in \Delta(\mathfrak{u} \cap \mathfrak{s})$ and an $L \cap K$ type η occurring in W such that

$$\eta = w(\delta + \rho_c) - \rho_c - 2\rho(\mathfrak{u} \cap \mathfrak{s}) - \sum_{\beta \in \Delta(\mathfrak{u} \cap \mathfrak{s})} n_\beta \beta.$$

The multiplicity of δ is at most the number of such expressions (counted with the multiplicity of η in W). Furthermore, if W has finite length, then

$$\sum (-1)^i m(\delta, \mathscr{R}^{S-i} W) = \sum_{\gamma} m(\eta, W) \sum_{w \in W_k} (-1)^{\mathfrak{l}(w)} P(w(\delta + \rho_c) - \rho_c - 2\rho(\mathfrak{u} \cap \mathfrak{s}) - \eta)$$

where $\mathfrak{l}(w)$ is the length of w, m is the multiplicity and P is the Kostant partition function with respect to the roots of $\mathfrak{u} \cap \mathfrak{s}$.

3. Let $\gamma_G = (\bar{\lambda}(\mu), \nu)$ be the Langlands data for a $(\mathfrak{g}, K)$ module. Let $\pi(\gamma_G)$ be the corresponding induced representation with lowest K-type μ. By abuse of notation if γ_G is the parameter for a limit of discrete series we still denote by $\pi(\gamma_G)$ the corresponding limit of discrete series. Assume $\mathfrak{g} \supseteq \mathfrak{b}^0$ and set $\gamma_{\mathfrak{q}}^L = \gamma_G - \rho(\mathfrak{u})$, $\mu_L = \mu - 2\rho(\mathfrak{u} \cap \mathfrak{s})$. Then

$$\mathcal{R}^i(\pi(\gamma_{\mathfrak{q}}^L)) = \begin{cases} 0 & i \neq S \\ \pi(\gamma_G) & i = S \end{cases}$$

whenever $\mathrm{Re}\langle(\bar{\lambda}(\mu), \nu), \alpha\rangle \geqq 0$ for all $\alpha \in \Delta(\mathfrak{u})$ (cf. [17] a special case of 8.2.15).

4. If $0 \to W_1 \to W_2 \to W_3 \to 0$ is an exact sequence then there is a long exact sequence

$$\ldots \to \mathcal{R}^i W_1 \to \mathcal{R}^i W_2 \to \mathcal{R}^i W_3 \to \mathcal{R}^{i+1} W_1 \to \ldots .$$

Corollary 1.2. Let $\mathrm{Re}\,\nu > 0$ and $\bar{\pi}_L = \bar{\pi}((\bar{\lambda}(\mu_L), \nu) - \rho(\mathfrak{u}))$ be the Langlands quotient of $\pi(\gamma_{\mathfrak{q}}^L)$. Then under the assumption in 3

$$\mathcal{R}^i(\bar{\pi}_L) = \begin{cases} 0 & i \neq S \\ \bar{\pi}(\gamma_G) & i = S. \end{cases}$$

Proof. The proof is by descending induction on the length of $\bar{\lambda}(\mu)$. We give a sketch of it. By 3 it is true for $\langle(\bar{\lambda}(\mu), \bar{\lambda}(\mu))\rangle$ maximal because of 6.6.6 of [17]. Otherwise we have an exact sequence

$$0 \to \bar{\pi}(\gamma_{\mathfrak{q}}^L) \to \pi(\gamma_{\mathfrak{q}}^L) \to X_L \to 0$$

and the composition factors are higher in the induction than $\pi(\gamma_{\mathfrak{q}}^L)$ (6.6.6 of [17]). Therefore

$$\mathcal{R}^i(X_L) = 0 \quad \text{and} \quad \mathcal{R}^i \pi(\gamma_{\mathfrak{q}}^L) = 0, \qquad i < S.$$

Using the long exact sequence in 4 we get $\mathcal{R}^i(\bar{\pi}(\gamma_{\mathfrak{q}}^L)) = 0$ if $i < S$, and also $\mathcal{R}^S \bar{\pi}(\gamma_{\mathfrak{q}}^L) \supset \bar{\pi}(\gamma_G)$ because of the induction hypothesis. Because of Theorem 1.1a the conclusion follows.

2. Cases With Only Complementary Series

We recall the definition of complementary series.

Definition 2.1 [9]. For a fixed M-parameter $\bar{\lambda}(\mu)$, the complementary series associated to it is defined as the set of $\pi(\bar{\lambda}(\mu), \nu)$, with $\nu > 0$ such that $\pi(\bar{\lambda}(\mu), \nu)$ is unitary and irreducible. The boundary of this set, for $\nu > 0$, is called the endpoints of the complementary series.

For convenience we will refer to *complementary series* as the *union of the two sets in Definition 2.1.*

The main result in this section is Theorem 2.3. In essence it reduces the problem of calculating complementary series of an arbitrary $\bar{\lambda}(\mu)$ to the corresponding question for the parameter of a spherical principal series. The main tools are the Dirac inequality and the reducibility results in Sect. 1. The technical statements of the reduction are Definition 2.2 and Theorem 2.3. It is enough to consider G and μ satisfying the following restrictions (the other cases are well known or can be reduced to this situation).

A. G is connected, simple, linear and has compact Cartan subgroups. Also $\mathfrak{g} \neq \mathfrak{sl}(2, \mathbf{R})$.

B. μ is not the lowest K-type of a discrete series or of a nondegenerate limit of discrete series.

We can exclude the case in which μ is the lowest K-type of a nondegenerate limit of discrete series because there is no complementary series attached to $\bar{\lambda}(\mu)$, in other words the nondegenerate limit of discrete series is the only unitary representation with that lowest K-type. This can be checked directly by using the Dirac inequality or by an argument as in [2].

Definition 2.2. Let $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ be a θ-stable parabolic subalgebra. We say that μ is L-trivial if $\mathfrak{q} \supseteq \mathfrak{b}^0$ and the restriction of μ_L to $[\mathfrak{l}, \mathfrak{l}] \cap \mathfrak{k}$ is trivial.

Theorem 2.3. *Let μ be a K-type which is not L-trivial for any L that contains a factor isomorphic to $Sp(m, 1)$, $m \geqq 2$. Then there exists an (explicitly constructed) θ-stable parabolic subalgebra $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ with the following properties:*

(1) *For $\nu>0$, $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ and $\bar{\pi}(\bar{\lambda}(\mu_L), \nu)$ are either both unitary or both nonunitary. $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ and $\bar{\pi}(\bar{\lambda}(\mu_L), \nu)$ are unitary if and only if they are complementary series.*

(2) *On $[\mathfrak{l}, \mathfrak{l}]$, $(\bar{\lambda}(\mu_L), \nu)$ is either the parameter of a spherical principal series or else $[\mathfrak{l}, \mathfrak{l}]=\mathfrak{sp}(m, 1)$ with $m \geqq 2$ and $\mu_L=a\varepsilon_{m+1}$ with $a>0$.*

(3) *For $\nu>0$, if $\bar{\pi}=\bar{\pi}(\bar{\lambda}(\mu), \nu)$ and $\bar{\pi}_L=\bar{\pi}(\bar{\lambda}(\mu_L), \nu)$ are both unitary, then*

$$\mathscr{R}^i \bar{\pi}_L = \begin{cases} 0 & \text{if } i<S \\ \bar{\pi} & \text{if } i=S. \end{cases}$$

Remarks. (a) The definition of the L-trivial K-type is made to avoid the case of the trivial representation which is isolated in $\mathrm{Sp}(n, 1)$ and $F_{4,1}$. The μ's that do not satisfy the hypotheses of the theorem will be classified in the course of the proof and a list of them appears at the start of Sect. 3. They will be handled separately in that section.

(b) Under the hypotheses on μ, the theorem reduces the question of whether $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ is unitary for the cases in (2) of the theorem. Then $L=G$ and the classification of the unitary spectrum is well known by [2] and [8, §7]. We will recall the necessary results as they are needed for the proof.

Proof. The proof will occupy the remainder of this section. Our objective will be to construct a θ-stable parabolic subalgebra $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ with the properties (2.1), (2.2), (2.3), (2.4) listed below. Before doing the actual construction, we show how the theorem follows once we have such a $\mathfrak{q}$. Then we establish a sufficient condition for (2.2) to be satisfied. Finally we give a case by case construction of $\mathfrak{q}$.

The subalgebra $\mathfrak{q}$ is to satisfy the following four properties:

(2.1) $\mathfrak{q} \supseteq \mathfrak{b}^0$.

(2.2) The Dirac inequality for $\bar{\pi}(\bar{\lambda}(\mu), \nu)$, stated in (2.5) below, gives an upper bound for unitary points $\nu > 0$ that is $\leqq$ the corresponding upper bound given by the Dirac inequality for $\bar{\pi}(\bar{\lambda}(\mu_L), \nu)$.

(2.3) $(\bar{\lambda}(\mu_L), \nu)$ is as in (2) of Theorem 2.3.

(2.4) If $(\bar{\lambda}(\mu_L), \nu)$ satisfies the Dirac inequality on $\mathfrak{l}$, then

$$\langle(\bar{\lambda}(\mu), \nu), \alpha\rangle \geqq 0 \qquad \text{for } \alpha \in \Delta(\mathfrak{u}).$$

For the representations given by (2) of Theorem 2.3 it can be checked using [8] and [2] that the upper bound for unitary points ν coincides with the endpoint of the complementary series and essentially with the one given by the Dirac inequality (see the detailed calculations in the proof). Call it ν_L. Due to conditions (2.1), (2.4) we can apply Theorem 1.1 to deduce that $\pi(\bar{\lambda}(\mu), \nu)$ and $\pi(\bar{\lambda}(\mu_L), \nu)$ are both irreducible for $\nu < \nu_L$. By (2.2) if $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ is unitary then $\nu \leqq \nu_L$. Thus $\nu = \nu_L$ is the endpoint of the complementary series, so in particular $\pi(\bar{\lambda}(\mu), \nu)$ is also reducible. The theorem now follows from Corollary 1.2.

Thus Theorem 2.3 is proved once we construct $\mathfrak{q}$ satisfying (2.1)–(2.4).

We now obtain a sufficient condition on $\mathfrak{q}$ such that $\mathfrak{q}$ satisfies (2.2) whenever (2.3) holds. This condition, (2.6) below, is easier to check in practice.

Recall that by hypothesis, μ is not the lowest K-type of a discrete series or a nondegenerate limit of discrete series. Then $\bar{\lambda}(\mu) = \mu + 2\rho_c - \rho(\psi)$ if $\mu + 2\rho_c - \rho(\psi)$ is already dominant (but singular with respect to ψ) or $\bar{\lambda}(\mu) = \mu + 2\rho_c - \rho(\psi) + \frac{1}{2}\beta$, where β is a simple noncompact root.

It is well known (for example [2]) that any unitary representation $\bar{\pi}$, with lowest K-type μ, has to satisfy the Dirac inequality

$$\|\gamma + \rho_c\|^2 \geqq \|(\bar{\lambda}(\mu), \nu)\|^2$$

where γ is a K-type of $\bar{\pi} \otimes s$ (s being the spin representation).

Because of (2.3) and the discussion right after (2.4) we will only need it in the following form

(2.5) $\|\sigma(\mu - \rho_n) + \rho_c\|^2 \geqq \|(\bar{\lambda}(\mu), \nu)\|^2$

where $\rho_n = \rho(\Phi) - \rho_c$ for some positive root system $\Phi \supset \Delta^+(\mathfrak{k})$ and $\sigma \in W_K$ is such that $\sigma(\mu - \rho_n)$ is dominant with respect to $\Delta^+(\mathfrak{k})$.

Consider all the systems Φ of the form $\Phi = \Phi(\mathfrak{l}) \cup \Delta(\mathfrak{u})$ where $\Phi(\mathfrak{l})$ is a positive system for $\mathfrak{l}$ compatible with $\Delta^+(\mathfrak{k})$, i.e. $\rho_c = \rho(\Phi(\mathfrak{l}) \cap \Delta(\mathfrak{k})) + \rho(\mathfrak{u} \cap \mathfrak{k})$. Set $\rho(\mathfrak{l}) = \rho(\Phi(\mathfrak{l}))$, $\rho(\mathfrak{l} \cap \mathfrak{k}) = \rho(\Phi(\mathfrak{l}) \cap \Delta(\mathfrak{k}))$ and $\rho_n(\mathfrak{l}) = \rho(\mathfrak{l}) - \rho(\mathfrak{l} \cap \mathfrak{k})$. Then $\rho_n = \rho_n(\Phi) = \rho_n(\mathfrak{l}) + \rho(\mathfrak{u} \cap \mathfrak{s})$.

(2.6) Our sufficient condition on $\mathfrak{q}$ for condition (2.2) to hold is that the σ used to make $\mu - \rho_n$ dominant is in $W_{L \cap K}$.

Then

(2.7) $\mu - \rho_n = \mu_L - \rho_n(\mathfrak{l}) + \rho(\mathfrak{u} \cap \mathfrak{s})$.

Since $\sigma \in W_{L\cap K}$ then $\sigma(\rho(\mathfrak{u}\cap\mathfrak{s}))=\rho(\mathfrak{u}\cap\mathfrak{s})$ and so we have

$$(2.8)\quad \begin{aligned}\sigma(\mu-\rho_n)+\rho_c&=\sigma(\mu_L-\rho_n(\mathfrak{l}))+\rho(\mathfrak{u}\cap\mathfrak{s})+\rho(\mathfrak{l}\cap\mathfrak{k})+\rho(\mathfrak{u}\cap\mathfrak{k})\\&=\sigma(\mu_L-\rho_n(\mathfrak{l}))+\rho(\mathfrak{l}\cap\mathfrak{k})+\rho(\mathfrak{u}).\end{aligned}$$

On the other hand

$$(2.9)\quad \mu+2\rho_c-\rho=\mu_L+2\rho(\mathfrak{l}\cap\mathfrak{k})-\rho(\mathfrak{l})+\rho(\mathfrak{u})$$

so $\bar{\lambda}(\mu)=\bar{\lambda}_L(\mu)+\rho(\mathfrak{u})$ because $\mathfrak{q}\supseteq\mathfrak{b}^0$.

Since $\rho(\mathfrak{u})$ is orthogonal to the roots of $\mathfrak{l}$ and $\beta\in\Delta(\mathfrak{l})$ we can write

$$(2.10)\quad \langle\sigma(\mu_L-\rho_n(\mathfrak{l}))+\rho(\mathfrak{l}\cap\mathfrak{k}),\ \rho(\mathfrak{u})\rangle=\langle\mu_L,\rho(\mathfrak{u})\rangle$$

and

$$(2.11)\quad \langle(\bar{\lambda}(\mu),\nu),\rho(\mathfrak{u})\rangle=\langle\mu_L+\rho(\mathfrak{u}),\rho(\mathfrak{u})\rangle.$$

Suppose that

$$(2.12)\quad \|\sigma(\mu_L-\rho_n(\mathfrak{l}))+\rho(\mathfrak{l}\cap\mathfrak{k})\|^2\geqq\|(\bar{\lambda}(\mu_L),\nu)\|^2$$

then (2.8)-(2.12) give

$$\|\sigma(\mu-\rho_n)+\rho_c\|^2\geqq\|(\bar{\lambda}(\mu),\nu)\|^2.$$

Thus $\mathfrak{q}$ satisfies condition (2.2) whenever (2.3) is satisfied. We now list the parabolic subalgebras $\mathfrak{q}$ for each μ.

We also remark that for calculating end points of complementary series it is enough to consider the simple noncompact factor of $\mathfrak{l}_0=\mathfrak{l}\cap\mathfrak{g}_0$ and μ_L restricted to it.

Construction of $\mathfrak{q}$ *for* $\mathfrak{g}=\mathfrak{so}(2n,1)$. We recall that a K-type is parametrized by

$$\mu=\sum_{k=1}^{n}\mu_k\varepsilon_k,\qquad \mu_1\geqq\ldots\geqq\mu_{n-1}\geqq|\mu_n| \text{ and } \mu_k-\mu_j\in\mathbf{Z},\ 2\mu_k\in\mathbf{Z},\ 1\leqq k,\ j\leqq n.$$

Because μ is not the lowest K-type of a discrete series or a limit of discrete series then $\mu_j\in\mathbf{Z}$ and $\mu_n=0$. Set $\psi=B_1$ and let $1\leqq t\leqq n$ be the smallest integer such that

$$\mu_t=\ldots=\mu_n=0.$$

Then $\mu+2\rho_c-\rho_1=\sum_{k=1}^{n}(\mu_k+n-k-1/2)\,\varepsilon_k$ and

$$(\bar{\lambda}(\mu),\nu)=\sum_{k=1}^{n-1}(\mu_k+n-k-1/2)\,\varepsilon_k+\nu\,\varepsilon_n.$$

Define $\mathfrak{q}$ by means of the element $\gamma=\sum_{k=1}^{t-1}(n-k)\,\varepsilon_k$, then

$$[\mathfrak{l}_0,\mathfrak{l}_0]=\mathfrak{so}(2(n-t+1),1)$$

and μ_L restricted to $[\mathfrak{l},\mathfrak{l}]\cap\mathfrak{k}$ is trivial.

The complementary series for $\pi(\bar{\lambda}(\mu_L), \nu)$ extends to the point $n-t+1/2$ ([8]). On the other hand condition (2.4) is

$$\mu_{t-1}+n-t+1/2 \geqq \nu \quad \text{if } 2 \leqq t \leqq n.$$

Because $\mu_{t-1}>0$ (if $t \geqq 2$) and the coefficient of ε_k in the expression of ρ_n^i is $1/2$ if $1 \leqq k \leqq n-1$, and $1/2$ or $-1/2$ if $k=n$, then $\mu-\rho_n^i$ can be made dominant by using only $\sigma \in W_{L \cap K}$. Thus (2.1)-(2.4) hold and the theorem follows.

Construction of $\mathfrak{q}$ *for* $\mathfrak{g}=\mathfrak{su}(n,1)$. In this case a K-type μ is parametrized by $\mu = \sum_{k=1}^{n+1} \mu_k \varepsilon_k$, where $\mu_k \in \mathbf{Z}$, $1 \leqq k \leqq n+1$, and $\mu_1 \geqq \ldots \geqq \mu_n$. Let $1 \leqq j \leqq n+1$ be the unique integer defined by:

$$\mu_{j-1}+n-2j+3>\mu_{n+1} \geqq \mu_j+n-2j+1.$$

Then $\mu+2\rho_c$ is dominant for A_{j-1} so we choose $\psi=A_{j-1}$.

$$\mu+2\rho_c-\rho_{j-1}=\sum_{k=1}^{j-1}\left(\mu_k+\frac{n}{2}-k\right)\varepsilon_k+\sum_{k=j}^{n}\left(\mu_k+\frac{n}{2}-k+1\right)\varepsilon_k$$
$$+\left(\mu_{n+1}-\frac{n}{2}+j-1\right)\varepsilon_{n+1}.$$

There are four cases.

If $\mu+2\rho_c-\rho_{j-1}$ is regular and dominant for A_{j-1}, then $(\bar{\lambda}(\mu), \nu)=\bar{\lambda}(\mu)=\mu+2\rho_c-\rho_{j-1}$ and μ is the lowest K-type of a discrete series representation.

If $\mu+2\rho_c-\rho_{j-1}$ is dominant for A_{j-1}, i.e.

$$\mu_{j-1}+n-2j+2 \geqq \mu_{n+1} \geqq \mu_j+n-2j+2 \quad \text{and} \quad \mu_{j-1}>\mu_j,$$

then the only unitary irreducible representation with the lowest K-type μ is a limit of discrete series as can be seen by [16] or using the Dirac operator which forces the real Langlands parameter to be zero. Excluding these two cases we are left with

a$_1$) $\mu_{n+1}=\mu_j+n-2j+2$ and $\mu_{j-1}=\mu_j$, for some j, $2 \leqq j \leqq n$.

a$_2$) $\mu_{n+1}=\mu_j+n-2j+1$, for some j, $1 \leqq j \leqq n$.

We do the reduction in two steps. In the first step (2.6) holds and in the second step we have to check (2.3) directly. Therefore (2.3) will hold combining the two steps.

Step 1. Let $0 \leqq t_0 \leqq j-1$ and $0 \leqq t_1 \leqq n-j$ be the largest integers such that

$$\mu_{j-t_0}=\ldots=\mu_j=\ldots=\mu_{j+t_1}=y.$$

Let now

$$\gamma=\sum_{k=1}^{j-t_0-1}(n-k+1)\varepsilon_k+\sum_{k=j-t_0}^{j+t_1}(n-j+t_0+1)\varepsilon_k+\sum_{k=j+t_1+1}^{n}(n-k+1)\varepsilon_k$$
$$+(n-j+t_0+1)\varepsilon_{n+1}.$$

Then the corresponding parabolic subalgebra $\mathfrak{q} \supseteq \mathfrak{b}^0$ and $[\mathfrak{l}_0, \mathfrak{l}_0] = \mathfrak{su}(t_0+t_1+1, 1)$

$$\mu_L|_{\mathfrak{u}(t_0+t_1+1)\cap\mathfrak{t}} = \begin{cases} \sum_{k=j-t_0}^{j+t_1} y\varepsilon_k + (y+1+t_1-t_0)\varepsilon_{n+1} & \text{in case } \mathrm{a}_1) \\ \sum_{k=j-t_0}^{j+t_1} y\varepsilon_k + (y+t_1-t_0)\varepsilon_{n+1} & \text{in case } \mathrm{a}_2). \end{cases}$$

Assume now that $\mathrm{a}_1)$ holds. Then

$$(\bar{\lambda}(\mu), v) = \sum_{k=1}^{j-1} \left(\mu_k + \frac{n}{2} - k\right)\varepsilon_k + \sum_{k=j}^{n} \left(\mu_k + \frac{n}{2} - k + 1\right)\varepsilon_k$$
$$+ \left(\mu_j + \frac{n}{2} - j + 1\right)\varepsilon_{n+1} + v(\varepsilon_{n+1} - \varepsilon_j)$$

and the inequality in (2.4) is equivalent to

$$v \leqq \min(\mu_{j-t_0-1} - \mu_j + t_0,\ \mu_j - \mu_{j+t_1+1} + t_1 + 1),$$

where the minimum has to be taken only on the terms which are defined. The first term is defined only if $t_0 \leqq j-2$ and the second only if $t_1 \leqq n-j-1$. If none of the terms are defined there is no condition on v. In the course of the proof we will often encounter an expression of this form, with one more term in the case of $\mathfrak{g} = \mathfrak{sp}(n, 1)$. The meaning of such an expression will be always subjected to the above restrictions on t_0 and t_1.

Similarly if $\mathrm{a}_2)$ holds then

$$(\bar{\lambda}(\mu), v) = \sum_{k=1}^{j-1} \left(\mu_k + \frac{n}{2} - k\right)\varepsilon_k + \left(\mu_j + \frac{n+1}{2} - j\right)\varepsilon_j$$
$$+ \sum_{k=j+1}^{n} \left(\mu_k + \frac{n}{2} - k + 1\right)\varepsilon_k + \left(\mu_j + \frac{n+1}{2} - j\right)\varepsilon_{n+1} + v(\varepsilon_{n+1} - \varepsilon_j)$$

and the inequality in (2.4) amounts to

$$v \leqq \min(\mu_{j-t_0-1} - \mu_j + t_0 + 1/2,\ \mu_j - \mu_{j+t_1+1} + t_1 + 1/2).$$

One checks easily that condition (2.6) is satisfied both in case $\mathrm{a}_1)$ and $\mathrm{a}_2)$. Thus condition (2.1) and (2.6) hold. On the other hand if μ satisfies $\mathrm{a}_1)$ (respectively $\mathrm{a}_2)$), then $\mu_L|_{[\mathfrak{l}_0, \mathfrak{l}_0]\cap\mathfrak{t}}$ satisfies $\mathrm{a}_1)$ (respectively $\mathrm{a}_2)$) for $\mathfrak{su}(t_1+t_0+1, 1)$ with $j = t_0+1$. Therefore we are left to consider the following situation ($m = t_0+t_1+1$ in the applications).

Step 2. Let $\mathfrak{g} = \mathfrak{su}(m, 1)$ and $\mu = \mu_{m+1}\varepsilon_{m+1}$ and either

$\mathrm{a}_1')$ $\mu_{m+1} = m - 2j + 2 \quad 2 \leqq j \leqq m$ or

$\mathrm{a}_2')$ $\mu_{m+1} = m - 2j + 1 \quad 1 \leqq j \leqq m.$

Let $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ be the θ-stable parabolic determined by

$$\gamma=\sum_{k=1}^{j-t-1}(m-k+1)\,\varepsilon_k+\sum_{k=j-t}^{j+\tilde{t}}(m-j+t+1)\,\varepsilon_k$$
$$+\sum_{k=j+\tilde{t}+1}^{n}(m-k+1)\,\varepsilon_k+(m-j+t+1)\,\varepsilon_{m+1}$$

where $t=\tilde{t}=\min(j-1, m-j)$ if $a_2')$ holds and $t=\min(j-1, m-j+1)$, $\tilde{t}=t-1$ if $a_1')$ holds. As before we easily obtain that $[\mathfrak{l}_0, \mathfrak{l}_0]=\mathfrak{su}(t+\tilde{t}+1, 1)$ and μ_L is the trivial representation. For this case condition (2.6) is not satisfied but (2.2) can be checked directly by exhibiting a specific ρ_n giving an upper bound that is smaller than the corresponding upper bound on $\mathfrak{l}$. Finally the inequality in (2.4) gives

$$\nu\leqq\begin{cases}t & \text{in case } a_1')\\ t+1/2 & \text{in case } a_2').\end{cases}$$

By [8, §7] the complementary series for L extends to the point t if $a_1')$ holds and $t+1/2$ if $a_2')$ holds.

Combining the two steps we get a parabolic subalgebra with $[\mathfrak{l}_0, \mathfrak{l}_0]=\mathfrak{su}(t+\tilde{t}+1, 1)$ satisfying (2.1)–(2.4). This completes the proof of the theorem in this case.

For the convenience of the reader we point out that if μ satisfies $a_1)$ then μ_L satisfies $a_1')$ and then in particular the complementary series for $\bar{\lambda}(\mu)$ extends up to (inclusive)

$$t=\min(j-1, m-j+1)$$

with $m=t_0+t_1+1$ and $j=t_0+1$, in other words up to $t=\min(t_0, t_{i+1})$. Similarly if μ satisfies $a_2)$ then μ_L satisfies $a_2')$ and the complementary series for $\bar{\lambda}(\mu)$ extends up to (inclusive)

$$t=\min(t_0, t_1)+1/2.$$

We now continue with the proof of the theorem.

Construction of $\mathfrak{q}$ *for* $\mathfrak{g}=\mathfrak{sp}(n,1)$. A K-type μ is parametrized by $\mu=\sum_{k=1}^{n+1}\mu_k\varepsilon_k$, where μ_k are integers and $\mu_1\geqq\ldots\geqq\mu_n\geqq 0$, $\mu_{n+1}\geqq 0$. The fact that μ is not L trivial for any L with a factor isomorphic to $sp(m, 1)$ means that μ_{n-1}, μ_n and μ_{n+1} are not simultaneously zero. (Otherwise we could choose the parabolic determined by $\gamma=\varepsilon_{n-1}+\varepsilon_n+\varepsilon_{n+1}$ and get a contradiction).

Let $1\leqq j\leqq n+1$ be defined by

$$\mu_{j-1}+2n-2j+4>\mu_{n+1}+2\geqq\mu_j+2n-2j+2.$$

Then $\mu+2\rho_c$ is dominant for C_{j-1} and we choose $\psi=C_{j-1}$.

$$\mu+2\rho_c-\rho_{j-1}=\sum_{k=1}^{j-1}(\mu_k+n-k)\,\varepsilon_k+\sum_{k=j}^{n}(\mu_k+n-k+1)\,\varepsilon_k$$
$$+(\mu_{n+1}-n+j)\,\varepsilon_{n+1}.$$

Exactly as in the case of $\mathfrak{su}(n, 1)$, because of the hypothesis on μ we are left with the following possibilities:

c$_1$) $\mu_{n+1}=\mu_j+2n-2j$, for some j, $1\leqq j\leqq n$

c$_2$) $\mu_{n+1}=\mu_j+2n-2j+1$ and $\mu_{j-1}=\mu_j$, for some j, $2\leqq j\leqq n$.

Let t_0, t_1, y and γ be defined as in the case $\mathfrak{g}=\mathfrak{su}(n,1)$ and first assume $y>0$. In this case we will show that Theorem (2.3) holds with $(\bar{\lambda}(\mu_L), \nu)$ a parameter for the spherical principal series. The reason for this assumption is that the σ needed to make the $\mu-\rho_n$'s dominant can be chosen in $W_{L\cap K}$, since the coefficients of ε_k in one such ρ_n are 0 or 1. Therefore in this case (2.6) is automatically satisfied. Let $\mathfrak{q}$ be the parabolic determined by γ (γ as in step 1 for $\mathfrak{su}(n, 1)$. Then

$$\mu_L|_{\mathfrak{u}(t_0+t_1+1,1)\cap\mathfrak{t}}=\begin{cases}\sum\limits_{k=j-t_0}^{j+t_1}(y-1)\,\varepsilon_k+(y-1+t_1-t_0)\,\varepsilon_{n+1} & \text{in case c}_1)\\ \sum\limits_{k=j-t_0}^{j+t_1}(y-1)\,\varepsilon_k+(y+t_1-t_0)\,\varepsilon_{n+1} & \text{in case c}_2).\end{cases}$$

Assume now that c$_1$) holds. Then

$$(\bar{\lambda}(\mu), \nu)=\sum_{k=1}^{j-1}(\mu_k+n-k)\,\varepsilon_k+(\mu_j+n-j+1/2)\,\varepsilon_j$$
$$+\sum_{k-j+1}^{n}(\mu_k+n-k+1)\,\varepsilon_k+(\mu_j+n-j+1/2)\,\varepsilon_{n+1}+\nu(\varepsilon_{n+1}-\varepsilon_j)$$

and the inequality in (2.4) amounts to

$$\nu\leqq\min(\mu_j-\mu_{j+t_1+1}+t_1+1/2,\ \mu_{j-t_1-1}-\mu_j+t_0+1/2,\ \mu_j+n-j+1/2).$$

Similarly if c$_2$) holds then

$$(\bar{\lambda}(\mu), \nu)=\sum_{k=1}^{j-1}(\mu_k+n-k)\,\varepsilon_k+(\mu_j+n-j+1)\,\varepsilon_j+\sum_{k=j+1}^{n}(\mu_k+n-k+1)\,\varepsilon_k$$
$$+(\mu_j+n-j+1)\,\varepsilon_{n+1}+\nu(\varepsilon_{n+1}-\varepsilon_j)$$

and the inequality in (2.4) is

$$\nu\leqq\min(\mu_j-\mu_{j+t_1+1}+t_1+1,\ \mu_{j-t_0-1}-\mu_j+t_0,\ \mu_j+n-j+1).$$

Therefore in both cases (2.1) and (2.6) are satisfied.

We now observe that if μ satisfies c$_1$) (respectively c$_2$)) then μ_L is as in step 2 for $su(n, 1)$ (with $m=t_0+t_1+1$) and satisfies a$'_2$) (respectively a$'_1$)) with $j=t_0+1$. Therefore in both cases we can apply step 2 to get the theorem, under the assumption $y>0$. ($(\bar{\lambda}(\mu_L), \nu)$ is the parameter of a spherical principal series.) As a consequence if μ satisfies c$_1$) (respectively c$_2$)) then the complementary series for $\bar{\lambda}(\mu)$ extends to the point (inclusive) $\min(t_0, t_1)+1/2$ (respectively $\min(t_0, t_1+1)$).

We now assume that $y=0$. In this case we will shown that the theorem is true with $[\mathfrak{l}_0, \mathfrak{l}_0]=\mathfrak{sp}(m,1)$, $m\geqq 2$ and $\mu_L=a\varepsilon_{m+1}$. In the course of the argument we will point out the cases for which $a=0$.

Because $y=0$ then $t_1=n-j$. Let $\gamma=\sum_{k=1}^{j-t_0-1}(n-k)\varepsilon_k$ and $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ be the corresponding parabolic subalgebra. Then $[\mathfrak{l}_0, \mathfrak{l}_0]\simeq\mathfrak{sp}(n-j+t_0+1, 1)$ and

$$\mu_L|_{[\mathfrak{l}_0,\mathfrak{l}_0]\cap\mathfrak{k}}=\begin{cases}2(n-j)\,\varepsilon_{n+1} & \text{if } \mathrm{c}_1) \text{ holds}\\ 2(n-j+1/2)\,\varepsilon_{n+1} & \text{if } \mathrm{c}_2) \text{ holds.}\end{cases}$$

If $j=n$ and $\mathrm{c}_1)$ holds then by assumption $t_0=0$ (otherwise $\mu_{n-1}=\mu_n=\mu_{n+1}=0$) and hence $[\mathfrak{l}_0, \mathfrak{l}_0]\simeq\mathfrak{so}(4,1)$ and μ_L is the trivial representation. In this case the complementary series for L extends to the point $3/2$. Otherwise following [2] it is not too difficult to see that the complementary series for π_L extends to the point $t_0+1/2$ if $\mathrm{c}_1)$ holds with $j\leqq n-1$, and to t_0 if $\mathrm{c}_2)$ holds.

On the other hand the inequality in (2.4) is

$$\begin{aligned} &v\leqq\mu_{j-t_0-1}+t_0+1/2 && \text{if } t_0\leqq j-2 \text{ and } \mathrm{c}_1) \text{ holds}\\ &v\leqq\mu_{j-t_0-1}+t_0 && \text{if } t_0\leqq j-2 \text{ and } \mathrm{c}_2) \text{ holds}\end{aligned}$$

(no condition if $t_0=j-1$).

With this choice of $\mathfrak{q}$ it is obvious that $\mu-\rho_n$ can be made dominant by an element of $W_{L\cap K}$, unless $\mu_{j-t_0-1}=1$ and $t_0\leqq j-2$. On the other hand if $\mu_{j-t_0-1}=1$ then by [2] we may still prove that if $\pi(\bar{\lambda}(\mu), v)$ is unitary then $v\leqq\bar{t}$ where $t_0+1/2\leqq\bar{t}<t_0+3/2$ if $\mathrm{c}_1)$ holds and $t_0\leqq\bar{t}<t_0+1$ if $\mathrm{c}_2)$ holds. Therefore (2.2) is always satisfied. The theorem now follows also in this case.

Remark 2.4. In the situation just described, i.e. $y=0$, we could have actually considered two cases:

i) $\mu_{j-1}=0$ and $t_0\leqq n-j=t_1$, or $j=1$ or $\mu_{j-1}>0$ (resp. $\mu_{j-2}=0$ and $t_0\leqq n-j+1=t_1+1$ or $j=2$ or $\mu_{j-2}>0$)
if $\mathrm{c}_1)$ holds (resp. if $\mathrm{c}_2)$ holds).

ii) $\mu_{j-1}=0$ and $t_0>n-j$ (resp. $t_0>n-j+1$) if $\mathrm{c}_2)$ holds).

The situation described in i) behaves exactly as the corresponding cases for $y\neq 0$. The only problem is that it is necessary to supply a direct proof of (2.2), since "σ" cannot be chosen in $W_{L\cap K}$. On the other hand for ii) we are forced to choose a parabolic for which the Levi factor contains a factor of type $\mathfrak{sp}(m,1)$. Therefore we prefer to treat the two cases together, in spite of the fact that they really behave differently.

Construction of $\mathfrak{q}$ *for* $\mathfrak{g}=\mathfrak{f}_{4,1}$. We recall that a K-type μ is parametrized by

$$\mu=\sum_{k=1}^{4}\mu_k\varepsilon_k=(\mu_1,\mu_2,\mu_3,\mu_4), \quad \text{where} \quad \mu_1\geqq\ldots\geqq\mu_4\geqq 0, \quad \mu_i-\mu_j\in\mathbf{Z}$$

and $2\mu_k\in\mathbf{Z}$. The assumption on μ implies that if $\mu_3=\mu_4=0$ then $\mu_1>\mu_2$. (Otherwise we could choose $\gamma=\varepsilon_1+\varepsilon_2$ and get a contradiction). Then,

$$\mu+2\rho_c=(\mu_1+7,\mu_2+5,\mu_3+3,\mu_4+1).$$

Assume first that $\mu+2\rho_c$ is dominant for F_3. Because of the assumption on ν,

$$\mu_1=\mu_2+\mu_3+\mu_4-2$$

and

$$(\bar{\lambda}(\mu),\nu)=(\mu_1+7/4+\nu/2,\mu_2+9/4-\nu/2,\mu_3+5/4-\nu/2,\mu_4+1/4-\nu/2).$$

Let $\gamma=\varepsilon_1+\varepsilon_2$, then $[\mathfrak{l}_0,\mathfrak{l}_0]=\mathfrak{sp}(2,1)$ and the positive roots for

$$\Delta(\mathfrak{l}\cap\mathfrak{k})\quad\text{are}\quad\{\varepsilon_1-\varepsilon_2,\varepsilon_3\pm\varepsilon_4,\varepsilon_3,\varepsilon_4\}.$$

The inequality in (2.4) gives

$$\nu\leqq\mu_2+\mu_4+5/2.$$

Let $\Delta^+(\mathfrak{l})=F_3\cap\Delta(\mathfrak{l})$. Then

$$\begin{aligned}\bar{\lambda}(\mu_L)&=\bar{\lambda}(\mathfrak{u})+\rho(\mathfrak{w}=(\mu_1-9/4,\mu_2-7/4,\mu_3+5/4,\mu_4+1/4)\\&=\tfrac{1}{2}(\mu_1+\mu_2-4)(1,1,0,0)+\tfrac{1}{2}(\mu_1-\mu_2-1/2)(1,-1,1,1)\\&\quad+\tfrac{1}{2}(\mu_3-\mu_4+1)(0,0,1,-1)\end{aligned}$$

and

$$\begin{aligned}\mu_L&=\tfrac{1}{2}(\mu_1+\mu_2-4)(1,1,0,0)+\tfrac{1}{2}(\mu_1-\mu_2)(1,-1,0,0)\\&\quad+(\mu_3+\mu_4)(0,0,0,1)+\mu_3(0,0,1,-1)\end{aligned}$$

where $(1,1,0,0)$ is orthogonal to the roots of $\mathfrak{l}$. In this case the Dirac inequality for μ_L is sharp (it gives $1/2$ as upper bound) and is obtained by taking $\mu_L-\rho(\Delta^+(\mathfrak{l}\cap\mathfrak{s}))$ and $\sigma=\mathrm{id}$ in (2.5) (see [2], Proposition 4.6). On the other hand

$$\begin{aligned}&\|\mu_L-\rho(\Delta^+(\mathfrak{l}\cap\mathfrak{s}))+\rho(\mathfrak{l}\cap\mathfrak{k})\|^2-\|(\bar{\lambda}(\mu_L),\nu)\|^2\\&=\|\mu-\rho_n^3+\rho_c\|^2-\|(\bar{\lambda}(\mu),\nu)\|^2\end{aligned}$$

where $\mu-\rho_n^3$ is a K-type. Therefore $\mathfrak{q}$ satisfies (2.2). We can now apply to $\mathfrak{sp}(2,1)$ and μ_L the reduction argument from $\mathfrak{sp}(n,1)$ to find a parabolic satisfying (2.1)-(2.4) with $(\bar{\lambda}(\mu_L),\nu)$ a parameter for a spherical principal series. (In the usual parametrization for $\mathfrak{sp}(2,1)$,

$$\mu_L=(\mu_3+\mu_4)\,\varepsilon_1+(\mu_3-\mu_4)\,\varepsilon_2+(\mu_3+\mu_4+2)\,\varepsilon_3$$

and μ_L satisfies c_1) with $j=1$. If $\mu_3+\mu_4>0$ then we are in the situation $y>0$ and so $(\bar{\lambda}(\mu_L),\nu)$ is indeed the parameter for a spherical principal series. If $y=0$ then because $j=1$, Remark 2.4i) will apply to yield the same conclusion.)

Assume now that $\mu+2\rho_c$ is dominant for F_1. Then because of the assumption on μ,

$$\mu_2+\mu_3=\mu_4+\mu_1.$$

In this case

$$(\bar{\lambda}(\mu),\nu)=(\mu_1+9/4-\nu/2,\mu_2+7/4+\nu/2,\mu_3+3/4+\nu/2,\mu_4+1/4-\nu/2).$$

We have two possibilities

f$_1$) $\mu_1=\mu_2=\mu_3=\mu_4$ and $2\mu_1\geqq 1$

f$_2$) μ not as in f$_1$.

Suppose f$_2$) holds. Let $\gamma=\varepsilon_1+\varepsilon_2$ and $\mathfrak{q}$ be the corresponding parabolic. Then $[\mathfrak{l}_0,\mathfrak{l}_0]=\mathfrak{sp}(2,1)$. Let $\Delta^+(\mathfrak{l})=F_1\cap\Delta(\mathfrak{l})$. Then

$$\bar{\lambda}(\mu_L)=\tfrac{1}{2}(\mu_1+\mu_2-4)(1,1,0,0)+\tfrac{1}{2}(\mu_1-\mu_2+1/2)(1,-1,1,-1)$$
$$+\tfrac{1}{2}(\mu_3+\mu_4+1)(0,0,1,1)$$

and

$$\mu_L=\tfrac{1}{2}(\mu_1+\mu_2-4)(1,1,0,0)+\tfrac{1}{2}(\mu_1-\mu_2)(1,-1,0,0)$$
$$+(\mu_3+\mu_4)(0,0,0,1)+\mu_3(0,0,1,-1).$$

The Dirac inequality for μ_L is sharp: it gives the point 1/2 if $\mu_1>\mu_2$ (cf. [2], 2.5 and 4.6) and the point 3/2 if $\mu_1=\mu_2$ (cf. [2] 5.1).

Moreover the same values are obtained by applying the Dirac inequality (2.5) to the $L\cap K$-type $\mu_L-\rho(\Delta^+(\mathfrak{l})\cap\mathfrak{s})$ in the first case and $s_{\varepsilon_1-\varepsilon_2}(\mu_L-\rho(\Delta(\mathfrak{l})\cap F_3\cap\mathfrak{s}))$ in the second case.

Because $\rho(\Delta^+(\mathfrak{l})\cap\mathfrak{s})=(0,0,1,0)$, $\rho(\Delta(\mathfrak{l})\cap F_3\cap\mathfrak{s})=(1/2,-1/2,1/2,1/2)$ one easily obtains

$$\|\mu_L-\rho(\Delta^+(\mathfrak{l}\cap\mathfrak{s})+\rho(\mathfrak{l}\cap\mathfrak{k})\|^2-\|(\bar{\lambda}(\mu_L,\nu)\|^2=\|\mu-\rho_n^1+\rho_c\|^2-\|(\bar{\lambda}(\mu),\nu)\|^2$$

and

$$\|s_{\varepsilon_1-\varepsilon_2}(\mu_L-\rho(\Delta(\mathfrak{l})\cap F_3\cap\mathfrak{s}))+\rho(\mathfrak{l}\cap\mathfrak{k})\|^2-\|(\bar{\lambda}(\mu_L),\nu)\|^2$$
$$=\|s_{\varepsilon_1-\varepsilon_2}(\mu-\rho_n^2)+\rho_c\|^2-\|(\bar{\lambda}(\mu),\nu)\|^2$$

where $\mu-\rho_n^1$ and $s_{\varepsilon_1-\varepsilon_2}(\mu-\rho_n^2)$ are K-types. Therefore $\mathfrak{q}$ satisfies (2.2). The inequality in (2.4) gives

$$\nu\leqq\mu_1-\mu_3+3/2.$$

We now repeat the argument for the chamber F_3 to complete the proof in case f$_2$). To be precise in the usual parametrization for $\mathfrak{sp}(2,1)$,

$$\mu_L=(\mu_3+\mu_4)\,\varepsilon_1+(\mu_3-\mu_4)\,\varepsilon_2+(\mu_3-\mu_4)\,\varepsilon_3,$$

$j=2$ and μ_L satisfies c$_1$). If $\mu_3>\mu_4$ then we are in the situation $y>0$, otherwise $y=0$ and the parabolic to which we apply a further reduction is of the form $\mathfrak{so}(4,1)$ (compare the beginning of the discussion after the assumption $y=0$ in the case of $\mathfrak{sp}(n,1)$. In both cases $(\bar{\lambda}(\mu_L),\nu)$ is in the final reduction the parameter of a spherical principal series.

Suppose now that f$_1$) holds. Let $\gamma=\varepsilon_1+\varepsilon_2+\varepsilon_3+\varepsilon_4$ then $[\mathfrak{l}_0,\mathfrak{l}_0]=\mathfrak{so}(6,1)$. Let

$$\Delta^+(\mathfrak{l})=F_1\cap\Delta(\mathfrak{l})\quad\text{and}\quad\Delta^+(\mathfrak{l}\cap\mathfrak{k})=F_1\cap\Delta(\mathfrak{l}\cap\mathfrak{k}).$$

$\mu_{L|_{\mathfrak{so}(6,1)\cap\mathfrak{k}}}$ is the trivial representation for $L\cap K$. Therefore the complementary series for π_L extends to the point 5/2 [8] and the inequality in (2.4) is equivalent to

$$\nu\leqq 2\mu_1+1/2.$$

If $\mu_1 \geqq 1$ then $s_{\varepsilon_1-\varepsilon_2}(\mu-\rho_n^1)$ gives the same estimate on $\mathfrak{g}$. If $\mu_1=1/2$, then the lower bound for the Dirac inequality (2.5) on G is $\bar{t}$, $7/2<\bar{t}<4$. This lower bound can be further improved by applying the Dirac operator to the K-type (3/2, 1/2, 1/2, 1/2) which is in the Langlands quotient at $\nu=7/2$. (The fact that (3/2, 1/2, 1/2, 1/2) is in the Langlands quotient can be checked by a lengthy but straight forward calculation of composition factors and multiplicity of K-types. We omit the details).

We now show that in fact the conclusion of Corollary 1.2 extends to $\nu=5/2$, even though $\mathrm{Re}\langle(\bar{\lambda}(\mu),\nu),\alpha)\rangle \geqq 0$ is not satisfied.

We first note that by [6] the composition series for the spherical principal series of $\mathfrak{so}(6,1)$ at $\nu=5/2$ gives an exact sequence

$$0 \to W_1 \to W \to W_2 \to 0$$

where W_2 is a one dimensional representation and W_1 is the Langlands quotient with parameter $(\bar{\lambda}(1/2,\,1/2,\,-1/2,\,-1/2),\,3/2)$. Then, by Corollary 1.2 for W_1 and Theorem 8.2.15 for W, we get

$$\mathscr{R}^i W = \begin{cases} 0 & i<S \\ \pi(\bar{\lambda}(1/2,1/2,1/2,1/2),5/2) & i=S \end{cases}$$

$$\mathscr{R}^S W_1 = \begin{cases} 0 & i<S \\ \bar{\pi}(\bar{\lambda}(2,2,1,1),\tfrac{3}{2}) & i=S. \end{cases}$$

Then the exact sequence 4 in Sect. 1 collapses to

$$0 \to \mathscr{R}^{S-1} W_2 \to \mathscr{R}^S W_1 \to \mathscr{R}^S W \to \mathscr{R}^S W_2 \to 0$$

(in particular, $\mathscr{R}^i W_2=0$ for $i<S$).

We first show that $\mathscr{R}^{S-1} W_2=0$. For this it is enough to see that Eq. 2 in Sect. 1 has no solution such that $w \in W_k^1$ has length 1. The only such w is $s_{\varepsilon_3+\varepsilon_4}$. But then if $\delta=(x_1,\,x_2,\,x_3,\,x_4)$ with $x_1 \geqq x_2 \geqq x_3 \geqq x_4 \geqq 0$, we get the following system of equations

$$\frac{x_1-x_2}{2}=n_4-n_5$$

$$\frac{x_1+x_4}{2}+1=n_3-n_5$$

$$\frac{x_1+x_3}{2}+1=n_2-n_5$$

$$\frac{x_2-x_4-x_3-x_1}{2}=5+n_1+2n_5$$

with n_1, n_2, n_3, n_4, n_5 non negative integers. This is a contradiction so

$$\mathscr{R}^{S-1} W_2=0.$$

It remains to show that $\mathscr{R}^S W_2=\bar{\pi}(\bar{\lambda}(\tfrac{1}{2},\,\tfrac{1}{2},\,\tfrac{1}{2},\,\tfrac{1}{2}),\,\tfrac{5}{2})$.

This can be seen either by a direct calculation of K-types or else by a computation of the length of the composition series of $\pi(\bar{\lambda}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}), \frac{5}{2})$. We omit the details. The proof for this case is now complete.

Assume now that $\mu+2\rho_c$ is dominant for F_2. Then unless $\mu_1=\mu_2+\mu_3+1$ and $\mu_4=0$ we either have that v is not strictly positive or else μ is also dominant for F_1 or F_3, cases we have already considered. So assume $\mu_1=\mu_2+\mu_3+1$, $\mu_4=0$. Then

$$(\bar{\lambda}(\mu), v)=(\mu_1+2+v/2, \mu_2+2-v/2, \mu_3+1-v/2, v/2).$$

Suppose first that $\mu_2 \geqq 1$. Then let $\gamma=\varepsilon_1+\varepsilon_2$ and $\mathfrak{q}$ be the corresponding parabolic. Then the Dirac inequality for μ_L is sharp and gives the value 1. By direct computation one can prove that the Dirac inequality (2.5) on $\mathfrak{g}$ has as lower bound $1 \leqq E<2$ if $\mu_2 \geqq 1$ and $3<E<4$ if $\mu_2=0$.

On the other hand because the inequality in (2.4) is equivalent to $\mu_2+2-v \geqq 0$ the result follows if $\mu_2 \geqq 1$. If $\mu_2=0$ one argues as follows. The representation in question is reducible at the point $v=2$ and at the point $v=3$ is reducible.

Then the Dirac inequality applied to the K-type $(1, 1, 1, 0)$ which is in $\bar{\pi}(\bar{\lambda}(\mu), v)$ for $1<v \leqq 3$, says that such representation is not unitary. (Again all these facts can be checked by a straight forward calculation of composition factors and multiplicity of K-types. We omit the details).

We now repeat the argument as for the previous chamber to prove Theorem 2.3 with $(\bar{\lambda}(\mu_L), v)$ a parameter for the spherical principal series. Theorem 2.3 is now completely proved.

3. Cases with Isolated Representations

In this section we investigate the representations not covered by Theorem 2.3. Excluding the case of the spherical principal series which is done in [8], we are to treat

I. $\mathfrak{g}=\mathfrak{sp}(n, 1)$, $\mu=\sum_{j=1}^{k} \mu_j \varepsilon_j$, $\mu_k>0$, $0<k<n-1$

II. $\mathfrak{g}=\mathfrak{f}_{4,1}$ $\quad \mu=\lambda(\varepsilon_1+\varepsilon_2)$, $\lambda>0$.

Proposition 3.1. *For μ as in case I or II, there is an (explicitly constructed) θ-stable parabolic subalgebra $\mathfrak{q}$ such that conclusion* (3) *of Theorem* 2.3 *holds. Moreover:*

(1′) *For $v>0$, $\bar{\pi}(\bar{\lambda}(\mu), v)$ and $\bar{\pi}(\bar{\lambda}(\mu_L), v)$ are either both unitary or both non-unitary.*

(2′) *On $[\mathfrak{l}, \mathfrak{l}]$, $(\bar{\lambda}(\mu), v)$ is the parameter of a spherical principal series.*

Proof. Case 1. Let $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ be the parabolic subalgebra determined by $\gamma=\sum_{i=1}^{k}(n-i+1)\varepsilon_i$. Then the inequality in (2.4) is satisfied for $v \leqq \mu_k+n-k-\frac{1}{2}$. On the other hand the Dirac inequality implies that $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ cannot be unitary unless

$$\nu \leqq n-k+\tfrac{1}{2} \quad \text{and} \quad n-k+\tfrac{1}{2} \leqq \mu_k+n-k-\tfrac{1}{2}$$

since $\mu_k > 0$ is an integer. Thus by Corollary 1.2, if $\bar{\pi}$ is unitary then $\bar{\pi} = \mathscr{R}^S(\bar{\pi}(\bar{\lambda}(\pi_L), \nu))$ and $\bar{\pi}_L$ is a quotient of a spherical principal series.

For $k<n-1$, $[\mathfrak{l}_0, \mathfrak{l}_0] \simeq \mathfrak{sp}(n-k, 1)$. In this case the complementary series extends to $\nu = n-k-\frac{1}{2}$. The trivial representation is the Langlands quotient at $\nu = n-k+\frac{1}{2}$ and is isolated. The same happens with $\bar{\pi}(\bar{\lambda}(\mu), \nu)$.

We now show that $\bar{\pi}(\bar{\lambda}(\mu), n-k+\frac{1}{2})$ is unitary. This was first done in [2]. It depended on knowing the multiplicities of K-types of $\bar{\pi}$ and the intertwining operator explicitly. We give a somewhat simpler proof based on the multiplicity formula in 1.4. (The idea of the proof is due to D. Vogan.)

Let δ be a K-type of $\bar{\pi}(\lambda(\mu), n-k+\frac{1}{2})$. Then

(1) $$\delta = 2\rho(\mathfrak{u} \cap \mathfrak{s}) + \sum n_i \beta_i \qquad \beta_i \in \Delta(\mathfrak{u} \cap \mathfrak{s}),\ n_i \in \mathbf{N}.$$

On the other hand, any K-type of $\pi(\bar{\lambda}(\mu), \nu)$ is of the form

(2) $$\delta = \gamma + 2\rho(\mathfrak{u} \cap \mathfrak{s}) + \sum m_i \beta_i \qquad \beta_i \in \Delta(\mathfrak{u} \cap \mathfrak{s}),\ m_i \in \mathbf{N}$$

and γ a K-type of the spherical principal series of $\mathrm{sp}(n-k, 1)$. But according to [6], γ must be of the form $\gamma = \dfrac{p+q}{2} \varepsilon_{k+1} + \dfrac{p-q}{2} \varepsilon_{k+2} + q \varepsilon_{n+1}$ where $p-q \in 2\mathbf{N}$, $q \in \mathbf{N}$ and the roots in $\Delta(\mathfrak{u} \cap \mathfrak{s})$ are of the form $\beta = \varepsilon_s \pm \varepsilon_{n+1}$ with $s \leqq k$. Then if δ is both of the form (1) and (2) then γ must be a combination of roots in $\Delta(\mathfrak{u} \cap \mathfrak{s})$ which implies $p-q=0$ and $p+q=0$, in other words $\gamma=0$.

Thus, $m(\delta, \bar{\pi}(\bar{\lambda}(\mu), n-k+\frac{1}{2}))=0$ or it is the same as $m(\delta, \pi(\bar{\lambda}(\mu), \nu))$. In addition $\bar{\pi}(\bar{\lambda}(\mu), n-k-\frac{1}{2})$ as a K-representation contains $\bar{\pi}(\bar{\lambda}(\mu), n-k+\frac{1}{2})$ since this is true on L.

The proof that $\bar{\pi}(\bar{\lambda}(\mu), n-k+\frac{1}{2})$ is unitary is now completed as in [2].

Case 2. Let $\mathfrak{q} = \mathfrak{l} + \mathfrak{u}$ be the parabolic subalgebra determined by $\gamma = \varepsilon_1 + \varepsilon_2$. Then $[\mathfrak{l}_0, \mathfrak{l}_0] \simeq \mathfrak{sp}(2, 1)$ and the inequality in (2.4) is satisfied up to $\nu \leqq \lambda + \frac{3}{2}$. The Dirac inequality implies that $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ is not unitary for $\nu > \frac{5}{2}$. Thus, the complementary series for $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ extends to $\nu = \frac{3}{2}$. Remains to see that $\bar{\pi}(\bar{\lambda}(\mu), \frac{5}{2})$ is unitary. The argument in Case I cannot be used. Instead, we use formula 1.4 for the multiplicity of a K-type in $\mathscr{R}^S \bar{\pi}_L$ to see that the K-types of $\bar{\pi}(\bar{\lambda}(\mu), \frac{5}{2})$ have a particularly simple form.

We get, for δ a K-type,

$$m(\delta, \bar{\pi}) = \sum_{w \in W_K} (-1)^{l(w)} P(w(\delta + \rho_c) - \rho_c - \lambda \varepsilon_1 - \lambda \varepsilon_2).$$

A combination of roots in $\Delta(\mathfrak{u} \cap \mathfrak{s})$ with non-negative integer coefficients is of the form $\alpha = \alpha_1 \varepsilon_1 + \alpha_1 \varepsilon_2 + \alpha_3 \varepsilon_3 + \alpha_4 \varepsilon_4$ where

$$\alpha_1 - \alpha_3 = n_1 + n_4$$

$$\alpha_1 + \alpha_2 = n_2 + n_1$$

$$\alpha_3 - \alpha_2 = n_3 - n_1.$$

Then $\delta=\mu_1\varepsilon_1+\mu_1\varepsilon_2+\mu_2\varepsilon_3+\mu_3\varepsilon_4$, $\mu_1\geqq\mu_2\geqq\mu_3\geqq 0$ and the μ_i are either all integers or half integers.

Since W_K is the group of permutations and sign changes in $\varepsilon_1,\ldots,\varepsilon_4$, $P(w(\delta+\rho_c)-\rho_c-\lambda\varepsilon_1-\lambda\varepsilon_2)\neq 0$ only if w leaves the first two coordinates fixed. Then $m(\delta,\bar{\pi})$ is the sum with the listed sign changes of the number of solutions (n_1, n_2, n_3, n_4) of the following equation

$(+)$ $(\mu_1-\lambda,\mu_1-\lambda,\mu_2,\mu_3)=\alpha$	$\alpha_1-\mu_2+1$ solutions if $\alpha_1\geqq\mu_2$, 0 otherwise
$(-)$ $(\mu_1-\lambda,\mu_1-\lambda,\mu_2,-\mu_3-1)=\alpha$	$\alpha_1-\mu_2+1$ if $\mu_2>\mu_3$, $\alpha_1\geqq\mu_2$; $\alpha_1-\mu_2$ if $\mu_2=\mu_3$, $\alpha_1\geqq\mu_2+1$ 0 otherwise
$(-)$ $(\mu_1-\lambda,\mu_1-\lambda,-\mu_2-3,\mu_3)=\alpha$	$\alpha_1-\mu_2-2$ if $\alpha_1\geqq\mu_2+3$; 0 otherwise
$(+)$ $(\mu_1-\lambda,\mu_1-\lambda,-\mu_2-3,-\mu_3-1)=\alpha$	$\alpha_1-\mu_2-2$ if $\alpha_1\geqq\mu_2+3$; 0 otherwise
$(-)$ $(\mu_1-\lambda,\mu_1-\lambda,\mu_3-1,\mu_2+1)=\alpha$	$\alpha_1-\mu_2$ if $\alpha_1\geqq\mu_2+1$; 0 otherwise
$(+)$ $(\mu_1-\lambda,\mu_1-\lambda,\mu_3-1,-\mu_2-2)=\alpha$	$\alpha_1-\mu_2-1$ if $\alpha_1\geqq\mu_2+2$; 0 otherwise
$(+)$ $(\mu_1-\lambda,\mu_1-\lambda,-\mu_3-2,\mu_2+1)=\alpha$	$\min(\alpha_1-\mu_2-1,\alpha_1-\mu_3-2)+1$ if $\alpha_1\geqq\mu_2+1$, $\alpha_1\geqq\mu_3+2$; 0 otherwise
$(-)$ $(\mu_1-\lambda,\mu_1-\lambda,-\mu_3-2,-\mu_2-2)=\alpha$	$\alpha_2-\mu_2-1$ if $\alpha_1\geqq\mu_2+2$; 0 otherwise.

We can now easily verify that $m(\delta,\bar{\pi})=0$ unless $\mu_1-\lambda=\mu_2=\mu_3$ in which case the multiplicity is 1.

Then the argument in [18] applies and we can conclude that $\bar{\pi}(\bar{\lambda}(\mu),\frac{5}{2})$ is unitary. The proof of the proposition is now complete.

Remark. It is possible to obtain the result on the multiplicity of the K-types in the same way as in [18] be computing the τ-invariant and then using the equations for the extremal K-types as in [18]. The calculations are about as lengthy as in our approach. Since this requires more notation just for this particular calculation, we have preferred to do a direct multiplicity calculation.

For the convenience of the reader we now summarize Theorem 2.3 and Proposition 3.1 in the following.

Theorem 3.2. *Let μ be a K-type and assume that μ is not the lowest K-type of a discrete series representation or a nondegenerate limit of discrete series. Then there is an (explicitly constructed) θ-stable parabolic subalgebra $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ depending on μ such that:*

(1) *For $\nu>0$, $\bar{\pi}(\bar{\lambda}(\mu),\nu)$ is unitary if and only if $\bar{\pi}(\bar{\lambda}(\mu_L),\nu)$ is unitary.*

(2) *On* $[\mathfrak{l}, \mathfrak{l}]$, $(\bar{\lambda}(\mu_L), \nu)$ *is the parameter of the spherical principal series or else* $[\mathfrak{l}, \mathfrak{l}] = \mathfrak{sp}(m, 1)$ *and* $\mu_L = a\,\varepsilon_{n+1}$ *with* $a > 0$.

(3) *For* $\nu > 0$, *if* $\bar{\pi}(\bar{\lambda}(\mu), \nu)$ *and* $\bar{\pi}(\bar{\lambda}(\mu_L), \nu)$ *are both unitary, then*

$$\mathscr{R}^i(\bar{\pi}(\bar{\lambda}(\mu_L), \nu)) = \begin{cases} 0 & \text{if } i < S \\ \bar{\pi}(\bar{\lambda}(\mu), \nu) & \text{if } i = S. \end{cases}$$

We now give a list of the unitary spectrum for each group. We have omitted the case of lowest K-types of discrete series or of a nondegenerate limit of discrete series (see 2, §2).

Table 1. Unitary parameters ν

	Complementary series	Isolated representations
$\underline{\mathfrak{g} = \mathfrak{so}(n, 1)}$ $\mu = \sum_{i=1}^{n} \mu_i \varepsilon_i$ such that $\mu_n = 0$ $t = \min\{1 \leqq j \leqq n; \mu_j = 0\}$	$\nu \leqq n - t + \frac{1}{2}$	
$\underline{\mathfrak{g} = \mathfrak{su}(n, 1)}$ $\mu = \sum_{i=1}^{n+1} \mu_i \varepsilon_i$ and $j \in \{1, \dots, n\}$ such that $\mu + 2\rho_c$ dominant for A_{j-1} but not for A_{j-2} a_1) $j \geqq 2$, $\mu_{n+1} = \mu_j + n - 2j + 2$ and $\mu_{j-1} = \mu_j$ a_2) $j \geqq 1$ and $\mu_{n+1} = \mu_j + n - 2j + 1$ $t_0 = \max\{0 \leqq i \leqq j-1; \mu_{j-i} = \mu_j\}$ $t_1 = \max\{0 \leqq i \leqq n-j; \mu_j = \mu_{j+i}\}$	$\nu \leqq \begin{cases} \min(t_0, t_1 + 1) \text{ if } a_1) \text{ holds} \\ \min(t_0, t_1) + \frac{1}{2} \text{ if } a_2) \text{ holds} \end{cases}$	
$\underline{\mathfrak{g} = \mathfrak{sp}(n, 1)}$ $\mu = \sum_{i=1}^{n+1} \mu_i \varepsilon_i$ such that $\mu + 2\rho_c$ dominant for C_{j-1} but not for C_{j-2} c_1) $j \geqq 1$ and $\mu_{n+1} = \mu_j + 2n - 2j$ c_2) $j \geqq 2$, $\mu_{n+1} = \mu_j + 2n - 2j + 1$ and $\mu_{j-1} = \mu_j$ $t_0 = \max\{0 \leqq i \leqq j-1; \mu_{j-i} = \mu_j\}$ $t_1 = \max\{0 \leqq i \leqq n-j; \mu_j = \mu_{j+i}\}$	$\nu \leqq \begin{cases} \min(t_0, t_1) + \frac{1}{2} \text{ if } c_1) \text{ holds and } \mu_j > 0 \\ t_0 + \frac{1}{2} \text{ if } c_1) \text{ holds and } \mu_j = 0,\ j \leqq n-1 \\ \frac{3}{2} \text{ if } c_1) \text{ holds, } j = n,\ \mu_{n-1} > \mu_n = 0 \\ n - k - \frac{1}{2} \text{ if } c_1) \text{ holds, } j = n, \\ \mu_k > \mu_{k+1} = \dots = \mu_{n-1} = \mu_n = 0, \\ 0 \leqq k \leqq n-2 \end{cases}$	$\nu = n - k + \frac{1}{2}$ if c_1) holds, $j = n$, $\mu_k > \mu_{k+1} = \dots = \mu_{n-1} = \mu_n = 0$, $0 \leqq k \leqq n-2$
$\underline{\mathfrak{g} = \mathfrak{f}_{4,1}}$ $\mu = \sum_{k=1}^{4} \mu_k \varepsilon_k$ such that f_1) $\mu_1 = \mu_2 + \mu_3 + \mu_4 + 2$ f_2) $\mu_1 = \mu_2 + \mu_3 + 1$, $\mu_4 = 0$ f_3) $\mu_2 + \mu_3 = \mu_4 + \mu_1$ and $\mu_1 > \mu_2$ or $\mu_1 = \mu_2 > \mu_3$ f_4) $\mu_1 = \mu_2 = \mu_3 = \mu_4$	$\nu \leqq \frac{1}{2}$ $\nu \leqq 1$ $\nu \leqq \frac{1}{2}$ if $\mu_1 < \mu_2$ $\nu \leqq \frac{3}{2}$ if $\mu_1 = \mu_2 > \mu_3$ $\nu \leqq \frac{5}{2}$	$\nu = \frac{5}{2}$ if f_3) holds and $\mu_1 = \mu_2 \geqq 0$, $\mu_3 = \mu_4 = 0$ $\nu = \frac{11}{2}$ if $\mu_1 = 0$

As a consequence of the results proved up to now and of the following lemma we are able to show in Corollary 3.5 that Zuckerman's conjecture is true for real rank one groups.

Lemma 3.4. *Let* $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ *be a* θ*-stable parabolic subalgebra and* π_L *an irreducible, finite dimensional representation of* L *with lowest* $(L\cap K)$ *type* μ_L *such that* $(\mu_L,\alpha)=0$ *for* $\alpha\in\Delta(\mathfrak{l})$ *and* $(\mu_L,\alpha)\geqq 0$ $\alpha\in\Delta(\mathfrak{u})$. *Assume that the Langlands parameter* $(\bar{\lambda}(\mu_L),\nu)$ *is such that*

$$\langle(\bar{\lambda}(\mu_L),\nu)+\rho(\mathfrak{w},\alpha\rangle>0 \quad \textit{for } \alpha\in\Delta(\mathfrak{u}).$$

Then $\mu=\mu_L+2\rho(\mathfrak{u}\cap\mathfrak{s})$ *is a* K*-type and*

$$\mathscr{R}^i\pi_L=\begin{cases}0 & i<S\\ \bar{\pi}(\bar{\lambda}(\mu),\nu) & i=S.\end{cases}$$

Proof. This is Corollary 1.2 once we prove that μ is a K-type, $\bar{\lambda}(\mu)=\bar{\lambda}(\mu_L)+\rho(\mathfrak{u})$ and $\mathfrak{q}\supseteq\mathfrak{b}^0$.

The fact that μ is a K-type, $\bar{\lambda}(\mu)=\bar{\lambda}(\mu_L)+\rho(\mathfrak{u})$ and $\mathfrak{q}\supseteq\mathfrak{b}^0$ are proved in [17] page 395 under the condition that $(\bar{\lambda}(\mu_L)+\rho(\mathfrak{u}),\alpha)>0$ for $\alpha\in\Delta(\mathfrak{u})$. To check this we argue as follows. Consider the Cayley transform that maps the simple noncompact root β into the real root. This maps $(\bar{\lambda}(\mu_L),\nu)$ onto a parameter on the maximally split Cartan subgroup $\mathfrak{g}=\mathfrak{t}^-+\mathfrak{a}$. By abuse of notation we call it $(\bar{\lambda}(\mu_L),\nu)$ again. The Cayley transform also leaves $\Delta(\mathfrak{u})$ invariant since $\beta\in\Delta(\mathfrak{l})$. Then

$$((\bar{\lambda}(\mu_L),\nu)+\rho(\mathfrak{u}),\alpha)>0$$

and

$$((\bar{\lambda}(\mu_L),\nu)+\rho(\mathfrak{u}),\theta\alpha)>0 \quad \text{for } \alpha\in\Delta(\mathfrak{u}).$$

But $\theta((\bar{\lambda}(\mu),\nu))=(\bar{\lambda}(\mu),-\nu)$ and $\theta\rho(\mathfrak{u})=\rho(\mathfrak{u})$ so

$$((\bar{\lambda}(\mu_L),0)+\rho(\mathfrak{u}),\alpha)=\tfrac{1}{2}((\bar{\lambda}(\mu_L),\nu)+\rho(\mathfrak{u})+\theta((\bar{\lambda}(\mu_L),\nu)+\rho(\mathfrak{u})),\alpha)>0$$

and the proof is complete.

Corollary 3.5. *Let* $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ *be a* θ*-stable parabolic subalgebra. Let* $\psi(\mathfrak{l})$ *be a positive chamber for* $\mathfrak{l}$ *such that* $2\rho(\mathfrak{l}\cap\mathfrak{k})$ *is dominant with respect to it and* $\psi=\psi(\mathfrak{l})\cup\Delta(\mathfrak{u})$. *Let* λ: $\mathfrak{q}\to\mathbf{C}$ *be a character such that*

$$\langle\lambda+\rho(\psi),\alpha\rangle>0 \quad \textit{for } \alpha\in\psi.$$

Let π^L be the representation of L corresponding to λ. Then

$$\mathscr{R}^i\pi^L=\begin{cases}0 & i\neq S\\ \pi & i=S\end{cases}$$

and π is an irreducible unitary representation.

Conversely, for any irreducible unitary representation π with non-singular integral infinitesimal character, there is $\mathfrak{q}$ and λ as above such that $\pi=\mathscr{R}^S\pi^L$.

Proof. Assume first that $\mathfrak{l}_0$ does not contain any factor of type $\mathfrak{sp}(m,1)$. If $\mathfrak{l}_0$ also does not contain any compact factor, then the result follows from the proof of Theorem 2.3. If $\mathfrak{l}_0$ does contain compact factors, one has to use induction by stages ([17] 6.3.6). We omit the details.

If $\mathfrak{l}_0$ contains a factor of type $\mathfrak{sp}(m,1)$, the result is contained in the proof of Theorem 3.1.

4. The L^2 Index and Unitarity

In this section we determine the representations that contribute to the index of the Dirac operator. We briefly review what is involved. Let $\Gamma \subseteq G$ be a discrete torsion free subgroup such that vol $(\Gamma\backslash G)<\infty$. Let η be a K-type with representation space V_η. Let s^+ and s^- be the two spin representations. We fix a positive chamber Δ^+ ($\tilde{B}_1$, in case $\mathfrak{so}(2n,1)$, $\tilde{A}_n$ in case $\mathfrak{su}(n,1)$, $\tilde{C}_n$ in case $\mathfrak{sp}(n,1)$, F_3F_4, in case $\mathfrak{f}_{4,1}$) such that $2\rho_c$ is dominant and normalize s^+ and s^- such that on T', the set of regular elements in T, the characters of $s^\pm$ satisfy

$$(\operatorname{ch} s^+ - \operatorname{ch} s^-)|_{T'} = \prod_{\alpha\in\Delta^+-\Delta(\mathfrak{k})} (e^{\alpha/2}-e^{-\alpha/2}).$$

Then there are two bundles $\varepsilon^\pm$ over $X=\Gamma\backslash G/K$ with fibers $E^\pm = V_\eta \otimes s^\pm$ and an elliptic operator D: $\varepsilon^+ \to \varepsilon^-$ called the Dirac operator. Then (cf. [12])

$$\operatorname{index} D = \sum_{\pi\in L^2_d(\Gamma\backslash G)} m_\Gamma(\pi)\{\dim \operatorname{Hom}_K(\pi, E^+) - \dim \operatorname{Hom}_K(\pi, E^-)\}$$

where $L^2(\Gamma\backslash G)$ is the discrete spectrum of $L^2(\Gamma\backslash G)$. In this section we calculate

$$m(\pi,\eta) = \dim \operatorname{Hom}_K(\pi, E^+) - \dim \operatorname{Hom}_K(\pi, E^-)$$

for all η and all π unitary. We need the following facts from [1].

1. $m(\pi,\eta) \neq 0$ only if the infinitesimal character of π, denoted by χ_π, is $\eta+\rho_c$.

2. Let $\operatorname{ch}\pi$ be the distribution character of π and for η a K-type, we denote by $\operatorname{ch}\eta$ its character. Then if $\chi_\pi = \chi_{\eta+\rho_c}$

$$\operatorname{ch}\pi|_{T'}\cdot\Delta = \sum_{w\in W_c} a_w e^{w(\eta+\rho_c)}$$

where $\Delta = \prod_{\alpha\in\Delta^+} (e^{\alpha/2}-e^{-\alpha/2})$.

Then

$$\operatorname{ch}\pi(\operatorname{ch} s^+ - \operatorname{ch} s^-)|_{T'} = \sum a_{\eta_i} \operatorname{ch}\eta_i|_{T'}$$

and

$$m(\pi,\eta) = (-1)^p a_\eta, \qquad p=\tfrac{1}{2}(\dim\mathfrak{g}-\dim\mathfrak{k}).$$

We will consider two cases. Let $\psi(\eta+\rho_c)$ be a positive system such that $\eta+\rho_c$ is dominant.

Case 1. $\lambda=\eta+\rho_c-\rho(\psi)$ is dominant with respect to ψ.

Case 2. $\lambda=\eta+\rho_c-\rho(\psi)$ is not dominant with respect to ψ.

We treat Case 1 first. It can be reduced to the computations in [19]. Let F_λ be the finite dimensional representation of $\mathfrak{g}$ with highest weight λ. Let F_λ^* be its dual and $H^*(\mathfrak{g}, K, \pi\otimes F_\lambda^*)$ be Lie algebra cohomology as defined in [4].

Lemma 4.1. *If* $m(\pi,\eta)\neq 0$ *and* π *is unitary irreducible then* $H^*(\mathfrak{g}, K, \pi\otimes F_\lambda^*)\neq 0$.

Proof. Since π is unitary, a standard argument is [4] shows that $H^*(\mathfrak{g}, K, \pi\otimes F_\lambda^*)\neq 0$ is equivalent to

a) $\chi_\pi=\chi_{F_\lambda}$

b) $\mathrm{Hom}_K(\pi\otimes F_\lambda^*, \Lambda^*\mathfrak{s})\neq 0$.

We check that these two properties are satisfied.

a) is clear from property 1 listed earlier. For b) we write

$$\Lambda^*\mathfrak{s}=(s^+\oplus s^-)\otimes(s^+\oplus s^-)=s\otimes s.$$

Then

$$\mathrm{Hom}_K(\pi\otimes F_\lambda^*, \Lambda^*\mathfrak{s})\cong\mathrm{Hom}_K(\pi, F_\lambda\otimes s\otimes s)=\mathrm{Hom}_K(\pi\otimes s^*, F_\lambda\otimes s).$$

Then $\mathrm{Hom}_K(\pi\otimes s^*, F_\lambda\otimes s)\neq 0$ by the property that $\eta\in\pi\otimes s^*|_K$.

We now recall the following proposition. For a character $\lambda\colon \mathfrak{l}\to C$ we denote by π_λ^L the corresponding representation.

Proposition 4.2 ([19] Sects. 5–7). Let π be unitary and F an irreducible finite dimensional representation of $\mathfrak{g}$. Suppose $H^*(\mathfrak{g}, K, \pi\otimes F)\neq 0$. Then there is a θ-stable parabolic subalgebra $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ and a one dimensional representation π_λ^L of L such that $\pi\cong\mathscr{R}^S\pi_\lambda^L$. We fix Δ_L^+ a positive system for which $2\rho(\mathfrak{l}\cap\mathfrak{k})$ is dominant. Then if $\psi=\Delta_L^+\cup\Delta(\mathfrak{u})$, λ must be the highest weight for F with respect to ψ.

We now compute $m(\pi,\eta)$ for such a representation. For any irreducible representation π^L of L we denote by $e(\pi^L)$ the virtual character

$$e(\pi^L)=\sum_i(-1)^i\,\mathrm{c}(\mathscr{R}^i\pi^L)$$

and refer to it as the Euler characteristic. It is a well defined map from the ring of virtual characters of L to the ring of virtual characters of G. Furthermore it commutes with coherent continuation.

Suppose

$$\mathrm{ch}\,\pi_\lambda^L=\sum C_j\,\mathrm{ch}\,\pi^L(\bar{\lambda}(\mu_j), \nu_j)=\sum C_j\,\mathrm{ch}\,\pi^L(\gamma_j)$$

with C_j integers and $\pi^L(\gamma_j)$, discrete series or principal series. Then

$$e(\pi_\lambda^L)=\sum C_j(\pi^L(\gamma_j)).$$

We note that $\gamma_j+\rho(\mathfrak{u})=w(\lambda+\rho)$ with $w\in W(\mathfrak{l})$. Since $\langle\lambda+\rho,\alpha\rangle>0$ for $\alpha\in\Delta(\mathfrak{u})$, $\langle\gamma_j+\rho(\mathfrak{u}),\alpha\rangle>0$. It follows that

$$e(\pi_\lambda^L)=(-1)^S\,\mathrm{ch}\,\mathscr{R}^S\pi_\lambda^L$$

and by using 8.2.15 in [17],

$$e(\pi^L(\gamma_j)) = (-1)^S \operatorname{ch} \mathscr{R}^S \pi^L(\gamma_j) = (-1)^S \operatorname{ch} \pi(\gamma_j + \rho(\mathfrak{u})).$$

Thus

$$\operatorname{ch} \pi = \sum C_j \operatorname{ch} \pi(\gamma_j + \rho(\mathfrak{u})).$$

Since we only need $\operatorname{ch}\pi|_{T'}$ we are only interested in the C_j's for which $\pi^L(\gamma_j)$ is a discrete series.

By comparing the expressions of such characters we can write for $\rho = \rho(\psi)$,

$$\operatorname{ch} \pi^L_{\lambda|T'} = (-1)^{P_L} \sum_{\psi_L \supseteq \Delta^+(\mathfrak{l} \cap \mathfrak{k})} \operatorname{ch} \pi(\psi_L, \lambda + \rho(\mathfrak{l}))|_{T'}.$$

Then

$$\operatorname{ch} \pi|_{T'} = (-1)^{P_L} \sum_{\substack{\psi \supseteq \Delta(\mathfrak{u}) \\ \psi \supseteq \Delta^+(\mathfrak{k})}} \operatorname{ch} \pi(\psi, \lambda + \rho)|_{T'},$$

so writing out the expression for $\operatorname{ch}\pi(\psi, \lambda+\rho)$ we get

$$m(\pi, \eta) = (-1)^{P_L} \varepsilon(\psi)$$

where ψ is the chamber for which $\eta + \rho_c$ is dominant and $\varepsilon(\psi) = (-1)^{|-\psi \cap \Delta^+|}$. Furthermore $\eta + \rho_c = w(\lambda + \rho_c)$ for a unique $w \in W(\mathfrak{l})$.

Case 2. In this case, $\eta + \rho_c$ must be singular. Since $\eta + \rho_c$ is regular with respect to $\Delta(\mathfrak{k})$, there cannot be two adjacent noncompact roots in the Dynkin diagram for the simple roots of ψ such that $\eta + \rho_c$ is singular with respect to both of them. Thus in real rank 1, $\eta + \rho_c$ is singular with respect to exactly one simple noncompact root in ψ say β. Then χ_π coincides with the infinitesimal character of a limit of discrete series.

We also note that if $w(\eta + \rho_c)$ is dominant with respect to $\Delta^+(\mathfrak{k})$ then $w(\eta + \rho_c)$ must be singular with respect to a compact root unless $w(\eta + \rho_c) = \eta + \rho_c$ and $w = \mathrm{id}$ or the simple reflection about β. Thus, if $\bar{\lambda}(\mu, \nu) = \gamma$ is the parameter of π, $\eta + \rho_c$ is the unique element in the W-orbit of γ which is dominant, regular with respect to $\Delta^+(\mathfrak{k})$ and

$$\operatorname{ch} \pi|_{T'} = (-1)^P m(\eta + \rho_c, \pi) \Delta^{-1} \Big(\sum_{w \in W_K} \varepsilon(w) e^{w(\eta + \rho_c)} \Big).$$

To compute $m(\eta + \rho_c, \pi)$ we argue as follows. If $\pi = \mathscr{R}^S \pi^L_\lambda$ for $\mathfrak{q} = \mathfrak{l} + \mathfrak{u}$ then λ must satisfy $(\lambda, \alpha) = 0$ for $\alpha \in \Delta(\mathfrak{l})$, $(\lambda + \rho, \alpha) \geqq 0$ for $\alpha \in \Delta(\mathfrak{u})$ and equality for at least one $\alpha \in \Delta(\mathfrak{u})$. In this case $m(\pi, \eta)$ is computed in the same way as in Case 1. We note that there is a unique $\tilde{w} \in W(\mathfrak{l})$ such that $\tilde{w}(\lambda + \rho) = \eta + \rho_c$. The existence of $\tilde{w}$ is clear by the calculation in Case 1. If there were w_1, $w_2 \in W(\mathfrak{l})$ then $w_1(\lambda + \rho) = w_2(\lambda + \rho)$ implies $\lambda + \rho(\mathfrak{u}) + w_1 \rho(\mathfrak{l}) = \lambda + \rho(\mathfrak{u}) + w_2 \rho(\mathfrak{l})$ so $w_1 = w_2$. Choose a chamber $\psi(\lambda + \rho)$ for which $\lambda + \rho$ is dominant. Then let $\psi = \tilde{w} \psi(\lambda + \rho)$. We have

$$m(\pi, \eta) = (-1)^{P_L} \varepsilon(\psi).$$

However, there are cases when $\pi \neq \mathscr{R}^S \pi_\lambda^L$. These are the cases $\mathfrak{g} = \mathfrak{sp}(n,1)$ and the representations with lowest K-type

$$\mu = \sum_{k=1}^{j-1} \mu_k \varepsilon_k + \mu_{n+1} \varepsilon_{n+1} \text{ where either } c_1\colon \mu_{n+1} = 2n-2j \text{ and}$$

$$(\bar{\lambda}(\mu), \nu) = \sum_{k=1}^{j-1} (\mu_k + n - k)\, \varepsilon_K + (n - j + \tfrac{1}{2} - \nu)\, \varepsilon_j$$

$$+ \sum_{k=j+1}^{n} (n-k+1)\, \varepsilon_k + (n-j+\tfrac{1}{2}+\nu)\, \varepsilon_{n+1},\ \nu = t_0 + \tfrac{1}{2},\ 1 \leqq j \leqq n-1$$

or c_2: $\mu_{n+1} = 2n-2j+1$ and

$$(\bar{\lambda}(\mu), \nu) = \sum_{k=1}^{j-1} (\mu_k + n - k)\, \varepsilon_k + (n-j+1\ \nu)_j$$

$$+ \sum_{k=j+1}^{n} (n-k+1)\, \varepsilon_k + (n-j+1+\nu)_{n+1}$$

$\nu = t_0$, $\mu_{j-1} = 0$ and $2 \leqq j \leqq n$.

In this case we have to calculate the character of π using the results in [3]. We sketch how to do the calculation for case c_1. In [3] the Langlands parameter is given by a $\gamma = \sum_{i=1}^{n+1} a_i e_i$ (cf. page 448 in [3], e_i is a certain basis of the split Cartan subalgebra) where $a_1 > a_2$, $a_1 \geqq -a_2$, $a_3 > a_4 > \ldots > a_{n+1} > 0$ $a_3, \ldots, a_n \in \mathbf{Z}$. The relation to $(\bar{\lambda}(\mu), \nu)$ is given by $a_1 = \nu + (n - j + \frac{1}{2})$, $a_2 = \nu - (n - j + \frac{1}{2})$ $a_K = \mu_{k-2} + n - k + 2$ for $3 \leqq k \leqq j+1$, $a_K = n - k + 2$ for $j+1 < k \leqq n+1$. We set $\nu = t_0 + \frac{1}{2}$, the first reducibility point. Write $\pi(\gamma)$ and $\bar{\pi}(\gamma)$ for the principal series determined by γ and the corresponding Langlands quotient.

Lemma 4.3. *Suppose that, for $\gamma = \sum_{i=1}^{n+1} a_i e_i$, there are $\tilde{i}$, $\tilde{j}$ such that*

$$a_2 = a_{\tilde{i}}, \quad a_{\tilde{j}+1} > a_1 > a_{\tilde{j}+2}, \quad 3 \leqq \tilde{i}, \quad \tilde{j} \leqq n+1.$$

Then

$$\operatorname{ch} \bar{\pi}(\gamma)|_{T'} = (-1)^{\tilde{i}+\tilde{j}+1} \operatorname{ch} \pi(C_{\tilde{i}-1}, \tilde{\gamma})|_{T'}$$

($C_{\tilde{i}-1}$ the chamber defined in Sect. 1 and $\bar{\pi}(C_{\tilde{i}-1}, \tilde{\gamma})$ the corresponding limit of discrete series) and

$$\tilde{\gamma} = \sum_{k=1}^{\tilde{j}-1} a_{k+2}\, \varepsilon_k + a_1\, \varepsilon_{\tilde{j}} + \sum_{k=\tilde{j}+1}^{n} a_{k+1}\, \varepsilon_k + a_{\tilde{i}}\, \varepsilon_{n+1}.$$

Proof. The proof is just a straightforward application of Theorem 4.2 in [3] and we refer to it for more details. What is involved is the following. $\gamma_0 = \gamma$ is of type V according to the quoted result. This just means that γ_0 is dominant with respect to a certain chamber with singularity determined by the coefficient of e_2. In any case due to the assumption we have a character identity of the form

$$\pi(\gamma_0)=\bar{\pi}(\gamma_0)+\bar{\pi}(\gamma_1) \quad \text{where } \gamma_1=s_{e_1-e_{j+2}}(\gamma_0).$$

Now we can write a relation of the same sort for the character of $\bar{\pi}(\gamma_1)$. Proceeding in this way we will be able to define parameters $\{\gamma_i\}_{i=0}^{k-1}$ such that

$$\pi(\gamma_i)=\bar{\pi}(\gamma_i)+\bar{\pi}(\gamma_{i+1})$$

and γ_k has a singularity determined (most of the time) by the coefficient of e_1. Because γ_k is of a particular sort,

$$\pi(\gamma_k)=\bar{\pi}(\gamma_k)+(\text{limit of discrete series}).$$

Solving all the equations for $\bar{\pi}(\gamma_0)$ and restricting the character to T' will give the required identity. For the reader willing to go through the details we define the parameters involved.

For $0\leqq k\leqq 2n-\tilde{i}-\tilde{j}+1$, let γ_k be defined by

$$\begin{aligned}
&\gamma_0=\gamma,\\
&\gamma_k=s_{e_1-e_{\tilde{j}+k+1}}\gamma_{k-1}, \quad 1\leqq k\leqq \tilde{i}-\tilde{j}-2\\
&\gamma_{\tilde{i}-\tilde{j}-1}=s_{e_1-e_{\tilde{i}}}s_{e_2-e_{i+1}}\gamma_{i-j-2}\\
&\gamma_k=s_{e_2-e_{k+\tilde{j}+2}}\gamma_{k-1}, \quad \tilde{i}-\tilde{j}\leqq k\leqq n-\tilde{j}-1\\
&\gamma_{n-\tilde{j}}=s_{2e_2}\gamma_{n-j-1}, \quad k=n-\tilde{j}\leqq 2n-\tilde{i}-\tilde{j}\\
&\gamma_k=s_{e_2+e_{2n-k-\tilde{j}+2}}\gamma_{k-1}, \quad n-\tilde{j}+1\leqq k\leqq 2n-\tilde{i}-\tilde{j}\\
&\bar{\gamma}=\gamma_{2n-\tilde{i}-\tilde{j}+1}, \quad k=2n-\tilde{i}-\tilde{j}+1
\end{aligned}$$

then the required identity is

$$\bar{\pi}(\gamma)=\sum_{k=0}^{2n-\tilde{i}-\tilde{j}}(-1)^k\pi(\gamma_k)+(-1)^{\tilde{i}+\tilde{j}}\pi^+(C_{\tilde{i}-\tilde{2}},\bar{\gamma}).$$

Lemma 4.4. *Suppose that for* $\gamma=\sum_{i=1}^{n+1}a_i e_i$ *there are* $\tilde{i}$, $\tilde{j}$ *such that* $-a_2=a_{\tilde{i}}$, $a_{\tilde{j}+1}>a_1>a_{\tilde{j}+2}$, $3\leqq\tilde{i},\tilde{j}\leqq n+1$.

Then

$$\operatorname{ch}\bar{\pi}(\gamma)|_{T'}=(-1)^{\tilde{i}+\tilde{j}+1}\operatorname{ch}\bar{\pi}(C_{\tilde{i}-2},\bar{\gamma})|_{T'}$$

where

$$\bar{\gamma}=\sum_{k=1}^{\tilde{j}-1}a_{k+2}\varepsilon_k+a_1\varepsilon_{\tilde{j}}+\sum_{k=j+1}^{n}a_{k+1}\varepsilon_k+a_{\tilde{i}}\varepsilon_{n+1}.$$

Proof. This is similar to Lemma 4.3. We omit the proof.

Similarly we get

Lemma 4.5. *Suppose* $\gamma=\sum_{k=1}^{n+1}a_k\varepsilon_k$ *satisfies* $a_2=0$ *and there is* $3\leqq\tilde{j}\leqq n+1$ *then*

$$\operatorname{ch}\bar{\pi}(\gamma)|_{T'}=0.$$

We now summarize these results in

Proposition 4.6. *As character identity we have: in case* c_1

$$\bar{\pi}(\bar{\lambda}(\mu), t_0+\tfrac{1}{2})|_{T'} = \begin{cases} -\bar{\pi}(C_{2n-2j-t_0+1}, \bar{\gamma})|_{T'}, m(\bar{\pi},\eta)=(-1)^{t_0} & t_0>n-j \\ -\bar{\pi}(C_{j+t_0}, \bar{\gamma})|_{T'}, m(\bar{\pi},\eta)=(-1)^{j+t_0+1} & t_0<n-j \\ 0 & t_0=n-j. \end{cases}$$

In case c_2,

$$\bar{\pi}(\bar{\lambda}(\mu), t_1)|_{T'} = \begin{cases} \bar{\pi}(C_{2n-j-t_0+2}, \bar{\gamma})|_{T'}, m(\bar{\pi},\eta)=(-1)^{t_0} & t_0>n-j+1 \\ \bar{\pi}(C_{j+t_0-1}, \bar{\gamma})|_{T'}, m(\bar{\pi},\eta)=(-1)^{j+t_0+1} & t_0<n-j+1 \\ 0 & t_0=n-j+1. \end{cases}$$

$\bar{\gamma}$ *is obtained from* $(\bar{\lambda}(\mu), v)$ *by conjugating by* W *so that it becomes dominant with respect to the corresponding chamber and* $\eta=\bar{\gamma}-\rho_c$.

Proof. Let's consider $(\bar{\lambda}(\mu), t_0+1/2)$.

If $t_0>n-j$ then $(\bar{\lambda}(\mu), t_0+1/2)$ satisfies the hypothesis of Lemma 4.3 with $\tilde{\imath}=2n-j-t_0+2$ and $\tilde{j}-j-t_0$. Therefore

$$\operatorname{ch}\bar{\pi}(\bar{\lambda}(\mu), t_0+1/2)|_{T'} = -(-1)^{2n-j-t_0+2+j-t_0} \operatorname{ch}\bar{\pi}(C_{2n-j-t_0+1}, \bar{\gamma})|_{T'}.$$

If $t_0<n-j$ then we are in the situation described by Lemma 4.4 with $\tilde{\imath}-j+t_0+2, \tilde{j}=\tilde{j}-t_0$. Therefore

$$\bar{\pi}(\bar{\lambda}(\mu), t_0+1/2)|_{T'} = -\bar{\pi}(C_{j+t_0}, \bar{\gamma})|_{T'}.$$

Similarly for $(\bar{\lambda}(\mu), t_0)$.

We summarize the results in the following theorems.

Theorem 4.7. *Suppose* $\bar{\pi}=\bar{\pi}(\bar{\lambda}(\mu), v)$ *is unitary such that* $\gamma=(\bar{\lambda}(\mu), v)$ *is regular. Then the* η*'s such that* $m(\bar{\pi},\eta)\neq 0$ *are given by the following procedure.*

Choose $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ as in Theorem 2.1 so that $\bar{\pi}=\mathscr{R}^S\pi_\lambda^L$ with $\lambda=\gamma-\rho(\mathfrak{u})$. Then $\eta+\rho_c=w\cdot\gamma$ with $w\in W(\mathfrak{l})$ and let ψ be the chamber for which $\eta+\rho_c$ is dominant. Then $m(\pi,\eta)=(-1)^{P_L}\varepsilon(\psi)$.

Conversely given η with $\eta+\rho_c$ regular, the unitary representations for which $m(\pi,\eta)\neq 0$ are given by the following.

Let $\lambda=\eta+\rho_c-\rho(\psi)$ and take $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ a θ-stable parabolic subalgebra such that $(\lambda,\alpha)=0$ for $\alpha\in\Delta(\mathfrak{u})$ and $(\lambda,\alpha)\geqq 0$ for $\alpha\in\Delta(\mathfrak{l})$. Then $\pi=\mathscr{R}^S\pi_\lambda^L$ (by 2.10 all such representations are unitary) and $m(\pi,\eta)=(-1)^{P_L}\varepsilon(\psi)$.

Theorem 4.8. *Suppose* $\bar{\pi}=\bar{\pi}(\bar{\lambda}(\mu), v)$ *is unitary and* $\gamma=(\bar{\lambda}(\mu), v)$ *is singular. If* $m(\bar{\pi},\eta)\neq 0$ *then there is a unique root such that* γ *is singular with respect to it, and a unique* K*-type* η *such that* $w\gamma=\eta+\rho_c$. $m(\bar{\pi},\eta)$ *is obtained as follows.*

A. $\mathfrak{g}=\mathfrak{sp}(n,1)$. The answer is given by Proposition 4.6.

B. In the other cases $\bar{\pi}=\mathscr{R}^S\pi_\lambda^L$ for $\mathfrak{q}=\mathfrak{l}+\mathfrak{u}$ as in Theorem 3. Let $\psi(\gamma)$ be a chamber for which γ is dominant. Then $m(\bar{\pi},\eta)=(-1)^{P_L}\varepsilon(\psi)$

where ψ is the unique chamber such that $\eta+\rho_c$ is dominant for ψ and there is $w\in W(\mathfrak{l})$ such that $\psi=w\psi(\gamma)$.

Conversely, given η, the unitary representations such that $m(\pi,\eta)\neq 0$ are given by

A. $\mathfrak{g}=\mathfrak{sp}(n,1)$, the cases in Proposition 4.6.

B. $\pi=\mathscr{R}^S\pi_\lambda^L$ such that $(\lambda+\rho,\alpha)\geqq 0$ for $\alpha\in\Delta(\mathfrak{u})$, $(\lambda,\alpha)=0$ for $\alpha\in\Delta(\mathfrak{l})$ and such that $\lambda+\rho$ is singular with respect to exactly one root.

References

1. Atiyah, M.F., Schmid, W.: A geometric construction of the discrete series. Invent. Math. **42**, 1-62 (1977)
2. Baldoni Silva, M.W.: The unitary dual of $Sp(n,1)$, $n\geqq 2$. Duke Math. Journal **48**, 549-583 (1981)
3. Baldoni Silva, M.W., Kraljevic, H.: Composition factors of the principal series representations of the group $Sp(n,1)$. TAMS **262**, 447-471 (1980)
4. Borel, A., Wallach, N.: Continuous cohomology, Discrete subgroups and representations of reductive groups. Ann. of Math. Studies, 94, Princeton University Press 1980
5. Hirai, T.: On irreducible representations of the Lorentz group of n-th order. Proc. Japan Acad. **38**, 83-87 (1962)
6. Johnson, K., Wallach, N.: Composition series and intertwining operators for the spherical principal series I. TAMS **229**, 137-173 (1977)
7. Kraljevic, H.: Representations of the universal covering group of the group $SU(n,1)$. Glasnik Mat. **8**, 23-72 (1973)
8. Kostant, B.: On the existence and irreducibility of certain series of representations. Bull. A.M.S. **75**, 627—642 (1969); Lie groups and their representations (Summer School of the Bolyal Jânos Mathematical Society), Halsted Press, New York 1975, pp. 231-329
9. Knapp, A., Stein, E.M.: Intertwining operators for semisimple groups. Ann. of Math. **93**, 489-578 (1971)
10. Knapp, A., Zuckerman, G.: Classification theorems for representations of semisimple Lie groups. Lecture Notes in Mathematics, Vol. 587, pp. 138-159. Berlin-Heidelberg-New York: Springer
11. Langlands, R.: On the classification of irreducible representations of real algebraic groups. Mimeographed notes, Inst. for Advanced Study, 1973
12. Moscovici, H.: L^2-index of elliptic operators on locally symmetric spaces of finite volume. Preprint
13. Miatello, R.: Alternating sum formulas for multiplicities in $L^2(\Gamma\backslash G)$ I and II. manuscript
14. Speh, B., Vogan, D.: Reducibility of generalized principal series representations. Acta Math. **145**, 227-299 (1980)
15. Thieleker, E.: On the quasi-simple irreducible representations of the Lorentz groups. TAMS **179**, 465-505 (1973)
16. Vogan, D.: The algebraic structure of the representation of semisimple Lie groups I. Ann. of Math. **109**, 1-60 (1979)
17. Vogan, D.: Representations of real reductive Lie groups. Progress in Mathematics, Birkhauser, 1981
18. Vogan, D.: Singular unitary representations. Lecture Notes in Math, vol. 880, pp. 506-535. Berlin-Heidelberg-New York: Springer 1980
19. Vogan, D.: Unitary representations with cohomology. Preprint

Oblatum 26-VI-1982

Invent. math. 72, 57–75 (1983)

Inventiones mathematicae

Characters of Irreducible Representations of the Lie Algebra of Vector Fields on the Circle

Alvany Rocha-Caridi*† and Nolan R. Wallach**

Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA

§ 0. Introduction

Let $\underline{n}$ denote the Lie algebra of vector fields on the circle whose nonpositive Fourier coefficients vanish. In [2] Gelfand conjectured that $H^k(\underline{n}, \mathbb{C})$ is 2-dimensional for each $k \geqq 1$. Goncharova in [3] proved the conjecture and, using the observation that $e_0 = \frac{1}{i}\frac{d}{d\theta}$ normalizes $\underline{n}$ actually calculated the action of e_0 in $H^k(\underline{n}, \mathbb{C})$. She showed that if $k \geqq 1$ then $H^k(\underline{n}, \mathbb{C})$ splits into two one-dimensional eigenspaces for e_0 with eigenvalues $-\frac{1}{2}(3k^2+k)$ and $-\frac{1}{2}(3k^2-k)$. The numbers $\frac{1}{2}(3k^2 \pm k)$, $k \in \mathbb{Z}_+$, are called Euler's pentagonal numbers.

Let $\underline{g}$ denote the Lie algebra of vector fields on the circle with finite Fourier series. Let $\underline{n}^-$ denote the space of complex conjugates of $\underline{n}$ and let $\underline{h} = \mathbb{C} e_0$. Then $\underline{g} = \underline{n}^- \oplus \underline{h} \oplus \underline{n}$, a triangular decomposition.

In this paper we consider irreducible highest weight modules for $\underline{g}$, $L(k)$, with integral highest weight k relative to e_0. As it turns out using Kac's formula for the determinant of the analogue of the Shapovalov form ([5]) $L(k)$ is a full Verma module unless $-k$ is a pentagonal number. Clearly $L(0) = \mathbb{C}$. Kac gave conjectural formulas (see [6] for a recent announcement of these conjectures) for the "characters" of the $L(k)$ for $-k$ pentagonal. In this paper we prove this conjecture in a strengthened form which also generalizes (although uses) Goncharova's result for $L(0)$. We show that for $p \geqq 1$, $H^p(\underline{n}, L(-\frac{1}{2}(3k^2 \pm k)))$ splits into two one-dimensional eigenspaces for e_0 with eigenvalues $-\frac{1}{2}(3(k+p)^2+(k+p))$ and $-\frac{1}{2}(3(k+p)^2-(k+p))$.

It is of some interest that the theory of these highest weight modules is quite similar to that of highest weight modules over Kac-Moody Lie algebras of rank 2. Our method involves proving analogues of our results for rank 2

* Partially supported by NSF MCS-8201260

** Partially supported by NSF MCS79-03153

† *Current Address:* School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA

Kac-Moody Lie algebras, for this algebra. Although the flow of arguments is quite similar to that of [9] the details are different and quite a bit more difficult. In particular, the analogues of the root hyperplanes for this algebra are quadratic surfaces in $\mathbb{C}^2$. We are forced in our proof of the analogue of Jantzen's character sum formula for the quotient of two Verma modules to use specific rational parametrizations of those varieties. This forces us to do a "local" version of Jantzen's filtrations which may be of independent interest. It is also of interest that although our goal is to obtain results for $\underline{g}$ we work on a central extension of $\underline{g}$, the Virasoro algebra. A critical tool in our analysis is Kac's formula for the determinant of the analogue of the Shapovalov form for the Virasoro algebra.

We now describe in detail the main results of this paper.

The algebra $\underline{g}$ of vector fields on the circle with finite Fourier series can also be described as the complex Lie algebra with basis $\{e_i\}_{i\in\mathbb{Z}}$, where

$$[e_i, e_j]=(j-i)\,e_{i+j}, \qquad i,j\in\mathbb{Z}.$$

Here we set $e_k=\frac{1}{i}e^{ik\theta}\frac{d}{d\theta}$, $k\in\mathbb{Z}$. $\underline{g}$ is also known as the Witt algebra.

We set $\underline{h}=\mathbb{C}\,e_0$, $\underline{n}=\bigoplus_{i\in\mathbb{N}}\mathbb{C}\,e_i$, and identify $\underline{h}^*$ with $\mathbb{C}$. If $\mu\in\mathbb{C}$ we denote by $M(\mu)$ and $L(\mu)$ the Verma module associated with $\underline{g}$, $\underline{h}$, $\underline{n}$ and μ, and its unique irreducible quotient, respectively. Relative to the action of $\underline{h}$, $M(\mu)$ decomposes as a direct sum of weightspaces:

$$M(\mu)=\bigoplus M(\mu)_{\mu-m} \qquad (m\in\mathbb{Z}_+)$$

where $M(\mu)_c=\{v\in M(\mu)\,|\,e_0\,v=c\,v\}$, $c\in\mathbb{C}$.

We fix a nonzero contravariant form $(\ ,\ )_\mu$ on $M(\mu)$ ([5]) and denote by $(\ ,\ )_{\mu,m}$ its restriction to $M(\mu)_{\mu-m}$. A formula for the determinant of the latter was obtained by V.G. Kac[1]:

$$\det(\ ,\ )_{\mu,m}=*\prod_{i=1}^{m}\left(\prod_{r|i}\left\{\mu+\frac{1}{24}\left[\left(3r-\frac{2i}{r}\right)^2-1\right]\right\}\right)^{p(m-i)}$$

where $*$ is a nonzero scalar (depending on the basis used) and p is the classical partition function. Using this formula V.G. Kac concludes that $M(\mu)$ is irreducible if and only if $\mu\neq-\frac{1}{24}(m^2-1)$, $m\in\mathbb{Z}_+$ ([5]). The integers of the form $\frac{1}{24}(m^2-1)$, $m\in\mathbb{Z}_+$, are the numbers $\frac{1}{2}(3k^2\pm k)$, $k\in\mathbb{Z}_+$. We let s_k (resp. t_k) be the non-negative integer $\frac{1}{2}(3k^2+k)$ (resp. $\frac{1}{2}(3k^2-k)$), $k\in\mathbb{Z}_+$, and let $P=\{s_k, t_k, k\in\mathbb{Z}_+\}$.

Let $\underline{n}^-=\bigoplus_{i\in\mathbb{N}}\mathbb{C}\,e_{-i}$. The cohomology $H^*(\underline{n},\mathbb{C})$, or equivalently, the homology $H_*(\underline{n}^-,\mathbb{C})$ was computed by Goncharova [3]. More precisely, she proved that

$$H_k(\underline{n}^-,\mathbb{C})_\nu=0 \qquad \text{unless } \nu=-s_k \text{ or } \nu=-t_k$$

and

$$H_k(\underline{n}^-,\mathbb{C})_\nu=\mathbb{C}(\nu) \qquad \text{if } \nu=-s_k \text{ or } \nu=-t_k.$$

[1] This formula (and the more general one for the Virasoro algebra) was announced in [5]; however, Kac has not, to the time of the publication of this article, circulated a proof

Here $H_k(\underline{n}^-, \mathbb{C})_\nu$ is the ν-weightspace of $H_k(\underline{n}^-, \mathbb{C})$ relative to the action of $\underline{h}$ and $\mathbb{C}(\nu)$ is the $\underline{h}$-module where $\underline{h}$ acts as ν.

In ([7, 9]) we obtained a resolution of the trivial module:

$$(0) \quad \ldots \to M(-s_k) \oplus M(-t_k) \xrightarrow{d_k} \ldots \to M(-2) \oplus M(-1) \xrightarrow{d_1} M(0) \xrightarrow{\varepsilon} L(0) \to 0$$

using (as Kac does in [6]) Goncharova's result. Our proof combines some general results obtained in [8] with Goncharova's computations in the form described above.

In this paper we construct resolutions of $L(-\nu)$, $\nu \in \{s_k, t_k\}$:

$$(k) \quad \ldots \to M(-s_i) \oplus M(-t_i) \xrightarrow{d_i} \ldots \xrightarrow{d_{k+2}} M(-s_{k+1}) \oplus M(-t_{k+1})$$
$$\xrightarrow{d_{k+1}} M(-\nu) \xrightarrow{\varepsilon_k} L(-\nu) \to 0$$

using (0) as a starting point. These two-term resolutions whose existence is the content of Theorem C below, imply immediately a generalization of Goncharova's result and yield also immediately the character formulas conjectured by V.G. Kac (see [6] for a recent reformulation of these conjectures). By Kac's above mentioned irreducibility criterion, Theorem C yields the characters of all irreducible integral highest weight modules over the Witt algebra.

In [9] we set up a conjectural scheme to prove Theorem C. This conjectural scheme was modeled after a similar program completely carried out in [9] for rank 2 symmetrizable Kac-Moody Lie algebras. With this scheme in mind, we constructed and resolved in [9, § 6] certain highest weight modules which we denote by $\tilde{L}(-\nu)$, $\nu \in P$. By [9, Theorem 6.14] the irreducibility of the $\tilde{L}(-\nu)$ is equivalent to the statement of Theorem C. We will prove Theorem C using this equivalence.

We will now state the main results of this article. In order to carry out the program devised in [9, § 6] we found it necessary to work over the universal central extension $\tilde{\underline{g}}$ of $\underline{g}$. $\tilde{\underline{g}}$ is the complex Lie algebra with basis $\{E'_0, E_i\}$ $(i \in \mathbb{Z})$ such that

$$[E_i, E_j] = (j-i) E_{i+j} + \tfrac{1}{12}(i^3 - i)\, \delta_{i,-j} E'_0 \quad \text{and} \quad [E_i, E'_0] = 0.$$

Let $\tilde{\underline{h}} = \mathbb{C}E_0 \oplus \mathbb{C}E'_0$, $\tilde{\underline{n}}_- = \bigoplus_{-i \in \mathbb{N}} \mathbb{C}E_i$. If $\mu \in \tilde{\underline{h}}$ and $\mu(E_0) = h$, $\mu(E'_0) = c$, we set $M(h, c) = M(\mu)$ where $M(\mu)$ is the Verma module with highest weight μ. We set $(\ ,\ )_{h,c} = (\ ,\ )_\mu$, where $(\ ,\ )_\mu$ is a fixed contravariant form on $M(\mu)$. We denote by $(\ ,\ )_{h,c,m}$ the restriction of $(\ ,\ )_{h,c}$ to the $\mu - m$ weightspace $M(\mu)_{\mu-m}$ of $M(\mu)$, $m \in \mathbb{Z}_+$. Kac's formula for the determinant of $(\ ,\ )_{h,c,m}$ states: Up to a non-zero multiple

$$\det(\ ,\ )_{h,c,m} = \prod_{i=1}^{m} \Big(\prod_{r \mid i,\, r^2 \leq i} \psi_{r,i/r}(h, c) \Big)^{p(m-i)}$$

where $\psi_{r,s}(h, c)$ is polynomial of degree 2 ([5])[2].

Our first result is the construction of elements giving embeddings among Verma modules over $\underline{g}$. These are cross-sections on the varieties $\psi_{r,s}(h, c) = 0$

---

[2] See Footnote [1]

and they are analogues of the elements constructed by Shapovalov [10] for semisimple Lie algebras where hyperplanes play the role of the varieties.

Theorem A. *Let $k>0$ be fixed, $v\in\{-s_k, -t_k\}$. There exists $\varepsilon_v>0$ such that if $D^2_{\varepsilon_v}=\{(h,c)\in\mathbb{C}^2 \mid |h-v|<\varepsilon_v,\ |c|<\varepsilon_v\}$ then there is a holomorphic map θ_v: $D^2_{\varepsilon_v}\to U(\underline{\tilde{n}}^-)_{-v-t_{k+1}}$ such that $\theta_v(h,\phi_v(h))\,v_{h,\phi_v(h)}$ gives the (unique up to scalar) embedding*

$$M(h-v-t_{k+1},\phi_v(h))\hookrightarrow M(h,\phi_v(h)).$$

Here, $v_{h,c}=1\otimes 1$ in $M(h,c)$ and $(h,\phi_v(h))$ is a rational parametrization of $\psi_{1,v+t_{k+1}}=0$.

Using the techniques of Jantzen [4] we note that the formula for the determinant of $(\ ,\ )_{\mu,m}$, $\mu\in\underline{h}^*$, $m\in\mathbb{Z}_+$, is equivalent to a character sum formula (§ 2). From this point of view the next theorem computes the determinant of a contravariant form on the quotient of two Verma modules as the quotient of two determinants. This result is given here as a character sum formula involving the terms of a filtration of a quotient of two Verma modules which Theorem A permits us to construct.

Theorem B. *Let $k\geq 1$, $v\in\{-s_k, -t_k\}$. Set $N(v)=M(v)/M(-t_{k+1})$. There is a filtration of submodules*

$$N(v)=N(v)_{(0)}\supset N(v)_{(1)}\supset N(v)_{(2)}\supset\dots$$

such that

(i) *$N(v)_{(1)}$ is the largest proper submodule of $N(v)$;*

(ii) *For every $i\in\mathbb{N}$, there is on $N(v)_{(i)}/N(v)_{(i+1)}$ a non-degenerate contravariant form;*

(iii) $$\sum_{i\geq 1}\operatorname{ch}N(v)_{(i)}=\sum_{l\in\mathbb{N},\,l\text{ odd}}(\operatorname{ch}M(-s_{k+l})+\operatorname{ch}M(-t_{k+l}))-\operatorname{ch}M(-t_{k+1})$$
$$-\sum_{j\in\mathbb{N},\,j\text{ odd}}(\operatorname{ch}M(-s_{k+1+j})+\operatorname{ch}M(-t_{k+1+j}));$$

(iv) $N(v)_{(1)}=(M(-s_{k+1})+M(-t_{k+1}))/M(-t_{k+1})$, $\quad N(v)_{(i)}=0,\ i>1.$

Using Theorem B and the same formal argument of [9, §§ 4, 5] we prove the irreducibility of $\tilde{L}(v)$ and hence the following

Theorem C. *Let $v\in\{-s_k, -t_k\}$ $k\in\mathbb{Z}_+$. There exists a resolution of $L(-v)$:*

$$(k)\quad \dots\to M(-s_i)\oplus M(-t_i)\xrightarrow{d_i}\dots\xrightarrow{d_{k+2}}M(-s_{k+1})\oplus M(-t_{k+1})$$
$$\xrightarrow{\eta_{k+1}}M(-v)\xrightarrow{\varepsilon_k}L(-v)\to 0.$$

(The maps are given explicitly in the proof.)

In § 1 we prove some general lemmas which will be used in the proofs of Theorems A–C. In § 2 we construct a variation of Jantzen's filtration of a Verma module and prove an analogue of his character sum formula. §§ 3, 4 and 5 are devoted to the proofs of Theorems A, B and C, respectively.

§ 1. Preliminaries

In this section we will prove some general results which will be applied in the subsequent sections. The first three lemmas will be used in the constructions of filtrations of Verma modules in §§ 2, 4. These filtrations are the counterparts of Jantzen's filtrations [4]. Their derivations here are a variation of Jantzen's construction that suit the algebras treated in this paper in addition to GCM Lie algebras.

Let V be a finite dimensional vector space over $\mathbb{C}$. Let $t_0 \in \mathbb{C}$. For each $\varepsilon > 0$ we denote by $D_\varepsilon(t_0)$ the open disc $\{t \in \mathbb{C} \mid |t - t_0| < \varepsilon\}$. We say that $f_1 \colon D_{\varepsilon_1}(t_0) \to V$ and $f_2 \colon D_{\varepsilon_2}(t_0) \to V$ have the same *germ* at t_0 if there is $\varepsilon > 0$ such that $f_1|_{D_\varepsilon(t_0)} \equiv f_2|_{D_\varepsilon(t_0)}$. We denote by $\mathcal{O}_{t_0}(V)$ the complex vector space of germs $\mathbf{f}$ of holomorphic functions $f \colon D_\varepsilon(t_0) \to V$ at t_0.

Let $\operatorname{End}(V)$ denote the space of endomorphisms of V. Let $\mathbf{A} \in \mathcal{O}_{t_0}(\operatorname{End}(V))$ and write $A_t = A(t)$. For each $k \in \mathbb{N}$ we set

$$\mathcal{O}_{t_0}^A(V)_k = \{\mathbf{f} \in \mathcal{O}_{t_0}(V) \mid A_t\, f(t) = (t - t_0)^k g(t),\ g \in \mathcal{O}_{t_0}(V) \text{ for all } t \in D_\varepsilon(t_0),\ \varepsilon > 0\}.$$

We also set $V_{t_0,k}^A = \{f(t_0) \mid \mathbf{f} \in \mathcal{O}_{t_0}^A(V)_k\}$. Then

$$V = V_{t_0,0}^A \supset V_{t_0,1}^A \supset V_{t_0,2}^A \supset \dots .$$

If $\mathbf{f} \in \mathcal{O}_{t_0}(\mathbb{C})$, let $f(t) = \sum_{i \geqq 0} a_i (t - t_0)^i$, $t \in D_{t_0}(\varepsilon)$, $\varepsilon > 0$. If $a_i \neq 0$ for at least one i, we denote by $\operatorname{ord}_{t_0} f(t)$ the smallest such index.

$$\text{Let } P_{ij} = \begin{array}{cc} \begin{array}{ccccccccccc} & & & i & & & & j & & & \end{array} & \\ \begin{bmatrix} 1 & & & \vdots & & & & \vdots & & & \\ & \ddots & & \vdots & & & & \vdots & & & \\ & & 1 & \vdots & & & & \vdots & & & \\ & & & 0 & \cdots & \cdots & \cdots & 1 & \cdots & \cdots & \cdots \\ & & & \vdots & 1 & & & \vdots & & & \\ & & & \vdots & & \ddots & & \vdots & & & \\ & & & \vdots & & & 1 & \vdots & & & \\ & & & 1 & \cdots & \cdots & \cdots & 0 & \cdots & \cdots & \cdots \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix} & \begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} \end{array} \quad \text{and } Q_{ij}(\beta(t)) = \begin{array}{cc} \begin{array}{ccccccc} & & & & j & & \end{array} & \\ \begin{bmatrix} 1 & & & & \vdots & & \\ & \ddots & & & \vdots & & \\ & & 1 & \cdots & \beta(t) & \cdots & \cdots \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} & \begin{array}{c} \\ \\ i \\ \\ \\ \\ \\ \end{array} \end{array}, \quad \beta \in \mathcal{O}_{t_0}(\mathbb{C}).$$

Lemma 1. *Let* $\mathbf{A} \in \mathcal{O}_{t_0}(\operatorname{End}(V))$ *be such that* $\det A_t \neq 0$ *for* $t \in D_\varepsilon(t_0) \setminus \{t_0\}$. *Let* $\{e_1, \dots, e_r\}$ *be a basis of* V. *Then there are* $\mathbf{R}, \mathbf{S} \in \mathcal{O}_{t_0}(\operatorname{End}(V))$ *such that* $\det R$ *and* $\det S$ *are non-zero constant functions and the matrix of* $R_t A_t S_t$ *relative to* $\{e_1, \dots, e_r\}$ *has diagonal form:*

$$D_t = \begin{bmatrix} a_1(t) & & 0 \\ & \ddots & \\ 0 & & a_r(t) \end{bmatrix}, \quad \text{where } 0 \leqq \operatorname{ord}_{t_0} a_1(t) \leqq \dots \leqq \operatorname{ord}_{t_0}(a_r(t)) < \infty.$$

Proof. We prove the result by induction on r. If $r = 1$ the result is clear. If $r > 1$, let $[b_{ij}(t)]$, $1 = i, j \leqq r$, $\mathbf{b}_{ij} \in \mathcal{O}_{t_0}(\mathbb{C})$, be the matrix of A_t relative to $\{e_1, \dots, e_r\}$. Let

$\operatorname{ord}_{t_0} b_{kl}(t)$ be a minimum. Then $C_t = P_{1k} A_t P_{1l}$ has $b_{kl}(t)$ on the first row and on the first column. Let $C_t = [c_{ij}(t)]$, $1 \leqq i, j \leqq r$. We note that $\frac{\mathbf{c}_{ij}}{\mathbf{c}_{11}} \in \mathcal{O}_{t_0}(\mathbb{C})$ for all $(i,j) \neq (1,1)$. Multiplying C_t on the left by

$$T_t = Q_{r1}\left(\frac{-c_{r1}(t)}{c_{11}(t)}\right) \dots Q_{21}\left(\frac{-c_{21}(t)}{c_{11}(t)}\right)$$

and on the right by

$$U_t = Q_{12}\left(\frac{-c_{12}(t)}{c_{11}(t)}\right) \dots Q_{1r}\left(\frac{-c_{1r}(t)}{c_{11}(t)}\right)$$

we obtain a matrix of the form

$$\begin{bmatrix} b_{k\ell} & 0 \cdots\cdots 0 \\ 0 & \\ \vdots & A^2 \\ 0 & \end{bmatrix}$$

Let V_2 be the span of $\{e_2, \dots, e_r\}$. By the induction assumption there are $\mathbf{R}^2, \mathbf{S}^2 \in \mathcal{O}_{t_0}(\operatorname{End}(V_2))$ such that $\det R^2$ and $\det S^2$ are nonzero constant functions and $R_t^2 A_t^2 S_t^2$ has a matrix representation of the form

$$\begin{bmatrix} a_2(t) & & 0 \\ & \ddots & \\ 0 & & a_r(t) \end{bmatrix}$$

with $0 \leqq \operatorname{ord}_{t_0} a_2(t) \leqq \dots \leqq \operatorname{ord}_{t_0} a_r(t) < \infty$. We now set $R^1 = TP_{1k}$,

$$S^1 = P_{1,l} U, \quad \hat{R}^2 = \begin{bmatrix} 1 & 0 \cdots\cdots 0 \\ 0 & \\ \vdots & R^2 \\ 0 & \end{bmatrix} \quad \text{and} \quad \hat{S}^2 = \begin{bmatrix} 1 & 0 \cdots\cdots 0 \\ 0 & \\ \vdots & S^2 \\ 0 & \end{bmatrix}$$

where R^2 and S^2 also denote the matrices of R^2 and S^2, respectively, relative to $\{e_2, \dots, e_r\}$. Now set $R = \hat{R}^2 R^1$ and $S = S^1 \hat{S}^2$. Q.E.D.

Lemma 2. *Let* $\mathbf{A} \in \mathcal{O}_{t_0}(\operatorname{End}(V))$ *be such that* $\det A_t \neq 0$ *for* $t \in D_\varepsilon(t_0) \setminus \{t_0\}$. *Then*

$$\sum_{k \geqq 1} \dim V_{t_0,k}^A = \operatorname{ord}_{t_0} \det A_t.$$

Proof. By Lemma 1 there are $\mathbf{R}, \mathbf{S} \in \mathcal{O}_{t_0}(\mathrm{End}(V))$ such that $\det R$ and $\det S$ are non-zero constants, and, relative to a basis $\{e_1, \dots, e_r\}$ of V, $R_t A_t S_t$ has a matrix representation

$$D_t = \begin{bmatrix} a_1(t) & & 0 \\ & \ddots & \\ 0 & & a_r(t) \end{bmatrix}$$

where $0 \leqq \mathrm{ord}_{t_0} a_1(t) \leqq \dots \leqq \mathrm{ord}_{t_0} a_r(t) < \infty$. Set $d_i = \mathrm{ord}_{t_0} a_i(t)$. Then $\mathrm{ord}_{t_0} \det A_t = \mathrm{ord}_{t_0} \det D_t = \sum_{i=1}^{r} d_i$. We note that $\mathcal{O}^D_{t_0}(V)_k = \{\mathbf{S}^{-1}\mathbf{f} \mid f \in \mathcal{O}^A_{t_0}(V)_k\}$ and hence $V^D_{t_0,k} = S^{-1}(t_0)\, V^A_{t_0,k}$. Let $\mathbf{f_j} \in \mathcal{O}^D_{t_0}(V)$ be such that $e_j = f_j(t_0)$, $1 \leqq j \leqq r$. Now $V^D_{t_0,k} = \sum_{d_j \geqq k} \mathbb{C}\, e_j$ and hence $\dim V^A_{t_0,k} = \dim V^D_{t_0,k} = \#\{j \mid k \leqq d_j\}$. This implies that

$$\sum_{k \geqq 1} \dim V^A_{t_0,k} = \sum_{k \geqq 1} \#\{j \mid k \leqq d_j\} = \sum_{j=1}^{r} d_j. \quad \text{Q.E.D.}$$

Let $\mathcal{L}^2(V, \mathbb{C})$ denote the space of all bilinear forms on V and let $\mathcal{L}^2_s(V, \mathbb{C})$ denote the space of all symmetric bilinear forms on V. Let $(\ ,\ )$ be a non-degenerate symmetric bilinear form on V such that $(A_t v, w) = (v, A_t w)$ for all $v, w \in V$. Set $B_t(v, w) = (A_t v, w)$. Then $\mathbf{B} \in \mathcal{O}_{t_0}(\mathcal{L}^2_s(V, \mathbb{C}))$. For $\mathbf{f}, \mathbf{g} \in \mathcal{O}_{t_0}(V)$, set $\langle f, g\rangle(t) = B_t(f(t), g(t))$. If $\mathbf{f}, \mathbf{g} \in \mathcal{O}^A_{t_0}(V)_k$, let $\langle f, g\rangle_k(t) = (t - t_0)^{-k} \langle f, g\rangle(t)$. Let $f(t_0) = 0$. Then

$$f(t) = (t - t_0)\, h(t), \qquad \mathbf{h} \in \mathcal{O}_{t_0}(V),$$

and

$$\begin{aligned} \langle f, g\rangle_k(t) &= (t-t_0)^{-k} \langle f, g\rangle(t) = (t-t_0)^{-k} B_t(f(t), g(t)) \\ &= (t-t_0)^{-k}(A_t f(t), g(t)) = (t-t_0)^{-k}(f(t), A_t g(t)) \\ &= (t-t_0)^{-k+1}(h(t), A_t g(t)) = (t-t_0)\,(h(t), a(t)), \qquad \mathbf{a} \in \mathcal{O}_{t_0}(\mathbb{C}). \end{aligned}$$

This implies that we may define for $v, w \in V^A_{t_0,k}$, $v = f(t_0)$, $w = g(t_0)$, $\mathbf{f}, \mathbf{g} \in \mathcal{O}^A_{t_0}(V)_k$, $\langle v, w\rangle_k = \langle f, g\rangle_k(t_0)$. The same argument just used shows that $\langle V^A_{t_0,k}, V^A_{t_0,k+1}\rangle = 0$. Hence $\langle\ ,\ \rangle_k$ induces a symmetric bilinear form on $V^A_{t_0,k}/V^A_{t_0,k+1}$ which we also denote by $\langle\ ,\ \rangle_k$.

Lemma 3. *If A satisfies the hypothesis of Lemmas* 1 *and* 2, *then $\langle\ ,\ \rangle_k$ is non-degenerate on $V^A_{t_0,k}/V^A_{t_0,k+1}$, $k \geqq 0$.*

Proof. Let the notation be as in the proof of Lemma 2. Set $v_i = S_{t_0} e_i$, $1 \leqq i \leqq r$. Then $V^A_{t_0,k} = \sum_{d_i \geqq k} \mathbb{C}\, v_i$. Let $f_i(t) = S_t e_i$. If $d_i = k$ and $d_j \geqq k$, then

$$\begin{aligned} \langle f_i, f_j\rangle_k(t) &= (t-t_0)^{-k}(A_t S_t e_i, f_j(t)) \\ &= (t-t_0)^{-k}(t-t_0)^{d_i}(h_i(t) R_t^{-1} e_i, f_j(t)) \\ &= (h_i(t) R_t^{-1} e_i, f_j(t)), \end{aligned}$$

where $a_i(t)=(t-t_0)^{d_i}h_i(t)$, $h_i(t_0)\neq 0$. Set $w_i=h_i(t_0)R_{t_0}^{-1}e_i$. Then $\langle v_i, v_j\rangle_k=(w_i, v_j)$. Furthermore, $(w_i, v_j)=0$ if $d_j>d_i$. Now, $V=\sum_{i=1}^{r}\mathbb{C}\,w_i$ and (,) is non-degenerate on V. By the above, the matrix $[(w_i, v_j)]$ has block form:

$$\begin{bmatrix} M_1 & & & 0 \\ & M_2 & & \\ & & M_3 & \\ * & & & \ddots \end{bmatrix}$$

with blocks $M_k=[(w_i, v_j)]$, $(d_i=d_j=k)$. Hence $\det[(w_i, v_j)]=\prod_k \det M_k$, which shows that $\det M_k\neq 0$ if $V^A_{t_0,k}\neq V^A_{t_0,k+1}$. Therefore $\det\langle\ ,\ \rangle_k=\det M_k\neq 0$ and $\langle\ ,\ \rangle_k$ is nondegenerate. Q.E.D.

Remark. Let $\mathbf{B}\in\mathcal{O}_{t_0}(\mathcal{L}^2_s(V,\mathbb{C}))$ and let (,) be a symmetric bilinear non-degenerate form on V. Then there is $\mathbf{A}\in\mathcal{O}_{t_0}(\mathrm{End}(V))$ such that $B_t(v,w)=(A_t v, w)$ for all $v, w\in V$. Let

$$\mathcal{O}^B_{t_0}(V)_k=\{f\in\mathcal{O}_{t_0}(V)\,|\,B_t(f(t), w)\in(t-t_0)^k\,\mathcal{O}_{t_0}(\mathbb{C}),\ \text{for all } w\in V\}.$$

It is clear that $\mathcal{O}^B_{t_0}(V)_k=\mathcal{O}^A_{t_0}(V)_k$, all $k\geqq 0$.

The following lemmas will be needed in § 3.

Lemma 4. *Let* $\mathbf{B}\in\mathcal{O}_{t_0}(\mathcal{L}^2_s(V,\mathbb{C}))$ *be such that* B_t *is non-degenerate for* $t\in D_\varepsilon(t_0)\setminus\{t_0\}$. *Let* (,) *be a symmetric bilinear non-degenerate form on* V. *Let* $\mathbf{A}\in\mathcal{O}_{t_0}(\mathrm{End}(V))$ *be such that* $B_t(v,w)=(A_t v, w)$ *for all* $v, w\in V$. *Let* $\det A_t=(t-t_0)^k q(t)$ *with* $q(t_0)\neq 0$, $\mathbf{q}\in\mathcal{O}_{t_0}(\mathbb{C})$. *Then* $\dim \operatorname{Rad} B_{t_0}\leqq k$.

Proof. Let $\{e_1, \dots, e_r\}$ be a basis of V. By Lemma 1 there are $\mathbf{R}, \mathbf{S}\in\mathcal{O}_{t_0}(\mathrm{End}(V))$ such that $\det R$ and $\det S$ are non-zero constant functions and the matrix of $R_t A_t S_t$ relative to $\{e_1, \dots, e_r\}$ has form

$$\begin{bmatrix} a_1(t) & & 0 \\ & \ddots & \\ 0 & & a_r(t) \end{bmatrix}$$

where $0\leqq \mathrm{ord}_{t_0} a_1(t)\leqq\dots\leqq \mathrm{ord}_{t_0} a_r(t)<\infty$. Let $d_i=\mathrm{ord}_{t_0} a_i(t)$. Now,

$$\begin{aligned}\det A_t(e_1\wedge\dots\wedge e_r)&=c\det R_t A_t S_t(e_1\wedge\dots\wedge e_r)\\ &=c\,R_t A_t S_t\,e_1\wedge\dots\wedge R_t A_t S_t\,e_r\\ &=(t-t_0)^{d_1+\dots+d_r}\,w(t)\,e_1\wedge\dots\wedge e_r,\end{aligned}$$

where $c \neq 0$ and $w(t_0) \neq 0$, $w \in \mathcal{O}_{t_0}(\mathbb{C})$. Let i_0 be the smallest index such that $d_{i_0} \neq 0$. Then $\dim \operatorname{Rad} B_{t_0} = r - i_0 + 1 \leqq d_{i0} + \ldots + d_r = k$. Q.E.D.

Lemma 5. *Let* $A \in \mathcal{O}_{t_0}(\operatorname{End}(V))$ *be such that* $\dim \operatorname{Ker} A_t = 1$ *for* $|t - t_0| < \varepsilon$, $\varepsilon > 0$. *Then there exist* $\xi \in \mathcal{O}_{t_0}(V)$ *and* $\varepsilon_1 > 0$ *such that* $\xi(t) \in \operatorname{Ker} A_t \setminus \{0\}$ *for all* $t \in D_{\varepsilon_1}(t_0)$.

Proof. Let $\mathbb{C} v_0 = \operatorname{Ker} A_{t_0}$ and let $\{v_0, v_1, \ldots, v_{r-1}\}$ be a basis of V. Let $w_i = A_{t_0} v_i$, $i = 1. \ldots, r-1$ and let $\{w_1, \ldots, w_r\}$ be a basis of V. Then $A_t(v_i) = \sum_{j=1}^{r} a_{ji}(t)\, w_j$, $i = 1, \ldots, r-1$. Now, $\det[a_{ij}(t_0)]$ $(1 \leqq i, j \leqq r-1)$ equals 1, hence $\det[a_{ij}(t)]$ $(1 \leqq i, j \leqq r-1)$ is non-zero for $t \in D_{\varepsilon_1}(t_0)$, $\varepsilon_1 > 0$. This implies that $\{A_t v_i\}$ $(1 \leqq i \leqq r-1)$ is a bais of $A_t V$ for all $t \in D_{\varepsilon_1}(t_0)$. We set $A_t v_0 = \sum_{i=1}^{r-1} \phi_i(t) A_t v_i$. Let $[b_{ij}(t)] = [a_{ij}(t)]^{-1}$ $(1 \leqq i, j \leqq r-1)$, $t \in D_{\varepsilon_1}(t_0)$. Then

$$\begin{aligned} B_t A_t v_0 &= \sum_{i=1}^{r-1} \phi_i(t) \sum_{k=1}^{r-1} \sum_{j=1}^{r} b_{kj}(t)\, a_{ji}(t)\, w_k \\ &= \sum_{i=1}^{r-1} \phi_i(t)\, w_i + q(t), \qquad q \in \mathcal{O}_{t_0}(\mathbb{C}\, w_r), \end{aligned}$$

where $B_t w_j = \sum_{k=1}^{r-1} b_{kj} w_k$, for $j = 1, \ldots, r-1$, and $B_t w_r = w_r$. Hence $\phi_i \in \mathcal{O}_{t_0}(\mathbb{C})$, $i = 1, \ldots, r-1$. Now set $\xi(t) = v_0 - \sum_{i=1}^{r-1} \phi_i(t)\, v_i$. Q.E.D.

§ 2. A Character Sum Formula for a Verma Module

We recall the definition of the *Witt algebra* $\underline{g}$: $\underline{g}$ is the complex Lie algebra with basis $\{e_i\}_{i \in \mathbb{Z}}$ such that

$$[e_i, e_j] = (j - i)\, e_{i+j}, \qquad i, j \in \mathbb{Z}.$$

We set $\underline{h} = \mathbb{C} e_0$, $\underline{n} = \bigoplus_{i \in \mathbb{N}} \mathbb{C} e_i$, $\underline{b} = \underline{h} \oplus \underline{n}$ and $\underline{n}^- = \bigoplus_{i \in \mathbb{N}} \mathbb{C} e_{-i}$. Let $\alpha \in \underline{h}^*$ be defined by $\alpha(e_0) = 1$. Let $Q = \mathbb{Z}\alpha$. For $\beta \in Q$ we set

$$\underline{g}_\beta = \{X \in \underline{g} \mid [H, X] = \beta(H)\, X \text{ for all } H \in \underline{h}\}.$$

Then $\underline{g}_{i\alpha} = \mathbb{C} e_i$, $\underline{g}$ has a grading $\underline{g} = \bigoplus_{\beta \in Q} \underline{g}_\beta$ with $\underline{g}_0 = \underline{h}$, $\dim \underline{g}_\beta = 1$ and $[\underline{g}_\beta, \underline{g}_\gamma] \subset \underline{g}_{\beta+\gamma}$, $\beta, \gamma \in Q$. Furthermore, there is an $\underline{h}$-invariant, non-degenerate pairing between the spaces $\underline{n}$ and $\underline{n}^-$ [9, Proposition 6.2]. Hence, $(\underline{g}, \underline{h})$ satisfies the conditions (T1) and (T2) of [9, § 1].

For reasons that will become clear in § 4 we need to consider the *Virasoro algebra* $\tilde{\underline{g}}$ which is the universal central extension of $\underline{g}$ (cf. [5]). $\tilde{\underline{g}}$ is the complex Lie algebra with basis $\{E_0', E_i\}_{i \in \mathbb{Z}}$ such that

$$[E_i, E_j]=(j-i)\,E_{i+j}+\tfrac{1}{12}(i^3-i)\,\delta_{i,-j}E_0' \quad \text{and} \quad [E_i, E_0']=0.$$

Let $\tilde{\underline{h}}=\mathbb{C}E_0+\mathbb{C}E_0'$ and $\tilde{\alpha}\in\tilde{\underline{h}}^*$ be defined by $\tilde{\alpha}(E_0)=1$, $\tilde{\alpha}(E_0')=0$. Let $\tilde{Q}=\mathbb{Z}\tilde{\alpha}$. For $\beta\in\tilde{Q}$ we set

$$\tilde{\underline{g}}_\beta=\{X\in\tilde{\underline{g}}\,|\,[H,X]=\beta(H)\,X \text{ for all } H\in\tilde{\underline{h}}\}.$$

Hence $\tilde{\underline{g}}$ has a grading: $\tilde{\underline{g}}=\bigoplus_{\beta\in\tilde{Q}}\tilde{\underline{g}}_\beta$ with $\tilde{\underline{g}}_0=\tilde{\underline{h}}$, $\dim\tilde{\underline{g}}_\beta<\infty$ and $[\tilde{\underline{g}}_\beta, \tilde{\underline{g}}_\gamma]\subset\tilde{\underline{g}}_{\beta+\gamma}$, $\beta,\gamma\in\tilde{Q}$, that is, the pair $(\tilde{\underline{g}},\tilde{\underline{h}})$ has a grading of the type considered in [9, §1] which also satisfies condition (T1) of [9, §1]. We identify $\tilde{Q}$ with $\mathbb{Z}$ and let $\tilde{\underline{n}}=\bigoplus_{i\in\mathbb{N}}\tilde{\underline{g}}_i$, $\tilde{\underline{n}}^-=\bigoplus_{i\in\mathbb{N}}\tilde{\underline{g}}_{-i}$ and $\tilde{\underline{b}}=\tilde{\underline{h}}\oplus\tilde{\underline{n}}$. Let $B(E_i,E_j)=\delta_{i,-j}$. Then B defines an $\tilde{\underline{h}}$-invariant pairing between the spaces $\tilde{\underline{n}}$ and $\tilde{\underline{n}}^-$. Hence, condition (T2) of [9, §1] is also satisfied by the pair $(\tilde{\underline{g}},\tilde{\underline{h}})$. We can therefore consider highest weight modules over $\tilde{\underline{g}}$ in appropriate categories of modules over $\tilde{\underline{g}}$, as in [9, §1]. If $\lambda\in\tilde{\underline{h}}^*$ we let $\mathbb{C}(\lambda)$ be the one-dimensional $\tilde{\underline{b}}$-module where $\underline{n}$ acts trivially and $\tilde{\underline{h}}$ acts via λ. We denote by $M(\lambda)$ the universal highest weight module or *Verma module*, $M(\lambda)=U(\tilde{\underline{g}})\otimes_{U(\tilde{\underline{b}})}\mathbb{C}(\lambda)$, associated with $\tilde{\underline{g}}$, $\tilde{\underline{h}}$, $\tilde{Q}^+$ and λ. Let $L(\lambda)$ denote the unique irreducible quotient of $M(\lambda)$. If $\lambda\in\tilde{\underline{h}}^*$ and M is a $\tilde{\underline{g}}$-module, we define

$$M_\lambda=\{v\in M\,|\,Hv=\lambda(H)\,v \text{ for all } H\in\tilde{\underline{h}}\}$$

the *λ-weightspace* of M. A *highest weight module* is a module M with a non-zero vector $v\in M_\lambda$ such that $\tilde{\underline{n}}v=0$ and $M=U(\tilde{\underline{g}})\,v$. λ is the *highest weight* of M. If M is a highest weight module, set $\operatorname{ch}M=\sum_{\mu\in\tilde{\underline{h}}^*}\dim M_\mu e^\mu$ where e^μ is a formal exponential.

Let σ be the involutive antiautomorphism of $\tilde{\underline{g}}$ such that $\sigma(E_i)=E_{-i}$, all $i\in\mathbb{Z}$, $\sigma(E_0')=E_0'$. Let σ also denote the extension of σ to an involutive anti-automorphism of $U(\tilde{\underline{g}})$. Let M be a $\tilde{\underline{g}}$-module. A symmetric, bilinear form $(\ ,\ )$ on M is *contravariant* (relative to σ) if $(Xv,w)=(v,\sigma(X)w)$ for all $X\in U(\tilde{\underline{g}})$, v, $w\in M$. The following counterpart of [4, Satz 1.6] is proved with the same type of arguments used in [4, Satz 1.6] (cf. [5]).

Proposition 1. (a) *Let M be a $\tilde{\underline{g}}$-module and let $(\ ,\ )$ be a contravariant form on M. Then $(M_\mu, M_\nu)=0$ if $\mu\neq\nu$.* (b) *Let M be a highest weight module. Then there is on M a non-zero contravariant form $(\ ,\ )$. $(\ ,\ )$ is unique up to a scalar multiple and* Rad$(\ ,\ )$ *is the largest proper submodule of M.*

Proposition 1 says, in particular, that there is a contravariant form on $M(\lambda)$, $\lambda\in\tilde{\underline{h}}^*$. We fix one, which we denote by $(\ ,\ )_\lambda$, by requiring that $(1\otimes 1, 1\otimes 1)_\lambda=1$. We recall the definition of $(\ ,\ )_\lambda$. Let β denote the projection of $U(\tilde{\underline{g}})$ onto the first summand relative to the decomposition

$$U(\tilde{\underline{g}})=U(\tilde{\underline{h}})\oplus(\tilde{\underline{n}}^-U(\tilde{\underline{g}})+U(\tilde{\underline{g}})\,\tilde{\underline{n}}).$$

We set $v_\lambda = 1\otimes 1$ in $M(\lambda)$. Let $(Xv_\lambda, Yv_\lambda)_\lambda = (\lambda\circ\beta)(\sigma(X)\,Y)$, for all $X, Y\in U(\underline{\tilde{g}})$. Then $(\ ,\ )_\lambda$ has the properties we sought. In particular, $(\ ,\ )_\lambda$ is non-degenerate if and only if $M(\lambda)=L(\lambda)$. We denote by $(\ ,\ )_{\lambda,m}$ the restriction of $(\ ,\ )_\lambda$ to $M(\lambda)_{\lambda-m}$, $m\in\mathbb{Z}_+$. (Here $(\lambda-m)(E_0)=\lambda(E_0)-m$ and $(\lambda-m)(E_0')=\lambda(E_0')$.)

Following [5], let for $c\in\mathbb{C}$, $r, s\in\mathbb{N}$,

$$r\neq s,\ \alpha_{r,s}^{\pm}(c) = -\tfrac{1}{48}\left[(13-c)(r^2+s^2)\pm\sqrt{c^2-26c+25}\,(r^2-s^2)-24rs-2+2c\right].$$

If $r\neq s$, $h\in\mathbb{C}$, let $\psi_{r,s}(h,c)=(h-\alpha_{r,s}^+(c))(h-\alpha_{r,s}^-(c))$ and let $\psi_{r,r}(h,c)=h+\frac{1}{24}(r^2-1)(1-c)$.

Theorem 2 [5]. *Let $\lambda\in\underline{\tilde{h}}^*$, $\lambda(E_0)=h$, $\lambda(E_0')=c$. Set $(\ ,\ )_{h,c,m}=(\ ,\ )_{\lambda,m}$. Then, up to a non-zero multiple,*

$$\det(\ ,\ )_{h,c,m}=\prod_{i=1}^{m}\Big(\prod_{r|i,\,r^2\leqq i}\psi_{r,i/r}(h,c)\Big)^{p(m-i)},^{3}$$

where p is the classical partition function.

Note. The Virasoro algebra defined above is isomorphic to the one in [5] via the isomorphism that sends each (non-central) generator to the negative of the corresponding (non-central) generator in [5]. This explains the sign differences in the above statement and the statement of [5].

Let $V=U(\underline{\tilde{n}}^-)=\bigoplus_{m\in\mathbb{Z}_+}U(\underline{\tilde{n}}^-)_{-m}$. Then $\dim(U(\underline{\tilde{n}}^-)_{-m})<\infty$ for each $m\in\mathbb{Z}_+$. The map T_λ that sends $X\in U(\underline{\tilde{n}}^-)$ to Xv_λ is a linear isomorphism of V and $M(\lambda)$ for each $\lambda\in\underline{\tilde{h}}^*$ which extends to an $\underline{\tilde{h}}$-module isomorphism of $V\otimes\mathbb{C}(\lambda)$ and $M(\lambda)$. Using this map we can view each $M(\lambda)$ as a representation (τ_λ, V) of $\underline{\tilde{g}}$ equipped with a contravariant form B_0 induced from $(\ ,\ )_\lambda$.

Let $\lambda\in\underline{\tilde{h}}^*$ be arbitrary but fixed. Then $(\pi_t, V)=(\tau_{\lambda+t}, V)$ is a family of representations of $\underline{\tilde{g}}$ with contravariant forms B_t induced from $(\ ,\ )_{\lambda+t}$ and the following hold:

(1) $\pi_t(H)_{|V_\mu}=((\lambda+\mu)(H)+t\,\tilde{\alpha}(H))\,I$, $H\in\underline{\tilde{h}}$.

(2) If $v\in V_\mu$, $X\in\underline{\tilde{g}}$ then $t\mapsto\pi_t(X)\,v\in V_\mu+\sum_{m\in S}V_{\mu+m}$, $S\subset\mathbb{Z}$, S finite, is polynomial.

(3) $t\mapsto B_{t|V_\mu\times V_\mu}$ is polynomial and $B_{t|V_\mu\times V_\mu}$ is nondegenerate on $0<|t|<\varepsilon_\mu$.

Let $\mathcal{O}_0(V)^F$ denote the complex vector space of germs $\mathbf{f}$ of holomorphic functions $f\colon D_\varepsilon\to\bigoplus_{\mu\in S}V_\mu$ at 0, $S\subset\underline{\tilde{h}}^*$, S finite. Let for each $k\in\mathbb{N}$

$$\mathcal{O}_0^B(V)_k^F=\{\mathbf{f}\in\mathcal{O}_0(V)^F\,|\,B_t(f(t),w)\in t^k\,\mathcal{O}_0(\mathbb{C}),\ w\in V\}$$

where $\mathcal{O}_0(\mathbb{C})$ is the space of germs of holomorphic $\mathbb{C}$-valued functions at 0 as in §1. If $\mathbf{f}\in\mathcal{O}_0(V)^F$, $X\in\underline{\tilde{g}}$, $t\in D_\varepsilon$, we set $\pi(X)\,\mathbf{f}(t)=\pi_t(X)\,(f(t))$ and $\boldsymbol{\pi}(\mathbf{X})\,\mathbf{f}=\pi(X)\,\mathbf{f}$. If $f\in\mathcal{O}_0^B(V)_k^F$ then $B_t(\pi_t(X)\ f(t),w)=B_t(f(t),\sigma(\pi_t(X))\,w)$, $X\in U(\underline{\tilde{g}})$, $w\in V$, $t\in D_\varepsilon$. Hence $\pi(X)\,\mathcal{O}_0^B(V)_k^F\subset\mathcal{O}_0^B(V)_k^F$. Let

---

[3] See Footnote [1]

$$V_{0,k}^{B,F}=\{f(0)|\mathbf{f}\in\mathcal{O}_0^B(V)_k^F\}.$$

Then $\pi_0(X)\,V_{0,k}^{B,F}\subset V_{0,k}^{B,F}$ and

$$V=V_{0,0}^{B,F}\supset V_{0,1}^{B,F}\supset\ldots\supset V_{0,k}^{B,F}\supset\ldots$$

is a filtration of V by π_0-invariant subspaces. We now set $(V_{0,k}^{B,F})_\mu=V_{0,k}^{B,F}\cap V_\mu$ and define

$$\operatorname{ch} V_{0,k}^{B,F}=\sum_\mu \dim(V_{0,k}^{B,F})_\mu e^\mu,$$

where e^μ is a formal exponential. Note that in the notation of § 1, $(V_{0,k}^{B,F})_\mu=(V_\mu)_{0,k}^{B(\mu)}$ where $B_t(\mu)$ is the restriction of B_t to $V_\mu\times V_\mu$ and $B(\mu)(t)=B_t(\mu)$, $B(\mu)\in\mathcal{O}_0(\mathscr{L}_s^2(V_\mu,\mathbb{C}))$. We now set $M(\lambda)_{(k)}=T_\lambda(V_{0,k}^{B,F})$ and $\operatorname{ch} M(\lambda)_{(k)}=e^\lambda \operatorname{ch} V_{0,k}^{B,F}$. Applying lemmas § 1.2 and § 1.3 to V_μ and B_μ (see the remark in § 1) we obtain

Proposition 3. *Let $\lambda\in\underline{h}^*$. There is a filtration of submodules $M(\lambda)=M(\lambda)_{(0)}\supset M(\lambda)_{(1)}\supset\ldots$ such that*

(i) *$M(\lambda)_{(1)}$ is the largest proper submodule of $M(\lambda)$;*

(ii) *For every $k\in\mathbb{N}$, there is on $M(\lambda)_{(k)}/M(\lambda)_{(k+1)}$ a non-degenerate contravariant form;*

(iii) $\sum\limits_{k>0}\operatorname{ch} M(\lambda)_{(k)}=\sum\limits_{m\in\mathbb{Z}_+} e^{\lambda-m}\operatorname{ord}_0(\det(\ ,\ )_{\lambda+t,m})$.

Let $M(h,c)=M(\lambda)$ where $\lambda(E_0)=h$ and $\lambda(E_0')=c$. Using Proposition 3 and a simple computation we obtain

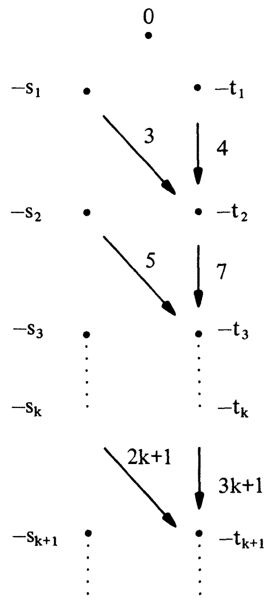
Proposition 4 [9, Proposition 6.4]. *Let for $k\in\mathbb{Z}_+$, $v=s_k$ or t_k, where $s_k=\frac{1}{2}(3k^2+k)$, $t_k=\frac{1}{2}(3k^2-k)$. Then*

$$\sum_{i>0}\operatorname{ch} M(-v,0)_{(i)}=\sum_{l\in\mathbb{N},\,l\,\mathrm{odd}}\operatorname{ch} M(-s_{k+l},0)+\operatorname{ch} M(-t_{k,l},0).$$

Remark. The filtration $(M(\lambda)_{(k)})_{k\in\mathbb{Z}_+}$ can also be derived using the technique of [4]. However the method used here is more general and applies to a situation (§ 4) needed in this article where the technique of [4] cannot be used directly.

§ 3. An Analogue to Shapovalov's Element

In this section we prove Theorem A (Introduction). We keep the notation of § 2 and let $s_k=\frac{1}{2}(3k^2+k)$, and $t_k=\frac{1}{2}(3k^2-k)$. We set $P_k=\{s_k,t_k\}$, $k\geqq 0$. Then $-t_{k+1}$ is the element in $-P_{k+1}$ closest to $-s_k$ and to $-t_k$ with $-s_k-(-t_{k+1})=2k+1$ and $-t_k-(-t_{k+1})=3k+1$. We consider the varieties $\psi_{1,2k+1}=0$ and $\psi_{1,3k+1}=0$ which are relevant to construct the elements giving the embeddings stated in Theorem A. These varieties correspond to hyperplanes in the semi-simple case and the embeddings considered correspond to the case of the transformation of a dominant integral weight by a simple root reflection in the semi-simple case. The following diagram is useful in keeping track of the embeddings considered above:



It is inspired by the Bruhat ordering diagram of [9, § 4].

We recall that, in the notation of § 2,

$$\begin{aligned}\psi_{r,s}(h,c) &= h^2-(\alpha^+_{r,s}(c)+\alpha^-_{r,s}(c))\,h+\alpha^+_{r,s}(c)\,\alpha^-_{r,s}(c)\\ &= h^2+\tfrac{1}{24}[(13-c)(r^2+s^2)-24rs-2+2c]\,h\\ &\quad+\frac{1}{(48)^2}\{[(13-c)(r^2+s^2)-24rs-2+2c]^2-(c^2-26c+25)(r^2-s^2)^2\}.\end{aligned}$$

Let $\mathscr{V}_{r,s}=\{(h,c)\in\mathbb{C}^2 \mid \psi_{r,s}(h,c)=0\}$.

For $k\in\mathbb{N}$ and $h\in\mathbb{C}$, $h\neq k$, we set

$$\beta_k(h)=24[h+\tfrac{1}{2}(3k^2+k)]\,[h+\tfrac{1}{3}(2k^2-k)]/(4k^2+4k)(h-k)$$

and for $k\in\mathbb{N}$ and $h\in\mathbb{C}$, $h\neq 3k$, we set

$$\gamma_k(h)=24[h+\tfrac{1}{2}(3k^2-k)]\,[h+\tfrac{1}{8}(27k^2+6k)]/(9k^2+6k)(2h-3k).$$

The following lemma is straightforward:

Lemma 1. *Let $k\in\mathbb{N}$. Then, for $h\neq k$, $(h,c)\in\mathscr{V}_{1,2k+1}$ if and only if $c=\beta_k(h)$, and for $h\neq\frac{3}{2}k$, $(h,c)\in\mathscr{V}_{1,3k+1}$ if and only if $c=\gamma_k(h)$.*

In particular, near (h,c), $h \neq k$, $\mathscr{V}_{1,2k+1}$ has a parametrization $(h, \beta_k(h))$ with $\beta_k(h)$ rational, and near (h,c), $h \neq \frac{3}{2}k$, $\mathscr{V}_{1,3k+1}$ has a parametrization $(h, \gamma_k(h))$ with $\gamma_k(h)$ rational.

We can now prove the

Lemma 2. *Let $k>0$ be fixed, $v \in \{-s_k, -t_k\}$. We set $\phi_{-s_k}(h) = \beta_k(h)$ and $\phi_{-t_k}(h) = \gamma_k(h)$. Let $m>0$. Then there is $\varepsilon_m > 0$ such that if $0<|h-v|<\varepsilon_m$ then $\psi_{r,s}(h, \phi_v(h)) \neq 0$ for all $1 \leqq r \leqq s$, $rs \leqq m$, $(r,s) \neq (1, v+t_{k+1})$.*

Proof. We enumerate the quadratic polynomials $\psi_{r,s}$ in the statement of the lemma as $\psi_1, \psi_2, \ldots, \psi_d$, and proceed to prove the lemma by contradiction. Therefore, since there are finitely many ψ_i, we assume that for some i, $1 \leqq i \leqq d$, there are infinitely many $h_j \in \mathbb{C}$ such that $\psi_i(\mathrm{h}_j, \phi_v(\mathrm{h}_j)) = 0$ for all j. Since ψ_i is polynomial and ϕ_v is rational this implies that $\psi_i(h, \phi_v(h)) = 0$ for all $h \in \mathbb{C}$. Therefore $\psi_i \in \mathbb{C}[h,c]\,\psi_{1,v+t_{k+1}}$. But the degrees of ψ_i and $\psi_{1,v+t_{k+1}}$ are the same, hence $\psi_i = a\psi_{1,v+t_{k+1}}$, $a \in \mathbb{C}$. Since the coefficient of h^2 is 1 in ψ_i and $\psi_{1,v+t_{k+1}}$ this implies that $\psi_i = \psi_{1,v+t_{k+1}}$, which contradicts our assumption on the ψ_i. Q.E.D.

Proof of Theorem A. We choose ε_v to be ε_m in Lemma 2 with $m = v + t_{k+1}$. Applying Lemma 2 we see that if $|h-v|<\varepsilon_v$ then $\psi_{r,s}(h, \phi_v(h)) \neq 0$ for all r,s such that $1 \leqq r \leqq s$, $rs \leqq v + t_{k+1}$, $(r,s) \neq (1, v+t_{k+1})$. Hence, if $|h-v|<\varepsilon_v$ then v is the only zero of $\det(\ ,\ )_{h, \phi_v(h), v+t_{k+1}}$ and it occurs with multiplicity $p((v+t_{k+1}) - (v+t_{k+1})) = p(0) = 1$, by Theorem 2.2. We now apply Lemma 1.4 to $B(h) = (\ ,\ )_{h, \phi_v(h), v+t_{k+1}}$ and $t_0 = v$ and obtain

$$\dim \operatorname{Rad}(\ ,\ )_{v,0,v+t_{k+1}} \leqq 1.$$

Since $v+t_{k+1}$ is the smallest natural number m such that $\det(\ ,\ )_{v,0,m} = 0$ ([9, Proposition 6.14]), we must have

$$M(v,0)_{-t_{k+1}} = \{v \in M(v,0)_{-t_{k+1}} \mid \tilde{n} v = 0\} = \operatorname{Rad}(\ ,\ )_{v,0,v+t_{k+1}}.$$

Therefore

$$\dim \operatorname{Rad}(\ ,\ )_{v,0,v+t_{k+1}} = 1.$$

Let A be as in Lemma 1.4, where $B(h) = (\ ,\ )_{h, \phi_v(h), v+t_{k+1}}$, $t_0 = v$. Applying Lemma 1.5 to A and $t_0 = v$ we obtain $\tilde{\theta}_v \in \mathcal{O}_v(V_{-v-t_{k+1}})$ such that $\tilde{\theta}_v(h) \in \operatorname{Rad}(\ ,\ )_{h, \phi_v(h), v+t_{k+1}} \setminus \{0\}$ for $h \in D_\varepsilon(v)$, $\varepsilon > 0$. Furthermore dim Rad$(\ ,\ )_{h, \phi_v(h), v+t_{k+1}} = 1$ for $h \in D_\varepsilon(v)$. This implies that $\underline{\tilde{n}}\, \tilde{\theta}_v(h)\, v_{h, \phi_v(h)} = 0$ for $h \in D_\varepsilon(v)$.

Let $u(h,c) = h$ and $v(h,c) = \psi_{1, v+t_{k+1}}(h,c)$. Then

$$\left(\frac{\partial(u,v)}{\partial(h,c)}\right) = \begin{pmatrix} 1 & 0 \\ * & \chi(v) \end{pmatrix}$$

where $\chi(v) = \frac{1}{24}(4k^2+4k)(h-k)$ if $v = -s_k$ and $\chi(v) = \frac{1}{24}(9\mathrm{k}^2+6k)(2h-3k)$ if $v = -t_k$. So $\left|\dfrac{\partial(u,v)}{\partial(h,c)}\right| \neq 0$ near $(v,0)$. We now set

$$\theta_\nu(h(u, v), c(u, v)) = \tilde{\theta}_\nu(u)$$

θ_ν clearly satisfies the conditions stated in Theorem A. This concludes the proof. Q.E.D.

§ 4. A Character Sum Formula for a Quotient of Verma Modules

In this section we prove Theorem B of the Introduction. We keep the notation of the preceeding section and let $\nu \in \{-s_k, -t_k\}$. Let $U \subset U(\underline{\tilde{n}}^-)$ be an $\underline{\tilde{h}}$-invariant subspace such that

$$U(\underline{\tilde{n}}^-) = U \oplus U(\underline{\tilde{n}}^-)\,\theta_\nu(\nu, 0).$$

Since θ_ν is holomorphic, the sum $U + U(\underline{\tilde{n}})\,\theta_\nu(h, \phi_\nu(h))$ is still direct for $|h-\nu| < \varepsilon$, $\varepsilon > 0$. We set

$$N(h) = N(h, \phi_\nu, \theta_\nu) = M(h, \phi_\nu(h))/U(\underline{\tilde{g}})\,\theta_\nu(h, \phi_\nu(h))\, v_{h, \phi_\nu(h)}.$$

Proof of Theorem B. Let $\bar{T}_h: U \to N(h)$ be the linear isomorphism defined by

$$\bar{T}_h(X) = X\, v_{h, \phi_\nu(h)} + U(\underline{\tilde{g}})\,\theta_\nu(h, \phi_\nu(h))\, v_{h, \phi_\nu(h)}$$

for $X \in U$, $|h-\nu| < \varepsilon$. $\bar{T}_h$ extends to an $\underline{\tilde{h}}$-module isomorphism of $U \otimes \mathbb{C}(h, \phi_\nu(h))$ and $N(h)$. Here we denote by $\mathbb{C}(h, c)$ the $U(\underline{\tilde{b}})$-module $\mathbb{C}(\lambda)$, where $\lambda(E_0) = h$ and $\lambda(E_0') = c$. Using this map we can define a representation $(\bar{\pi}_h, U)$ equivalent to $N(h)$ for $|h-\nu| < \varepsilon$. Now $(\ ,\ )_{h, \phi_\nu(h)}$ induces a contravariant form on $N(h)$, since $U(\underline{\tilde{g}})\,\theta_\nu(h, \phi_\nu(h))\, v_{h, \phi_\nu(h)}$ is contained in $\mathrm{Rad}(\ ,\ )_{h, \phi_\nu(h)}$ (see § 3). This contravariant form in turn induces a contravariant form $\bar{B}_h$, $|h-\nu| < \varepsilon$, on U. Therefore $(\bar{\pi}_h, U)$ is a family of representations of $\underline{\tilde{g}}$ with contravariant forms $\bar{B}_h$, $|h-\nu| < \varepsilon$, $\varepsilon > 0$, and the following hold:

(1) $\bar{\pi}_h(H)|_{U_\mu} = ((\lambda+\mu)(H))\, I$, $H \in \underline{\tilde{h}}$, where $\lambda(E_0) = h$, $\lambda(E_0') = \phi_\nu(h)$.

(2) If $v \in U_\mu$, $X \in \underline{\tilde{g}}$ then $h \mapsto \bar{\pi}_h(X)\, v \in U_\mu + \sum_{m \in S} U_{\mu - m}$, $S \subset \mathbb{Z}_+$, S finite, is holomorphic on $D_\varepsilon(\nu)$.

(3) $h \mapsto \bar{B}_{h|U_\mu \times U_\mu}$ is holomorphic on $D_\varepsilon(\nu)$.

$\bar{B}_{h,\mu} = \bar{B}_{h|U_\mu \times U_\mu}$ is non-degenerate on $0 < |h-\nu| < \varepsilon_\nu$. The second statement of (3) follows from Lemma 3.2. Proceedings exactly as in $\xi 2$ with U, $\bar{\pi}_h$ and $\bar{B}_h$ instead of V, π_t and B_t and ν instead of 0, we apply Lemmas 1.2 and 1.3 and obtain a filtration

$$N(\nu) = N(\nu)_{(0)} \supset N(\nu)_{(1)} \supset \dots$$

by submodules such that

(i) $N(\nu)_{(1)}$ is the largest proper submodule of $N(\nu)$.

(ii) For every $i \in \mathbb{N}$, there is on $N(\nu)_{(i)}/N(\nu)_{(i+1)}$ a non-degenerate contravariant form

(iii) $\sum_{i>0} \operatorname{ch} N(\nu)_{(i)} = \sum_{m \in \mathbb{Z}_+} e^{\nu-m} \operatorname{ord}_\nu \det \bar{B}_{h,-m}$ where $\operatorname{ch} N(\nu)_{(i)} = \sum_{\mu \in h^*} \dim(N(\nu)_{(i)})_\mu e^\mu$

and e a formal exponential.

To conclude the proof of Theorem B we need only to compute $\operatorname{ord}_\nu \det \bar{B}_{h,-m}$, $m \in \mathbb{Z}_+$.

We fix $m \in \mathbb{Z}_+$. Let $X_1, \ldots, X_{d_m}$ be a basis of U_{-m} and let $Y_1, \ldots, Y_{s_m}$ be a basis of $U(\underline{\tilde{n}}^-)_{-m+\nu+t_{k+1}}$. Let $D_{h-m,c}(h,c)$ denote the determinant of $(\ ,\ )_{h,c,m}$ relative to some basis of $M(h,c)_\mu$ where $\mu(E_0)=h-m$, $\mu(E_0')=c$. We denote by $\langle\ ,\ \rangle_{h,c}$ the contravariant form (denoted by B_0 in § 2, for h, c fixed) induced by $(\ ,\ )_{h,c}$ on V. Then

$$D_{h-m,c}(h,c) = q_\nu(h,c) \left[\begin{array}{c|c} \langle X_i, X_j\rangle_{h,c} & \langle X_i, Y_j\theta_\nu(h,c)\rangle_{h,c} \\ \hline \langle Y_j\theta_\nu(h,c), X_i\rangle_{h,c} & \langle Y_i\theta_\nu(h,c), Y_j\theta_\nu(h,c)\rangle_{h,c} \end{array}\right]$$

where q_ν is holomorphic near $(\nu, 0)$. We assume, without loss of generality, that $q_\nu(\nu, 0)=1$. Let $D'_{h-m,c}(h,c) = \det(\langle X_i, X_j\rangle_{h,c})$. We note that since for $X \in \underline{\tilde{n}}$, $X\theta_\nu(h(u,0), c(u,0))\, v_{h(u,0),c(u,0)} = 0$ then

$$\begin{aligned} X\theta_\nu(h(u,v), c(u,v))\, v_{h(u,v),c(u,v)} &= v(h,c)\, z(h,c)\, v_{h(u,v),c(u,v)} \\ &= \psi_{1,\nu+t_{k+1}}(h,c)\, z(h,c)\, v_{h(u,v),c(u,v)}, \end{aligned}$$

where $z(h,c)$ is holomorphic near $(\nu, 0)$. Using this and the definition of $\langle\ ,\ \rangle_{h,c}$ (§ 2) we obtain

$$\begin{aligned} D_{h-m,c}(h,c) = {} & q(h,c)\, D'_{h-m,c}(h,c) \det(\langle Y_i\theta_\nu(h,c), Y_j\theta_\nu(h,c)\rangle_{h,c}) \\ & + \psi_{1,\nu+t_{k+1}}(h,c)^{p(m-\nu-t_{k+1})+1}\, g(h,c) \end{aligned}$$

where g is holomorphic near $(\nu, 0)$. Now,

$$\langle Y_i\theta_\nu(h,c), Y_j\theta_\nu(h,c)\rangle_{h,c} = \langle \sigma(Y_j)\, Y_i\theta_\nu(h,c), \theta_\nu(h,c)\rangle_{h,c}.$$

Using the above remark we obtain:

$$\begin{aligned} & D_{h-m,c}(h,c) \\ & \quad = q(h,c)\, D'_{h-m,c}(h,c)\, D_{h-m+\nu+t_{k+1},c}(h-\nu-t_{k+1}, c)\, \langle \theta_\nu(h,c), \theta_\nu(h,c)\rangle_{h,c}^{p(m-\nu-t_{k+1})} \\ & \quad\quad + \psi_{1,\nu+t_{k+1}}(h,c)^{p(m-\nu-t_{k+1})+1}\, Q(h,c) \end{aligned}$$

where Q is holomorphic near $(\nu, 0)$. Therefore we have

$$\begin{aligned} & \left.\frac{D_{h-m,c}(h,c)}{\psi_{1,\nu+t_{k+1}}(h,c)^{p(m-\nu+t_{k+1})}}\right|_{(h,c)=(h,\phi_\nu(h))} \\ & \quad = q(h,\phi_\nu(h))\, D'_{h-m,\phi_\nu(h)}(h,\phi_\nu(h))\, D_{h-m+\nu+t_{k+1},\phi_\nu(h)}(h-\nu-t_{k+1}, \phi_\nu(h)) \\ & \quad\quad \cdot \left(\left.\frac{\langle \theta_\nu(h,c), \theta_\nu(h,c)\rangle_{h,c}}{\psi_{1,\nu+t_{k+1}}(h,c)}\right|_{(h,c)=(h,\phi_\nu(h))}\right)^{p(m-\nu-t_{k+1})} \end{aligned}$$

for h near v. From the proof of Theorem A it is clear that

$$\left.\frac{\langle\theta_v(h,c),\theta_v(h,c)\rangle_{h,c}}{\psi_{1,v+t_{k+1}}(h,c)}\right|_{(h,c)=(v,\phi_v(h))}\neq 0.$$

Thus

$$\begin{aligned}\operatorname{ord}_v D'_{h-m,\phi_v(h)}(h,\phi_v(h)) &= \operatorname{ord}_v D_{h-m,\phi_v(h)}(h,\phi_v(h)) - p(m-v-t_{k+1}) \\ &\quad - \operatorname{ord}_v D_{h-m+v+t_{k+1},\phi_v(h)}(h-v-t_{k+1},\phi_v(h)).\end{aligned}$$

Now,

$$\operatorname{ord}_v \det \bar{B}_{h,-m} = \operatorname{ord}_v D'_{h-m,\phi_v(h)}(h,\phi_v(h)),$$

$$\operatorname{ord}_v D_{h-m,\phi_v(h)}(h,\phi_v(h)) = \operatorname{ord}_0 \det(\ ,\ )_{v+t,0,m}$$

and

$$\operatorname{ord}_v D_{h-m+v+t_{k+1},\phi_v(h)}(h-v-t_{k+1},\phi_v(h)) = \operatorname{ord}_0 \det(\ ,\ )_{-t_{k+1}+t,0,m}.$$

The statements (i), (ii) and (iii) of Theorem B now follow as in the proof of Proposition 2.4. (iv) is obtained in the proof of Theorem C which is given in the next section. Q.E.D.

§ 5. The Characters of Irreducible Integral Highest Weight Modules

We are now ready to prove Theorem C stated in the Introduction, therefore completing the program initiated in [9] (cf. [7]).

We recall the results of [9] which lead to the proof of Theorem C presented below.

Theorem 1 ([9], cf. [7]). (i) *For each $k\geqq 1$ there exist $\underline{g}$-module embeddings I_k: $M(-s_k)\to M(0)$, J_k: $M(-t_k)\to M(0)$ such that* $\operatorname{im} I_k+\operatorname{im} J_k=\operatorname{im} I_{k-1}\cap\operatorname{im} J_{k-1}$, $k\geqq 2$.

(ii) *There exists a resolution of the trivial $\underline{g}$-module $\mathbb{C}$:*

$$(0)\ldots\to M(-s_k)\oplus M(-t_k)\xrightarrow{d_k}\ldots\xrightarrow{d_2} M(-s_1)\oplus M(-t_1)\xrightarrow{d_1} M(0)\xrightarrow{\varepsilon}\mathbb{C}\to 0$$

where $d_k(x,y)=(x+y,-(x+y))$, $k\geqq 2$, $d_1(x,y)=x+y$ and ε is the canonical surjective homomorphism. (For simplicity we identify $M(-s_k)$ (resp. $M(-t_k)$) with its image under I_k (resp. J_k) in $M(0)$).

Note. In [6] the existence of a resolution of type (1) is also stated, but no mention is made there to the differential maps used.

Using the identifications in Theorem 1 (ii) and letting $v\in\{-s_k,-t_k\}$, $k\in\mathbb{Z}_+$, we set, for $i\in\mathbb{N}$,

$$M(v)^i=M(-s_{k+i})+M(-t_{k+i}).$$

Corollary 2 ([9]). *$M(-s_k)+M(-t_k)=M(-s_{k-1})\cap M(-t_{k-1})$, $k\in\mathbb{N}$. If $v\in\{-s_k, -t_k\}$ then*

$$M(v)\supset M(v)^1\supset M(v)^2\supset\dots$$

is a $\underline{g}$-module filtration and

$$\sum_{i\geqq 1}\operatorname{ch}M(v)^i=\sum_{l\in\mathbb{N},\, l\,\mathrm{odd}}\operatorname{ch}M(-s_{k+l})+\operatorname{ch}M(-t_{k+l}).$$

Let $\tilde{L}(v)=M(v)/(M(-s_{k+1})+M(-t_{k+1}))$, $v\in\{-s_k, -t_k\}$.

Corollary 3 ([9]). *Let $v\in\{-s_k, -t_k\}$, $k\in\mathbb{Z}_+$. There exists a resolution of $\tilde{L}(v)$:*

$$\dots\to M(-s_i)\oplus M(-t_i)\xrightarrow{d_i}\dots\xrightarrow{d_{k+2}}M(-s_{k+1})\oplus M(-t_{k+1})$$

$$\xrightarrow{\eta_{k+1}}M(v)\xrightarrow{\varepsilon_k}\tilde{L}(v)\to 0$$

where the d_i are as in Theorem 1, $\eta_{k+1}(x,y)=x+y$ and ε_k is the quotient map.

Proof of Theorem C. We assert that $\tilde{L}(v)\simeq L(v)$.

Indeed, let $N(v)^1=(M(-s_{k+1})+M(-t_{k+1}))/M(-t_{k+1})\subset N(v)$ and $N(v)^i=0$ if $i>1$. Then $\operatorname{ch}N(v)/N(v)^1=\operatorname{ch}\tilde{L}(v)$, by Corollary 2. Now, $\sum_{i>0}\operatorname{ch}N(v)^i=\operatorname{ch}N(v)^1=\operatorname{ch}\tilde{L}(-s_{k+1})$ also by Corollary 2. By Theorem B and Corollary 2 we obtain

$$\begin{aligned}\sum_{i>0}\operatorname{ch}N(v)_{(i)}=&\sum_{i\in\mathbb{N},\, i\,\mathrm{odd}}(\operatorname{ch}M(-s_{k+i})+\operatorname{ch}M(-t_{k+i}))\\&-\sum_{i\in\mathbb{N},\, i\,\mathrm{odd}}(\operatorname{ch}M(-s_{k+i+1})+\operatorname{ch}M(-t_{k+i+1}))-\operatorname{ch}M(-t_{k+1})=\operatorname{ch}\tilde{L}(-s_{k+1}).\end{aligned}$$

Now, $N(v)_{(i)}\supset N(v)^i$ for all $i>0$, by Theorem B. Set $q_i=\operatorname{ch}(N(v)_{(i)}/N(v)^i)$ for $i>0$ and assume that $q_i>0$ for some $i>0$. Then $\sum_{i>0}\operatorname{ch}N(v)^i+\sum_{i>0}q_i=\sum_{i>0}\operatorname{ch}N(v)_{(i)}$. By the above, $\sum_{i>0}q_i=0$, which contradicts our assumption. Hence $q_i=0$ for all $i>0$. In particular, $N(v)^1=N(v)_{(1)}$. Hence $N(v)/N(v)^1$ is irreducible and so $\tilde{L}(v)$ is irreducible. This proves our assertion. Theorem C now follows from Corollary 3. Q.E.D.

Using ([9], Theorem 6.14) we obtain in particular

Corollary 4. *Let $v\in\{-s_k, -t_k\}$, $k\in\mathbb{Z}_+$. Then*

(i) $\operatorname{ch}L(v)=\operatorname{ch}M(v)+(-1)^k\sum_{i>k}(-1)^i(\operatorname{ch}M(-s_i)+\operatorname{ch}M(-t_i))$

(ii) $\operatorname{ch}M(-v)=\operatorname{ch}L(-v)+\sum_{i\in\mathbb{N}}(\operatorname{ch}L(-s_{k+i})+\operatorname{ch}L(-t_{k+i}))$.

Corollary 5. *Let $v\in\{-s_k, -t_k\}$, $k\in\mathbb{Z}_+$, and let $p\geqq 1$. Then $H_p(\underline{n}^-, L(v))=\mathbb{C}(-\frac{1}{2}(3(k+p)^2+(k+p))\oplus\mathbb{C}(-\frac{1}{2}(3(k+p)^2-(k+p))$ as an $\underline{h}$-module.*

Proof. Tensoring the $U(\underline{n}^-)$-free resolution (k) with $\mathbb{C}$ over $U(\underline{n}^-)$ we note that the resulting differential maps are all zero. The result is now obvious. Q.E.D.

Remark. Using the irreducibility of $L(v)$, $v\in\{-s_k, -t_k\}$, $k\in\mathbb{Z}_+$, and standard arguments, we may identify $H_p(\underline{n}^-, L(v))$ with $H^p(\underline{n}, L(v))$. Hence Corollary 5 also computes the cohomologies $\bar{H}^p(\underline{n}, L(v))$ as stated in the Introduction.

References

1. Garland, H., Lepowsky, J.: Lie algebra homology and the Macdonald-Kac formulas. Invent. Math. **34**, 37–76 (1976)
2. Gelfand, I.M.: The cohomology of infinite dimensional Lie algebras; some questions of integral geometry. Actes, Congrès intern. math., 1970, Tome 1, p. 95–111
3. Goncharova, L.V.: The cohomologies of Lie algebras of formal vector fields on the line. Functs. Anal. Prilozhen., **7**, No. 2, 6–14 (1973); No. 3, 33–44 (1973)
4. Jantzen, J.E.: Moduln mit einem höchsten Gewicht. Lecture Notes in Mathematics, vol. 750. Berlin-New-York-Heidelberg: Springer 1979
5. Kac, V.G.: Highest weight representations of infinite-dimensional lie algebras. Proceedings of the International Congress of Mathematicians, Helsinki, 1978
6. Kac, V.G.: Some problems on infinite dimensional Lie algebras and their representations (for AMS meeting at Amherst, 1981). Preprint
7. Rocha-Caridi, A.: Resolutions of irreducible highest weight modules over infinite dimensional graded Lie algebras, Proceedings of the 1981 Conference on Lie algebras and related topics. Lecture Notes in Mathematics, Vol. 933. Berlin-Heidelberg-New York: Springer 1982
8. Rocha-Caridi, A., Wallach, N.R.: Projective modules over graded Lie algebras I. Math. Z **180**, 151–177 (1982)
9. Rocha-Caridi, A., Wallach, N.R.: Highest weight modules over graded lie algebras: Resolutions. Filtrations and character formulas. To appear in the transactions of the A.M.S.
10. Shapovalov, N.N.: On a bilinear form on the universal enveloping algebra of a complex semisimple Lie algebra. Functional Anal. Appl. **6**, 307–312 (1972)

Oblatum 17-VII-1982

Note Added in Proof

1) A proof of Kac's formula, for the Shapovalov determinant has been recently given by B.L. Feigin and D.B. Fuchs, functs. Anal. prilozhen. 16, N° 2 (1982), 47–63.
2) The characters of *all* $L(h,0)$, $h\in\mathbb{C}$, as well as the characters of *all* $L(h,c)$, $h\in\mathbb{C}$, $c=25$, 26, have been recently obtained by the authors.

Invent. math. 71, 77–129

Inventiones mathematicae

Intersection Homology II

Mark Goresky [1]* and Robert MacPherson [2]**

[1] Department of Mathematics, Northeastern University, Boston, MA 02115, USA
[2] Department of Mathematics, Brown University, Providence, RI 02912, USA

In [19, 20] we introduced topological invariants $IH_*^{\bar{p}}(X)$ called intersection homology groups for the study of singular spaces X. These groups depend on the choice of a perversity $\bar{p}$: a perversity is a function from $\{2, 3, \ldots\}$ to the non-negative integers such that both $\bar{p}(c)$ and $c-2-\bar{p}(c)$ are positive and increasing functions of c (2.1). The group $IH_i^{\bar{p}}(X)$ is defined for spaces X called pseudomanifolds: a pseudomanifold of dimension n is a space that admits a stratification

$$X = X_n \supset X_{n-2} \supset X_{n-3} \supset \ldots \supset X_1 \supset X_0$$

such that $X_n - X_{n-2}$ is an oriented dense n manifold and $X_i - X_{i-1}$ for $i \leqq n-2$ is an i manifold along which the normal structure of X is locally trivial (§1.1).

The groups $IH_*^{\bar{p}}(X)$ are the total homology groups of a subcomplex $IC_*^{\bar{p}}(X)$ of the ordinary locally finite chains $C_*(X)$. We recall the definition ([20], §1.3) (which uses a fixed stratification of X)

$$IC_i^{\bar{p}}(X) = \left\{ \begin{array}{l} i \text{ chains } c \text{ that intersect each } X_{n-k} \text{ for } k>0 \text{ in a set of} \\ \text{dimension at most } i-k+\bar{p}(k) \text{ and whose boundary } \partial c \\ \text{intersects each } X_{n-k} \text{ for } k>0 \text{ in a set of dimension at} \\ \text{most } i-k-1+\bar{p}(k). \end{array} \right\}$$

Since the conditions $IC_*^{\bar{p}}(X)$ are local, the $IC_*^{\bar{p}}(U)$ for U open in X form a sheaf of chain complexes, denoted $\mathbf{IC}_*^{\bar{p}}(X)$. The purpose of this paper is to study this sheaf of chain complexes. Because sheaves of cochain complexes are more familiar, we renumber by $\mathbf{IC}_{\bar{p}}^{-i} = \mathbf{IC}_i^{\bar{p}}(X)$. By studying this complex of intersection chains, $\mathbf{IC}_{\bar{p}}^{\bullet}$, we obtain results about $IH_*^{\bar{p}}(X)$ because the hypercohomology group $\mathscr{H}^{-i}(\mathbf{IC}_{\bar{p}}^{\bullet})$ is $IH_i(X)$.

The change of point of view from the groups $IH_*^{\bar{p}}(X)$ to the sheaves $\mathbf{IC}^{\bullet}$ was suggested to us by Deligne and Verdier. It leads to many advantages, some of which we now list.

* Partially supported by National Science and Engineering Research Council of Canada and the Sloan Foundation

** Partially supported by National Science Foundation

(1) We consider $\mathbf{IC}^{\bullet}$ as an object in the derived category $D^b(X)$ of the category of sheaves of $\mathbb{Z}$ modules on X. This allows us to bring the functorial apparatus of $D^b(X)$ to bear. (This apparatus is reviewed in Chap. 1.)

(2) Motivated by the vanishing properties of the stalk homology of $\mathbf{IC}^{\bullet}$, Deligne gave [14] a second construction of it using the standard operations of sheaf theory. This construction is the key idea in this paper:

Theorem (§ 3.5). *Let* $\tau_{\leqq k}$: $D^b(X) \to D^b(X)$ *be the truncation functor* (§ 1.12) *which kills the stalk homology of degree* $>k$. *Suppose* U_k *is* $X - X_{n-k}$ *and* i_k *is the inclusion* $U_k \to U_{k+1}$. *Then there is a canonical isomorphism in* $D^b(X)$

$$\mathbf{IC}^{\bullet}(X) = \tau_{\leqq p(n)-n} R i_{n*} \dots \tau_{\leqq p(3)-n} R i_{3*} \tau_{\leqq p(2)-n} R i_{2*} \mathbb{Z}_{U_2}[n].$$

This construction works in any context where both sheaf theory and stratifications have been developed. So it produces intersection homology groups for topological pseudomanifolds (the approach of [20] required a piecewise linear structure) and for algebraic varieties in any characteristic.

(3) There is a stratification free characterization of $\mathbf{IC}^{\bullet}$.

Theorem (§4.1). *For any topological pseudomanifold* X, *there is a constructible complex* (§1.11) $\mathbf{IC}^{\bullet}$ *in* $D^b(X)$ *which is uniquely characterized up to canonical isomorphism in* $D^b(X)$ *by the conditions:*

(a) $\mathbf{IC}^{\bullet} | X - \Sigma = \mathbb{Z}_{X-\Sigma}[n]$ *for some subset* $\Sigma \subset X$ *of dimension* $n-2$.

(b) *The homology of every stalk vanishes in dimension* $< -n$.

(c) $\dim\{x \in X \mid H^m(\mathbf{IC}^{\bullet}_x) \neq 0\} \leqq n - \min\{c \mid p(c) = n+m\}$ *for all* $m \geqq -n+1$.

(d) $\dim\{x \in X \mid H^m_x(\mathbf{IC}^{\bullet}) \neq 0\} \leqq n - \min\{c \mid c-2-p(c) = -m\}$ *for all* $m \leqq -1$.

Here $\mathbf{IC}^{\bullet}_x$ is the stalk and $H^m_x(\mathbf{IC}^{\bullet})$ (§1.7) may be thought of as the compact support hypercohomology of a small open regular neighbourhood around x.

This yields the following axiomatic characterization of intersection homology, which does not depend on derived categories:

If $\mathbf{S}^{\bullet}$ is a constructible complex of fine sheaves satisfying (a) through (d) above, then the cohomology of the complex

$$\dots \to \Gamma(X; \mathbf{S}^{i-1}) \to \Gamma(X; \mathbf{S}^i) \to \Gamma(X; \mathbf{S}^{i+1}) \dots$$

is naturally isomorphic to $IH^{\bar{p}}_*(X)$.

This characterization implies the topological invariance of the intersection homology groups. In particular they are independent of the stratification of X.

(4) Sheaf theory allows one to give local (sheaf theoretic) expressions for global facts on hypercohomology. For example the canonical maps of intersection homology theory ($\bar{p}(c) \leqq \bar{q}(c)$ for all c)

$$H^*(X) \xrightarrow{\alpha} IH^{\bar{p}}_*(X) \xrightarrow{\eta} IH^{\bar{q}}_*(X) \xrightarrow{\omega} H_*(x)$$

can be defined locally by giving maps in $D^b(X)$ of the corresponding sheaves of cochain complexes (§5.1, §5.5). Since such a map can be completed to a distinguished triangle, the vanishing of the third term of the triangle gives a local criterion for α, η and ω to be isomorphisms (§5.5, §5.6).

Similarly, the intersection pairing (where $\bar{p}(c)+\bar{q}(c)=c-2$ for all c)

$$IH_*^{\bar{p}}(X) \otimes IH_*^{\bar{q}}(X) \to H_*(X)$$

results from a pairing of the corresponding objects in $D^b(X)$ (§5.2). Poincaré duality also has a local expression:

Theorem (§5.3). *If $\bar{p}(c)+\bar{q}(c)=c-2$ for all c, then if X is oriented*

$$\mathbf{IC}_{\bar{p}}^{\bullet} \otimes \mathbb{Q} \cong R\mathbf{Hom}^{\bullet}(\mathbf{IC}_{\bar{q}}^{\bullet}, \mathbb{D}_X^{\bullet}) \otimes \mathbb{Q}$$

where $\mathbb{D}_X^{\bullet}$ is the dualizing complex on X.

The Verdier duality theorem (§1.7) shows that this implies Poincaré duality on the hypercohomology

$$IH_*^{\bar{p}}(X) \otimes \mathbb{Q} \cong \mathrm{Hom}(IH_*^{\bar{q}}(X), \mathbb{Q})$$

provided X is compact.

There is a perversity $\bar{m}$ for which the intersection homology groups (with rational coefficients) are particularly important for the study of a complex analytic variety X. This is the *middle* perversity $\bar{m}(c)=\frac{c-2}{2}$. This makes sense since complex analytic varieties admit stratifications with only (real) even dimensional strata.

This paper contains several results on the middle group $IH_*^{\bar{m}}(X;\mathbb{Q})$. Among these are:

(1) Self duality: if $i+j=n$, then

$$IH_i^{\bar{m}}(X) \cong \mathrm{Hom}(IH_j^{\bar{m}}(X), \mathbb{Q})$$

this results from Verdier local duality as explained above.

(2) Kunneth theorem (§6.3):

$$IH_*^{\bar{m}}(X \times Y) \cong IH_*^{\bar{m}}(X) \otimes IH_*^{\bar{m}}(Y).$$

(3) Lefschetz hyperplane theorem:

Theorem (§7.1). *Suppose X is an n dimensional subvariety of complex projective space and H is a generic hyperplane. Then the map*

$$\alpha_*\colon IH_i^{\bar{m}}(X \cap H) \to IH_i^{\bar{m}}(X)$$

(*where α_* is the homomorphism induced by the normally nonsingular inclusion $X \cap H \to X$* (§5.4)) *is an isomorphism for $i<n-1$ and is a surjection for $i=n-1$.*

This paper contains three axiomatic characterizations of the complex of intersection chains $\mathbf{IC}_{\bar{p}}^{\bullet}$. Their ranges of validity are summarized as follows:

1. $[\mathrm{AX1}]_R$(§3.3) uses a (fixed) stratification
valid for any perversity $\bar{p}$
and any coefficient ring R.
2. [AX2] (§4.1) stratification independent
valid for any perversity $\bar{p}$
and any coefficient ring R.

3. [AX3] (§6.1) stratification independent
valid for middle perversity $\bar{m}$
and field coefficients.

We would like to thank J. Cheeger, W. Fulton, O. Gabber, S. Gelfand, C. McCrory, P. Siegel, J.-L. Verdier, and K. Vilonen for valuable conversations. We have especially profited from conversations with P. Deligne whose ideas are fundamental to this paper. A number of people have contributed corrections and improvements to the first draft of this paper. We mention in particular P. Deligne, A. Borel, and J.-L. Verdier. We are grateful to the IHES for their hospitality and support.

Table of Contents

Each chapter $n+1$ begins with a section §n.0 which specifies the assumptions on the space, its stratification (if any), the coefficient ring, and the perversity.

Intersection homology with coefficients in a local system is treated in a series of paragraphs marked $\mathscr{L}$, which are distributed throughout the paper.

Chapter 1. Sheaf Theory

This chapter (except for §1.1 and §1.2) consists of a summary (without proofs) of the theory of derived categories. We have included this material for the convenience of the reader. It is not necessary to absorb all of Chap. 1 before beginning to read this paper. The main references for this chapter are [3, 6, 17, 23, 25, 41, 42, 43]. Additional material on sheaf theory may be found in [4, 5, 9, 38, 39].

We describe a single version of the derived category of the category of sheaves (on the topological spaces which are defined in §1.1) which is closed under the standard operations of derived category theory, and which is rich enough for all the examples and applications we wish to consider.

1.1. *Topological Pseudomanifolds*

Definition. A 0-dimensional topologically stratified Hausdorff space is a countable collection of points with the discrete topology.

An n-dimensional *topological stratification* of a paracompact Hausdorff topological space X is a filtration by closed subsets

$$X=X_n\supset X_{n-1}\supset X_{n-2}\supset\ldots\supset X_1\supset X_0$$

such that for each point $p\in X_i-X_{i-1}$ there exists a distinguished neighborhood N of p in X, a compact Hausdorff space L with an $n-i-1$ dimensional topological stratification,

$$L=L_{n-i-1}\supset\ldots\supset L_1\supset L_0\supset L_{-1}=\phi$$

and a homeomorphism

$$\phi\colon \mathbb{R}^i\times\operatorname{cone}^\circ(L)\to N$$

which takes each $\mathbb{R}^i\times\operatorname{cone}^\circ(L_j)$ homeomorphically to $N\cap X_{i+j+1}$. Here, $\operatorname{cone}^\circ(L)$ denotes the open cone, $L\times[0,1)/(l,0)\sim(l',0)$ for all l, $l'\in L$. We use the convention that $L_{-1}=\phi$ and $\operatorname{cone}^\circ(\phi)=$one point.

Thus, if X_i-X_{i-1} is not empty, it is a manifold of dimension i. It is called the i-dimensional *stratum* of X and is denoted S_i.

If X admits a topological stratification then it is locally compact, satisfies conditions *hlc* and *clc* of Borel-Moore [6] and is a *cs* space in the sense of Siebenmann [36]. Every compact topologically stratified space can be embedded in Euclidean space.

A topologically stratified Hausdorff space X is *purely n-dimensional* if $X_n - X_{n-1}$ is dense in X. In this case, all notions of dimension (topological dimension [24], cohomological dimension [5] §1.5 p. 18, Borel-Moore dimension [6] coincide and equal n. A topologically stratified Hausdorff space X is purely n-dimensional if and only if every open subset of X is n-dimensional in any of the above senses.

Definition. A *topological pseudomanifold* of dimension n is a purely n-dimensional stratified paracompact Hausdorff topological space X which admits a stratification

$$X = X_n \supset \ldots \supset X_1 \supset X_0$$

such that $X_{n-1} = X_{n-2}$ (i.e., $S_n = X - X_{n-2}$ is an n-dimensional manifold which is dense in X). We use the symbol Σ to denote the singularity subset X_{n-2}.

The following types of spaces admit topological stratifications: complex algebraic varieties, complex analytic varieties, real analytic varieties, semi-algebraic and semi-analytic sets, subanalytic sets, Whitney stratified sets [40], abstract (or Thom-Mather) stratified sets [30], and piecewise linear spaces.

The following types of spaces are topological pseudomanifolds: irreducible (or equidimensional) complex algebraic or analytic varieties, the locus of points p in a normal n-dimensional real algebraic variety such that every neighborhood of p has topological dimension n, n-dimensional triangulated spaces so that each $n-1$ simplex is contained in exactly two n-simplices.

Throughout this paper we assume that all spaces are topological pseudomanifolds.

1.2. *Stratified Maps*

Let X and Y be stratified topological pseudomanifolds.

Definition. A continuous map $f\colon X \to Y$ is *stratified* if it satisfies the following two conditions:

(C1) For any connected component S of any stratum $Y_k - Y_{k-1}$, the set $f^{-1}(S)$ is a union of connected components of strata of X.

(C2) For each point $p \in Y_i - Y_{i-1}$ there exists a neighborhood N of p in Y_i, a topologically stratified space

$$F = F_k \supset F_{k-1} \supset \ldots \supset F_{-1} = \phi$$

and a stratum preserving homeomorphism

$$F \times N \to f^{-1}(N)$$

which commutes with the projection to N.

Remark. If $f\colon X \to Y$ is a subanalytic map between subanalytic sets then there exist stratifications of X and Y such that f is stratified. It is not possible however, to stratify the spaces in an arbitrary diagram of subanalytic (or even complex analytic) maps in such a way that each map is stratified.

1.3. *Complexes of Sheaves*

Let X be a topological pseudomanifold. Throughout this paper, R will denote a regular Noetherian ring with finite Krull dimension. (We shall be mainly concerned with the cases $R=\mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{C}$.)

Throughout this paper we shall use the word *sheaf* to mean a topological sheaf of R-modules. Sheaves on X will be denoted $\mathbf{A}$, $\mathbf{B}$, etc., and bounded complexes of sheaves on X will be denoted $\mathbf{A}^{\bullet}$, $\mathbf{B}^{\bullet}$, etc. The constant sheaf on X is denoted $\mathbf{R}_X$. Whenever convenient, we identify a sheaf $\mathbf{A}$ with the complex $\mathbf{A}^{\bullet}$ given by $\mathbf{A}^p=0$ for $p\neq 0$ and $\mathbf{A}^0=\mathbf{A}$.

Let X be a topological pseudomanifold and let $\mathbf{A}^{\bullet}$ be a bounded complex of sheaves of R-modules on X, i.e., a sequence

$$\dots\to\mathbf{A}^{p-1}\xrightarrow{d}\mathbf{A}^{p}\xrightarrow{d}\mathbf{A}^{p+1}\to\dots$$

with $d\circ d=0$ and $\mathbf{A}^p=0$ for $|p|$ sufficiently large. The sheaf of sections associated with $\mathbf{A}^{\bullet}$ assigns to any open set U the chain complex

$$\dots\to\Gamma(U;\mathbf{A}^{p-1})\to\Gamma(U;\mathbf{A}^{p})\to\Gamma(U;\mathbf{A}^{p+1})\to\dots.$$

The p^{th} cohomology sheaf $\mathbf{H}^p(\mathbf{A}^{\bullet})$ associated with $\mathbf{A}^{\bullet}$ is the sheafification of the presheaf whose sections over an open set U is the p^{th} homology group of this chain complex. The stalk at a point $x\in X$ of the sheaves $\mathbf{A}^p$ and $\mathbf{H}^p(\mathbf{A}^{\bullet})$ are denoted $\mathbf{A}^p_x$ and $\mathbf{H}^p(\mathbf{A}^{\bullet})_x$ respectively. In particular $\mathbf{H}^p(\mathbf{A}^{\bullet}_x)\cong\mathbf{H}^p(\mathbf{A}^{\bullet})_x$. The complex $\mathbf{A}[n]$ is defined by $\mathbf{A}[n]^p=\mathbf{A}^{p+n}$. The restriction of $\mathbf{A}^{\bullet}$ to a subspace $Y\subset X$ is denoted $\mathbf{A}^{\bullet}|Y$.

1.4. *Constructible Sheaves*

Definition. A sheaf $\mathbf{S}$ on X is called locally constant if every point $x\in X$ has a neighborhood U such that the restriction maps

$$\mathbf{S}_x\leftarrow\Gamma(U;\mathbf{S})\to\mathbf{S}_y$$

are isomorphisms for all $y\in U$.

A complex of sheaves is called *cohomologically locally constant* (CLC) if the associated local cohomology sheaves are locally constant.

Let $X_0\subset X_1\subset\dots\subset X_n=X$ be a filtration by closed subsets. A complex of sheaves $\mathbf{A}^{\bullet}$ on X is said to be *constructible* with respect to this filtration if, for each j, $\mathbf{A}^{\bullet}|(X_j-X_{j-1})$ is CLC, and has finitely generated stalk cohomology.

If X has a subanalytic structure, then the complex $\mathbf{A}^{\bullet}$ is said to be *subanalytically constructible* if it is constructible with respect to some filtration of X by closed subanalytic subsets. One defines *PL*-constructible and algebraically constructible complexes of sheaves similarly. $\mathbf{A}^{\bullet}$ is *topologically constructible* if it is bounded and is constructible with respect to some topological stratification of X.

Note. All complexes of sheaves considered in this paper will be topologically constructible.

If $\mathbf{A}^{\bullet}$ is a complex of sheaves on X which is constructible with respect to a given stratification and if $f\colon X \to Y$ is a stratified map then $f_* \mathbf{A}^{\bullet}$ is constructible with respect to the given stratification of Y. A similar remark holds for f^*.

Theorem. *Suppose X is a topological pseudomanifold and $\mathbf{A}^{\bullet}$ is a topologically constructible complex of sheaves on X. Then $\mathbf{A}^{\bullet}$ is* perfect *in the sense of* [3] *Exp.* 9, cohomologically constructible *in the sense of Verdier* [42, 45], *and satisfies* condition (P, Q) of Wilder ([47, 6]). *Therefore $\mathcal{H}^i(X; \mathbf{A}^{\bullet})$ is finitely generated, when X is compact.* ([6] *Prop.* 6.8.)

Proof. We claim that for any $x \in X$ there is a neighborhood basis $U_1 \supset U_2 \supset U_3 \supset \ldots$ such that for each i and m, the restriction map

$$\mathcal{H}^i(U_m; \mathbf{A}^{\bullet}) \to \mathcal{H}^i(U_{m+1}; \mathbf{A}^{\bullet})$$

is an isomorphism. It follows that $\mathcal{H}^i(U_m; \mathbf{A}^{\bullet}) \cong \mathbf{H}^i(\mathbf{A}^{\bullet})_x$, but $\mathbf{H}^i(\mathbf{A}^{\bullet})_x$ is a finitely generated R-module.

By [3] Exp. 9 §5.1 this (plus the fact that R is a regular Noetherian ring) will imply $\mathbf{A}^{\bullet}$ is perfect.

By [6] Prop. 6.8 this will imply $\mathbf{A}^{\bullet}$ satisfies the Wilder condition, which implies $\mathcal{H}^i(X, \mathbf{A}^{\bullet})$ is finitely generated for compact X.

By [V] Theorem 8, this will imply $\mathbf{A}^{\bullet}$ satisfies condition CC of Verdier.

Proof of Claim. Fix $x \in X$ and let N be a distinguished neighborhood $N \cong \mathbb{R}^i \times \mathrm{cone}^\circ(L)$ as in §1.1. Let Y be the join $S^{i-1} * L$, stratified in the obvious way (i.e., $S^{i-1} \subset S^{i-1} * L$ is a stratum but $L \subset S^{i-1} * L$ is not a union of strata unless $i=0$). This determines a stratification of the open cone,

$$\mathrm{cone}^\circ(Y) = Y \times [0, 1)/(y, 0) \sim (y', 0) \qquad \text{for all } y, y' \in Y.$$

Choose a stratum preserving homeomorphism $\psi\colon \mathrm{cone}^\circ(Y) \to N$, with $\psi(\text{vertex}) = x$. Let $U_m \subset N$ be the smaller neighborhood $U_m = \psi\left(Y \times \left[0, \frac{1}{m}\right)\right)$ we will now verify the claim for $m=1$, the other cases being very similar. Define a 1-parameter family of stratum preserving stretching embeddings,

$$G\colon\ U_2 \times [\tfrac{1}{2}, 1] \to U_1$$

$$G((y, t), s) = (y, t/s).$$

For any $s \in [\frac{1}{2}, 1]$ let $i_s\colon U_2 \to U_2 \times [\frac{1}{2}, 1]$ be the inclusion at the level s. We must show that

$$\mathcal{H}^*(U_2; i_1^* G^* \mathbf{A}^{\bullet}) \cong \mathcal{H}^*(U_2; i_{\frac{1}{2}}^* G^* \mathbf{A}^{\bullet}).$$

In fact, $i_1^* G^* \mathbf{A}^{\bullet}$ and $i_{\frac{1}{2}}^* G^* \mathbf{A}^{\bullet}$ are both quasi-isomorphic to $R\pi_* G^* \mathbf{A}^{\bullet}$ where $\pi\colon U_2 \times [\frac{1}{2}, 1] \to U_2$ is the projection to the second factor. To see this, consider the effect on stalk cohomology of the natural map

$$R\pi_* G^* \mathbf{A}^{\bullet} \to R\pi_* i_{s*} i_s^* G^* \mathbf{A}^{\bullet} = i_s^* G^* \mathbf{A}^{\bullet}.$$

For any $x'=(y,t)\in U_2$ the stalk cohomology of $i_s^* G^* \mathbf{A}^{\bullet}$ is simply $\mathbf{H}^{\bullet}(\mathbf{A}^{\bullet})_{(g,t/s)}$. However, by [17] 4.17.1 the stalk cohomology of $R\pi_* G^* \mathbf{A}^{\bullet}$ is $\mathscr{H}^*(\pi^{-1}(x'); G^* \mathbf{A}^{\bullet})$. Since $\pi^{-1}(x')$ is an interval, the following lemma now implies that this natural map is an isomorphism on stalk cohomology (where $\mathbf{B}^{\bullet} = G^* \mathbf{A}^{\bullet}$).

Lemma. *Let $\mathbf{B}^{\bullet}$ be a complex of sheaves on the unit interval I. Suppose the cohomology sheaves $\mathbf{H}^q(\mathbf{B}^{\bullet})$ are locally constant. Then for any $t\in I$, the restriction map*

$$\mathscr{H}^q(I;\mathbf{B}^{\bullet}) \to H^q(\mathbf{B}^{\bullet})_t$$

is an isomorphism.

Proof of Lemma. The sheaves $\mathbf{H}^q(\mathbf{B}^{\bullet})$ are constant on I. The spectral sequence for the hypercohomology of $\mathbf{B}^{\bullet}$ collapses, since

$$E_2^{p,q} = H^p(I; \mathbf{H}^q(\mathbf{B}^{\bullet})) = H^p(I; \mathbf{H}^q(\mathbf{B}^{\bullet})_t) = 0$$

unless $p=0$ and $E_2^{0,q} = \mathbf{H}^q(\mathbf{B}^{\bullet})_t$.

1.5. *Quasi-Isomorphisms and Injective Resolutions*

A sheaf map ϕ: $\mathbf{A}^{\bullet} \to \mathbf{B}^{\bullet}$ which commutes with the differentials, is called a quasi-isomorphism if the induced map $\mathbf{H}^p(\phi)$: $\mathbf{H}^p(\mathbf{A}^{\bullet}) \to \mathbf{H}^p(\mathbf{B}^{\bullet})$ is an isomorphism for each p. If ϕ is a quasi-isomorphism and if each $\mathbf{B}^p$ is injective, then $\mathbf{B}^{\bullet}$ is called an injective resolution of $\mathbf{A}^{\bullet}$. Injective resolutions exist for any complex of sheaves of R-modules and are uniquely determined up to chain homotopy.

We now recall the "canonical" bounded injective resolution of a bounded complex of sheaves ([6] §1.3, [4] p. 32, [17] *I* §1.4, *II* §7.1). First we describe the canonical resolution of a sheaf $\mathbf{B}$.

For each $x\in X$ let $\mathbf{B}_x \to I(x)$ be the canonical embedding of the stalk of $\mathbf{B}$ into an injective R-module, as described in [4] p. 32 and [17] I §1.4. (If R is a field, $\mathbf{B}_x$ will already be injective, so we may simplify the construction by taking $I(x) = \mathbf{B}_x$.) We obtain a canonical embedding of $\mathbf{B}$ into the injective sheaf $\mathbf{I}^0$ where

$$\mathbf{I}^0(U) = \prod_{x\in U} I(x).$$

Similarly the cokernel of $\mathbf{B} \to \mathbf{I}^0$ has a canonical embedding into an injective sheaf $\mathbf{I}^1$. Continuing this way gives a resolution

$$\mathbf{B} \to \mathbf{I}^0 \xrightarrow{d^0} \mathbf{I}^1 \to \dots.$$

By [44] and [3] Exp. 2 Theorem 4.3, the sheaf $\mathbf{ker}\, d^{p+n+1}$ is injective (where $n = \dim(X)$ and $p = \dim(R)$). Define the canonical resolution of $\mathbf{B}$ to be

$$0 \to \mathbf{I}^0 \to \mathbf{I}^1 \to \dots \to \mathbf{I}^{p+n} \to \mathbf{ker}\, d^{p+n+1} \to 0.$$

This construction is functorial in $\mathbf{B}$.

For any bounded complex of sheaves $\mathbf{A}^{\bullet}$, let

$$0 \to \mathbf{J}^{m,0} \to \mathbf{J}^{m,1} \to \dots \to \mathbf{J}^{m,p+n+1} \to 0$$

be the canonical injection resolution of $\mathbf{A}^m$. These $\mathbf{J}^{m,r}$ form a double complex with differentials $\mathbf{J}^{m,r} \to \mathbf{J}^{m+1,r}$ induced from the differential $\mathbf{A}^m \to \mathbf{A}^{m+1}$. The *canonical injective resolution* $\mathbf{A}^{\bullet} \to \mathbf{I}^{\bullet}$ is the single complex

$$\mathbf{I}^p = \bigoplus_{m+r=p} \mathbf{J}^{m,r}$$

which is associated to this double complex [17] I §2.6.

1.6. *Hypercohomology*

The p^{th} *hypercohomology group* $\mathscr{H}^p(X;\mathbf{A}^{\bullet})$ of a complex of sheaves $\mathbf{A}^{\bullet}$ is defined to be the p^{th} cohomology group of the cochain complex

$$\dots \to \Gamma(X;\mathbf{I}^{p-1}) \to \Gamma(X;\mathbf{I}^p) \to \Gamma(X;\mathbf{I}^{p+1}) \to \dots$$

where $\mathbf{I}^{\bullet}$ is the canonical injective resolution of $\mathbf{A}^{\bullet}$. This group is naturally isomorphic to the p^{th} cohomology group of the single complex which is associated to the double complex $C^p(X;\mathbf{A}^q)$ [17] II §4.6.

The double complex $C^p(X;\mathbf{A}^q)$ gives rise to a spectral sequence for hypercohomology, with

$$E^2_{pq} = H^p(X;\mathbf{A}^q) \to \mathscr{H}^{p+q}(X;\mathbf{A}^{\bullet}).$$

1.7. *Sheaf* Hom

If $\mathbf{A}$ and $\mathbf{B}$ are sheaves on X, let $\mathrm{Hom}(\mathbf{A},\mathbf{B})$ denote the abelian group of all (global) sheaf maps $\mathbf{A} \to \mathbf{B}$. Let $\mathbf{Hom}(\mathbf{A},\mathbf{B})$ be the sheaf whose sections over an open set U are $\Gamma(U;\mathbf{Hom}(\mathbf{A},\mathbf{B})) = \mathrm{Hom}(\mathbf{A}|U,\mathbf{B}|U)$. If $\mathbf{A}^{\bullet}$ and $\mathbf{B}^{\bullet}$ are complexes of sheaves, let $\mathbf{Hom}^{\bullet}(\mathbf{A}^{\bullet},\mathbf{B}^{\bullet})$ be the single complex of sheaves which is obtained from the double complex $\mathbf{Hom}^{p,q}(\mathbf{A}^{\bullet},\mathbf{B}^{\bullet}) = \mathbf{Hom}(\mathbf{A}^p,\mathbf{B}^q)$ in the usual way.

A cocycle $\xi \in \Gamma(X;\mathbf{Hom}^k(\mathbf{A}^{\bullet},\mathbf{B}^{\bullet}))$ is a chain map from $\Gamma(X;\mathbf{A}^{\bullet})$ to $\Gamma(X;\mathbf{B}^{\bullet}[k])$ which commutes with the differentials of $\mathbf{A}^{\bullet}$ and $\mathbf{B}^{\bullet}$. ξ is a coboundary if it is chain homotopic to 0.

If $\mathbf{S}_1$ and $\mathbf{S}_2$ are sheaves, we may consider them to be complexes in degree 0 (i.e. $S_1^p = 0$ for $p \neq 0$) with $d=0$. Then the sheaf $\mathbf{Ext}^i(\mathbf{S}_1,\mathbf{S}_2)$ is equal to the i^{th} cohomology sheaf associated to $\mathbf{Hom}^{\bullet}(\mathbf{S}_1,\mathbf{I}^{\bullet})$ where $\mathbf{I}^{\bullet}$ is an injective resolution of $\mathbf{S}_2$.

1.8. *The Constructible Derived Category* [42, 43, 23]

Let $K(X)$ denote the category whose objects $\mathbf{A}^{\bullet}$ are topologically constructible bounded complexes of sheaves on X and whose morphisms ϕ: $\mathbf{A}^{\bullet} \to \mathbf{B}^{\bullet}$ are homotopy classes of sheaf maps which commute with the differentials. Let

$INJ(X)$ denote the category of constructible bounded complexes of injective sheaves and chain homotopy classes of sheaf maps. There is an equivalence of categories between $INJ(X)$ and the constructible *derived* category $D^b(X)$, induced by the canonical functors

$$INJ(X) \to K(X) \to D^b(X).$$

An object in $D^b(X)$ is a doubly bounded complex of topologically constructible sheaves. A morphism in $D^b(X)$ from $\mathbf{A}^\bullet$ to $\mathbf{B}^\bullet$ is an equivalence class of diagrams of chain maps $\mathbf{A}^\bullet \leftarrow \mathbf{C}^\bullet \to \mathbf{B}^\bullet$ where $\mathbf{A}^\bullet \leftarrow \mathbf{C}^\bullet$ is a quasi-isomorphism. Two such diagrams

$$\mathbf{A}^\bullet \xleftarrow{f_1} \mathbf{C}_1^\bullet \xrightarrow{g_1} \mathbf{B}^\bullet, \qquad \mathbf{A}^\bullet \xleftarrow{f_2} \mathbf{C}_2^\bullet \xrightarrow{g_2} \mathbf{B}^\bullet$$

are considered equivalent if there is a homotopy commutative diagram

$$\begin{tikzcd}
& \mathbf{C}_1^\bullet \arrow[dl, "f_1"'] \arrow[dr, "g_1"] & \\
\mathbf{A}^\bullet & \mathbf{C}^\bullet \arrow[l, "f"'] \arrow[r] \arrow[u] \arrow[d] & \mathbf{B}^\bullet \\
& \mathbf{C}_2^\bullet \arrow[ul, "f_2"] \arrow[ur, "g_2"'] &
\end{tikzcd}$$

where f is a quasi-isomorphism.

Two complexes $\mathbf{A}^\bullet$ and $\mathbf{B}^\bullet$ are isomorphic in $D^b(X)$ if there is a morphism $\mathbf{A}^\bullet \leftarrow \mathbf{C}^\bullet \to \mathbf{B}^\bullet$ where both arrows are quasi-isomorphisms. In this case we say $\mathbf{A}^\bullet$ and $\mathbf{B}^\bullet$ are different *incarnations* of the same isomorphism class of objects in $D^b(X)$.

If $\mathbf{B}^\bullet$ is injective, then

$$\mathrm{Hom}_{D^b(X)}(\mathbf{A}^\bullet, \mathbf{B}^\bullet) = H^0(X; \mathbf{Hom}^\bullet(\mathbf{A}^\bullet, \mathbf{B}^\bullet)).$$

Remark. Although a chain map ϕ: $\mathbf{A}^\bullet \to \mathbf{B}^\bullet$ which induces isomorphisms on the associated cohomology sheaves becomes an isomorphism in $D^b(X)$, there exist sheaf maps which induce the 0 map on cohomology but which are not 0 in $D^b(X)$. (However, see §1.13.)

1.9. *Derived Functors*

A covariant additive functor T from complexes of sheaves to an abelian category gives rise to its right derived functor RT defined on $D^b(X)$ by the formula

$$RT(\mathbf{A}^\bullet) = T(\mathbf{I}^\bullet)$$

where $\mathbf{I}^\bullet$ is the canonical injective resolution of $\mathbf{A}^\bullet$ (see § 1.5).

This procedure applies to the functors $\mathbf{Hom}(\mathbf{B}^{\bullet},*)$, Γ (global sections), Γ_c (global sections with compact support), f_* (direct image), and $f_!$ (direct image with proper supports), where $f\colon X\to Y$ is a continuous map. For a closed subspace $Z\subset X$, the functor Γ_Z assigns to any complex of sheaves $\mathbf{S}^{\bullet}$ the complex of global sections of $\mathbf{S}^{\bullet}$ which vanish on $X-Z$. The i^{th} cohomology group of $R\Gamma_Z(\mathbf{S}^{\bullet})$ is denoted $\mathscr{H}^i_Z(\mathbf{S}^{\bullet})$.

In certain cases we may substitute a simpler resolution for $\mathbf{I}^{\bullet}$ (Hartshorne [23]). If $T=\Gamma$ then $RT(\mathbf{A}^{\bullet})=T(\mathbf{J}^{\bullet})$ where $\mathbf{J}^{\bullet}$ is a flabby or a fine resolution of $\mathbf{A}^{\bullet}$. If $T=\mathbf{Hom}^{\bullet}(\mathbf{B}^{\bullet},*)$ we may take $\mathbf{I}^{\bullet}$ to be any flabby or fine resolution of $\mathbf{A}^{\bullet}$ by sheaves of *injective* R-modules. Since f^* is exact we have $Rf^*(\mathbf{A}^{\bullet})\cong Lf^*(\mathbf{A}^{\bullet})\cong f^*(\mathbf{A}^{\bullet})$. If f is an inclusion of a subspace then $f_!$ is exact so $Rf_!(\mathbf{A}^{\bullet})\cong f_!(\mathbf{A}^{\bullet})$.

Define $\mathbf{A}^{\bullet}\otimes\mathbf{B}^{\bullet}$ to be the single complex which is associated to the double complex $\mathbf{A}^p\otimes\mathbf{B}^q$. Define the derived functor $\mathbf{A}^{\bullet}\overset{L}{\otimes}?$ by the formula

$$\mathbf{A}^{\bullet}\overset{L}{\otimes}\mathbf{B}^{\bullet}=\mathbf{A}^{\bullet}\otimes\mathbf{J}^{\bullet}$$

where $\mathbf{J}^{\bullet}\to\mathbf{B}^{\bullet}$ is a resolution of $\mathbf{B}^{\bullet}$ whose stalks are flat R-modules. If R is a field then $\mathbf{A}^{\bullet}\overset{L}{\otimes}\mathbf{B}^{\bullet}\cong\mathbf{A}^{\bullet}\otimes\mathbf{B}^{\bullet}$.

To verify that these functors are defined on the constructible derived category, we need the following

Proposition. *If $\mathbf{A}^{\bullet}$ and $\mathbf{B}^{\bullet}\in D^b(X)$ are constructible with respect to a given stratification of X, then so are*

$$R\mathbf{Hom}^{\bullet}(\mathbf{A}^{\bullet},\mathbf{B}^{\bullet}) \quad \textit{and} \quad \mathbf{A}^{\bullet}\overset{L}{\otimes}\mathbf{B}^{\bullet}.$$

Furthermore $Rf_*\mathbf{A}^{\bullet}$ and $Rf_!\mathbf{A}^{\bullet}$ are constructible with respect to a given stratification of Y, whenever $f\colon X\to Y$ is stratified with respect to these stratifications.

1.10. *Triangles* ([23] p. 20, 32)

$D^b(X)$ is not an abelian category, but it has "distinguished triangles" as a replacement for exact sequences. A triangle of morphisms in $D^b(X)$,

$$\begin{tikzcd} \mathbf{A}^{\bullet} \arrow[rr] & & \mathbf{B}^{\bullet} \arrow[dl] \\ & \mathbf{C}^{\bullet} \arrow[ul, "{[1]}"] & \end{tikzcd}$$

is called distinguished if it is isomorphic (in $D^b(X)$) to a diagram of sheaf maps

$$\begin{tikzcd} \tilde{\mathbf{A}}^{\bullet} \arrow[rr, "\phi"] & & \tilde{\mathbf{B}}^{\bullet} \arrow[dl] \\ & \mathbf{M}^{\bullet} \arrow[ul, "{[1]}"] & \end{tikzcd}$$

where $\mathbf{M}^\bullet$ is the (algebraic) mapping cone of ϕ and $\tilde{\mathbf{B}}^\bullet \to \mathbf{M}^\bullet \to \tilde{\mathbf{A}}^\bullet[1]$ are the canonical maps.

For example, a short exact sequence of complexes of sheaves becomes a distinguished triangle in $D^b(X)$.

Any edge of a distinguished triangle determines the third member up to (non canonical) isomorphism in $D^b(X)$.

A distinguished triangle determines a long exact sequence on the associated cohomology sheaves,

$$\to \mathbf{H}^p(\mathbf{A}^\bullet) \to \mathbf{H}^p(\mathbf{B}^\bullet) \to \mathbf{H}^p(\mathbf{C}^\bullet) \to \mathbf{H}^{p+1}(\mathbf{A}^\bullet) \to \mathbf{H}^{p+1}(\mathbf{B}^\bullet) \to$$

as well as a long exact sequence on hypercohomology.

If $\mathbf{F}^\bullet$ is a complex of sheaves, and

$$\begin{tikzcd} \mathbf{A}^\bullet \arrow[rr] & & \mathbf{B}^\bullet \arrow[dl] \\ & \mathbf{C}^\bullet \arrow[ul, "{[1]}"] & \end{tikzcd}$$

is a distriangle in $D^b(X)$, then we have distinguished triangles

$$\begin{tikzcd} R\mathbf{Hom}^\bullet(\mathbf{F}^\bullet, \mathbf{A}^\bullet) \arrow[rr] & & R\mathbf{Hom}^\bullet(\mathbf{F}^\bullet, \mathbf{B}^\bullet) \arrow[dl] \\ & R\mathbf{Hom}^\bullet(\mathbf{F}^\bullet, \mathbf{C}^\bullet) \arrow[ul, "{[1]}"] & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} R\mathbf{Hom}^\bullet(\mathbf{A}^\bullet, \mathbf{F}^\bullet) \arrow[dr, "{[1]}"'] & & R\mathbf{Hom}^\bullet(\mathbf{B}^\bullet, \mathbf{F}^\bullet) \arrow[ll] \\ & R\mathbf{Hom}^\bullet(\mathbf{C}^\bullet, \mathbf{F}^\bullet) \arrow[ur] & \end{tikzcd}$$

1.11. *Exact Sequence of a Pair*

Let j: $Y \to X$ be the inclusion of a *closed* subspace. Denote by i: $U \to X$ the inclusion of the open complement. If $\mathbf{A}^\bullet$ is a complex of sheaves on X, there are distinguished triangles in $D^b(X)$,

$$\begin{tikzcd} R i_! i^* \mathbf{A}^\bullet \arrow[rr] & & \mathbf{A}^\bullet \arrow[dl] \\ & R j_* j^* \mathbf{A}^\bullet \arrow[ul, "{[1]}"] & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} R j_* j^! \mathbf{A}^\bullet \arrow[rr] & & \mathbf{A}^\bullet \arrow[dl] \\ & R i_* i^* \mathbf{A}^\bullet \arrow[ul, "{[1]}"] & \end{tikzcd}$$

The second triangle can be obtained from the first by Verdier duality. (see §1.12) In the case $\mathbf{A}^\bullet \cong \mathbb{Z}_X$ the hypercohomology exact sequences are simply the long exact cohomology sequences for $H^*(X, Y)$ and $H^*(X, U)$ respectively. If $\mathbf{A}^\bullet \cong \mathbb{D}^\bullet_X$ we obtain the long exact homology sequences for $H_*(X, U)$ and $H_*(X, Y)$ respectively.

1.12. *Duality*

Let X be a topological pseudomanifold.

In [6] Borel and Moore defined the dual $\mathfrak{D}(\mathbf{A}^\bullet)$ of any complex of sheaves $\mathbf{A}^\bullet$ and showed (when R is a Dedeking ring) that for any open set $U \subset X$ the cohomology groups $\mathscr{H}_c^i(U:\mathbf{A}^\bullet)$ and $\mathscr{H}^i(U;\mathfrak{D}(\mathbf{A}^\bullet))$ are dual. This means there is a natural (split) exact sequence

$$0\to \operatorname{Ext}(\mathscr{H}_c^{q+1}(U;\mathbf{A}^\bullet),R)\to \mathscr{H}^{-q}(U;\mathfrak{D}(\mathbf{A}^\bullet))$$
$$\to \operatorname{Hom}(\mathscr{H}_c^{q}(U;\mathbf{A}^\bullet),R)\to 0.$$

(Here, $\mathscr{H}_c^q$ denotes the hypercohomology with compact supports, i.e., $R^q\Gamma_c$.) In fact, this property characterizes $\mathfrak{D}(\mathbf{A}^\bullet)$ up to quasi-isomorphism. It implies, for example, that if X is compact and R is a field

$$\mathscr{H}^q(X;\mathbf{A}^\bullet)=\operatorname{Hom}(\mathscr{H}^{-q}(X;\mathfrak{D}(\mathbf{A}^\bullet)),R).$$

In [42] Verdier showed that there is a complex of sheaves $\mathbb{D}_X^\bullet$ (called the dualizing complex) such that

$$\mathfrak{D}(\mathbf{A}^\bullet)\cong R\mathbf{Hom}^\bullet(\mathbf{A}^\bullet,\mathbb{D}_X^\bullet)$$

for any bounded complex $\mathbf{A}^\bullet$. He identified $\mathbb{D}_X^\bullet=\mathfrak{D}(\mathbf{R}_X)$. If $\mathbf{B}^\bullet\cong\mathfrak{D}(\mathbf{A}^\bullet)$ then the corresponding pairing

$$\mathbf{B}^\bullet\overset{L}{\otimes}\mathbf{A}^\bullet\to\mathbb{D}_X^\bullet$$

is said to be a Verdier dual pairing.

If $\mathbf{A}^\bullet$ is a bounded topologically constructible complex of sheaves on X then there is a natural isomorphism in $D^b(X)$,

$$\mathbf{A}^\bullet\cong\mathfrak{D}(\mathfrak{D}(\mathbf{A}^\bullet)).$$

We now describe two important functors $Rf_!$: $D^b(X)\to D^b(Y)$ and $f^!$: $D^b(Y)\to D^b(X)$ which were defined by Verdier for any continuous map $f\colon X\to Y$ between locally compact topological spaces. $Rf_!$ is the right derived functor of the direct image functor with proper supports, $f_!$.

$\Gamma(U,f_!\mathbf{A}^\bullet)=\{\gamma\in\Gamma(f^{-1}(U),\mathbf{A}^\bullet)|\text{support of }\gamma\text{ is proper over }U\}$. The stalk $\mathbf{H}^*(Rf_!\mathbf{A}^\bullet)_y$ is the hypercohomology with compact supports of the fibre $f^{-1}(y)$, with coefficients in $\mathbf{A}^\bullet$. If f is proper than $Rf_*=Rf_!$.

If $\mathbf{I}^\bullet$ is a complex of injective sheaves on Y, $f^!(\mathbf{I}^\bullet)$ is defined to be the sheaf associated to the presheaf whose sections over an open set $U\subset X$ are $\Gamma(U;f^!\mathbf{I}^\bullet)=\operatorname{Hom}^\bullet(f_!\mathbf{K}_U^\bullet,\mathbf{I}^\bullet)$ where $\mathbf{K}_U^\bullet$ is the standard injective resolution of the constant sheaf $\mathbf{R}_U$. The *Verdier duality theorem* is a canonical isomorphism in $D^b(Y)$,

$$Rf_*R\mathbf{Hom}^\bullet(\mathbf{A}^\bullet,f^!\mathbf{B}^\bullet)\cong R\mathbf{Hom}^\bullet(Rf_!\mathbf{A}^\bullet,\mathbf{B}^\bullet)$$

for any $\mathbf{A}^\bullet\in D^b(X)$ and $\mathbf{B}^\bullet\in D^b(Y)$.

Remarks on $f^!$ and $\mathbb{D}_X^\bullet$

If $f\colon Z\to X$ is the inclusion of a closed subspace then $\mathscr{H}^i(Z;f^!\mathbf{A}^\bullet)$ is denoted $\mathscr{H}_Y^i(X;\mathbf{A}^\bullet)$. This group is also constructed as a derived functor in §1. There is a natural isomorphism

$$\mathbb{D}_X^\bullet\cong f^!\mathbf{R}_{pt}$$

where f: $X \to$point. The complex $\mathbb{D}_X^{\cdot}$ is quasi-isomorphic to the complex of sheaves of singular chains on X which is associated to the complex of presheaves

$$\Gamma(U;\mathbf{C}^{-p}) = C_p(X, X-U; R).$$

The associated cohomology sheaves of $\mathbb{D}_X^{\cdot}$ are nonzero in negative degree only, with stalks $\mathbf{H}^{-p}(\mathbb{D}_X^{\cdot})_x = H_p(X, X-x; R)$. The hypercohomology groups $\mathscr{H}^*(X;\mathbb{D}_X^{\cdot})$ equal the ordinary homology groups with closed support of X (i.e., Borel-Moore homology). The spectral sequence associated with the complex $\mathbb{D}_X^{\cdot}$ is the Grothendreck-Zeeman spectral sequence [48].

For any homology n-manifold X, $\mathbb{D}_X^{\cdot}[-n]$ is naturally isomorphic to the *orientation sheaf* of X. If X is a smooth oriented manifold then $\mathbb{D}_X^{\cdot}[-n]$ is naturally isomorphic to the complex of differential forms on X.

In terms of duality, the functors $f^!$ and $Rf_!$ may be described by

$$f^!\mathbf{B}^{\cdot} \cong \mathfrak{D}_X(f^*\mathfrak{D}_Y(\mathbf{B}^{\cdot}))$$
$$Rf_!\mathbf{A}^{\cdot} \cong \mathfrak{D}_Y(Rf_*\mathfrak{D}_X(\mathbf{A}^{\cdot}))$$

where f: $X \to Y$ is a continuous map between topological pseudomanifolds, $\mathbf{A}^{\cdot} \in D^b(X)$ and $\mathbf{B}^{\cdot} \in D^b(Y)$.

If $\mathbf{A}^{\cdot}$ is a topologically constructible complex of sheaves on X, j: $x \to X$ is the inclusion of a point, and N is a distinguished neighborhood of x, of the type considered in §1.1 and §1.3 then

$$H^p(j^*\mathbf{A}^{\cdot}) \cong \mathscr{H}^p(N;\mathbf{A}^{\cdot}) = \mathbf{H}^p(\mathbf{A}^{\cdot})_x,$$
$$H^p(j^!\mathbf{A}^{\cdot}) \cong \mathscr{H}_c^p(N;\mathbf{A}^{\cdot}).$$

We call these groups the stalk homology and the costalk homology (respectively) of $\mathbf{A}^{\cdot}$ at x.

Proposition (1.12). *The dualizing complex* $\mathbb{D}_X^{\cdot}$ *is constructible with respect to any topological stratification of the topological pseudomanifold* X.

The proof follows from the local product structure of a topological pseudomanifold and the fact that

$$\mathbb{D}_{V\times W}^{\cdot} = \pi_1^*\mathbb{D}_V^{\cdot} \otimes \pi_2^*\mathbb{D}_W^{\cdot}$$

where π_1 and π_2 are the projections of $V \times W$ to the first and second factors respectively.

§1.13. *Standard Identities*

Suppose X and Y are topological pseudomanifolds with fixed stratifications and f: $X \to Y$ is a stratified map. Fix $\mathbf{A}^{\cdot} \in D^b(X)$ and $\mathbf{B}^{\cdot}$, $\mathbf{C}^{\cdot}$, $\mathbf{E}^{\cdot} \in D^b(Y)$ which are constructible with respect to these stratifications. Then there are natural isomorphisms in $D^b(X)$ and $D^b(Y)$,

(1) $\mathfrak{D}(\mathbf{A}^\bullet) = R\mathbf{Hom}^\bullet(\mathbf{A}^\bullet, \mathbb{D}^\bullet_X)$

(2) $\mathbb{D}^\bullet_X = \mathfrak{D}(\mathbf{R}_X) = f^!(\mathbb{D}^\bullet_Y)$

(3) $\mathbf{A}^\bullet \cong \mathfrak{D}(\mathfrak{D}(\mathbf{A}^\bullet))$

(4) $R\,\mathbf{Hom}^\bullet(\mathbf{B}^\bullet \overset{L}{\otimes} \mathbf{C}^\bullet, \mathbf{E}^\bullet) \cong R\,\mathbf{Hom}^\bullet(\mathbf{B}^\bullet, R\,\mathbf{Hom}^\bullet(\mathbf{C}^\bullet, \mathbf{E}^\bullet))$

(5) $f^!\mathbf{B}^\bullet \cong \mathfrak{D}_X f^* \mathfrak{D}_Y \mathbf{B}^\bullet$

(6) $Rf_!\mathbf{A}^\bullet \cong \mathfrak{D}_Y Rf_* \mathfrak{D}_X \mathbf{A}^\bullet$

(7) $f^*(\mathbf{B}^\bullet \overset{L}{\otimes} \mathbf{C}^\bullet) \cong f^*\mathbf{B}^\bullet \overset{L}{\otimes} f^*\mathbf{C}^\bullet$

(8) $f^!R\,\mathbf{Hom}^\bullet(\mathbf{B}^\bullet, \mathbf{C}^\bullet) \cong R\,\mathbf{Hom}^\bullet(f^*\mathbf{A}^\bullet, f^!\mathbf{B}^\bullet)$

(9) $Rf_*R\,\mathbf{Hom}^\bullet(f^*\mathbf{B}^\bullet, \mathbf{A}^\bullet) \cong R\,\mathbf{Hom}^\bullet(\mathbf{B}^\bullet, Rf_*\mathbf{A}^\bullet)$

(10) $Rf_*R\,\mathbf{Hom}^\bullet(\mathbf{A}^\bullet, f^!\mathbf{B}^\bullet) \cong R\,\mathbf{Hom}^\bullet(Rf_!\mathbf{A}^\bullet, \mathbf{B}^\bullet)$

(11) If $f\colon Y \times Z \to Y$ is the projection to the first factor, then

$$f^*R\,\mathbf{Hom}^\bullet(\mathbf{B}^\bullet, \mathbf{C}^\bullet) \cong R\,\mathbf{Hom}^\bullet(f^*\mathbf{B}^\bullet, f^*\mathbf{C}^\bullet)$$

(12) If X is a subset of Y with inclusion $f\colon X \to Y$ then

$$\begin{aligned} &X \text{ open in } Y \Rightarrow f^!\mathbf{B}^\bullet \cong f^*\mathbf{B}^\bullet \\ &X \text{ closed in } Y \Rightarrow Rf_!\mathbf{A}^\bullet \cong Rf_*\mathbf{A}^\bullet. \end{aligned}$$

(13) Fibre square:
If

$$\begin{tikzcd} Z \arrow[r, "\bar{f}"] \arrow[d, "\bar{\pi}"'] & Z \arrow[d, "\pi"] \\ X \arrow[r, "f"] & Y \end{tikzcd}$$

is a fibre square, then

$$R\bar{f}_*\bar{\pi}^*\mathbf{A}^\bullet \cong \pi^* Rf_*\mathbf{A}^\bullet \quad \text{for any } \mathbf{A}^\bullet \in D^b(X).$$

Further identities for CLC sheaves (§ 1.4):

(14) If $\mathbf{B}^\bullet$ is CLC on Y then

$$R\,\mathbf{Hom}^\bullet(\mathbf{C}^\bullet, \mathbf{E}^\bullet \overset{L}{\otimes} \mathbf{B}^\bullet) \cong R\,\mathbf{Hom}^\bullet(\mathbf{C}^\bullet, \mathbf{E}^\bullet) \overset{L}{\otimes} \mathbf{B}^\bullet$$

(15) If $f\colon X^n \to Y^m$ is an inclusion of one oriented homology manifold in another one, and if $\mathbf{B}^\bullet$ is CLC on Y, then $f^!\mathbf{B}^\bullet$ is CLC on X and

$$f^!\mathbf{B}^\bullet \cong f^*\mathbf{B}^\bullet[m-n].$$

(16) If $\mathbf{B}^\bullet$ and $\mathbf{C}^\bullet$ are CLC on Y, then
   (a) Each $y \in Y$ has a neighborhood U such that $\mathbf{B}^\bullet|\mathbf{U}$ is quasi-isomorphic to a complex of constant sheaves

(b) $R\,\mathbf{Hom}^{\cdot}(\mathbf{B}^{\cdot}, \mathbf{C}^{\cdot})$ is CLC

(c) $(R\,\mathbf{Hom}^{\cdot}(\mathbf{B}^{\cdot}, \mathbf{C}^{\cdot}))_x \cong R\,\mathbf{Hom}^{\cdot}(\mathbf{B}^{\cdot}_x\ \mathbf{C}^{\cdot}_x)$

(d) If $\mathbf{H}^p(\mathbf{B}^{\cdot})=0$ for all but one value of p, then $\mathbf{B}^{\cdot}$ is naturally isomorphic (in $D^b(Y)$) to a local coefficient system.

(17) If $\mathbf{B}^{\cdot}$ is CLC on $Y\times\mathbb{R}^n$ then restriction of sections induces a quasi-isomorphism

$$\pi^* R\pi_* \mathbf{B}^{\cdot} \xrightarrow{\cong} \mathbf{B}^{\cdot}$$

where π: $Y\times\mathbb{R}^n \to Y$ is the projection to the first factor. (This is because the stalk cohomology of $\pi^* R\pi_* \mathbf{B}^{\cdot}$ at any point (y, t) is equal to the hypercohomology of the restriction of $\mathbf{B}^{\cdot}$ to $\pi^{-1}(y)$. Since $\mathbf{B}^{\cdot}$ is CLC its cohomology sheaves are constant on $\pi^{-1}(y)$ and the spectral sequence for this hypercohomology group collapses, i.e., $E_2^{pq}=\mathscr{H}^p(\mathbb{R}^n; \mathbf{H}^q(\mathbf{B}^{\cdot}))=0$ unless $p=0$ and $E_2^{0q}=(\mathbf{H}^q(\mathbf{B}^{\cdot}))_{(y,t)}$. See lemma following Theorem 1.4).

1.14. *Truncation* ([13, 14, 23])

If $\mathbf{A}^{\cdot}$ is a complex of sheaves on X, define new complexes

$$(\tau_{\leqq p}\mathbf{A}^{\cdot})^n \equiv \begin{cases} \mathbf{A}^n & \text{if } n<p \\ \ker d^p & \text{if } n=p \\ 0 & \text{if } n>p, \end{cases}$$

$$(\tau^{\geqq p}\mathbf{A}^{\cdot})^n \equiv \begin{cases} 0 & \text{if } n<p \\ \operatorname{coker} d^{p-1} & \text{if } n=p \\ \mathbf{A}^n & \text{if } n>p. \end{cases}$$

These functors $\tau_{\leqq p}$ and $\tau^{\geqq p}$ determine "truncation" functors on the derived category $D^b(X)$. Notice, however, that $\tau_{\leqq p}\mathbf{A}^{\cdot}$ is naturally quasi-isomorphic to the complex

$$(\tilde{\tau}_{\leqq p}\mathbf{A}^{\cdot})^n \equiv \begin{cases} \mathbf{A}^n & \text{if } n\leqq p \\ \operatorname{Im} d^p & \text{if } n=p+1 \\ 0 & \text{if } n>p+1 \end{cases}$$

while $\tau^{\geqq p}\mathbf{A}^{\cdot}$ is quasi-isomorphic to the complex

$$(\tilde{\tau}^{\geqq p}\mathbf{A}^{\cdot})^n \equiv \begin{cases} 0 & \text{if } n<p-1 \\ \operatorname{Im} d^{p-1} & \text{if } n=p-1 \\ \mathbf{A}^n & \text{if } n\geqq p. \end{cases}$$

Theorem. *Suppose $\mathbf{A}^{\cdot}$ and $\mathbf{B}^{\cdot}$ are complexes of sheaves on X. Then,*

1. $\tau_{\leqq p}\tau_{\leqq q}\mathbf{A}^{\cdot}=\tau_{\leqq \min(p,q)}\mathbf{A}^{\cdot}$.
2. $(\tau_{\leqq p}\mathbf{A}^{\cdot})_x=\tau_{\leqq p}(\mathbf{A}^{\cdot}_x)$ *where $\mathbf{A}^{\cdot}_x$ denotes the stalk at $x\in X$.*
3. $\mathbf{H}^k(\tau_{\leqq p}\mathbf{A}^{\cdot})_x=\begin{cases}\mathbf{H}^k(\mathbf{A}^{\cdot})_x & \textit{for } k\leqq p\\ 0 & \textit{for } k>p.\end{cases}$
4. *If ϕ: $\mathbf{A}^{\cdot}\to\mathbf{B}^{\cdot}$ is a sheaf map which induces isomorphisms on the associated cohomology sheaves,*

$$\phi_*\colon \mathbf{H}^n(\mathbf{A}^\bullet)\cong \mathbf{H}^n(\mathbf{B}^\bullet) \quad \textit{for all } n\leqq p$$

then $\tau_{\leqq p}\phi\colon \tau_{\leqq p}\mathbf{A}^\bullet\to\tau_{\leqq p}\mathbf{B}^\bullet$ *is a quasi-isomorphism.*

5. *If* $f\colon X\to Y$ *is a continuous map, and* $\mathbf{C}^\bullet$ *is a complex of sheaves on* Y, *then*

$$\tau_{\leqq p}f^*(\mathbf{C}^\bullet)\cong f^*\tau_{\leqq p}(\mathbf{C}^\bullet).$$

6. *For any* $\mathbf{A}^\bullet\in D^b(X)$ *there is a distinguished triangle*

$$\begin{tikzcd}[column sep=small] \tau_{\leqq p}\mathbf{A}^\bullet \arrow[rr] & & \mathbf{A}^\bullet \arrow[dl] \\ & \tau^{\geqq p+1}\mathbf{A}^\bullet \arrow[ul, "{[1]}"] & \end{tikzcd}$$

7. *If* R *is a fiield and* $\mathbf{A}^\bullet$ *is a CLC complex of sheaves of* R*-modules on* X, *then there are natural quasi-isomorphisms*

$$\tau^{\geqq -p}R\,\mathbf{Hom}^\bullet(\mathbf{A}^\bullet,\mathbf{R}_X)\to\tau^{\geqq -p}R\,\mathbf{Hom}^\bullet(\tau_{\leqq p}\mathbf{A}^\bullet,\mathbf{R}_X)\leftarrow R\,\mathbf{Hom}^\bullet(\tau_{\leqq p}\mathbf{A}^\bullet,\mathbf{R}_X).$$

Deligne has also defined a "truncation over a closed subset" functor:

Definition. Let Y be a closed subset of X and let $\mathbf{S}^\bullet$ be a complex of sheaves on X. Fix an integer p. Then $\tau^Y_{\leqq p}\mathbf{S}^\bullet$ is the sheafification of the presheaf $\mathbf{T}^\bullet$, where

(a) for $i<p$, $T^i(U)=\Gamma(U;\mathbf{S}^i)$

(b) for $i=p$, $T^i(U)=\begin{cases}\Gamma(U;\mathbf{S}^i) & \text{if } U\cap Y=\phi\\ \ker d\subset\Gamma(U;\mathbf{S}^i) & \text{if } U\cap T\neq\phi\end{cases}$

(c) for $i>p$, $T^i(U)=\begin{cases}\Gamma(U;\mathbf{S}^i) & \text{if } U\cap Y=\phi\\ \mathbf{0} & \text{if } U\cap Y\neq\phi.\end{cases}$

The stalk of the associated cohomology sheaf is

$$\mathbf{H}^i(\tau^Y_{\leqq p}\mathbf{S}^\bullet)_x=\begin{cases}0 & \text{if } x\in Y \text{ and } i>p\\ \mathbf{H}^i(\mathbf{S}^\bullet)_x & \text{otherwise.}\end{cases}$$

The functor $\tau^Y_{\leqq p}$ passes to a functor on the derived category $D^b(X)$.

1.15. *Lifting Morphisms*

Proposition. *Let* $f\colon \mathbf{A}^\bullet\to\mathbf{B}^\bullet$ *be a morphism in* $D^b(X)$. *Suppose* $\mathbf{H}^i(\mathbf{A}^\bullet)=0$ *for* $i>p$ *and* $\mathbf{H}^i(\mathbf{B}^\bullet)=0$ *for* $i<p$.

Then the canonical map

$$\mathrm{Hom}_{D^b(X)}(\mathbf{A}^\bullet,\mathbf{B}^\bullet)\to\mathrm{Hom}(\mathbf{H}^p(\mathbf{A}^\bullet),\mathbf{H}^p(\mathbf{B}^\bullet))$$

is a bijection.

Proof. Up to quasi-isomorphism, $\mathbf{A}^{\bullet}$ and $\mathbf{B}^{\bullet}$ can be represented by complexes

$$\cdots \longrightarrow \mathbf{A}^{p-1} \xrightarrow{d^{p-1}} \mathbf{A}^{p} \longrightarrow 0 \longrightarrow \cdots$$

$$\cdots \longrightarrow 0 \longrightarrow \mathbf{I}^{p} \xrightarrow{d^{p}} \mathbf{I}^{p+1} \longrightarrow \cdots$$

where $\mathbf{I}^{d}$ is injective for all d. By injectivity, any morphism in $D^{b}(X)$, $f\colon \mathbf{A}^{\bullet} \to \mathbf{B}^{\bullet}$ corresponds to an actual map $\bar{f}$ between these complexes, i.e., a map

$$\operatorname{coker} d^{p-1} = \mathbf{H}^{p}(\mathbf{A}^{\bullet}) \to \ker d^{p} = \mathbf{H}^{p}(\mathbf{B}^{\bullet}) \quad \square$$

Proposition. *Suppose* $\mathbf{A}^{\bullet}$, $\mathbf{B}^{\bullet}$, *and* $\mathbf{C}^{\bullet}$ *are objects in* $D^{b}(X)$ *and that* $\mathbf{H}^{n}(\mathbf{A}^{\bullet}) = 0$ *for all* $n \geqq p+1$. *Let* $\psi\colon \mathbf{B}^{\bullet} \to \mathbf{C}^{\bullet}$ *be be a morphism such that* $\psi_{*}\colon \mathbf{H}^{n}(\mathbf{B}^{\bullet}) \to \mathbf{H}^{n}(\mathbf{C}^{\bullet})$ *is an isomorphism for all* $n \leqq p$. *Then the map induced by* ψ,

$$\operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{B}^{\bullet}) \to \operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{C}^{\bullet})$$

is an isomorphism. In particular, any $\phi\colon \mathbf{A}^{\bullet} \to \mathbf{C}^{\bullet}$ *has a unique lift (in* $D^{b}(X)$*)* $\tilde{\phi}\colon \mathbf{A}^{\bullet} \to \mathbf{B}^{\bullet}$ *such that* $\phi = \psi \circ \tilde{\phi}$.

Proof. Let $M^{\bullet}$ denote the algebraic mapping cylinder of ψ. From the long exact sequence on cohomology which is associated to the triangle

$$\begin{tikzcd} \mathbf{B}^{\bullet} \arrow[rr, "\psi"] & & \mathbf{C}^{\bullet} \arrow[dl, "\theta"] \\ & \mathbf{M}^{\bullet} \arrow[ul, "{[1]}"] & \end{tikzcd}$$

we see that $\mathbf{H}^{n}(\mathbf{M}^{\bullet}) = 0$ for all $n \leqq p-1$. Furthermore, for $\phi\colon \mathbf{A}^{\bullet} \to \mathbf{C}^{\bullet}$ the composition $\theta \circ \phi\colon \mathbf{A}^{\bullet} \to \mathbf{M}^{\bullet}$ induces the 0 map on $\mathbf{H}^{p}(\mathbf{A}^{\bullet}) \to \mathbf{H}^{p}(\mathbf{M}^{\bullet})$. The preceding lemma implies that the map induced by θ

$$\operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{C}^{\bullet}) \to \operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{M}^{\bullet}) \cong \operatorname{Hom}(\mathbf{H}^{p}(\mathbf{A}^{\bullet}), \mathbf{H}^{p}(\mathbf{M}^{\bullet}))$$

is the 0 map. Similarly $\operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{M}^{\bullet})[-1] = 0$. The conclusion now follows from the exact sequence

$$\begin{aligned} &\to \operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{M}^{\bullet})[-1] \to \operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{B}^{\bullet}) \\ &\to \operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{C}^{\bullet}) \to \operatorname{Hom}_{D^{b}(X)}(\mathbf{A}^{\bullet}, \mathbf{M}^{\bullet}) \end{aligned}$$

§ 2. $IH_{*}^{\bar{p}}$ as Hypercohomology, for P.L. Pseudomanifolds

In this chapter we show how the construction of $IH_{*}^{\bar{p}}(X)$ from [20] actually defines a complex of sheaves on X (the complex of $(\bar{p}, *)$-allowable piecewise linear chains). We calculate the local cohomology groups of this complex and find that

$$IH_{i}^{\bar{p}}(X, X-x) = 0 \qquad \text{if } i \leqq n - p(k) - 1$$

whenever x lies in a stratum of codimension k.

This vanishing condition is an essential property in the axiomatic characterization of the intersection homology sheaf.

2.0. In this chapter we will assume X is a piecewise-linear pseudomanifold with a fixed (P.L.) stratification

$$\phi = X_{-1} \subset X_0 \subset X_1 \subset \ldots \subset X_{n-2} = \Sigma \subset X$$

as in [20]. We also fix a commutative ring with unit, R.

A *perversity* is a sequence of integers $\bar{p} = (p_2, p_3, \ldots, p_n)$ with $p_2 = 0$ and $p_c \leqq p_{c+1} \leqq p_c + 1$. We shall often write $p(c)$ instead of p_c. There are several perversities of particular importance:

the zero perversity $\bar{0} = (0, 0, 0, \ldots, 0)$
the lower middle perversity $\bar{m} = (0, 0, 1, 1, 2, 2, 3, \ldots)$
the upper middle perversity $\bar{n} = (0, 1, 1, 2, 2, 3, 3, \ldots)$
the logarithmic perversity $\bar{l} = (0, 1, 2, 2, 3, 3, 4, 4, \ldots)$
the sublogarithmic perversity $\bar{s} = (0, 0, 1, 1, 2, 2, 3, \ldots)$
the top perversity $\bar{t} = (0, 1, 2, 3, 4, 5, \ldots)$.

For any perversity $\bar{p}$, the *complementary* perversity is defined to be

$$\bar{t} - \bar{p} = (0, 1 - p_1, 2 - p_2, 3 - p_3, \ldots).$$

For example, $\bar{0}$ and $\bar{t}$ are complementary, as are $\bar{m}$ and $\bar{n}$.

In this chapter we fix a perversity $\bar{p}$, and define a complex of sheaves $\mathbf{IC}^{\bullet}_{\bar{p}}$. However we shall usually omit the subscript $\bar{p}$ for notational convenience.

2.1. *The Complex of Sheaves,* $\mathbf{IC}^{\bullet}_{\bar{p}}$

The treatment in this section is parallel to that of [20] (p. 138) which may be consulted for further details. Define the complex of sheaves of P.L. chains $\mathbf{C}^{\bullet}$ on X by specifying the sections $\Gamma(U; \mathbf{C}^{\bullet})$ over any open subset $U \subset X$ as follows: If T is a locally finite triangulation of U let $C_i^T(U)$ denote the group of locally finite i-dimensional simplicial chains (with R-coefficients), with respect to this triangulation. The support of a chain $\xi \in C_i^T(U)$ is denoted $|\xi|$. Let $C_i(U)$ denote the limit of the $C_i^T(U)$ taken over all locally finite triangulations T of U. If $V \subset U$ and T is a locally finite triangulation of U, there exists a locally finite triangulation S of V such that each simplex of S is contained in a unique simplex of T. Any chain ξ in $C_i^T(U)$ thus gives rise by restriction to a chain ξ' in $C_i^S(V)$ such that $|\xi'| = |\xi| \cap V$. Taking limits over all locally finite compatible triangulations defines a *restriction homomorphism* $C_i(U) \to C_i(V)$ and thus defines a presheaf.

Definition. $\Gamma(U;\, \mathbf{C}^{-k}) \equiv C_k(U)$.

Remark. If $A \subset U$ is a (relatively closed) i-dimensional P.L. subset of U, and if $B \subset A$ is an $i-1$ dimensional P.L. subset then there is a one to one correspondence between those chains $\alpha \in \Gamma(U;\, \mathbf{C}^{-k})$ such that $|\alpha| \subset A$ and $|\partial\alpha| \subset B$, and between (Borel-Moore) homology classes with infinite supports, $\tilde{\alpha} \in H_i^{\infty}(A, B)$. The chain $\partial\alpha$ corresponds to the class $\partial_*(\tilde{\alpha}) \in H_{i-1}^{\infty}(B, \phi)$ under the connecting homomorphism.

$\mathbf{C}^{\cdot}$ is a complex of fine sheaves on X and is quasi-isomorphic to the dualizing complex $\mathbf{D}_X^{\cdot}$.

Define $\mathbf{IC}_{\bar{p}}^{-k}$ to be the subsheaf of $\mathbf{C}^{-k}$ whose sections over an open set $U \subset X$ consist of all locally finite P.L. chains $\xi \in \Gamma(U; \mathbf{C}^{-k})$ such that $|\xi|$ is $(\bar{p}, i)$-allowable and $|\partial \xi|$ is $(\bar{p}, i-1)$-allowable, with respect to the filtration $U \cap X_0 \subset U \cap X_1 \subset \ldots \subset U \cap \Sigma \subset U$. This means that for each c,

$$\dim(|\xi| \cap U \cap X_{n-c}) \leqq i - c + p(c),$$
$$\dim(|\partial \xi| \cap U \cap X_{n-c}) \leqq i - 1 - c + p(c).$$

Thus, $\mathbf{IC}_{\bar{p}}^{\cdot}$ is a complex of fine sheaves on X whose hypercohomology $\mathscr{H}^*(X; \mathbf{IC}_{\bar{p}}^{\cdot})$ is canonically isomorphic to $IH_*^{\bar{p}}(X; R)$. The associated cohomology sheaf $\mathbf{H}^{-k}(\mathbf{IC}_{\bar{p}}^{\cdot})$ is called the local intersection homology sheaf and is denoted $\mathbf{IH}_k^{\bar{p}}$. The stalk at $x \in X$ of this sheaf is denoted $IH_k^{\bar{p}}(X, X-x)$ or $IH_k^{\bar{p}}(X, X-x; R)$.

2.2. $\mathscr{L}$. *Local Coefficient Systems*

Let $\mathbf{F}$ be a local coefficient system of R modules on $X - \Sigma$. [37] (In other words, $\mathbf{F}$ is a sheaf of R modules which, viewed as an étale space over $X - \Sigma$, is a locally trivial fiber bundle. $\mathbf{F}$ is determined by the data: a base point in C and a representation of $\pi_1(C)$ for each connected component C of $X - \Sigma$.)

Consider an open subset $U \subset X$ and a locally finite triangulation T of U. Since $\mathbf{F}$ may not be defined on all of U, it is impossible to define a group $C_i^T(U, \mathbf{F})$ of i chains with coefficients in $\mathbf{F}$. Nevertheless, for any perversity $\bar{p}$, one can define $IC_i^{\bar{p}, T}(U, \mathbf{F})$ as the group of locally finite i-dimensional simplicial chains ξ with coefficients in $\mathbf{F}$ that satisfy allowability conditions: for each $c \geqq 2$,

$$\dim(|\xi| \cap U \cap X_{n-c}) \leqq i - c + \bar{p}(c),$$
$$\dim(|\partial \xi| \cap U \cap X_{n-c}) \leqq i - 1 - c + \bar{p}(c).$$

One can also define a boundary map $IC_i^{\bar{p}, T}(U, \mathbf{F}) \to IC_{i-1}^{\bar{p}, T}(U, \mathbf{F})$. This is because if Δ is any i simplex with nonzero coefficient in ξ, both the interior of Δ and the interiors of all the $i-1$ dimensional faces of Δ lie entirely in $X - \Sigma$ by the allowability conditions. This is all one needs to define by the usual methods the simplicial chain complex with local coefficients.

Definition. The complex $\mathbf{IC}_{\bar{p}}^{\cdot}(\mathbf{F})$ of intersection chains with coefficients in X is given by $\Gamma(U, \mathbf{IC}_{\bar{p}}^i(\mathbf{F})) = \lim IC_i^{\bar{p}, T}(U, \mathbf{F})$ where the limit is taken over locally finite compatible triangulations of U. The intersection homology groups with coefficients in $\mathbf{F}$, $IH_*^{\bar{p}}(X; \mathbf{F})$, are the hypercohomology, or global section cohomology of $\mathbf{IC}_{\bar{p}}^{\cdot}(\mathbf{F})$.

2.3. *Indexing Schemes*

There are four ways in the literature to index the dimension of an intersection homology group:

(a) Homology subscripts, as in [20]: a subscript k indicates chains of dimension k.

(b) Homology superscripts, as in Chap. 3 of this paper: a superscript $-k$ indicates chains of dimension k.

(c) Cohomology superscripts, as in [7, 16] a superscript l indicates chains of codimension l, i.e., dimension $n-l$.

(d) Beilinson-Bernstein-Deligne-Gabber scheme: a superscript l indicates chains of codimension $\frac{n}{2}+l$.

For an n-dimensional compact oriented pseudomanifold these schemes compare as follows: $IH_k(X)$ in scheme (a) is isomorphic to $\mathscr{H}^{-k}(X;\mathbf{IC}^\cdot)$ in scheme (b), $IH^{n-k}(X)$ in scheme (c), and $IH^{\frac{n}{2}-k}(X)$ in scheme (d).

2.4. *Calculation of the Local Intersection Homology*

Recall that a P.L. stratification of X determines a filtered P.L. isomorphism of a neighborhood U of each point $x\in X_k - X_{k-1}$ with $x * S^{k-1} * L_X$, where S^{k-1} is the (P.L.) $k-1$ sphere and L_x is a filtered P.L. space called the *link of the stratum* $X_k - X_{k-1}$ at the point x. (If x and y are points in the same connected component of $X_k - X_{k-1}$ then L_x and L_y are P.L. isomorphic.)

Proposition. *The stalk at the point* $x\in X_k - X_{k-1}$ *of the local intersection homology sheaf is*

$$IH_j(X, X-x) = IH_x^{-j} = \begin{cases} IH_{j-k-1}(L_x) & \text{if } j \geqq n-p(n-k) \\ 0 & \text{if } j \leqq n-p(n-k)-1. \end{cases}$$

Intuition. If $j<n-p(n-k)$ then any j-dimensional cycle in $IC_j^{\bar{p}}(X)$ will intersect the stratum $X_k - X_{k-1}$ in a subset of dimension less than k, and can therefore (by transversality) be moved away from the point $\{x\}$. So it represents 0 in the local homology group. If $j \geqq n-p(n-k)$ then any j-dimensional cycle which contains $\{x\}$ also contains a neighborhood of $\{x\}$ in the stratum $X_k - X_{k-1}$, so it is locally the product of $\mathbb{R}^k$ with the cone over a $j-k-1$ dimensional cycle in the link L of the stratum.

Proof. $S^{k-1} * L_x$ inherits a stratification from that of X. We shall find an isomorphism between the stalk at x of the sheaf $\mathbf{IC}^{-j}$ and the group $IC_{j-1}(S^{k-1} * L_x)$.

Any P.L. chain $\eta \in IC_{j-1}(S^{k-1} * L_x)$ gives rise to a germ of a chain near x, by forming the join $x * \eta$. Conversely, if ξ is a germ of a chain near x then in a sufficiently small neighborhood U it can be expressed as a sum of simplices each of which contains x as a vertex. Pseudo-radial retraction along cone lines then determines a P.L. chain $\eta \in IC_{j-1}(S^{k-1} * L_x)$ such that $(x*\eta)\cap U = |\xi| \cap U$. Thus, $\mathbf{IC}_*^{-j} \cong IC_{j-1}(S^{k-1} * L_x)$.

We now compute $IC_*(S^{k-1} * L_x)$. If $A \in C_q(S^{k-1})$ and $B \in IC_{j-q-1}(L_x)$ then $A * B \in IC_j(S^{k-1} * L_x)$ provided either (a) $q \leqq j-(n-k)+p(n-k)-1$ or (b) $q \leqq j-(n-k)+p(n-k)$ and $\partial B = 0$. Letting

$$\tau IC_j(L_x) = \begin{cases} IC_j(L_x) & \text{if } j \geqq n-k-p(n-k) \\ \ker \partial & \text{if } j = n-k-1-p(n-k) \\ 0 & \text{if } j < n-k-1-p(n-k) \end{cases}$$

we obtain a chain map

$$C_*(S^{k-1}) \otimes \tau I C_*(L_x) \to I C_*(S^{k-1} * L_x)$$

which is given by the join of chains. In fact, this map induces isomorphisms on homology because it has a homotopy inverse ψ which is determined by the following (perversity-preserving) rule: If η is a j-simplex in $S^{k-1} * L_x$, let $\psi(\eta) = A * B$ where $A = \eta \cap S^{k-1}$ and B is the image of $\eta - A$ under pseudo-radial retraction along join lines, to L_x.

Applying the Kunneth formula,

$$\begin{aligned} IH_j(X, X-x) &= IH_{j-1}(S^{k-1} * L_x) \\ &= \bigoplus_{q=0}^{k-1} H_q(S^{k-1}) \otimes H_{j-q-2}(\tau I C_*(L_x)) \\ &= \begin{cases} IH^{\bar{p}}_{j-k-1}(L_x) & \text{if } j \geqq n - p(n-k) \\ 0 & \text{if } j < n - p(n-k). \end{cases} \end{aligned}$$

2.5. *Attaching Property of* $\mathbf{IC}^{\bullet}$

The following proposition will be needed in the next chapter (§ 3.5) where its significance will become apparent.

Define $U_k = X - X_{n-k}$. Let $i_k\colon U_k \to U_{k+1}$ and $j_k\colon (U_{k+1} - U_k) \to U_{k+1}$ be the inclusions. Set $\mathbf{IC}^{\bullet}_k = \mathbf{IC}^{\bullet} | U_k$.

Proposition. *The natural homomorphism*

$$\mathbf{IC}^{\bullet}_{k+1} \to R i_{k*} i_k^* \mathbf{IC}^{\bullet}_{k+1}$$

induces isomorphisms

$$\mathbf{H}^m(j_k^* \mathbf{IC}^{\bullet}_{k+1}) \to \mathbf{H}^m(j_k^* R i_{k*} i_k^* \mathbf{IC}^{\bullet}_{k+1})$$

for all $m \leqq \bar{p}(k) - n$.

Proof. Since $\mathbf{IC}^{\bullet}$ is fine, we have

$$R i_{k*} i_k^* \mathbf{IC}^{\bullet} | U_{k+1} = i_{k*} i_k^* \mathbf{IC}^{\bullet} | U_{k+1}.$$

Note that sections of $\mathbf{IC}^{\bullet} | U_{k+1}$ consist of chains in U_{k+1} which can be triangulated with finitely many simplices near $x \in X_{n-k}$, and which satisfy a perversity restriction there. On the other hand, sections of $i_{k*} i_k^*(\mathbf{IC}^{\bullet} | U_{k+1})$ consist of chains *in* U_k which can be triangulated with locally (in U_k) finitely many simplices and which do not necessarily satisfy a perversity restriction near X_{n-k}. Thus $\mathbf{IC}^{\bullet} | U_{k+1}$ is a complex of subsheaves of $i_{k*} i_k^*(\mathbf{IC}^{\bullet} | U_{k+1})$ and we shall now show that the inclusion induces isomorphisms on the cohomology sheaves of dimensions $m \leqq p(k) - n$, thus establishing the above claim. It suffices to study the cohomology at a point $x \in X_{n-k}$. The inclusion of stalks

$$\mathbf{IC}^{\bullet}_x \to [i_{k*} i_k^*(\mathbf{IC}^{\bullet} | U_{k+1})]_x$$

is a chain map. There is a map back which can be defined as follows: For $j \geqq n - p(k)$, let $\xi \in [i_{k*} i_k^*(\mathbf{IC}^{-j} | U_{k+1})]_x$. Such a germ has a representation which is a j-dimensional P.L. chain (not necessarily compact) contained in U_{k+1}. Choose a local P.L. filtered product neighborhood of x, $U=D^{n-k}\times c(L)$ where D^{n-k} is the (P.L.) $n-k$ disc and $c(L)$ denotes the cone on a filtered space L, the *link* of the $n-k$-dimensional stratum. U may be chosen so that ξ is transverse to $D^{n-k}\times L$, i.e., so that $\dim(\xi\cap D^{n-k}\times L)\leqq j-1$ (McCrory [29]), and $\xi\cap D^{n-k}\times L$ can be triangulated with *finitely* many simplices. Let $\pi\colon U\to D^{n-k}$ denote the projection to the first factor and let $\eta=c(\xi\cap(D^{n-k}\times L))$ be the P.L. mapping cylinder of $\pi|(|\xi|\cap D^{n-k}\times L)$. η can be oriented using the product orientation from $|\xi|\times[0,1]$, so it defines a chain. We claim $\eta\in\mathbf{IC}_x^{-j}$ i.e., that $|\eta|$ is $(\bar{p},j)$-allowable, since $\dim(\eta\cap X_{n-k})\leqq n-k$. Note that when $\partial\xi=0$ we get $\partial\eta=0$ and in fact η represents the same class as ξ in $[\mathbf{H}^{-j}(i_{k*}i_k^*\mathbf{IC}^\bullet|U_{k+1})]_x$ since $\xi-\eta$ is the boundary of the (infinite) chain $c(\xi\cap U)$. Thus the above sheaf inclusion induces a surjective on local cohomology, and a similar (relative) argument shows it is also injective on local cohomology.

§ 3. Sheaf Theoretic Construction of $\mathbf{IC}^\bullet$

In [14] Deligne suggested a new method for constructing the complex $\mathbf{IC}^\bullet$. His procedure constructs (for any perversity $\bar{p}$) a complex of sheaves $\mathbf{P}^\bullet$ on any topological pseudomanifold by starting with the constant sheaf on the nonsingular part and using standard sheaf theoretic operations.

In this chapter we show that $\mathbf{P}^\bullet$ is naturally isomorphic to $\mathbf{IC}^\bullet$ (provided both are constructed with respect to the same stratification - an assumption which is lifted in the next chapter). This result was suggested by Deligne.

In § 3.1 we give Deligne's construction of the complex $\mathbf{P}^\bullet$. In § 3.3 we list axioms which uniquely characterize this complex of sheaves (up to quasi-isomorphism). In § 3.4 we verify that the complex $\mathbf{IC}^\bullet$ from § 2.1 satisfies these axioms and is therefore quasi-isomorphic to $\mathbf{P}^\bullet$.

3.0. Throughout this chapter, X will be an n-dimensional topological pseudomanifold. We *fix* a topological stratification [§ 1.1]

$$\phi=X_{-1}\subset X_0\subset\ldots\subset X_{n-2}=\Sigma\subset X.$$

In § 3.5 we will also assume that X has a P.L. structure and the topological stratification is also a P.L. stratification.

In this chapter we fix a perversity $\bar{p}$, and a regular Noetherian ring R of finite Krull dimension. The word *sheaf* will mean a sheaf of R-modules.

The complex resulting from Deligne's construction [§ 3.1] is denoted $\mathbf{P}^\bullet$ and in § 3.5 the complex of P.L. intersection chains on X is denoted $\mathbf{IC}_{\bar{p}}^\bullet$. Theorem 3.5 asserts that these complexes are canonically isomorphic in $D^b(X)$ whenever they are both defined. In subsequent chapters we will use $\mathbf{IC}^\bullet$ (or $\mathbf{IC}_{\bar{p}}^\bullet$) to denote this isomorphism class of objects, for any topological pseudomanifold.

3.1. *Deligne's Construction*

Consider the filtration by open sets,

$$U_1=U_2\subset U_3\subset\ldots\subset U_{n+1}=X$$

where $U_k = X - X_{n-k}$ and $i_k\colon U_k \to U_{k+1}$ is the inclusion. Define complexes $\mathbb{P}_k^\bullet \in D^b(U_k)$ inductively as follows

$$\mathbb{P}_2^\bullet = \mathbb{D}_{X-\Sigma}^\bullet \cong \mathbf{R}[n] \quad \text{on} \quad U_2 = X - \Sigma$$
$$\mathbb{P}_{k+1}^\bullet = \tau_{\leqq p(k)-n} R i_{k*} \mathbb{P}_k^\bullet \quad \text{for } k \geqq 2.$$

Definition. Deligne's construction is the complex $\mathbb{P}^\bullet = \mathbb{P}_{n+1}^\bullet$ which is defined by this process. In other words,

$$\mathbb{P}^\bullet = \tau_{\leqq p(n)-n} R i_{n*} \dots \tau_{\leqq p(3)-n} R i_{3*} \tau_{\leqq p(2)-n} R i_{2*} \mathbf{R}_{X-\Sigma}[n].$$

$\mathscr{L}$. We could equally well have started with a system of local coefficients $\mathbf{F}$ on $X-\Sigma$ in place of the constant sheaf $\mathbf{R}$, so

$$\mathbb{P}_2^\bullet = \mathbf{F}[n] \quad \text{on} \quad U_2 = X - \Sigma$$
$$\mathbb{P}_{k+1}^\bullet = \tau_{\leqq p(k)-n} R i_{k*} \mathbb{P}_k^\bullet \quad \text{for } k \geqq 1$$
$$\mathbb{P}^\bullet(\mathbf{F}) = \mathbb{P}_{n+1}^\bullet.$$

The resulting complex $\mathbb{P}^\bullet(\mathbf{F})$ is called the "intersection homology chains with coefficients in $\mathbf{F}$".

Lemma. $\mathbb{P}^\bullet$ *is constructible with respect to the given stratification of* X.

Proof. Clearly $\mathbb{P}_2^\bullet$ is constructible on U_2. Suppose we have shown that $\mathbb{P}_k^\bullet$ is constructible on U_k. Fix $x \in U_{k+1} - U_k$ and let N be a distinguished neighborhood of x in U_{k+1}, of the type considered in § 1.1, i.e., there is a stratum preserving homeomorphism

$$N \cong \mathbb{R}^k \times V$$

where $V = \operatorname{cone}^\circ(L)$ for some stratified space L. Let $N^\circ = N \cap U_k \cong \mathbb{R}^k \times V^\circ$ where $V^\circ = \operatorname{cone}(L) - \text{vertex}$. Consider the fibre square

$$\begin{tikzcd} \mathbb{R}^k \times V^\circ \arrow[r, "i"] \arrow[d, "\pi"] & \mathbb{R}^k \times V \arrow[d, "\pi"] \\ V^\circ \arrow[r, "i"] & V \end{tikzcd}$$

By § 1.13.17, § 1.13.13, and induction,

$$\begin{aligned} R i_*(\mathbb{P}_k^\bullet | N^\circ) &\cong R i_*(\pi^* R\pi_* \mathbb{P}_k^\bullet | N^\circ) \\ &\cong \pi^* R i_* R\pi_*(\mathbb{P}_k^\bullet | N^\circ) \end{aligned}$$

which shows that $R i_{k*} \mathbb{P}_k^\bullet$ is CLC on each stratum of U_{k+1}. It follows that $\tau_{\leqq p(k)-n} R i_{k*} \mathbb{P}_k^\bullet$ is also CLC on each stratum of U_{k+1}, which completes the induction.

3.2. *The Attaching Map*

If $\{X_k\}$ denotes the filtration of the space X, let $U_k = X - X_{n-k}$ denote the complementary filtration by open sets, with inclusions $i_k\colon U_k \to U_{k+1}$. Let $j_k\colon (U_{k+1} - U_k) \to U_{k+1}$ be the inclusion of the stratum of codimension k into U_{k+1}. Let $\mathbf{S}^\bullet$ be a complex of sheaves on X which is constructible with respect to the filtration $\{X_k\}$ (see § 1.11) and let $\mathbf{S}_k^\bullet = \mathbf{S}^\bullet | U_k$.

Definition. The attaching map of degree m associated with the complex $\mathbf{S}^\bullet$ over the stratum $X_{n-k}-X_{n-k-1}$ is the sheaf map

$$A_m\colon \mathbf{H}^m(j_k^*\,\mathbf{S}^\bullet_{k+1})\to\mathbf{H}^m(j_k^*\,R\,i_{k*}\,i_k^*\,\mathbf{S}^\bullet_{k+1})$$

which is obtained by restricting the natural morphism

$$\mathbf{S}^\bullet_{k+1}\to R\,i_{k*}\,i_k^*\,\mathbf{S}^\bullet_{k+1}$$

to this stratum, and taking the induced map on cohomology sheaves.

We shall say the sheaf $\mathbf{S}^\bullet$ is r-attached across this stratum, if A_m is an isomorphism for all $m\leqq r$.

3.3. *Axioms* [AX1]

Definition. Let $\mathbf{S}^\bullet$ be a complex of sheaves on X, which is constructible with respect to the stratification $\{X_k\}$ and let $\mathbf{S}^\bullet_k=\mathbf{S}^\bullet|(X-X_{n-k})$. We shall say $\mathbf{S}^\bullet$ satisfies the axioms [AX1] (with perversity $\bar{p}$, and with respect to the stratification $\{X_k\}$) provided:

(a) Normalization: $\mathbf{S}^\bullet|(X-\Sigma)\cong\mathbf{F}[n]$ where F is a local coefficient system on $X-\Sigma$.

(b) Lower bound: $\mathbf{H}^i(\mathbf{S}^\bullet)=0$ for all $i<-n$.

(c) Vanishing condition: $\mathbf{H}^m(\mathbf{S}^\bullet_{k+1})=0$ for all $m>p(k)-n$.

(d) Attaching: $\mathbf{S}^\bullet$ is $p(k)-n$ attached across each stratum of codimension k, i.e., the attaching maps

$$\mathbf{H}^m(j_k^*\,\mathbf{S}^\bullet_{k+1})\to\mathbf{H}^m(j_k^*\,R\,i_{k*}\,i_k^*\,\mathbf{S}^\bullet_{k+1})$$

are isomorphisms for all $k\geqq 2$ and all $m\leqq\bar{p}(k)-n$.

Definition. We shall say $\mathbf{S}^\bullet$ satisfies $[\mathrm{AX1}]_R$ it if satisfies [AX1] with $\mathbf{F}=\mathbf{R}_{X-\Sigma}$ = the constant sheaf, in part (a).

3.4. *Alternate Formulations of* AX1[d]

We may replace axiom (d) with

(d′) $\quad \mathbf{H}^m(j_k^!\,\mathbf{S}^\bullet_{k+1})=0 \quad$ for all $\quad k\geqq 2 \quad$ and all $\quad m\leqq\bar{p}(k)-n+1.$

It is easy to see that (d′)⇒(d) using the long exact cohomology sequence associated with the distinguished triangle

$$\begin{tikzcd} j_k^*\,\mathbf{S}^\bullet_{k+1} \arrow[rr] & & j_k^*\,R\,i_{k*}\,i_k^*\,\mathbf{S}^\bullet_{k+1} \arrow[dl, "{[1]}"] \\ & j_k^!\,\mathbf{S}^\bullet_{k+1} \arrow[ul] & \end{tikzcd}$$

The same sequence also give (c) and (d)⇒(d′).

Furthermore AX1[d′] is equivalent to

(d″) for all $k\geqq 2$, for all $x\in X_{n-k}-X_{n-k-1}$ and for all $m\leqq p(k)-k+1$, we have

$$H^m(j_x^! \mathbf{S}^\bullet)=0$$

where $j_x: \{x\}\to X$ is the inclusion of a point.

To see this, factor j_x into a composition

$$x \xrightarrow[u_x]{} X_{n-k}-X_{n-k-1} \xrightarrow[j]{} X$$

Then $j_x^! \mathbf{S}^\bullet \cong u_x^! j^! \mathbf{S}^\bullet \cong u_x^* j^! \mathbf{S}^\bullet[n-k]$ by § 1.13.15. Thus the cohomology of this complex vanishes in dimensions $m \leqq p(k)-k+1$ iff the stalk cohomology of $j^! \mathbf{S}^\bullet$ vanishes in dimensions $\leqq p(k)-n+1$.

This reformulation of AX1[d] is useful because it is equivalent to AX2[d] which will appear in the next chapter.

3.5. [AX1] *Characterizes Deligne's Construction*

Theorem. *The functor* $\mathbb{P}^\bullet$ *which assigns to any locally trivial sheaf* $\mathbf{F}$ *on* $X-\Sigma$, *the complex*

$$\mathbb{P}^\bullet(\mathbf{F})=\tau_{\leqq p(n)-n} R i_{n*} \ldots \tau_{\leqq p(2)-n} R i_{2*} \mathbf{F}[n]$$

defines an equivalence of categories between

(a) *the category of locally constant sheaves on* $X-\Sigma$ *and*

(b) *the full subcategory of* $D^b(X)$ *whose objects are all complexes of sheaves which satisfy the axioms* [AX1].

The inverse functor $\mathbf{L}$ assigns to any constructible complex of sheaves $\mathbf{S}^\bullet$ which satisfy [AX1] the locally constant sheaf $\mathbf{L}(\mathbf{S}^\bullet)=\mathbf{H}^{-n}(\mathbf{S}^\bullet|(X-\Sigma))$.

Proof. In fact we will show that for each $k \geqq 2$ the functor

$$\mathbb{P}_k^\bullet=\tau_{\leqq p(k)-n} R i_{k*}$$

defines an equivalence of categories between

(a) the full subcategory C_k of $D^b(U_k)$ whose objects are complexes of sheaves which satisfy the axioms [AX1] on U_k, and

(b) the full subcategory C_{k+1} of $D^b(U_{k+1})$ whose objects are complexes of sheaves which satisfy the axioms [AX1] on U_{k+1}.

The inverse functor $\mathbf{L}_k$ is i_k^*.

This will suffice because $\mathbb{P}^\bullet=\mathbb{P}_n^\bullet \circ \ldots \circ \mathbb{P}_3^\bullet \circ \mathbb{P}_2^\bullet$ and $\mathbf{L}=\mathbf{L}_2 \circ \mathbf{L}_3 \circ \ldots \circ \mathbf{L}_n$. Using § 1.13.16(d), $\mathbf{L}_2(\mathbf{S}_3^\bullet)=\mathbf{H}^{-n}(\mathbf{S}_3^\bullet|(X-\Sigma))$ where $\mathbf{S}_3^\bullet=\mathbf{S}_3^\bullet|U_3$.

Clearly L_k is a functor from C_{k+1} to C_k. $\mathbb{P}_k^\bullet$ is a functor from C_k to C_{k+1} for the following reasons: For any $\mathbf{A}^\bullet \in C_k$, $\mathbb{P}_k^\bullet(\mathbf{A}^\bullet)=\tau_{\leqq p(k)-n} R i_{k*} \mathbf{A}^\bullet$ satisfies [AX1](a)(b)(c) by construction. [AX1](d) is also satisfied because the attaching map is the composition

$$\tau_{\leqq p(k)-n} j_k^* R i_{k*} \mathbf{A}^\bullet \cong j_k^* \mathbb{P}_k^\bullet(\mathbf{A}^\bullet) \to j_k^* R i_{k*} i_k^* \mathbb{P}_k^\bullet(\mathbf{A}^\bullet) \cong j_k^* R i_{k*} \mathbf{A}^\bullet$$

which induces isomorphisms on stalk cohomology in the dimensions $m \leqq p(k)-n$.

Clearly $\mathbf{L}_k \mathbb{P}_k^{\boldsymbol{\cdot}}(\mathbf{S}_k^{\boldsymbol{\cdot}}) = \mathbf{S}_k^{\boldsymbol{\cdot}}$. For each object $\mathbf{A}^{\boldsymbol{\cdot}}$ in C_{k+1} we must construct a quasi-isomorphism $T_{k+1}(\mathbf{A}^{\boldsymbol{\cdot}})$: $\mathbf{A}^{\boldsymbol{\cdot}} \cong \mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L}_k(\mathbf{A}^{\boldsymbol{\cdot}})$ which is natural as a transformation from the identity to $\mathbb{P}_k^{\boldsymbol{\cdot}} \circ \mathbf{L}_k$, i.e., for any morphism f: $\mathbf{A}^{\boldsymbol{\cdot}} \to \mathbf{B}^{\boldsymbol{\cdot}}$ in the category C_{k+1}, the following diagram commutes:

$$\begin{tikzcd}
\mathbf{A}^{\boldsymbol{\cdot}} \arrow[r, "f"] \arrow[d, "T_{k+1}(\mathbf{A}^{\boldsymbol{\cdot}})"'] & \mathbf{B}^{\boldsymbol{\cdot}} \arrow[d, "T_{k+1}(\mathbf{B}^{\boldsymbol{\cdot}})"] \\
\mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L}_k(\mathbf{A}^{\boldsymbol{\cdot}}) \arrow[r, "\mathbb{P}_k^{\boldsymbol{\cdot}} L_k(f)"'] & \mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L}_k(\mathbf{B})
\end{tikzcd}$$

Consider the natural map $\mathbf{A}^{\boldsymbol{\cdot}} \to R i_{k*} i_k^* \mathbf{A}^{\boldsymbol{\cdot}}$. Define $T_{k+1}(\mathbf{A}^{\boldsymbol{\cdot}})$ to be the composition

$$\mathbf{A}^{\boldsymbol{\cdot}} \cong \tau_{\leqq p(k)-n} \mathbf{A}^{\boldsymbol{\cdot}} \to \tau_{\leqq p(k)-n} R i_{k*} i_k^* \mathbf{A}^{\boldsymbol{\cdot}} = \mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L}_k(\mathbf{A}^{\boldsymbol{\cdot}})$$

This is a quasi isomorphism over U_k. We must check that it is also a quasi isomorphism over $U_{k+1} - U_k$. By [AX1](d) the morphism

$$j_k^* \mathbf{A}^{\boldsymbol{\cdot}} \to j_k^* R i_{k*} i_k^* \mathbf{A}^{\boldsymbol{\cdot}}$$

induces isomorphisms on cohomology sheaves in all dimensions $m \leqq p(k) - n$. Thus

$$j_k^* \mathbf{A}^{\boldsymbol{\cdot}} \cong j_k^* \tau_{\leqq p(k)-n} \mathbf{A}^{\boldsymbol{\cdot}} \cong j_k^* \tau_{\leqq p(k)-n} R i_{k*} i_k^* \mathbf{A}^{\boldsymbol{\cdot}}$$

which completes the check.

For any morphism f: $\mathbf{A}^{\boldsymbol{\cdot}} \to \mathbf{B}^{\boldsymbol{\cdot}}$ in C_{k+1} we have a diagram

$$\begin{tikzcd}
\mathbf{A}^{\boldsymbol{\cdot}} \arrow[rrr, "f"] \arrow[dr] \arrow[dd, "T_{k+1}(\mathbf{A}^{\boldsymbol{\cdot}})"'] & & & \mathbf{B}^{\boldsymbol{\cdot}} \arrow[dl] \arrow[dd, "T_{k+1}(\mathbf{B}^{\boldsymbol{\cdot}})"] \\
& R i_{k*} \mathbf{L} \mathbf{A}^{\boldsymbol{\cdot}} \arrow[r] & R i_{k*} \mathbf{L} \mathbf{B}^{\boldsymbol{\cdot}} & \\
\mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L}(\mathbf{A}^{\boldsymbol{\cdot}}) \arrow[ur] \arrow[rrr, "\mathbb{P}_k^{\boldsymbol{\cdot}} L(f)"'] & & & \mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L}(\mathbf{B}^{\boldsymbol{\cdot}}) \arrow[ul, "\theta"]
\end{tikzcd}$$

By induction the right and left triangles and top and bottom trapezoids commute. We must show that the outside square commutes. It is clear that

$$\theta \circ \mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L}(f) \circ T_{k+1}(\mathbf{A}^{\boldsymbol{\cdot}}) = \theta \circ T_{k+1}(\mathbf{B}^{\boldsymbol{\cdot}}) \circ f.$$

However, according to § 1.15 composition with θ induces an isomorphism

$$\operatorname{Hom}_{D^b(U_{k+1})}(\mathbf{A}^{\boldsymbol{\cdot}}, \mathbb{P}_k^{\boldsymbol{\cdot}} \mathbf{L} \mathbf{B}^{\boldsymbol{\cdot}}) \to \operatorname{Hom}_{D^b(U_{k+1})}(\mathbf{A}^{\boldsymbol{\cdot}}, R i_{k*} \mathbf{L}(\mathbf{B}^{\boldsymbol{\cdot}}))$$

so we can cancel the θ from the above equation. This completes the proof.

Corollary. *If a constructible complex $\mathbf{S}^{\boldsymbol{\cdot}}$ satisfies* $[\mathrm{AX1}]_R$, *then $\mathbf{S}^{\boldsymbol{\cdot}}$ is naturally quasi-isomorphic to Deligne's complex $\mathbb{P}^{\boldsymbol{\cdot}} = \mathbb{P}^{\boldsymbol{\cdot}}(\mathbf{R}_{X-\Sigma})$ which was defined in* § 3.1. *If in addition all the $\mathbf{S}^i$ are fine, the cohomology groups of the complex*

$$\to \Gamma(X; S^{i-1}) \to \Gamma(X; S^i) \to \Gamma(X; S^{i+1}) \to \ldots$$

are naturally isomorphic to the intersection homology groups $IH_*^{\bar{p}}(X)$.

Example. J. Cheeger has shown [10] that if X is a Riemannian pseudomanifold with conical singularities then the complex of locally L^2 differential forms on $X-\Sigma$ is a complex of fine sheaves on X which satisfies [AX1]. This complex is defined by

$\Gamma(U;\Omega^p)\equiv$ those differential p-forms ω on $U\cap(X-\Sigma)$ such that for every point $x\in U$ there is a neighborhood V_x such that

$$\int_{V_x\cap(X-\Sigma)}\omega\wedge *\omega<\infty \quad\text{and}\quad \int_{V_x\cap(X-\Sigma)} d\omega\wedge *(d\omega)<\infty.$$

The following example explains the use of the word "attaching": the complex

$$\bigoplus_i \mathbf{H}^i(\mathbf{IC}^\cdot)[-i]$$

satisfies $[\mathrm{AX1}]_R$ except for the attaching axiom. It clearly has the same homology sheaves as $\mathbf{IC}^\cdot$ but is in general not isomorphic to $\mathbf{IC}^\cdot$ since $\mathbf{IC}^\cdot$ is indecomposable (see Corollary 2 in § 4.1).

3.6. $\mathbf{IC}^\cdot$ *Satisfies the Axioms* [AX1]

Theorem. *If X is a P.L. pseudomanifold with a fixed P.L. stratification, then the sheaf of intersection chains* $\mathbf{IC}^\cdot$ *satisfies the axioms* $[\mathrm{AX1}]_R$ *with respect to that stratification.*

Proof. Axioms AX1(a)(b) are obviously satisfied. Axiom [AX1](c) and constructibility were verified in § 2.4. [AX1](d) was verified in § 2.5.

Corollary. *If X is a P.L. pseudomanifold with a fixed P.L. stratification then the sheaf of piecewise linear intersection chains as constructed in* [19] (*and* § 2.1) *is naturally quasi isomorphic to* $\mathbb{P}^\cdot$ *as constructed by Deligne's procedure in* § 3.1.

§ 4. Topological Invariance of $\mathbf{IC}^\cdot$

In this chapter we shall show that the intersection homology groups $IH_*^{\bar p}(X)$ are topological invariants and they do not depend on the choice of stratification of X. In fact, we shall show for any homeomorphism $f\colon X\to Y$, that the complexes $\mathbb{P}_X^\cdot$ and $f^*\,\mathbb{P}_X^\cdot$ are quasi-isomorphic.

A key ingredient of the proof is the construction of the canonical $\bar p$ filtration $X_0^{\bar p}\subset X_1^{\bar p}\subset\ldots\subset X_{n-2}^{\bar p}\subset X_n^{\bar p}=X$ of X. This depends on the choice of a perversity $\bar p$ but not on a previous choice of a stratification of X - it is a purely topological invariant of X. The filtration $\{X_i^{\bar p}\}$ is a sort of "homological stratification". For example, $X_i^{\bar p}-X_{i-1}^{\bar p}$ is a R-homology manifold of dimension i. It may be thought of as the "coarsest stratification" with respect to which Deligne's construction gives $\mathbf{IC}^\cdot$ - any topological stratification is a refinement of it. The role of the canonical $\bar p$ filtration in the proof is to compare objects of $D^b(X)$ satisfying axioms AX1 with respect to two different topological stratifications (which may not have a common refinement).

This chapter contains a set of axioms [AX2] which uniquely characterize the complex $\mathbb{P}^\bullet$ up to quasi-isomorphism but which do not refer to a choice of stratification of X. These axioms involve the concepts of local support and co-support of a complex of sheaves, which we now describe.

If $\mathbf{S}^\bullet$ is a complex of sheaves on X, and $j_x\colon \{x\}\to X$ is the inclusion of a point, there is a homomorphism

$$\mathscr{H}^m(X;\mathbf{S}^\bullet)\to\mathscr{H}^m(Rj_{x*}j_x^*\mathbf{S}^\bullet)\cong H^m(j_x^*\mathbf{S}^\bullet).$$

If a class $\xi\in\mathscr{H}^m(X;\mathbf{S}^\bullet)$ does not vanish under this homomorphism then any cycle representative of ξ must contain the point x. Thus, $H^m(j_x^*\mathbf{S}^\bullet)$ represents local classes which "cannot be pulled away from the point x", and we say

$$\{x\in X\,|\,H^m(j_x^*\mathbf{S}^\bullet)\neq 0\}$$

is the *local support* set of the complex $\mathbf{S}^\bullet$ (in dimension m).

Similarly, there is a homomorphism

$$H^m(j_x^!\mathbf{S}^\bullet)=\mathscr{H}^m(Rj_{x*}j_x^!\mathbf{S}^\bullet)\to\mathscr{H}^m(X;\mathbf{S}^\bullet).$$

A class $\eta\in\mathscr{H}^m(X;\mathbf{S}^\bullet)$ is in the image of this homomorphism if some cycle representative of η is completely contained in a neighborhood of x. Thus $H^m(j_x^!\mathbf{S}^\bullet)$ represents local classes which are "supported near x" and we say

$$\{x\in X\,|\,H^m(j_x^!\mathbf{S}^\bullet)\neq 0\}$$

is the local co-support set of the complex $\mathbf{S}^\bullet$ (in dimension m). The axioms [AX2] place restrictions on the size of the local support and co-support sets.

4.0. Throughout this chapter we shall assume X is an n-dimensional topological pseudomanifold (but we do not fix any particular stratification of X). We shall also fix a perversity $\bar{p}$ and let q denote the complementary perversity, $q(k)=k-2-p(k)$ for all $k\geqq 2$. By a *sheaf* on X, we shall mean a sheaf of R-modules, where R is a fixed finite dimensional regular Noetherian ring.

4.1. *Axioms* [AX2]

For the perversity $\bar{p}$, let p^{-1} denote the sub-inverse of p, i.e.,

$$p^{-1}(l)=\min\{c\,|\,p(c)=l\}.$$

We use the convention that $\min(\phi)=\infty$. Recall that $\bar{p}$ is a function from the set $\{2,3,4,\ldots\}$ to the nonnegative integers. It satisfies inequalities which imply that if it takes two integer values then it also takes each value between them.

Let $\bar{q}$ denote the complementary perversity, $q(k)=k-2-p(k)$ and let q^{-1} denote the sub-inverse of $\bar{q}$.

Suppose X is an n-dimensional topological pseudomanifold. For each $x\in X$, let $j_x\colon\{x\}\to X$ denote the inclusion.

Definition. A topologically constructible complex of sheaves $\mathbf{S}^{\bullet}$ on X satisfies axioms [AX2] provided:

(a) Normalization – There is a closed subset $Z \subset X$ such that $\mathbf{S}^{\bullet}|(X-Z) \cong \mathbf{R}_{X-Z}[n]$ and $\dim(Z) \leqq n-2$.

(b) Lower bound
$\mathbf{H}^m(\mathbf{S}^{\bullet}) = 0$ for all $m < -n$.

(c) Support condition
For all $m \geqq -n+1$, $\dim\{x \in X \mid H^m(j_x^* \mathbf{S}^{\bullet}) \neq 0\} \leqq n - p^{-1}(m+n)$.

(d) Cosupport condition
For all $m \leqq -1$, $\dim\{x \in X \mid H^m(j_x^! \mathbf{S}^{\bullet}) \neq 0\} \leqq n - q^{-1}(-m)$.

Where "dim" denotes the topological dimension of Hurewicz and Wallman [24]. We use the convention that a set of negative dimension (including $-\infty$) must be empty.

Uniqueness Theorem. *Up to canonical isomorphism there exists a unique complex in $D^b(X)$ which satisfies axioms* [AX2]. *It is given by* $\mathbf{IC}^{\bullet}$, *constructed* (*as in* § 2.1 *and* § 3.1) *with respect to any stratification of* X.

The proof is in § 4.3

Corollary 1. *For any topological pseudomanifold X, the groups $IH_*^{\bar{p}}(X)$ are topological invariants and exist independently of the choice of a stratification of X.*

$\mathscr{L}$. Let Z be X_{n-2} for some topological stratification $X = X_n \supset \ldots \supset X_0$ of X and let $\mathbf{F}$ be a local system on $X-Z$. Then if we replace axiom [AX2](a) by

(a′) $$\mathbf{S}^{\bullet}|(X-Z) \cong \mathbf{F}[n]$$

the uniqueness theorem still holds with $\mathbf{IC}^{\bullet}$ replaced by $\mathbf{IC}^{\bullet}(\mathbf{F})$.

Definition. A topological pseudomanifold X is *irreducible* if $X - \Sigma$ is connected for some (and hence for any) stratification of X.

Corollary 2. *If X is irreducible and $\mathbf{F}$ is a local coefficient system on $X-\Sigma$ which is indecomposable as a local system, then $\mathbf{IC}^{\bullet}(\mathbf{F})$ is indecomposable, i.e., if $\mathbf{IC}^{\bullet} \cong \mathbf{A} \oplus \mathbf{B}^{\bullet}$ in $D^b(X)$, then either $\mathbf{A}^{\bullet} \cong \mathbf{zero}^{\bullet}$ or $\mathbf{B}^{\bullet} \cong \mathbf{zero}^{\bullet}$.*

Proof. Both $\mathbf{A}^{\bullet}$ and $\mathbf{B}^{\bullet}$ satisfy the axioms [AX2] except for axiom (a).

Remark. Assume $\mathbf{S}^{\bullet}$ is a complex of sheaves which is constructible with respect to a topological stratification $X_0 \subset X_1 \subset \ldots \subset X_n = X$, and satisfies [AX2]. Then it also satisfies the axioms [AX2]′ where the sets in statements (c) and (d) are assumed to be unions of strata and where "$X-Z$" is replaced by "$X - X_{n-2}$". This is because $\mathbf{S}^{\bullet}|(X - X_{n-2})$ is a local coefficient system and $\pi_1(X - X_{n-2} - Z) \to \pi_1(X - X_{n-2})$ is surjective since Z has topological codimension 2

4.2. *Construction of the Canonical $\bar{p}$-Filtration*

Let $U^{\bar{p}}$ be the largest open set on which $\mathbb{D}_X^{\bullet}$ is CLC (§ 1.4). Let $\Sigma^{\bar{p}} = X - U^{\bar{p}}$, and let $n-m$ be the topological dimension of $\Sigma^{\bar{p}}$. Then $m \geqq 2$ because for any topological stratification of X as a topological pseudomanifold, $\Sigma^{\bar{p}}$ will be a union of connected components of interiors of strata (by Prop. 1.12), and the stratum of dimension n is contained in $U^{\bar{p}}$.

Define $X^{\bar{p}}_{n-2}=\Sigma^{\bar{p}}=X-U^{\bar{p}}$.

Suppose by induction that $X^{\bar{p}}_{n-k}$, $X^{p}_{n-k+1},\ldots,X^{\bar{p}}_{n-2}\subset X$ has been defined. Let $U^{\bar{p}}_k=X-X^{\bar{p}}_{n-k}$ and let $\mathbf{P}^{\bullet}_k\in D^b(U^{\bar{p}}_k)$ be the complex obtained from Deligne's construction using the filtration by $\{X^{\bar{p}}_i\}$. Let $h_k\colon U^{\bar{p}}_k\to X$ be the inclusion. Let V' be the largest open subset of $X^{\bar{p}}_{n-k}$ on which $\mathbf{D}^{\bullet}_{X^{\bar{p}}_{n-k}}$ and $(Rh_{k*}\mathbf{P}^{\bullet}_k)|X^{\bar{p}}_{n-k}$ are both CLC. Let V be the union of the connected components of V' which have topological dimension $n-k$.

Define $X^{\bar{p}}_{n-k-1}=X^{\bar{p}}_{n-k}-V$.

This completes the inductive step in the definition.

Proposition. 1. *This procedure terminates after finitely many steps. In fact,* $\dim(X^{\bar{p}}_{n-k-1})\leqq n-k-1$.

2. *For any topological stratification of X, each $X^{\bar{p}}_{j-k-1}$ is a closed union of connected components of strata.*

3. *Each $Z_{n-k}=X^{\bar{p}}_{n-k}-X^{\bar{p}}_{n-k-1}$ is either empty or else it is an $n-k$ dimensional R-homology manifold.*

4. *Let $\mathbf{Q}^{\bullet}$ be the complex of sheaves obtained by applying Deligne's construction to the filtration $\{X^{\bar{p}}_i\}$. Then for any k, $\mathbf{Q}^{\bullet}|Z_{n-k}$ is CLC.*

Proof. We prove all these propositions simultaneously by induction. Suppose they are true for all integers $<k$. Fix a topological stratification $X_0\subset X_1\subset\ldots\subset X_n=X$ of X.

Lemma. *$Rh_{k*}\mathbf{Q}_k|X^{\bar{p}}_{n-k}$ is constructible with respect to this stratification.*

Proof of Lemma. We must show this complex is CLC on each stratum of $X^{\bar{p}}_{n-k}$. Choose any point $x\in X^{\bar{p}}_{n-k}$ and let S_r be the stratum which contains x. Choose a conical filtered space $V=V_n\supset V_{n-1}\supset\ldots\supset V_r=$a point, and a continuous homeomorphism of $V\times\mathbb{R}^r$ to a neighborhood N of x in X (see the definition of a stratification in § 1.1). Let $\pi\colon N\to V$ denote the resulting projection to the first factor. V inherits a filtration

$$V_0=V^{\bar{p}}_r\subset V^{\bar{p}}_{n-k}\subset V^{\bar{p}}_{n-k+1}\subset\ldots\subset V^{\bar{p}}_n$$

such that $X^{\bar{p}}_{n-j}=V^{\bar{p}}_{n-j}\times\mathbb{R}^r$.

We now use a tilda to denote the intersection of a subset with N, and we use a bar to denote the projection of such a subset to V, as follows:

Let $\tilde{U}^p=U^p\cap N$, $\tilde{U}^{\bar{p}}_j=U^{\bar{p}}_j\cap N$.

Let $\tilde{i}_j\colon\tilde{U}^{\bar{p}}_j\to\tilde{U}^{\bar{p}}_{j+1}$ and $\tilde{h}_j\colon\tilde{U}^{\bar{p}}_j\to N$ denote the inclusions.

Let $\bar{U}^p=\pi(\tilde{U}^{\bar{p}})$, $\bar{U}^{\bar{p}}_j=\pi(\tilde{U}^{\bar{p}}_j)$.

Let $\bar{i}_j\colon\bar{U}^{\bar{p}}_j\to\bar{U}^{\bar{p}}_{j+1}$ and $h_j\colon\bar{U}^{\bar{p}}_j\to V$ denote the inclusions.

It follows from the inductive hypothesis (2) that $\tilde{U}^{\bar{p}}_j=\pi^{-1}(\bar{U}^{\bar{p}}_j)$.

Therefore,

$$\begin{aligned}(Rh_{k*}\mathbf{Q}^{\bullet}_k)|N&\cong Rh_{k*}(\mathbf{Q}^{\bullet}_k|N)\\&\cong R\tilde{h}_{k*}\tau_{\leqq p(k-1)-n}R\tilde{i}_{k-1*}\ldots\tau_{\leqq p(2)-n}R\tilde{i}_{2*}\mathbf{R}_{\tilde{U}^{\bar{p}}}[n]\\&\cong R\tilde{h}_{k*}\tau_{\leqq p(k-1)-n}R\tilde{i}_{k-1*}\ldots\tau_{\leqq p(2)-n}R\tilde{i}_{2*}\pi^{*}\mathbf{R}_{\bar{U}^{\bar{p}}}[n].\end{aligned}$$

Now the π^* moves to the left changing tildas to bars, giving

$$\cong R\tilde{h}_{k*}\pi^*\tau_{\leq p(k-1)-n}R\bar{i}_{k-1*}\ldots\tau_{\leq p(2)-n}R\bar{i}_{2*}\mathbf{R}_{\bar{U}^{\bar{p}}}[n]$$
$$\cong \pi^* R\bar{h}_{k*}\tau_{\leq p(k)-n}R\bar{i}_{k-1*}\ldots\tau_{\leq p(2)-n}R\bar{i}_{2*}\mathbf{R}_{\bar{U}^{\bar{p}}}[n]$$

which is CLC when restricted to $\pi^{-1}(V_0)$.

This lemma implies that the set V' above is a union of connected components of strata and it contains the $n-k$ dimensional strata in X. Thus V is also a union of connected components of strata which contains all the $n-k$ dimensional strata in V'. This proves (1) and (2). Property (3) is guaranteed by the condition that $\mathbb{D}^{\bullet}_{X^{\bar{p}}_{n-k}}$ be CLC on $V' \supset V = Z_{n-k}$. Finally, $Rh_{k*}\mathbf{Q}^{\bullet}_k|Z_{n-k}$ is CLC so $\tau_{\leq p(k)-n}Rh_{k*}\mathbf{Q}^{\bullet}_k|Z_{n-k} = \mathbf{Q}^{\bullet}|Z_{n-k}$ is also CLC, which proves (4).

4.3. *Proof of Topological Invariance* (Theorem 4.1)

For any topological stratification $\{X_j\}$ of X, define $S\{X_j\}$ to be the full subcategory of $D^b(X)$ consisting of complexes which are construtible with respect to $\{X_j\}$. Define $\mathbf{Q}^{\bullet}$ to be the object (in $D^b(X)$) obtained by Deligne's construction with respect to the canonical $\bar{p}$ filtration.

The proof of Theorem 4.1 will follow from three statements:

1. An object $\mathbf{A}^{\bullet}$ in $S\{X_j\}$ satisfies [AX2] if and only if it satisfies $[\mathrm{AX1}]_R$ with respect to the stratification $\{X_j\}$ (see Lemma 1 below).
2. $\mathbf{Q}^{\bullet}$ satisfies [AX2] (Lemma 2 below).
3. $\mathbf{Q}^{\bullet}$ is an object in $S\{X_j\}$ for any stratification $\{X_j\}$ of X (Proposition 4.2).

Statement (2) clearly guarantees the existence part of Theorem 4.1.

Uniqueness. Let $\mathbf{S}^{\bullet}$ be a complex of sheaves which satisfies [AX2]. By assumption $\mathbf{S}^{\bullet}$ is constructible with respect to some topological stratification $\{X_j\}$ of X. By statement (3), $\mathbf{Q}^{\bullet}$ is also an object in $S\{X_j\}$. By statement (1), both $\mathbf{S}^{\bullet}$ and $\mathbf{Q}^{\bullet}$ satisfy $[\mathrm{AX1}]_R$ with respect to this stratification, so by Theorem 3.5 and its corollary, $\mathbf{S}^{\bullet}$ and $\mathbf{Q}^{\bullet}$ are canonically isomorphic in $D^b(X)$.

Now let $\mathbf{A}^{\bullet}$ be the object obtained from Deligne's construction with respect to any other topological stratification of X. By Theorem 3.5, it satisfies $[\mathrm{AX1}]_R$ with respect to that stratification and by statement (1) it also satisfies axioms [AX2]. Thus it is canonically isomorphic to $\mathbf{Q}^{\bullet}$. Q.E.D.

Lemma 1. *Let $\{X_j\}$ be a topological stratification of X and suppose $\mathbf{S}^{\bullet}$ is an object in $S\{X_j\}$. Then $\mathbf{S}^{\bullet}$ satisfies* $[\mathrm{AX1}]_R$ *if and only if it satisfies* [AX2].

Proof. $[\mathrm{AX1}]_R$(a)(b) is equivalent to [AX2](a)(b) using the remark in §4.1. [AX1](c)$\Rightarrow$[AX2](c) as follows:

By constructibility the set $\{x \in X \mid H^m(j_x^*\mathbf{S}^{\bullet}) \neq 0\}$ is a union of strata. AX1(c) states that these strata may not include $X - X_{n-k}$ if $p(k) - n < m$. Thus, the only allowable strata are contained in $X - X_{n-k}$ for $p(k) - n \geq m$, or $k \geq p^{-1}p(k) \geq p^{-1}(m+n)$. This set has dimension less than or equal to $n-k \leq n - p^{-1}(m+n)$ which verifies AX2(c). The same calculation gives AX2(c) $\Rightarrow$AX1(c).

By § 3.4, [AX1](c)(d)$\Leftrightarrow$[AX1](d'') which is equivalent to [AX2](d) by a counting argument analogous to the one in the preceding paragraph.

Lemma 2. *Let* $\mathbf{Q}^{\bullet}$ *be the object obtained by applying Deligne's construction* (§ 2.1) *to the canonical* $\bar{p}$*-filtration* $\{X_i^{\bar{p}}\}$ *of* X. *Then* $\mathbf{Q}^{\bullet}$ *satisfies* [AX2].

Proof. $\mathbf{Q}^{\bullet}$ satisfies [AX2](a)(b)(c) by construction. It remains to verify axiom [AX2](d).

We will verify that

$$x \in X_l^{\bar{p}} - X_{l-1}^{\bar{p}} \Rightarrow H^m(j_x^! \mathbf{Q}^{\bullet}) = 0 \qquad \text{for all } m \leqq -q(n-l)-1.$$

This will suffice because if

$$(X_l^{\bar{p}} - X_{l-1}^{\bar{p}}) \cap \{x \mid H^m(j_x^! \mathbf{Q}^{\bullet}) \neq 0\} \neq \phi$$

then

$$m > -q(n-l)-1$$

so

$$q^{-1}(-m) \leqq q^{-1} q(n-l) \leqq n-l$$

or

$$l \leqq n - q^{-1}(-m).$$

Verification. Let $j\colon X_l^{\bar{p}} - X_{l-1}^{\bar{p}} \to U$ and $i\colon U - X_{l-1}^{\bar{p}} \to U$ be the inclusions, where $U = X - X_{l-1}^{\bar{p}}$.

Consider the long exact sequence on the stalk cohomology at x, which is associated to the distinguished triangle

$$\begin{tikzcd} j_* j^! \mathbf{Q}^{\bullet}|U \arrow[rr] & & \mathbf{Q}^{\bullet}|U \arrow[dl] \\ & Ri_* i^* \mathbf{Q}^{\bullet}|U \arrow[ul, "{[1]}"] & \end{tikzcd}$$

since $\mathbf{Q}^{\bullet}|U \cong \tau_{\leqq p(n-l)-n} Ri_* i^* \mathbf{Q}^{\bullet}|U$ we have

$$\mathbf{H}^m(j_* j^! \mathbf{Q}^{\bullet})_x = 0 \qquad \text{for } m \leqq -n + p(n-l) + 1.$$

Now factor j_x into a composition

$$x \xrightarrow[u_x]{} X_l^{\bar{p}} - X_{l-1}^{\bar{p}} \xrightarrow[j]{} U = X - X_{l-1}^{\bar{p}}.$$

Then $j_x^! \mathbf{Q}^{\bullet} = u_x^! j^! \mathbf{Q}^{\bullet} = u_x^* j^! \mathbf{Q}^{\bullet}[l]$ since $X_l^{\bar{p}} - X_{l-1}^{\bar{p}}$ is a homology manifold. Thus the cohomology of this complex vanishes in dimensions $m \leqq l - n + p(n-l) + 1 = -q(n-l)+1$, as desired.

§ 5. Basic Properties of $IH^{\bar{p}}(X)$

In this chapter we prove the basic results of intersection homology without assuming X has a P.L. structure, using the methods of sheaf theory.

5.0. Throughout this chapter, X will denote an n-dimensional topological pseudomanifold, but we do not necessarily fix a stratification of X. Since we will be considering several perversities at once, we will denote the complex of intersection chains with perversity $\bar{p}$ by $\mathbf{IC}_{\bar{p}}^{\bullet}$.

Fix a regular Noetherian ring R of finite dimension. By *sheaf* we shall mean a sheaf of R-modules.

In some parts of this chapter we will assume that X has an R-orientation.

Definition. An R-orientation for X is a chosen quasi-isomorphism

$$\mathbb{D}^\bullet_{X-\Sigma} \to \mathbf{R}_{X-\Sigma}[n].$$

If $\operatorname{char}(R) \neq 2$ then an R-orientation of X is equivalent to an orientation of $X-\Sigma$ in the usual topological sense.

5.1. *The Maps from Cohomology and to Homology*

Choose an orientation on X.

Let $j\colon \Sigma \to X$ be the inclusion of the singularity set of X (for some topological stratification of X) and let $i\colon X-\Sigma = U \to X$ be the inclusion of the non-singular part.

Definition. The "cap product with the orientation class" is the morphism $\phi\colon \mathbf{R}_X[n] \to \mathbb{D}^\bullet_X$ which is obtained as the canonical lift (in $D^b X$) of the orientation:

$$\mathbf{R}_X[n] \to Ri_*\mathbf{R}_{X-\Sigma}[n] \xrightarrow{\cong} Ri_*\mathbb{D}^\bullet_{X-\Sigma}.$$

The lift exists and is unique in $D^b(X)$ because in the distinguished triangle,

$$\begin{tikzcd}
& Rj_*\mathbb{D}^\bullet_\Sigma[1] \arrow[rr, "{[1]}"] & & \mathbb{D}^\bullet_X \arrow[dl] \\
\mathbf{R}_X[n] \arrow[r] & Ri_*\mathbf{R}_{X-\Sigma} \arrow[r] & Ri_*\mathbf{R}_U[n] \arrow[ul] &
\end{tikzcd}$$

the cohomology sheaves associated to $Rj_*\mathbb{D}^\bullet_\Sigma[1]$ vanish in dimensions $t \leqq -n$ (see § 1.15).

Proposition. *Suppose X is oriented. Let $\mathbf{IC}^\bullet$ denote the complex of intersection chains on X with respect to the perversity $\bar{p}$. Let $i\colon X-\Sigma \to X$ denote the inclusion. There are unique morphism in $D^b(X)$,*

$$\mathbf{R}_X[n] \to \mathbf{IC}^\bullet \to \mathbb{D}^\bullet_X$$

such that the induced morphism

$$i^*\mathbf{R}_X[n] \to i^*\mathbf{IC}^\bullet$$

is the evident one and the induced morphism

$$i^*\mathbf{IC}^\bullet \to i^*\mathbb{D}^\bullet_X$$

is given by the orientation. These morphism factor the cap product with the orientation.

Proof. Choose a stratification $\{X_k\}$ of X. With notation as in § 3.1, suppose by induction that $\mathbb{P}_k^\bullet \to \mathbb{D}_{U_k}^\bullet$ has been constructed. We obtain a morphism

$$\mathbb{P}_{k+1}^\bullet \to R i_{k*} \mathbb{P}_k^\bullet \to R i_{k*} \mathbb{D}_{U_k}^\bullet$$

which has a unique lift to $\mathbb{D}_{U_{k+1}}^\bullet$ by § 1.15 since the local cohomology sheaves associated to $R j_{k*} \mathbb{D}_{X_{n-k}-X_{n-k-1}}^\bullet$ vanish in dimensions $t \leqq k-n-1$. Similarly a morphism $\mathbf{R}_{U_k} \to \mathbb{P}_k^\bullet$ (defined by induction) gives rise to a morphism

$$\mathbf{R}_{U_{k+1}} \to R i_{k*} i_k^* \mathbf{R}_{U_k} \to R i_{k*} \mathbb{P}_k^\bullet$$

which has a unique lift to $\tau_{\leqq p(k)-n} R i_{k*} \mathbb{P}_k^\bullet$ by § 1.15.

$\mathscr{L}$. For all stratified pseudomanifolds X, oriented or not, there is an orientation local system $\mathcal{O}$ on the nonsingular part $X - \Sigma$ of X

$$\mathcal{O} \cong \mathbf{H}^{-n}(\mathbb{D}^\bullet) | X - \Sigma.$$

(If X is oriented then $\mathcal{O} \cong \mathbf{R}_{(X-\Sigma)}$.) In general there are canonical morphisms

$$\mathbf{R}_X[n] \to \mathbf{IC}^\bullet \quad \text{and} \quad \mathbf{IC}^\bullet(\mathcal{O}) \to \mathbb{D}^\bullet.$$

5.2. *Construction of the Intersection Pairings*

Suppose $\bar{l} + \bar{m} \leqq \bar{p}$ are perversities. We shall define a product morphism

$$\mathbf{IC}_{\bar{l}}^\bullet \overset{L}{\otimes} \mathbf{IC}_{\bar{m}}^\bullet \to \mathbf{IC}_{\bar{p}}^\bullet[n].$$

The product is defined using a stratification, but turns out to be independent of the stratification. (In fact, one will obtain the same product morphism by following this construction, using the common refinement of the canonical $\bar{l}$, $\bar{m}$, $\bar{p}$ filtrations, in place of the stratification.)

For a stratification $\{X_k\}$ of X, let $\mathbf{L}^\bullet$, $\mathbf{M}^\bullet$, and $\mathbb{P}^\bullet$ denote the complex from § 3.1 associated to the perversities $\bar{l}$, $\bar{m}$, and $\bar{p}$ respectively. Using notation as in § 3.1, we shall define morphisms $\mathbf{L}_k^\bullet \overset{L}{\otimes} \mathbf{M}_k^\bullet \to \mathbb{P}_k^\bullet$ inductively over $U_k = X - X_{n-k}$ as follows:

On $U_2 = X - \Sigma$ the morphism is multiplication,

$$\mathbf{R}_{X-\Sigma}[n] \overset{L}{\otimes} \mathbf{R}_{X-\Sigma}[n] \to \mathbf{R}_{X-\Sigma}[n][n].$$

Now suppose μ_k: $\mathbf{L}_k^\bullet \otimes \mathbf{M}_k^\bullet \to \mathbb{P}_k^\bullet[n]$ has been constructed. We must define a morphism

$$(\tau_{\leqq l(k)-n} R i_{k*} \mathbf{L}_k^\bullet) \overset{L}{\otimes} (\tau_{\leqq m(k)-n} R i_{k*} \mathbf{M}_k^\bullet) \to \tau_{\leqq p(k)-n} R i_{k*} \mathbb{P}_k^\bullet[n].$$

The pairing μ_k induces morphisms

$$\begin{aligned}(\tau_{\leqq l(k)-n} R i_{k*} \mathbf{L}_k^\bullet) \overset{L}{\otimes} (\tau_{\leqq m(k)-n} R i_{k*} \mathbf{M}_k^\bullet) &\to R i_{k*} \mathbf{L}_k^\bullet \overset{L}{\otimes} R i_{k*} \mathbf{M}_k^\bullet \\ &\to R i_{k*} (\mathbf{L}_k^\bullet \overset{L}{\otimes} \mathbf{M}_k^\bullet) \\ &\to R i_{k*} (\mathbb{P}_k^\bullet[n]).\end{aligned}$$

By § 1.15, this composition has a canonical lift (in $D^b(U_{k+1})$) to $\tau_{\leq p(k)-n} Ri_{k*}\mathbb{P}_k^{\bullet}[n]$ since the cohomology sheaves associated to

$$(\tau_{\leq l(k)-n} Ri_{k*}\mathbf{L}_k^{\bullet}) \overset{L}{\otimes} (\tau_{\leq m(k)-n} Ri_{k*}\mathbf{M}_k^{\bullet})$$

vanish in dimensions $j \geq l(k)+m(k)-2n+1$.

Remark. The compatibility between the intersection pairings defined for different choices of $\bar{l}, \bar{m}$, and $\bar{p}$ is easily checked. In particular these products are compatible with the cup product $(\mathbf{R}_X \overset{L}{\otimes} \mathbf{R}_X \to \mathbf{R}_X)$ and the cap product $(\mathbf{R}_X \overset{L}{\otimes} \mathbb{D}_X^{\bullet} \to \mathbb{D}_X^{\bullet})$.

Corollary. *Let X be a topological pseudomanifold. If $\bar{l}+\bar{m} \leq \bar{p}$ there exist canonical "intersection" pairings*

$$IH_i^{\bar{l}}(X) \overset{L}{\otimes} IH_j^{\bar{m}}(X) \to IH_{i+j-n}^{\bar{p}}(X).$$

These pairings are compatible with the cup and cap products.

$\mathscr{L}$. *Remark.* It was not necessary to have an orientation of X in the preceding construction.

$\mathscr{L}$. We could also have started with local coefficient systems $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{F}_3$ on $X-\Sigma$ and a product $\mathbf{F}_1 \otimes \mathbf{F}_2 \to \mathbf{F}_3$. This gives rise to "intersection pairings"

$$IH_i^{\bar{l}}(X; \mathbf{F}_1) \overset{L}{\otimes} IH_j^{\bar{m}}(X; \mathbf{F}_2) \to IH_{i+j-n}^{\bar{p}}(X; \mathbf{F}_3).$$

5.3. $\mathbf{IC}^{\bullet}$ *and Verdier Duality*

In this paragraph we shall assume the coefficient ring R is a field, which we now denote by k. In this section (except for the last paragraph) we shall assume X is k-orientable and that a k orientation has been chosen.

One of the most important properties of $\mathbf{IC}^{\bullet}$ is the duality between $\mathbf{IC}_{\bar{p}}^{\bullet}$ and $\mathbf{IC}_{\bar{q}}^{\bullet}$ when $\bar{p}+\bar{q}=\bar{t}$. In particular if X has even codimension strata, $\mathbf{IC}_{\bar{m}}^{\bullet}$ is self dual (for example, if X is a complex analytic variety). For the reader who is only interested in the statement that the sheaf $\mathbb{P}^{\bullet}$ as constructed by Deligne satisfies this duality, the following rather simple proof can be extracted from Chap. 3: $\mathbb{P}^{\bullet}$ is characterized by [AX1] (with $\mathbf{F}=\mathbf{k}_X$) as shown in § 3.5. We may replace axiom [AX1](d) with [AX1](d″) as shown in § 3.4. This set of axioms is "self dual": [AX1](c) for a complex $\mathbf{A}^{\bullet}$ and a perversity $\bar{p}$ is equivalent to [AX1](d″) for the dual complex $\mathfrak{D}(\mathbf{A}^{\bullet})$ and perversity $\bar{q}=\bar{t}-\bar{p}$, and vice versa.

In the following detailed argument we will use the axioms [AX2] (which we believe are more natural) instead of [AX1].

Definition. A pairing $\mathbf{A}^{\bullet} \overset{L}{\otimes} \mathbf{B}^{\bullet} \to \mathbb{D}_X^{\bullet}[n]$ of objects in $D^b(X)$ (where $n=\dim(X)$) is called a *Verdier dual pairing* if it induces an isomorphism in $D^b X$,

$$\text{(§ 1.12)} \qquad \mathbf{A}^{\bullet} \xrightarrow{\cong} R\mathbf{Hom}^{\bullet}(\mathbf{B}^{\bullet}, \mathbb{D}_X^{\bullet})[n]$$

Theorem. *Suppose $\bar{p}+\bar{q}=\bar{t}$ are perversities. Then the intersection pairing of § 5.2, followed by the map to homology*

$$\mathbf{IC}^{\bullet}_{\bar{p}} \overset{L}{\otimes} \mathbf{IC}^{\bullet}_{\bar{q}} \to \mathbf{IC}^{\bullet}_{\bar{t}}[n] \to \mathbb{D}^{\bullet}_X[n]$$

is a Verdier dual pairing.

Corollary. *If X is compact, the pairings*

$$IH^{\bar{p}}_{*}(X;k) \otimes IH^{\bar{q}}_{*}(X;k) \to H_{*}(X;k) \to k$$

induce isomorphisms

$$IH^{\bar{p}}_{i}(X;k) \cong \mathrm{Hom}(IH^{\bar{q}}_{n-i}(X;k), k) \qquad ([6]).$$

Remark. Field coefficients are used in an essential way during the following proof of Theorem 5.3. The dualizing complex of a point $\{x\}$ with coefficients in k is

$$\mathbb{D}^m_{\{x\}} = \begin{cases} k & \text{for } m=0 \\ 0 & \text{for } m \neq 0 \end{cases}$$

and this complex is injective as a complex of k-modules.

Proof of Theorem 5.3. Let $\mathbf{S}^{\bullet} = R\mathbf{Hom}^{\bullet}(\mathbf{IC}^{\bullet}_{\bar{q}}, \mathbb{D}^{\bullet}_X)[n]$. The intersection pairing induces an isomorphism

$$\mathbf{IC}^{\bullet}_{\bar{p}}|U = \mathbf{k}_U[n] \xrightarrow{\cong} \mathbf{S}^{\bullet}|U$$

where $U = X - \Sigma$ is the nonsingular set. We must check that this isomorphism extends to a quasi-isomorphism over the rest of X. It suffices to check that $\mathbf{S}^{\bullet}$ satisfies the axioms AX2(c) and (d).

Let $j_x\colon \{x\} \to X$ denote the inclusion of a point. Then

$$\begin{aligned} j_x^* \mathbf{S}^{\bullet} &= j_x^* \mathfrak{D}(\mathbf{IC}^{\bullet}_{\bar{q}})[n] \\ &= j_x^!(\mathfrak{D}(\mathfrak{D}\mathbf{IC}^{\bullet}_{\bar{q}})[n] \\ &\cong \mathrm{Hom}(j_x^! \mathbf{IC}^{\bullet}_{\bar{q}}, k)[n]. \end{aligned}$$

Therefore $H^m(j_x^* \mathbf{S}^{\bullet}) = \mathrm{Hom}(H^{-m-n}(j_x^!(\mathbf{IC}^{\bullet}_{\bar{q}}), k)$.

Since $\mathbf{IC}^{\bullet}_{\bar{q}}$ satisfies AX2(d), the set of points $x \in X$ for which this group is nonzero, has dimension $\leq n - p^{-1}(m+n)$, which verifies AX2(c).

Similarly, $H^m(j_x^! \mathbf{S}^{\bullet}) \cong \mathrm{Hom}(H^{-m-n} j_x^* \mathbf{IC}^{\bullet}_{\bar{q}}, k)$. The set of points $x \in X$ for which this group is nonzero has dimension $\leq n - q^{-1}(m)$ which verifies AX2(d).

$\mathscr{L}$. We now drop the assumption that X is oriented and we let $\mathcal{O}$ be the orientation local system of k modules on $X - \Sigma$ considered in § 5.1 $\mathscr{L}$. A pairing $\mathbf{F}_1 \otimes \mathbf{F}_2 \to \mathcal{O}$ of local systems on $X - \Sigma$ is called perfect if the induced mapping $\mathbf{F}_1 \to \mathrm{Hom}(\mathbf{F}_2, \mathcal{O})$ is an isomorphism. The above proof gives the

Theorem. *Suppose $\bar{p}+\bar{q}=\bar{t}$ are perversities and the pairing $\mathbf{F}_1 \otimes \mathbf{F}_2 \to \mathcal{O}$ is perfect. Then the intersection pairing followed by the map to homology*

$$\mathbf{IC}^{\bullet}_{\bar{p}}(\mathbf{F}_1) \overset{L}{\otimes} \mathbf{IC}_{\bar{q}}(\mathbf{F}_2) \to \mathbf{IC}^{\bullet}_{\bar{t}}(\mathcal{O})[u] \to \mathbb{D}^{\bullet}_X[u]$$

is a Verdier dual pairing.

5.4. *Functoriality for Normally Nonsingular Maps*

5.4.1. *Normally Nonsingular Inclusions* [15, 18].

Definition. An inclusion of oriented topological pseudomanifolds $\alpha: Y \to X$ is said to be *normally nonsingular* with codimension c, if Y has a c-dimensional tubular neighborhood in X, i.e., an open neighborhood $N \subset X$ and a retraction $\pi: N \to Y$ such that (π, N, Y) is homeomorphic to an $\mathbb{R}^c$-vectorbundle over Y (where Y is identified with the 0-section).

For example, suppose X is a Whitney stratified subset of some manifold M, and $Y = H \cap X$ where H is a smooth submanifold of M which is transverse to each stratum of X. Then the inclusion $Y \to X$ is normally nonsingular with codimension $c = \dim(M) - \dim(H)$.

Theorem. *Suppose $\alpha: Y \to X$ is a normally nonsingular inclusion with codimension c. Fix a perversity $\bar{p}$ and let $\mathbf{IC}_X^{\bullet}$ and $\mathbf{IC}_Y^{\bullet}$ denote the intersection homology complex on X and Y respectively. Then there are canonical isomorphisms $\alpha^* \mathbf{IC}_X^{\bullet} \cong \mathbf{IC}_Y^{\bullet}[c]$ and $\alpha^! \mathbf{IC}_X^{\bullet} \cong \mathbf{IC}_Y^{\bullet}$.*

Proof. Let $\pi: N \to Y$ denote the tubular neighborhood of Y in X and suppose $\dim(X) = n$. From the topological invariance of $\mathbf{IC}_X^{\bullet}$ we have a quasi-isomorphism $\mathbf{IC}_X^{\bullet}|N \cong \pi^* \alpha^* \mathbf{IC}_X^{\bullet}$.

We shall now check the axioms [AX2] for the complex $\alpha^* \mathbf{IC}_X^{\bullet}[-c]$ on Y. To verify AX2(c) we must find

$$\begin{aligned} \beta &= \dim\{y \in Y \mid H^m(j_y^* \alpha^* \mathbf{IC}_X^{\bullet}[-c]) \neq 0\} \\ &= \dim\{y \in Y \mid H^{m-c}(j_y^* \alpha^* \mathbf{IC}_X^{\bullet}) \neq 0\} \end{aligned}$$

where $j_y: \{y\} \to Y$ is the inclusion of a point. Suppose $x \in N$ and $\pi(x) = y$. Then

$$j_x^* \mathbf{IC}_X^{\bullet} \cong j_x^* \pi^* (\alpha^* \mathbf{IC}_X^{\bullet}) \cong (\pi j_x)^* (\alpha^* \mathbf{IC}_X^{\bullet}) \cong j_y^* \alpha^* \mathbf{IC}_X^{\bullet}.$$

Consequently,

$$\begin{aligned} n - p^{-1}(m - c + n) &\geqq \dim\{x \in X \mid H^{m-c} j_x^* \mathbf{IC}_X^{\bullet} \neq 0\} \\ &\geqq \dim\{y \in Y \mid H^{m-c} j_y^* \alpha^* \mathbf{IC}_X^{\bullet} \neq 0\} + c \end{aligned}$$

which shows $\beta \leqq n - c - p^{-1}(m - c + n)$ as desired.

To verify AX2(d) we must calculate

$$\begin{aligned} \gamma &= \dim\{y \in Y \mid H^m(j_y^! \alpha^* \mathbf{IC}_X^{\bullet}[-c]) \neq 0\} \\ &= \dim\{y \in Y \mid H^{m-c}(j_y^! \alpha^* \mathbf{IC}_X^{\bullet}[-c]) \neq 0\}. \end{aligned}$$

If $\pi(x) = y$, then

$$j_x^! \mathbf{IC}_X^{\bullet} \cong j_x^! \pi^* \alpha^* \mathbf{IC}_X^{\bullet} \cong j_x^! \pi^! \alpha^* \mathbf{IC}_X^{\bullet}[-c] \cong j_y^! \alpha^* \mathbf{IC}_X^{\bullet}[-c].$$

Therefore $H^m(j_x^! \mathbf{IC}_X^{\bullet}) \cong H^{m-c}(j_y^! \alpha^* \mathbf{IC}_X^{\bullet})$ and

$$\begin{aligned} n - q^{-1}(-m) &\geqq \dim\{x \in X \mid H^m(j_x^! \mathbf{IC}_X^{\bullet}) \neq 0\} \\ &\geqq \dim\{y \in Y \mid H^{m-c}(j_y^! \alpha^* \mathbf{IC}_X^{\bullet}) \neq 0\} + c \end{aligned}$$

which shows $\gamma \leqq n - c - q^{-1}(-m)$ as desired.

The remaining axioms in [AX2] may be easily verified.

Part (2) of the proposition follows because

$$\begin{aligned}
\alpha^! \mathbf{IC}_X^\bullet &\cong \mathfrak{D}_Y(\alpha^* \mathfrak{D}_X(\mathbf{IC}_X^\bullet)) \\
&\cong \mathfrak{D}_Y(\alpha^* \mathfrak{D}_N(\mathbf{IC}_X^\bullet | N)) \\
&\cong \mathfrak{D}_Y(\alpha^* R\mathbf{Hom}^\bullet(\pi^* \alpha^* \mathbf{IC}_X^\bullet, \pi^* \alpha^* \mathbb{D}_X^\bullet)) \\
&\cong \mathfrak{D}_Y(\alpha^* \pi^* R\mathbf{Hom}^\bullet(\alpha^* \mathbf{IC}_X^\bullet, \alpha^* \mathbb{D}_X^\bullet)) \\
&\cong \mathfrak{D}_Y(R\mathbf{Hom}^\bullet(\alpha^* \mathbf{IC}_X^\bullet, \alpha^! \mathbb{D}_X^\bullet)[c]) \\
&\cong \mathfrak{D}_Y(\mathfrak{D}_Y(\alpha^* \mathbf{IC}_X^\bullet))[-c] \\
&\cong \alpha^* \mathbf{IC}_X^\bullet[-c] = \mathbf{IC}_Y^\bullet
\end{aligned}$$

5.4.2. *Normally Nonsingular Projections*

Definition. An oriented topological fibre bundle $\pi\colon Y \to X$ is normally nonsingular with codimension $(-c)$ if the fibre $\pi^{-1}(x)$ is a topological manifold of dimension c.

Theorem. *Let $\pi\colon Y \to X$ be a normally nonsingular fibration with codimension $-c$. Fix a perversity $\bar{p}$ and let $\mathbf{IC}_X^\bullet$ and $\mathbf{IC}_Y^\bullet$ denote the intersection homology complexes on X and Y. Then $\pi^* \mathbf{IC}_X^\bullet \cong \mathbf{IC}_Y^\bullet[-c]$ and $\pi^! \mathbf{IC}_X^\bullet = \mathbf{IC}_Y^\bullet$.*

The proofs are similar to those in § 5.4.1.

5.4.3. *Normally Nonsingular Maps*

Definition. A normally nonsingular map $f\colon Y \to X$ between oriented topological pseudomanifolds, is one which can be factored as a composition of a normally nonsingular inclusion, followed by a normally nonsingular fibration. The relative dimension of f is defined to be the sum of the codimensions of the two factors. Topological pseudomanifolds and normally nonsingular maps form a category (see [15]).

Definition. Let $f\colon Y \to X$ be a proper normally nonsingular map of relative dimension c. Then the induced homomorphisms

$$f_*\colon IH_k^{\bar{p}}(Y) \to IH_k^{\bar{p}}(X)$$

and

$$f^*\colon IH_k^{\bar{p}}(X) \to IH_{k-c}^{\bar{p}}(Y)$$

are constructed as follows.

Consider the canonical "adjunction morphisms"

$$Rf_! f^! \mathbf{IC}_X^\bullet \to \mathbf{IC}_X^\bullet \quad \text{and} \quad \mathbf{IC}_X^\bullet \to Rf_* f^* \mathbf{IC}_X^\bullet.$$

By Theorems 5.4.1 and 5.4.2 these become morphisms

$$Rf_! \mathbf{IC}_Y^\bullet \to \mathbf{IC}_X^\bullet \quad \text{and} \quad \mathbf{IC}_X^\bullet \to Rf_* \mathbf{IC}_Y^\bullet[c]$$

Since f is proper, $Rf_! = Rf_*$. Taking hypercohomology gives homomorphisms

$$IH_k^{\bar{p}}(Y) = \mathscr{H}^{-k}(X; Rf_* \mathbf{IC}_Y^\bullet) \to \mathscr{H}^{-k}(X; \mathbf{IC}_X^\bullet) = IH_k^{\bar{p}}(X)$$

and

$$IH_k^{\bar{p}}(X) = \mathscr{H}^{-k}(X; \mathbf{IC}_X^\bullet) \to \mathscr{H}^{-k}(X; Rf_* \mathbf{IC}_Y^\bullet[c]) = IH_{k-c}^{\bar{p}}(Y)$$

Proposition. *$IH_k^{\bar{p}}$ is both a covariant functor (via f_*) and a contravariant functor (via f^*) on the category of topological pseudomanifolds and normally non-singular maps.*

$\mathscr{L}$. If f: $Y \to X$ is a normally nonsingular map of topological pseudo-manifolds, then X and Y can be stratified so that the inverse image of the largest stratum $X - \Sigma$ of X is the largest stratum of Y. If c is the relative dimension of f, then for any local system $\mathbf{F}$ on $X - \Sigma$,

$$\pi^* \mathbf{IC}_X^\bullet(\mathbf{F}) \cong \mathbf{IC}_Y^\bullet(\pi^* \mathbf{F})[-c]$$

and

$$\pi^! \mathbf{IC}_X^\bullet(\mathbf{F}) \cong \mathbf{IC}_Y^\bullet(\pi^* \mathbf{F} \otimes \mathbf{Hom}(\pi^* \mathcal{O}_X, \mathcal{O}_Y)).$$

5.5. *The Obstruction Sequence for Comparing two Perversities*

It is clear from Deligne's construction that whenever $\bar{p} \leqq \bar{q}$ are perversities, there is a canonical morphism $\mathbf{IC}_{\bar{p}}^\bullet \to \mathbf{IC}_{\bar{q}}^\bullet$. Thus we obtain a distinguished triangle

$$\begin{tikzcd}
\mathbf{IC}_{\bar{p}}^\bullet \arrow[rr] & & \mathbf{IC}_{\bar{q}}^\bullet \arrow[dl] \\
& \mathbf{S}^\bullet \arrow[ul, "{[1]}"] &
\end{tikzcd}$$

and a long exact sequence on hypercohomology,

$$\to IH_i^{\bar{p}}(X) \to IH_i^{\bar{q}}(X) \to \mathscr{H}^{-i}(X; \mathbf{S}^\bullet) \to IH_{i-1}^{\bar{p}}(X) \to \ldots$$

which is called the obstruction sequence because $\mathscr{H}^{-i}(X; \mathbf{S}^\bullet)$ is the obstruction to lifting a class from $IH_i^{\bar{q}}(X)$ to $IH_i^{\bar{p}}(X)$.

Now fix a stratification $\{X_k\}$ of X.

Proposition. *Suppose $p(c) = q(c)$ for all $c \neq k$, and $q(k) = p(k) + 1$. Then:*

(1) *$spt\, \mathbf{H}^*(\mathbf{S}^\bullet) = closure(spt\, \mathbf{H}^*(\mathbf{S}^\bullet) \cap (X_{n-k} - X_{n-k-1}))$ where spt denotes the support of a sheaf.*

(2) *If $x \in X_{n-k} - X_{n-k-1}$ then*

$$\begin{aligned} &\mathbf{H}^i(\mathbf{S}^\bullet)_x = 0 && \text{for all } i \neq q(k) - n, \\ &\mathbf{H}^i(\mathbf{S}^\bullet)_x \cong \mathbf{H}^i(\mathbf{IC}_{\bar{q}}^\bullet)_x && \text{if} \quad i = q(k) - n. \end{aligned}$$

(3) *If $\mathbf{H}^{q(k)-n}(\mathbf{IC}_{\bar{q}}^\bullet)_x = 0$ for all $x \in X_{n-k} - X_{n-k-1}$ then $\mathbf{IC}_{\bar{p}}^\bullet \to \mathbf{IC}_{\bar{q}}^\bullet$ is a quasi isomorphism.*

The proof follows directly from Deligne's construction.

$\mathscr{L}$. The results of this section hold equally well when $\mathbf{IC}^\bullet_{\bar{p}}$ is replaced by $\mathbf{IC}^\bullet_{\bar{p}}(F)$ and $\mathbf{IC}^\bullet_{\bar{q}}$ is replaced by $\mathbf{IC}^\bullet_{\bar{q}}(F)$.

5.6. *Normal Varieties, Local Complete Intersection, and Witt Spaces*

5.6.1. *Witt Spaces.* Let k be a field and let X be a n-dimensional stratified piecewise linear pseudomanifold [20].

Definition. X is a k-Witt space if $IH^{\bar{m}}_l(L_x; k)=0$ for all $x \in X_{n-2l-1} - X_{n-2l-2}$. Here, L_x is the link of the stratum containing x, and $\bar{m}=(0, 0, 1, 1, 2, 2, \ldots)$.

Proposition (P. Siegel [35]). *X is a k-Witt space if and only if the canonical morphism $\mathbf{IC}^\bullet_{\bar{m}} \to \mathbf{IC}^\bullet_{\bar{n}}$ is a quasi isomorphism. Thus, if X is a Witt space $IH^{\bar{m}}_n(X; k)$ is self dual.*

The proof follows from the identification (§2.2) of the stalk $H^{\bar{n}(2l+1)-n}(I\,C^\bullet_{\bar{n}})_x$ with $IH_l(L_x)$ for any $x \in X_{n-2l-1} - X_{n-2l-2}$. Then apply Proposition 5.5(2).

5.6.2. We also have obstruction groups relating perversities $\bar{0}$ and $\bar{t}$ to cohomology and homology (if X is oriented),

$$\begin{tikzcd} \mathbf{R}_X \arrow[rr] & & \mathbf{IC}^\bullet_{\bar{0}} \arrow[dl] \\ & \mathbf{S}^\bullet_1 \arrow[ul, "{[1]}"] & \end{tikzcd} \qquad \begin{tikzcd} \mathbf{IC}^\bullet_{\bar{t}} \arrow[rr] & & \mathbb{D}^\bullet_X \arrow[dl] \\ & \mathbf{S}^\bullet_2 \arrow[ul, "{[1]}"] & \end{tikzcd}$$

The cohomology sheaf $\mathbf{H}^i(\mathbf{S}^\bullet_1)$ vanishes except for $i=0$ and $\mathbf{H}^i(\mathbf{S}_2)$ vanishes except for $i=-n$. For a point $p \in X$ the stalks of $\mathbf{H}^0(\mathbf{S}^\bullet_1)$ and $\mathbf{H}^{-n}(\mathbf{S}^\bullet_2)$ at p are both free R modules of rank $r-1$ where r is the local Betti number, rank $H_n(X, X-p; R)$, at p. The number r has two other interpretations: If $X = X_n \supset X_{n-1} \supset \ldots$ is any stratification of X and N is a distinguished neighborhood of p (see §1.1), then r is the number of connected components of $(X - X_{n-2}) \cap N$. If X is a complex analytic variety, then r is the number of analytic branches at p.

Definition. A *normal* topological pseudomanifold X of dimension n is one such that rank $H_n(X, X-p; R)=1$ for all $p \in X$.

By Zariski's main theorem, a normal complex algebraic variety is normal as a topological pseudomanifold. By the above remarks, we have the

Proposition. *For a normal oriented n-dimensional topological pseudomanifold X, we have*

$$\mathbf{R}_X \cong \mathbf{IC}^\bullet_{\bar{0}} \quad \text{and} \quad \mathbf{IC}^\bullet_{\bar{t}} \cong \mathbb{D}^\bullet_X$$

so

$$H^i(X) \cong IH^{\bar{0}}_{n-i}(X) \quad \textit{and} \quad IH^{\bar{t}}_i(X) \cong H_i(X).$$

5.6.3. *Local Complete Intersections*

Proposition. *Let Y be a compact complex algebraic variety which is normal and is a local complete intersection. Let $\bar{p}$ be a perversity such that $p(k) \geqq \frac{k}{2}$ for each $k \geqq 4$. Then for all i we have $IH_i^{\bar{p}}(Y) \cong H_i(Y)$.*

Proof. Let us say that an n-dimensional triangulable space X is a space of type Q if it is a normal pseudomanifold, has a stratification by even codimension and orientable strata, and if for each $x \in X_{n-c} - X_{n-c-1}$ the local homology groups $H_i(X, X-x)$ vanish for all $i \leqq n-1-\frac{c}{2}$. Hamm [22] shows that a normal local complete intersection is a space of type Q, and we shall now show that the conclusion of the proposition holds for any n-dimensional space X of type Q.

For each $x \in X_{n-c} - X_{n-c-1}$ the link L_x is an $n-1$ dimensional space of type Q so by induction, the proposition applies to L_x. Thus,

$$IH_i^{\bar{p}}(X, X-x) \cong IH_{i-1}^{\bar{p}}(L_x) \cong H_{i-1}(L_x) \cong H_i(X, X-x) = 0$$

provided $i \leqq n-1-\frac{c}{2}$ and $p(k) \geqq \frac{k}{2}$ for all k.

Applying Proposition 5.5 to any string of perversities between $\bar{p}$ and $\bar{t}$ (where $t(c)=(c-2)$) we conclude that for all i,

$$IH_i^{\bar{p}}(X) \cong IH_i^{\bar{t}}(X) \cong H_i(X)$$

since X is normal.

§ 6. The Middle Group

6.0. In this chapter, X will denote an *oriented* topological pseudomanifold, except in the paragraph marked $\mathscr{L}$ but we do not fix any particular stratification of X. Except in § 6.2 we will assume the ring R is a field, which we now denote by k. We shall consider the middle perversities $\bar{m}=(0, 0, 1, 1, 2, 2, \ldots)$ and $\bar{n}=(0, 1, 1, 2, 2, 3, \ldots)$ with their corresponding complexes $\mathbf{IC}_{\bar{m}}^{\bullet}$ and $\mathbf{IC}_{\bar{n}}^{\bullet}$ in the derived category of the category of sheaves of k-vectorspaces.

In § 6.2 and § 6.3, X will be a k-Witt space, so $\mathbf{IC}_{\bar{m}}^{\bullet} = \mathbf{IC}_{\bar{n}}^{\bullet}$ which will be denoted simply $\mathbf{IC}^{\bullet}$ or $\mathbf{IC}_X^{\bullet}$.

6.1. *Axioms* [AX3]

Definition. Let $\mathbf{S}^{\bullet}$ be a topologically constructible complex of sheaves (of k-modules) on X. We shall say $\mathbf{S}^{\bullet}$ satisfies axioms [AX3] provided:

(a) Normalization
There is a closed subspace $Z \subset X$ such that $\mathbf{S}^{\bullet}|(X-Z) \cong \mathbf{k}_{X-Z}[n]$ and $\dim(Z) \leqq n-2$.

(b) Lower bound

$\mathbf{H}^m(\mathbf{S}^\bullet)=0 \quad$ for all $m<-n$.

(c) Support condition

$\dim\{x\in X \mid \mathbf{H}^i(\mathbf{S}^\bullet)_x \neq 0\} \leqq -2i-n-2 \quad$ for $i\geqq -n+1$.

(d) Duality

There is an isomorphism $\mathbf{S}^\bullet \cong \mathfrak{D}(\mathbf{S}^\bullet)[n]$.

Theorem. *If the complex* $\mathbf{S}^\bullet$ *satisfies axioms* [AX3] *then there is a natural isomorphism in* $D^b(X)$, $\mathbf{S}^\bullet \cong \mathbf{IC}^\bullet_{\bar{n}}$. *It follows that* $\mathbf{IC}^\bullet_{\bar{m}} \cong \mathbf{IC}^\bullet_{\bar{n}}$ *so* X *is a k-Witt space.*

Proof. We shall show $\mathbf{S}^\bullet \cong \mathbf{IC}^\bullet_{\bar{m}}$ by verifying the axioms [AX2]. Note that $m^{-1}(c)=2c+2$.

Axioms AX2(a)(b)(c) are obviously satisfied. For any $x\in X$ we have

$$\begin{aligned} j_x^! \mathbf{S}^\bullet &\cong R\,\mathbf{Hom}^\bullet(j_x^* R\,\mathbf{Hom}^\bullet(\mathbf{S}^\bullet, \mathbb{D}^\bullet_X), \mathbb{D}^\bullet_{\{x\}}) \\ &\cong \mathrm{Hom}(j_x^* \mathbf{S}^\bullet[-n], k) \\ &\cong \mathrm{Hom}(j_x^* \mathbf{S}^\bullet, k)[n] \end{aligned}$$

so $H^i(j_x^! \mathbf{S}^\bullet) \cong \mathrm{Hom}(H^{-n-i}(j_x^* \mathbf{S}^\bullet), k)$.

The set of points for which this does not vanish has (by AX3(c)) dimension $\leqq n+2i-2 \leqq n+2i-1 = n-p^{-1}(-i)$ where $\bar{p}$ is the perversity complementary to $\bar{m}$. This verifies axiom AX2(d), so $\mathbf{S}^\bullet = \mathbf{IC}^\bullet_{\bar{m}}$. By duality we also obtain $\mathbf{S}^\bullet = \mathbf{IC}^\bullet_{\bar{n}} \cong \mathbf{IC}^\bullet_{\bar{m}}$.

$\mathscr{L}$. **Definition.** A complex of sheaves $\mathbf{S}^\bullet$ is a *middle intersection homology sheaf* if for some stratification of X and for some local coefficient system $\mathbf{F}$ on $X-\Sigma$,

$$\mathbf{S}^\bullet = \mathbf{IC}^\bullet_{\bar{m}}(\mathbf{F}) = \mathbf{IC}^\bullet_{\bar{n}}(\mathbf{F}).$$

Theorem. *A complex of sheaves* $\mathbf{S}^\bullet$ *is a middle intersection homology sheaf if and only if both* $\mathbf{S}^\bullet$ *and* $\mathfrak{D}(\mathbf{S}^\bullet)$ *satisfy* [AX3](b) (*lower bound*) *and* [AX3](c) (*support condition*).

In this case, $\mathfrak{D}(\mathbf{S}^\bullet) = \mathbf{IC}^\bullet_{\bar{m}}(\mathbf{Hom}(\mathbf{F}, \mathcal{O}))$.

6.2. *Small Maps and Resolutions*

In contrast to the rest of this chapter, the results in this section are valid over an arbitrary finite dimensional regular Noetherian ring R of coefficients.

Definition. A proper surjective algebraic map $f\colon Y\to X$ between irreducible complex n-dimensional algebraic varieties is *homologically small* if for all $q>-2n$.

$$\mathrm{cod}_{\mathbb{C}}\{x\in X \mid \mathbf{H}^q(Rf_* \mathbf{IC}^\bullet_Y)_x \neq 0\} > q+2n.$$

It follows that there is a Zariski open set $U\subset X$ such that $f|f^{-1}(U)$ is a finite covering projection. The above criterion is satisfied, for example, if Y is the normalization of X or if f is a *small map*, i.e., if Y is nonsingular and for all $r>0$,

$$\operatorname{cod}_{\mathbb{C}}\{x \in X \mid \dim_{\mathbb{C}} f^{-1}(x) \geqq r\} > 2r.$$

Examples of Small Maps. If X is one or two dimensional than a small map $f\colon Y \to X$ must be a finite map. If X is a threefold then the fibres of a small map f must be zero dimensional except possibly over a set of isolated points in X where the fibres may be at most curves.

Theorem. *Let $f\colon Y \to X$ be a homologically small map of degree 1. Then $Rf_* \mathbf{IC}^\bullet_Y \cong \mathbf{IC}^\bullet_X$ and in particular $\mathbf{IH}^{\bar m}_*(Y) \cong \mathbf{IH}^{\bar m}_*(X)$.*

Proof. The complex $Rf_* \mathbf{IC}^\bullet_Y$ satisfies the criteria (AX2) of Theorem 4.1.

Definition. A *small resolution* $f\colon Y \to X$ is a resolution of singularities which is a small map.

Corollary. *If $f\colon Y \to X$ is a small resolution then the intersection homology groups of X equal the (ordinary) homology groups of Y, and $Rf_* \mathbf{R}_Y \cong \mathbf{IC}^\bullet_X$.*

Remark. Small resolutions do not always exist, and are not necessarily unique when they do exist. However, if X has several small resolutions, their cohomologies are isomorphic as groups, but not as rings. Two small resolutions must have the same signature and Euler characteristic.

$\mathscr{L}$. **Theorem.** *If $f\colon Y \to X$ is homologically small, then*

$$Rf_* \mathbf{IC}^\bullet_Y = \mathbf{IC}^\bullet_X(R^0(f \mid U)_* \mathbf{R}_{f^{-1}(U)}),$$

where U is the Zariski open set over which f is a covering projection.

6.3. *Kunneth Formula*

For $k = \mathbb{R}$ the following proposition was first proved by Cheeger using L^2 cohomology [10].

Proposition. *Suppose X and Y are Witt spaces. Then*

$$IH^{\bar m}_l(X \times Y) = \bigoplus_{a+b=l} IH^{\bar m}_a(X) \otimes IH^{\bar m}_b(Y).$$

In particular, the signature of $X \times Y$ is the product of the signatures of X and Y.

Proof. Let $\pi_1\colon X \times Y \to X$ and $\pi_2\colon X \times Y \to Y$ be the projections. It suffices to prove that $\mathbf{IC}^\bullet_{X \times Y} \cong \pi_1^* \mathbf{IC}^\bullet_X \overset{L}{\otimes} \pi_2^* \mathbf{IC}^\bullet_Y$, which is done by verifying the axioms. The support condition is easy to verify, and duality holds because

$$\begin{aligned}
&R\,\mathbf{Hom}^\bullet(\pi_1^* \mathbf{IC}^\bullet_X \overset{L}{\otimes} \pi_2^* \mathbf{IC}^\bullet_Y, \mathbb{D}^\bullet_{X \times Y}) \\
&\quad \cong R\,\mathbf{Hom}^\bullet(\pi_1^* \mathbf{IC}^\bullet_X \overset{L}{\otimes} \pi_2^* \mathbf{IC}^\bullet_Y, \pi_1^* \mathbb{D}^\bullet_X \overset{L}{\otimes} \pi_2^* \mathbb{D}^\bullet_Y) \\
&\quad \cong R\,\mathbf{Hom}^\bullet(\pi_1^* \mathbf{IC}^\bullet_X, \pi_1^* \mathbb{D}^\bullet_X) \overset{L}{\otimes} R\,\mathbf{Hom}^\bullet(\pi_2^* \mathbf{IC}^\bullet_Y, \pi_2^* \mathbb{D}^\bullet_Y) \\
&\quad \cong \pi_1^* \mathbf{IC}^\bullet_X \overset{L}{\otimes} \pi_2^* \mathbf{IC}^\bullet_Y[-n-m]
\end{aligned}$$

where $n = \dim(X)$ and $m = \dim(Y)$.

$\mathscr{L}$. Similarly, the characterization of middle intersection homology sheaves in § 6.1 $\mathscr{L}$ can be used to show

$$\mathbf{IC}^{\bullet}_{X\times Y}(\pi_1^*\mathbf{F}_1\otimes\pi_2^*\mathbf{F}_2)\cong\pi_1^*\mathbf{IC}^{\bullet}_X(\mathbf{F}_1)\overset{L}{\otimes}\pi_2^*\mathbf{IC}^{\bullet}_Y(\mathbf{F}_2)$$

so

$$IH_l^{\bar m}(X\times Y;\pi_1^*\mathbf{F}_1\otimes\pi_1^*\mathbf{F}_2)\cong\bigoplus_{a+b=l} IH_a^{\bar m}(X;\mathbf{F}_1)\otimes IH_b^{\bar m}(Y;\mathbf{F}_2).$$

§ 7. The Lefschetz Theorem on Hyperplane Sections

The purpose of this chapter is to show that the classical theorem of Lefschetz (on the homology of a hyperplane section of a nonsingular projective variety) continues to hold in the singular case provided we replace homology by intersection homology. This holds for a range of perversities which include the middle perversity $\bar m$ and the logarithmic perversity $\bar l$. In § 7.4 we deduce as corollaries of this theorem some results about ordinary homology and cohomology.

Our original proof of this theorem proceeded by replacing Thom's Morse-theoretic argument in the nonsingular case ([2, 31]) with a stratified Morse theoretic argument in the singular case (using the techniques of [21]). The proof we present here was pointed out to us by Deligne (who had also observed that the theorem is true). It is essentially the same as the proof in the theorem of Artin [1]. We have also made use of some ideas of K. Vilonen [46].

7.0 Throughout this chapter, X will denote a complex projective n dimensional variety with its canonical orientation, and all homology groups will be understood to take coefficients in a finite dimensional regular Noetherian ring R.

7.1. **Theorem.** *Suppose X is a complex n-dimensional algebraic variety embedded in complex projective space. Fix a perversity $\bar p$ such that $p(c)\leqq c/2$. Let $Y=H\cap X$ where H is a hyperplane which is transverse to each stratum of a Whitney stratification of X. Then the normally nonsingular inclusion α: $Y\to X$ induces isomorphisms α_*: $IH_k^{\bar p}(Y)\xrightarrow{\cong}IH_k^{\bar p}(X)$ for all $k<n-1$ and a surjection $\alpha_*: IH_{n-1}^{\bar p}(Y)\to IH_{n-1}^{\bar p}(X)$.*

7.2. *Intersection Homology of Affine Varieties*

This section contains the technical tools needed in the proof of the Lefschetz theorem.

Lemma. *Suppose* $\mathbf{A}$ *is an algebraically constructible sheaf on* C. *Then* $H^2(\mathbb{C},\mathbf{A})=0$.

Proof. $\mathbf{A}$ is locally trivial except on a finite set of points ([8] Exp. 7, 8). Let K be the union of the line segments joining the origin to each of these points. Then $H^*(K,\mathbf{A}|K)\cong H^*(\mathbb{C},\mathbf{A})$. But K is one dimensional.

Proposition. *Let* $\mathbf{S}^\bullet$ *be a complex of algebraically constructible sheaves on* $\mathbb{C}^n$, *which satisfies a support condition*

$$\dim_{\mathbb{C}}\{x \in \mathbb{C}^n \mid \mathbf{H}^m(\mathbf{S}^\bullet)_x \neq 0\} \leqq f(m)$$

where $f\colon \mathbb{Z} \to \mathbb{Z}$ *is a nonincreasing integer valued function. Then* $\mathscr{H}^j(\mathbb{C}^n; \mathbf{S}^\bullet)=0$ *for all* $j > \max\{m+f(m) \mid f(m) \geqq 0\}$. *The same holds if we replace* $\mathbb{C}^n$ *by an arbitrary affine algebraic variety.*

Proof. If $n=1$ the proposition follows from studying the spectral sequence associated to $\mathbf{S}^\bullet$ with E^2 term

$$E^2_{rs} = H^r(\mathbb{C}; \mathbf{H}^s(\mathbf{S}^\bullet)) \Rightarrow \mathscr{H}^{r+s}(\mathbb{C}; \mathbf{S}^\bullet).$$

For $r \geqq 2$ we have $E^2_{rs}=0$ by the previous lemma. Furthermore,

$$\begin{aligned} E^2_{0s} &= H^0(\mathbb{C}; \mathbf{H}^s(\mathbf{S}^\bullet))=0 \quad \text{if } s > f^{-1}(0),\\ E^2_{1s} &= H^1(\mathbb{C}; \mathbf{H}^s(\mathbf{S}^\bullet))=0 \quad \text{if } s > f^{-1}(1)\end{aligned}$$

where $f^{-1}(r) = \max\{m \mid f(m) \geqq r\}$ is the "super-inverse" of f. Consequently

$$\bigoplus_{r+s=j} E^2_{rs}=0 \quad \text{if } j > \max_{r=0,1}\,[r+f^{-1}(r)] = \max\{f(m)+m \mid 0 \leqq f(m) \leqq 1\}.$$

which implies $\mathscr{H}^j(\mathbb{C}; \mathbf{S}^\bullet)=\mathbf{0}$ in this range also.

Remark. The super-inverse, f^{-1}, may be interpreted as follows: Suppose $\mathbf{S}^\bullet$ is constructible with respect to a stratification $X_0 \subset X_1 \subset \ldots \subset X_n = \mathbb{C}^n$ by complex k dimensional algebraic varieties X_k. The support condition means that for any point $x \in X_r - X_{r-1}$ the stalk cohomology $\mathbf{H}^k(\mathbf{S}^\bullet)_x$ vanishes for all $k > f^{-1}(r)$.

We now proceed by induction on n. Suppose $\mathbf{S}^\bullet$ is a complex of sheaves on $\mathbb{C}^n$ which satisfies the hypotheses of the proposition. Choose a stratification of $\mathbb{C}^n$,

$$X_0 \subset X_1 \subset \ldots \subset X_{n-1} \subset X_n = \mathbb{C}^n$$

by complex i-dimensional algebraic varieties X_i such that $\mathbf{S}^\bullet$ is constructible with respect to this stratification (§1.4).

We can find a linear projection $\pi\colon \mathbb{C}^n \to \mathbb{C}^{n-1}$ with the property that the stalk homology of the complex of sheaves $\mathbf{B}^\bullet = R\pi_* \mathbf{S}^\bullet$ at any point $y \in \mathbb{C}^{n-1}$ may be identified with the hypercohomology of the fibre $\pi^{-1}(Y)$, i.e.,

$$\mathbf{H}^i(R\pi_*(\mathbf{S}^\bullet))_y = \mathscr{H}^i(\pi^{-1}(y); \mathbf{S}^\bullet).$$

Such a $\pi\colon \mathbb{C}^n \to \mathbb{C}^{n-1}$ may be constructed as follows: Embed $\mathbb{C}^n$ in $\mathbb{CP}^n$ by adding a $\mathbb{CP}^{n-1}$ at infinity. Complete the stratification $X_0 \subset X_1 \subset \ldots \subset X_n$ of $\mathbb{C}^n$ to a stratification of $\mathbb{CP}^n$ by stratifying the $\mathbb{CP}^{n-1}$ at infinity (i.e. without refining the stratification of $\mathbb{C}^n$). Let p be a point in the largest stratum of the $\mathbb{CP}^{n-1}$ at infinity, and define π by projection along the parallel lines which pass through p. This fibration of $\mathbb{C}^n$ by lines contains a subbundle D of compact discs such that X_{n-1} is contained in the interior of D, Now, $R\pi_* \mathbf{S}^\bullet = R\pi_* \mathbf{S}^\bullet | D$ and $\pi | D$ is proper. By [17] §4.17.1 we have,

$$\mathbf{H}^i(R\pi_*(\mathbf{S}^\bullet))_y = \mathscr{H}^i(\pi^{-1}(y) \cap D; \mathbf{S}^\bullet) = \mathscr{H}^i(\pi^{-1}(y); \mathbf{S}^\bullet).$$

We will now apply the case $n=1$ to the computation of these stalk cohomology groups.

It is possible to find a stratification $Y_0 \subset \ldots \subset Y_{n-1} = \mathbb{C}^{n-1}$ by complex i-dimensional algebraic varieties Y_i, and a refinement $X_0^1 \subset X_1^1 \subset \ldots \subset X_n^1 = \mathbb{C}^n$ of the stratification $\{X_i\}$ of $\mathbb{C}^n$ such that $\pi\colon \mathbb{C}^n \to C^{n-1}$ takes strata to strata. (In other words, stratify the map π).

Suppose $y \in Y_r - Y_{r-1}$. We claim that $H^k(R\pi_* \mathbf{S}^\cdot)_y = 0$ for all $k > \max[f^{-1}(r), 1+f^{-1}(r+1)]$. To see this, consider the stratification of $\pi^{-1}(y)$ obtained from intersecting with the $\{X_i'\}$. Clearly $\mathbf{S}^\cdot | \pi(y)$ is constructible with respect to this stratification. A point $x \in \pi^{-1}(y)$ which lies in a zero dimensional stratum of $\pi^{-1}(y)$ may be an element of any stratum of X except those strata in X_{r-1}. Therefore $\mathbf{H}^k(\mathbf{S}^\cdot)_x = 0$ for all

$$k > \max[f^{-1}(r), f^{-1}(r+1), \ldots, f^{-1}(n)] = f^{-1}(r).$$

A point $x' \in \pi^{-1}(y)$ which lies in the one-dimensional stratum of $\pi^{-1}(y)$ may be an element of any stratum in $X - X_r$. Therefore, $\mathbf{H}^k(\mathbf{S}^\cdot)_{x'} = 0$ for all

$$k > \max[f^{-1}(r+1), f^{-1}(r+2), \ldots, f^{-1}(n)] = f^{-1}(r+1).$$

Applying the case $n=1$ to the fibre $\pi^{-1}(y)$ we obtain $\mathscr{H}^k(\pi^{-1}(y); \mathbf{S}^\cdot) = 0$ for all $k > \max[1+f^{-1}(r+1), f^{-1}(r)]$.

We will now apply the induction hypothesis to the complex of sheaves $R\pi_* \mathbf{S}^\cdot$ on $\mathbb{C}^{n-1}$. This complex satisfies a support condition

$$\dim\{y \in \mathbb{C}^{n-1} \mid \mathbf{H}^m(R\pi_* \mathbf{S}^\cdot)_y \neq 0\} \leqq g(m)$$

where, for every $r \geqq 0$

$$g^{-1}(r) \leqq \max[1+f^{-1}(1+r), f^{-1}(r)].$$

Therefore $H^k(\mathbb{C}^n; \mathbf{S}^\cdot) = H^k(\mathbb{C}^{n-1}; R\pi_* \mathbf{S}^\cdot) = 0$ whenever

$$k > \max\{m+g(m) \mid 0 \leqq g(m) \leqq n-1\} = \max\{r+g^{-1}(r) \mid 0 \leqq r \leqq n-1\}.$$

This condition will be satisfied if

$$k > \max\{r+f^{-1}(r) \mid 0 \leqq r \leqq n\} = \max\{m+f(m) \mid 0 \leqq f(m) \leqq n\}$$

as claimed.

7.3. *Proof of the Lefschetz Theorem*

Fix a perversity $\bar{p}$ with $p(c) \leqq c/2$ and let $\mathbf{IC}_X^\cdot$ denote the associated intersection homology complex on X. By §5.4.1 the triangle

$$\begin{tikzcd} R\alpha_* \alpha^! \mathbf{IC}_X^\cdot \arrow[rr] & & \mathbf{IC}_X^\cdot \arrow[dl] \\ & Ri_* i^* \mathbf{IC}_X^\cdot \arrow[ul, "{[1]}"] & \end{tikzcd} \quad \text{becomes} \quad \begin{tikzcd} R\alpha_* \mathbf{IC}_Y^\cdot \arrow[rr] & & \mathbf{IC}_X^\cdot \arrow[dl] \\ & Ri_* i^* \mathbf{IC}_X^\cdot \arrow[ul, "{[1]}"] & \end{tikzcd}$$

where i: $(X-Y)\to X$ is the inclusion of the complement of the hyperplane section. The long exact sequence on hypercohomology associated with this triangle is

$$\to IH_i^{\bar p}(Y)\to IH_i^{\bar p}(X)\to \mathscr{H}^{-i}(R\,i_* i^*\mathbf{IC}_X^\bullet)\to IH_{i-1}^{\bar p}(Y)\to$$

so we must show $\mathscr{H}^{-i}(i^*\mathbf{IC}_X^\bullet)=\mathscr{H}^{-i}(R\,i_* i^*\mathbf{IC}_X^\bullet)=0$ for all $i\leqq n-1$.

Since X can be stratified by algebraic subvarieties, the complex $\mathbf{IC}_X^\bullet$ is algebraically constructible. Therefore the complex $i^*\mathbf{IC}_X^\bullet$ on the affine variety $X-X\cap H$ satisfies the hypotheses of Prop. 7.2 with the support condition given by axioms (AX2):

$$\dim_{\mathbb{C}}\{x\,|\,\mathbf{H}^{-2n}(i^*\mathbf{IC}_X^\bullet)_x\neq 0\}\leqq n$$

$$\dim_{\mathbb{C}}\{x\,|\,\mathbf{H}^{m}(i^*\mathbf{IC}_X^\bullet)_x\neq 0\}\leqq n-\frac{p^{-1}(m+2n)}{2}\qquad \text{for } m\geqq -2n+1.$$

Since $p(c)\leqq\frac{c}{2}$ we have $\frac{p^{-1}(m+2n)}{2}\geqq m+2n$ so

$$\dim_{\mathbb{C}}\{x\,|\,\mathbf{H}^{m}(i^*\mathbf{IC}_X^\bullet)_x\neq 0\}\leqq -m-n\qquad \text{for } m\geqq -2n.$$

The conclusion of Prop. 7.2 now reads

$$\mathscr{H}^{j}(X-Y;i^*\mathbf{IC}_X^\bullet)=0\qquad \text{for all } j>-n \text{ as desired.}$$

Remark. Decreasing the perversity does not give better bounds on j because the dimension of the support of $\mathbf{H}^{-2n}(i^*\mathbf{IC}_X^\bullet)$ is always n. However, if we increase the perversity past $p(c)=c/2$ the Lefschetz theorem continues to hold, although only for a smaller range of dimensions j. For a general perversity $\bar p\geqq c/2$ we have the following theorem:

$$IH_i(Y)\to IH_i(X) \text{ is an isomorphism for } i<j^*$$

and is a surjection for $i=j^*$, where

$$j^*=\max_m\left[m+n-\frac{p^{-1}(m+2n)}{2}\right]-1.$$

7.4. *Consequences in Ordinary Homology*

7.4.1. **Corollary.** *Let X be a normal n-dimensional projective variety and let $Y=X\cap H$ be a generic hyperplane section of X. Then the Gysin homomorphism (in ordinary cohomology)*

$$H^k(Y)\to H^{k+2}(X)$$

is an isomorphism for $k>n-1$ and is a surjection for $k=n-1$.

Proof. Take $p=0$ in Theorem 7.1. (This corollary can also be proved using the result of [27] and [32] on the vanishing of cohomology of Stein spaces.)

The following corollary was proved by Kato [26], Oka [34] and Ogus [33], and Kaup [27], [28].

7.4.2. **Corollary.** *Let X be a local complete intersection which is normal and let $Y=H\cap X$ be a generic hyperplane section of X. Then the homomorphism induced by inclusion*

$$H_i(Y)\to H_i(X)$$

is an isomorphism for $i<n-1$ and a surjection for $i=n-1$.

Proof. Let $p(c)=c/2$. Then by Prop. 2.3.3 we have $IH_i^{\bar{p}}(X)\cong H_i(X)$ for all i. The same holds for Y. Therefore the Lefschetz theorem (7.1) for intersection homology implies the same result in ordinary homology. The following corollary was pointed out to us by Horrocks [11].

7.4.3. **Corollary.** *Let X be a normal projective algebraic variety and let $\beta_1=\operatorname{rank}(H_1(X))$. Then β_1 is even.*

Proof. For a normal variety, $IH_1^{\bar{p}}(X)\cong H_1(X)$ for any perversity. Apply the Lefschetz theorem to successive hyperplane sections of X until we arrive at a two-dimensional variety Z with isolated singularities. Then $\beta_1(X)=\beta_1(Z)$ which is even (one verifies this directly).

7.4.4. *Remark.* The Lefschetz theorem for the middle perversity $\bar{m}$ is discussed in [11] as evidence that $IH_*^{\bar{m}}(X)$ has a pure Hodge structure.

§ 9. Generalized Deligne's Construction and Duality

We have already proved that if $\bar{p}(c)+\bar{q}(c)=c-2$ for all c, then the sheaves $\mathbf{P}^\bullet$ and $\mathbf{Q}^\bullet$ resulting from Deligne's construction with perversities $\bar{p}$ and $\bar{q}$ respectively have a canonical Verdier dual pairing. The proof was dispersed throughout §3, §4, and §5. Here we use the techniques of those chapters to study directly the relation between a single step in Deligne's construction and Verdier duality.

9.0. In this chapter R is a finite dimensional regular Noetherian ring (however in Sect. 9.2 we will assume R is a field). The space X is a stratified n-dimensional topological pseudomanifold and $U\subset X$ is an open union of connected components of strata. The closed subspace $Y=X-U$ is also a union of connected components of strata. Let i: $U\to X$ and j: $Y\to X$ denote the inclusions. Let $D_c^b(X)$ denote the derived category of sheaves of R-modules which are constructible with respect to this stratification

9.1. Fix an integer p. Let $\mathscr{C}(p)$ be the full subcategory of $D_c^b(X)$ whose objects $\mathbf{B}^\bullet$ satisfy the following conditions:

(a) $\mathbf{H}^i(j^*\mathbf{B}^\bullet)=0$ for all $i\geqq p$,

(b) $\mathbf{H}^i(j^!\mathbf{B}^\bullet)=0$ for all $i\leqq p$.

Theorem. *The functor $\tau^Y_{\leqq p-1}Ri_*$: $D_c^b(U)\to D_c^b(X)$ (see §1.4 and §1.14) takes its values in $\mathscr{C}(p)$. This functor determines an equivalence of categories $D_c^b(U)\cong\mathscr{C}(p)$ whose inverse is i^*.*

Proof. $\tau^Y_{\leqq p-1}Ri_*\mathbf{A}^\bullet$ is constructible by an argument similar to that in Lemma 3.1. The equivalence of categories argument is similar to that in the proof of Theorem 3.5.

9.2. Now suppose R is a field and Y is a k-dimensional R-homology manifold (for some $k<n$). Define $\mathscr{C}'_1$ to be the full subcategory of $D^b_c(U)$ whose objects $\mathbf{A}^\bullet$ satisfy

(c) $j^* R i_* \mathbf{A}^\bullet$ is CLC (§1.4),

(d) $j^! R i_* \mathbf{A}^\bullet$ is CLC.

Let $\mathscr{C}'_2(p)$ be the full subcategory of $D^b_c(X)$ whose objects $\mathbf{B}^\bullet$ satisfy conditions (a) and (b) of §9.1 and satisfy $\mathbf{B}^\bullet|U$ is in $\mathscr{C}'_1$.

Theorem. *The functor $\tau^Y_{\leqq p-1} R i_*: \mathscr{C}'_1 \to \mathscr{C}'_2(p)$ is an equivalence of categories.*

(The proof is similar to the proof of Theorem 9.1.)

Corollary. *If $p+q=-k$ then there is a one to one correspondence between pairings on U*

$$\sigma\colon\ \mathbf{A}^\bullet \otimes \mathbf{B}^\bullet \to \mathbb{D}^\bullet_U$$

and pairings on X,

$$\sigma'\colon\ \tau^Y_{\leqq p-1} R i_* \mathbf{A}^\bullet \otimes \tau^Y_{\leqq q-1} R i_* \mathbf{B}^\bullet \to \mathbb{D}^\bullet_X$$

which is given by the rule:

$$\sigma \text{ corresponds to } \sigma' \text{ if } \sigma'|U=\sigma.$$

Furthermore, σ is a Verdier dual pairing on U if and only if σ' is a Verdier dual pairing on X.

Proof of Corollary. Since $\tau^Y_{\leqq q-1} R i_* \mathbf{B}^\bullet$ is an object in $\mathscr{C}'_2(q)$, $R\mathbf{Hom}^\bullet(\tau^Y_{\leqq q-1} R i_* \mathbf{B}^\bullet, \mathbb{D}^\bullet_X)$ is an object in $\mathscr{C}'_2(p)$. This is a calculation as in §5.3. Therefore a morphism

$$\mathbf{A}^\bullet \to R\mathbf{Hom}^\bullet(\mathbf{B}^\bullet, \mathbb{D}^\bullet_U)$$

corresponds to a morphism

$$\tau^Y_{\leqq p-1} R i_* \mathbf{A}^\bullet \to R\mathbf{Hom}^\bullet(\tau^Y_{\leqq q-1} R i_* \mathbf{B}^\bullet, \mathbb{D}^\bullet_X).$$

Furthermore, isomorphisms correspond to isomorphisms.

References

1. Artin, M.: Théorème de finitude pour un morphisme propre: dimension cohomologique des schemas algébriques affines, EGA4, expose XIV. Lecture Notes in Mathematics vol. 305. Berlin-Heidelberg-New York: Springer 1973
2. Andreotti, A., Frankel, T.: The Lefschetz theorem on hyperplane sections. Annals of Mathematics **69**, 713–717 (1959)
3. Barthel, G., Georinger, G., Duskin, J., Fittler, R., Acuña, H., Gergondey, R., Verdier, J.L., Zisman, M.: Dualité de Poincaré. Séminaire Heidelberg-Strasbourg 1966/67. Publ. I.R.M.A., 5, rue R. Descartes, 67-Strasbourg
4. Bredon, G.: Sheaf theory. New York: McGraw-Hill, 1967
5. Borel, A.: Seminar on transformation groups. Annals of Mathematics Studies, no. 46, Princeton University Press, Princeton, NJ 1960
6. Borel, A., Moore, J.C.: Homology theory for locally compact spaces. Michigan Math J. **7**, 137–159 (1960)

7. Borho, W., MacPherson, R.: Représentations des groupes de Weyl et homologie d'intersection pour les variétés nilpotentes. C.R. Acad. Sci. Paris, t. 292 Ser. I, 707–710 (1981)
8. Cartan, H., Chevalley, C.: Séminaire de Géometrie Algébrique. École Normale Superieure, Paris 1956
9. Cartan, H., Eilenberg, S.: Homological algebra. Princeton University Press, Princeton, NJ 1956
10. Cheeger, J.: On the Hodge theory of Riemannian pseudomanifolds. In: Geometry of the laplace operator. Proceedings of Symposia in pure Mathematics **36**, 91–146 (1980). Amer. Math. Soc. Providence, RI
11. Cheeger, J., Goresky, M., MacPherson, R.: L^2 Cohomology and intersection homology of singular algebraic varieties. Seminar on differential geometry, Yau, S.T. (ed.) Princeton University Press, Princeton, NJ 1982
12. Deligne, P.: Théorème de Lefschetz et critères de dégénérescence de suites spectrales. Publ. Math. I.H.E.S. **35**, 107–126 (1968)
13. Deligne, P.: Théorie de Hodge II, Publ. Math. I.H.E.S. **40**, 21 (1971)
14. Deligne, P.: Letter to D. Kazhdan and G. Lusztig dated 20 April 1979
15. Fulton, W., MacPherson, R.: Categorical framework for the study of singular spaces. Memoirs of the Amer. Math. Soc. vol. 243 A.M.S., Providence, RI 1981
16. Gelfand, S., MacPherson, R.: Verma modules and Schubert cells: a dictionary, Seminaire d'Algebra. Lecture Notes in Mathematics vol. 924. Berlin-Heidelberg-New York: Springer 1982
17. Godement, R.: Topologie algébrique et théorie des faisceaux. Paris: Hermann, 1958
18. Goresky, M.: Whitney stratified chains and cochains. Trans. Amer. Math. Soc. **267**, 175–196 (1981)
19. Goresky, M., MacPherson, R.: La dualité de Poincaré pour les espaces singuliers. C.R. Acad. Sci. t. **284**, (Serie A) 1549–1551 (1977)
20. Goresky, M., MacPherson, R.: Intersection homology theory. Topology **19**, 135–162 (1980)
21. Goresky, M., MacPherson, R.: Stratified Morse theory. Proceedings of A.M.S. conference in Singularities at Arcata Calif. 1981
22. Hamm, H.: Lokale topologische Eigenschaften komplexer Räume. Math. Ann. **191**, 235–252 (1971)
23. Hartshorne, R.: Residues and duality. Lecture Notes in Mathematics vol. 20. Berlin-Heidelberg-New York: Springer 1966
24. Hurewicz, W., Wallman, H.: Dimension theory. Princeton University Press, Princeton, NJ 1941
25. Iverson, B.: Cohomology of sheaves, preprint, Aarhus, Denmark (1976)
26. Kato, M.: Partial Poincaré duality for k-regular spaces and complex algebraic sets. Topology **16**, 33–50 (1977)
27. Kaup, L.: Nachr. Adad. Wiss. Göttingen Math-Phys. Kl. **II**, 213–224, 1966
28. Kaup, L.: Exakte Sequenzen für globale und lokale Poincaréhomomorphismen. Real and Complex Singularities (P. Holm, ed.) Sitjthoff and Noordhoff Publ., Norway 1978
29. McCrory, C.: Stratified general position. Algebraic and geometric topology. Lecture Notes in Mathematics, vol. 664, pp. 142–146. Berlin-Heidelberg-New York: Springer 1978
30. Mather, J.: Stratifications and mappings. Dynamical systems Peixoto, M.M. (ed.). New York: Academic Press 1973
31. Milnor, J.: Morse theory, Annals of Mathematics Studies no. 51, Princeton University Press (1969), Princeton, New Jersey
32. Narasimhan, R.: On the homology groups of Stein Spaces. Invent. Math. **2**, 377–385 (1967)
33. Ogus, A.: Local cohomological dimension of algebraic varieties. Annals of Mathematics **98**, 327–366 (1973)
34. Oka, M.: On the cohomology structure of projective varieties, in Manifolds – Tokyo, 1973. Hattori, A. (ed.). University of Tokyo Press, 1975
35. Siegel, P.: Witt spaces, a geometric cycle theory for KO homology at odd primes. Ph.D. thesis (M.I.T.), 1979. (To appear in Amer. J. of Math.)
36. Siebenmann, L.: Deformations of homeomorphisms on stratified sets. Comm. Math. Helvetici **47**, 123–163 (1972)
37. Steenrod, N.: The topology of fibre bundles. Princeton Mathematical series no. 14, Princeton University Press, Princeton, NJ 1951
38. Swan, R.: The theory of sheaves. Chicago Lectures in Mathematics, University of Chicago Press, Chicago, 1964
39. Tennison, B.: Sheaf Theory. London Mathematical Society Lecture Note Series vol. 20, Cambridge University Press, 1975

40. Thom, R.: Ensembles et morphismes stratifiés. Bull, Amer. Math. Soc. **75**, 240–284 (1969)
41. Verdier, J.L.: Le théorème de dualité de Poncaré. C.R. Acad. Sci. **256** (1963)
42. Verdier, J.L.: Dualité dans la cohomologie des espaces localment compactes. Séminar Bourbaki vol. **300** (1965)
43. Verdier, J.L.: Categories derivées, État 0. SGA $4\frac{1}{2}$. Lecture Notes in Mathematics vol. 569. Berlin-Heidelberg-New York: Springer 1977
44. Verdier, J.L.: Dimension des espaces localement compacts. C.R. Acad. Sci. Paris t. **261**, 5293–5296 (1965)
45. Verdier, J.L.: Faisceaux constructibles sur un espace localement compact. C.R. Acad. Sci. Paris t. **262** (1966)
46. Vilonen, K.: Master's thesis. Brown University, Providence, R.I. (1980)
47. Wilder, R.L.: Topology of manifolds. Amer. Math. Soc. Publ. no. **32**, Providence, RI (1949)
48. Zeeman, C.: Dihomology I and II. Proc. London Math. Soc. (1) **12**, 609–638, 639–689 (1962)

Oblatum VI-1981 & XII-1982

Invent. math. 72, 131-152 (1983)

Inventiones mathematicae

Sur le noyau de Bergman des domaines de Reinhardt

D. Boichu et G. Coeuré

Universite des Sciences et Techniques de Lillie I, U.E.R. de Mathematiques Pures et Appliques, F-59655 Villeneuve d'Asco Cedex, France

Introduction

Etant donné un domaine strictement pseudo-convexe Ω, borné, à frontière Ω^* de classe C^∞, on sait depuis les travaux de Kerzman [14], Fefferman [11] et de Boutet de Montvel [7] que son noyau de Bergman K_Ω se prolonge en une fonction C^∞ sur $\bar{\Omega}\times\bar{\Omega}$ en dehors de la diagonale de $\Omega^*\times\Omega^*$. Sur cette diagonale:

$$K_\Omega(z,z)=F(z)\frac{1}{\rho(z)^{n+1}}+G(z)\operatorname{Log}\rho(z).$$

F et G sont des fonctions C^∞ sur $\mathbb{C}^n$, la fonction ρ définit Ω par $\rho>0$ et $d\rho \neq 0$ sur Ω^*.

L'article de Hörmander [13] détermine complètement F par la relation: $F(z)=\frac{n!}{(4\pi)^n}(\det L_z)\cdot\|d\rho\|^2$, où L_z est la forme de Levi au point z de Ω^*. Aucune propriété générale de G n'est connue.

Dans le cas particulier où Ω est un domaine de Reinhardt, un développement asymptotique nouveau permettant d'étudier G est établi dans les paragraphes 1 à 3 de cet article. On en déduit au paragraphe 4, une expression explicite du terme dominant de la singularité logarithmique de K_Ω en dimension 2: la frontière peut être définie par $|z_2|=\exp\cdot y(\operatorname{Log}|z_1|)$, alors G est $\sum_{k=1}^{k=6}\frac{L_k(y'')}{(y'')^{k+2}}$, où L_k est un opérateur différentiel à coefficients constants d'ordre $6-k$ et homogène de degré k. On en déduit au paragraphe 5 que les seuls domaines de Reinhardt, strictement pseudoconvexes, bornés, complets et réguliers dans $\mathbb{C}^2$, pour lesquels $G\equiv 0$, sont les ellipsoïdes; ce qui établit dans ce cas particulier, une conjecture de Ramadamov [18].

Notations. Un point z de $\mathbb{C}^n$ sera souvent écrit sous la forme $z=(z',z_n)$ avec z' dans $\mathbb{C}^{n-1}$. Si φ est une fonction d'une variable, $\varphi(z)$ désignera le point de coordonnées $\varphi(z_i)$.

Ω est un domaine de Reinhardt pseudo-convexe, complet à frontière de classe C^1; son image $\tilde{\Omega}$ par l'application: $\{z \to x = \operatorname{Log}|z|\}$ est un ouvert de $\mathbb{R}^n$ qui peut être défini par $x_n < y(x')$, où y est une fonction concave sur un ouvert convexe W de $\mathbb{R}^{n-1}$.

Etant donné un point (z, ξ) de $\mathbb{C}^n \times \mathbb{C}^n$ dont aucune coordonnée est nulle, on posera:

$$z'\bar{\xi}' = \exp[2a + 2i\theta]; \qquad z_n \bar{\xi}_n = \exp[2y(a) + 2\lambda + 2i\Psi]. \tag{1}$$

Au voisinage d'un point (z^0, z^0) de $\Omega^* \times \Omega^*$ dont aucune des coordonnées est nulle, a, θ, Ψ sont des fonctions C^∞ de z et ξ; λ est une fonction de z et ξ ayant la même régularité que la fonction y; Ω est défini par $\lambda(z, z) < 0$.

Etant donnée une fonction $g(t)$ indéfiniment dérivable, on posera:

$$\frac{d^k}{dt^k}\left(\frac{1}{g}\right) = \frac{Q_k(g)}{g^{k+1}}$$

avec $Q_k(g) = \sum_{|\alpha| = k} C_\alpha g^{(\alpha_1)} \dots g^{(\alpha_k)}$, où les C_α sont des constantes convenables.

1. Une propriété pseudo-locale de K_Ω

K_Ω est somme de la série:

$$\frac{1}{(2\pi)^n} \sum_{\substack{m \geqq 1 \\ p \geqq 1}} \frac{2p \cdot (z'\bar{\xi}')^{m-1} (z_n \bar{\xi}_n)^{p-1}}{\int_W \exp[2mu + 2py(u)]\, du}, \quad \text{avec } m \text{ dans } \mathbb{N}^{n-1} \text{ et } mu = \sum_{i=1}^{n-1} m_i u_i. \tag{2}$$

Posons $r(a, u-a) = y(u) - [y(a) + y'(a)(u-a)]$; en utilisant (2), on obtient:

$$K_\Omega = \frac{1}{(2\pi)^n} \frac{1}{z\bar{\xi}} S_y, \quad \text{avec} \quad z\bar{\xi} = (z_1 \bar{\xi}_1) \dots (z_n \bar{\xi}_n)$$

$$S_y(z, \xi) = \sum_{\substack{m \geqq 1 \\ p \geqq 1}} \frac{2p \exp[2mi\theta + 2p(\lambda + i\Psi)]}{\int_W \exp[2\mu v + 2pr(a, v)]\, du} \tag{3}$$

$$\mu = m + py'(a) \quad \text{et} \quad v = u - a.$$

On pose pour tout $\varepsilon > 0$

$$I_\varepsilon = \left\{(m, p) \in \mathbb{N}^{n-1} \times \mathbb{N};\ \left| y'(a_0) + \frac{m}{p} \right| < \varepsilon \right\}.$$

Lorsque y est strictement concave, chaque composante de $y'(a)$ est strictement négative; alors I_ε n'est pas vide. On notera $S_{y,\varepsilon}$ la série (2) sommée seulement sur I_ε.

Lemme 1. *Soit y_i $(i=1,2)$ deux fonctions de classe C^1 et strictement concaves sur des ouverts convexes W_i de $\mathbb{R}^{n-1}$ qui coïncident dans un voisinage de a_0.*

Il existe $\varepsilon>0$ tels que:

$S_{y_1}=S_{y_{2,\varepsilon}}+$(fonction analytique sur un voisinage V_ε de (z^0,z^0) dans $\mathbb{C}^n\times\mathbb{C}^n$).

Démonstration. En utilisant (2), la valeur absolue du terme général de chacune de ces séries est majorée pour α assez petit par:

$$\frac{p\,e^{2p\lambda}}{\int\limits_{|v|<\alpha}\exp[2|\mu|\,v_1-p\varepsilon'\,|v|]\,dv}.$$

Lorsque (m,p) n'est pas dans I_ε et a appartient à un voisinage V_ε de a_0, il existe $\varepsilon'>0$ tel que $|\mu|\geqq 2\varepsilon'\sqrt{|m|^2+p^2}$. On constate alors que le terme général est un:

$$\mathcal{O}[p^{n-1}\cdot(|m|^2+p^2)^{\frac{1}{2}}\cdot\exp[(2\lambda-\varepsilon')(|m|^2+p^2)^{\frac{1}{2}}]].$$

La série converge normalement pour $\lambda<\frac{\varepsilon'}{3}$ et a dans V_ε, ainsi $S_{y_i}=S_{y_{i,\varepsilon}}$ +(fonction analytique dans un voisinage de (z_0,z_0)).

Le terme général de la série $S_{y_{1,\varepsilon}}-S_{y_{2,\varepsilon}}$ est majoré en module par:

$$\sum_{I_\varepsilon}\sum_{i=1,2}2p\,e^{2p\lambda}\frac{\int\limits_{W_i-V}\exp[\mu_i(u-a)+2pr_i(a,u-a)]\,d\mu}{[\int\limits_{V}\exp[\mu(u-a)+2pr(a,u-a)]\,du]^2}.$$

$-a_0+V$ est un voisinage relativement compact de l'origine tel que les fonctions y_i coïncident sur $-a_0+2\bar{V}$. Etant donné un voisinage relativement compact V_1 de a_0 tel que $\bar{V}_1$ soit contenu dans V, puisque les fonctions y_i sont strictement concaves, il existe une constante $k>0$ tel que $\frac{r_i(a,u-a)}{|u-a|}\leqq -k$ pour tout u dans W_i-V et a dans V_1.

Choisissons ε strictement plus petit que k, l'intégrale du numérateur est, pour $(m,p)\in I_\varepsilon$ et a dans V_1, majorée par:

$$\int\limits_{W_i-V}\exp[2p(\varepsilon-k)\,|u-a|]\,du=\mathcal{O}(e^{-\theta p})$$

pour un θ convenable, indépendant de a dans V_1.

Le dénominateur est un $\mathcal{O}\left(\frac{1}{p^{n-1}}\right)^2$ uniformément en a, dans V_1.

Enfin, le cardinal de l'ensemble $\{m;\,(m,p)\in I_\varepsilon\}$ est un $\mathcal{O}(p^{n-1})$. $S_{y_{1,\varepsilon}}-S_{y_{2,\varepsilon}}$ est donc une fonction analytique pour $a\in V_1$ et $\lambda<\theta$.

Corollaire. *Etant donnés deux domaines de Reinhardt, bornés et complets, à frontières de classe C^1, dont les images par l'application* $\log|z|$ *sont strictement convexes et les frontières coïncident dans un voisinage de z_0, leurs noyaux de Bergman diffèrent d'une fonction analytique dans un voisinage de z_0 dans $\mathbb{C}^n$.*

2. Développement asymptotique de K_Ω

On suppose ici que y est une fonction concave de classe C^3 sur $\mathbb{R}^{n-1}$.
On utilisera les notations suivantes:

$$\begin{aligned} h(a,y,v) &= \int_0^1 (1-\tau)\,[y''(a+\tau v)-y''(a)]\,d\tau \\ g(a,m,p,y,t) &= \int_{\mathbb{R}^{n-1}} \exp\{2\mu v + p[y''(a)+2t\,h(a,y,v)](v)\}\,dv. \end{aligned} \tag{3'}$$

Etant donné un voisinage V de l'origine et une fonction χ de classe C^∞, valant 1 sur V et à support dans $2V$, on posera

$$g_V(a,m,p,y,t) = \int_{\mathbb{R}^{n-1}} \chi(v)\exp\{2\mu v + p[y''(a)+2t\,h(a,y,v)](v)\}\,dv.$$

Après développement de $y(u)$ par la formule de Taylor à l'ordre 2 au point a, la série S_y se présente sous la forme:

$$S_y = \sum_{m\geqq 1,\, p\geqq 1} \frac{2p\exp[2mi\theta + 2p(\lambda + i\Psi)]}{g(a,m,p,y,1)}.$$

Lemme 2. *On suppose:*

(i) $y''(a_0)$ *strictement négative; $m(a)$ désignera la plus petite valeur propre de* $-y''(a)$.

(ii) $\|y''(a)-y''(a_0)\| \leqq \alpha \leqq \frac{1}{2} m(a_0)$, *pour tout a.*

(iii) V *est un voisinage compact de l'origine et* $m(x) \geqq \frac{1}{2} m(a_0)$ *sur* $a_0 + V$.

Etant donné un opérateur D^l, d'ordre l, à coefficients constants sur la variable a, pour a dans a_0+V, on a:

$$\left|D^l \cdot \frac{d^k}{dt^k} g\right| \leqq p^{\frac{l-k-n}{2}} \cdot A\left(\frac{|\mu|}{\sqrt{p}}\right) \cdot \exp\left[(-y''(a)-\alpha t)^{-1}\left(\frac{\mu}{\sqrt{p}}\right)\right]; \tag{4}$$

$$g \geqq p^{\frac{1-n}{2}} \cdot \exp\left[(-y''(a)+\alpha t)^{-1}\left(\frac{\mu}{\sqrt{p}}\right)\right] \tag{5}$$

$$\left|\frac{d^k}{dt^k}(g-g_V)\right|_{t=0} \leqq \exp(-\varepsilon' p), \tag{6}$$

pour un $\varepsilon'>0$ lorsque $(m,p)\in I_\varepsilon$ et $6\varepsilon \leqq \inf\{m(a_0)\,|v|,\ v$ décrivant $\mathbb{R}^{n-1}-V\}$;

$$\left|D^l \cdot \frac{d^k}{dt^k}\frac{1}{g}\right| \leqq p^{\frac{n+l-k-1}{2}} \cdot A\left(\frac{|\mu|}{\sqrt{p}}\right) \cdot \exp\left\{\frac{[2(l+k)+1]\,\alpha - m(a)}{m(a)^2-\alpha^2}\,\frac{|\mu|^2}{p}\right\}. \tag{7}$$

A est une fonction à croissance polynomiale.

Démonstration. La condition (ii) et l'ordre de régularité imposé à y fait de $h(a,y,v)(v)$ un $\mathcal{O}(|v|^3)$. Il existe donc une constante K tel que pour a dans a_0

$+V$, on a:

$$\left|D^l\cdot\frac{d^k}{dt^k}g\right|\leqq K p^{l+k}\int|v|^{3k+l}\cdot(1+|v|)^l\exp[2\mu\cdot v+p(y''(a)+t\alpha)(v)]\,dv.$$

La quadrature de cette intégrale gaussienne conduit à (4). L'inégalité (5) s'obient de la même façon.

La seconde inégalité dans (ii) et le choix de V dans (iii) entraîne:

$$\left|\frac{d^k}{dt^k}(g-g_V)\right|_{t=0}\leqq K\,p^k\int_{\complement V}|v|^{3k}\exp\left[2\mu\cdot v-\frac{m(a_0)}{6}|v|^2\cdot p\right]dv.$$

Sur I_ε, la norme de μ est majorée par εp; le choix de ε dans (3) impose

$$\exp\left[2\mu\cdot v-\frac{m(a_0)}{6}\cdot|v|^2 p\right]\leqq\exp[-\varepsilon p\,|v|]$$

sur $\complement V$. Une quadrature donne l'estimation (6) annoncée.

Etablissons (7):

$$D^l\frac{d^k}{dt^k}\frac{1}{g}=\frac{Q_{k+l}(g)}{g^{k+l+1}},$$

où $Q_{k+l}(g)$ est de la forme $\sum_\beta C_\beta g^{(\beta_1)}\dots g^{(\beta_\lambda)}$, avec $\lambda=l+k=|\beta|$, les ordres totaux de dérivation en a et t étant respectivement égaux à l et k. En utilisant (1) et (2) on obtient la majoration suivante:

$$\left|D^l\cdot\frac{d^k}{dt^k}\cdot\frac{1}{g}\right|\leqq p^{\frac{n+l-k-1}{2}}\cdot A\left(\frac{|\mu|}{\sqrt{p}}\right)\exp\left\{\left[\frac{l+k}{[-y''(a)-t\alpha]}-\frac{l+k+1}{[-y''(a)+t\alpha]}\right]\left(\frac{\mu}{\sqrt{p}}\right)\right\}$$

$$\leqq p^{\frac{n+l-k-1}{2}}\cdot A\left(\frac{|\mu|}{\sqrt{p}}\right)\exp\left[\frac{[2(l+k)+1]\,\alpha-m(a)}{m(a)^2-\alpha^2}\,\frac{|\mu|^2}{p}\right].\quad\square$$

Théorème 1. *On suppose que Ω est logarithmiquement strictement convexe, borné, à frontière de classe C^1. On se place en un point z^0 où Ω^* est strictement pseudo-convexe et de classe C^{3+l} ($l\geqq 0$).*

Il existe une suite positive ε_k et une suite V_k de voisinages de l'origine dans $\mathbb{R}^{n-1}$, toutes deux décroissantes, pour lesquels on pose:

$$g_k(a,m,p,y,t)=\int\chi_k(v)\exp\{2\mu\cdot v+p[y''(a)+2th(a,y,v)](v)\}\,dv.$$

χ_k est une fonction C^∞ à support dans V_k et valant 1 dans un voisinage de l'origine.

$$Q_k(g_k)=Q_k(a,m,p,y)\quad\text{est défini par}\quad\left(\frac{dt}{dt^k}\frac{1}{g_k}\right)_{t=0}=\frac{Q_k(g_k)}{g_k^{k+1}}$$

$$L_k=\frac{1}{k!}\sum_{I_{\varepsilon_k}}2p\exp[(2mi\theta+2p(\lambda+i\Psi)]\cdot\frac{Q_k(a,m,p,y)}{[g(a,m,p,y,0)]^{k+1}};$$

Alors:

$$K_\Omega(z,\xi)=\frac{(2\pi)^{-n}}{z\cdot\bar{\xi}}\sum_{k\leqq q} L_k+R_q. \tag{7'}$$

R_q est une fonction de classe C^l, dans un voisinage de (z^0,z^0) dans $\bar{\Omega}\times\bar{\Omega}$, pour $q\geqq 3(n+l+1)$.

Démonstration. Pour tout entier k il existe une fonction y_k ayant sur $\mathbb{R}^{n-1}$ tout entier la même régularité que y et un voisinage V_k de l'origine dans $\mathbb{R}^{n-1}$ tels que

(i) y_k coïncide avec y sur a_0+V_k;

(ii) $\|y_k''(a)-y''(a_0)\|\leqq\frac{1}{3}\frac{m(a_0)}{3k+1}$ sur $\mathbb{R}^{n-1}$;

(iii) $m(a)\geqq\frac{1}{2}m(a_0)$ sur a_0+V_k.

On sait de plus, d'après le lemme 1, qu'il existe une suite $\varepsilon_k>0$, que l'on peut supposer décroissante et majorée par $\frac{1}{6}\inf\{m(a_0)\,|v|,\ v$ décrivant $\mathbb{R}^{n-1}-V_k\}$ telle que S_y diffère de S_{y_k,ε_k} par une fonction analytique dans un voisinage de (z^0,z^0) dans $\mathbb{C}^n\times\mathbb{C}^n$.

Introduisons $S_{y_k,t}=\sum\dfrac{2p\exp[2mi\theta+2p(\lambda+i\Psi)]}{g(a,m,p,y_k,t)}$.

La formule de Taylor conduit à:

$$S_{y_q}=S_{y_q,t=1}=\sum_{k<q}\frac{1}{k!}\left(\frac{d^k}{dt^k}S_{y_q,t}\right)_{t=0}+\frac{1}{(q-1)!}\int_0^1(1-\tau)^{q-1}\frac{d^q}{d\tau^q}S_{y_q,\tau}\,d\tau$$

On vérifie que pour $q\geqq 3(n+l-1)$, le reste de Taylor sommé sur I_{ε_q} est une fonction de classe C^l dans un voisinage de (z_0,z_0) dans $\bar{\Omega}\times\bar{\Omega}$. En effet:

Soit D^l un opérateur différentiel, d'ordre $\leqq l$, à coefficients constants, sur les variables a, θ, λ, Ψ. Sur I_{ε_q}, chaque composante de m est un $\mathcal{O}(p)$ et pour p fixé son cardinal est un $\mathcal{O}(p^{n-1})$. Aussi, sur I_{ε_q}, le terme général de $D^l\cdot\dfrac{d^q}{dt^q}S_{y_q,t}$ est, après sommation en m, majoré selon l'estimation (7) du lemme 2, par:

$$p^{\frac{3n+3-q-1}{2}}\cdot A\left(\frac{\mu}{\sqrt{p}}\right)\exp\left\{\frac{[2(l+p)+1]\,\alpha-m(a)}{m(a)^2-\alpha^2}\,\frac{|\mu|^2}{p}\right\}.$$

Choisissons $q=3(n+l+1)$; les conditions (i) et (ii) sur y_q font que le coefficient de $\dfrac{|\mu|^2}{p}$ dans l'exponentielle est minoré par un nombre strictement négatif sur a_0+V_q. Aussi, après sommation en m sur I_{ε_q}, le terms général de $D^l\,\dfrac{d^q}{dt^q}S_{y_q,t}$ est un $\mathcal{O}\left(\dfrac{1}{p^2}\right)$ uniformément pour $\lambda\leqq 0$ et a dans a_0+V_q. La propriété du reste est donc établie.

Il reste à étudier la série $\dfrac{1}{k!}\left(\dfrac{d^k}{dt^k}S_{y_k,t,\varepsilon_q}\right)_{t=0}-L_k$.

Elle admet comme série majorante, la somme:

$$\sum_{\varepsilon_q} 2p \cdot \frac{|Q_k[g(a,m,p,y_q,0)] - Q_k[g_k(a,m,p,y,0)]|}{[g(a,m,p,y,0)]^{k+1}}$$
$$+ \sum_{I_{\varepsilon_k} - I_{\varepsilon_q}} 2p \cdot \frac{|Q_k[g_k(a,m,p,y,0)]|}{[g(a,m,p,y,0)]^{k+1}}.$$

Puisque y_q coïncide avec y sur $a_0 + V_q$, les estimations (4) et (6) du lemme 2 permettent de majorer le numérateur du terme général de la première série par

$$\exp\left[\frac{k-1}{m(a)} \frac{|\mu|^2}{p}\right] \cdot \exp(-\varepsilon'_q \cdot p).$$

En utilisant l'inégalité (5) de ce même lemme, on constate que le terme général de la première série est à décroissance exponentielle en p. Sa somme est donc C^∞ dans un voisinage de (z^0, z^0) dans $\mathbb{C}^n \times \mathbb{C}^n$.

Utilisons les inégalités (4) et (5), le terme général de la seconde série est majoré par:

$$\mathcal{O}(p^{\frac{n+1-k}{2}})\, \mathcal{O}\left(\frac{|\mu|}{\sqrt{p}}\right)^{3k} \exp\left[\frac{-1}{m(a)} \frac{|\mu|^2}{p}\right].$$

Or sur $I_{\varepsilon_k} - I_{\varepsilon_q}$, la norme de μ est minorée par $p \cdot \varepsilon_q$ et majorée par $p \cdot \varepsilon_k$. Sa somme est donc aussi une fonction C^∞ dans un voisinage de (z^0, z^0) dans $\mathbb{C}^n \times \mathbb{C}^n$. □

3. Forme intégrale de L_k

$$[g(a,m,p,y,0)]^{-1}$$
$$= [\mathcal{L}(e^{py''(a)})(2\mu)]^{-1} = \left|\frac{p}{\pi}\right|^{n-1} |\det y''(a)|\, \mathcal{F}(e^{py''(a)})(2\mu)$$
$$= \left(\frac{p}{\pi}\right)^{n-1} |\det y''(a)| \int_{\mathbb{R}^{n-1}} \exp[2i(\tau+iv)\,\mu + py''(a)(\tau+iv)]\, d\tau.$$

Les notations $\mathcal{L}$ et $\mathcal{F}$ désignent les transformations de Laplace et Fourier. Rappelons que:

$$g_k^{(q)}(a,m,p,y,0) = (2p)^q \cdot \int \chi_{V_k}(v)\, h^q(a,y,v)(v) \exp[2\mu v + py''(a)(v)]\, dv.$$

Etant donnée l'expression $\frac{d^k}{dt^k}\left(\frac{1}{g}\right) = [\sum C_\alpha g^{(\alpha_1)} \dots g^{(\alpha_k)}] \cdot \frac{1}{g^{k+1}}$, on pose à partir de maintenant:

$$Q_k(h(v)) = \sum C_\alpha \cdot h^{\alpha_1}(a,y,v_1)(v_1) \dots h^{\alpha_k}(a,y,v_k)(v_k). \tag{8}$$

On pose aussi:

$$A=\exp\left[2i\left(\theta+\sum_{j=i}^{j=k}\tau_j+t\right)\right] \quad \text{à valeurs dans } \mathbb{C}^{n-1}; \tag{9}$$

$$\begin{aligned} B=\exp\Bigg[\sum_{j=1}^{j=k} & 2i\, y'(a)(\tau_j)+y''(a)(\tau_j,\tau_j+2iv_j) \\ & +y''(a)(t)+2i y'(a)(t)+2\lambda+2i\Psi\Bigg]. \end{aligned} \tag{10}$$

Après avoir rassemblé les produits d'intégrales intervenant dans L_k sous une même intégrale multiple, on trouve la forme suivante pour L_k

$$L_k=\frac{|\det y''(a)|^{k+1}}{k!\,(2\pi)^{(n-1)(k+1)}}$$
$$\cdot\left(\frac{\partial}{\partial\lambda}\right)^{n(k+1)}\sum_{I_{\varepsilon_k}}\int_{(\mathbb{R}^{n-1})^k\times(\mathbb{R}^{n-1})^k\times\mathbb{R}^{n-1}}\chi_k(v)\cdot Q_k(h(v))\,A^m B^p\,dv\,d\tau\,dt.$$

χ_k est une fonction à support compact valant 1 dans un voisinage de l'origine de $(\mathbb{R}^{n-1})^k$.

On peut supposer que I_{ε_k} est défini par $\left\{(m,p);\ \frac{\alpha_k}{\beta_k}<\frac{m}{p}\leqq\frac{\alpha'_k}{\beta_k}\right\}$, où (α_k,β_k) et (α'_k,β_k) sont convenablement choisis dans $\mathbb{N}^{n-1}\times\mathbb{N}$ et vérifiant en particulier:

$$\frac{\alpha_k}{\beta_k}<-y'(a_0)<\frac{\alpha'_k}{\beta_k}.$$

On omettra dans la suite l'indice k pour alléger les notations, et l'on désignera par α_i (resp. α'_i) les composantes de α (resp. α').

Lemme 3. *On a la représentation suivante:*

$$\sum_{I_\varepsilon}A^m\cdot B^p=A^m\cdot B^p=\frac{B\,P(A,B)}{\prod_I(1-A_I^\alpha A_J^{\alpha'}B^\beta)},\quad \text{où:} \tag{11}$$

- P est un polynôme en A et B.
- I décrit les parties de $\{1,2,\dots,n-1\}$, J est le complémentaire de I
- On note: $x_I^\alpha=\prod_{i\in I}x_i^{\alpha_i}$.

Démonstration. Posons $p=N\beta+R$ dans la division euclidienne de p par β; alors m varie dans l'intervalle $\left[N\alpha+E\left(\frac{\alpha R}{\beta}\right)+1,\ N\alpha'+E\left(\frac{\alpha' R}{\beta}\right)\right]$ où E désigne la partie entière.

Soit $A=(A_i)$, les composantes de A; posons:

$$x_i=A_i^{N\alpha_i+E\left(\frac{\alpha_i R}{\beta}\right)+2},$$

$$y_i=A_i^{N\alpha'_i+E\left(\frac{\alpha'_i R}{\beta}\right)+1}.$$

Lorsque chaque $A_i \neq 1$, on a:

$$\sum_m A^m = \prod_{i=1}^{i=n-1} \frac{x_i - y_i}{1 - A_i} = \sum_I \frac{(-1)^{|J|} x_I \cdot y_J}{\prod (1 - A_i)}.$$

Après sommation en N et R, on obtient pour un polynôme $\tilde{P}$ convenable, divisible par chaque $1 - A_i$

$$\frac{\tilde{P}(A, B)\, B}{\prod_I (1 - A^\alpha A_J^{\alpha'} B^\beta) \cdot \prod_i (1 - A_i)}. \quad \square$$

Sous les hypothèses du théorème 1, on obtient ainsi une représentation intégrale des parties singulières L_k du noyau de Bergman que l'on fixe dans le théorème suivant:

Théorème 2. *Pour tout entier k, il existe* ε_k *dans* $(\mathbb{R}_+)^{n-1}$ *et une fonction* χ_k *à support compact valant* 1 *dans un voisinage de l'origine de* $(\mathbb{R}^{n-1})^k$ *tels que:*

– *pour tout triplet d'entiers* $(\alpha_k, \alpha_k', \beta_k)$ *dans* $\mathbb{N}^{n-1} \times \mathbb{N}^{n-1} \times \mathbb{N}$ *vérifiant*

$$-y'(a_0) - \varepsilon_k < \frac{\alpha_k}{k} < -y'(a_0) < \frac{\alpha_k'}{k} < -y'(a_0) + \varepsilon_k.$$

On a:

$$L_k = \frac{|det\; y''(a)|^{k+1}}{k!\,(\pi)^{(n-1)(k+1)}} \cdot \left(\frac{\partial}{\partial \lambda}\right)^{n(k+1)} \int_{(\mathbb{R}^{n-1})^k \times (\mathbb{R}^{n-1})^k \times \mathbb{R}^{n-1}} \chi_k(v) \frac{B \cdot P_k(A, B)\, Q_k(h(v))}{\prod_I (1 - A_I^\alpha A_J^{\alpha'} B^\beta)}\, dv\, d\tau\, dt$$

où les fonctions A, B, Q_k, P_k *ont été définis en* (9), (10), (8), *et* (11), *on notera que* P_k *est un polynôme en A et B.*

Remarque. Les seules singularités des noyaux contenues dans la représentation intégrale des L_k ont lieu lorsque

$$\lambda = v = t = 0, \quad \text{et}$$

$$2\Big[\sum_{i \in I} \alpha_{i,k} \theta_i + \sum_{i \in \mathscr{J}} \alpha_{j,k}' \theta_i + \beta_k \Psi\Big] = 0 \pmod{2\pi}. \tag{11'}$$

Lorsque $|z_0| \neq |\xi_0|$, compte tenu de la stricte concavité de y, $\lambda(z_0, \xi_0)$ est strictement négatif. Ainsi K_Ω est analytique au voisinage d'un tel point.

Lorsque $|z_0| = |\xi_0|$ et $z_0 \neq \xi_0$, en choisissant convenablement la dernière coordonnée, 2Ψ est alors différent de 0 (mod 2π). On peut alors choisir α, α', β tels que $\frac{\alpha}{\beta}$ et $\frac{\alpha'}{\beta}$ soient aussi proche que l'on veut de $-y'(a_0)$ et tel que (11') n'ait pas lieu.

On retrouve donc le résultat connu selon lequel K_Ω est C^∞ sur $\bar{\Omega} \times \bar{\Omega}$ en dehors de la diagonale.

4. Calcul du terme dominant de la singularité logarithmique en dimension 2

Ω est donc un domaine de Reinhardt, complet, strictement pseudo-convexe, à frontière C^∞, borné de $\mathbb{C}^2$.

Avec la représentation de Ω définie en (1), près d'un point $z^0 \in \Omega^*$ à coordonnées non nulles, la formule de Feffermann s'écrit localement:

$$K(z,z)=K(a,\lambda)=\frac{\phi_1(a,\lambda)}{(\lambda)^3}+\phi_2(a,\lambda)\operatorname{Log}(-\lambda)$$

avec $\phi_1, \phi_2 \in \mathscr{C}^\infty([a_0-\alpha, a_0+\alpha]\times[\lambda_0, 0])$ où $a_0=\operatorname{Log}|z_1^0|$; $\alpha>0$; $\lambda_0<0$.

On se propose de calculer $\Phi_2(a_0,0)$.

Dans ce cas particulier, L_k s'écrit sous la forme:

$$L_k(z)=\left(\frac{|y''(a)|}{2\pi}\right)^{k+1}\frac{\partial^{2(k+1)}G_k(z)}{\partial\lambda^{2(k+1)}}, \quad \text{avec} \tag{12}$$

$$G_K=\int_{\mathbb{R}^k\times\mathbb{R}^k\times\mathbb{R}}\frac{\chi_k(v)\,Q_k(h(v))\,P_k(A,B)\,B}{(1-A^{\alpha_k}B^{\beta_k})(1-A^{\alpha'_k}B^{\beta_k})}\,dv\,d\tau\,dt. \tag{13}$$

Le polynôme P_k, défini par le lemme 3, est ici en deux variables; il ressort facilement de sa construction que:

$$P_k(1,1)=(\alpha'_k-\alpha_k)\cdot\beta_k.$$

On considère un changement de base orthonormale dans $\mathbb{R}^{k+1}$ (variables $\tau_1,\dots,\tau_k,t$) tel que le nouveau premier vecteur de base soit $1/\sqrt{k+1}(1,\dots,1)$. Alors la nouvelle première coordonnée est $1/\sqrt{k+1}\left(\sum_{j=1}^{k}\tau_j+t\right)$. On la note τ_1, les autres étant $\tau_2,\dots,\tau_{k+1}$ sans qu'il y ait risque de confusion.

w étant l'image de $(v_1,\dots,v_k,0)$ par cette transformation orthonormale, on constate aisément que la transformation: $(v_1,\dots,v_k)\to(w_2,\dots,w_{k+1})$ est linéaire et inversible de $\mathbb{R}^k$ dans $\mathbb{R}^k$. Dans ces conditions, A et B prennent dans (12) la forme:

$$A=\exp(2i\sqrt{k+1}\,\tau_1)$$
$$B=\exp(2i\sqrt{k+1}\,y'(a)\,\tau_1+y''(a)\,\tau^2+2iy''(a)\,\tau\cdot w+2\lambda). \tag{13'}$$

Première intégration dans G_k. Dorénavant, pour alléger, on omet l'indice k dans α_k, α'_k, β_k. On introduit aussi les coefficients non universels:

$$b=-2\sqrt{k+1}(\alpha+\beta y'(a))$$
$$c=-\beta y''(a)$$
$$d=-2\beta y''(a).$$

En tenant compte de la stricte pseudo-convexité b, c, d sont tous trois strictement positifs.

Avec ces notations, il vient:

$$A^\alpha B^\beta = \exp[-ib\tau_1 - c\tau^2 - id\tau w + 2\beta\lambda].$$

Le résultat intermédiaire qui suit réalise une première réduction des singularités.

Lemme 4. *Il existe des réels* $\tau_2^0, \dots, \tau_{k+1}^0 > 0$; $\lambda_0 < 0$; *il existe une fonction* Ψ_1 *analytique à valeurs complexes définie sur un voisinage* V *de* 0 *dans* $\mathbb{R}^k(\tau_2, \dots, \tau_{k+1}) \times \mathbb{R}^k(v) \times \mathbb{R}(\lambda)$ *tels que:*

(a) $$ib\Psi_1 + c\Psi_1^2 + id\Psi_1 w_1 = -c\sum_{j=2}^{k+1} \tau_j^2 - id\sum_{j=2}^{k+1} \tau_j w_j + 2\beta\lambda. \tag{14}$$

(b) *à une fonction analytique près sur* $[0, \lambda_0]$:

$$G_k(\lambda) = \frac{\pi\beta}{\sqrt{k+1}} \int_{\substack{|\tau| < \tau_j^0 \\ j=2,\dots,k+1}} \frac{\chi(v)\, Q_k(h(v))\, d\tau_2, \dots, d\tau_{k+1}\, dv}{[i(b + dw_1) + 2c\Psi_1]\, \Psi_1} \tag{15}$$

(w_j *est la j-ième coordonnée de w défini dans* (13′)).

Démonstration. $1 - A^\alpha B^\beta$ est une fonction analytique réelle en $\tau_1, \tau_2, \dots, \tau_{k+1}, v, \lambda$ réels qu'on peut prolonger en une fonction analytique complexe pour τ_1 complexe, les autres variables restant réelles. Cette fonction est nulle en $\tau = v = \lambda = 0$ et sa dérivée holomorphe par rapport à τ_1 est:

$$\frac{\partial(1 - A^\alpha B^\beta)}{\partial\tau_1}(0) = (A^\alpha B^\beta)(-ib - 2c\tau_1 - idw_1)|_{(0)} = -ib \neq 0.$$

Le théorème des fonctions implicites assure l'existence d'un voisinage V de $\tau_2 = \dots = \tau_{k+1} = 0$, $v = 0$, $\lambda = 0$, d'une fonction analytique à valeurs complexes $\tau_1 = \Psi_1(\tau_2, \dots, v, \lambda)$ sur V telle que $1 - A^\alpha B^\beta$ s'annule identiquement sur V.

Cette identité est (14) a priori modulo $2i\pi$, puis sans modulo par continuité. D'où (a).

On a réalisé une intégration dans (13) par rapport à τ_1 en choisissant un contour donnant lieu a une fonction de λ régulière par la méthode des résidus.

La partie réelle de (14) est:

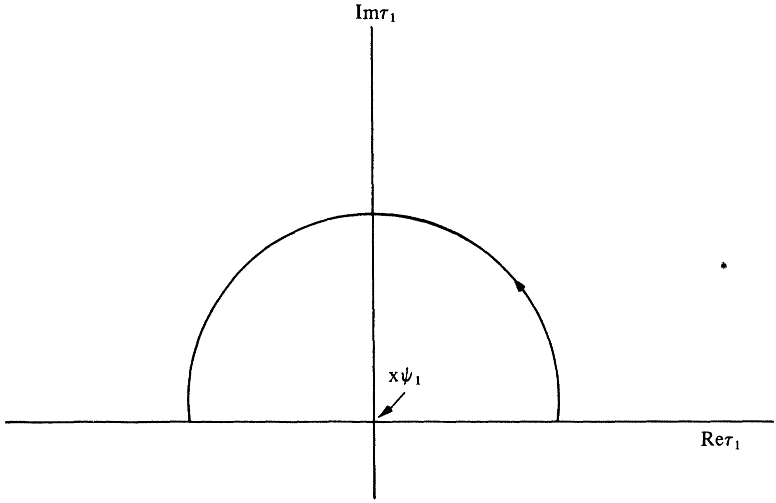
$$\operatorname{Im}\Psi_1(b + dw_1 + c\operatorname{Im}\Psi_1) = c\left((\operatorname{Re}\Psi_1)^2 + \sum_{j=2}^{k+1} \tau_j^2\right) - 2\beta\lambda. \tag{16}$$

Comme w_1 est linéaire en v et $b > 0$; par la continuité de Ψ_1, pour v, $\tau_2, \dots, \tau_{k+1}$, λ assez petits, on a:

$$b + dw_1 + c\operatorname{Im}\Psi_1 > 0.$$

La second membre de (16) étant positif il vient $\operatorname{Im}\Psi_1 > 0$.

Dans le plan complexe des τ_1, $\Psi_1(\tau_2, \dots, \tau_{k+1}, v, \lambda)$ est localisé dans le demi-plan supérieur. C'est le seul zéro de $1 - A^\alpha B^\beta$ dès que $\tau_1, \tau_2, \dots, \tau_{k+1}, v, \lambda$ sont assez petits (d'après le théorème des fonctions implicites). On limite le domaine



d'intégration dans G_k pour assurer l'existence d'un seul zéro à $1-A^\alpha B^\beta$ pour chaque $\tau_2,\dots,\tau_{k+1}$, v, λ à savoir $\Psi_1(\tau_2,\dots,\tau_{k+1},v,\lambda)$. L'intégrale restante pour l'obtention de G_k comme on l'a déjà remarquée donne naissance à une fonction analytique en λ sur $[0,\lambda_0]$ pour un certain $\lambda_0<0$.

On peut maintenant choisir un contour dans le plan des τ_1 formé d'un demi-cercle supérieur tel que, pour chaque $\tau_2,\dots,\tau_{k+1}$, v, λ assez petits, Ψ_1 est dans son intérieur. Il résulte du théorème des résidus que l'intégration sur l'axe $\mathrm{Re}\,\tau_1$ dans le calcul de G_k est remplacé par une intégration sur un demi-cercle non situé à proximité du pôle $\Psi_1(\tau_2,\dots,\lambda)$ de l'intégrant qui donne naissance à une nouvelle fonction analytique en λ sur $[0,\lambda_0]$, avec en plus un résidu qui est:

$$\frac{2i\pi\,\chi(v)\,Q_k(h(v))\,P(A,B)\,B}{(1-A^{\alpha'-\alpha})\cdot\dfrac{\partial(1-A^\alpha B^\beta)}{\partial\tau_1}}\quad\text{au point}\quad \Psi_1(\tau_2,\dots,\lambda).$$

Mais $\dfrac{\partial(1-A^\alpha B^\beta)}{\partial\tau_1}(\Psi_1)=-i(b+dw_1)-2c\Psi_1$.

Une intégration du résidu en $\tau_2,\dots,\tau_{k+1}$, v, λ donne:

$$2i\pi\int_{\substack{|\tau_j|<\tau_j^0\\ j=2,\dots,k+1}}\frac{\chi(v)\,Q_k(h(v))\,P_k(A,B)\,B\,d\tau_2,\dots,d\tau_{k+1}\,dv}{(-i(b+dw_1)-2c\Psi_1)(1-A^{\alpha'-\alpha})}.$$

$\chi(v)$ aura été le cas échéant modifié pour avoir un support plus petit. A, B dans l'intégrant dépendent maintenant de $\Psi_1,\tau_2,\dots,\tau_{k+1},v,\lambda$.

L'intégrant de (17) se simplifie. En effet:

$A=1+(\Psi_1\times$ fonction analytique en $\Psi_1)$

$(1-A^{\alpha'-\alpha})(\Psi_1)=-2\sqrt{k+1}\quad i(\alpha'-\alpha)\,\Psi_1(1+\Psi_1\times$ fonction analytique en $\Psi_1)$.

Comme Ψ_1 vérifie $A^\alpha B^\beta=1$; en posant $B=\exp\Theta$ il résulte $\Theta=$constante $\times\Psi_1$ et $B-1=\Psi_1\times$(fonction analytique en Ψ_1). Ainsi

$$P_k(A,B)=P_k(1,1)+(A-1)\,Q_1(A,B)+(B-1)\,Q_2(A,B)$$

où Q_1, Q_2 sont des polynômes en A, B. Alors $P_k(A,B)=P_k(1,1)+\Psi_1\times$fonction analytique. Le lemme suit par remplacement dans (17) de $1-A^{\alpha'-\alpha}$ et $P_k(A,B)$ par les expressions ci-dessus sachant que $P_k(1,1)=(\alpha'-\alpha)\,\beta$.

5. Etude de L_0

A ce stade, le comportement asymptotique de L_0 se dégage.

Proposition. *L_0 a un développement asymptotique de la forme:*

$$\frac{|y''(a)|}{4}\cdot\frac{1}{\lambda^3}+\textit{fonction analytique sur } [0,\lambda_0] \qquad \lambda_0<0.$$

Démonstration. On calcule dans ce qui suit à une fonction analytique en λ près sur $[0,\lambda_0]$. Pour $k=0$, (15) donne:

$$G_0=\frac{\pi\beta}{(ib+2c\,\Psi_1(\lambda))\,\Psi_1(\lambda)}$$

Ψ_1 est définie par $ib\Psi_1+c\Psi_1^2=2\beta\lambda$ c'est-à-dire

$$\Psi_1(\lambda)=\frac{i(-b+\sqrt{b^2-8\beta\lambda c})}{2c} \quad \text{car} \quad \operatorname{Re}\Psi_1>0.$$

Il vient:

$$G_0=-\frac{2\pi\beta c}{\sqrt{b^2-8\beta\lambda c}(-b+\sqrt{b^2-8\beta\lambda c})}=\frac{\pi}{4\lambda}\left(1+\frac{b}{\sqrt{b^2-8\beta\lambda c}}\right)$$
$$=\frac{\pi}{4\lambda}(1+1+\lambda g(\lambda)); \quad g \text{ analytique en } 0.$$

D'après (12), on obtient la proposition en dérivant deux fois G_0.

La partie singulière L_k, $1\leq k$. Maintenant, on réalise une nouvelle intégration dans G_k donnée par (15) par un changement de contour après avoir localisé les zéros de Ψ_1 qui forment un pincement. On accède à L_k par des dérivations en λ suivant (12).

La partie singulière sera logée dans une intégrale absolument convergente se prêtant à une étude du comportement asymptotique.

Lemme 5. *Pour $k\geqq 1$, la partie singulière de L_k est dans:*

$$C_k\beta^{2k+6}\,y''(a)^{k+4}\int_U\frac{Q_k\left(h\left(\frac{v}{d}\right)\right)dv\,d\tau_3,\ldots,d\tau_{k+1}}{\left[-8b\beta\lambda+w_2^2+\sum_{j=3}^{k+1}(\tau_j^2+w_j^2)\right]^{\frac{1}{2}+2(k+1)}}. \tag{18}$$

C_k est une constante universelle indépendante de α, β, a. U est un voisinage borné de 0.

Démonstration. (a) La fonction $\Psi_1(\tau_2, \ldots, \tau_{k+1}, v, \lambda)$ est analytique en τ_2 réel, elle admet un prolongement holomorphe pour τ_2 complexe. La relation (14) reste vraie. Il résulte que $\Psi_1(\tau_2, \ldots) = 0$, ($\tau_2$ étant complexe) si et seulement si

$$c\tau_2^2 + id\tau_2 w_2 - \left\{2\beta\lambda - c\sum_{j=3}^{k+1} \tau_j^2 - id\sum_{j=3}^{k+1} \tau_j w_j\right\} = 0. \tag{19}$$

Ψ_1 est une fonction holomorphe de τ_2 s'annulant aux zéros de (19) qui sont:

$$\frac{i}{2c}\left(-dw_2 \pm \sqrt{d^2 w_2^2 - 4c\{\ldots\}}\right). \tag{20}$$

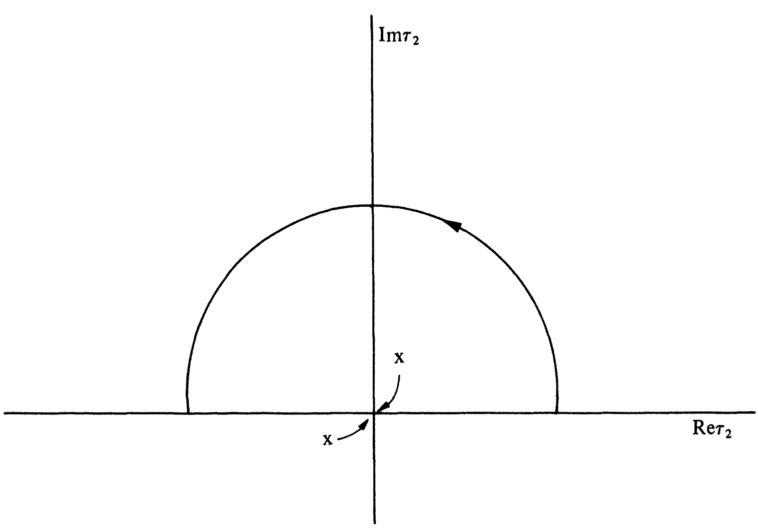
L' accolade dans (20) est la même que dans (19) et a une partie réelle négative de sorte que le radical définit la racine carrée holomorphe associée à la détermination principale du logarithme. En définitive (20) définit deux fonctions analytiques en $\tau_3, \ldots, \tau_{k+1}, \lambda$ qui tendent vers 0 avec $v, \tau_3, \ldots, \tau_{k+1}, \lambda$.

(b) la localisation des zéros précédents s'appuie sur la propriété suivante:

Soit le trinôme $z^2 + 2iaz + \gamma = 0$, a réel, $\operatorname{Re}\gamma > 0$. Alors $\operatorname{Re}\sqrt{a^2 + \gamma} > |a|$ et les deux racines du trinôme sont de part et d'autre de l'axe réel en dépendant analytiquement de a et γ.

Cette propriété se démontre en supposant $a = 1$ en revenant à la définition de la racine carrée holomorphe associée à la détermination principale du logarithme.

(c) Ψ_1 porte la partie singulière dans (15). Pour $\tau_3, \ldots, \tau_{k+1}$, v, τ fixés d'après (14) et (b), $\Psi_1 = 0$ admet deux racines en pincement de l'axe $\operatorname{Re}\tau_2$. Ces deux racines définissent deux fonctions analytiques à valeurs dans les deux demi-espaces séparés strictement par $\operatorname{Re}\tau_2$. On utilise une nouvelle fois un demi-cercle supérieur et le théorème des résidus. On appelle $\Psi_2(\tau_3, \ldots, v, \lambda)$ la racine



de $\Psi_1=0$ dans le demi-plan supérieur. On a modulo une fonction analytique en $\lambda, v, \dots$:

$$\int_{|\tau_2|<\tau_2^0} \frac{d\tau_2}{|i(b+dw_1)+2c\Psi_1|\,\Psi_1} = \frac{2i\pi}{[i(b+dw_1)+2c\Psi_1(\Psi_2)]\,\dfrac{\partial\Psi_1}{\partial\tau_2}(\Psi_2)}.$$

Sachant que $\Psi_1(\Psi_2)=0$, en dérivant dans (14) par rapport à τ_2 il vient:

$$i(b+dw_1)\frac{d\Psi_1}{d\tau_2} = -2c\Psi_2 - idw_2.$$

Le résidu est alors $2i\pi/-2c\Psi_2-idw_2$. En utilisant (20) et en intégrant par rapport à $\tau_3,\dots,\tau_{k+1},v$ on obtient pour $k\geqq 1$ a une fonction analytique en λ près:

$$G_k(\lambda) = \frac{-2\pi^2\beta}{\sqrt{k+1}} \int_{\mathcal{O}} \frac{Q_k(h(v))\,d\tau_3,\dots,dv}{\sqrt{d^2w_2^2-4c\left\{2\beta\lambda - c\sum_{j=3}^{k+1}\tau_j^2 - id\sum_{j=3}^{k+1}\tau_j w_j\right\}}} \tag{21}$$

$\mathcal{O}$ étant un voisinage borné de l'origine.

(d) L_k s'obtient à partir de G_k en dérivant $2(k+1)$ fois par rapport à λ. En dérivant (21) sous le signe somme et en effectuant les changements de variable $v\mapsto dv$, $\tau\mapsto 2c\tau$ il vient:

$$\int \frac{Q_k\left(h\left(\dfrac{v}{d}\right)\right)dv\,d\tau_3,\dots,d\tau_{k+1}}{\left[-8c\beta\lambda+w_2^2+\sum_{j=1}^{k+1}\tau_j^2+2i\sum_{j=3}^{k+1}\tau_j w_j\right]^{\frac{1}{2}+k+2}} \tag{22}$$

avec la constante multiplicative:

$$-\frac{2\pi^2}{\sqrt{k+1}}\,\frac{1}{2}\left(\frac{1}{2}+1\right)\dots\left(\frac{1}{2}+2k+1\right)\frac{\beta(-8c\beta)^{2(k+1)}}{d^k(2c)^{k-1}}.$$

On omettra dorénavant les constantes universelles, ainsi c et d pourront être remplacés par $\beta y''(a)$. Les β sont conservés, leur disparition dans les coefficients asymptotiques constitue un indice de correction des calculs.

La partie singulière de L_k s'obtient à partir de l'expression précédente en multipliant par $y''(a)^{k+1}$ (voir (12)). Il résulte la constante $\beta^{2k+6}\,y''(a)^{k+4}$ à un facteur multiplicatif universel près.

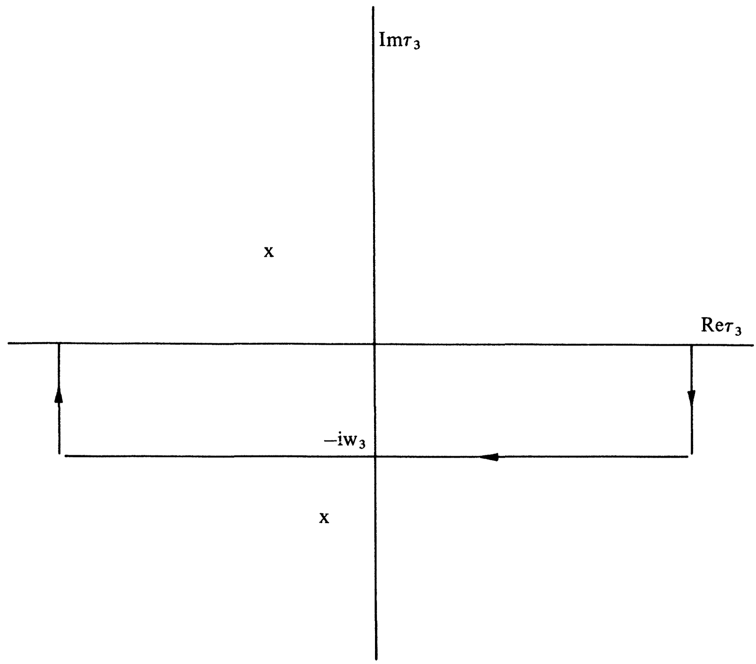
Pour étudier (22), on pose

$$g(\tau_3) = \lambda_1 + w_2^2 + \sum_{j=3}^{k+1}\tau_j^2 + 2i\sum_{j=3}^{k+1}\tau_j w_j$$

$$\lambda_1 = -8c\beta\lambda$$

alors:

$$g(\tau_3) = \lambda_1 + w_2^2 + \sum_{j=3}^{k+1}\{(\tau_j+iw_j)^2+w_j^2\}.$$



Comme trinôme holomorphe de τ_3 complexe, g a des zéros extérieurs à la bande définie par l'axe réel et $\operatorname{Im}\tau_3 = -w_3$. (On applique (b) aux zéros de g, en distinguant deux cas suivant le signe de w_3). Pour τ_3 dans cette bande $|\operatorname{Im}\tau_3 + w_3| \leqq |w_3|$. Comme

$$\operatorname{Re} g(\tau_3) = \lambda_1 + w_2^2 + \sum_{j=4}^{k+1} \tau_j^2 + w_3^2 + (\operatorname{Re}\tau_3)^2 - (\operatorname{Im}\tau_3 + w_3)^2,$$

il résulte que $\operatorname{Re} g(\tau_3) > 0$. Ainsi, la racine carrée utilisée (associée à la détermination principale du logarithme) (dans (22)) permet d'avoir $\tau_3 \mapsto \sqrt{g(\tau_3)}$ holomorphe sur la bande ci-dessus. Il est permis a une fonction régulière en λ près d'intégrer sur le contour $t_3 \mapsto t_3 - iw_3$, t_3 réel limité à $|t_3| < \tau_3^0$. En utilisant une méthode analogue pour les variables $\tau_4, \ldots, \tau_{k+1}$ (22) donne l'intégrale cherchée.

Evaluation asymptotique de L_k, $1 \leqq k$. On aura à utiliser le résultat élémentaire suivant; on pose:

$$I_{q,p}(\lambda) = \int_0^1 \frac{u^q\,du}{(\lambda + u^2)^{\frac{1}{2}+p}} \qquad p,\ q \text{ entiers } \geqq 0;\ \lambda > 0.$$

Alors près de 0_+, $I_{q,p}$ a le comportement suivant:

$q > 2p$, $I_{q,p}$ borné;

q impair $I_{q,p}(\lambda) = (\text{fonction analytique en } \lambda) + \dfrac{C}{\lambda^{p-\frac{q}{2}}}$;

q pair $I_{q,p}(\lambda)=(\text{fonction analytique en } \lambda)+C\cdot I_{0,p-\frac{q}{2}}$;

$I_{0,0}(\lambda)=-\frac{1}{2}\operatorname{Log}\lambda+(\text{fonction analytique en } \lambda)$;

$p>0$, $I_{0,p}(\lambda)=\dfrac{g_p(\lambda)}{\lambda^p}$; g_p analytique en λ.

Ainsi, un terme en logarithme dans l'évaluation asymptotique de $I_{q,p}$ apparaît si et seulement si $q=2p$.

Il est maintenant possible de préciser les singularités dans chaque L_k.

Rappelons que $h(v)=(h(v_1),\ldots,h(v_k))$. On mettra la fonction $h(v_i)$ sous la forme $v_i^3\,U(v_i)$; on a d'après (3') $U^{(p)}(0)=C_p\,y^{(p+3)}(a)$, où C_p est une constante. On pose: $v^3\,U(v)=(v_1^3\,U(v_1),\ldots,v_k^3\,U(v_k))$.

On pose aussi $\lambda_1=-8c\beta\lambda$. Le changement de variables linéaire inversible $(w_2,\ldots,w_{k+1})\mapsto(v_1,\ldots,v_k)$ est noté T.

En utilisant l'homogénéïté de Q_k, l'intégrale (18) devient à une constante multiplicative universelle près:

$$\frac{1}{\beta^{k-6}}\,y''(a)^{2k-4}\int_{\mathcal{O}}\frac{Q_k\left(v^3\,U\left(\frac{v}{d}\right)\right)dv\,d\tau_3,\ldots,d\tau_{k+1}}{\left(\lambda_1+\sum_{j=2}^{k+1}w_j^2+\sum_{j=3}^{k+1}\tau_j^2\right)^{\frac{1}{2}+2(k+1)}}. \tag{23}$$

Lemme 6. *Avec les notations ci-dessus:*

(i) *L_k est borné en λ pour $k>6$;*

(ii) *la fonction puissance dominante $1/\lambda^3$ apparaît uniquement dans L_0.*

(iii) *Le coefficient du terme en logarithme dans L_k; $1\leqq k\leqq 6$, est:*

$$\frac{H^{(6-k)}(0)}{y''(a)^{k+2}} \tag{24}$$

où

$$H(\rho)=\int_{S^{2k-2}}Q_k(T(w')^3\,U(\rho\,T(w')))\,d\sigma.$$

S^{2k-2} est la sphère unité de $\mathbb{R}^k(w)\times\mathbb{R}^{k-1}(\tau_3,\ldots,\tau_{k+1})$ et $d\sigma$ est l'élément d'aire usuel sur S^{2k-2}.

$$(w_2,\ldots,w_{k+1},\tau_3,\ldots,\tau_{k+1})=\rho(w'_2,\ldots,w'_k,\tau'_3,\ldots,\tau'_{k+1})=\rho(w',\tau')$$
$$\text{et}\quad (w',\tau')\in S^{2k-2}.$$

Démonstration. La forme du dénominateur de l'intégrant dans (23) suscite le passage en polaires dans $\mathbb{R}^{2k-1}$. On note $(w,\tau)=\rho(w',\tau')$ comme il est rappelé dans l'énoncé de la proposition. (23) devient:

$$\frac{1}{\beta^{k-6}\,y''(a)^{2k-4}}\int_0^{\rho_0}\frac{\rho^{5k-2}\,d\rho}{(\lambda_1+\rho^2)^{\frac{1}{2}+2(k+1)}}\left\{\int_{S^{2k-2}}Q_k\left(T(w')^3\,U\left(\frac{\rho}{d}\,T(w')\right)\right)d\sigma_{w',\tau'}\right\}$$

L'expression entre accolades est une fonction régulière en ρ qu'on note $H\left(\frac{\rho}{d}\right)$. En tenant compte du comportement des intégrales $I_{q,p}$ dans la recherche du terme en logarithme, si on effectue un développement de Taylor en 0 de H, il convient de ne s'intéresser qu'au terme d'ordre $6-k$ en ρ (seul celui-ci donne une intégrale du type $I_{q,p}$ avec $q=2p$). Ce terme est $\frac{H^{(6-k)}(0)}{d^{6-k}}$ à une constante près. En outre, si $k>6$ l'expression ci-dessus est bornée en λ_1, donc en λ. Pour $1 \leqq k \leqq 6$, le coefficient du terme logarithme voit disparaître la contribution de β, et c'est (24). On remarque aussi que les fonctions puissances intervenant pour $k=1,\ldots,6$ sont $1/\lambda^2$ et $1/\lambda$. La proposition est prouvée. □

On est maintenant en mesure de calculer le terme dominant de la singularité logarithmique du noyau de Bergman de Ω.

Un polynôme en k variables, de degré $6-k$, à coefficients réels, de la forme

$$\sum_{\substack{i_1 \leqq i_2 \leqq \ldots \leqq i_k \\ i_1+\ldots+i_k=6-k}} \alpha_{i_1\ldots i_k} X_1^{i_1} \ldots X_k^{i_k}$$

est noté P_{6-k}^k.

Pour f, fonction régulière d'une variable, on pose alors

$$P_{6-k}^k[f] = \sum \alpha_{i_1,\ldots,i_k} f^{(i_1)} \ldots f^{(i_k)}.$$

Théorème 3. *Soit Ω un domaine de Reinhardt complet, borné, strictement pseudo-convexe à frontière C^∞ de $\mathbb{C}^2$. La frontière dans le plan des logarithmes est définie par une fonction y. Pour $z=(z_1,z_2)\in\Omega$, on pose*

$$a=\operatorname{Log}(|z_1|) \quad et \quad \lambda = \operatorname{Log}|z_2| - y(\operatorname{Log}|z_1|).$$

Alors, il existe des polynômes P_{6-k}^k, $k=1,2,\ldots,6$, dont les coefficients ne dépendent que de Ω, tels qu'en tout point de la frontière de Ω dont aucune coordonnée est nulle, le coefficient dominant de la singularité logarithmique du noyau de Bergman de en ce point est donnée par la formule

$$\frac{1}{|z_1^0|^2\,|z_2^0|^2}\left(\sum_{k=1}^{6} P_{6-k}^k[y'''](a)/y''(a)^{k+2}\right).$$

Démonstration. Pour obtenir $\phi_2(a_0,0)$ il faut expliciter l'expression $H^{(6-k)}(0)$, $1 \leqq k \leqq 6$ du théorème précédent. Q_k est une somme de monômes homogènes d'ordre k. Prenons en un $X_1 X_2 \ldots X_k$ par exemple, la méthode est la même pour les autres; il s'agit de trouver le coefficient de ρ^{6-k} dans le développement de Taylor en 0 de

$$T_1(w')^3 \ldots\ldots T_k(w')^3\, U(\rho\, T_1(w')) \ldots U(\rho\, T_k(w')).$$

(T_i est la $i^{\text{ème}}$ fonction coordonnée de l'application T). Ce coefficient est modifié par une constante universelle après intégration en w'. C'est aussi bien le coefficient de ρ^{6-k} de la fonction $\rho \to U(\rho)^k$. Pour cela, on a

$$U(\rho) = \sum_{p=0}^{6-k} \frac{U^{(p)}(a)}{p!}\rho^p + \mathcal{O}(\rho^{6-k}).$$

D'après la symétrie du propos si $i_i, \ldots, i_k$ sont des entiers tels que $i_1 \leqq i_2 \leqq \ldots \leqq i_k$ avec $i_1 + \ldots + i_k = 6-k$, on obtient dans $\left(\sum_{p=0}^{6-k} \frac{U^{(p)}(0)}{p!} \rho^p\right)^k$ pour terme associé au choix $i_1, \ldots, i_k$, le coefficient $U^{(i_1)}(0) \ldots U^{(i_k)}(0)$ à une constante près. Pour tous les choix possibles des indices $i_1, \ldots, i_k$ sachant que $U(0) = y''(a_0)$ il apparaît $P^k_{6-k}[y'''](a_0)$. L'expression complète du terme $\phi_2(a_0, 0)$ résulte de (7') développé à un ordre q assez grand. Tout est prouvé. □

6. Conjecture de Ramadamov

L'absence de terme dominant en logarithme au bord par le noyau de Bergman se traduit par une équation différentielle ordinaire satisfaite par le bord du domaine en coordonnées logarithmiques du type:

$$(E): \quad \sum_{k=1}^{6} P^k_{6-k}[y''']/(y'')^{k+2} = 0; \qquad P^1_5[y'''] = y^{(\mathrm{VIII})}.$$

Il résulte alors immédiatement du théorème de Cauchy sur les équations différentielles à second membre analytique le résultat suivant:

Proposition. *Soit $\Omega \subset \mathbb{C}^2$ domaine strictement pseudo-convexe de Reinhardt, complet, borné à frontière C^∞. Si près d'un point du bord à coordonnées non nulles le noyau de Bergman est sans terme en logarithme alors localement la frontière est analytique réelle.*

Ramadamov a conjecturé que les seuls domaines strictement pseudo-convexes bornés de $\mathbb{C}^n$ ayant un noyau de Bergman sans terme en logarithme sont les boules.

Nous allons démontrer la véracité de cette affirmation dans C^2 au moins pour les domaines de Reinhardt complets bornés à frontière régulière. Plus précisément, on a:

Théorème 4. *Un domaine de Reinhardt, borné de $\mathbb{C}^2$, strictement pseudo-convexe, complet, à frontière C^∞, dont le terme dominant de la singularité logarithmique de son noyau de Bergman est identiquement nul sur la frontière est un ellipsoïde d'équation $a|z_1|^2 + b|z_2|^2 \leqq 1$.*

Démonstration. Après une affinité sur chacune des coordonnées, on peut supposer que y est défini sur $(-\infty, 0)$ avec $y(-\infty) = 0$ et $y(0) = -\infty$. Nous dirons que Ω est normalisé.

Puisque y est strictement concave, y'' et y' sont liés par la relation $y'' = -e^{p(y')}$, où p est une fonction régulière sur $]-\infty, 0[$. Un calcul élémentaire montre, à partir de (E), que p vérifie l'équation

$$(E_1): \quad p^{(\mathrm{VI})} = \sum a(p')^{\alpha_1}(p'')^{\alpha_2}(p''')^{\alpha_3}(p^{\mathrm{IV}})^{\alpha_4}(p^{\mathrm{V}})^{\alpha_5}$$

les a_α sont des constantes universelles; la somme est prise sur les 5-uplets α tels que $\alpha_1 + 2\alpha_2 + \ldots + 5\alpha_5 = 6$.

On notera p_0 la fonction p associée à la boule; p_0 est évidemment une solution de (E_1); d'autre part, un calcul simple conduit à $p_0(y') = \mathrm{Log}\, 2y'(y'-1)$.

Lemme 6. *Soit p une solution de (E_1) associée à un domaine strictement pseudo-convexe complet, normalisé, à frontière régulière: Posons $q=p-p_0$. Alors, quel que soit l'ordre de dérivation, on a vers $-\infty$:*

$$q^{(i)}=o\left(\frac{1}{(y')^i}\right).$$

Démonstration. Près du point $(1,0)$, le bord s'écrit

$$|z_1|^2=1-a|z_2|^2+|z_3|^3 h_1(|z_2|);$$

h_1 étant C^∞ près de 0. La stricte pseudo-convexité en $(1,0)$ impose $a>0$. Avec $x=\operatorname{Log}|z_1|$; $y=\operatorname{Log}|z_2|$ l'équation du bord devient:

$$\begin{aligned} 2x &= \operatorname{Log}(1-ae^{2y}+e^{3y}h_1(e^y)) \\ &= -ae^{2y}+e^{3y}h_2(e^y). \end{aligned} \tag{25}$$

h_2 comme les fonctions h ultérieures seront C^∞ en 0. En dérivant deux fois cette relation, on obtient:

$$-e^{p(y')}=y''=-y'^2\frac{[-2ae^{2y}+e^{3y}h_4(e^y)]}{[-ae^{2y}+e^{3y}h_5(e^y)]}=-y'^2[2+e^y h_6(e^y)]$$

ou encore

$$p(y')=\operatorname{Log}(2y'^2)+e^y h_7(e^y). \tag{26}$$

On relie e^y et y' en dérivant (25):

$$e^y=\frac{1}{a\sqrt{-y'}}(1+e^y h_8(e^y)).$$

En posant $v=e^y$, $u=\dfrac{1}{\sqrt{-y'}}$, l'expression précédente permet d'utiliser le théorème des fonctions implicites en $u=0$; $v=0$. Il existe ainsi une fonction g de classe C^∞ près de 0 avec $g(0)=0$, telle que $e^y=g\left(\dfrac{1}{\sqrt{-y'}}\right)$. L'égalité (21) devient alors près de $-\infty$:

$$p(y')=\operatorname{Log}(2y'^2)+\frac{1}{\sqrt{-y'}}h_9\left(\frac{1}{\sqrt{-y'}}\right).$$

Par ailleurs, $p_0(y')=\operatorname{Log}2y'(y'-1)+\dfrac{1}{\sqrt{-y'}}h_{10}\left(\dfrac{1}{\sqrt{-y'}}\right)$. Il en résulte: $q(y')$

$$=\frac{1}{\sqrt{-y'}}r\left(\frac{1}{\sqrt{-y'}}\right),$$

r, C^∞ en 0. On en déduit par dérivations successives le résultat cherché.

Lemme 7. *Soit p solution de E_1 sur $]-\infty,0[$; $q=p-p_0$. Si q vérifie l'estimation $q^{(i)}=\mathcal{O}\left(\frac{1}{q'^i}\right)$ en $-\infty$; $i=1,\ldots,5$; alors pour tout $y'\in]-\infty,0[$ on a $q''=0$.*

Démonstration. En remplaçant p par p_0+q dans E_1, il vient:

$$p_0^{\mathrm{VI}}+q^{\mathrm{VI}}=a_\alpha(p'_0+q')^{\alpha_1},\ldots,(p_0^{\mathrm{V}}+q^{\mathrm{V}})^{\alpha_5}$$

à l'aide de la formule multinomiale de Newton; sachant que p_0 est solution de E_1 on a:

$$q^{v'}=\sum_{k=0}^{5}\phi_k(p'_0,\ldots,p_0^{(k)},q',\ldots,q^{(6-k)})$$

ou ϕ_k est polynomiale en p'_0, $p''_0,\ldots$; q', $q'',\ldots$ avec exactement k dérivations sur p_0; $6-k$ dérivations sur q. Le cas $k=6$ disparaît du fait que p_0 est solution de E_1. Intégrons successivement cinq fois l'expression précédente de a fixé à y' pour obtenir une primitive cinquième.

$$g'(y')=\int_a^{y'}dt_5\int_a^{t_5}dt_4\int_a^{t_4}dt_3\int_a^{t_3}\int_a^{t_2}\sum_{k=0}^{5}\phi_K\,dt_1$$
$$+\left(\left(\frac{(y'-a)^4}{4!}q^{(\mathrm{V})}(a)+\ldots+q'(a)\right)\right).$$

Par Fubini, l'intégrale multiple est $\int_a^{y'}\frac{(t-a)^4}{4!}(\sum\phi_k)\,dt$.

Faisons tendre maintenant y' vers $-\infty$. $p_0^{(i)}\sim\frac{1}{y'^i}$, (on revient à l'expression de p_0), grâce à l'hypothèse $\phi_k\sim\frac{1}{(y')^6}$. Par conséquent, l'intégrale multiple est convergente; aussi $q'(y')\to 0$ par l'hypothèse. Il résulte que $\frac{(y'-a)^4}{4!}q^{(\mathrm{V})}+\ldots$ $+q'(a)$ doit être borné quand $y'\to-\infty$. On en déduit $q^{\mathrm{V}}(a)=\ldots=q''(a)=0$. Le raisonnement vaut quelque soit $a\in]-\infty,0[$. □

On achève maintenant la démonstration du théorème 4; d'après les lemmes 6 et 7, on a: $p'=p'_0+k$ (k=constante). Soit $\tilde{p}$ la fonction "p" associée au domaine $\tilde{\Omega}$ déduit de Ω par la symétrie $\{z_1|z_2\}$; on a évidemment aussi:

$$p'=p'_0+C\quad (C=\text{constante}).$$

On vérifie, d'autre part, sans difficulté que $\tilde{p}$ est lié à par la relation: $p'\left(\frac{1}{u}\right)$ $=[3u-u^2p'(u)]$.

Il vient donc:

$$3u-u^2p'_0(u)-ku^2=p'_0\left(\frac{1}{u}\right)+k+C.$$

En substituant la valeur de p'_0, on constate que $ku^2+k+C=0$. Ainsi $k=C$ $=0$ et ainsi $p=p_0+\operatorname{Log}\alpha$ avec $\alpha>0$.

Cette fonction p définit le domaine $|z_1|^{2\alpha}+|z_2|^{2\alpha}<1$. La stricte pseudo-convexité impose $\alpha=1$, ce qui termine la démonstration.

References

1. D'Angelo, J.: A note on the Bergman Kernel. Duke Math. Journal **45**, (n° 2) 259–265 (1978)
2. Bell, S.: Biholomorphic mapping and the $\bar{\partial}$-problem. Annals of Math. **114**, 103–113 (1981)
3. Bell, S., Boas, H.: Regularity of the Bergman projection in weakly pseudoconvex domains. Math. Ann. **257**, 23–30 (1981)
4. Bell, S., Ligocka, E.: A simplification and extension of Fefferman's theorem on biholomorphic mappings. Invent. Math. **57**, 283–289 (1980)
5. Boichu, D.: Conjecture de Ramadanov sur le noyau de Bergman Thèse de 3ème cycle, Université de Lille I, juin 1982
6. Bonami, A., Lohoue, N.: Projecteurs de Bergman et Szegö pour une classe de domaines faiblement pseudo-convexes et estimations L^p. Prépublications, Université de Paris-Sud
7. Boutet de Monvel, L., Sjöstrand, J.: Sur la singularité des noyaux de Bergman et de Szegö. S.M.F. Astérisque **34-35**, 123–164 (1976)
8. Coeuré, G.: Sur le noyau de Bergman des domaines de Reinhart. Publication IRMA-Lille-Vol. **3**, (fasc. 5, II) 1–16 (1981)
9. Diederich, K.: Some recent developments in the theory of the Bergman kernel function: a survey. Proceedings of Symposia in Pure Mathematics, Vol. 30 (1977)
10. Diederich, K., Fornaess, J.: Pseudo-convex domains with real-analytic boundary. Ann. of Math. **107**, 371–384 (1978)
11. Fefferman, C.: The Bergman Kernel and biholomorphic mapping of pseudoconvex domains. Invent. Math. **26**, 1–65 (1974)
12. Hörmander, L.: Introduction to complex analysis in several variables. North Holland, Amsterdam (1973)
13. Hörmander, L.: L^2-Estimates and existence theorems for the $\bar{\partial}$-operator. Acta Math. **113**, 89–152 (1965)
14. Kerzman, N.: The Bergman kernel function. Differentiability at the boundary. Math. Ann. **195**, 149–158 (1972)
15. Kohn, J.J.: Harmonic integrals on strongly pseudo-convex manifolds I and II. Ann. Math. **78**, 112–148 (1963); **79**, 450–472 (1964)
16. Kohn, J.J.: Boundary behavior of $\bar{\partial}$ on weakly pseudo-convex manifolds of dimension two. J. Differential Geometry **6**, 523–542 (1972)
17. Kohn, J.J.: Subellipticité of the $\bar{\partial}$ Neumann problem on pseudo-convex domains sufficient conditions. Acta Math. **142**, 79–122 (1979)
18. Ramadanov, I.P.: A characterisation of the balls in C^n by means of the Bergman kernel. Compte rendu de l'Académie bulgare des Sciences, Tome 34, (n° 7) (1981)
19. Skwarczynski, M.: Biholomorphic invariants related to the Bergman functions Wroclawska Drukarnia Naukowa, 1980

Oblatum 6-VII-1982

Invent. math. 72, 153-158 (1983)

Inventiones mathematicae

Addendum to "On the Variation in the Cohomology of the Symplectic Form of the Reduced Phase Space"

J.J. Duistermaat and G.J. Heckman*

Mathematisch Instituut der RUU, Budapestlaan 6, Utrecht, Netherlands
Mathematisch Instituut der RUL, Wassenaarseweg 80, Leiden, Netherlands

1. Introduction

Let M be a compact symplectic manifold with symplectic form σ, and let T be a torus acting on M in a Hamiltonian way. The symbol X is used both for an element of the Lie algebra $\underline{t}$ of T and for the corresponding vector field on M. A Hamiltonian action means that there is a momentum mapping

$$J: M \to \underline{t}^* \tag{1.1}$$

having the properties

$$(\sigma|X) = -dJ_X, \qquad X \in \underline{t}, \tag{1.2}$$

$$\sigma(X, Y) = 0, \qquad X, Y \in \underline{t}. \tag{1.3}$$

Here we have used the notation $(\sigma|X)$ for the inner product of the vector field X with the form σ, and $J_X(m) = \langle X, J(m) \rangle$ for the X-component of J.

In an earlier paper [4] it was shown that the push forward $J_*(dm)$ of the Liouville measure dm on M under the momentum mapping J is a piecewise polynomial measure on $\underline{t}^*$. Moreover, in case X has isolated isolated zeros on M an explicit formula for the integral

$$\int_M e^{i\langle X, J(m)\rangle}\, dm \tag{1.4}$$

was obtained using the method of stationary phase. The goal of this note is to extend this formula to the case that the zeros of X are not necessarily isolated.

If we write N for the zero set of X and $i: N \to M$ for the inclusion, then $i^*(\sigma)$ is a symplectic form on N. The normal bundle E of N in M has the structure of a symplectic vector bundle. Denote this symplectic structure on E by τ. By linearization X induces a fiber preserving automorphism $LX: E \to E$ leaving τ invariant. Here invariant is meant in the infinitesimal sense. We can

* Partially supported by NSF Grant MCS-8203769

choose an automorphism $i: E \to E$ commuting with LX such that $i^2 = -\mathrm{id}$ and $\tau(i.,.)$ is a Riemannian structure on E. Conclusion is that E becomes a complex vector bundle over N, and LX a complex automorphism of E.

Choose an LX-invariant connection D on E, and let Ω be the curvature matrix. Now we can formulate the following

Theorem. *If $\lambda_1, \ldots, \lambda_r$ are the different weights for the action of LX on E, then*

$$\int_M e^{J_X} e^{\sigma} = \int_N \frac{e^{i^* J_X} e^{i^* \sigma}}{\det\left(\frac{LX + \Omega}{2\pi i}\right)} \tag{1.5}$$

for $X \in \underline{t}_{\mathbb{C}} = \underline{t} \otimes_{\mathbb{R}} \mathbb{C}$ such that

$$\langle X, \lambda_j \rangle \neq 0, \quad j = 1, \ldots, r. \tag{1.6}$$

Here J_X and LX are extended linearly for $X \in \underline{t}_{\mathbb{C}}$.

Decompose $E = \bigoplus E_\lambda$ according to different weights λ, and let Ω_λ be the curvature matrix of E_λ. Clearly LX acts on E_λ by multiplication with $\langle X, \lambda \rangle$. Now

$$\begin{aligned}
\left\{\det\left(\frac{LX+\Omega}{2\pi i}\right)\right\}^{-1} &= \left\{\det\left(\frac{LX}{2\pi i}\right)\right\}^{-1} \{\det(1 + LX^{-1}\Omega)\}^{-1} \\
&= \left\{\det\left(\frac{LX}{2\pi i}\right)\right\}^{-1} \{\prod_\lambda \det(1 + LX^{-1}\Omega_\lambda)\}^{-1} \\
&= \left\{\det\left(\frac{LX}{2\pi i}\right)\right\}^{-1} \{\sum_{k \geq 0} \alpha_k(X)\}^{-1} \\
&= \left\{\det\left(\frac{LX}{2\pi i}\right)\right\}^{-1} \{\sum_{k \geq 0} \beta_k(X)\}.
\end{aligned}$$

Here $\alpha_k(X)$ and $\beta_k(X)$ are differential forms on N of degree $2k$ with $\alpha_0(X) = \beta_0(X) = 1$. Their cohomology classes in $H^{2k}(N, \mathbb{C})$ are polynomials in the Chern classes of the various E_λ. The coefficients of these polynomials are rational functions in X, and homogeneous of degree $-k$. Because $i^* J_X$ is locally constant on N we obtain that the integrand in the right hand side of (1.5) is a closed form on N whose cohomology class is a polynomial in the cohomology class of $i^* \sigma$ and the Chern classes of E_λ. The coefficients of this polynomial are meromorphic in X, and analytic for $X \in \underline{t}_{\mathbb{C}}$ satisfying (1.6). Because the left hand side of (1.5) is an analytic function in X on all of $\underline{t}_{\mathbb{C}}$ it is sufficient to prove (1.5) for $X \in \underline{t}$ satisfying (1.6).

The proof of the theorem follows with minor adaptions [2] where a similar formula is obtained for M a complex manifold, L a holomorphic line bundle on M with Chern class $[\sigma]$ and X a holomorphic vector field on M which acts on L.

That Bott's ideas could be extended from the holomorphic to the symplectic case was observed in [1], at least in the situation that X has isolated zeros.[1]

---

1 After this note was written we learned that the case of non isolated zeros has also been treated in a somewhat more general context in: N. Berline et M. Vergne, Classes caractéristiques équivariantes. Formule de localisation et chomologie équivariante, to appear in the Comptes Rendus

The purpose of this paper is just to carry out the explicit calculation needed for this extension. The proof presented here is self-contained modulo the citation of formula (2.28).

2. Proof of the Theorem

Fix a T-invariant Riemannian metric g on M. Using the exponential mapping we obtain a diffeomorphism ψ from a neighborhood U of the zero section in E onto a neighborhood $\psi(U)$ of N in M. We can take U invariant under multiplication μ_ε in the fiber by positive constants $\varepsilon \leqq 1$. The push forward under ψ of the linear vector field LX on U is equal to X. Suppose $L\theta$ is a 1-form on $E \setminus N$ having the properties

$$(L\theta | LX) = 1, \tag{2.1}$$

$$(dL\theta, LX) = 0, \tag{2.2}$$

$$\mu_\varepsilon^* L\theta = L\theta. \tag{2.3}$$

Given (2.1) the condition (2.2) is equivalent to $\mathscr{L}_{LX}(L\theta) = 0$. Using the 1-form $g(X, X)^{-1} g(X, .)$ on $M \setminus N$ and a partition of unity we obtain a 1-form θ on $M \setminus N$ satisfying

$$(\theta | X) = 1, \tag{2.4}$$

$$(d\theta | X) = 0 \tag{2.5}$$

and such that the pull back under ψ of the form θ on $\psi(U)$ is equal to $L\theta$.

Consider the $(2n-1)$-form

$$v = -\sum_{k=1}^{n} (-1)^k e^{Jx} \theta \wedge (d\theta)^{k-1} \wedge \frac{\sigma^{n-k}}{(n-k)!} \tag{2.6}$$

on $M \setminus N$. An easy computation shows that

$$(dv | X) = \left(e^{Jx} \frac{\sigma^n}{n!} \middle| X \right) \tag{2.7}$$

which in turn implies that

$$dv = e^{Jx} \frac{\sigma^n}{n!} \tag{2.8}$$

on $M \setminus N$. For positive ε denote by B_ε the ε-ball bundle in E, and by S_ε the boundary of B_ε. Clearly

$$\int_M e^{Jx} \frac{\sigma^n}{n!} = \lim_{\varepsilon \downarrow 0} \int_{M \setminus \psi(B_\varepsilon)} e^{Jx} \frac{\sigma^n}{n!} \tag{2.9}$$

which by Stokes' theorem is equal to

$$\lim_{\varepsilon \downarrow 0} \int_{\psi(S_\varepsilon)} \sum_{k=1}^{n} (-1)^k e^{Jx} \theta \wedge (d\theta)^{k-1} \wedge \frac{\sigma^{n-k}}{(n-k)!}. \tag{2.10}$$

The mapping $\psi\mu_\varepsilon$ is a diffeomorphism from the unit normal sphere bundle S to $\psi(S_\varepsilon)$. Applied to forms which are smooth on all of U $\lim_{\varepsilon\downarrow 0}(\psi\mu_\varepsilon)^*$ is equal to $(i\pi)^* = \pi^* i^*$. Here $\pi: E\to N$ is the natural projection. The 1-form $(\psi\mu_\varepsilon)^*\theta = L\theta$ does not depend on ε. Hence (2.10) is equal to

$$\int_S \sum_{k=1}^{n} (-1)^k e^{\pi^* i^* J x} L\theta \wedge (dL\theta)^{k-1} \wedge \frac{(\pi^* i^* \sigma)^{n-k}}{(n-k)!}. \tag{2.11}$$

In order to construct the 1-form $L\theta$ on $E\backslash N$ we choose a connection

$$D: \Gamma(E)\to\Gamma(T^*N\otimes E) \tag{2.12}$$

on E. For an exposition of the theory of connections and characteristic classes see Chern's book [3]. From now on we will replace N by one of its connected components. Let q be the rank of E as a complex vector bundle.

Let $V\subset N$ be an open set, and ${}^t s=(s_1, \ldots, s_q)$ a frame field over V. This gives a trivialization of E over V by

$$\begin{gathered} V\times\mathbb{C}^q\to\pi^{-1}(V) \\ (n,z)\to\sum z_i s_i(n), \quad {}^t z=(z_1, \ldots, z_q). \end{gathered} \tag{2.13}$$

If $s'=gs$ is a new frame field over V, then

$$z={}^t g z', \qquad dz={}^t dg z' + {}^t g dz'. \tag{2.14}$$

The connection matrix ω of D relative to the frame field s is defined by

$$Ds=\omega s. \tag{2.15}$$

The transformation formula for a change of frame field is

$$\omega' g = dg + g\omega. \tag{2.16}$$

Consider the vector valued 1-form Dz on E given by

$$Dz = dz + {}^t\omega z. \tag{2.17}$$

The form Dz transforms under a change of frame field according to

$$Dz={}^t g Dz'. \tag{2.18}$$

Moreover, an easy calculation shows that

$$d(Dz)={}^t\Omega z - {}^t\omega\wedge Dz \tag{2.19}$$

where

$$\Omega = d\omega - \omega\wedge\omega \tag{2.20}$$

is the curvature matrix relative to s.

Let H be a Hermitean structure on E. Assume that the connection D is admissable with respect to H. Hence, for a unitary frame field s over V we have

$$\omega+{}^t\bar{\omega}=0, \qquad \Omega+{}^t\bar{\Omega}=0. \tag{2.21}$$

From now on s and $s'=gs$ will be unitary frame fields over V. Because X generates a compact group we may assume that LX is leaving both the Hermitean structure H and the connection D invariant, i.e.

$$LX+{}^t\overline{LX}=0, \tag{2.22}$$

$$[LX,\omega]=0, \qquad [LX,\Omega]=0. \tag{2.23}$$

Moreover we will choose the frame fields s and $s'=gs$ in such a way that $d(LX)=0$.

Consider the 1-form

$$L\theta=(z,z)^{-1}(z,{}^tLX^{-1}Dz) \tag{2.24}$$

on $E\backslash N$. Clearly $L\theta$ is independent of the choice of the frame field s, and satisfies the conditions (2.1), (2.2) and (2.3). Using (2.19), (2.21) and (2.23) one has

$$L\theta\wedge(dL\theta)^{k-1}=(z,z)^{-k}(z,{}^tLX^{-1}Dz)\wedge\{(Dz,{}^tLX^{-1}Dz)+(z,{}^tLX^{-1}\,{}^t\Omega z)\}^{k-1}.$$

Hence (2.11) can be rewritten as

$$\int_N e^{i^*Jx}e^{i^*\sigma}\sum_{k=q}^{n}\sum_{j=1}^{k}(-1)^k\binom{k-1}{j-1}$$
$$\times\int_{S^{2q-1}}(z,{}^tLX^{-1}Dz)\wedge(Dz,{}^tLX^{-1}Dz)^{j-1}\wedge(z,{}^tLX^{-1}\,{}^t\Omega z)^{k-j}. \tag{2.26}$$

Here S^{2q-1} is the unit sphere in $\mathbb{C}^q$. The second integrand has to be integrated point wise for $n\in N$. The outcome of this integral is independent of the choice of the frame field s. But for fixed $n\in V$ one can always choose a frame field s, which is horizontal at n. Therefore, in the summation over j in (2.26) all summands vanish except the one with $j=q$. Hence (2.26) becomes

$$\int_N e^{i^*Jx}e^{i^*\sigma}(-1)^q\int_{S^{2q-1}}\frac{(z,{}^tLX^{-1}dz)\wedge(dz,{}^tLX^{-1}dz)^{q-1}}{\{(z,z)+(z,{}^tLX^{-1}\,{}^t\Omega z)\}^q}. \tag{2.27}$$

Using a formula of Bott ([2], p. 327)

$$\int_{S^{2q-1}}\frac{(z,dz)\wedge(dz,dz)^{q-1}}{\{(z,z)+t(z,Sz)\}^q}=\frac{1}{\det(1+tS)}\int_{S^{2q-1}}(z,dz)\wedge(dz,dz)^{q-1} \tag{2.28}$$

for any $q\times q$ Hermitean matrix S, we can rewrite (2.27) as

$$\int_N\frac{e^{i^*Jx}e^{i^*\sigma}(-1)^q\det({}^t\overline{LX}^{-1})}{\det(1+{}^tLX^{-1}\,{}^t\Omega)}\int_{S^{2q-1}}(z,dz)\wedge(dz,dz)^{q-1}. \tag{2.29}$$

With our orientation of $\mathbb{C}^q$ we get

$$\int_{S^{2q-1}} (z, dz) \wedge (dz, dz)^{q-1} = \int_{B^{2q}} (dz, dz)^q = (2\pi i)^q \tag{2.30}$$

and because LX^{-1} is skew-Hermitean (2.29) can be written as

$$\int_N \frac{e^{i^* J_X}\, e^{i^* \sigma}}{\det\left(\dfrac{LX+\Omega}{2\pi i}\right)}. \tag{2.31}$$

This finishes the proof of the theorem.

References

1. Berline, N., Vergne, M.: Zeros d'un champ de vecteurs et classes caracteristiques equivariantes. Preprint M.I.T. 1982
2. Bott, R.: A residue formula for holomorphic vector fields. J. Diff. Geometry **1**, 311–330 (1967)
3. Chern, S.S.: Complex manifolds without potential theory. Van Nostrand Math. Studies #**15** (1967)
4. Duistermaat, J.J., Heckman, G.J.: On the variation in the cohomology of the reduces phases space. Invent. Math. **69**, 259–268 (1982)

Oblatum 2-XI-1982

Invent. math. 72, 159–199 (1983)

Inventiones mathematicae

F-Isocrystals and De Rham Cohomology. I

P. Berthelot[1] * and A. Ogus[2] **

[1] U.E.R. de Mathématiques, Université de Rennes, Campus de Beaulieu, F-35042 Rennes Cedex, France
[2] Department of Mathematics, University of California, Berkeley, CA 94720, USA

0. Introduction

This paper is devoted to a study of crystalline cohomology tensored with $\mathbb{Q}$. Its main theme is that, after tensoring with $\mathbb{Q}$, one obtains much stronger "analytic continuation" properties, in senses we shall try to make precise, than when working with modules over the Witt ring. For example, if V is a complete discrete valuation ring with perfect residue field k of characteristic p, we can think of Spec V as an analogue of a small disc in the classical case, whose size depends on the absolute index of ramification e. If X is a smooth proper V-scheme, a central "analytic continuation" result of crystalline cohomology asserts that the cohomology of the family $H^i_{\mathrm{DR}}(X/V)$ depends only on the closed fiber X_0, not on the lifting X, provided e is strictly less than p. If $p \leqq e < \infty$, we show here that this is still true "up to isogeny," i.e. $H^i_{\mathrm{DR}}(X/V) \otimes \mathbb{Q}$ depends only on X_0, and in fact is canonically isomorphic to $H^i_{\mathrm{cris}}(X_0/W) \otimes V \otimes \mathbb{Q}$, as had already been conjectured in [1].

The main idea in our approach is the systematic exploitation of the action of the absolute Frobenius endomorphism on crystalline cohomology – an idea which, in fact, goes back to Dwork [8, 9] and which was made very explicit for us in [13]. The technical key is (1.3), which asserts that, with suitably hypotheses, the relative Frobenius morphism $F_{X/S}$ induces an isogeny on the crystalline cohomology of a smooth family X/S of schemes in characteristic p.

* Laboratoire Associé au C.N.R.S. n° 305; partially supported by N.S.F. grant n° MCS80-23848
** Partially supported by N.S.F. grant n° MCS82-02651 and the Sloan Foundation

We in fact also show in (1.9) that, under suitable hypotheses, the F-crystal structure $F^*_{X/S}: F^*_{X/S} R^q f_* \mathcal{O}_{X/S} \to R^q f_* \mathcal{O}_{X/S}$ is nondegenerate in the sense of [21]: there is a morphism of crystals:

$$V_{X/S}: R^q f_* \mathcal{O}_{X/S} \to F^*_{X/S} R^q f_* \mathcal{O}_{X/S}$$

such that $V_{X/S} \circ F^*_{X/S}$ and $F^*_{X/S} \circ V_{X/S}$ are multiplication by a certain power of p.

The second section studies the crystalline cohomology of infinitesimal deformations. Using the results of § 1, we prove that if $f: X \to S$ is a smooth morphism in characteristic p, if $S_0 \hookrightarrow S$ is a nilpotent immersion, and if $S \hookrightarrow T$ is a suitable divided power thickening, then $\mathbb{R} f_*(\mathcal{O}_{X/T}) \otimes^{\mathbb{L}} \mathbb{Q}$ depends functorially only on $X \times_S S_0$. This says that crystalline cohomology is, up to isogeny, invariant under deformations. (Note that no divided power structure on $S_0 \hookrightarrow S$ or on $S_0 \hookrightarrow T$ is required.) We easily deduce from this the existence of the isomorphism

$$\sigma_{\text{cris}}: H^i_{\text{DR}}(X/V) \otimes \mathbb{Q} \xrightarrow{\sim} H^i_{\text{cris}}(X_0/W) \otimes V \otimes \mathbb{Q}$$

alluded to above, and verify several compatibilities. In addition, we give estimates on the relative positions of the two lattices $H^i_{\text{DR}}(X/V)$ and $H^i_{\text{cris}}(X_0/W) \otimes V$ in these vector spaces, and explain an explicit example.

The remaining sections of Part I of this paper are devoted to compatibilities and applications. In the third section, for example, we verify in (3.5) that σ_{cris} is compatible with formation of Chern classes of line bundles and show in (3.8) that if $L_0 \in \text{Pic}(X_0)$, then a suitable (explicit) power of L_0 lifts to X iff $\sigma^{-1}_{\text{cris}}(c_1(L_0))$ belongs to $F^1_{\text{Hodge}} H^2_{\text{DR}}(X/V)$. We also explain in (3.12) some conditions under which ramification is required for the lifting of line bundles and prove some analogous results for lifting morphisms of p-divisible groups and abelian varieties (3.15), and K3 surfaces (3.23).

In § 4, we study the $W \otimes \mathbb{Q}$-module structure and action of Frobenius on $H^*_{\text{DR}}(X/V) \otimes \mathbb{Q}$ induced by the isomorphism σ_{cris}. To keep track of these in a convenient way, we follow an idea of Deligne and introduce a group called the "crystalline Weil group" and a canonical semilinear action of this group on $H^*_{\text{DR}}(X/V) \otimes \mathbb{Q}$ which summarizes these structures. Combining our results with some results of Messing and Gillet, we find that this entire package of data depends only upon the generic fiber of X/V. In the case of potentially good reduction, an interesting formula ((4.7)) expresses the inertial part of this action in terms of the geometric action of the inertia group on the closed fiber. This last result had in fact already been stated by Messing, in a somewhat different form. We end this section by conjecturing in (4.8) that an absolute Hodge cycle in De Rahm cohomology is invariant under the crystalline Weil group (which would be a consequence of the Hodge conjecture). We are able to verify this conjecture for abelian varieties of CM type with ordinary reduction.

Part I of this paper ends with a technical appendix illustrating the existence of torsion in divided power envelopes. This explains some of the technical difficulties we encounter in the early sections. Moreover, it leads to the apparently pathological appearance of torsion in $H^0_{\text{cris}}(S/W)$ for certain (singular) schemes S, and to a counterexample to the full faithfulness of the Dieudonné crystal associated to p-divisible groups over a general base.

Part II of this paper will be devoted to a further elaboration of the results here. Its main ingredient will be the study of the F-isocrystals arising from families of varieties, enabling us to "analytically continue" F-isocrystals to certain p-adic neighborhoods. We shall especially study the F-isocrystals arising from families of abelian varieties and K3 surfaces, and shall use these techniques to verify Conjecture (4.8) in a very subtle case. For more details, we refer the reader to the introduction of part II.

Here are some conventions that will be in use throughout both parts of this paper.

We use k to denote a perfect field of characteristic $p>0$, $W(k)$ or W to denote its Witt ring, and V for a finite extension of W, i.e. a finite flat W-algebra which is also a discrete valuation ring (DVR). All morphisms of schemes are quasi-compact and quasi-separated. All formal V-schemes will have the p-adic topology, unless otherwise stated.

We shall have to deal systematically with crystalline cohomology in the limit, using the techniques of [3, § 7]. If $S \subseteq T$ is a divided power (PD) thickening and if X is an S-scheme, then $\mathrm{Cris}(X/T)$ will denote the site consisting of the T-PD thickenings of open subsets of X on which p is nilpotent. (This is the site denoted $\mathrm{Cris}(X/\hat{T})$ in [3, § 7].)

We would like to express our gratitude to P. Deligne (who first proved the existence of σ_{cris} for abelian varieties), W. Messing, and N. Katz, for several helpful discussions. We also want to reiterate our dept to the seminal work of B. Dwork. The authors would like very much to thank respectively the University of California at Berkeley and the Institute for Advanced Study at Princeton for their hospitality, which has made this work possible.

1. The Relative Frobenius Morphism

(1.1) If X is a scheme in characteristic p, we let F_X denote its absolute Frobenius endomorphism. If $f: X \to S$ is a morphism in characteristic p, we have the familiar relative Frobenius diagram

$$\begin{tikzcd} X \arrow[r, "F_{X/S}"] \arrow[dr, "f"'] & X^{(p/S)} \arrow[r, "W_{X/S}"] \arrow[d, "f^{(p/S)}"] & X \arrow[d, "f"] \\ & S \arrow[r, "F_S"] & S. \end{tikzcd}$$

When there seems to be no confusion possible, we shall write $f': X' \to S$ instead of $f^{(p/S)}: X^{(p/S)} \to S$.

Let $\Sigma =: \mathrm{Spec}(\mathbb{Z}_p)$ endowed with its canonical PD-ideal (p). Our object is to study the morphism

(1.1.1) $$F^*_{X/S}: \mathbb{R}f'_{\mathrm{cris}*}\mathcal{O}_{X'/\Sigma} \to \mathbb{R}f_{\mathrm{cris}*}\mathcal{O}_{X/\Sigma}$$

in $(S/\Sigma)_{\mathrm{cris}}$ and the corresponding morphisms on the cohomology sheaves. Actually, we shall only do this indirectly. We shall first study the behavior of the morphism on crystalline cohomology relative to a "p-adic base" T, where T

is a formal scheme (with the p-adic topology), and $S \hookrightarrow T$ is a closed immersion defined by a PD-ideal $(\mathscr{I}, \gamma)$ compatible with the standard divided power structure on $(p) \subseteq W$. Note that if $\mathcal{O}_T$ is p-torsion free, as we shall henceforth assume, this last condition is automatic.

A key example is the following:

(1.2) **Example.** Let k be a perfect field of characteristic p, $W = W(k)$ the ring of Witt vectors with coefficients in k, V a finite extension of W of (arbitrary) ramification index e, and $S = \operatorname{Spec}(V/pV) \cong \operatorname{Spec}(k[t]/(t^e))$. Then we can regard $T = \operatorname{Spf}(V)$ as a PD-thickening of S. Note that if $e > 1$, the absolute Frobenius morphism of S cannot be lifted to an endomorphism of T.

(1.3) **Theorem.** *Suppose that X/S is smooth (of finite type), S is noetherian, and $S \hookrightarrow (T, \mathscr{I}, \delta)$ is a PD-immersion of S in a formal scheme. Let $u_{X/T}\colon (X/T)_{\mathrm{cris}} \to X_{\mathrm{Zar}}$ be the canonical projection, and*

$$f_{X/T}\colon (X/T)_{\mathrm{cris}} \xrightarrow{u_{X/T}} X_{\mathrm{Zar}} \xrightarrow{f_{\mathrm{Zar}}} T_{\mathrm{Zar}}$$

be the composed morphism. If $\mathcal{O}_T$ is p-torsion free, the natural map

$$F_{X/T} =: F^*_{X/S}\colon \ \mathbb{R} f'_{X'/T*}\mathcal{O}_{X'/T} \otimes^{\mathbb{L}} \mathbb{Q} \to \mathbb{R} f_{X/T*}\mathcal{O}_{X/T} \otimes^{\mathbb{L}} \mathbb{Q}$$

induced by $F_{X/S}$ is an isomorphism.

Proof. For any abelian sheaf F on X_{Zar}, the presheaf $U \mapsto F(U) \otimes \mathbb{Q}$ is a sheaf, since X_{Zar} is noetherian. This implies that the natural map $(f_{\mathrm{Zar}*}F) \otimes \mathbb{Q} \to f_{\mathrm{Zar}*}(F \otimes \mathbb{Q})$ is an isomorphism, and it follows that the same is true in the derived category: the morphism

$$(\mathbb{R} f_{\mathrm{Zar}*}F) \otimes^{\mathbb{L}} \mathbb{Q} \to \mathbb{R} f_{\mathrm{Zar}*}(F \otimes^{\mathbb{L}} \mathbb{Q})$$

is an isomorphism for any $F \in D(X_{\mathrm{Zar}})$. Thus, since $f_{X/T} = f_{\mathrm{Zar}} \circ u_{X/T}$, it suffices to prove that the arrow

$$F^*_{X/S}\colon \mathbb{R}\, u_{X'/T*}\mathcal{O}_{X'/T} \otimes^{\mathbb{L}} \mathbb{Q} \to \mathbb{R}\, u_{X/T*}\,\mathcal{O}_{X/T} \otimes^{\mathbb{L}} \mathbb{Q} \tag{1.3.1}$$

is a quasi-isomorphism.

This statement is local on T and on X. We may assume, for instance, that X lifts to a smooth formal scheme Y over T. Suppose for a moment that the absolute Frobenius endomorphism F_S lifts to an endomorphism F_T – necessarily a PD morphism, since $\mathcal{O}_T$ is torsion free – and that F_X lifts to an F_T-endomorphism F_Y of Y. Then if $Y' =: Y \times_{F_T} T$, we get a lifting $F_{Y/T}\colon Y \to Y'$ of the relative Frobenius morphism $F_{X/S}$, and the action $F^*_{Y/T}$ of $F_{Y/T}$ on de Rham cohomology represents the action of $F_{X/S}$ on crystalline cohomology. Now, it is easy to check that in [3, (8.1)–(8.24)], one can replace the p-adic base by any p-torsion free formal scheme, so that these results remain valid in our setting. Then [3, (8.3)] tells us the precise image of $F^*_{Y/T}$, from which we can easily deduce the theorem.

To eliminate the hypothesis that F_S lift to T, we can try to choose a suitable lifting Y' of X' and a lifting $F_{Y/T}$ of $F_{X/S}$ for which [3, (8.3)] is still true. This is not difficult, and the only real difference occurs in the proof of the very first lemma [3, (8.5)], which we must generalize to the following:

(1.4) **Lemma.** *If X/S and T are as above, let Y/T be any smooth lifting of X/S, let $\bar{Y}/\bar{T}$ be its reduction mod p, and let $F_{Y/T}\colon Y \to Y'$ be any lifting of the relative Frobenius map: $F_{\bar{Y}/\bar{T}}\colon \bar{Y}/\bar{T} \to \bar{Y}'/\bar{T}$. Then the action of $F^*_{Y/T}$ on $\Omega^j_{Y'/T}$ is divisible by p^j, and there is a commutative diagram:*

$$\begin{tikzcd}
\Omega^j_{Y'/T} \arrow[r, "\varphi^j"] \arrow[d, "\pi^j"'] & \mathscr{H}^j(F_{Y/T*}\Omega^\bullet_{\bar{Y}/\bar{T}}) \arrow[r, "\sim"] & \mathscr{H}^j(F_{X/S*}\Omega^\bullet_{X/S}) \\
\Omega^j_{\bar{Y}'/\bar{T}} \arrow[r, "\sim"] \arrow[ur, "C^{-1}"] & \Omega^j_{X'/S} \arrow[ur, "C^{-1}"'] &
\end{tikzcd}$$

where π^j is the natural projection, C^{-1} is the inverse Cartier operator, and $\varphi^j(\omega)$ is the image in $\mathscr{H}^j$ of $p^{-j}F^*_{Y/T}(\omega)$.

Proof. First of all, $\Omega^1_{\bar{Y}'/\bar{T}} = W^*_{\bar{Y}/\bar{T}}(\Omega^1_{\bar{Y}/\bar{T}})$, hence is generated as an $\mathcal{O}_{\bar{Y}'}$-module by elements of the form $W^*_{\bar{Y}/\bar{T}}(d\bar{\alpha})$, where $\bar{\alpha}$ is a local section of $\mathcal{O}_{\bar{Y}}$. For each $\bar{\alpha}$, choose a lifting α' of $W^*_{\bar{Y}/\bar{T}}(\bar{\alpha})$ in $\mathcal{O}_{Y'}$. Then the set of all these $d\alpha'$ generates $\Omega^1_{Y'/T}$ as an $\mathcal{O}_{Y'}$-module, since it does so mod p. Choose a lifting α of $\bar{\alpha}$ to $\mathcal{O}_Y$; then there is a $\beta \in \mathcal{O}_Y$ such that $F^*_{Y/T}(\alpha') = \alpha^p + p\beta$. Hence $F^*_{Y/T}(d\alpha') = p\alpha^{p-1}d\alpha + p\,d\beta$, and in particular $F^*_{Y/T}$ is divisible by p on $\Omega^1_{Y'/T}$. Moreover, $p^{-1}F^*_{Y/T}(d\alpha') = \alpha^{p-1}d\alpha + d\beta$, and $C^{-1}(d\bar{\alpha}') = C^{-1}(dW^*_{\bar{Y}/\bar{T}}(\bar{\alpha}))$ = the cohomology class of $\bar{\alpha}^{p-1}d\bar{\alpha}$. This proves the lemma with $j=1$, and the general case follows by taking exterior powers. □

We can now proceed exactly as in [3] to obtain the following generalization of [(8.3)].

(1.5) **Proposition.** *With the notations of* (1.4), *let $N^\bullet_{Y/T}$ be the largest subcomplex of $F_{Y/T*}\Omega^\bullet_{Y/T}$ such that $N^k_{Y/T}$ is contained in $p^kF_{Y/T*}\Omega^k_{Y/T}$ for all k. Then $F^*_{Y/T}$ factors through a quasi-isomorphism: $\psi_{Y/T}\colon \Omega^\bullet_{Y'/T} \to \mathrm{N}^\bullet_{Y/T}$.*

Since the inclusion $N^\bullet_{Y/T} \subseteq F_{Y/T*}\Omega^\bullet_{Y/T}$ obviously becomes a quasi-isomorphism when tensored with $\mathbb{Q}$, this proves Theorem (1.3). □

For the purposes of the remainder of this paper, Theorem (1.3) is adequate, and the reader may, if he wishes, proceed to Sect. 2. However, it still remains for us to return to the original problem of studying the morphism (1.1.1) in $(S/\Sigma)_{\mathrm{cris}}$; moreover, it is natural to ask for precise p-adic estimates for $F_{X/T}$ in Theorem (1.3). For these purposes, we shall construct an inverse, up to isogeny, to $F_{X/T}$.

(1.6) **Theorem.** *Under the hypothesis of* (1.3), *assume that the relative dimension of X/S is less than or equal to n. Then there exists a morphism*

$$V_{X/T}\colon \mathbb{R}f_{X/T*}\mathcal{O}_{X/T} \to \mathbb{R}f'_{X'/T*}\mathcal{O}_{X'/T}$$

such that $V_{X/T} \circ F_{X/T}$ and $F_{X/T} \circ V_{X/T}$ are multiplication by p^n. The morphism $V_{X/T}$ is functorial with respect to X/T: if X_1/S_1 and $S_1 \hookrightarrow T_1$ satisfy the same hypothesis *as X/S, $S \hookrightarrow T$, then for any commutative diagram*

$$\begin{array}{ccccc} X_1 & \xrightarrow{f_1} & S_1 & \hookrightarrow & T_1 \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle v} \\ X & \xrightarrow{f} & S & \hookrightarrow & T, \end{array}$$

the following square is commutative:

$$\begin{array}{ccc} \mathbb{L}v^*\mathbb{R}f_{X/T*}\mathcal{O}_{X/T} & \longrightarrow & \mathbb{R}f_{X_1/T_1*}\mathcal{O}_{X_1/T_1} \\ \downarrow{\scriptstyle \mathbb{L}v^*(V_{X/T})} & & \downarrow{\scriptstyle V_{X_1/T_1}} \\ \mathbb{L}v^*\mathbb{R}f'_{X'/T*}\mathcal{O}_{X'/T} & \longrightarrow & \mathbb{R}f'_{X'_1/T_1*}\mathcal{O}_{X'_1/T_1}. \end{array}$$

Proof. Recall from [3, (8.19)] that the operation of replacing a bounded p-torsion free complex of sheaves $A^\bullet$ by the largest subcomplex $N(A^\bullet)$ such that $N^k \subseteq p^k A^k$ passes over to the derived category. Thus, in the above (global) setting, it makes sense to form $N\mathbb{R}u_{X/T*}\mathcal{O}_{X/T}$, and there is a canonical map: $N\mathbb{R}u_{X/T*}\mathcal{O}_{X/T} \to \mathbb{R}u_{X/T*}\mathcal{O}_{X/T}$. Using the method of [3], one finds easily that (1.5) globalizes to the following result.

(1.7) **Proposition.** *The map $F_{X/T}: \mathbb{R}u_{X'/T*}\mathcal{O}_{X'/T*} \to \mathbb{R}u_{X/T*}\mathcal{O}_{X/T}$ factors naturally through a quasi-isomorphism:*

$$\psi_{X/T}: \mathbb{R}u_{X'/T*}\mathcal{O}_{X'/T} \to N\mathbb{R}u_{X/T*}\mathcal{O}_{X/T}. \quad \square$$

To define $V_{X/T}$, it is necessary to use the calculus of gauges and cogauges [3, (8.7) ff.]. In this language, the functor N is the functor $\mathbb{L}\eta$, where η is the cogauge defined by $\eta(i)=i$. For any $k \geqq 0$, let $\eta_k = \inf(\eta, k)$. There is a natural transformation: $\mathbb{L}\eta \to \mathbb{L}\eta_k$. I claim that for X/T as above, and with $K_{X/T} =: \mathbb{R}u_{X/T*}\mathcal{O}_{X/T}$, the map $\mathbb{L}\eta K_{X/T} \to \mathbb{L}\eta_n K_{X/T}$ is an isomorphism. This can be checked locally, e.g. when X lifts to a smooth formal scheme of relative dimension n. In this case $K_{X/T}$ can be represented by a complex of length n, and the statement is obvious from the definitions.

It is easy to see from the caculus of cogauges that there are natural commutative diagrams:

$$\begin{tikzcd} \mathrm{id} \arrow[rr, "p^n"] \arrow[dr, "v_n"'] & & \mathrm{id} \\ & \mathbb{L}\eta_n \arrow[ur, "i_n"'] & \end{tikzcd} \qquad \begin{tikzcd} \mathbb{L}\eta_n \arrow[rr, "p^n"] \arrow[dr, "i_n"'] & & \mathbb{L}\eta_n \\ & \mathrm{id}\,. \arrow[ur, "v_n"'] & \end{tikzcd}$$

Define $V_{X/T}: K_{X/T} \to K_{X'/T}$ to be v_n followed by the inverse of the composite:

$$K_{X'/T} \xrightarrow{\psi} \mathbb{L}\eta K_{X/T} \xrightarrow{\sim} \mathbb{L}\eta_n K_{X/T}.$$

It is clear from the diagrams that this map has the desired properties. The functoriality follows from the functoriality of each of the constructions involved. $\square$

We now return to our original study of the morphism (1.1.1). Unfortunately we shall have to make some restrictive hypothesis on the base S in order for our proof to work. We begin with the following technical result:

(1.8) **Lemma.** *Let Σ'_1 be a perfect scheme in characteristic p, $\Sigma' = \mathrm{Spf}(W(\mathcal{O}_{\Sigma'_1}))$, and let S be a Σ'_1-scheme. If $j\colon S \hookrightarrow Y$, $j'\colon S \hookrightarrow Y'$ are two closed immersions of S into smooth formal schemes over Σ', and T, T' the formal completions (along (p)) of the divided power neighborhoods of S in Y, Y', then $\mathcal{O}_{T'}$ is p-torsion free if and only if $\mathcal{O}_T$ is p-torsion free.*

Proof. Replacing the immersion $S \hookrightarrow Y'$ by the diagonal immersion, we may assume that there exists a smooth morphism of formal schemes $\pi\colon Y' \to Y$ such that $j = \pi \circ j'$. Since the assertion is local on S, we can reduce the problem to the two cases: Y' is étale over Y, or Y' is the formal affine line over Y. If $\mathcal{I}$ is the ideal of S in Y, and if Y' is étale over Y, we may assume that the ideal of S in Y' is $\mathcal{I}\mathcal{O}_{Y'}$ and that $\mathcal{O}_Y/\mathcal{I}^n \xrightarrow{\sim} \mathcal{O}_{Y'}/\mathcal{I}^n\mathcal{O}_{Y'}$ for all n. Since the divided power neighborhoods of S in the reductions Y_n, Y'_n of Y, Y' modulo p^n depend only upon the infinitesimal neighborhoods of S in Y_n, Y'_n [3, (3.20.7)], it follows that $T'_n \xrightarrow{\sim} T_n$ for all n. hence $T' \xrightarrow{\sim} T$. On the other hand, if Y' is the formal affine line over Y, we can find a section $s\colon Y \to Y'$ such that $j' = s \circ j$. If t generates the ideal of s, we get $\mathcal{O}_{Y'} \simeq \mathcal{O}_Y\{t\}$, and the ideal $\mathcal{I}'$ of S in Y' is $(\mathcal{I}, t)$. It is then immediate to check that $\mathcal{O}_{T'}$ is the algebra of formal power series $\sum_{q\in\mathbb{N}} a_q t^{[q]}$ where $a_q \in \mathcal{O}_T$ and $a_q \to 0$ if $q \to \infty$, which proves the lemma. $\square$

(1.9) **Theorem.** *Let Σ'_1 be a perfect scheme in characteristic p, $\Sigma' = \mathrm{Spf}(W(\mathcal{O}_{\Sigma'_1}))$, and let S be a Σ'_1-scheme. Assume that, locally on S, there exists an embedding $S \hookrightarrow Y$ into a Σ'-smooth formal scheme such that the formal completion T of the divided power neighborhood of S in Y is p-torsion free. If $f\colon X \to S$ is smooth of relative dimension at most n, there exists for all i a canonical homomorphism*

$$V_{X/S}\colon R^i f_{\mathrm{cris}*}\mathcal{O}_{X/\Sigma} \to R^i f'_{\mathrm{cris}*}\mathcal{O}_{X'/\Sigma}$$

such that $V_{X/S} \circ F^*_{X/S}$ and $F^*_{X/S} \circ V_{X/S}$ are multiplication by p^n. The homomorphism $V_{X/S}$ is functorial in (Σ'_1, S, X).

Proof. Since the morphism $S \to \Sigma'_1$ induces an isomorphism of sites $\mathrm{Cris}(S/\Sigma) \xrightarrow{\sim} \mathrm{Cris}(S/\Sigma')$ [5, 1.1.13], we may replace Σ by Σ'. Let us choose a covering S_α of S and embeddings $S_\alpha \hookrightarrow Y_\alpha$, where Y_α is a smooth formal scheme over Σ'. Let T_α be the completed divided power neighborhood of S_α in Y_α, and $X_\alpha = X \times_S S_\alpha$. The smoothness of Y_α and the universal property of the divided power neighborhoods imply that, for any $(U, T', \delta) \in \mathrm{Cris}(S/\Sigma')$, there exists, locally on T', a Σ'-morphism $h_\alpha\colon T' \to T_\alpha$. If $X_U =: f^{-1}(U)$, let us observe first that there exists a base changing isomorphism

$$\mathbb{L}\, h_\alpha^* \,\mathbb{R} f_{X_\alpha/T_\alpha *}\, \mathcal{O}_{X_\alpha/T_\alpha} \xrightarrow{\sim} \mathbb{R} f_{X_U/T' *}\, \mathcal{O}_{X_U/T'}.$$

Indeed, h_α factors through the reduction $T_{\alpha,n}$ of T_α modulo p^n for some n, so that by the usual base changing theorem [3, (7.8)] we may assume that $T' = T_{\alpha,n}$, and the claim follows from [3, (7.24.2)]. (Note that T_α is not noetherian in general, so that the reference is slightly abusive; but one sees easily that, since are using the p-adic topology on $\mathcal{O}_{T_\alpha}$, all results from [3, (7.18)-(7.24)] remain valid except (7.24.3), which we shall not need.) On the other hand, there exists for any sheaf E on $\mathrm{Cris}(X/\Sigma')$ a canonical isomorphism

$$\mathbb{R}f_{\mathrm{cris}*}(E)_{(U,T',\delta)} \xrightarrow{\sim} \mathbb{R}f_{X_U/T'*}(E|_{(X_U/T')_{\mathrm{cris}}});$$

in particular, the value on T' of the morphism $F^*_{X/S}$ can be identified with the pull-back by h_α of the morphism

$$F_{X\alpha/T_\alpha}\colon \mathbb{R}f'_{X'_\alpha/T_\alpha*}\,\mathcal{O}_{X'_\alpha/T_\alpha} \to \mathbb{R}f_{X_\alpha/T_\alpha*}\,\mathcal{O}_{X_\alpha/T_\alpha}.$$

Now since $\mathcal{O}_{T_\alpha}$ is torsion free, Theorem (1.6) defines a morphism

$$V_{X_\alpha/T_\alpha}\colon \mathbb{R}f_{X_\alpha/T_\alpha*}\,\mathcal{O}_{X_\alpha/T_\alpha} \to \mathbb{R}f'_{X'_\alpha/T_\alpha*}\,\mathcal{O}_{X'_\alpha}.$$

We can take its pull-back by h_α, and thus we obtain by the above discussion a morphism

$$(\mathbb{R}f_{\mathrm{cris}*}\,\mathcal{O}_{X/\Sigma})_{(U,T,\delta)} \to (\mathbb{R}f'_{\mathrm{cris}*}\,\mathcal{O}_{X'/\Sigma})_{(U,T,\delta)}.$$

This morphism does not depend locally upon the choice of h_α or on the embedding $S_\alpha \hookrightarrow Y_\alpha$: if $x \in U$ belongs to $S_\alpha \cap S_\beta$, and if (Y_β, h_β) is another choice in a neighborhood of x, let $S_\alpha \cap S_\beta \to Y_\alpha \times_{\Sigma'} Y_\beta$ be the diagonal embedding, $T_{\alpha\beta}$ the corresponding completed divided power neighborhood, $h\colon T \to T_{\alpha\beta}$ a Σ'-PD-morphism extending $U \cap S_\alpha \cap S_\beta \to Y_\alpha \times_{\Sigma'} Y_\beta$. Thanks to (1.8), $\mathcal{O}_{T_{\alpha\beta}}$ is p-torsion free, and $V_{X_{\alpha\beta}/T_{\alpha\beta}}$ is defined. The functoriality assertion in (1.6) implies that

$$h^*_\alpha(V_{X_\alpha/T_\alpha}) = h^*_{\alpha\beta}(V_{X_{\alpha\beta}/T_{\alpha\beta}}) = h^*_\beta(V_{X_\beta/T_\beta}).$$

Thus we can glue the various maps $h^*_\alpha(V_{X_\alpha/T_\alpha})$ on the cohomology sheaves and obtain $V_{X/S}\colon R^i f_{\mathrm{cris}*}\mathcal{O}_{X/\Sigma} \to R^i f_{\mathrm{cris}*}\mathcal{O}_{X'/\Sigma}$. Its functoriality results from the functoriality of $V_{X/T}$ in (1.6). $\square$

(1.10) *Remarks.* (i) The hypothesis of (1.9) is satisfied if S is a complete intersection over Σ' [5, 2.3.3]. On the other hand, we shall show in an appendix that the divided power envelopes have some p-torsion in general, even for singularities as simple as $\mathrm{Spec}(k[X,Y]/(X^2, XY, Y^2))$.

(ii) If $S \hookrightarrow T$ is a PD-thickening such that p is nilpotent, and $u\colon X \to Y$ a finite, locally free morphism between two smooth S-schemes $f\colon X \to S$, $g\colon Y \to S$, one can define a trace morphism

$$u_*\colon \mathbb{R}f_{X/T*}\mathcal{O}_{X/T} \to \mathbb{R}\,g_{Y/T*}\mathcal{O}_{Y/T}$$

such that $u_* \circ u^*$ is multiplication by $\deg(u)$. In the case where $u = F_{X/S}$, this should give another construction of $V_{X/S}$, without any hypothesis on S. However, it is not known in general whether $F^*_{X/S} \circ F_{X/S*}$ is multiplication by $p^{\dim(X)}$.

Another method for constructing $V_{X/S}$ should be given by the formalism of the relative de Rham-Witt complex [Illusie, correspondance], but the construction of the relative de Rham-Witt complex seems itself to require the same hypothesis on the absence of torsion.

(iii) If $f: X \to S$ is smooth and each $R^i f_{\text{cris}*} \mathcal{O}_{X/\Sigma}$ is flat, then it is a crystal on S, and its formation commutes with base change. The relative Frobenius map then induces an F-crystal structure:

$$\Phi_{X/S}: F_S^* R^i f_{\text{cris}*} \mathcal{O}_{X/\Sigma} \to R^i f_{\text{cris}*} \mathcal{O}_{X/\Sigma}.$$

Under the rather restrictive hypotheses of (1.9), we can conclude that there exists a

$$V_{X/S}: R^i f_{\text{cris}*} \mathcal{O}_{X/\Sigma} \to F_S^* R^i f_{\text{cris}*} \mathcal{O}_{X/\Sigma}$$

such that the compositions $V_{X/S} \circ \Phi_{X/S}$ and $\Phi_{X/S} \circ V_{X/S}$ are equal to multiplication by p^i. (Such F-crystals are called "nondegenerate" in [21].) We should also remark that satisfying finiteness results for the values of $R^i f_{\text{cris}*} \mathcal{O}_{X/\Sigma}$ on non-noetherian objects of $\text{Cris}(S/\Sigma)$ are not known, without further restrictive hypotheses (e.g. liftability of X or smoothness of S). □

2. Infinitesimal Deformations

Let us consider the same situation as in (1.3): we denote by $f: X \to S$ a smooth morphism of noetherian schemes in characteristic p, and by $S \hookrightarrow T$ a PD-immersion of S in a formal scheme T such that $\mathcal{O}_T$ is p-torsion free. Let $S_0 \hookrightarrow S$ be a closed subscheme defined by a nilpotent ideal, $X_0 = X \times_S S_0$. We begin by applying (1.3) to show that the crystalline cohomology of X/T can be made "functorial up to isogeny" in X_0; in particular, it depends only, up to isogeny, upon X_0. An important consequence is the comparison theorem between the crystalline cohomology of a smooth scheme over a perfect field and the de Rham cohomology of a formal lifting over an arbitrarily ramified extension of the corresponding ring of Witt vectors.

(2.1) **Theorem.** *Under the previous hypothesis, let $f: X \to S$, $g: Y \to S$ be two smooth morphisms of noetherian schemes, $u: X_0 \to Y_0$ an S-morphism. There exist morphisms*

$$u^*: \mathbb{R} g_{Y/T*}(\mathcal{O}_{Y/T}) \otimes^{\mathbb{L}} \mathbb{Q} \to \mathbb{R} f_{X/T*}(\mathcal{O}_{X/T}) \otimes^{\mathbb{L}} \mathbb{Q}, \tag{2.1.1}$$

$$u^*: \mathbb{R}\Gamma_{\text{cris}}(Y/T) \otimes^{\mathbb{L}} \mathbb{Q} \to \mathbb{R}\Gamma_{\text{cris}}(X/T) \otimes^{\mathbb{L}} \mathbb{Q}, \tag{2.1.2}$$

which are canonical in the following sense: if $h: Z \to S$ is another smooth morphism of noetherian schemes, and $v: Y_0 \to Z_0$ an S_0-morphism, then

$$(v \circ u)^* = u^* \circ v^*. \tag{2.1.3}$$

If u is the reduction to S_0 of $\tilde{u}: X \to Y$, then

$$u^* = \tilde{u}^*. \tag{2.1.4}$$

For each i, we consequently obtain morphisms

$$u^*: R^i g_{Y/T*}(\mathcal{O}_{Y/T}) \otimes \mathbb{Q} \to R^i f_{X/T*}(\mathcal{O}_{X/T}) \otimes \mathbb{Q},$$
$$u^*: H^i_{\text{cris}}(Y/T) \otimes \mathbb{Q} \to H^i_{\text{cris}}(X/T) \otimes \mathbb{Q}.$$

Proof. As we have seen in the proof of (1.3), the noetherian hypothesis implies that (2.1.2) follows from (2.1.1) by taking global sections on T.

Let $X^{(1)}=X'$, $X^{(n)}=X^{(n-1)\prime}$, etc. It follows from (1.3) and induction that for each n we have a canonical isomorphism

(2.1.5) $$F_{X/S}^{(n)*}\colon \mathbb{R}f^{(n)}_{X^{(n)}/T*}(\mathcal{O}_{X^{(n)}/T})\otimes^{\mathbb{L}}\mathbb{Q}\xrightarrow{\sim}\mathbb{R}f_{X/T*}(\mathcal{O}_{X/T})\otimes^{\mathbb{L}}\mathbb{Q}$$

(canonical in the sense that it commutes with the functoriality morphisms induced by morphisms of S-schemes). Let $\mathcal{J}$ be the ideal of S_0 in $S, j\colon S_0\hookrightarrow S$. For large enough n, $\mathcal{J}^{p^n}=0$; hence F_S^n factors through a morphism $\rho^{(n)}\colon S\to S_0$, and the following diagram commutes:

$$\begin{tikzcd} S \arrow[r, "F_S^n"] & S \\ S_0 \arrow[u, "j"] \arrow[r, "F_{S_0}^n"'] & S_0 . \arrow[u, "j"'] \arrow[from=1-1, to=2-2, "\rho^{(n)}"] \end{tikzcd}$$

Pulling the morphism $u\colon X_0\to Y_0$ back through $\rho^{(n)}$, we get a morphism $\rho^{(n)*}(u)\colon X^{(n)}\to Y^{(n)}$. If $n'\geqq n$, $\rho^{(n')}=\rho^{(n)}\circ F_S^{n'-n}$, and hence $\rho^{(n')*}(u)=F_S^{n'-n*}(\rho^{(n)*}(u))$. This relation, together with the canonicity of $F_S^{n'-n*}$, implies that if we define u^* to be the composed morphism

$$\begin{tikzcd} \mathbb{R}g_{Y/T*}(\mathcal{O}_{Y/T})\otimes^{\mathbb{L}}\mathbb{Q} \arrow[r, "(F_{Y/S}^{(n)*})^{-1}", "\sim"'] \arrow[d, dashed, "u^*"'] & \mathbb{R}g^{(n)}_{Y^{(n)}/T*}(\mathcal{O}_{Y^{(n)}/T})\otimes^{\mathbb{L}}\mathbb{Q} \arrow[d, "\rho^{(n)*}(u)^*"] \\ \mathbb{R}f_{X/T*}(\mathcal{O}_{X/T})\otimes^{\mathbb{L}}\mathbb{Q} & \mathbb{R}f^{(n)}_{X^{(n)}/T*}(\mathcal{O}_{X^{(n)}/T})\otimes^{\mathbb{L}}\mathbb{Q}, \arrow[l, "F_{X/S}^{(n)*}"', "\sim"] \end{tikzcd}$$

u^* does not depend upon the choice of n.

Since $\rho^{(n)*}(v\circ u)=\rho^{(n)*}(v)\circ\rho^{(n)*}(u)$, the relation (2.1.3) follows from the functoriality of crystalline cohomology. The last assertion follows from the definition of u^* and the canonicity of $F_{X/S}^{(n)*}$. □

(2.2) **Corollary.** *Assume that X and X' are two deformations of X_0 over S. Then there exist canonical isomorphisms*

(2.2.1) $$\delta_{X,X'}\colon \mathbb{R}f_{X'/T*}(\mathcal{O}_{X'/T})\otimes^{\mathbb{L}}\mathbb{Q}\xrightarrow{\sim}\mathbb{R}f_{X/T*}(\mathcal{O}_{X/T})\otimes^{\mathbb{L}}\mathbb{Q},$$

(2.2.2) $$\delta_{X,X'}\colon \mathbb{R}\Gamma_{\mathrm{cris}}(X'/T)\otimes^{\mathbb{L}}\mathbb{Q}\xrightarrow{\sim}\mathbb{R}\Gamma_{\mathrm{cris}}(X/T)\otimes^{\mathbb{L}}\mathbb{Q},$$

such that:

(i) *if X'' is a third deformation of X_0 over S, then*

(2.2.3) $$\delta_{X,X''}=\delta_{X,X'}\circ\delta_{X',X''};$$

(ii) *if Y, Y' are two deformations of a smooth S_0-scheme Y_0, and $u\colon X_0\to Y_0$ an S_0-morphism, then*

(2.2.4) $$\delta_{X,X'}\circ u^*=u^*\circ\delta_{Y,Y'}.$$

Proof. The existence of $\delta_{X,X'}$ and the relation (2.2.3) follow from (2.1) applied to $u=v=\mathrm{Id}_{X_0}$. The relation (2.2.4) results from

$$\delta_{X,X'} \circ u^* = \mathrm{Id}_{X_0}^* \circ u^* = (u \circ \mathrm{Id}_{X_0})^* = (\mathrm{Id}_{Y_0} \circ u)^* = u^* \circ \delta_{Y,Y'}. \quad \square$$

(2.3) *Remarks.* (i) It is possible to extend the construction of u^* to the case where S, T vary. Let us consider a commutative diagram

$$\begin{tikzcd} S'_0 \arrow[r, hook] \arrow[d] & S' \arrow[r, hook] \arrow[d] & T' \arrow[d, "w"] \\ S_0 \arrow[r, hook] & S \arrow[r, hook] & T, \end{tikzcd}$$

where $S'_0 \hookrightarrow S' \hookrightarrow T'$ satisfy the same hypothesis as $S_0 \hookrightarrow S \hookrightarrow T$, and w is a PD-morphism. Let $f\colon X \to S'$, $g\colon Y \to S$ be two smooth morphisms, $X_0 =: X \times_{S'} S'_0$, $Y_0 =: Y_S S_0$, $Y' =: Y \times_S S'$, $Y'_0 =: Y_0 \times_{S_0} S'_0 \simeq Y' \times_{S'} S'_0$, $g'\colon Y' \to S'$. There exists a functiorality morphism

$$\mathbb{L} w^* \mathbb{R} g_{Y/T*}(\mathcal{O}_{Y/T}) \otimes^{\mathbb{L}} \mathbb{Q} \to \mathbb{R} g'_{Y'/T'*}(\mathcal{O}_{Y'/T'}) \otimes^{\mathbb{L}} \mathbb{Q}. \tag{2.3.1}$$

If $u\colon X_0 \to Y_0$ is a morphism over $S'_0 \to S_0$, it admits a factorization $u'\colon X_0 \to Y'_0$. We can then define

$$u^*\colon \mathbb{L} w^* \mathbb{R} g_{Y/T*}(\mathcal{O}_{Y/T}) \otimes^{\mathbb{L}} \mathbb{Q} \to \mathbb{R} f_{X/T'*}(\mathcal{O}_{X/T'}) \otimes^{\mathbb{L}} \mathbb{Q}$$

to be the composite of u'^* and (2.3.1), and it is easy to check that the properties of (2.1) and (2.2) remain valid.

(ii) At the cost of some notational complexity we can eliminate X and Y from Theorem (2.1) altogether in the following way. If $S_0 \hookrightarrow S \subseteq T$ are as in the situation of (2.1) and if $f_0\colon X_0 \to S_0$ is smooth, then for each $n \gg 0$ we get an object $D^{(n)}(X_0/T)$ in the derived category of $\mathcal{O}_T \otimes \mathbb{Q}$-modules. Each $D^{(n)}$ is functorial in X_0, and if $n' \geqq n$ there is a natural isomorphism: $D^{(n')}(X_0/T) \to D^{(n)}(X_0/T)$; moreover if $f\colon X \to S$ is a smooth lifting of $f_0\colon X_0 \to S_0$, then there is a natural isomorphism

$$D^{(n)}(X_0/T) \to \mathbb{R} f_{X/T*}(\mathcal{O}_{X/T}) \otimes^{\mathbb{L}} \mathbb{Q}, \quad \text{for every } n \gg 0,$$

compatible with the "transition" isomorphisms above. Indeed, to construct $D^{(n)}$, just note that if $n \gg 0$, we can use diagram (2.1.6) to define a smooth $f^{(n)}\colon X^{(n)} \to S$ by pulling back $f_0\colon X_0 \to S$ along $\rho^{(n)}\colon S \to S_0$; we then let $D^{(n)}(X_0/T) =: \mathbb{R} f^{(n)}_{X^{(n)}/T*}(\mathcal{O}_{X^{(n)}/T}) \otimes^{\mathbb{L}} \mathbb{Q}$. If $n' \geqq n$, the relative Frobenius map $X^{(n)} \to X^{(n')}$ induces the transition isomorphism: $D^{(n')} \to D^{(n)}$. If $f\colon X \to S$ is a smooth lifting of f_0, (2.1.6) shows that $X^{(n)}$ is just the Frobenius pull-back of X/S, and hence relative Frobenius also induces the isomorphism: $D^{(n)}(X_0/T) \to \mathbb{R} f_{X/T*}(\mathcal{O}_{X/T}) \otimes^{\mathbb{L}} \mathbb{Q}$. $\square$

(2.4) **Theorem.** *Let V be a complete discrete valuation ring with perfect residue field k of characteristic p and fraction field K of characteristic zero, and let $\mathbf{X}$ be a smooth formal V-scheme with special fiber X_0 over k. Then there are* *canonical isomorphisms:*

(2.4.1) $$\sigma_{\mathrm{cris}}\colon \mathbb{R}\Gamma(\mathbf{X}, \Omega^{\cdot}_{\mathbf{X}/V}) \otimes^{\mathbb{L}}_{V} K \xrightarrow{\sim} \mathbb{R}\Gamma_{\mathrm{cris}}(X_0/W) \otimes^{\mathbb{L}}_{W} K,$$

(2.4.2) $$\sigma_{\mathrm{cris}}\colon H^i_{\mathrm{DR}}(\mathbf{X}/V) \otimes_V K \xrightarrow{\sim} H^i_{\mathrm{cris}}(X_0/W) \otimes_W K.$$

Proof. The map $(W,(p),\gamma) \to (V,(p),\gamma)$ is of course a PD-morphism. Let $R =: V/(p)$; R has a unique structure of local artinian k-algebra, with residue field k. If $S =: \operatorname{Spec} R$, $T =: \operatorname{Spf} V$, $\Sigma =: \operatorname{Spf} W$, we can view T as a Σ-PD-thickening of S. Let $\Sigma_n =: \operatorname{Spec} W/p^n W$, $T_n =: \operatorname{Spec} V/p^n V$, $X =: \mathbf{X} \times_T S$, $\overline{X}_0 =: X_0 \times_{\Sigma_1} S$. The base changing theorem for crystalline cohomology [3, (7.8)] implies that there is a natural isomorphism

$$\mathbb{R}\Gamma_{\mathrm{cris}}(X_0/\Sigma_n) \otimes^{\mathbb{L}}_{W/p^nW} V/p^n V \xrightarrow{\sim} \mathbb{R}\Gamma_{\mathrm{cris}}(\overline{X}_0/T_n)$$

for all n. Since V is a flat W-module, we have an isomorphism of functors $\otimes^{\mathbb{L}}_{W/p^nW} V/p^n V \simeq \otimes^{\mathbb{L}}_{W} V$. Passing to the limit, we get an isomorphism:

$$\mathbb{R} \varprojlim (\mathbb{R}\Gamma_{\mathrm{cris}}(X_0/\Sigma_n) \otimes^{\mathbb{L}}_{W} V) \xrightarrow{\sim} \mathbb{R} \varprojlim \mathbb{R}\Gamma_{\mathrm{cris}}(\overline{X}_0/T_n).$$

Since V is a finitely generated free W-module, $\otimes^{\mathbb{L}}_{W} V$ commutes with $\mathbb{R} \varprojlim$ [3, (B2.3)], and so by [3, (7.22.2)] we obtain an isomorphism:

(2.4.3) $$\mathbb{R}\Gamma_{\mathrm{cris}}(X_0/\Sigma) \otimes^{\mathbb{L}}_{W} V \xrightarrow{\sim} \mathbb{R}\Gamma_{\mathrm{cris}}(\overline{X}_0/T).$$

On the other hand, X is a smooth S-scheme and $\mathbf{X}$ is a smooth formal lifting of X to the S-PD-thickening T of S. By [3, (7.4)] and another limit argument, we obtain a canonical isomorphism:

(2.4.4) $$\mathbb{R}\Gamma(\mathbf{X}, \Omega^{\cdot}_{\mathbf{X}/T}) \xrightarrow{\sim} \mathbb{R}\Gamma_{\mathrm{cris}}(X/T).$$

Moreover, X and $\overline{X}_0$ are two deformations of X_0, so that (2.2) gives us an isomorphism:

$$\mathbb{R}\Gamma_{\mathrm{cris}}(X/T) \otimes^{\mathbb{L}} \mathbb{Q} \xrightarrow{\sim} \mathbb{R}\Gamma_{\mathrm{cris}}(\overline{X}_0/T) \otimes^{\mathbb{L}} \mathbb{Q}.$$

Thus, if we combine this with the isomorphisms deduced from (2.4.3) and (2.4.4) by tensoring with $\mathbb{Q}$, we obtain the isomorphism of the theorem.

The word "canonical" in the statement means the following. If $h\colon \mathbf{X} \to \mathbf{Y}$ is a morphism of smooth V-schemes, there is a commutative diagram:

(2.4.5) $$\begin{array}{ccc} \mathbb{R}\Gamma(\mathbf{Y}, \Omega^{\cdot}_{\mathbf{Y}/V}) \otimes^{\mathbb{L}} K & \xrightarrow{\sim} & \mathbb{R}\Gamma_{\mathrm{cris}}(Y_0/W) \otimes^{\mathbb{L}} K \\ {\scriptstyle \mathbb{R}\Gamma_{\mathrm{DR}}(h)} \downarrow & & \downarrow {\scriptstyle \mathbb{R}\Gamma_{\mathrm{cris}}(h_0)} \\ \mathbb{R}\Gamma(\mathbf{X}, \Omega^{\cdot}_{\mathbf{X}/V}) \otimes^{\mathbb{L}} K & \xrightarrow{\sim} & \mathbb{R}\Gamma_{\mathrm{cris}}(X_0/W) \otimes^{\mathbb{L}} K. \end{array}$$

This follows from the canonicity of the maps used to construct σ_{cris}. □

Applying Grothendieck's comparison theorem, we get the following corollary:

(2.5) **Corollary.** *If* $\mathbf{X}$ *is a smooth proper* V*-scheme with special fiber* X_0, *there is a canonical isomorphism:*

$$\sigma_{\mathrm{cris}}\colon H^i_{\mathrm{DR}}(\mathbf{X}/V)\otimes_V K \xrightarrow{\sim} H^i_{\mathrm{cris}}(X_0/W)\otimes_W K. \quad \square$$

(2.6) *Remarks.* (i) Let $X'_n \subseteq \mathbf{X}$ be the closed subscheme defined by $(\pi)^n$. Then for $n \geqq e/(p-1)$, the ideal $(\pi)^n \subseteq V$ has a (unique) PD structure, and we can define the crystalline cohomology $H^*_{\mathrm{cris}}(X'_n/V)$ with respect to this PD structure. In fact, for $n' \geqq n \geqq e/(p-1)$, we have a commutative diagram of canonical isomorphisms

$$\begin{tikzcd} H^*_{\mathrm{cris}}(X'_{n'}/V) \arrow[d, "\wr"] \arrow[dr, "\sim"] & \\ H^*_{\mathrm{cris}}(X'_n/V) \arrow[r, "\sim"] & H^*_{\mathrm{DR}}(\mathbf{X}/V). \end{tikzcd}$$

Since $X'_e = X$. it follows from our construction of σ_{cris} that for $n \geqq e$, the following diagram commutes:

$$\begin{tikzcd} H^*_{\mathrm{cris}}(X'_n/V)\otimes K \arrow[r, "\sim"] \arrow[d, "\wr"] & H^*_{\mathrm{DR}}(\mathbf{X}/V)\otimes K \arrow[d, "\wr"] \\ H^*_{\mathrm{cris}}(X'_e/V)\otimes K \arrow[r, "\sim"] & H^*_{\mathrm{cris}}(X_0/W)\otimes K. \end{tikzcd}$$

(ii) Under the hypothesis of (2.4), let us assume that there exists a smooth formal W-scheme $\mathbf{Y}$ such that $\mathbf{X} \simeq \mathbf{Y} \otimes_W V$. There exist canonical isomorphisms:

$$\mathbb{R}\Gamma(\mathbf{Y}, \Omega^{\cdot}_{\mathbf{Y}/W}) \simeq \mathbb{R}\Gamma_{\mathrm{cris}}(X_0/W),$$
$$\mathbb{R}\Gamma(\mathbf{X}, \Omega^{\cdot}_{\mathbf{X}/V}) \simeq \mathbb{R}\Gamma(\mathbf{Y}, \Omega^{\cdot}_{\mathbf{Y}/W}) \otimes^{\mathbb{L}}_W V,$$

hence

(2.6.1) $$\mathbb{R}\Gamma(\mathbf{X}, \Omega^{\cdot}_{\mathbf{X}/V}) \simeq \mathbb{R}\Gamma_{\mathrm{cris}}(X_0/W) \otimes^{\mathbb{L}}_W V.$$

Then σ_{cris} is merely deduced from (2.6.1) by tensoring with $\mathbb{Q}$: indeed, with the notations of the proof of (2.4), there exists a canonical isomorphism $X \simeq \bar{X}_0$, which induces $\delta_{X,\bar{X}_0}$ by (2.1.4).

(2.7) **Proposition** (compatibility with base change). *Let* $V \to V'$ *be a local homomorphism of complete discrete valuation rings with residue fields* k, k', *let* $k \to k'$, $W(k) \to W(k')$, *and* $K \to K'$ *be the induced maps; let* $\mathbf{X}$ *be a smooth formal* V*-scheme and let* $\mathbf{X}' =: \mathbf{X} \times_{\mathrm{Spf}\, V} \mathrm{Spf}\, V'$. *Then the following diagram commutes:*

$$\begin{tikzcd} \mathbb{R}\Gamma(\mathbf{X}, \Omega^{\cdot}_{\mathbf{X}/V}) \otimes^{\mathbb{L}} K \arrow[r] \arrow[d, "\wr"] & \mathbb{R}\Gamma(\mathbf{X}', \Omega^{\cdot}_{\mathbf{X}'/V'}) \otimes^{\mathbb{L}} K' \arrow[d, "\wr"] \\ \mathbb{R}\Gamma_{\mathrm{cris}}(X_0/W(k)) \otimes^{\mathbb{L}} K \arrow[r] & \mathbb{R}\Gamma_{\mathrm{cris}}(X'_0/W(k')) \otimes^{\mathbb{L}} K'. \end{tikzcd}$$

Proof. As above, let $S =: \mathrm{Spec}(V/pV)$, $S' =: \mathrm{Spec}(V'/pV')$, $T =: \mathrm{Spf}\, V$, $T' =: \mathrm{Spf}\, V'$, $X =: \mathbf{X} \times_T S$, $X' =: \mathbf{X}' \times_{T'} S' \simeq X \times_S S'$. Applying (2.3) and (2.2.4) to the projection

$X' \to X$, we get a commutative diagram

$$\begin{array}{ccc}
\mathbb{R}\Gamma_{\mathrm{cris}}(X/T) \otimes^{\mathbb{L}} \mathbb{Q} & \longrightarrow & \mathbb{R}\Gamma_{\mathrm{cris}}(X'/T') \otimes^{\mathbb{L}} \mathbb{Q} \\
\downarrow \wr & & \downarrow \wr \\
\mathbb{R}\Gamma_{\mathrm{cris}}(\bar{X}_0/T) \otimes^{\mathbb{L}} \mathbb{Q} & \longrightarrow & \mathbb{R}\Gamma_{\mathrm{cris}}(\bar{X}'_0/T') \otimes^{\mathbb{L}} \mathbb{Q}.
\end{array}$$

Therefore, the proposition follows from the naturality of the base change isomorphism, and of the isomorphism $\mathbb{R}\Gamma_{\mathrm{cris}}(X/T) \simeq \mathbb{R}\Gamma(\mathbf{X}, \Omega^{\bullet}_{\mathbf{X}/T})$. $\square$

It is natural to ask how far the isomorphism σ_{cris} is from preserving the lattice structures. Our method, suitably refined, gives some information. Suppose $\mathbf{X}/V$ is a smooth proper V-scheme (or formal scheme), and use σ_{cris} to identify $H^i_{\mathrm{DR}}(\mathbf{X}/V) \otimes K$ with $H^i_{\mathrm{cris}}(X_0/W) \otimes K$. Then $L_{\mathrm{DR}} =: H^i_{\mathrm{DR}}(\mathbf{X}/V)/\text{torsion}$ and $L_{\mathrm{cris}} =: H^i_{\mathrm{cris}}(X_0/W) \otimes V/\text{torsion}$ become two V-lattices in the same K-vector space, and it makes sense to ask how far apart they are.

(2.8) **Theorem.** *With the above notations, let l be the smallest integer greater than or equal to $\log_p(e/p-1)$. Then:*

(2.8.1) *If h is the highest Hodge slope of X_0/k in degree i, $p^{lh} L_{\mathrm{cris}} \subseteq L_{\mathrm{DR}}$ and $p^{lh} L_{\mathrm{DR}} \subseteq L_{\mathrm{cris}}$.*

(2.8.2) *If t is the ordinate of the end point of the Hodge polygon of X_0/k in degree i, the lengths of $L_{\mathrm{DR}}/L_{\mathrm{DR}} \cap L_{\mathrm{cris}}$ and of $L_{\mathrm{cris}}/L_{\mathrm{DR}} \cap L_{\mathrm{cris}}$ are bounded by elt.*

Proof. Let n be the smallest integer greater than or equal to $e/p-1$. Then the ideal $(\pi^n) \subseteq V$ has divided powers (c.f. (3.9) below, for instance). Let

$$S =: \operatorname{Spec} V/(\pi^n), \qquad X =: \mathbf{X} \times_{\operatorname{Spec} V} S, \qquad X_0 =: \mathbf{X} \times_{\operatorname{Spec} V} \operatorname{Spec} k,$$

and $\bar{X}_0 =: X_0 \times_{\operatorname{Spec} k} S$. We have canonical isomorphisms:

$$H^i_{\mathrm{DR}}(\mathbf{X}/V) \simeq H^i_{\mathrm{cris}}(X/V), \qquad H^i_{\mathrm{cris}}(X_0/W) \otimes V \simeq H^i_{\mathrm{cris}}(\bar{X}_0/V),$$

just as before. Moreover, S is a scheme in characteristic p, and F^l_S factors through $\operatorname{Spec} k$, so we have a canonical isomorphism: $\bar{X}^{(l)}_0 \simeq X^{(l)}$; hence a commutative diagram:

$$\begin{array}{ccccc}
H^i_{\mathrm{DR}}(\mathbf{X}/V) \otimes K & \longleftarrow & H^i_{\mathrm{cris}}(X/V) & \xleftarrow{F^{(l)}_{X/V}} & H^i_{\mathrm{cris}}(X^{(l)}/V) \\
\sigma_{\mathrm{cris}} \downarrow \wr & & & & \downarrow \wr \\
H^i_{\mathrm{cris}}(X_0/W) \otimes K & \longleftarrow & H^i_{\mathrm{cris}}(\bar{X}_0/V) & \xleftarrow{F^{(l)}_{\bar{X}_0/V}} & H^i_{\mathrm{cris}}(\bar{X}^{(l)}_0/V).
\end{array}$$

Let L' denote the image of $H^i_{\mathrm{cris}}(X^{(l)}/V) \simeq H^i_{\mathrm{cris}}(\bar{X}^{(l)}_0/V)$ in

$$H^i_{\mathrm{DR}}(\mathbf{X}/V) \otimes K \simeq H^i_{\mathrm{cris}}(X_0/W) \otimes K,$$

so that $L' \subseteq L_{\mathrm{DR}} \cap L_{\mathrm{cris}}$. We shall prove that p^{lh} annihilates L_{cris}/L' and L_{DR}/L', and that the length of each of these V-modules is less than or equal to p^{elt}.

Clearly it suffices to prove these results for the cokernels of $F^{(l)}_{X_0/V}$ and $F^{(l)}_{X/V}$ respectively. The map $F^{(l)}_{X_0/V}$ is obtained from $F^{(l)}_{X/W}$ by tensoring with V, and hence the required estimates on its cokernel can easily be deduced from known results about Hodge polygons [3, (8.36)]. We can use the techniques of [3, §8] and the results of §1 to obtain similar estimates for $F^{(l)}_{X/V}$ (and to give a direct proof for $F^{(l)}_{X_0/V}$).

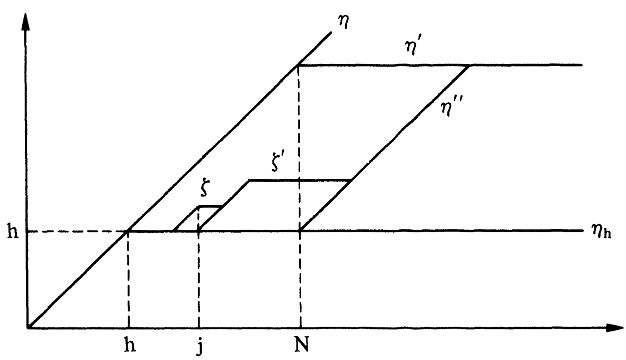
We shall prove that $p^h H^i_{\mathrm{cris}}(X/V)$ is contained in the image of $F_{X/V}$. (It is clear that we can then iterate this result and obtain the estimates necessary for (2.8.1).) We know by (1.7) that, in the derived category, $F_{X/V}$ factors through a quasi-isomorphism

$$\psi_{X/V}\colon \mathbb{R}u_{X'/V*}\mathcal{O}_{X'/V} \xrightarrow{\sim} \mathbb{L}\eta\,\mathbb{R}u_{X/V*}\mathcal{O}_{X/V},$$

where η is the cogauge: $\eta(i)=i$. This reduces us to a problem in the calculus of cogauges (c.f. [3, (8.22) ff.]). Let us abbreviate $\mathbb{R}u_{X/V*}\mathcal{O}_{X/V}$ by $K^\bullet$; let h be the constant cogauge with value h and $\eta_h =: \mathrm{Inf}(\eta, h)$. We have a commutative diagram:

$$\begin{array}{ccccc} & & \mathbb{L}hK^\bullet & \longleftarrow & K^\bullet \\ & & \downarrow & & \downarrow{\scriptstyle p^h} \\ \mathbb{L}\eta K^\bullet & \longrightarrow & \mathbb{L}\eta_h K^\bullet & \longrightarrow & K^\bullet. \end{array}$$

After passing to cohomology, we see that it will suffice to prove that the map $H^i(\mathbb{L}\eta K^\bullet)\to H^i(\mathbb{L}\eta_h K^\bullet)$ is surjective. It is clear that we may replace η and η_h by the gauges η' and η'' shown below, for $N \gg 0$, without changing anything. One can then pass from η'' to η' by a chain of simple augmentations. If ζ and ζ' lie between η'' and η' and if ζ is a simple augmentation of ζ' at j, then necessarily $j>h$.



We have a triangle [3, (8.23.6)]:

$$0\to\mathbb{L}\zeta\to\mathbb{L}\zeta'\to\mathscr{H}^j(K^\bullet\otimes\mathbb{F}_p)[-j]\to 0,$$

hence an exact sequence

$$\ldots\to H^i(\mathbb{L}\zeta)\to H^i(\mathbb{L}\zeta')\to H^{i-j}(\mathscr{H}^j(K^\bullet\otimes\mathbb{F}_p))\to\ldots.$$

Let $S' =: V/pV$ and Y' be the pull-back by $F_{S'}$ of $\mathbf{X} \times S'$. By the Cartier isomorphism, $\mathscr{H}^j(K^{\cdot} \otimes \mathbb{F}_p) \xrightarrow{\sim} \Omega^j_{Y'/S'}$, and so

$$H^{i-j}(\mathscr{H}^j(K^{\cdot} \otimes \mathbb{F}_p)) \xrightarrow{\sim} H^{i-j}(\Omega^j_{Y'/S'}).$$

Since $j > h$, our hypothesis implies that $H^{i-j}(X'_0, \Omega^j_{X'_0/k}) = 0$, and it is easy to see that $H^{i-j}(Y', \Omega^j_{Y'/S})$ also vanishes. This implies that each $H^i(\mathbb{L}\,\zeta) \to H^i(\mathbb{L}\,\zeta')$ is surjective, hence also $H^i(\mathbb{L}\,\eta) \to H^i(\mathbb{L}\,\eta_h)$ is surjective, and (2.8.1) follows.

We shall leave the details of the proof of (2.8.2) to the reader. First one checks that the length of $H^q(Y', \Omega^p_{Y'/S'})$ as a V-module is less than or equal to $e h^{p,q}$. Then one follows the argument of [3, (8.38)] in the present context, obtaining the estimate as claimed. □

(2.9) *Remark.* Suppose S is a formally smooth formal W-scheme with the p-adic topology and (H, ∇, F) is an F-crystal in the sense of Katz [13]. Then, as Katz explains, if θ_1 and θ_2 are two V-valued points of S which are congruent modulo the maximal ideal of V, the connection ∇ induces a canonical isomorphism $\varepsilon(\theta_2, \theta_1)\colon \theta_2^* H \otimes K \xrightarrow{\sim} \theta_1^* H \otimes K$. If $f\colon \mathbf{X} \to S$ is a smooth proper morphism of formal schemes and if the De Rahm cohomology sheaves $R^q f_* \Omega^{\cdot}_{\mathbf{X}/S}$ are locally free, then they inherit an F-crystal structure from crystalline cohomology, and we find ourselves in the above situation, with $H = R^q f_* \Omega^{\cdot}_{\mathbf{X}/S}$ and ∇ its Gauss-Manin connection. Suppose θ_1 factors through a W-valued point θ_0 of S, so that we have canonical isomorphisms: $\theta_0^* H \xrightarrow{\sim} H^q_{\mathrm{cris}}(X_0/W)$, $\theta_1^* H \xrightarrow{\sim} H^q_{\mathrm{cris}}(X_0/W) \otimes V$, where X_0 is the reduction of $\theta_i^* \mathbf{X}$ to k. Then if we combine $\varepsilon(\theta_2, \theta_1)$ with these isomorphisms, we construct a commutative diagram

$$\begin{array}{ccc} K \otimes \theta_2^* H & \xrightarrow[\sim]{\varepsilon(\theta_2,\theta_1)} & K \otimes \theta_1^* H \\ \uparrow \wr & & \downarrow \wr \\ K \otimes H^q_{\mathrm{DR}}(\theta_2^* \mathbf{X}/V) & \longrightarrow & K \otimes H^q_{\mathrm{cris}}(X_0/W). \end{array}$$

It is straightforward to check that the bottom arrow in the above square is exactly our isomorphism σ_{cris}. This was the method used by Deligne to contruct σ_{cris} in the special case of abelian varities.

(2.10) *Remark.* Using the above remark, it is easy to give an explicit example of σ_{cris} and hence to see that it really does not preserve the lattices L_{DR} and L_{cris}. Our example seems to show that the explicit bounds given in (2.8) are fairly sharp.

Let X_0/k be an ordinary elliptic curve, and let $\mathbf{X}/S$ be its universal formal deformation. From the theory of canonical coordinates [7] we know that there exist a basis $\{\eta, \omega\}$ for $H^1_{\mathrm{DR}}(\mathbf{X}/S)$ and a parameter t for S (so that $S \xrightarrow{\sim} \operatorname{Spf} W[[t]]$) such that $\nabla \eta = 0$ and $\nabla \omega = (dt/1+t) \otimes \eta$. If $S(1)$ denotes the p-adic completion of the divided power envelope of the diagonal of S and if $\xi =: 1 \otimes t - t \otimes 1$, then the isomorphism $\varepsilon\colon p_2^* H \xrightarrow{\sim} p_1^* H$ on $S(1)$ is given by the usual rule:

$$\varepsilon(p_2^*(x)) = \sum p_1^*(\nabla(d/dt)^n(x)) \otimes \xi^{[n]}.$$

If we define $\theta_1 \in S(V)$ by $t \mapsto 0$ and $\theta_2 \in S(V)$ by $t \mapsto \tau$ (with τ any element of the maximal ideal of V), we easily calculate:

$$\varepsilon(\theta_2^*(\eta)) = \theta_1^*(\eta), \qquad \varepsilon(\theta_2^*(\omega)) = \theta_1^*(\omega) + \log(1+\tau) \cdot \theta_1^*(\eta).$$

If τ is a uniformizer of V and e is the absolute ramification index, $\log(1+\tau)$ need not be integral as soon as $e > p$.

3. Obstructions and the Hodge Filtration

Let k be a perfect field, $W = W(k)$, V a finite totally ramified extension of W, with fraction field K; let $\mathbf{X}$ be a smooth proper formal V-scheme, and let X_0 be its closed fiber. In this section we investigate the relationship between the obstruction to extending data from X_0 to $\mathbf{X}$ and the isomorphism (2.4.2)

$$\sigma_{\mathrm{cris}}: H^*_{\mathrm{DR}}(\mathbf{X}/V) \otimes_V K \to H^*_{\mathrm{cris}}(X_0/W) \otimes_W K.$$

We begin by reviewing the theory of obstructions and Chern classes of line bundles [11, 7.4; 2; 7, 2.2].

(3.1) To remain in the setting of §1, let T be a formal scheme, or a scheme on which p is nilpotent, endowed with a PD ideal $(\mathscr{I}, \gamma)$, and let $f: X \to T$ be a T-scheme on which p is nilpotent and such that γ extends to X (e.g. X is an S-scheme, where $S = V(\mathscr{I})$). There is an exact sequences of sheaves in $(X/T)_{\mathrm{cris}}$:

(3.1.1) $$0 \to \mathscr{I}_{X/T} \to \mathscr{O}_{X/T} \to i_* \mathscr{O}_X \to 0.$$

where $i: X_{\mathrm{Zar}} \to (X/T)_{\mathrm{cris}}$ is the canonical "immersion" [3, (5.19)]. Since $\mathscr{I}_{X/T}$ is a sheaf of nilideals, (3.1.1) has a multiplicative analog

(3.1.2) $$0 \to 1 + \mathscr{I}_{X/T} \to \mathscr{O}^*_{X/T} \to i_* \mathscr{O}^*_X \to 0.$$

Let δ be the canonical PD structure on $\mathscr{I}_{X/T}$, and x a local section of $\mathscr{I}_{X/T}$. Since $\mathscr{I}_{X/T}$ is a p-torsion sheaf, the series

$$\log(1+x) =: \sum_{n=1}^{\infty} (-1)^{n-1} (n-1)! \, \delta_n(x)$$

is actually a finite sum, and defines a homomorphism of abelian sheaves

$$\log: 1 + \mathscr{I}_{X/T} \to \mathscr{I}_{X/T}.$$

Since i_* is exact, there exist canonical isomorphisms

$$R^1 f_* \mathscr{O}^*_X \simeq R^1 f_{X/T*} i_* \mathscr{O}^*_X, \qquad \mathrm{Pic}(X) \simeq H^1(X, \mathscr{O}^*_X) \simeq H^1(X/T, i_* \mathscr{O}^*_X),$$

and the coboundaries associated to (3.1.2) give homomorphisms

$$R^1 f_* \mathscr{O}^*_X \to R^2 f_{X/T*}(1 + \mathscr{I}_{X/T}), \qquad \mathrm{Pic}(X) \to H^2(X/T, 1 + \mathscr{I}_{X/T}),$$

Composing with log and the inclusion $\mathscr{I}_{X/T} \subset \mathcal{O}_{X/T}$, we obtain homomorphisms

$$c_{\mathrm{cris}}: R^1 f_* \mathcal{O}_X^* \to R^2 f_{X/T*} \mathcal{O}_{X/T}, \qquad c_{\mathrm{cris}}: \mathrm{Pic}(X) \to H^2_{\mathrm{cris}}(X/T)$$

which define the crystalline Chern class of a line bundle.

(3.2) Now suppose that $\mathbf{X}/T$ is a smooth formal scheme (resp. a smooth scheme if p is nilpotent on T); let $S =: V(\mathscr{I})$, and $X = \mathbf{X} \times_T S$. On $\mathbf{X}$ we have the multiplicative De Rham complex:

$$\Omega^\times_{\mathbf{X}/T}: \mathcal{O}^*_{\mathbf{X}} \xrightarrow{d\log} \Omega^1_{\mathbf{X}/T} \to \Omega^2_{\mathbf{X}/T} \to \dots,$$

where $d\log(\alpha) = d\alpha/\alpha$. Let $\mathscr{K}^\times_{\mathbf{X}/T}$ (resp. $\mathscr{I}^\bullet_{\mathbf{X}/T}$) denote the kernel of the obvious map of complexes $\Omega^\times_{\mathbf{X}/T} \to \mathcal{O}^*_X$ (resp. $\Omega^\bullet_{\mathbf{X}/T} \to \mathcal{O}_X$). Using log in degree zero and the identity maps in higher degrees, we obtain a morphism of complexes:

$$\log^\bullet: \mathscr{K}^\times_{\mathbf{X}/T} \to \mathscr{I}^\bullet_{\mathbf{X}/T}. \tag{3.2.1}$$

The short exact sequences:

$$0 \to \mathscr{K}^\times_{\mathbf{X}/T} \to \Omega^\times_{X/T} \to \mathcal{O}^*_X \to 0,$$
$$0 \to \mathscr{K}^0_{\mathbf{X}/T} \to \mathcal{O}^*_{\mathbf{X}} \to \mathcal{O}^*_X \to 0,$$

give coboundaries which fit in a commutative diagram (as well as a similar diagram for the cohomology):

$$\begin{CD}
R^1 f_* \mathcal{O}^*_X @>{\partial}>> \mathbb{R}^2 f_* \mathscr{K}^\times_{\mathbf{X}/T} @>{\log^\bullet}>> \mathbb{R}^2 f_* \mathscr{I}^\bullet_{\mathbf{X}/T} @>>> \mathbb{R}^2 f_* \Omega^\bullet_{\mathbf{X}/T} \\
@| @VVV @VVV @VVV \\
R^1 f_* \mathcal{O}^*_X @>{\partial}>> R^2 f_* \mathscr{K}^0_{\mathbf{X}/T} @>{\log}>> R^2 f_* \mathscr{I}\mathcal{O}_{\mathbf{X}} @>{i}>> R^2 f_* \mathcal{O}_{\mathbf{X}}.
\end{CD} \tag{3.2.2}$$

(3.3) **Lemma.** *Via the canonical isomorphism $R^2 f_{X/T*}\mathcal{O}_{X/T} \simeq \mathbb{R}^2 f_* \Omega^\bullet_{\mathbf{X}/T}$ (resp. $H^2_{\mathrm{cris}}(X/T) \simeq H^2_{\mathrm{DR}}(\mathbf{X}/T)$), the Chern class $c_{\mathrm{cris}}(L)$ of a line bundle L on X corresponds to the image of the class of L under the top horizontal arrow in* (3.2.2).

Proof. Let $L(\Omega^\bullet_{\mathbf{X}/T})$ be the complex on $\mathrm{Cris}(X/T)$ deduced from $\Omega^\bullet_{\mathbf{X}/T}$ by linearization; there exists a canonical homomorphism $\mathcal{O}_{X/T} \to L(\mathcal{O}_{\mathbf{X}})$, and $L(\Omega^\bullet_{\mathbf{X}/T})$ is a resolution of $\mathcal{O}_{X/T}$, by the Poincaré lemma [3, (6.12)]. There exists also a surjective homomorphism $L(\mathcal{O}_{\mathbf{X}}) \to i_* \mathcal{O}_X$; let $\mathscr{K}$ be its kernel, which is a PD-ideal in $L(\mathcal{O}_{\mathbf{X}})$ such that $(\mathcal{O}_{X/T}, \mathscr{I}_{X/T}) \to (\mathrm{L}(\mathcal{O}_{\mathbf{X}}), \mathscr{K})$ is a PD-morphism. Let $L(\mathcal{O}_{\mathbf{X}})^*$ be the abelian sheaf of invertible sections of $L(\mathcal{O}_{\mathbf{X}})$, and $L(\Omega_{\mathbf{X}})^\times$ the complex:

$$L(\mathcal{O}_{\mathbf{X}})^* \xrightarrow{d\log} L(\Omega^1_{\mathbf{X}/T}) \xrightarrow{L(d)} L(\Omega^2_{\mathbf{X}/T}) \to \dots;$$

let $\mathscr{K}^\bullet$, $\mathscr{K}^\times$ be the complexes:

$$\mathscr{K} \xrightarrow{L(d)} L(\Omega^1_{\mathbf{X}/T}) \xrightarrow{L(d)} L(\Omega^2_{\mathbf{X}/T}) \to \dots,$$

$$1 + \mathscr{K} \xrightarrow{d\log} L(\Omega^1_{\mathbf{X}/T}) \xrightarrow{L(d)} L(\Omega^2_{\mathbf{X}/T}) \to \dots.$$

Since $\mathcal{O}_{X/T} \to L(\mathcal{O}_{\mathbf{X}})$ is a PD-morphism, there exists a commutative diagram:

$$\begin{tikzcd}
1+\mathscr{I}_{X/T} \arrow[r, "\log"] \arrow[d] & \mathscr{I}_{X/T} \arrow[r] \arrow[d] & \mathcal{O}_{X/T} \arrow[d] \\
1+\mathscr{K} \arrow[r, "\log"] & \mathscr{K} \arrow[r] & L(\mathcal{O}_{\mathbf{X}}).
\end{tikzcd}$$

The commutative diagram of exact sequences of complexes

$$\begin{tikzcd}
0 \arrow[r] & 1+\mathscr{I}_{X/T} \arrow[r] \arrow[d] & \mathcal{O}^*_{X/T} \arrow[r] \arrow[d] & i_* \mathcal{O}^*_X \arrow[r] \arrow[d, equal] & 0 \\
0 \arrow[r] & \mathscr{K}^\times \arrow[r] & L(\Omega_{\mathbf{X}/T})^\times \arrow[r] & i_* \mathcal{O}^*_X \arrow[r] & 0
\end{tikzcd}$$

gives a commutative diagram:

$$\begin{tikzcd}
R^1 f_* \mathcal{O}^*_X \arrow[r, "\partial"] \arrow[d, equal] & R^2 f_{X/T*}(1+\mathscr{I}_{X/T}) \arrow[r, "\log"] \arrow[d] & R^2 f_{X/T*}\mathscr{I}_{X\,T} \arrow[r] \arrow[d, "\wr"] & R^2 f_{X/T*}\mathcal{O}_{X/T} \arrow[d, "\wr"] \\
R^1 f_* \mathcal{O}^*_X \arrow[r, "\partial"] & \mathbb{R}^2 f_{X/T*}\mathscr{K}^\times \arrow[r, "\log^\bullet"] & \mathbb{R}^2 f_{X/T*}\mathscr{K}^\bullet \arrow[r] & \mathbb{R}^2 f_{X/T*}L(\Omega^\bullet_{\mathbf{X}/T}),
\end{tikzcd} \tag{3.3.1}$$

where the isomorphisms follow from the Poincaré lemma, and the top horizontal arrow is c_{cris}. On the other hand, the sheaves $i_* \mathcal{O}^*_X$, $L(\Omega^i_{\mathbf{X}/T})$, $\mathscr{K}$ are acyclic for the projection $u_{X/T*}$ on the Zariski topos, and so are $L(\mathcal{O}_{\mathbf{X}})^*$, $1+\mathscr{K}$ (by the same argument as in [3, proof of (7.1)]. Their direct images are respectively $\mathcal{O}^*_X$, $\Omega^i_{\mathbf{X}/T}$, $\mathscr{I}\mathcal{O}_{\mathbf{X}}$, $\mathcal{O}^*_{\mathbf{X}}$, $\mathscr{K}^0_{\mathbf{X}/T}$, and since $f_{X/T} = f \circ u_{X/T}$ we get a commutative diagram

$$\begin{tikzcd}
R^1 f_* \mathcal{O}^*_X \arrow[r, "\partial"] \arrow[d, equal] & \mathbb{R}^2 f_{X/T*}\mathscr{K}^\times \arrow[r, "\log^\bullet"] \arrow[d, "\wr"] & \mathbb{R}^2 f_{X/T*}\mathscr{K}^\bullet \arrow[r] \arrow[d, "\wr"] & \mathbb{R}^2 f_{X/T*}L(\Omega^\bullet_{\mathbf{X}/T}) \arrow[d, "\wr"] \\
R^1 f_* \mathcal{O}^*_X \arrow[r, "\partial"] & \mathbb{R}^2 f_* \mathscr{K}^\times_{\mathbf{X}/T} \arrow[r, "\log^\bullet"] & \mathbb{R}^2 f_* \mathscr{I}^\bullet_{\mathbf{X}/T} \arrow[r] & \mathbb{R}^2 f_* \Omega^\bullet_{\mathbf{X}/T}
\end{tikzcd} \tag{3.3.2}$$

where the bottom horizontal arrow is the top one in (3.2.2). This proves the lemma. □

Since the Chern class in De Rham cohomology c_{DR} is defined by the map $d\log\colon R^1 f_* \mathcal{O}^*_{\mathbf{X}} \to R^2 f_* F^1 \Omega_{\mathbf{X}/T}$, where

$$^\cdot F^i \Omega^\bullet_{\mathbf{X}/T} = : 0 \to \Omega^i_{\mathbf{X}/T} \xrightarrow{d} \Omega^{i+1}_{\mathbf{X}/T} \to \dots$$

with $\Omega^i_{\mathbf{X}/T}$ in degree i, and since this map is equal to the coboundary defined by the exact sequence

$$0 \to F^1 \Omega^\bullet_{\mathbf{X}/T} \to \Omega^\times_{\mathbf{X}/T} \to \mathcal{O}^*_{\mathbf{X}} \to 0,$$

the lemma implies the following proposition:

(3.4) **Proposition** [2, Prop. 2.3]. *If $\mathscr{L} \in \mathrm{Pic}(\mathbf{X})$ lifts $L \in \mathrm{Pic}(X)$, then $c_{\mathrm{DR}}(\mathscr{L})$ and $c_{\mathrm{cris}}(L)$ correspond via the isomorphism*

$$\mathbb{R}^2 f_* \Omega^{\cdot}_{\mathbf{X}/T} \simeq R^2 f_{X/T_*} \mathcal{O}_{X/T} \qquad (\text{resp. } H^2_{\mathrm{DR}}(\mathbf{X}/T) \simeq H^2_{\mathrm{cris}}(X/T)).$$

In particular, $c_{\mathrm{cris}}(L)$ maps to zero in $R^2 f_* \mathcal{O}_{\mathbf{X}}$. □

We now check that the functoriality morphisms defined by (2.1) are compatible with Chern classes.

(3.5) **Proposition.** *Under the hypothesis of* (2.1), *let L be a line bundle on Y, L' a line bundle on X, L_0 and L'_0 their reductions on Y_0 and X_0, and let us assume that $L'_0 \simeq u^*(L_0)$. Then $u^*(c_{\mathrm{cris}}(L) \otimes 1) = c_{\mathrm{cris}}(L') \otimes 1$ in $\mathbb{R}^2 f_{X/T*}(\mathcal{O}_{X/T}) \otimes \mathbb{Q}$.*

Proof. Let n be such that F^n_S factors through $\rho^{(n)}: S \to S_0$. Then u^* is defined by the commutative square

$$\begin{tikzcd}
\mathbb{R} g_{Y/T*}(\mathcal{O}_{Y/T}) \overset{\mathbb{L}}{\otimes} \mathbb{Q} \arrow[r, "(F^{(n)*}_{Y/S})^{-1}", "\sim"'] \arrow[d, "u^*"'] & \mathbb{R}\, g^{(n)}_{Y^{(n)}/T*}(\mathcal{O}_{Y^{(n)}/T}) \overset{\mathbb{L}}{\otimes} \mathbb{Q} \arrow[d, "\rho^{(n)*}(u)^*"] \\
\mathbb{R} f_{X/T*}(\mathcal{O}_{X/T}) \overset{\mathbb{L}}{\otimes} \mathbb{Q} & \mathbb{R} f^{(n)}_{X^{(n)}/T*}(\mathcal{O}_{X^{(n)}/T}) \overset{\mathbb{L}}{\otimes} \mathbb{Q}. \arrow[l, "F^{(n)*}_{X/S}"']
\end{tikzcd}$$

Let $L^{(n)}$, $L'^{(n)}$ be the inverse images of L, L' through the projections $W^{(n)}_{Y/S}: Y^{(n)} \to Y$, $W^{(n)}_{X/S}: X^{(n)} \to X$; we have

$$F^{(n)*}_{Y/S}(L^{(n)}) = F^{(n)*}_{Y/S} W^{(n)*}_{Y/S}(L) \simeq L^{p^n},$$

and similarly for L'. On the other hand, the factorization of F^n_S through S_0 defines similar factorizations for $W^{(n)}_{Y/S}$ and $W^{(n)}_{X/S}$, and the isomorphism $L'_0 \simeq u^*(L_0)$ gives an isomorphism $L'^{(n)} \simeq \rho^{(n)*}(u)^*(L^{(n)})$. The usual functoriality of Chern classes gives therefore

$$\begin{aligned} p^n\, u^*(c_{\mathrm{cris}}(L) \otimes 1) &= F^{(n)*}_{X/S} \circ \rho^{(n)*}(u)^*(c_{\mathrm{cris}}(L^{(n)}) \otimes 1) \\ &= F^{(n)*}_{X/S}(c_{\mathrm{cris}}(L'^{(n)})) \\ &= p^n\, c_{\mathrm{cris}}(L'). \quad \square \end{aligned}$$

(3.6) **Corollary.** *Under the hypothesis of* (2.2), *let L and L' be two line bundles on X and X', such that their reductions L_0 and L'_0 on X_0 are isomorphic. Then*

$$\delta_{X,X'}(c_{\mathrm{cris}}(L') \otimes 1) = c_{\mathrm{cris}}(L) \otimes 1.$$

Proof. Apply (3.5) to $u = \mathrm{Id}_{X_0}$. □

(3.7) **Corollary.** *Under the hypothesis of* (2.4), *let $X \subset \mathbf{X}$ be the closed subscheme defined by the PD-ideal $(p) \subset V$. If $\mathscr{L}$ is a line bundle on $\mathbf{X}$ (resp. L on X) with reduction L_0 on X_0, the isomorphism*

$$\sigma_{\mathrm{cris}}: H^2_{\mathrm{DR}}(\mathbf{X}/V) \otimes_V K \xrightarrow{\sim} H^2_{\mathrm{cris}}(X_0/W) \otimes_W K$$

(*resp.* $\sigma'_{\text{cris}}: H^2_{\text{cris}}(X/V) \otimes_V K \simeq H^2_{\text{cris}}(X_0/W) \otimes_W K$)
takes $c_{\text{DR}}(\mathscr{L}) \otimes 1$ (*resp.* $c_{\text{cris}}(L) \otimes 1$) *to* $c_{\text{cris}}(L_0) \otimes 1$.

Proof. Since σ_{cris} is deduced from σ'_{cris} by composition with the canonical isomorphism $H^2_{\text{DR}}(\mathbf{X}/V) \otimes K \simeq H^2_{\text{cris}}(X/V) \otimes K$, the first assertion follows from the second thanks to (3.4), But σ'_{cris} is obtained by composition of (2.4.3) and (2.2.2) (for $X' = \bar{X}_0$), and the corollary follows from the functoriality of c_{cris} and from (3.6). □

(3.7.1) *Remark.* It is clear from the "splitting principle" that the results (3.4) to (3.7) extend to the Chern classes of vector bundles of arbitrary rank [2].

We now are ready for our first obstruction result. We keep the same hypothesis as in (2.4) (i.e., V is a complete discrete valuation ring of unequal characteristics, with perfect residue field k and fraction field K). We denote by $F^i H^*_{\text{DR}}(\mathbf{X}/V) =: \operatorname{Im}(H^*(\mathbf{X}, F^i \Omega^{\cdot}_{\mathbf{X}/V}) \to H^*_{\text{DR}}(\mathbf{X}/V))$ the Hodge filtration on De Rham cohomology. Recall that if $\mathbf{X}$ is a proper formal V-scheme, $H^2(\mathbf{X}, \mathcal{O}_{\mathbf{X}})$ is a finitely generated V-module, and hence its p-torsion submodule has finite length.

(3.8) **Theorem.** *If* $\mathbf{X}$ *is a smooth proper formal V-scheme, let t be a nonnegative integer such that the p-torsion in* $H^2(\mathbf{X}, \mathcal{O}_{\mathbf{X}})$ *is killed by* p^t. *Define l as follows:*

$l =$ *: the smallest integer greater than* $\log_p(e/(p-1))$ *if* $p > 2$;

$l =$ *: the smallest integer greater than or equal to* $\log_2(e) + 1$ *if* $p = 2$. *Then if* L_0 *is a line bundle on the special fiber* X_0 *of* $\mathbf{X}$, $L_0^{p^{(t+l)}}$ *lifts to a line bundle on* $\mathbf{X}$ iff $c_{\text{cris}}(L_0) \in H^2_{\text{cris}}(X_0/W)$ *corresponds, via* σ_{cris}, *to an element of* $F^1 H^2_{\text{DR}}(\mathbf{X}/V) \otimes K$.

Proof. Recall that if n is a positive integer and π is a uniformizing parameter for V, $\operatorname{ord}_\pi(\pi^n/n!) = \dfrac{p-1-e}{p-1} n + e \dfrac{\Sigma(n)}{p-1}$, where $\Sigma(n)$ is the sum of the p-adic digits of n. From this one can easily deduce:

(3.9) **Lemma.** *The ideal* $(\pi)^n \subseteq V$ *has divided powers iff* $n \geqq e/(p-1)$, *and the divided powers are* π*-adically nilpotent iff* $n > e/(p-1)$. □

Let n be the smallest integer greater than $e/(p-1)$. Let $X' \subseteq \mathbf{X}$ be the closed subscheme defined by $I =: (\pi)^n$.

(3.10) **Lemma.** *If* L_0 *is any line bundle on* X_0, $L_0^{p^l}$ *extends to a line bundle* L' *on* X'.

Proof. First suppose that $p > 2$. Then $p \in (\pi)^n$, so X' is a scheme of characteristic p. The closed subscheme $X_0 \subseteq X'$ is defined by the nilpotent ideal $(\pi)/(\pi^n) \otimes \mathcal{O}_{\mathbf{X}}$, which is generated by the image ε of $\pi \otimes 1$. Since p^l is an integer greater than $e/(p-1)$, $p^l \geqq n$, hence $(F_{X'})^l(\varepsilon) = 0$, and there is a map $\rho: X' \to X_0$ such that $\rho \circ \operatorname{inc} = (F_{X_0})^l$. Then $\rho^*(L_0)$ is a line bundle on X' extending $L_0^{p^l}$.

Now suppose $p = 2$. In this case the subscheme $X'' \subseteq \mathbf{X}$ defined by $(\pi)^e$ is in characteristic 2, and $2^{l-1} \geqq e$, so the above argument implies that $L_0^{2^{l-1}}$ extends to X''. The embedding $X'' \subseteq X'$ is defined by a square zero ideal I which is killed by multiplication by 2, so standard obstruction theory implies that the cokernel of $\operatorname{Pic}(X') \to \operatorname{Pic}(X'')$ is also killed by 2. Thus, $L_0^{2^l}$ extends to X'. □

Now suppose that $c_{\mathrm{cris}}(L_0)$ lies in $F^1 H^2_{\mathrm{DR}}(\mathbf{X}/V)\otimes K$, and let L' be a line bundle on X' lifting $L_0^{p^l}$. Then $c_{\mathrm{cris}}(L'_0)=p^l c_{\mathrm{cris}}(L_0)$, so it too lies in $F^1 H^2_{\mathrm{DR}}(\mathbf{X}/V)\otimes K$. We have a commutative diagram:

$$\begin{array}{ccc} H^*_{\mathrm{cris}}(X'/V) & \longrightarrow & H^*_{\mathrm{DR}}(\mathbf{X}/V) \\ \downarrow & & \downarrow \\ H^*_{\mathrm{cris}}(X_0/W)\otimes V\otimes K & \xleftarrow{\sigma_{\mathrm{cris}}} & H^*_{\mathrm{DR}}(\mathbf{X}/V)\otimes K. \end{array}$$

It follows that the image η of $c_{\mathrm{cris}}(L')$ via the map:

$$H^2_{\mathrm{cris}}(X'/V)\to H^2_{\mathrm{DR}}(\mathbf{X}/V)\to H^2(\mathbf{X},\mathcal{O}_{\mathbf{X}})$$

is torsion. By the commutativity of (3.2.2) we see that $i\circ\log\circ\partial(L')$ is torsion in $H^2(\mathbf{X},\mathcal{O}_{\mathbf{X}})$. The ideal $\mathscr{I}^0_{\mathbf{X}/T}=I\mathcal{O}_{\mathbf{X}}$ is principal, generated by π^n, and hence we can identify $H^2(\mathscr{I}^0_{\mathbf{X}/T})$ with $\mathcal{O}_{\mathbf{X}}$ and i with multiplication by π^n. Thus, $\log\circ\partial(L')\in H^2(\mathcal{O}_{\mathbf{X}})$ is torsion, hence killed by p^t. If $L''=:(L')^{p^t}$, $\log\circ\partial(L'')$ therefore maps to zero in $H^2(\mathscr{I}^0_{\mathbf{X}/T})$. Since $n>e/(p-1)$, the divided power structure γ on (π^n) is π-adically nilpotent, so the map $\log\colon \mathscr{K}^{\cdot}_{\mathbf{X}/T}\to\mathscr{I}^0_{\mathbf{X}/T}$ is an isomorphism, with inverse $\exp\colon x\mapsto \Sigma\,\gamma_n(x)$. Thus, $\partial(L'')=0$, and hence L'' lifts to an element of $H^1(\mathbf{X},\mathcal{O}^*_{\mathbf{X}})$. □

(3.11) **Corollary.** *With the assumptions of* (3.8), *the p-torsion part of the cokernel of the map* $\mathrm{Pic}(\mathbf{X})\to\mathrm{Pic}(X_0)$ *is killed by* p^{l+t}. □

By way of contrast, we offer a result showing that in some cases ramification is required to lift line bundles. Here we assume the residue field k is algebraically closed.

(3.12) **Proposition.** *Suppose* $\mathbf{X}/W$ *is smooth and proper with torsion free Hodge groups, and suppose that its Hodge to De Rham spectral sequence degenerates at* E_1. *Let* L_0 *be a line bundle on* X_0 *which is not a* p^{th} *power but whose Hodge Chern class in* $H^1(X_0,\Omega^1_{X_0/k})$ *vanishes, and suppose in addition that one of the following holds:*

a) X_0 *is a surface and* $\mathrm{ord}_p(L_0\cdot L_0)=1$.

b) $(F^{\cdot}H^2_{\mathrm{DR}}(\mathbf{X}/W),F^*_{X_0})$ *is strongly divisible, i.e.,* $F^*_{X_0}$ *acting on* $H^2_{\mathrm{cris}}(X_0/W)\cong H^2_{\mathrm{DR}}(\mathbf{X}/W)$ *is divisible by* p^2 *on* F^2_{Hodge}. (*Note: this is automatic if* $p\neq 2$.)

Then no non trivial power of L_0 *lifts to* $\mathbf{X}$.

Proof. Suppose on the contrary that L_0^n lifts to $\mathbf{X}$. Then $\sigma^{-1}_{\mathrm{cris}}[n\,c_{\mathrm{cris}}(L_0)]\in H^2_{\mathrm{cris}}(X_0/W)$ lies in $F^1_{\mathrm{Hodge}}H^2_{\mathrm{DR}}(\mathbf{X}/W)$, and since $H^2(\mathbf{X},\mathcal{O}_{\mathbf{X}})$ is torsion free, the same is true of $c_{\mathrm{cris}}(L_0)$. On the other hand, our assumption that the Hodge Chern class of L_0 vanishes implies that the image of $c_{\mathrm{cris}}(L_0)\bmod p$ lies in $F^2_{\mathrm{Hodge}}H^2_{\mathrm{DR}}(X_0/k)$, so $\sigma^{-1}_{\mathrm{cris}}[c_{\mathrm{cris}}(L_0)]=\xi+p\,\eta$, where $\xi\in F^2_{\mathrm{Hodge}}H^2_{\mathrm{DR}}(\mathbf{X}/W)$. Since $\sigma^{-1}_{\mathrm{cris}}[c_{\mathrm{cris}}(L_0)]\in F^1_{\mathrm{Hodge}}H^2_{\mathrm{DR}}(\mathbf{X}/W)$ and since $H^2(\mathbf{X},\mathcal{O}_{\mathbf{X}})$ is torsion free, $\eta\in F^1_{\mathrm{Hodge}}H^2_{\mathrm{DR}}(\mathbf{X}/W)$. Hence $(L_0\cdot L_0)=(\xi+p\,\eta\cdot\xi+p\,\eta)=p^2(\eta\cdot\eta)$, which contradicts a). Also $F^*_{X_0}c_{\mathrm{cris}}(L_0)=p\,c_{\mathrm{cris}}(L_0)$, so $F^*_{X_0}(\xi)+p\,F^*_{X_0}(\eta)=p(\xi+p\,\eta)$. Since $F^*_{X_0}(\xi)\in p^2H^2$ and $F^*_{X_0}(\eta)\in pH^2$, we see that

$p(\xi+p\eta)\in p^2H^2$, and hence $c_{\mathrm{cris}}(L_0)\in pH^2_{\mathrm{cris}}(X_0/W)$. By [18, (1.5)], this contradicts our assumption that L_0 is not a p^{th} power. $\square$

(3.12.1) *Remark.* Situations as in the above proposition do arise in nature. For example let E be a supersingular elliptic curve and let $X_0=E\times E$. Then $H^1_{\mathrm{cris}}(X_0/W)$ has a basis $\{\omega_1,\omega_2,\eta_1,\eta_2\}$ which is transformed by $F^*_{X_0}$ into $\{p\eta_1,p\eta_2,\omega_1,\omega_2\}$. The map $c_{\mathrm{cris}}\colon \mathrm{NS}(X_0)\otimes\mathbb{Z}_p\to H^2_{\mathrm{cris}}(X_0/W)(1)^{F^*_{X_0}}$ is an isomorphism, and hence there is an element l_0 in $\mathrm{NS}(X_0)\otimes\mathbb{Z}_p$ such that $c_{\mathrm{cris}}(l_0)=\omega_1\wedge\omega_2+p\eta_1\wedge\eta_2$. Let L_0 be an element in $\mathrm{NS}(X_0)$ with the same image in $\mathrm{NS}(X_0)\otimes\mathbb{F}_p$ as l_0. Then it is clear that $c_{\mathrm{cris}}(L_0)$ maps to the image of $\omega_1\wedge\omega_2$ in $F^2H^2_{\mathrm{DR}}(X_0/k)$. Moreover, since it is easy to see that every lifting X of X_0 satisfies the strong divisibility property b), we see that there is no lifting X of X_0 to W to which any power of L_0 extends. On the other hand, one can choose liftings of X_0 defined over a ramified extension of W to which L_0 lifts - c.f. [18, 2.3] and [16].

(3.12.2) *Remark.* Let us also point out that once a lifting $\mathbf{X}$ of X_0 is given over V, and if V' is an extension of V and a power of some $L_0\in\mathrm{Pic}(X_0)$ lifts to $\mathbf{X}\times_{\mathrm{Spec}\,V}\mathrm{Spec}\,V'$, then already a power of L_0 lifts to $\mathbf{X}$. This follows immediately from (3.8).

(3.13) We shall now give a result similar to (3.8) for homomorphisms of abelian schemes. We can prove it for p-divisible groups as well, provided we extend (2.4) to p-divisible groups as follows.

Let $S=\mathrm{Spec}(V/pV)$, $S_0=\mathrm{Spec}(k)$, $\Sigma=\mathrm{Spec}(W)$, $T=\mathrm{Spec}(V)$, $\Sigma_n=\mathrm{Spec}(W/p^nW)$, $T_n=\mathrm{Spec}(V/p^nV)$ (so that $S_0=\Sigma_1$, $S=T_1$). If $\mathbf{G}$ is a p-divisible group on V, with reductions G on S and G_0 and S_0, the Dieudonné crystal $\mathbb{D}(\mathbf{G})$ of $\mathbf{G}$ is by definition the Dieudonné crystal $\mathbb{D}(G)$ of G [15; 5, 3.3.6]. Let

$$D(G_0)=:\Gamma(S_0/\Sigma,\mathbb{D}(G_0))\simeq\varprojlim_n\mathbb{D}(G_0)_{(S_0,\Sigma_n)},$$

$$D(\mathbf{G})=:\Gamma(S/T,\mathbb{D}(\mathbf{G}))\simeq\varprojlim_n\mathbb{D}(G)_{(S,T_n)}.$$

(Recall that $D(G_0)$ is semi-linearly isomorphic to the usual Dieudonné module $M(G_0)$.)

(3.14) **Proposition.** *With the above notations, there exists a canonical isomorphism*

$$\sigma_{\mathrm{cris}}\colon D(\mathbf{G})\otimes_V K\xrightarrow{\sim}D(G_0)\otimes_W K.$$

Proof. Let $\bar{G}_0$ be the pull-back of G_0 through the natural projection $S\to S_0$. Because the Dieudonné crystal commutes with base change, we have a canonical isomorphism

$$D(\bar{G}_0)\simeq D(G_0)\otimes_W V.$$

On the other hand, if n is such that F^n_S factors through S_0, we get an isomorphism $\bar{G}_0^{(n)}\xrightarrow{\sim}G^{(n)}$. We can then define an isomorphism $D(G)\otimes K\xrightarrow{\sim}D(\bar{G}_0)\otimes K$ by the commutative square

$$\begin{array}{ccc} D(G)\otimes K & \xrightarrow[\sim]{F_G^{(n)-1}} & D(G^{(n)})\otimes K \\ \downarrow\wr & & \downarrow\wr \\ D(\bar{G}_0)\otimes K & \xleftarrow[\sim]{F_{G_0}^{(n)}} & D(G_0^{(n)})\otimes K, \end{array}$$

where $F_G^{(n)}$ and $F_{G_0}^{(n)}$ are isomorphisms because of the existence of the Verschiebung, and the proposition follows. □

(3.15) **Theorem.** *Let* **X** *and* **Y** *be formal abelian schemes (resp. p-divisible groups) over V, and $f_0: X_0 \to Y_0$ a morphism. Use σ_{cris} to transport the Hodge filtration on De Rham cohomology (resp. Dieudonné crystals) of* **X** *and* **Y** *to crystalline cohomology (resp. Dieudonné crystals) of X_0 and Y_0, and suppose that $H^1_{\text{cris}}(f_0)\otimes \mathrm{Id}_K$ (resp. $D(f_0)\otimes \mathrm{Id}_K$) preserves the Hodge filtrations. Then if l is defined as in* (3.8), *$p^l f_0$ lifts to a morphism* $\mathbf{X}\to\mathbf{Y}$.

Note. Messing has pointed out to us that this result can be obtained fairly directly from the methods of his thesis, and, except for the explicit bound on l, has been known for a long time.

Proof. The proofs for abelian schemes and p-divisible groups are identical. We choose to explain the case of abelian schemes, although it follows in fact from the case of p-divisible groups by the Serre-Tate theorem [22, § 5, Th. 4].

Recall that if X, Y are smooth schemes over a scheme S, with reductions X_0, Y_0 over a subscheme $S_0 \subset S$ defined by a square zero ideal $\mathscr{I}$, the obstruction to extending an S_0-morphism $f_0: X_0 \to Y_0$ to an S-morphism $f: X\to Y$ is an element of $\mathrm{Ext}^1_{\mathcal{O}_{X_0}}(f_0^*\Omega^1_{Y_0/S_0}, \mathscr{I}\mathcal{O}_{X_0}) \simeq H^1(X_0, \mathscr{I}\otimes f_0^*\Omega^{1\vee}_{Y_0/S_0})$. If X, Y are abelian schemes, this obstruction is additive in f_0: this results immediately from its functoriality properties and the fact that $\Omega^1_{Y_0/S_0}\simeq \omega_{Y_0}\otimes_{\mathcal{O}_{S_0}}\mathcal{O}_{Y_0}$, where ω_{Y_0} is the sheaf of translation invariant differential forms (on which the addition of Y_0 acts as the diagonal map).

Returning to our situation, let n be the smallest integer greater than $e/(p-1)$, and let $X'\subset\mathbf{X}$, $Y'\subset\mathbf{Y}$ be the closed subschemes defined by (π^n).

(3.16) **Lemma.** *If $f_0: X_0\to Y_0$ is any homomorphism, $p^l f_0$ lifts to a homomorphism $X'\to Y'$.*

Proof. If $p\neq 2$, let $S'' =: V(\pi^n)$, $X'' =: X'$, $Y'' =: Y'$, and $m =: l$. If $p=2$, let $S'' =: S$, $X'' =: X$, $Y'' =: Y$, and $m =: l-1$. It is enough to show in each case that $p^m f_0$ lifts to a map $X''\to Y''$ (since when $p=2$ the obstruction to lift $p^m f_0$ to a map $X'\to Y'$ is killed by p).

We have relative Frobenius morphisms $F^{(m)}_{X''/S''}: X''\to X''^{(m)}$, $F^{(m)}_{Y''/S''}: Y''\to Y''^{(m)}$; let $V^{(m)}_{Y''/S''}: Y''^{(m)}\to Y''$ be the iterated Verschiebung. Since the morphism $F^m_{S''}: S''\to S''$ factors through $\rho^{(m)}: S''\to S_0$, $f_0^{(m)}$ extends to a map $g: X''^{(m)}\to Y''^{(m)}$. Let $f'' = V^{(m)}_{Y''/S''}\circ g\circ F^{(m)}_{X''/S''}: X''\to Y''$. Then $f_0'' = V^{(m)}_{Y_0/S_0}\circ f_0^{(m)}\circ F^{(m)}_{X_0/S_0} = p^m f_0$. □

Now if $f': X'\to Y'$ lifts $p^l f_0$, we see that $H^1_{\text{cris}}(f')$: $H^1_{\text{cris}}(Y'/V)\to H^1_{\text{cris}}(X'/V)$ induces $p^l H^1_{\text{cris}}(f_0)$: $H^1_{\text{DR}}(\mathbf{Y}/V)\otimes\mathbb{Q}\to H^1_{\text{DR}}(\mathbf{X}/V)\otimes\mathbb{Q}$, and since $H^1(\mathbf{X},\mathcal{O}_{\mathbf{X}})$ is torsion free and $H^1_{\text{cris}}(f_0)$ preserves the Hodge filtration, so does $H^1_{\text{cris}}(f')$. We have chosen n large enough so that the divided power structure on $(\pi)^n$ is p-adically nilpotent. This implies that f' extends to a morphism $\mathbf{X} \to \mathbf{Y}$. To see this, it is enough to construct a compatible family of maps $f_i: X'_i \to Y'_i$ for all $i \geqq n$, where X'_i and Y'_i are defined by $(\pi)^i$, such that $f_n = f'$. But the ideal $(\pi^n)/(\pi^i) \subseteq V/(\pi^i) := V_i$ has *nilpotent* divided powers, and $H^1_{\mathrm{cris}}(f_n)$: $H^1_{\mathrm{cris}}(Y'_n/V_i) \to H^1_{\mathrm{cris}}(X'_n/V_i)$ preserves the Hodge filtrations defined by the liftings X'_i and Y'_i. Thus, the existence of f_i is guaranteed by the following result of Messing [15], which ends the proof of (3.15):

(3.17) **Theorem.** *If X and Y are abelian schemes (resp. p-divisible groups) defined over a nilpotent* PD *thickening $S \hookrightarrow T$ and if $f_0: X_0 \to Y_0$ is a homomorphism between their reductions to S such that $H^1_{\mathrm{cris}}(f_0)$ preserves the Hodge filtrations, then f_0 lifts uniquely to homomorphism $X \to Y$.*

Messing's proof is based on group theory, so it seems worthwhile in the case of abelian schemes to give a proof based on crystalline cohomology and deformation theory; this method has the advantage of being applicable to several other situations, c.f. (3.23). (The principles of this proof are also well known, but do not appear in the literature in any explicit form.) Prior to doing so, let us give an example showing that the value of l in (3.15) is the best possible; this example is an elaboration of a related one due to Serre [14, §1, e)].

(3.18) **Example** ("quasi canonical liftings"). Let X_0 be an ordinary elliptic curve over k, let X be its canonical lifting to W, and let $X^{(n)}$ be its pull back by $F^n_W: W \to W$; let $F^{(n)}: X \to X^{(n)}$ and $V^{(n)}: X^{(n)} \to X$ be the canonical liftings of the iterated Frobenius and Verschiebung. The group scheme $\mathrm{Ker}(p^n)$ on $X^{(n)}$ is canonically isomorphic to $\mathbb{Z}/p^n\mathbb{Z} \times \mu_{p^n}$, and $\mathbb{Z}/p^n\mathbb{Z} \times 0 = \mathrm{Ker}(V^{(n)})$, so $X \cong X^{(n)}/(\mathbb{Z}/p^n\mathbb{Z} \times 0)$.

Let V be a finite extension of W containing the $p^{n^{\mathrm{th}}}$ roots of unity with $e = (p-1)p^{n-1}$. If ζ is a primitive $p^{n^{\mathrm{th}}}$ root of unity, $\zeta \in \mu_{p^n}(V)$ defines a map of group schemes over V, $\zeta: \mathbb{Z}/p^n\mathbb{Z} \to \mu_{p^n}$, which specializes to the zero map over k. The graph of ζ is a subgroupscheme Γ of $\mathbb{Z}/p^n\mathbb{Z} \times \mu_{p^n} \cong_{p^n} X^{(n)}$, flat over V. Let $Y = X^{(n)}/\Gamma$. We have a diagram:

$$\begin{tikzcd} 0 \arrow[r] & (\mathbb{Z}/p^n\mathbb{Z}, 0) \arrow[r] & X^{(n)} \arrow[r, "V(n)"] \arrow[d, equal] & X \arrow[r] & 0 \\ 0 \arrow[r] & \Gamma \arrow[r] & X^{(n)} \arrow[r, "\mathrm{proj}"] & Y \arrow[r] & 0. \end{tikzcd}$$

Over k, Γ becomes $(\mathbb{Z}/p^n\mathbb{Z}, 0)$, so we can find an isomorphism $f_0: X_0 \to Y_0$ such that $f_0 \circ V_0^{(n)} = \mathrm{proj}_0$. Clearly $p^n f_0$ lifts to a map $X \to Y$. On the other hand, if $g: X \to Y$ lifted $p^{n-1} f_0$, then we would have: $(g \circ V^{(n)})_0 = p^{n-1} f_0 \circ V_0^{(n)} = p^{n-1}\,\mathrm{proj}_0$, so by rigidity $g \circ V^{(n)} = p^{n-1} \circ \mathrm{proj} = \mathrm{proj} \circ p^{n-1}$. Then $\mathrm{Ker}(V^{(n)})$ would be killed by $\mathrm{proj} \circ p^{n-1}$, and hence $p^{n-1}(\mathrm{Ker}\, V^{(n)}) = (p^{n-1})\mathbb{Z}/p^n\mathbb{Z} \times 0$ would be contained in Γ. But $((p^{n-1}), 0)$ is not in Γ, since ζ is a primitive $p^{n^{\mathrm{th}}}$ root of unity. $\square$

(3.19) Our proof of (3.17) will result from a relation between the obstruction to extending f_0 and its action on the Hodge filtrations of X and Y. Let us consider the following general situation. Suppose $S \hookrightarrow T$ is a PD immersion of affine schemes defined by the PD ideal (I, γ), and suppose X and Y are smooth T-schemes with reductions X_0 and Y_0 to S. If $f_0: X_0 \to Y_0$ is a morphism, it induces a map $H^*_{\mathrm{cris}}(f_0)$: $H^i_{\mathrm{cris}}(Y_0/T) \to H^i_{\mathrm{cris}}(X_0/T)$. Using the canonical isomorphism between crystalline and DeRham cohomology, we can view this as a map: $H^i_{\mathrm{cris}}(f_0)$: $H^i_{\mathrm{DR}}(Y/T) \to H^i_{\mathrm{DR}}(X/T)$. (Warning: this map *depends* on the PD structure γ.)

Let us assume that the Hodge spectral sequences of X/T and Y/T degenerate at E_1 and are locally free, so that the Hodge and DeRham cohomology sheaves commute with base change. Then the map $H^i_{\mathrm{cris}}(f_0) \otimes \mathrm{id}_{\mathcal{O}_S}$ can be identified with $H^i_{\mathrm{DR}}(f_0)$, and hence it preserves the Hodge filtrations. It follows that $H^i_{\mathrm{cris}}(f_0)$ induces a map:
$F^1_{\mathrm{Hodge}} H^i_{\mathrm{DR}}(Y/T) \to \mathrm{gr}^0_F H^i_{\mathrm{DR}}(X/T) \otimes I = H^i(X, \mathcal{O}_X) \otimes I$. Assume also that I is a square zero ideal: then this map factors through a map:

$$\rho(f_0, \gamma): F^1 H^i_{\mathrm{DR}}(Y_0/S) \to H^i(X_0, \mathcal{O}_{X_0}) \otimes I.$$

On the other hand, the obstruction $\mathrm{ob}(f_0)$ to extending f_0 to an S-morphism $X \to Y$ is an element of $\mathrm{Ext}^1_{\mathcal{O}_{X_0}}(f_0^* \Omega^1_{Y_0/S}, \mathcal{O}_{X_0} \otimes I)$, and hence defines by cup-product a map:

$$\mathrm{ob}(f_0) \cup : H^{i-1}(Y_0, \Omega^1_{Y_0/S}) \to H^i(X_0, \mathcal{O}_{X_0}) \otimes I.$$

Consider the following diagram:

$$\text{(3.19.1)} \qquad \begin{tikzcd} F^1 H^i_{\mathrm{DR}}(Y_0/S) \arrow[r, "\rho(f_0,\gamma)"] \arrow[d, "\mathrm{proj}"'] & H^i(X_0, \mathcal{O}_{X_0}) \otimes I \arrow[d, "\mathrm{proj}"] \\ H^{i-1}(Y_0, \Omega^1_{Y_0/S}) \arrow[r] \arrow[ur, "\mathrm{ob}(f_0)"] & H^i(X_0, \mathcal{O}_{X_0}) \otimes I/I^{[2]}. \end{tikzcd}$$

(3.20) **Proposition.** *After composition with* proj, *the above diagram becomes commutative. That is,* $\mathrm{proj} \circ \rho(f_0, \gamma) = \mathrm{proj} \circ \mathrm{ob}(f_0) \circ \mathrm{proj}$.

Proof. Recall that the morphism of ringed topoi $f_{0\,\mathrm{cris}}: (X_0/T)_{\mathrm{cris}} \to (Y_0/T)_{\mathrm{cris}}$ induces a map: $f^{-1}_{0\,\mathrm{cris}}(\mathcal{J}^{[k]}_{Y_0/T}) \to \mathcal{J}^{[k]}_{X_0/T}$ for all k, and that, thanks to [3, (7.2.1)] and our hypothesis, on the Hodge spectral sequence,
$H^i_{\mathrm{cris}}(X_0/T, \mathcal{J}^{[k]}_{X_0/T}) \cong \sum_{a+b=k} I^{[a]} F^b_X H^i_{\mathrm{DR}}(X/T)$, where $F^\cdot_X$ is the Hodge filtration of X/T. Moreover, the inclusion i: $X_0 \hookrightarrow X$ induces an exact functor $i_{\mathrm{cris}*}$, and $i_{\mathrm{cris}*} \mathcal{O}_{X_0/T} \cong \mathcal{O}_{X/T}$, so that we have an isomorphism:
$H^i_{\mathrm{cris}}(X/T, \mathcal{O}_{X/T}) \to H^i_{\mathrm{cris}}(X_0/T, \mathcal{O}_{X_0/T})$, and similarly for $Y_0 \subseteq Y$. Furthermore, there is an exact sequence: $0 \to \mathcal{J}_{X/T} \to i_{\mathrm{cris}*} \mathcal{J}_{X_0/T} \to i_{X/T*}(I \otimes \mathcal{O}_X) \to 0$, and the following diagram is commutative:

$$\begin{tikzcd} H^i(Y/T, \mathcal{J}_{Y/T}) \arrow[d] & F^1_Y H^i_{\mathrm{DR}}(Y/T) \arrow[l, "\cong"'] \arrow[r, "\rho(f_0,\gamma)"] & H^i(X, \mathcal{O}_X) \otimes I \arrow[d, "\wr"] \\ H^i(Y/T, i_* \mathcal{J}_{Y_0/T}) \arrow[r] \arrow[d] & H^i(X/T, i_* \mathcal{J}_{X_0/T}) \arrow[r] \arrow[d] & H^i(X, \mathcal{O}_X \otimes I) \arrow[d] \\ H^i(Y/T, \mathcal{O}_{Y/T}) \arrow[r, "f_0^*"] & H^i(X/T, \mathcal{O}_{X/T}) \arrow[r] & H^i(X, \mathcal{O}_X). \end{tikzcd}$$

This makes it clear that the map $\rho(f_0, \gamma)$ is induced by the following local map:

$$\begin{aligned}\mathbb{R}u_{Y/T*}\mathscr{I}_{Y/T} &\to \mathbb{R}u_{Y/T*}i_*\mathscr{I}_{Y_0/T} \cong \mathbb{R}u_{Y_0/T*}\mathscr{I}_{Y_0/T} \to \mathbb{R}f_{0*}\mathbb{R}u_{X_0/T*}\mathscr{I}_{X_0/T}\\ &\cong \mathbb{R}f_{0*}\mathbb{R}u_{X/T*}i_*\mathscr{I}_{X_0/T} \to \mathbb{R}f_{0*}\mathbb{R}u_{X/T*}i_{X/T*}(\mathscr{O}_X\otimes I) = \mathbb{R}f_{0*}(\mathscr{O}_X\otimes I).\end{aligned}$$

It is clear that there is a commutative diagram:

$$\begin{tikzcd}\mathscr{I}_{Y_0/T} \arrow[r] \arrow[d] & \mathscr{I}_{Y_0/T}/\mathscr{I}^{[2]}_{Y_0/T} \arrow[d]\\ f_{0\,\mathrm{cris}*}\mathscr{I}_{X_0/T} \arrow[r] & f_{0\,\mathrm{cris}*}\mathscr{I}_{X_0/T}/\mathscr{I}^{[2]}_{X_0/T},\end{tikzcd}$$

and that the map $i_{\mathrm{cris}*}\mathscr{I}_{X_0/T} \to i_{X/T*}(\mathscr{O}_X\otimes I)$ induces a map $i_{\mathrm{cris}*}\,\mathscr{I}_{X_0/T}/\mathscr{I}^{[2]}_{X_0/T} \to i_{X/T*}(\mathscr{O}_X\otimes I/I^{[2]})$. Thus we obtain a commutative diagram:

$$\begin{tikzcd}\mathbb{R}u_{Y/T*}(\mathscr{I}_{Y/T}) \arrow[r] \arrow[d] & \mathbb{R}f_{0*}(\mathbb{R}u_{X_0/T*}\mathscr{I}_{X_0/T}) \arrow[r] \arrow[d] & \mathbb{R}f_{0*}(\mathscr{O}_X\otimes I) \arrow[d]\\ \mathbb{R}u_{Y/T*}(\mathscr{I}_{Y/T}/\mathscr{I}^{[2]}_{Y/T}) \arrow[r] & \mathbb{R}f_{0*}(\mathbb{R}u_{X_0/T*}\mathscr{I}_{X_0/T}/\mathscr{I}^{[2]}_{X_0/T}) \arrow[r] & \mathbb{R}f_{0*}(\mathscr{O}_X\otimes I/I^{[2]}),\end{tikzcd}$$

where the top horizontal arrow induces $\rho(f_0, \gamma)$. It follows from the proof of the filtered Poincaré lemma [3, (6.13) and (7.2)] that $\mathbb{R}u_{Y/T*}(\mathscr{I}_{Y/T}/\mathscr{I}^{[2]}_{Y/T}) \cong \Omega^1_{Y/T}[-1]$, so we have a morphism (in the derived category):

$$\tilde{\rho}(f_0, \gamma)\colon f_0^*\,\Omega^1_{Y_0/T}[-1] \to \mathscr{O}_{X_0}\otimes I/I^{[2]}.$$

This "is" an element of $H^0\mathbb{R}\,\mathrm{Hom}(f_0^*\,\Omega^1_{Y_0/S_0}[-1],\ \mathscr{O}_{X_0}\otimes I/I^{[2]}) \cong \mathrm{Ext}^1_{\mathscr{O}_{X_0}}(f_0^*\,\Omega^1_{Y_0/S_0}, \mathscr{O}_{X_0}\otimes I/I^{[2]})$. Thus it is clear that (3.20) reduces to the following:

(3.21) **Theorem.** *$\tilde{\rho}(f_0, \gamma)$ is the image $\bar{\xi}$ of the obstruction class $\xi = \mathrm{ob}(f_0)$.*

Proof. First of all we need a convenient expression for the obstruction class ξ. Let $Z =: X \times_T Y$, and regard X_0 as a closed subscheme of Z via the map $(\mathrm{inc}, \mathrm{inc}\circ f_0)$. Then to lift the morphism f_0 is the same as to lift X_0 to a closed subscheme of Z, flat over T (for then the projection to X will automatically be an isomorphism).

Let $\mathscr{I}$ be the ideal of X_0 in Z and $\mathscr{I}_0$ the ideal of X_0 in Z_0. As explained in [19, (2.7)], there is an exact sequence of $\mathscr{O}_{X_0}$-modules:

$$(3.21.1) \qquad 0 \to I\mathscr{O}_{X_0} \to \mathscr{I}/\mathscr{I}^2 \to \mathscr{I}_0/\mathscr{I}_0^2 \to 0.$$

Moreover, there is a natural one-one correspondence between splittings of (3.21.1) and flat liftings of X_0 in Z. Thus, the image ξ of (3.21.1) in $\mathrm{Ext}^1_{\mathscr{O}_{X_0}}(\mathscr{I}_0/\mathscr{I}_0^2, I\mathscr{O}_X)$ "is" the obstruction to extending f_0. Of course, $\mathscr{I}_0/\mathscr{I}_0^2$ in naturally isomorphic to $f_0^*\,\Omega^1_{Y_0/S}$.

Now let D be the PD envelope of X_0 in $X \times Y$, and let $\bar{\mathscr{I}}$ be the ideal of X_0 in D. We have:

$$\mathbb{R}u_{X_0/T*}\,\mathscr{I}^{[k]}_{X_0/T} \cong F^k_{X_0}\,\Omega^{\cdot}_{D/T},$$

where $F^k_{X_0}\Omega^{\cdot}_{D/T}$ is the complex: $\bar{\mathscr{J}}^{[k]}\to\bar{\mathscr{J}}^{[k-1]}\Omega^1_{D/T}\to\dots$. In particular, the map $\pi_X: D\to X$ induces a natural quasi-isomorphism: $\mathrm{gr}^1_{F_{X_0}}\Omega^{\cdot}_{X/T}\xrightarrow{\sim}\mathrm{gr}^1_{F_{X_0}}\Omega^{\cdot}_{D/T}$. It is clear that $\tilde{\rho}(f\circ\gamma)$ is the composite:

$$f_0^*(\mathrm{gr}^1_{F_{Y_0}}\Omega^{\cdot}_{Y_0/S})\xrightarrow{\pi_Y^*}\mathrm{gr}^1_{F_{X_0}}\Omega^{\cdot}_{D/T}\xleftarrow[\sim]{\pi_X^*}\mathrm{gr}^1_{F_{X_0}}\Omega^{\cdot}_{X/T}\to\mathcal{O}_{X_0}\otimes I/I^{[2]}.$$

Now we have a commutative diagram:

$$\begin{tikzcd}
\xi\colon 0 \arrow[r] & \mathcal{O}_{X_0}\otimes I \arrow[r]\arrow[d] & \mathscr{J}/\mathscr{J}^2 \arrow[r]\arrow[d] & f_0^*\Omega^1_{Y_0/S} \arrow[r]\arrow[d] & 0 \\
\bar{\xi}\colon 0 \arrow[r] & \mathcal{O}_{X_0}\otimes I/I^{[2]} \arrow[r] & \bar{\mathscr{J}}/\bar{\mathscr{J}}^{[2]} \arrow[r] & f_0^*\Omega^1_{Y_0/S}. &
\end{tikzcd}$$

Let $A^{\cdot}$ be the complex: $\bar{\mathscr{J}}/\bar{\mathscr{J}}^{[2]}\to f_0^*\Omega^{\cdot}_{Y_0/S}$. Then we have a commutative diagram:

$$\begin{tikzcd}
\mathcal{O}_{X_0}\otimes I/I^{[2]} \arrow[r, "\sim"] & A^{\cdot} \\
 & f_0^*\Omega^{\cdot}_{Y_0/S}[-1]. \arrow[ul, dashed, "\bar{\xi}"] \arrow[u, dashed]
\end{tikzcd}$$

On the other hand, $\mathrm{gr}^1_{F_{X_0}}\Omega^{\cdot}_{D/T}$ is the complex: $\bar{\mathscr{J}}/\bar{\mathscr{J}}^{[2]}\to\Omega^{\cdot}_{X_0/S}\oplus f_0^*\Omega^1_{Y_0/S}$, which maps in an obvious way to $A^{\cdot}$. It is now easy to see that the following diagram commutes:

$$\begin{tikzcd}
f_0^*\Omega^1_{Y_0/S}[-1] \arrow[r] & \mathrm{gr}^1_{F_{X_0}}\Omega^{\cdot}_{D/T} \arrow[d] & \mathrm{gr}^1_{F_{X_0}}\Omega^{\cdot}_{X/T} \arrow[l, "\sim"'] \arrow[d] \\
 & A^{\cdot} & \mathcal{O}_{X_0}\otimes I/I^{[2]}. \arrow[l, "\sim"']
\end{tikzcd}$$

This implies that $\bar{\xi}=\tilde{\rho}(f_0,\gamma)$ in the derived category. □

(3.22) *Proof of* (3.17). Since the ideal of S in T is PD nilpotent, we can proceed inductively and assume that $I^{[2]}=0$. By rigidity of abelian schemes, the homomorphism lifting f_0 is unique, and we can therefore assume S affine. The homomorphism $H^1(Y_0,F^1\Omega^{\cdot}_{Y_0/S})\to H^0(Y_0,\Omega^1_{Y_0/S})$ is surjective, and the vanishing of $\rho(f_0,\gamma)$ implies that the cup product by $\mathrm{ob}(f_0)$: $H^0(Y_0,\Omega^1_{Y_0/S})\to H^1(X_0,\mathcal{O}_{X_0})\otimes I$ is zero. Since $\Omega^1_{Y_0/S}\simeq\omega_Y\otimes_{\mathcal{O}_S}\mathcal{O}_{Y_0}$, the theorem follows from the identification of the two groups $\mathrm{Ext}^1_{\mathcal{O}_{X_0}}(f_0^*\Omega^1_{Y_0/S},\mathcal{O}_{X_0}\otimes I)$ and

$$\mathrm{Hom}(H^0(Y_0,\Omega^1_{Y_0/S}),H^1(X_0,\mathcal{O}_{X_0})\otimes I)\quad\text{with}\quad\Gamma(S,\omega_Y)^\vee\otimes H^1(X_0,\mathcal{O}_{X_0})\otimes I.\quad□$$

(3.23) *Remark.* It is straightforward to use the techniques of (3.17) in other contexts when "local Torelli" holds. For example, if X and Y are $K3$ surfaces over a *nilpotent* PD thickening $S\subseteq T$ and $f_0: X_0\to Y_0$ is an isomorphism between their reductions to S such that $H^2_{\mathrm{cris}}(f_0)$ preserves the Hodge filtrations, then f_0 extends to an isomorphism $X \to Y$. This provides a proof of a key step in the proof of [18, (2.5)] that an automorphism α of a K3 surface over a field of characteristic $p>2$ acting trivially on its crystalline cohomology is necessarily the identity. (The references given there are rather vague and inconclusive.) We would like to take this opportunity to point out a subtlety in low characteristics. It is shown in [18, (2.3)] that if L is a polarization of X, there is a complete DVR V with $e(V) \leqq 2$ to which (X, L) can be lifted. Since $H^2_{\mathrm{cris}}(\alpha)$ automatically preserves the Hodge filtration of the lifting, we see that if the divided power structure on the maximal ideal of V is nilpotent, α will lift also, and hence be the identity by the well-known result in De Rham cohomology in characteristic zero. The necessary nilpotence will hold *unless* $p=3$ and $e=2$ – which means that X is a superspecial K3 surface in characteristic 3, by the results of [18, (2.3)]. Thus, the proof breaks down in this case. [Let us note that if Tate's conjecture is true, the Picard number of X is 22, and there will exist *another* polarization L' such that (X, L') lifts without ramification, so the same argument will again work.] However, we can argue as follows in this case. Let $V =: W(\sqrt{p})$, let $(\mathbf{X}, \mathscr{L})$ be a lifting of (X, L) to $T =: \operatorname{Spec} V$, and let $S =: \operatorname{Spec} k$, $S' =: \operatorname{Spec} V/pV$, $\mathbf{X}' = X \times_T S'$. Then $R =: V/pV \cong k[\varepsilon]/(\varepsilon^2)$, and the ideal (ε) can be endowed with the nilpotent PD structure δ, where $\delta_n(\varepsilon) = 0$ for all $n \geqq 2$. (This is *not* the divided power structure on (ε) compatible with the divided power structure on the maximal ideal of V.) There is an evident PD morphism $(W, \gamma) \to (R, \delta)$, and it follows from the base changing theorem for crystalline cohomology that we have a functorial isomorphism: $R \otimes H^2_{\mathrm{cris}}(X/W) \xrightarrow{\cong} H^2_{\mathrm{cris}}(X/(R, \delta))$. Since $H^2_{\mathrm{cris}}(\alpha/W)$ is the identity, the same is true of $H^2_{\mathrm{cris}}(\alpha/(R, \delta))$. Now the canonical isomorphism: $H^2_{\mathrm{cris}}(X/(R, \delta)) \cong H^2_{\mathrm{DR}}(X'/S')$ induces a Hodge filtration on $H^2_{\mathrm{cris}}(X/(R, \delta))$ which will be invariant under $\mathrm{id} = H^2_{\mathrm{cris}}(\alpha/(R, \delta))$, and hence the local Torelli theorem and (3.21) will imply that α lifts to an automorphism α' of X'/S'. We also have canonical isomorphisms: $H^2_{\mathrm{cris}}(X'/V) \xrightarrow{\cong} H^2_{\mathrm{cris}}(X/V) \xrightarrow{\cong} H^2_{\mathrm{cris}}(X/W) \otimes V$ under which the action of α' corresponds to that of α, and so it follows that α' acts trivially on $H^2_{\mathrm{cris}}(X'/V)$. But now $X' \subseteq \mathbf{X}$ is defined by the PD nilpotent ideal (3), and so we can conclude that α' lifts to $\mathbf{X}$. □

4. The Crystalline Weil Group

If X is a smooth proper V-scheme and X_k is its reduction $\bmod(\pi)$, the isomorphism $\sigma_{\mathrm{cris}}: H^*_{\mathrm{DR}}(X/V) \otimes \mathbb{Q} \to H^*_{\mathrm{cris}}(X_k/W) \otimes V \otimes \mathbb{Q}$ furnishes us with a $W \otimes \mathbb{Q}$-structure on the $V \otimes \mathbb{Q}$-vector space $H^*_{\mathrm{DR}}(X/V) \otimes \mathbb{Q}$. This structure, together with the action of Frobenius on crystalline cohomology, is in a sense analogous to the $\mathbb{Q}$-vector space structure on DeRham cohomology over the complex field furnished by singular cohomology. To keep track of these data, we follow an idea of Deligne by introducing the so-called "crystalline Weil group."

For the sake of simplicity, we shall restrict our attention to the following situation. Let k be a perfect field of characteristic $p>0$, let $W(k)$ be its Witt ring, and let $K(k)$ be the fraction field of $W(k)$. We denote by $\bar{K}(k)$ or just $\bar{K}$ an algebraic closure of $K(k)$; the valuation of $K(k)$ prolongs uniquely to a valuation of $\bar{K}(k)$, and the residue field $\bar{k}$ of $\bar{K}(k)$ is an algebraic closure of k. Let $K_{nr}(k)$ or just K_{nr} denote the maximal unramified extension of $K(k)$ in $\bar{K}(k)$; its residue field is also $\bar{k}$. If we start with some finite extension K of $K(k)$, we will choose $\bar{K}(k)$ to be an algebraic closure of K.

(4.1) **Definition.** *The "crystalline Weil group of $\bar{K}(k)$", denoted $W_{\mathrm{cris}}(\bar{K}(k))$, is the group of automorphisms of $\bar{K}(k)$ covering some integral power of the Frobenius automorphism F_{nr} of $K_{nr}(k)$.*

Note that if $\psi \in W_{\mathrm{cris}}(\bar{K}(k))$, there is a unique integer $\deg(\psi)$ such that ψ acts as $F_{nr}^{\deg(\psi)}$ on $K_{nr}(k)$. Since F_{nr} preserves the valuation of $K(k)$, ψ is automatically continuous, hence acts on $\bar{k}$; evidently its action on $\bar{k}$ is $F_{\bar{k}}^{\deg(\psi)}$, which also determines $\deg(\psi)$ uniquely. Since $\bar{K}$ is algebraically closed, the map $\deg\colon W_{\mathrm{cris}}(\bar{K}) \to \mathbb{Z}$ is surjective, and we have an exact sequence:

$$1 \to I_{\mathrm{cris}}(\bar{K}) \to W_{\mathrm{cris}}(\bar{K}) \xrightarrow{\deg} \mathbb{Z} \to 1, \tag{4.1.1}$$

where $I_{\mathrm{cris}}(\bar{K}) =: \mathrm{Ker}(\deg)$ is just the inertial Galois group $\mathrm{Gal}(\bar{K}/K_{nr})$. Thus, $\bar{K}^{I_{\mathrm{cris}}(\bar{K})} = K_{nr}$ and $K^{W_{\mathrm{cris}}(\bar{K})} = K_{nr}^{F_{nr}} = \mathbb{Q}_p$. We can also let $W_{\mathrm{cris}}(\bar{K})$ act by continuity on the completion $\hat{\bar{K}}$ of $\bar{K}$; it is well-known [24] that $\hat{\bar{K}}^{I_{\mathrm{cris}}(\bar{K})} = \hat{K}_{nr} = W(\bar{k}) \otimes \mathbb{Q}_p$, so again we get $\hat{\bar{K}}^{W_{\mathrm{cris}}(\bar{K})} = \mathbb{Q}_p$.

If $\bar{K}'$ is another algebraic closure of $K(k)$, any $K(k)$-isomorphism $\sigma\colon \bar{K} \to \bar{K}'$ is continuous and induces an isomorphism: $K_{nr} \to K'_{nr}$. Since $F'_{nr} \circ \sigma = \sigma \circ F_{nr}$, we see that $\psi \mapsto \sigma\psi\sigma^{-1}$ defines an isomorphism:

$$W_{\mathrm{cris}}(\sigma)\colon W_{\mathrm{cris}}(\bar{K}) \to W_{\mathrm{cris}}(\bar{K}')$$

compatible with deg. If $\tau\colon \bar{K}' \to \bar{K}''$ is another isomorphism,

$$W_{\mathrm{cris}}(\tau) \circ W_{\mathrm{cris}}(\sigma) = W_{\mathrm{cris}}(\tau\sigma).$$

If H is a $\bar{K}$-vector space, by a "semi-linear action of $W_{\mathrm{cris}}(\bar{K})$ on H" we mean a map $\rho\colon W_{\mathrm{cris}}(\bar{K}) \to \mathrm{Aut}_{\mathbb{Q}_p}(H)$ such that $\rho(\psi)(ax) = \psi(a)\rho(\psi(x))$ for $a \in \bar{K}$, $x \in H$, $\psi \in W_{\mathrm{cris}}(\bar{K})$. Let $\bar{K}(i)$ denote the $\bar{K}$-vector space $\bar{K}$ on which $W_{\mathrm{cris}}(\bar{K})$ acts by $\rho(\psi)(a) = p^{-\deg(\psi)i}\psi(a)$, and let $H(i) =: H \otimes_{\bar{K}} \bar{K}(i)$ with the tensor product action.

We shall say that a smooth proper $\bar{K}$-scheme X has "good reduction" iff there exist a finite extension K' of $K(k)$ in $\bar{K}$ with ring of integers V', a smooth proper V'-scheme $\mathbf{X}'$, and an isomorphism $\alpha\colon \mathbf{X}'_{\bar{K}} \xrightarrow{\sim} X$, where $\mathbf{X}'_{\bar{K}} =: \mathbf{X}' \times_{\operatorname{Spec} V'} \operatorname{Spec} \bar{K}$. In this situation let k' be the residue field of V' and $\mathbf{X}_{k'} =: \mathbf{X}' \times_{\operatorname{Spec} V'} \operatorname{Spec} k'$.

(4.2) **Theorem.** *Suppose X is a smooth proper $\bar{K}$-scheme with good reduction. There is a unique semi-linear action ρ_{cris} of $W_{\mathrm{cris}}(\bar{K})$ on $H^*_{\mathrm{DR}}(X/\bar{K})$ with the following properties.*

(4.2.1) *It is functorial in X. In fact, if $f\colon Y \circ\!\!\to X$ is a correspondence*[1] *of degree d, the map $H^*_{\mathrm{DR}}(f)\colon H^*_{\mathrm{DR}}(Y/\bar{K}) \to H^{*+2d}_{\mathrm{DR}}(X/\bar{K})(+d)$ is compatible with the actions of $W_{\mathrm{cris}}(\bar{K})$.*

[1] i.e., a closed subset of $Y \times X$ of codimension $d + \dim(Y)$. Recall that a morphism $f\colon X \to Y$ induces a correspondence $Y \circ\!\!\to X$ of degree zero, viz. the transpose of the graph of f.

(4.2.2) *If $\mathbf{X}'$ is any smooth proper V'-scheme, the image of the map:*

$$H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes K_{nr}\subseteq H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes K' \xrightarrow{\sigma^{-1}_{\mathrm{cris}}} H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K')\subseteq H^*_{\mathrm{DR}}(\mathbf{X}'_{\bar K}/\bar K)$$

consists precisely of the invariants of $I_{\mathrm{cris}}(\bar K)$, and the action of $W_{\mathrm{cris}}(\bar K)/I_{\mathrm{cris}}(\bar K)$ $\cong \mathbb{Z}$ on these invariants is given by the action of absolute Frobenius.

Proof. In order to obtain the unicity and functoriality asserted, we shall have to use some results of a forthcoming paper [10] of Messing and Gillet on cycle classes in crystalline cohomology tensored with $\mathbb{Q}$. For the reader's convenience, we shall state here a summary of what we need.

(4.3) **Theorem** (Messing-Gillet). *If X_0 and Y_0 are smooth proper k-schemes, a correspondence z_0: $Y_0 \circ\!\!\rightarrow X_0$ of degree d induces a morphism*

$$z_0^*.\ H^i_{\mathrm{cris}}(Y_0/W)\otimes\mathbb{Q}\rightarrow H^{i+2d}_{\mathrm{cris}}(X_0/W)\otimes\mathbb{Q},$$

compatible with composition of correspondances. If $\mathbf{X}$ and $\mathbf{Y}$ are smooth proper V-schemes with reductions X_0 and Y_0 and z: $\mathbf{Y}_K \circ\!\!\rightarrow \mathbf{X}_K$ is a correspondance of degree d with specialization z_0, we have a commutative diagram

$$\begin{array}{ccc}
H^i_{\mathrm{DR}}(z)\colon H^i_{\mathrm{DR}}(\mathbf{Y}_K/K) & \longrightarrow & H^{i+2d}_{\mathrm{DR}}(\mathbf{X}_K/K)\\
\wr\big\downarrow & & \wr\big\downarrow\\
H^i_{\mathrm{cris}}(z_0)\otimes\mathrm{Id}\colon H^i_{\mathrm{cris}}(Y_0/W)\otimes K & \longrightarrow & H^{i+2d}_{\mathrm{cris}}(X_0/W)\otimes K\\
\big\uparrow & & \big\uparrow\\
H^i_{\mathrm{cris}}(z_0)\colon H^i_{\mathrm{cris}}(Y_0/W)\otimes\mathbb{Q} & \longrightarrow & H^{i+2d}_{\mathrm{cris}}(X_0/W)\otimes\mathbb{Q}.\ \square
\end{array}$$

(4.4) We now begin the proof of (4.2). Choose a smooth proper $\mathbf{X}'$ over a suitable V' and an isomorphism $\alpha\colon \mathbf{X}'_{\bar K}\xrightarrow{\sim} X$. Using α^* and $\sigma_{\mathrm{cris}}\otimes\mathrm{id}_{\bar K}$ we get an isomorphism

$$\sigma_\alpha\colon H^*_{\mathrm{DR}}(X/\bar K)\xrightarrow{\sim} H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes\bar K.$$

Notice that if $K'\subseteq K''\subseteq\bar K$, with K'' also finite over $K(k)$, and if $\mathbf{X}''=:\mathbf{X}'\times_{\mathrm{Spec}\,V'}\mathrm{Spec}\,V''$, we get by (2.7) a commutative diagram

(4.4.1)
$$\begin{tikzcd}
H^*_{\mathrm{DR}}(X/\bar K) \arrow[r, "\sim"] \arrow[dr, "\sim"'] & H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes\bar K \arrow[d, "\wr"]\\
 & H^*_{\mathrm{cris}}(\mathbf{X}''_{k''}/W(k''))\otimes\bar K.
\end{tikzcd}$$

If now $\psi\in W_{\mathrm{cris}}(\bar K)$ has degree d, ψ covers the d^{th} power of the Frobenius automorphism of $W(k')$. The absolute Frobenius automorphism $F_{\mathbf{X}'_{k'}}$ induces a semi-linear automorphism Φ of $H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes K(k')$, and hence $\Phi^d\otimes\psi$ makes sense and forms a ψ-linear automorphism of $H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes\bar K$. Using the isomorphism σ_α, we can transfer this to an automorphism of $H^*_{\mathrm{DR}}(X/\bar K)$.

If we check that the action is functorial in $X/\bar K$, it will follow that it is independent of the choices. Notice first that (4.4.1) shows that it is independent of the choice of K', in the sense that we can replace K' by a finite extension K''. Now suppose that f: $Y \circ\!\!\rightarrow X$ is a correspondence over $\bar{K}$, where X and Y have good reduction. Suppose $\mathbf{X}'$ and $\mathbf{Y}'$ smooth and proper over V', V'', respectively, and that $\alpha: \mathbf{X}'_{\bar{K}} \xrightarrow{\sim} X$, $\beta: \mathbf{Y}'_{\bar{K}} \xrightarrow{\sim} Y$ are isomorphisms. By the previous remark, we can assume without loss of generality that $K' = K''$, and also that f is defined by a correspondence $f': \mathbf{Y}'_{K'} \circ\!\!\rightarrow \mathbf{X}'_{K'}$. Then we get by (4.3) a commutative diagram:

$$\begin{array}{ccc} H^*_{\mathrm{DR}}(\mathbf{Y}'_{K'}/K') & \longrightarrow & H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K') \\ \Big\downarrow{\scriptstyle \sigma_{\mathrm{cris}}} & & \Big\downarrow{\scriptstyle \sigma_{\mathrm{cris}}} \\ H^*_{\mathrm{cris}}(\mathbf{Y}'_{k'}/W(k')) \otimes K' & \longrightarrow & H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k')) \otimes K' \end{array}$$

which implies the functoriality we claimed. Property (4.4.2) is apparent from the definition, and the uniqueness is also clear. □

Perhaps it is worth remarking that if $\bar{K}'$ is another algebraic closure of $K(k)$ and $\sigma: \bar{K} \to \bar{K}'$ is a $K(k)$-isomophism, the map $H^*_{\mathrm{DR}}(X/\bar{K}) \to H^*_{\mathrm{DR}}(X_{\bar{K}'}/\bar{K}')$ induced by σ is compatible with the actions, via the map $W_{\mathrm{cris}}(\sigma)$: $W_{\mathrm{cris}}(\bar{K}) \to W_{\mathrm{cris}}(\bar{K})$.

(4.5) *Remark.* In fact, we have an additional functoriality: if $(\mathbf{X}', \alpha)$ and $(\mathbf{Y}', \beta)$ are as above, and if f: $\mathbf{Y}'_{k'} \to \mathbf{X}'_{k'}$ is a correspondence between their closed fibers, then f induces a map: $H^*_{\mathrm{cris}}(\mathbf{Y}'_{k'}/W(k')) \to H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))$ which is compatible with the action of absolute Frobenius. Therefore, it induces a map: $H^*_{\mathrm{DR}}(Y/\bar{K}) \to H^*_{\mathrm{DR}}(X/\bar{K})$ compatible with the actions of $W_{\mathrm{cris}}(\bar{K})$.

(4.6) *Remark.* In some situations, it is more convenient to work with finite extensions than with algebraically closed fields. To do this, we introduce the crystalline Weil category W_{cris}. An object in W_{cris} is a finite extension K' of $K(k)$; if its residue field is k', we have canonically $K(k') \subseteq K'$, (the maximal unramified extension of $K(k)$ in K'). A morphism $K' \to K''$ is a triple (d, a, b) where d is an integer, a: $K' \to K''$ is a K-linear map, and b: $K' \to K''$ is a map of fields inducing commutative diagrams:

$$\begin{array}{ccccc} K' & \xrightarrow{b} & K'' & & \\ \Big\uparrow & & & \nwarrow & \\ K(k') & \xrightarrow{F^d} & K(k') & \xrightarrow{a^*} & K(k''), \end{array} \qquad \begin{array}{ccccc} K' & \xrightarrow{b} & K'' & & \\ \Big\uparrow & & & \nwarrow & \\ K(k') & \xrightarrow{a^*} & K(k'') & \xrightarrow{F^d} & K(k''). \end{array}$$

Composition in W_{cris} is defined in the evident fashion. If (d, a, b): $K' \to K''$ is a morphism, we have a commutative diagram:

$$\begin{array}{ccc} H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K') & \longrightarrow & K'' \otimes_a H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K') \\ \Big\downarrow{\scriptstyle \sigma_{\mathrm{cris}}} & & \Big\downarrow{\scriptstyle \mathrm{id} \otimes \sigma_{\mathrm{cris}}} \\ K' \otimes H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k')) & \longrightarrow & K'' \otimes H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k')). \end{array}$$

Thus, the map $H^*_{\mathrm{cris}}(F_{\mathbf{X}'_{k'}})^d \otimes b$ defines a map:

$$\rho_{\mathrm{cris}}(d,a,b)\colon H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K') \to K'' \otimes_a H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K').$$

If we fix an algebraic closure $\bar{K}$ of $K(k)$ and consider the "limit" over all $K' \hookrightarrow \bar{K}$, we can recover the entire action of $W_{\mathrm{cris}}(\bar{K})$ in this way.

(4.7) In the remainder of this section, we shall investigate a few properties of the action of W_{cris}, with the hope of developing some sort of feeling for what this structure means.

Suppose that K is a finite extension of $K(k)$ contained in $\bar{K}$ and that X is a smooth proper K-scheme. Then $H^*_{\mathrm{DR}}(X_{\bar{K}}/\bar{K}) \cong H^*_{\mathrm{DR}}(X/K) \otimes_K \bar{K}$, and hence there is an evident semi-linear action ρ_{DR} of $\mathrm{Gal}(\bar{K}/K)$, with invariants $H^*_{\mathrm{DR}}(X/K)$. We shall say that X/K has potentially good reduction if $X_{\bar{K}}$ has good reduction. In this case $W_{\mathrm{cris}}(\bar{K})$ also acts on $H^*_{\mathrm{DR}}(X_{\bar{K}}/\bar{K})$, and it will be instructive to compare the actions ρ_{DR} and ρ_{cris}. Before we do so, some preliminaries are necessary.

First of all, we had better establish a "sign" convention for Galois groups. If K'/K is a Galois extension, $\mathrm{Gal}(K'/K)$ is the group of automorphisms of K' over K, which operates on the *left* on K' and on the the *right* on $\mathrm{Spec}\,K'$. If $f\colon K \to K'$ is a map of rings and X is a K-scheme, we let fX denote the K'-scheme defined by base change using the map $\mathrm{Spec} f$; thus if $g\colon K' \to K''$ is another map, there is a canonical isomorphism $\theta_{g,f}\colon (g \circ f)(X) \to g(fX)$, satisfying the usual cocycle condition. In particular, if $f\colon K \to K'$ is a Galois extension, if X is a K-scheme, and $g \in \mathrm{Gal}(K'/K)$, we obtain in this way a K'-isomorphism $\gamma_g\colon X_{K'} \to g(X_{K'})$, where $X_{K'} =: fX$. If h is another element of $\mathrm{Gal}(K'/K)$, the cocycle condition implies that $h(\gamma_g) \circ \gamma_h = \gamma_{hg}\colon X_{K'} \to hgX_{K'}$.

For example, suppose that X/K has potentially good reduction. Then there exist a finite Galois extension K' of K with ring of integers V', a smooth proper V'-scheme $\mathbf{X}'$, and an isomorphism $\alpha\colon \mathbf{X}'_{K'} \to X_{K'}$. Thanks to α, we get an isomorphism $\gamma'_g\colon \mathbf{X}'_{K'} \to g\mathbf{X}'_{K'}$, so that the following diagram commutes:

$$\text{(4.7.1)}\qquad \begin{array}{ccc} X_{K'} & \xrightarrow{\gamma_g} & gX_{K'} \\ \uparrow{\scriptstyle \alpha} & & \uparrow{\scriptstyle g(\alpha)} \\ \mathbf{X}'_{K'} & \xrightarrow{\gamma'_g} & g\mathbf{X}'_{K'} \end{array}$$

The isomorphisms γ'_g also satisfy the cocycle conditions. Moreover, if $g_{k'}$ is the image of g in the residual Galois group $\mathrm{Gal}(k'/k)$, we can specialize γ'_g to a correspondence of degree zero $\gamma'_{g,k'}\colon g_{k'}\mathbf{X}'_{k'} \circ\!\!\rightarrow \mathbf{X}'_{k'}$. These specialized correspondences still satisfy the cocycle conditions. In particular, if g belongs to the inertia group $I(K'/K)$, $g_{k'} = \mathrm{id}$, $g_{k'}\mathbf{X}'_{k'}$ is canonically isomorphic to $\mathbf{X}'_{k'}$, and we can view $\gamma'_{g,k'}$ as an endomorphism of $\mathbf{X}'_{k'}$. The cocycle condition implies that the action is "on the left," that is, $\gamma'_{h,k'} \circ \gamma'_{g,k'} = \gamma_{hg,k'}$. By functoriality, we obtain a *linear* action of $I(K'/K)$ on $H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k')) \otimes \mathbb{Q}$, on the left. Using the isomorphisms σ_{cris} and α we can transfer this action to a (linear) action δ_{DR} of

$I(K'/K')$ on $H^*_{\mathrm{DR}}(X_{K'}/K')$, hence also on $H^*_{\mathrm{DR}}(X_{\bar{K}}/\bar{K})$, if $\bar{K}$ is an algebraic closure of K'. Note that by (4.3), the representations δ and ρ_{cris} commute with one another. We view δ_{DR} as a representation of $\mathrm{Gal}(\bar{K}/K)$ by the map $\mathrm{Gal}(\bar{K}/K) \to \mathrm{Gal}(K'/K)$.

(4.8) **Theorem.** *If X/K is smooth, proper, and has potentially good reduction, and if $\bar{K}$ is an algebraic closure of K, let ρ_{cris}, ρ_{DR}, and δ_{DR} be the actions of $W_{\mathrm{cris}}(\bar{K})$, $\mathrm{Gal}(\bar{K}/K)$, and $I(\bar{K}/K)$ described above. Then if $g \in I(\bar{K}/K)$ is regarded as an element of $W_{\mathrm{cris}}(\bar{K})$, we have the following:*

$$\rho_{\mathrm{DR}}(g) = \delta_{\mathrm{DR}}(g) \circ \rho_{\mathrm{cris}}(g).$$

Proof. Suppose g is any member of $\mathrm{Gal}(K'/K)$, and use the notation of the discussion above. We have the following diagram:

$$\begin{tikzcd}
H^*_{\mathrm{DR}}(X/K)\otimes K' \arrow[r] & H^*_{\mathrm{DR}}(X_{K'}/K') \arrow[r, "\alpha^*"] & H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K') \arrow[r, "\sigma_{\mathrm{cris}}"] & H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes K' \\
H^*_{\mathrm{DR}}(X/K)\otimes K' \arrow[r] \arrow[u, "\mathrm{id}"] & H^*_{\mathrm{DR}}(gX_{K'}/K') \arrow[r, "(g\alpha)^*"] \arrow[u, "\gamma_g^*"'] & H^*_{\mathrm{DR}}(g\mathbf{X}'_{K'}/\mathbf{K}') \arrow[r, "\sigma_{\mathrm{cris}}"] \arrow[u, "\gamma_g'^*"'] & H^*_{\mathrm{cris}}(g_k\mathbf{X}'_{k'}/W(k'))\otimes K' \arrow[u, "\gamma'^*_{g,k'}"'] \\
H^*_{\mathrm{DR}}(X/K)\otimes K' \arrow[r] \arrow[u, "\mathrm{id}\otimes g"] & H^*_{\mathrm{DR}}(X_{K'}/K') \arrow[r, "\alpha^*"] \arrow[u] & H^*_{\mathrm{DR}}(\mathbf{X}'_{K'}/K') \arrow[r, "\sigma_{\mathrm{cris}}"] \arrow[u] & H^*_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes K'. \arrow[u, "\beta(g)"']
\end{tikzcd}$$

Let us analyze this diagram square by square, starting from the top left. Its horizontal arrows are the standard base change maps; it commutes because of the definition of $\gamma_g\colon X_{K'} \to gX_{K'}$. (Check it on the subspace $H^*_{\mathrm{DR}}(X/K)$ of $H^*_{\mathrm{DR}}(X/K)\otimes K'$.) The commutativity of the next square to the right is immediate from the definition of γ_g, and the last square on the top right commutes because of (4.3). In the second row, the horizontal arrows of the left most square are the base change maps again, and its right vertical map comes from the g-morphism: $gX_{K'} \to X_{K'}$. It commutes because the base change maps of DeRahm cohomology are natural. The next square on the right commutes for the same reason, and the last one is (2.7).

In particular, the left-most composed vertical arrow is the action of $\rho_{\mathrm{DR}}(g)$. If $g \in I(K'/K)$, then g_k is the identity and $\gamma'^*_{g,k'} = \delta_{\mathrm{DR}}(g)$, while $\beta(g)$ corresponds to $\rho_{\mathrm{cris}}(g)$. This proves the theorem. □

If k is a finite field, we can generalize this somewhat. Namely, in this case let $W_{\mathrm{cris}}(\bar{K}/K) =: W_{\mathrm{cris}}(\bar{K}) \cap \mathrm{Gal}(\bar{K}/K)$, and note that $W_{\mathrm{cris}}(\bar{K}/K)$ is dense in $\mathrm{Gal}(\bar{K}/K)$. Just as above, we obtain an action of all of $W_{\mathrm{cris}}(\bar{K}/K)$ on $\mathbf{X}'_{\bar{k}'}/k'$ (by correspondences), and hence an action δ of $W_{\mathrm{cris}}(\bar{K}/K)$ on $H^*_{\mathrm{DR}}(X_{\bar{K}}/\bar{K})$. The same proof shows that (4.8) is still true, for any $g \in W_{\mathrm{cris}}(\bar{K}/K)$.

The following corollary, which in fact is essentially equivalent to the theorem, was first stated (independently) by Messing.

(4.9) **Corollary.** *With the above notations, assume that K'/K is Galois and totally ramified, and extend the action of $I(K'/K)$ on $H^*_{\mathrm{cris}}(X_0/W)\otimes\mathbb{Q}$ to*

$H^*_{\mathrm{cris}}(X_0/W)\otimes K'$ *by letting it act (semi-linearly) on* K' *as usual. Then* σ_{cris} *induces an isomorphism*

$$H^*_{\mathrm{DR}}(X/K) \xrightarrow{\sim} (H^*_{\mathrm{cris}}(X_0/W)\otimes K')^{I(K'/K)}. \quad \square$$

(4.10) Using the isomorphism σ_{cris} and the group W_{cris}, we can generalize the conjectures in [20] concerning the relationship of absolute Hodge cycles to crystalline cohomology. This approach was suggested by Deligne in his seminar in 1978 in the context of abelian varieties, for which the isomorphism σ_{cris} was already known.

Let $\bar{K}$ be as above and let $X_1, \ldots, X_r$ be a finite family of smooth proper $\bar{K}$-schemes with good reduction. If $i<0$, let

$$H^i_{\mathrm{DR}}(X_j/\bar{K}) =: \mathrm{Hom}[H^{-i}_{\mathrm{DR}}(X_j/\bar{K}), \bar{K}],$$

with the usual action of $W_{\mathrm{cris}}(\bar{K})$ on Hom. If $i_1, \ldots, i_r$ is a corresponding family of integers, we let X denote the pair $((X_1, \ldots, X_r), (i_1, \ldots, i_r))$, and

$$H_{\mathrm{DR}}(X/\bar{K}) =: \bigotimes_j H^{i_j}_{\mathrm{DR}}(X_j/\bar{K}),$$

on which $W_{\mathrm{cris}}(\bar{K})$ acts. Since $\bar{K}$ is a field of characteristic zero, it makes sense to speak of the absolute Hodge cycles of X [6, 20]; recall that these form a finite dimensional $\mathbb{Q}$-vector space $H_{\mathrm{AH}}(X)$ and that $H_{\mathrm{AH}}(X)\otimes_{\mathbb{Q}} \bar{K} \hookrightarrow H_{\mathrm{DR}}(X/\bar{K})$.

(4.11) **Conjecture.** *If* $\xi \in H_{\mathrm{AH}}(X)$, *its image* ξ_{DR} *in* $H_{\mathrm{DR}}(X/\bar{K})$ *is invariant under* $W_{\mathrm{cris}}(\bar{K})$. *In particular, if* $\mathbf{X}/V'$ *is smooth and* $\xi \in H_{\mathrm{AH}}(\mathbf{X}_{K'}/K')$, *then* $\sigma_{\mathrm{cris}}(\xi_{\mathrm{DR}}) \in H_{\mathrm{cris}}(\mathbf{X}_{k'}/W(k'))\otimes K'$ *in fact lies in* $H_{\mathrm{cris}}(\mathbf{X}_{k'}/W(k'))\otimes K(k')$ *and is fixed by the action of absolute Frobenius.*

Suppose now that X is a finite family of smooth projective K-schemes with good reduction and indices, where K is a finite extension of $K(k)$. Then there is a natural representation ρ_{AH} of $\mathrm{Gal}(\bar{K}/K)$ in the finite dimensional $\mathbb{Q}$-vector space $H_{\mathrm{AH}}(X_{\bar{K}})$, and a morphism of $\mathrm{Gal}(\bar{K}/K)$-modules: $H_{\mathrm{AH}}(X_{\bar{K}}) \to H_{\mathrm{et}}(X_{\bar{K}}, \mathbb{Q}_l)$. Moreover, if we let $\mathrm{Gal}(\bar{K}/K)$ act on $H_{\mathrm{AH}}(X_{\bar{K}})\otimes \bar{K}$ by the *semi-linear* tensor product action, the map: $H_{\mathrm{AH}}(X_{\bar{K}})\otimes_{\mathbb{Q}} \bar{K} \to H_{\mathrm{DR}}(X_{\bar{K}}/\bar{K})$ takes the action ρ_{AH} to the action ρ_{DR} [6]. Using these facts we can obtain a clearer understanding of the inertial part of the above conjecture. Choose as above a finite extension K' of K and a smooth projective family $\mathbf{X}'$ over V' with $\mathbf{X}'_{\bar{K}} \cong \mathbf{X}_{\bar{K}}$. Then it is well-known that via the map: $H(\mathbf{X}'_{\bar{k}}, \mathbb{Q}_l) \cong H(X_{\bar{K}}, \mathbb{Q}_l)$, the action of $I(\bar{K}/K)$ through its action δ_{DR} on $\mathbf{X}'_{\bar{k}}$ corresponds to its action ρ_l on $H(X_{\bar{K}}, \mathbb{Q}_l)$ via $I(\bar{K}/K) \hookrightarrow \mathrm{Gal}(\bar{K}/K)$. Conjectually [23], the character of this action is an integer independent of l, (including p if we use crystalline cohomology).

(4.12) **Proposition.** *The following are equivalent:*

i) *The image of* $H_{\mathrm{AH}}(X_{\bar{K}})$ *in* $H_{\mathrm{DR}}(X_{\bar{K}})$ *is (pointwise) fixed by* $I(\bar{K}/K) \subseteq W_{\mathrm{cris}}(\bar{K})$.

i)$^{\mathrm{bis}}$ *The image of* $H_{\mathrm{AH}}(X_{\bar{K}})$ *in* $H_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes \bar{K} \cong H_{\mathrm{DR}}(X_{\bar{K}})$ *in fact lies in* $H_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k'))\otimes K_{nr}\cdot K$.

ii) *The map* $H_{\mathrm{AH}}(X_{\bar{K}}) \to H_{\mathrm{DR}}(X_{\bar{K}})$ *is compatible with the (linear) action* ρ_{AH} *of* $I(\bar{K}/K) \subseteq \mathrm{Gal}(\bar{K}/K)$ *on* $H_{\mathrm{AH}}(X_{\bar{K}})$ *and the (linear) action* δ_{DR} *of* $I(\bar{K}/K)$ *on* $H_{\mathrm{DR}}(X_{\bar{K}})$.

Proof. The equivalence of i) and i)$^{\mathrm{bis}}$ is obvious. The equivalence of i) and ii) follows immediately from (4.8). □

(4.13) **Example.** If X/K has good reduction, $I(\bar{K}/K)$ acts trivially on $H_{\mathrm{AH}}(X_{\bar{K}}) \subseteq H_{\mathrm{et}}(X_{\bar{K}}, \mathbb{Q}_l)$ and on $H_{\mathrm{DR}}(X_{\bar{K}}/\bar{K})$. Thus, the image of $H_{\mathrm{AH}}(X_{\bar{K}}) \to H_{\mathrm{DR}}(X_{\bar{K}}/\bar{K})$ is contained in $H_{\mathrm{DR}}(X/K) \otimes K_{nr}$.

(4.14) **Proposition.** *Suppose that $\psi \in W_{\mathrm{cris}}(\bar{K})$ has degree $d \geqq 0$ and that there is a ψ-linear endomorphism f of $X/\bar{K}$ lifting the d^{th} power of its absolute Frobenius endomorphism over $\bar{k}$. Then $H_{\mathrm{AH}}(X)$ is fixed by ψ.*

Proof. Choose a finite extension K' of K with ring of integers V' such that there is a smooth proper $\mathbf{X}'$ over V' such that $\mathbf{X}'_{\bar{K}} \cong X$. Let K'' contain the compositum of K' and $\psi(K')$ in $\bar{K}''$, and let ψ' denote the map $K' \to K''$ induced by $\psi: \bar{K} \to \bar{K}$, or the corresponding map $V' \to V''$. Let $\mathbf{X}''$ be the V''-scheme obtained by base change and the "inclusion" $V' \subseteq V''$, and let $\mathbf{X}'''$ be the V''-scheme obtained by base change using the map $\psi': V' \to V''$. Then $\mathbf{X}'''_{\bar{K}} \cong \psi(\mathbf{X}''_{\bar{K}}) \cong \psi(X)$ in our previous notation. The ψ-linear morphism $f: X \to X$ corresponds to a $\bar{K}$-linear morphism $f_{/\bar{K}}: X \to \psi(X)$. By choosing K'' large enough, we may assume that $f_{/\bar{K}}$ descends to a K''-morphism $f_{/K''}: \mathbf{X}''_{K''} \to \mathbf{X}'''_{K''}$. Since the map of residue fields $\psi'_{k'}: k' \to k''$ corresponds to the inclusion followed by $F^d_{k''}$, we have $\mathbf{X}'''_{k''} \cong \mathbf{X}''^{(d)}_{k''}$, and the assumption on f is that the correspondence $f_{/k''}: \mathbf{X}'''_{k''} \multimap \mathbf{X}''_{k'}$ is the d^{th} power of the relative Frobenius morphism. Hence by (2.7) and (4.3), the following diagram commutes:

$$\begin{tikzcd}
H_{\mathrm{DR}}(\mathbf{X}'_{K'}/K') \arrow[r, "\sigma_{\mathrm{cris}}"] \arrow[d, "\text{base change}"'] & H_{\mathrm{cris}}(\mathbf{X}'_{k'}/W(k')) \otimes K' \arrow[d, "\text{base change}"] \\
H_{\mathrm{DR}}(\mathbf{X}'''_{K''}/K'') \arrow[r, "\sigma_{\mathrm{cris}}"] \arrow[d, "f_{/K''}"'] & H_{\mathrm{cris}}(\mathbf{X}'''_{k''}/W(k'')) \otimes K'' \arrow[d, "F^{(d)*}_{X''_{k''}}"] \\
H_{\mathrm{DR}}(\mathbf{X}''_{K''}/K'') \arrow[r, "\sigma_{\mathrm{cris}}"] & H_{\mathrm{cris}}(\mathbf{X}''_{k''}/W(k'')) \otimes K''.
\end{tikzcd}$$

After tensoring with $\bar{K}$, we see that $f^*: H_{\mathrm{DR}}(X/\bar{K}) \to H_{\mathrm{DR}}(X/\bar{K})$ is $\rho_{\mathrm{cris}}(\psi)$.

Now by the very definition of absolute Hodge cycles, the notion is functorial even for morphisms covering nontrivial automorphisms of the ground field, such as f. Thus, we have commutative diagrams:

$$\begin{tikzcd}
H_{\mathrm{AH}}(X) \arrow[r, hook] \arrow[d, "f_{\mathrm{AH}}"'] & H_{\mathrm{DR}}(X/\bar{K}) \arrow[d, "f^*_{\mathrm{DR}}"] \\
H_{\mathrm{AH}}(X) \arrow[r, hook] & H_{\mathrm{DR}}(X/\bar{K})
\end{tikzcd} \qquad
\begin{tikzcd}
H_{\mathrm{AH}}(X) \arrow[r, hook] \arrow[d, "f_{\mathrm{AH}}"'] & H_{\mathrm{et}}(X, \mathbb{Q}_l) \arrow[d, "f_l^*"] \\
H_{\mathrm{AH}}(X) \arrow[r, hook] & H_{\mathrm{et}}(X, \mathbb{Q}_l).
\end{tikzcd}$$

But we also have a commutative diagram:

$$\begin{tikzcd}
H_{\mathrm{et}}(X, \mathbb{Q}_l) \arrow[r, "\sim"] \arrow[d, "f_l^*"'] & H_{\mathrm{et}}(\mathbf{X}'_{\bar{k}}, \mathbb{Q}_l) \arrow[d, "f_{\bar{k}}^*"] \\
H_{\mathrm{et}}(X, \mathbb{Q}_l) \arrow[r, "\sim"] & H_{\mathrm{et}}(\mathbf{X}'_{\bar{k}}, \mathbb{Q}_l).
\end{tikzcd}$$

Since $f_{\bar{k}}=(F_{X'_{\bar{k}}})^d$ and *absolute* Frobenius acts *trivially* on l-adic cohomology, we see that $f_{\mathrm{AH}}=\mathrm{id}$. □

(4.15) **Corollary.** *If each* X_i *is an abelian variety of* CM *type with ordinary reduction, then* $H_{\mathrm{AH}}(X)$ *is fixed by* $W_{\mathrm{cris}}(\bar{K})$.

Proof. An abelian variety of CM type is defined over a number field and has potentially good reduction everywhere. Thus we may assume that each X_i has good reduction and that the residue field k is finite. Moreover, we may replace each X_i by any isogenous abelian variety, and in particular we may assume that each X_i is simple. It is clear from the preceeding result that it suffices to prove that X_i is isogenous to an abelian variety defined over $W(k)$ to which the absolute Frobenius endomorphism lifts, i.e., to the canonical lifting of its reduction. This is well-known, but here is a proof. Let $A=X_i$, let E be the CM field acting on A and assume that the full ring of integers $\mathcal{O}$ of E acts on A. Then $H^1(A_{\mathbb{C}}, \mathbb{Q})$ is a free E-vector space of rank one, and hence $H^1_{\mathrm{et}}(A_{\bar{K}'}, \mathbb{Q}_l)$ is a free $E\otimes\mathbb{Q}_l$-module of rank one. We have $E\hookrightarrow \mathrm{End}(A_k)$, and if $k=\mathbb{F}_{p^d}$, $\varphi=:(F_{A_k})^d$ is a k-linear endomorphism of A_k, i.e., belongs to $\mathrm{End}(A_k)$. This φ acts on $H^1(A_{\bar{k}}, \mathbb{Q}_l)$, and since $H^1(A_{\bar{k}}, \mathbb{Q}_l)$ is a free $E\otimes\mathbb{Q}_l$-module of rank one, there is an $e\in E\otimes\mathbb{Q}_l$ such that $H^1(e, \mathbb{Q}_l)=H^1(\varphi, \mathbb{Q}_l)$. But $\mathrm{End}(A_k)/E\hookrightarrow(\mathrm{End}(A_k)/E)\otimes\mathbb{Q}_l$ and $\mathrm{End}(A_k)\otimes\mathbb{Q}_l\subseteq\mathrm{End}\, H^1_{\mathrm{et}}(A_k, \mathbb{Q}_l)$, so φ in fact lies in E. Since φ is integral over $\mathbb{Z}$, it lies in $\mathcal{O}$. Then by [15], A is isogenous to the canonical lifting of A_k. □

Appendix: Torsion in PD Envelopes

(A.1) **Proposition.** *Let R be an algebra over the localization of $\mathbb{Z}$ at (p), let $I\subseteq R$ be an ideal, and let $\gamma_p\colon I\to I$ be a function satisfying:*

1) $\gamma_p(x+y)=\gamma_p(x)+\gamma_p(y)+\sum_{i=1}^{p-1}\frac{1}{i!}\frac{1}{(p-i)!}x^i y^{p-i}$.
2) $\gamma_p(ax)=a^p\gamma_p(x)$.
3) $p!\,\gamma_p(x)=x^p$.

Then there is a unique PD *structure* $\{\gamma_n\}$ *on* I *extending* γ_p.

Proof. The uniqueness is proved in [1, 1.2.5]. For the existence, suppose that N is a positive integer and that we have operators γ_n for all $n<N$ satisfying all the axioms, viz:

(i) $\gamma_n(x+y)=\sum_{i+j=n}\gamma_i(x)\gamma_j(y)$ if $n<N$.

(ii) $\gamma_a(x)\gamma_b(x)=((a,b))\gamma_{a+b}(x)$ if $a+b<N$ and $((a,b))=:\frac{(a+b)!}{a!\,b!}$.

(iii) $\gamma_n(\lambda x)=\lambda^n\gamma_n(x)$ if $n<N$.

(iv) $\gamma_a(\gamma_b(x))=C_{a,b}\gamma_{ab}(x)$ if $ab<N$ and $C_{a,b}=\frac{(ab)!}{a!(b!)^a}$.

Then we can define γ_N satisfying the same axioms, by the following rules:

Case 0: $N\leqq p$. If $N<p$, let $\gamma_N(x)=\frac{1}{N!}x^N$, if $N=p$, use γ_p. It is trivial to check the axioms.

Case 1: $p \nmid N$ and $N>p$. Set $\gamma_N(x)=N^{-1}x\gamma_{N-1}(x)$. Note that by (ii), $x\gamma_i(x)=(i+1)\gamma_{i+1}(x)$ for $0\leqq i\leqq N-1$. Now let's check:

$$\begin{aligned}\text{(i)}\ \gamma_N(x+y)&=N^{-1}(x+y)\sum_{i=0}^{N-1}\gamma_i(x)\gamma_{N-i-1}(y)\\ &=N^{-1}\sum_{i=0}^{N-1}(i+1)\gamma_{i+1}(x)\gamma_{N-i-1}(y)+N^{-1}\sum_{i=0}^{N-1}(N-i)\gamma_i(x)\gamma_{N-i}(y)\\ &=\gamma_N(x)+N^{-1}\sum_{i=1}^{N-1}i\gamma_i(x)\gamma_{N-i}(y)+N^{-1}\sum_{i=1}^{N-1}(N-i)\gamma_i(x)\gamma_{N-i}(y)+\gamma_N(y)\\ &=\sum_{i+j=N}\gamma_i(x)\gamma_j(y).\end{aligned}$$

(ii) We may assume that $a+b=N$, so that one of $\{a,b\}$ is not divisible by p; say $p\nmid a$. Then:

$$\begin{aligned}\gamma_a(x)\gamma_b(y)&=a^{-1}x\gamma_{a-1}(x)\gamma_b(x)=a^{-1}((a-1,b))x\gamma_{a+b-1}(x)\\ &=a^{-1}((a-1,b))(a+b)\gamma_{a+b}(x)=((a,b))\gamma_{a+b}(x).\end{aligned}$$

(iii) Is trivial.

(iv) We may assume that $ab=N$, with $b>1$. Since $p\nmid a$ and $p\nmid b$,

$$\begin{aligned}\gamma_a(\gamma_b(x))&=\gamma_a(b^{-1}x\gamma_{b-1}(x))=b^{-a}x^a\gamma_a(\gamma_{b-1}(x))\\ &=b^{-a}(a!)\gamma_a(x)\,C_{a,b-1}\gamma_{ab-a}(x)\\ &=b^{-a}(a!)\,C_{a,b-1}((a,ab-a))\gamma_{ab}(x)=C_{ab}\gamma_{ab}(x).\end{aligned}$$

Case 2: $N=pm$ with $m>1$. In this case, one sees easily that $C_{m,p}$ is a p-adic unit, so we can define $\gamma_N(x)=C_{m,p}^{-1}\gamma_m(\gamma_p(x))$. Let us again check the axioms. The most difficult is (i), for which we shall need some notation. If $\mathbf{m}$ is a multi-index $(m_0,m_1,\ldots,m_p)$, let $|\mathbf{m}|=m_0+m_1+\ldots+m_p$, $a(\mathbf{m})=m_1+2m_2+\ldots+pm_p$, $b(\mathbf{m})=pm_0+(p-1)m_1+\ldots+m_{p-1}$.

Claim: There are universal constants $N_{\mathbf{m}}\in\mathbb{Z}$ (made explicit below) such that $\gamma_N(x+y)=C_{m,p}^{-1}\sum_{|\mathbf{m}|=m}N_{\mathbf{m}}\gamma_{a(\mathbf{m})}(x)\gamma_{b(\mathbf{m})}(y)$, (provided only that the γ_n's satisfy the axioms for $n<N$ and that γ_N is defined as above).

Granted the claim, let us prove property (i). Note that if $|\mathbf{m}|=m$, then $a(\mathbf{m})+b(\mathbf{m})=mp=N$. If $a+b=N$, let $N_{a,b}=:\sum\{N_{\mathbf{m}}\colon|\mathbf{m}|=m,\ a(\mathbf{m})=a,\ b(\mathbf{m})=b\}$. Then the claim can be rewritten:

$$\gamma_N(x+y)=C_{m,p}^{-1}\sum_{a+b=N}N_{a,b}\gamma_a(x)\gamma_b(y).$$

Thus, (i) reduces to the assertion that whenever $a+b=m$, $N_{a,b}=C_{m,p}$. Although this could be checked explicitly, we can also use the following argument. The ideal $(x,y)\subseteq\mathbb{Q}[x,y]$ admits a unique divided power structure, and it follows that the formula in the claim and formula (i) are both true in $\mathbb{Q}[x,y]$. Since $\{\gamma_a(x)\gamma_b(y)\}$ are linearly independent over $\mathbb{Q}$, we must have $N_{a,b}=C_{m,p}$ in $\mathbb{Q}$, hence also in $\mathbb{Z}$.

Perhaps it is unnecessary to make the ugly formula for $N_{\mathbf{m}}$ explicit, but here it is. If $(b_1,\ldots,b_k)$ is a sequence of positive integers, we let $((b_1,\ldots,b_k))=:\dfrac{(b_1+\ldots b_k)!}{b_1!\ldots b_k!}$. Then:

$$N_{\mathbf{m}}=:\left[\prod_{i=0}^{p}(m_i!)\,C_{m_i,i}\,C_{m_i,p-i}\right]((pm_0,(p-1)m_1,\ldots))((m_1,2m_2,\ldots,pm_p)).$$

Now let us prove the claim. We use formula (i) first for γ_p and then for γ_m to obtain

$$\gamma_N(x+y)=C_{m,p}^{-1}\gamma_m(\gamma_p(x+y))=C_{m,p}^{-1}\gamma_m\left(\sum_{i=0}^{p}\gamma_i(x)\gamma_{p-i}(y)\right)=C_{m,p}^{-1}\sum_{|\mathbf{m}|=m}\alpha_{\mathbf{m}},$$

$$\text{where}\ \alpha_{\mathbf{m}}=\prod_{i=0}^{p}\gamma_{m_i}(\gamma_i(x)\gamma_{p-i}(y)).$$

We now must calculate each term in the product. If $1 \leqq i \leqq p-1$, we have

$$\begin{aligned}\gamma_{m_i}(\gamma_i(x)\gamma_{p-i}(y)) &= (m_i)!\,\gamma_{m_i}(\gamma_i(x))\,\gamma_{m_i}(\gamma_{p-i}(y))\\ &= (m_i)!\,C_{m_i,i}\,\gamma_{im_i}(x)\,C_{m_i,p-i}\,\gamma_{(p-i)m_i}(y).\end{aligned}$$

For $i=0$ and $i=p$ we get, respectively:

$$\gamma_{m_0}(\gamma_p(y)) = C_{m_0,p}\,\gamma_{pm_0}(y) = (m_0)!\,C_{m_0,p}\,C_{m_0,0}\,\gamma_{pm_0}(y)$$

and

$$\gamma_{m_p}(\gamma_p(x)) = C_{m_p,p}\,\gamma_{pm_p}(x) = (m_p)!\,C_{m_p,0}\,C_{m_p,p}\,\gamma_{pm_p}(x),$$

which are of the same form. Thus we can write:

$$\alpha_{\mathbf{m}} = \prod_{i=0}^{p} (m_i!)\,C_{m_i,i}\,C_{m_i,p-i}\,\gamma_{im_i}(x)\,\gamma_{(p-i)m_i}(y).$$

Since $\prod_{i=0}^{p} \gamma_{im_i}(x)\,\gamma_{(p-i)m_i}(y) = ((pm_0,\dots,m_{p-1}))\,((m_1,2m_2,\dots,pm_p))\,\gamma_{a(\mathbf{m})}\,\gamma_{b(\mathbf{m})}$, the claim follows. This completes the proof of (i).

To prove (ii), we first suppose that $a=1$ and $b=N-1$. We have to check that $x\gamma_{N-1}(x) = N\gamma_N(x)$. We have

$$\begin{aligned}N\gamma_N(x) &= C_{m,p}^{-1}\,pm\gamma_m(\gamma_p(x)) = C_{m,p}^{-1}\,p\gamma_p(x)\,\gamma_{m-1}(\gamma_p(x))\\ &= C_{m,p}^{-1}\,x\gamma_{p-1}(x)\,C_{m-1,p}\,\gamma_{mp-p}(x)\\ &= C_{m,p}^{-1}\,C_{m-1,p}((p-1,mp-p))\,x\gamma_{mp-1}(x) = x\gamma_{mp-1}(x).\end{aligned}$$

Now suppose that one of $\{a,b\}$ is greater than 1 and not divisible by p - say a. Then $\gamma_a(x)\gamma_b(x) = a^{-1}x\gamma_{a-1}(x)\gamma_b(x) = a^{-1}((a-1,b))\,x\gamma_{N-1}(x) = a^{-1}((a-1,b))\,N\gamma_N(x)$ by the previous case. Since $a^{-1}((a-1,b))N = ((a,b))$, (ii) follows in this case also. Finally, if $a=pa'$ and $b=pb'$, we have

$$\begin{aligned}\gamma_a(x)\,\gamma_b(x) &= C_{a',p}^{-1}\,C_{b',p}^{-1}\,\gamma_{a'}(\gamma_p(x))\,\gamma_{b'}((\gamma_p(x))\\ &= C_{a',p}^{-1}\,C_{b',p}^{-1}((a',b'))\,\gamma_{a'+b'}(\gamma_p(x))\\ &= C_{a',p}^{-1}\,C_{b',p}^{-1}((a',b'))\,C_{a'+b',p}\,\gamma_N(x)\\ &= ((a,b))\,\gamma_N(x).\end{aligned}$$

Property iii) is trivial, and iv) is easy. □

(A.2) **Proposition.** *If $Z =: \operatorname{Spec} k[x,y]/(x^2,xy,y^2)$, then $H^0_{\mathrm{cris}}(Z/W(k))$ contains p-torsion.*

Proof. We do the case $p \neq 2$ and leave the modifications when $p=2$ to the reader.

Let J be the ideal $(x^2,xy,y^2) \subseteq W[x,y]$, and let γ be the standard divided power structure on (p). The first step is to show that the PD envelope $(\mathscr{D},\bar{J},\gamma)$ of $\mathcal{O}_Z$ in $A =: W[x,y]$ (compatible with γ on (p)) contains p-torsion. Let $\tau \in \bar{J}$ be the element $\tau =: \gamma_p(x^2)\gamma_p(y^2) - [\gamma_p(xy)]^2$. We claim that $p\tau = 0$. Indeed, $p!\,\gamma_p(x^2)\gamma_p(y^2) = (x^2)^p\gamma_p(y^2) = \gamma_p(x^2y^2) = (xy)^p\gamma_p(xy) = (p!)[\gamma_p(xy)]^2$. The hard part is to show that $\tau \neq 0$. To do this, it suffices to find a PD-algebra (B,K,δ), and a homomorphism $\theta\colon A \to B$ mapping the ideal (x^2,xy,y^2) to K, such that $\delta_p\theta(x^2)\,\delta_p\theta(y^2) \neq [\delta_p\theta(xy)]^2$.

Take $B = k[x,y,u,v]/(x^3,y^3,x^2y,xy^2,u^2,v^2,xu,xv,yu,yv)$, and let $K \subseteq B$ be the ideal generated by (x^2,y^2,xy,u,v). As a k-vector space K has a basis: (x^2,xy,y^2,u,v,uv). Define a Frobenius linear map $\delta_p\colon K \to K$ by $\delta_p(ax^2+bxy+cy^2+du+ev+fuv) =: a^pu + c^pv$.

Claim: δ_p extends to a PD structure on K.

We use (A.1). If $\alpha_1 = a_1x^2 + b_1xy + \dots$ and $\alpha_2 = a_2x^2 + \dots$, then $\delta_p(\alpha_1+\alpha_2) = (a_1^p + a_2^p)u + (c_1^p + c_2^p)v$. On the other hand, $K^3 = 0$, so $\alpha_1^i\alpha_2^j = 0$ if $i+j=p$. This implies that the right hand side of 1) reduces to $\delta_p(\alpha_1) + \delta_p(\alpha_2)$, which checks 1). The other two properties are clear. Now with the obvious map $A \to B$ and the obvious notation, we see that in B, $\delta_p(x^2)\delta_p(y^2) = uv$, while $\delta_p(xy)^2 = 0$.

Note that since $pB = 0$, the map $\mathscr{D} \to B$ factors through $\mathscr{D} \otimes \mathbb{Z}/p\mathbb{Z}$, hence $\tau \notin p\mathscr{D}$ and hence τ maps to a nonzero element of the p-adic completion of $\mathscr{D}$.

It remains only to check that $d\tau=0$. Writing $\alpha^{[n]}$ for $\gamma_n(\alpha)$, we calculate:

$$\tau=(x^2)^{[p]}(y^2)^{[p]}-((xy)^{[p]})^2=(x^2)^{[p]}(y^2)^{[p]}-((p,p))(xy)^{[2p]}.$$

$$\begin{aligned}\partial\tau/\partial x &= (2x)(x^2)^{[p-1]}(y^2)^{[p]}-((p,p))\,y(xy)^{[2p-1]}\\ &=\frac{2}{(p-1)!}x^{2p-1}(y^2)^{[p]}-((p,p))((p-1,p))^{-1}(xy)^{[p]}(xy)^{[p-1]}y\\ &=2x^{[p-1]}x^p(y^2)^{[p]}-2(xy)^{[p]}x^{[p-1]}y^p\\ &=2x^{[p-1]}(x^p(y^2)^{[p]}-(xy^2)^{[p]})\\ &=0.\end{aligned}$$

Since a similar calculation implies that $\partial\tau/\partial y=0$, we find $d\tau=0$, hence τ induces a torsion element of $H^0_{\text{cris}}(Z/W)$. □

(A.3) *Consequence: The absolute Frobenius endomorphism $F_Z^*\colon H^0_{\text{cris}}(Z/W)\to H^0_{\text{cris}}(Z/W)$ is not injective.*

Proof. Let $\varphi\colon W[x,y]\to W[x,y]$ send (x,y) to (x^p,y^p), and let $\bar{\varphi}\colon \mathscr{D}_J(A)\to\mathscr{D}_J(A)$ be the induced map. Then the map $\operatorname{Ker}(d)\to\operatorname{Ker}(d)$ induced by $\bar{\varphi}$ can be identified with F_Z^*, so it suffices to check that $\bar{\varphi}(\tau)=0$. But

$$\begin{aligned}\bar{\varphi}(\tau)&=\bar{\varphi}\{\gamma_p(x^2)\gamma_p(y^2)-[\gamma_p(xy)]^2\}\\ &=\gamma_p(x^{2p})\gamma_p(y^{2p})-[\gamma_p(x^py^p)]^2\\ &=\{(xy)^{2p-2}\}^p[\gamma_p(x^2)\gamma_p(y^2)-\gamma_p(xy)]^2\\ &=0\end{aligned}$$

since $\{(xy)^{2p-2}\}^p$ is divisible by p. □

(A.4) **Consequence.** *Over the base Z, the Dieudonné functor $\mathbb{D}$ from the category of p-divisible groups to the category of (F,V)-crystals is not fully faithful.*

Proof. Let G be the p-divisible group associated to the constant deformation of an ordinary elliptic curve over Z. Its Dieudonné crystal $\mathbb{D}$ is determined by its value H on the object $D=\operatorname{Spf}\hat{\mathscr{D}}_J(A)$; it is a free $\mathcal{O}_D$-module of rank 2 with horizontal basis $\{\omega,\eta\}$ in which the (F,V)-crystal structure is given by $\Phi(\omega^\sigma)=p\omega$, $\Phi(\eta^\sigma)=\eta$, $V(\omega)=\omega^\sigma$, $V(\eta)=p\eta^\sigma$ (where the exponent σ denotes the pull-back by $\bar{\varphi}$). Consider the endomorphism ε of H given by $\omega\mapsto\tau\eta$, $\eta\mapsto 0$. Since $d\tau=0$, this endomorphism is horizontal, and it is clear that $\Phi\circ\varepsilon^\sigma=\varepsilon\circ\Phi$, $\varepsilon^\sigma\circ V=V\circ\varepsilon$. Since $\varepsilon(0)=0$ but $\varepsilon\neq 0$, ε cannot be induced by a morphism $G\to G$.

References

1. Berthelot, P.: Cohomologie cristalline des schémas de caractéristique $p>0$. Lecture Notes in Math. vol. 407. Berlin-Heidelberg-New York: Springer 1974
2. Berthelot, P., Illusie, L.: Classes de Chern en cohomologie cristalline I et II. C.R. Acad. Sc. Paris **270**, série A, 1695 and 1750 (1970)
3. Berthelot, P., Ogus, A.: Notes on crystalline cohomology. Math. Notes, vol. 21. Princeton University Press (1978)
4. Berthelot, P., Messing, W.: Théorie de Dieudonné cristalline I. In: Journées de Géométrie Algébrique de Rennes 1978, Astérisque **63**, 17–37 (1979)
5. Berthelot, P., Breen, L., Messing, W.: Théorie de Dieudonné cristalline II, Lecture Notes in Math. **930**, Springer 1982
6. Deligne, P.: Hodge cycles on abelian varieties (notes by J.S. Milne). In: Hodge cycles, motives, and Shimura varieties. Lecture Notes in Math. vol. 900. Berlin-Heidelberg-New York: Springer 1982
7. Deligne, P., Illusie, L.: Cristaux ordinaires et coordonnées canoniques, in Surfaces algébriques, Lecture Notes in Math. **868**, Springer 1981

8. Dwork, B.: Normalized period matrices I. Ann. of Math. **94**, 377–388 (1971)
9. Dwork, B.: Normalized period matrices II. Ann. of Math. **98**, 1–57 (1973)
10. Gillet, H., Messing, W.: Riemann-Roch and cycle classes in crystalline cohomology (to appear)
11. Grothendieck, A.: Crystals and the De Rham cohomology of schemes (notes by J. Coates and O. Jussila). In: Dix exposés sur la cohomologie des schémas. North-Holland 1968
12. Hartshorne, R.: On the De Rham cohomology of algebraic varieties. Publ. Math. I.H.E.S. **45**, 6–98 (1975)
13. Katz, N.: Travaux de Dwork. In: Séminaire Bourbaki 1971–72, exposé 409. Lecture Notes in Math. vol. 317, 167–200 (1973)
14. Mazur, B.: Rational isogenies of prime degree. Invent. Math. **44**, 129–162 (1978)
15. Messing, W.: The crystals associated to Barsotti-Tate groups. Lecture Notes in Math. vol. 264. Berlin-Heidelberg-New York: Springer 1972
16. Norman, P.: Lifting abelian varieties. Invent. Math. **64**, (1981)
17. Ogus, A.: F-crystals and Griffiths transversality. In: Proceedings of the International Symposium on Algebraic Geometry at Kyoto 1977. Kinokuniya 1978
18. Ogus, A.: Supersingular K3 crystals. In: Journées de Géométrie Algébrique de Rennes 1978. Astérisque **64**, 3–87 (1979)
19. Ogus, A.: Griffiths transversality in crystalline cohomology. Ann. of Math. **108**, 395–419 (1978)
20. Ogus, A.: Hodge cycles and crystalline cohomology. In: Hodge cycles, motives and Shimura Varieties. Lecture Notes in Math. vol. 900. Berlin-Heidelberg-New York: Springer 1982
21. Saavedra Rivano, N.: Catégories tannakiennes. Lecture Notes in Math. vol. 265. Berlin-Heidelberg-New York: Springer 1972
22. Serre, J-P.: Groupes p-divisibles. In: Séminaire Bourbaki 1966–67, exposé 318, Benjamin
23. Serre, J-P., Tate, J.: Good reduction of abelian varieties. Ann. of Math. **88**, 492–517 (1968)
24. Tate, J.: p-divisible groups. In: Proc. Conf. on Local Fields. Nuffic Summer School at Driebergen. Berlin-Heidelberg-New York: Springer 1967

Oblatum 29-I-1982 & 20-X-1982

Invent. math. 72, 201–220 (1983)

Inventiones mathematicae

Artin Groups and Infinite Coxeter Groups

K.I. Appel and P.E. Schupp*

University of Illinois at Urbana-Champaign, Department of Mathematics, Urbana, IL 61801, USA

1. Introduction

A Coxeter matrix $\mathcal{M}$ over a finite set I is a symmetric matrix with entries $m_{ij} \in \mathbb{N} \cup \{\infty\}$, where $m_{ii} = 1$ for $i \in I$ and $m_{ij} \geqq 2$ for $i \neq j \in I$. (Many of our results will not depend on the cardinality of I, but we shall assume that I is finite for simplicity.)

The Artin group G defined by $\mathcal{M}$ is the group with generating set $\{a_i : i \in I\}$ and for each pair $i \neq j$ with $m_{ij} < \infty$ a defining relation of the form

$$a_i a_j a_i \ldots = a_j a_i a_j \ldots$$

saying that the alternating string of a_i's and a_j's of length m_{ij} beginning with a_i is equal to the alternating string of length m_{ij} beginning with a_j. This definition of an Artin group was introduced by Brieskorn [3] as a generalization of the braid groups. The braid group B_n on n strings has the well-known presentation with generators $a_1, \ldots, a_n$ and defining relations

$$a_i a_j = a_j a_i, \qquad |i-j| \geqq 2$$
$$a_i a_{i+1} a_i = a_{i+1} a_i a_{i+1}, \qquad i = 1, \ldots, n-1.$$

Associated with the Artin group G is the Coxeter group $\bar{G}$ obtained by setting the squares of the generators equal to the identity. The defining relations then become $a_i^2 = 1$, $i \in I$, and $(a_i a_j)^{m_{ij}} = 1$ for $i \neq j$, $m_{ij} < \infty$. Thus $\bar{G}$ is the Coxeter group defined by the matrix $\mathcal{M}$. An Artin group is said to be of *finite type* if the associated Coxeter group $\bar{G}$ is finite. (For the braid group B_n the associated Coxeter group is the symmetric group S_{n+1}.) Brieskorn and Saito [4], and, independently, Deligne [6] have investigated Artin groups of finite type. In 1969, Garside [7] solved the conjugacy problem for braid groups. Brieskorn,

* This work was supported by NSF grants MCS 7902512 and MCS 7703452. The authors would like to thank the referee for an extremely careful reading of the paper and many helpful suggestions

Saito, and Deligne showed that any Artin group of finite type has solvable conjugacy problem. Brieskorn and Saito showed that their solution of the conjugacy problem is applicable only when the Artin group is of finite type. For detailed discussions of braid and Artin groups, see Birman [1] and Len [8].

The theorem of Coxeter [5] classifying the finite Coxeter groups is well-known. If a Coxeter group $\bar{G}$ on more than two generators is finite then "few" of the m_{ij} can exceed two. In particular, for any three distinct generators a_i, a_j, a_k at least one of m_{ij}, m_{jk}, and m_{ik} must equal two. By contrast, we say that an Artin group or a Coxeter group is of *large type* if $m_{ij} \geqq 3$ for all $i \neq j$, and is of *extra-large type* if $m_{ij} \geqq 4$ for all $i \neq j$.

We have discovered that Artin and Coxeter groups of extra-large type can be studied by extending some of the techniques of small cancellation theory. These methods can be further refined to handle groups of large type. With the exception of Theorem 3 (which would need to be modified), the theorems we prove are very probably true for all Artin and Coxeter groups. An extension of our techniques to the general case seems quite difficult.

The theorems stated below, except for the solvability of the conjugacy problem, are all known for Coxeter groups. Most follow from Tits' elegant solution of the word problem for Coxeter groups [12]. See also Bourbaki [2]. An interesting feature of our technique is that there is little difference in the arguments for Artin groups and for Coxeter groups. When stating results applicable to both Artin and Coxeter groups we use G to denote an Artin or Coxeter group with generating set $\{a_i: i \in I\}$ and Coxeter matrix $\mathcal{M}$. If $J \subseteq I$, let G_J denote the subgroup of G generated by $\{a_j: j \in J\}$, and let $\mathcal{M}_J$ denote the restriction of $\mathcal{M}$ to $J \times J$. If H is a subgroup of G one says that the generalized word problem for H in G is solvable if there is an algorithm which, when given an element $g \in G$ decides if $g \in H$.

Theorem 1. *Let G be an Artin or Coxeter group of extra-large type. If $J \subseteq I$ then G_J has a presentation defined by the Coxeter matrix $\mathcal{M}_J$ and the generalized word problem for G_J in G is solvable. If $J, K \subseteq I$ then $G_J \cap G_K = G_{(J \cap K)}$.*

Theorem 2. *An Artin group of extra-large type is torsion-free.*

There is a "Freiheitssatz" for Artin groups.

Theorem 3. *Let G be an Artin group of extra-large type. Then the set $\{a_i^2: i \in I\}$ freely generates a free subgroup of G.*

Professor Tits has pointed out to us the similarity of Theorem 3 and Theorem 2.5 of [13]. He conjectures that for an arbitrary Artin group the only relations between the a_i^2 are the obvious commutations. This seems likely but our methods do not shed light on the general case.

Theorem 4. *An Artin or Coxeter group of extra-large type has solvable conjugacy problem.*

Our proofs use the results and methods of small cancellation theory. This theory had its origins in Dehn's work on Fuchsian groups. For a Fuchsian group G the entire Cayley graph lies in the hyperbolic plane and Dehn used hyperbolic geometry to solve the word and conjugacy problems for finitely generated Fuchsian groups.

Dehn showed that a non-trivial word w equal to the identity in G contains more than half of a defining relator. Let $|u|$ denote the length of the word u and suppose that $w=w_1 s w_2$ and a defining relator r of G has the form st with $|s|>|t|$. Then, in G, $w=w'=w_1 t^{-1} w_2$ and $|w'|<|w|$. This shortening process gives an algorithm, now called *Dehn's algorithm*, for solving the word problem. Dehn also found that conjugate elements of G are conjugate by a "short" conjugating element.

In general, there is no nice geometric representation of a group. The method of cancellation diagrams, introduced by R.C. Lyndon [9] in 1966, gives geometric insight into rather general situations. A cancellation diagram is a "local slice" of a Cayley graph which is small enough to be planar. In the small cancellation situation the diagram permits one to initate some of the arguments of hyperbolic geometry.

It turns out that there is a very nice interaction between the forms of defining relators for Artin and Coxeter groups of extra-large type and the geometry of cancellation diagrams. We shall show that an extended version of Dehn's algorithm solves the word problem for such groups. We also obtain a strong analogue of Dehn's results on the conjugacy problem.

2. Small Cancellation Theory and the Word Problem for Coxeter Groups

Small cancellation theory has become an important part of combinatorial group theory. (For a survey see [11].) We shall need various results from this theory. (For detailed proofs see [10].)

We write a group G as the quotient of a free group F $(=F\langle X\rangle)$, say $G=F/N$, where N is the normal closure of a set $\mathscr{R}$. We assume that $\mathscr{R}$ is *symmetrized*, that is, all elements of $\mathscr{R}$ are cyclically reduced and $\mathscr{R}$ is closed under cyclic permutations and taking of inverses.

If $\mathscr{R}$ is a symmetrized subset of F then an *$\mathscr{R}$-diagram* M is a diagram in the plane such that every edge e of M is *labelled* by an element $1\neq\varphi(e)\in F$ and such that if D is a region of M and the edges of the boundary of D are (in counterclockwise order) $e_1, \dots, e_k$, then the word $\varphi(e_1)\dots\varphi(e_k)$ is reduced without cancellation and is an element of R. We state the first basic result of small cancellation theory.

Result I. Let N be the normal subgroup of F generated by a symmetrized set $\mathscr{R}$. If w is a freely reduced word of F then $w\in N$ if and only if there is a connected, simply connected $\mathscr{R}$-diagram M in the plane such that the label on the boundary of M is w.

In order to use Result I to actually prove results one needs some sort of hypothesis on $\mathscr{R}$. If r_1 and r_2 are elements of $\mathscr{R}$ with $r_1\neq r_2^{-1}$, $r_1=c_1 b$, and $r_2=b^{-1}c_2$ then b is called a *piece*. A piece is thus simply a subword of an element of $\mathscr{R}$ that can be cancelled in the multiplication of two non-inverse elements of $\mathscr{R}$. The hypotheses of "small cancellation" assert that pieces are relatively small parts of elements of $\mathscr{R}$. We say that $\mathscr{R}$ satisfies *Condition $C(p)$* if no element of $\mathscr{R}$ is a product of fewer than p pieces; $\mathscr{R}$ satisfies *Condition*

$T(q)$ if whenever $3<n\leqq q$ and $r_1,\ldots,r_n\in\mathcal{R}$, and no successive r_i,r_{i+1} form an inverse pair (including r_n,r_1), at least one of the products $r_1r_2,\ldots,r_{n-1}r_n,r_nr_1$ is reduced without cancellation. (Note that $T(3)$ is vacuous.)

The meaning of these conditions will become clear when we interpret their geometric significance for $\mathcal{R}$-diagrams. Let M be a diagram. If D is a region of M then a sequence $e_1,\ldots,e_n$ of consecutive edges that contains all the edges in the boundary ∂D of D and is of minimal length is called a *boundary cycle* of D. The *degree*, $d(D)$, of D is the number of edges in a boundary cycle of D. A *boundary vertex or edge* is a vertex or edge in the boundary ∂M of M. Vertices are interior if they are not on the boundary. The interior degree, $i(D)$, of a region D is the number of edges in a boundary cycle of D that are interior edges of M. A region D is *almost interior* if ∂D has no edges on the boundary of M (but ∂D may contain vertices of M). The *degree*, $d(v)$, of a vertex v is the number of edges incident to v. If interior edges e_1 and e_2 meet at vertex v of degree two, then we can delete v and combine e_1 and e_2 into a single edge e with label $\varphi(e)=\varphi(e_1)\,\varphi(e_2)$. Henceforth we shall consider only diagrams in which all interior vertices have degree at least three.

We need one more definition. Let D_1 and D_2 be regions (not necessarily distinct) of a diagram M such that an edge e lies in $\partial D_1\cap\partial D_2$. Let $e\delta_1$ and δ_2e^{-1} be boundary cycles of D_1 and D_2 respectively. M is *reduced* if in such a situation one never has $\varphi(\delta_2)=(\varphi(\delta_1))^{-1}$. (For D_1 and D_2 distinct, this is equivalent to saying that if one deletes e and combines D_1 and D_2 into a single region D then the label on the boundary of D is not freely equal to 1.) The geometric significance of the small cancellation conditions now follows easily.

Result II. Let $\mathcal{R}$ satisfy $C(p)$ and $T(q)$ and let M be a reduced $\mathcal{R}$-diagram. If D is an almost interior region of M then $d(D)\geqq p$ and if v is an interior vertex of M then $d(v)\geqq q$.

Since we are mainly interested in diagrams, we shall say that a diagram M satisfies $C(p)$ and $T(q)$ if M satisfies the conclusion of Result II. These properties have strong consequences for the "rate of growth" of such diagrams. If D is a region of a diagram M we say that D is a *simple boundary region* of M if $\partial D\cap\partial M$ is non-empty (although perhaps consisting only of a vertex) and connected and if the edges in $\partial D\cap\partial M$ are consecutively traversed in some boundary cycle of M. We use $\Sigma^*(k-i(D))$ to denote the sum of $(k-i(D))$ over the simple boundary regions of M. An unpleasant complication that can arise in the treatment of diagrams is the possibility of the existence of vertices of degree one. Such vertices arise if there is a *spike*, i.e., a part of the boundary curve of the diagram consisting of edges that lie on no regions and whose vertices a vertex of degree one. Since the part of a boundary label that lies on a spike freely reduces to the identity, the hypothesis that boundary words are freely reduced is sufficient to make spikes impossible.

The next result is a discrete analogue of the Gauss-Bonnet curvature formula.

Result III. Let M be a connected, simply connected diagram with no vertices of degree one and more than one region. If M satisfies $C(6)$ then $\Sigma^*_M[4-i(D)]\geqq 6$. If M satisfies $C(4)$ and $T(4)$ then $\Sigma^*_M[3-i(D)]\geqq 4$.

We will call a region D of M a *Dehn region* if D is a simple boundary region and $i(D) \leqq 3$. If M satisfies $C(7)$ a similar result shows that if M has more than one region then M has at least two Dehn regions so Dehn's algorithm applies.

Now we consider Coxeter groups. Let $\bar{F}_i = \langle a_i;\ a_i^2 = 1\rangle$ and let $\bar{F} = \underset{i \in I}{\ast} \bar{F}_i$. Thus $\bar{F}$ is a free product of cyclic groups of order two, and we identify a generator and its inverse. A word $w = a_{i_1} \ldots a_{i_n}$ is reduced if no two successive generators have the same subscript. The length of w is n. Subject to the condition that each generator is its own inverse, one can construct $\mathscr{R}$-diagrams over $\bar{F}$ exactly as over a free group.

If $m_{ij} < \infty$, let $r_{ij} = (a_i a_j)^{m_{ij}}$. Since each generator is equal to its inverse in $\bar{F}$ there are exactly two distinct cyclic permutations of r_{ij}, namely r_{ij} and $r_{ji} = r_{ij}^{-1} = (a_j a_i)^{m_{ij}}$. Let $\mathscr{R} = \{r_{ij} \mid i \neq j \in I\}$. If $r, r' \in \mathscr{R}$ and $r' \neq r^{-1}$ then at most one letter is cancelled in the product rr'. Thus in a minimal diagram for a Coxeter group the label on an edge common to two distinct regions has length one.

A region D of M is *self-identified* if there is an edge e of D such that both e and e^{-1} are traversed in a boundary cycle of D. A connected, simply connected $\mathscr{R}$-diagram M is *minimal* if there is no such $\mathscr{R}$-diagram with fewer regions that has the same boundary label as M.

Lemma 1. *Every minimal connected, simply connected $\mathscr{R}$-diagram for a Coxeter group G of large type is reduced and does not contain any self-identified region.*

Proof. Suppose that M is such a diagram and that e is an edge on the common boundary of D_1 and D_2 making M not reduced. If $D_1 \neq D_2$ then, as usual e can be removed, combining D_1 and D_2 into a single region D with boundary label 1, contradicting minimality. Hence $D_2 = D_1$ and D_1 has a boundary cycle of the form $e\eta e^{-1}\gamma$ where η bounds a submap K of M that contains at least one region. Among all edges that make M non-reduced let e be such that K has as few regions as possible. Then K is reduced and thus is a $C(6)$ map. Hence K contains a Dehn region E. Thus E and D_1 have three consecutive edges in common and hence their labels form an inverse pair, contradicting minimality. □

The proof that Dehn's algorithm solves the word problem for Coxeter groups of extra-large type is now immediate. If w is a non-trivial reduced word that is equal to 1 in G let M be a minimal connected, simply connected R-diagram with boundary label w. Lemma 1 and the hypothesis that G is of extra-large type say that M is a $C(8)$ diagram. Let D be a Dehn region of M. Then w can be written $w_1 s w_2$ where $s = \varphi(\partial D \cap \partial M)$ and some defining relator r has the form $r = st$ with t a product of not more than three pieces. But pieces all have length one so $|t| \leqq 3$ and $|s| \geqq 5$ and Dehn's algorithm applies.

3. Further Small Cancellation Theory

Diagrams that satisfy $C(4)$ and $T(4)$ play a crucial role in our study of Artin groups. In this section we examine the structure of such diagrams. Let M be a connected, simply connected diagram satisfying $C(4)$ and $T(4)$. A *singleton strip* of M is a simple boundary region D of M satisfying $i(D) \leqq 1$. A *compound strip* of M is a subdiagram S of M consisting of regions $D_1, \ldots, D_n$, $n \geqq 2$ which are all simple boundary regions of M successively encountered in the given order in the counterlockwise direction along ∂M such that $i(D_1) = i(D_n) = 2$ and $i(D_k) = 3$ for $1 < k < n$, and D_{k-1} and D_k meet along a single interior edge of S. A *strip* is either a singleton strip or a compound strip.

Two strips are called *disjoint* if they have no regions in common. If S is a strip in M then $\partial M \cap \partial S$ is called the *outside boundary* of S while $\partial S \cap \partial(M - S)$ is called the *base* of S. (Note that the base of a singleton strip of interior degree 0 is a single point.) If S is a compound strip and e is an edge separating two regions of S then e is called an *interior edge* of S. If v is a vertex on $\partial(M - S)$ but not on ∂S then v is called a *base vertex* of S.

Lemma 2. *Let M be a connected, simply connected diagram with no spikes and more than one region. If M satisfies $C(4)$ and $T(4)$ then either M contains two singleton strips, or four distinct strips, possibly overlapping, of which at least two are disjoint.*

Proof. The proof is by induction on the number of regions of M. First, suppose that the boundary of M is a simple closed curve and that every boundary region of M is a simple boundary region. From Result III we know that $\Sigma^*_M[3 - i(D)] \geqq 4$ and we examine how this can happen. Beginning at some boundary vertex we traverse the boundary of M in counterlockwise order, at each boundary region computing the partial sum $s = \Sigma^*(3 - i(D))$ over those simple boundary regions encountered so far. Since M has more than one region and the boundary of M is a simple closed curve, M has no region with interior degree zero. A region D of interior degree $3 - k$ contributes k to s.

A boundary region D with $i(D) = 3 + k$, $k > 0$ can be thought of as cancelling the most recent contributions of regions of interior degree less than three.

First suppose there are no singleton strips. We proceed around the boundary of M in the counterclockwise direction beginning at a region of smallest interior degree. Each time a region of interior degree two is encountered it is followed by a sequence (possibly empty) of regions of interior degree three and then either a region of interior degree two or one of interior degree at least four. If a region of interior degree two is encountered a strip is completed (although its initial region may be the terminal region of a previous strip). If a region of interior degree four or more is encountered then the contribution of the previous region of interior degree two to $S^*[3 - i(D)]$ is cancelled. Thus, if M has no singleton strip there must be four uncancelled interior degree two contributions, hence four compound strips.

If M has a singleton strip D then let M' be the map obtained by removing the region D from M. If M' has only one region E then M has singleton strips D and E. If M' has more than one region then M' satisfies one of the conclusions of the lemma and M satisfies the same conclusion.

Now, still assuming that the boundary of M is a simple closed curve we suppose that there is a boundary region E of M with $\partial E \cap \partial M$ disconnected.

Then $M-E$ has at least two components say C_1 and C_2. The diagrams $J_i = C_i \cup E$ each have more than one region. By induction each J_i satisfies the lemma. At most one of the strips in each J_i can involve E. Thus the lemma holds for M regardless of which case of the conclusion holds for each J_i.

Finally, if the boundary of M is not a simple closed curve then M has two subdiagrams M_1 and M_2 that are extremal disk in the sense that the boundary of each M_i is a simple closed curve and M_i is connected to the rest of M by a single vertex, v_i. Each M_i is either a single region, and thus a singleton strip or contains more than one region and satisfies the inductive conclusion. Attachement at v_i can make at most one strip of M_i not encountered consecutively in the boundary of M. Thus M satisfies the conclusion of the lemma. $\square$

4. Two Generator Artin Groups

We always consider an Artin (or Coxeter) group with generating set $A = \{a_i : i \in I\}$ and defining matrix $\mathcal{M}$. Let $F = \langle A \rangle$ be the free group on A with the free group length function. If w is a freely reduced word on A then the length of w, written $|w|$, is the number of letters in w. Let F_i be the infinite cyclic group generated by a_i. Then F also has the free product structure $F = \mathop{\ast}\limits_{i \in I} F_i$ and the corresponding free product normal form. If $w = a_{j_1}^{e_1} \dots a_{j_n}^{e_n}$ where each $j_t \neq j_{t-1}$ and each $e_t \neq 0$, then the free product or *syllable length* of w, written $\|w\|$, is n.

In this section we examine words that represent the identity in a two generator Artin group. We use $\bar{a}_i$ to denote the inverse of generator a_i. Let G_{ij} be the two generator one-relator group $\langle a_i, a_j; \hat{\mathcal{R}}_{ij} \rangle$ where $\hat{\mathcal{R}}_{ij}$ is the symmetrized set obtained from the defining relator

$$r_{ij} = a_i a_j a_i \dots \bar{a}_j \bar{a}_i \bar{a}_j \quad \text{if } 2 \leqq m_{ij} < \infty,$$

and $\hat{\mathcal{R}}_{ij}$ is empty if $m_{ij} = \infty$.

Lemma 3. *The set $\hat{\mathcal{R}}_{ij}$ satisfies the small cancellation conditions $C(4)$ and $T(4)$.*

Proof. An element $r \in \hat{\mathcal{R}}_{ij}$ is a cyclic permutation of r_{ij} or its inverse. Thus, there are exactly two places in r where successive occurrences of generators have opposite sign. If m_{ij} is even $a_j \bar{a}_i$ and $\bar{a}_j a_i$ occur in cyclic permutations of r_{ij}, while if m_{ij} is odd they occur in r_{ij}^{-1}. Otherwise $a_i \bar{a}_j$ and $\bar{a}_i a_j$ occur.

Call a subword of r consisting of successive occurrences of generators of opposite signs a *special subword.* If $r_1, r_2 \in \hat{\mathcal{R}}_{ij}$ and a special subword of r_1 or r_2 is cancelled in the product $r_1 r_2$ then $r_2 = r_1^{-1}$. Similarly, if a subword of r_1 of length m_{ij} is cancelled in the product $r_1 r_2$ then $r_2 = r_1^{-1}$. Thus, if $r_2 \neq r_1^{-1}$ the piece c of r_1 that is cancelled in $r_1 r_2$ cannot contain a special subword and cannot overlap both special subwords since they begin m_{ij} letters apart in r. Thus no element of $\hat{\mathcal{R}}_{ij}$ can be a product of fewer than four pieces. The distinct generators a_i and a_j occur alternately in elements of $\hat{\mathcal{R}}_{ij}$. Thus if

$r_1, r_2, r_3 \in \hat{\mathscr{R}}_{ij}$ there cannot be cancellation in all three of the products $r_1 r_2$, $r_2 r_3$ and $r_3 r_1$ for such cancellation would require two successive occurrences of the same generator. Thus $T(4)$ is satisfied. $\square$

At this point we need another result from small cancellation theory.

Result IV. Let $G = \langle X; \mathscr{R} \rangle$ be a finitely presented group where $\mathscr{R}$ satisfies either $C(6)$, or $C(4)$ and $T(4)$. Then the word and conjugacy problems for G are solvable.

We see from Lemma 3 and Result IV that the group G_{ij} has solvable conjugacy problem. Since G_{ij} is a one-relator group whose defining relator is not a proper power, G_{ij} is torsion-free.

Corollary 1. *The two generator group G_{ij} has solvable word and conjugacy problems and is torsion-free.*

If M is an $\hat{\mathscr{R}}_{ij}$-diagram we assume (as usual) that there are no interior vertices of degree two but we also assume that the boundary edges of M are subdivided so that each boundary edge is labelled by a generator or its inverse. The proof of Lemma 3 shows that if M is reduced and if e is an interior edge of M then the label $\varphi(e)$ on e cannot contain a special subword. Thus, for every region D of M there are vertices v_1 and v_2 that separate the occurrences of generators of opposite exponent in the special subwords of the boundary label r of D. We call such vertices *separating* vertices of D. If $\delta_1 \delta_2$ is a boundary cycle of D where δ_1 and δ_2 begin and end at the separating vertices then all generators occurring in δ_1 have one exponent and all those occurring in δ_2 have the opposite exponent.

We note another property of separating vertices for future use.

Lemma 4. *If D_1 and D_2 have a common edge incident to a vertex v of M then v is not a separating vertex for both D_1 and D_2.*

Proof. If v were a separating vertex for both D_1 and D_2 then $D_1 \cup D_2$ would have boundary label 1, contradicting minimality. $\square$

Lemma 5. *Let S be a strip of a reduced $\hat{\mathscr{R}}_{ij}$-diagram M, let s be the label on the outside boundary of S, and let b be the label on the base of S. Then $\|s\| \geqq m_{ij} + 1$ and the length and syllable length of s each exceed the corresponding quantity for b by at least two. Furthermore, some region of S has both of its separating vertices on the outside boundary of S.*

Proof. We prove the lemma by induction on the number of regions in S. If S is a singleton strip the conclusion is immediate. If S consists of two regions D_1 and D_2 meeting along an interior edge e with base vertex v then, by Lemma 4, v can be a separating vertex for at most one of D_1 and D_2. Thus, at least one and possibly both of D_1 and D_2 have both separating vertices on the outside boundary of S. Although formally there are a number of cases to consider depending on location of separating vertices all analyses are quite similar to the case of Fig. 1 that we will consider. (In the figure separating vertices are indicated by $\ast$.) Here let s_i be the label on $\partial D_i \cap \partial M$, let b_i be the label on

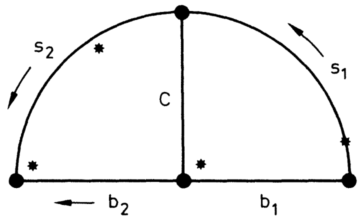


Fig. 1

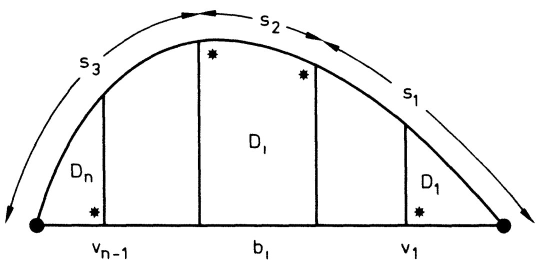


Fig. 2

that part of the base of S that is on D_i, and let c be the label on the edge e. Since the length of a label between separating vertices of a region is m_{ij}, $|s_2| \geqq |b_2| + |c|$. Because b_1 labels an edge shorter than the distance between separating vertices $|b_1| < m_{ij}$ and thus $|c| + |s_1| \geqq |b_1| + 2$. Adding these inequalities yields $|s_1| + |s_2| \geqq |b_1| + |b_2| + 2$ as desired. To check the inequality on syllable length note that from the form of the relators, each new letter encountered on s or b is also a change of syllable *except* at the two vertices of degree three in s which are the endpoints of the edge e and at these vertices there is no syllable change in either s or b. Thus the syllable length of each of s and b is exactly one less than the corresponding length and so the corresponding inequality holds.

We now consider the case in which S consists of regions $D_1, \ldots, D_n$, $n \geqq 3$. Note that if $1 < i < n$ then $D_{i+1}, \ldots, D_n$ is a strip of the diagram M' obtained by removing regions $D_1, \ldots, D_i$ from M. Similarly, $D_1, \ldots, D_i$ is a strip of the diagram obtained by removing $D_{i+1}, \ldots, D_n$. We first consider the case in which the end regions D_1 and D_n have base vertices v_1 and v_{n-1} as separating vertices (see Fig. 2). Then, by Lemma 4, v_1 and v_{n-1} are not separating vertices of D_2 and D_{n-1}, respectively. It is easy to see that there must be a region D_i, $1 < i < n$, that has both its separating vertices on the outside boundary of S. Let s_1 be the label on the part of the outside boundary of S that is on the boundary of $D_1 \cup \ldots \cup D_{i-1}$. Let s_2 be the label on $\partial D_i \cap \partial M$ and let s_3 be the label on the part of the outside boundary of S that is on the boundary of $D_{i+1} \cup \ldots \cup D_n$. Let b_1, b_2, b_3 be labels on the corresponding parts of the base of S and let c_1, c_2 be the labels on the interior edges of D_i. By the induction hypothesis $|s_1| + |c_1| \geqq |b_1| + 2$ and $|s_3| + |c_2| \geqq |b_3| + 2$. Since both separating vertices of D_i are on the outer boundary of S, $|s_2| \geqq |c_1| + |b_2| + |c_2|$. Adding these inequalities yields $|s_1|+|s_2|+|s_3| \geqq |b_1|+|b_2|+|b_3|+4$. As in the previous discussion the corresponding inequality for syllable lengths follows immediately. If one of the end regions, say D_1, has both of its separating vertices on the outside boundary of S then one can divide S into two parts D_1 and $D_2 \cup \ldots \cup D_n$ and apply the induction hypothesis. □

An important consequence of Lemma 5 is the following length estimate on non-trivial words that are equal to the identity in G_{ij}.

Lemma 6. *If w is a non-trivial reduced word and $w=1$ in G_{ij} then $\|w\| \geqq 2m_{ij}$.*

Proof. We may assume that w is cyclically reduced in the free product sense. If M has one region then w is a defining relator and thus has exactly $2m_{ij}$ syllables. If M has more than one region then M has two disjoint strips, S_1 and S_2, by Lemma 1. By Lemma 5, the label s_h, the outside boundary of S_h satisfies $\|s_h\| \geqq m_{ij}+1$. There is a cyclic permutation w' of w in which both s_1 and s_2 occur as subwords. The last syllable of s_1 and the first syllable of s_2 may be a single syllable of w' and another syllable may be lost in conjugating back to w but $\|w\| \geqq 2(m_{ij}+1)-2=2m_{ij}$. □

Next, we obtain a lemma that makes the extended Dehn algorithm effective.

Lemma 7. *Let $w=1$ in G_{ij} and write $w=w_1 w_2$. a) If $\|w_1\| \leqq m_{ij}$ then $|w_1| \leqq |w_2|$. b) If $\|w_1\| < m_{ij}$ then $|w_1| < |w_2|$.*

Proof. The proofs of a) and b) are parallel. For a) read occurrences of $<*$ as $\leqq$ for b) read $<*$ as $<$.

The proof is by induction on the number of regions in the minimal $\mathscr{R}_{ij}$-diagram M for w. If M has one region then syllables are single letters and the result holds. If M has more than one region then M has two disjoint strips S_1 and S_2. By Lemma 4 each strip S_h has a region E_h both of whose separating vertices are on the outside boundary of S_h. Let $s_h=\varphi(\partial E_h \cap \partial M)$. If either s_1 or s_2 is contained in w_2 the result follows by induction in the following way. Suppose that s_2 is contained in w_2. Let M' be the diagram obtained by removing the region E_2 from M. Then M' has boundary label $w_1 w_2'$ where w_2' is obtained from w_2 by replacing the subword s_2 by t_2, where the boundary label on E_2 is $s_2 t_2^{-1}$. Since both separating vertices of E_2 are on ∂M, $|t_2| \leqq |s_2|$. By induction $|w_1| <* |w_2'| \leqq |w_2|$.

The general case follows from the special case by "moving w_1 off one of the E_h" in the following way. Now $\|w_1\| <* m_{ij}$, each $\|s_h\| \geqq m_{ij}$, and the label on $(\partial S_h \cap \partial M)$ has at least $m_{ij}+1$ syllables by Lemma 5. Certainly then, w_1 cannot contain both the first letter of s_1 and the last letter of s_2. For definiteness, suppose that w_1 contains k letters of s_1 but not the first letter. Suppose that one begins reading w_1 at the vertex v_1 and finishes reading w_1 at the vertex v_2.

The k-th boundary vertex, say v_1', encountered after v_1 is the last vertex in $\partial E_1 \cap \partial M$. Let v_2' be the k-th vertex encountered after v_2. Then M has a boundary cycle with label $w_1' w_2'$ where w_1' is the label on the boundary path from v_1' to v_2'. Thus w_1 and w_2 have been shifted by length k so $|w_1'|=|w_1|$, $|w_2'| =|w_2|$ and s_1 is not a subword of w_1'. Hence $|w_1'| <* |w_2'|$ by the argument above and the result follows. □

5. Artin Groups on More than Two Generators

Now we are ready to describe our method for studying an Artin group G. If $i \neq j$ let $\mathscr{R}_{ij}$ be the set of all non-trivial words that are cyclically reduced in the free group sense and that are equal to the identity in G_{ij}. (An element $r \in \mathscr{R}_{ij}$ will be called a relator of *type* (i,j).) Let $\mathscr{R} = \bigcup_{i,j} \mathscr{R}_{ij}$ be the union of the $\mathscr{R}_{ij}$ for all pairs $i,j \in I$ with $i \neq j$. Then, of course, the Artin group G has the infinitely related presentation $\langle A; \mathscr{R} \rangle$. We will often note that parallel results are true for Coxeter groups. For Coxeter groups, $\mathscr{R}_{ij}$ will be taken to be the symmetrized set determined by r_{ij}. The reason for this seemingly more complicated presentation for Artin groups is given by the following lemma.

Lemma 8. *If M is a minimal connected, simply connected $\mathscr{R}$-diagram for an Artin or Coxeter group such that $m_{ij} \geqq m$ for all $i,j \in I$ then M satisfies $C(2m)$.*

Proof. We prove the lemma in the case in which the boundary word of M is cyclically reduced. The general case is easily obtained from this case by elimination of spikes and adjacent edges with inverse labels. First we show that if e is an interior edge of M lying on two distinct regions D_1 and D_2 of M then its label $\varphi(e)$ has the form a_i^n for some generator a_i. For if $\varphi(e)$ involves two distinct generators a_i and a_j then the labels on D_1 and D_2 are both of type (i,j). But we claim that this cannot happen if M is a minimal diagram. First note that removing the edge e combines D_1 and D_2 into a single region D whose boundary label u is a word on the generators of G_{ij}. Now let u' be a cyclically reduced conjugate of u and let D' be a region with boundary label u'. We construct a diagram M' with the same boundary label as that of M and with D replaced by D' by identifying adjacent pairs of edges with inverse labels in $M - D$ and inserting D' in the resulting diagram. Thus we have contradicted the minimality of M. Hence the label on an interior edge separating two regions has syllable length one. But, by Lemma 6, each $r \in \mathscr{R}$ satisfies $\|r\| \geqq 2m$ so, if M has no self-identified region, the conclusion of the lemma holds.

Next, exactly as in the proof of Lemma 1, we show that there are no self-identified regions in M. For suppose that D is a self-identified region. Then there is a loop η of ∂D bounding a submap K of M that contains at least one region. Among all self-identified regions, choose one such that K has as few regions as possible. Then K has no self-identified regions and thus satisfies $C(2m)$ as shown in the previous paragraph. But K has some Dehn region D_1 whence $\partial(D_1)$ has three consecutive syllables in common with the label on ∂D. But this means the labels on D_1 and D have the same type, contradicting minimality and thus we see that M can have no self-identified region. □

Corollary 2. *Every minimal connected, simply connected $\mathscr{R}$-diagram for an Artin or Coxeter group of large type satisfies $C(6)$. Every such diagram for an Artin or Coxeter group of extra-large type satisfies $C(8)$.*

If M is a $C(8)$-diagram without vertices of degree one then M has a Dehn region D. Let s be the label on $\partial D \cap \partial M$, s is a subword of the boundary label w of M, and let sb be the boundary label of D. Since the label on an interior edge cannot exceed one syllable, $\|s\| \geqq 5$ and $\|b\| \leqq 3$. Lemma 7 implies that

$|b|<|s|$ so replacing s by b in w yields a shorter word w'. Thus Dehn's algorithm with respect to the set $\mathcal{R}$ solves the word problem for G. When $sb\in\mathcal{R}$ with $|b|<|s|$, replacement of s by b in w and free reduction of the resulting word will be called an *$\mathcal{R}$-reduction* of w. A word w is called *$\mathcal{R}$-reduced* if no $\mathcal{R}$-reduction is applicable to w; and w is called *cyclically $\mathcal{R}$-reduced* if no $\mathcal{R}$-reduction is applicable to any cyclic conjugate of w.

6. Artin Groups and Coxeter Groups of Extra-Large Type

In the results to be established in this section there is little difference between the cases of Artin and Coxeter groups - the geometry of the appropriate cancellation diagrams is essentially the same. Let $\langle A;\mathcal{R}\rangle$ denote an Artin or Coxeter group with generating set $A=\{a_i\colon i\in I\}$. For an Artin group G, let $\mathcal{R}=\bigcup_{i,j}\mathcal{R}_{ij}$ be as above. For uniformity of notation, if G is a Coxeter group, now let $\mathcal{R}=\{r_{i,j}\colon i\neq j\}$. If $J\subseteq I$ let $A_J=\{a_j\colon j\in J\}$, let $\mathcal{R}_J=\{r\mid r\in\mathcal{R}_{ij};\ i,j\in J\}$ and let G_J be the group with presentation $\langle A_J;\mathcal{R}_J\rangle$. As usual, we assume that there are no interior vertices of degree two in $\mathcal{R}$-diagrams. In a minimal simply connected $\mathcal{R}$-diagram, M, the label on an interior edge cannot involve two distinct generators. It is convenient to add boundary vertices of degree two so that the label on a boundary edge of M involves only one generator. We begin by considering the embedding of G_J in G.

Lemma 9. *Let G be an Artin or Coxeter group of large type. Let w be a word on A_J. If $w=1$ in G then $w=1$ in G_J.*

Proof. Consider the set J as fixed. If the lemma were false we could pick among the words w on A_J such that $w=1$ in G but $w\neq 1$ in G_J one that was the boundary label of an $\mathcal{R}$-diagram with as few regions as possible. Let M be such a smallest diagram for w. We may assume that M has no vertices of degree one since the label on a spike is equal to the identity in both G and G_J. Since G is of large type, M satisfies $C(6)$ and thus there must be a Dehn region D in M. Since the label of $\partial D\cap\partial M$ has at least three syllables, both generators in the label of ∂D occur in w, hence D is labelled by some $r\in\mathcal{R}_J$. But this yields a contradiction since removing D from M produces a smaller diagram with boundary label equal to w both in G and in G_J. $\square$

Corollary 3. *Each group G_J is embedded in G by the obvious map $a_j\to a_j$.*

Lemma 10. *Let G be an Artin or Coxeter group of extra-large type. If v is an $\mathcal{R}$-reduced word that represents an element of G_J then v is a word on A_J.*

Proof. Assume the lemma false, and among all counterexamples choose v as short as possible. Then v is freely reduced and neither begins nor ends with A_J-symbols. Thus if u_0 is a freely reduced word on A_J with $vu_0=1$ then vu_0 is cyclically reduced and there is a connected, simply connected $\mathcal{R}$-diagram without vertices of degree 1 and boundary label vu_0. The connected, simply connected $\mathcal{R}$-diagrams without vertices of degree 1 and with boundary label vw where w is a word on A_J is thus non-empty.

Choose M from this set with fewest number of regions and suppose that the boundary label of M is uv. Since M has at least one region, M has a boundary cycle $\alpha\beta$ where α and β have labels u and v respectively. If M consisted of a single region then u would have to be a single syllable and so v would not be $\mathscr{R}$-reduced. Thus M has a Dehn region, D. If $\sigma=\partial D\cap\partial M$ contained only edges of β then v would not be $\mathscr{R}$-reduced. Thus σ must contain edges of both α and β (by minimality of number of regions). But now, using Result III we see that M must have exactly two Dehn regions, each of interior degree one. But this is impossible since v neither begins nor ends with A_J-symbols and the boundary label of the Dehn regions must consist of A_J-symbols since they contain subwords of u of length at least two. □

To prove Theorem 1 we note the following. Given a word w we can effectively calculate an $\mathscr{R}$-reduced word v with $v=w$ in G. Now v does not contain any generators not contained in w and $v\in G_J$ only if v is a word on A_J. □

Theorem 2. *An Artin group of extra-large type is torsion-free.*

Proof. Suppose $w\in G$ and u is a cyclically $\mathscr{R}$-reduced conjugate of w. If u is non-empty, suppose that u has finite order n. We may assume that u involves at least three generators. For if u involves only two generators, say a_i and a_j then $u^n=1$ in the one relator group G_{ij}. But a one-relator group whose relator is not a proper power is torsion-free.

Let M be a minimal $\mathscr{R}$-diagram for u^n. Since u involves more than two generators, M has more than one region. Thus M has a Dehn region D. The label on $\partial D\cap\partial M$ involves only two generators, hence is a subword of a cyclic conjugate of u. But replacing s by the base b of D is an $\mathscr{R}$-reduction, contradicting the assumption that u is cyclically $\mathscr{R}$-reduced. □

Theorem 3. *Let G be an Artin group of extra-large type. Let H be the subgroup of G generated by the set $S_I=\{a_i^2\mid i\in I\}$. Then H is freely generated by S_I.*

Proof. If H were not freely generated by S_I then some word w all of whose syllables have length at least two would be equal to 1 in G. Let M be a minimal $\mathscr{R}$-diagram for such a word w. M must have a Dehn region D.

Now D itself, as a $\mathscr{R}_{ij}$-diagram either consists of a single region or has strips as shown in Lemma 1. Let s be the label on $\partial D\cap\partial M$. If D consists of a single region then s has at least five syllables of length one, at most two of which (at the ends of s) can be part of longer syllables, contradicting the definition of w. Suppose that D contains at least two regions. Then consider the strips of D. Since $m_{ij}\geqq 4$, a singleton strip must have at least five consecutive single generator syllables. By Lemma 4 at least one region of a compound strip must have both separating vertices on the boundary and there are four single generator syllables between each pair of separating vertices. Now we must take into consideration the fact that the end syllables of these sequences may join with syllables on the same generator on the labels of neighboring regions to see whether all of the remaining single generator syllables on the label of D can be among the three syllables on interior edges. Simple counting shows us that this is impossible if D has a singleton strip.

Suppose there are no singleton strips. If there were only two regions of D with both separating vertices on the boundary and these were adjacent with common adjacent syllable there would be no guarantee of a single generator syllable on the boundary, but the existence of four strips rules this out. Thus the theorem is proved. □

7. The Conjugacy Problem

We next investigate the conjugacy problem. Here the appropriate cancellation diagram is the annulus. An *annular* $\mathscr{R}$-diagram M is an $\mathscr{R}$-diagram whose complement has exactly two components. If u and v are cyclically reduced words on A_I, not conjugate in F, and neither is equal to the identity in G then u and v are conjugate in G if and only if there is an annular $\mathscr{R}$-diagram whose outer boundary is labelled by u and whose inner boundary is labelled by v. (We depart from the convention of [10] in that we read the labels on both the outer and inner boundaries of M counterclockwise.) Such an $\mathscr{R}$-diagram is called a *conjugacy* diagram for u and v. A conjugacy diagram M for u and v is *minimal* if it contains a number of regions that is smallest among all conjugacy diagrams for u and v.

In the study of the word problem we could be sure that minimal connected simply connected $\mathscr{R}$-diagrams do not contain any self-identified regions. Note, however, that self-identified regions certainly can appear in conjugacy diagrams for Artin and Coxeter groups. For example, if $a_1a_2a_1a_2a_1\bar{a}_2\bar{a}_1\bar{a}_2\bar{a}_1\bar{a}_2$ is a defining relator then a_1 is conjugate to a_2 and the single relator diagram in Fig. 3 is clearly a minimal conjugacy diagram for a_1 and a_2. This example also illustrates the fact that a "large" part of the boundary of a region may be involved in a self-identification. We will consider the cases of conjugacy diagrams with and without self-identified regions separately. Under the hypothesis that the group considered is of extra-large type we shall prove a very strong structure theorem for minimal conjugacy diagrams labelled by cyclically $\mathscr{R}$-reduced words. Our proof is based on the following structure theorem from small cancellation theory.

Result V. Let M be a reduced annular $\mathscr{R}$-diagram without vertices of degree one satisfying the following hypotheses.

(i) If D is an almost interior region then $d(D) \geqq 7$.

(ii) If D is a region of M and ∂D intersects only one boundary of M then $i(D) \geqq 4$.

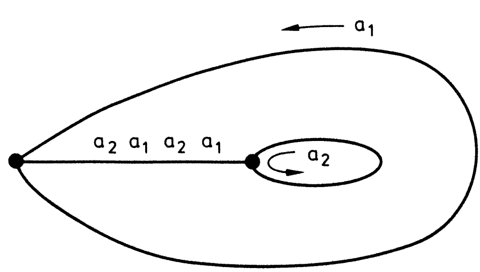


Fig. 3

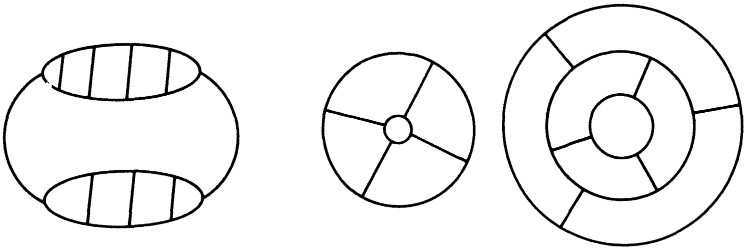


Fig. 4

Then M satisfies one of the following conditions:

(a) Every region D has edges on both boundaries of M and the intersection of ∂D with the inner (or outer) boundary of M is a consecutive part of that boundary. Furthermore, $i(D) \leqq 2$.

(b) Every region D of M has edges on only one boundary of M and $i(D) = 4$. Further, every boundary edge of M is on the boundary of a region of M. (In other words, the inner and outer boundaries have no edges in common.)

In short, Result V says that M "looks like" one of the cases illustrated in Fig. 4.

Suppose that G is of extra-large type. We now show that if u and v are non-trivial cyclically $\mathscr{R}$-reduced words and M is a minimal conjugacy diagram for u and v with no self-identified regions then M satisfies the hypotheses of Result V. First, the minimality of M implies that M is reduced, exactly as in Lemmas 2 and 9. Hypothesis (i) is satisfied since the fact that G is of extra-large type implies that a reduced $\mathscr{R}$-diagram satisfies $C(8)$. To verify hypothesis (ii), let D be a region of M that intersects only one boundary of M, say the outer boundary σ labelled by u. First suppose that $\partial D \cap \sigma$ is a consecutive part of σ. If $i(D) \leqq 3$ then D has a boundary label st where $\|t\| \leqq 3$ and thus $\|s\| \geqq 5$ and s is the label on $\partial D \cap \sigma$. This contradicts the assumption that u is cyclically $\mathscr{R}$-reduced. Next, we show that $\partial D \cap \sigma$ must be a consecutive part of σ. For, if not, then $M - \bar{D}$ contains a simply connected component C that has at least one region. Then $J = C \cup \bar{D}$ is a connected, simply connected $\mathscr{R}$-diagram with more than one region. Hence J has at least two regions that are simple boundary regions of J and have interior degrees less than or equal to three in J. One such region E is different from D. Thus $\partial E \cap \sigma$ is a consecutive part of σ and $i(E) \leqq 3$ in M. As above, this contradicts the assumption that u is cyclically $\mathscr{R}$-reduced.

The assumption that G is of extra-large type implies $C(8)$ on minimal $\mathscr{R}$-diagrams. This fact, together with the form of the defining relators allows us to eliminate the situation that arises from case (b) of the conclusion of Result V by considering words that are reduced in a special sense. Consider the situation of Fig. 5 below. Suppose $\mathscr{R}$ contains relators $r_1 = s_1 c_1 b_1 t_1$ and $r_2 = s_2 c_2 b_2 c_1^{-1}$ with both $\|c_1 b_1 t_1\| \leqq 4$ and $\|c_2 b_2 c_1^{-1}\| \leqq 4$. Then the replacement in w of a subword $s = s_1 s_2$ by $c_2 b_2 b_1 t_1$ is called a *special $\mathscr{R}$-reduction of w.*

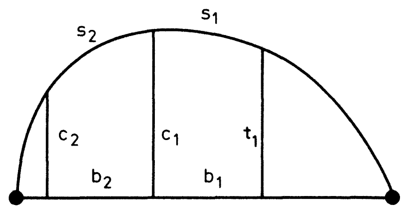


Fig. 5

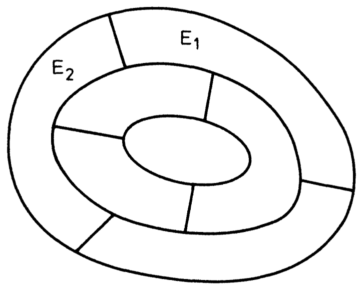


Fig. 6

Since G is of extra-large type, $|s_1| \geqq |c_1 b_1 t_1|$, $|s_2| \geqq |c_2 b_2 c_1^{-1}|$, by Lemma 7. Thus $|s_1 s_2| \geqq |c_1 b_1 t_1| + |c_2 b_2 c_1^{-1}| > |b_1 t_1| + |c_2 b_2| = |c_2 b_2 b_1 t_1|$, so a special $\mathscr{R}$-reduction will shorten w. We say that w is *specially cyclically $\mathscr{R}$-reduced* if every cyclic permutation w^* of w is $\mathscr{R}$-reduced and furthermore no special $\mathscr{R}$-reduction becomes shortened the length of w^*.

Lemma 11. *Let u and v be specially cyclically $\mathscr{R}$-reduced words that are conjugate in a group G of extra-large type. If M is a minimal conjugacy diagram for u and v that contains no self-identified regions then for every region D of M, ∂D has edges on both boundaries of M, the intersection of ∂D with each boundary of M is a consecutive part of that boundary and $i(D) \leqq 2$.*

Proof. The lemma states that M satisfies conclusion (a) of Result V. We show that conclusion (b) cannot hold (see Fig. 6).

If conclusion (b) holds then M consists of two bands of regions and each region has interior degree four. Let E_1 and E_2 be two consecutive regions in the outer band of M. Let M' be the diagram obtained by removing regions E_1 and E_2 from M. Then M' has boundary labels u' and v where u' is conjugate to u. Furthermore u' is obtainable from u by a special $\mathscr{R}$-reduction, contradicting the assumption that u is specially cyclically $\mathscr{R}$-reduced. □

Suppose that M is a minimal conjugacy diagram for u and v that satisfies conclusion (a) of Result V. If e is any interior edge of M then e lies on the boundaries of two distinct regions D_1 and D_2 of M. The label c on e cannot involve two distinct generators for otherwise D_1 and D_2 would be regions of the same type. Hence $c = a_i^k$ for some i. From the diagram obtained by cutting M open along the edge e we obtain the result that there are cyclic per-

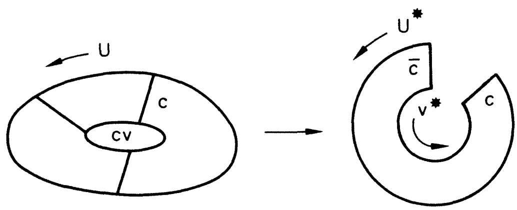


Fig. 7

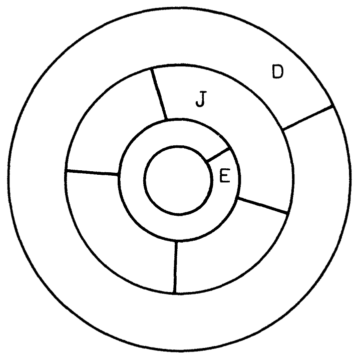


Fig. 8

mutations u^* and v^* of u and v respectively that are conjugate by a power of a generator. (In the Coxeter case, $c=1$ or $c=a_i$.) □

We now determine the structure of conjugacy diagrams that contain self-identified regions.

Lemma 12. *Let G be of extra-large type. Let u and v be non-trivial specially cyclically $\mathscr{R}$-reduced words of G that are conjugate in G and let M be a minimal conjugacy diagram for u and v. If M contains a self-identified region then all regions of M are self-identified.*

Proof. Suppose M contains both self-identified regions and non-self-identified regions. First, suppose that M has distinct self-identified regions D and E, say with E interior to D, such that some non-self-identified region of M lies between D and E (see Fig. 8). We may choose D and E so that no self-identified region lies between them. Let J be the submap of M that lies between the inner part of the boundary of D and the outer part of the boundary of E. Then J is a non-empty annular $\mathscr{R}$-diagram with no self-identified region. If K is a region of J and $\partial K \cap \partial D$ is a consecutive part of ∂D then the syllable length of the label $\varphi(\partial K \cap \partial D)$ cannot be greater than one for otherwise J would contain a region of the same type as D having a common boundary with D. This would violate the minimality of M. The same statement can be made replacing D by E. Thus J satisfies the hypothesis of Result V. Hence J must contain a region having two consecutive syllables on its outer boundary and two consecutive syllables on its inner boundary, again contradicting the minimality of M. Therefore no such J can exist.

Thus the hypotheses on M require that it contain a self-identified region D such that there are regions between D and one of the boundaries of M (say the outer boundary for definiteness) but none of these are self-identified. Let J be the submap of M bounded by the outer boundaries of D and M. J is a non-empty annular $\mathscr{R}$-diagram. Since u is specially cyclically $\mathscr{R}$-reduced and D is self-identified, J satisfies the hypotheses of Result V and hence the conclusion of Result V shows that M cannot be minimal. Thus M cannot contain both self-identified and non-self-identified regions. $\square$

Let M be a minimal conjugacy diagram in which all regions are self-identified. Let D and E be distinct adjacent regions of M. The syllable length of the label s on $\partial D \cap \partial E$ must be one, otherwise D and E would be of the same type, violating minimality. Thus s is a power of some generator a_i.

Consider the Coxeter group case. Here s is a single generator. The labels on the inner and outer boundaries must also be single generators. For suppose that u is the label on the outer boundary σ of M. Then σ is part of the boundary of a self-identified region D. Let e be the interior edge of M with D lying on both sides of e, say with e incident to a vertex p of σ. Because of the form of the defining relators of a Coxeter group G, if one reads to outer boundary label u^* of M that begins and ends at p, the same letter must be both first and last letter of u^*. If u^* contained any other letters then $u^* = a_i \tilde{u} a_i = a_i^{-1} \tilde{u} a_i$ would not be cyclically reduced, hence u is a single letter. The same observation applied to the inner boundary of G. Thus, in the Coxeter case, a self-identified conjugacy diagram establishes the conjugacy of two distinct generators of G. For m_{ij} even a self-identified region labelled by $(a_i a_j)^{m_{ij}}$ would have both inner and outer boundary labelled by the same generator. This cannot occur in a minimal diagram. Thus self-identified regions that occur in a minimal conjugacy diagram are labelled by relators r_{ij} where m_{ij} is odd. Let Γ be the Coxeter graph of the presentation. The vertex set of Γ is I. If $i,j \in I$ and $m_{ij} < \infty$ then the vertices i and j are connected by an edge with label m_{ij}. A path α in Γ is *odd* if all the edges in α have odd labels. Two generators a_i and a_j are thus conjugate in G if and only if i and j are connected by an odd path in Γ.

We summarize our results concerning Coxeter groups of extra-large type.

Theorem 4′. *Let G be a Coxeter group of extra-large type. Let u and v be specially cyclically $\mathscr{R}$-reduced words on the generators of G. Then u and v are conjugate in G if and only if*

(i) *both u and v have length greater than one and there exist cyclic permutations u^* and v^* of u and v respectively such that $u^* = v^*$ in G or there is a generator a_i occurring in both u and v such that $u^* = a_i v^* a_i$; or*

(ii) *both u and v are single generators whose indices are connected by an odd path in the Coxeter graph Γ of G.*

Next, we turn to Artin groups and first consider the case of minimal conjugacy diagrams M with self-identified regions. First of all, M might have just one region, say of type (i,j). Since $\mathscr{R}_{ij}$ contains all the relators of G_{ij} this situation corresponds to the case in which the boundary labels u and v are conjugate in G_{ij}. Result IV says that conjugacy in G_{ij} is decidable.

Next, suppose M has more than one region. Let r_{ij} be the defining relator for G_{ij}. Note that if m_{ij} is even then both generators have exponent sum zero in r_{ij}, while if m_{ij} is odd then the exponent sum on a_i in r_{ij} is the negative of that of a_j. This property extends to all relators in $\mathscr{R}_{ij}$. Therefore, if a word w in G_{ij} is a conjugate of either a_i^n or a_j^n for some n, there are at most two possibilities and these are completely determined by the spelling of w. Furthermore, for fixed n, whether or not w is conjugate to a_i^n in G_{ij} is decidable. In particular if w is also a power of a generator, say a_i^k then for m_{ij} even, the only possibility is $a_i^k \sim a_j^k$ (we write $c \sim d$ for the phrase "c is conjugate to d") while if m_{ij} is odd the only possibility is $a_i^k \sim a_j^{-k}$. If u and v are the boundary labels of M then u and v must be conjugate in the following way. First, u is a word of some G_{ij} and is conjugate to a power of a generator, say $u \sim a_i^n$ while v is a word of some G_{kl} that is conjugate to (say) a_k^m. Further, if $n=m$ there must be an odd path of even length connecting i and k in the Coxeter graph Γ while if $n=-m$ there is an odd path of odd length connecting i and k in Γ.

Now consider a minimal conjugacy diagram M for specially cyclically-$\mathscr{R}$-reduced words u and v that has no self-identified regions, and so has the structure described by Lemma 11. Each region D_k of M is labelled by an element $r_{(i,j)}$ of some $\mathscr{R}_{ij}$. The structure of M guarantees that if a relator of type (i,j) occurs then a_i and a_j occur in both u and v. Thus the types of relators that occur can be determined from u and v. We now consider the "fine structure" of M. Since $r_{(i,j)}=1$ in G_{ij} there is a minimal connected, simply connected $\mathscr{R}_{ij}$-diagram D_k with boundary label $r_{(i,j)}$. We think of the region D_k as being replaced by the diagram D_k. This may pinch together parts of M but we obtain an annular diagram M' in which each region is labelled by a single defining relator of G.

Lemma 13. *The diagram M' satisfies $C(4)$ and $T(4)$.*

Proof. That M' satisfies $C(4)$ is immediate from Lemma 4 and the fact that relators of different types can have only a single generator in common. We must show that M has no interior vertex of degree three. Lemma 4 shows that an interior vertex of degree three could only occur in the situation (cf. Fig. 9) where some part of the common boundary of D_1 and D_2 in the diagram M has two regions, say E_1 and E_2, of D_1 incident at v and only one region K of D_2 incident at v.

Generators a_i and a_j occur alternately in defining relators, but this situation implies that the boundary label on K contains two consecutive occurrences of the same generator, a contradiction. □

Given u and v, whether or not u and v are conjugate in G can now be effectively determined by (say) first performing the tests required to determine

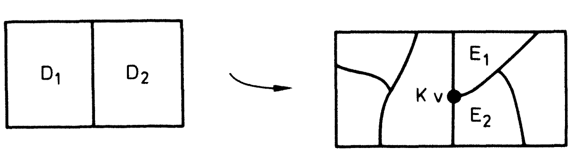


Fig. 9

if the two words are conjugate via a diagram with self-identified regions as suggested above. If they are not, then we need only test whether there is an annulus satisfying the conclusion of Lemma 11 with u and v as boundary labels. The argument above shows that these tests are effective. Thus we have the following theorem.

Theorem 4″. *Let G be an Artin group of extra-large type. Then G has solvable conjugacy problem. Furthermore, if u and v are specially cyclically reduced words involving at least three generators then u and v are conjugate in G if and only if there exist cyclic permutations u^* and v^* of u and v respectively, a generator a_i occurring in both u and v and an exponent n such that $u^* = a_i^n v^* a_i^{-n}$ in G.*

One might note that we appealed to Result IV rather than providing a bound on the exponent n in the statement of Theorem 4″. The techniques of this paper do not naturally lead to such a bound but one can be found by using a special case of technical results used to extend this theorem to groups of large type.

Remark. The embedding of the group G_{ij} in an Artin or Coxeter group G need not be a Frattini embedding; that is, elements that are not conjugate in G_{ij} may be conjugate in G. For example, if m_{ij} is even while m_{ik} and m_{jk} are odd then a_i is conjugate to a_j in G but not in G_{ij}. On the other hand, our analysis shows that if all the m_{ij} are even or infinity then the embedding of each G_{ij} in G is a Frattini embedding.

References

1. Birman, J.: Braids, Links, and Mapping Class Groups. Princeton, N.J.: Princeton University Press 1974
2. Bourbaki, N.: Groupes et algèbres de Lie, Chapitres 4, 5, et 6, Eléments de Mathématique XXXIV. Paris: Hermann 1968
3. Brieskorn, E.: Die Fundamentalgruppe des Raumes der Regulären Orbits einer endlichen komplexen Spielsgruppe. Invent. Math. **12**, 57–61 (1971)
4. Brieskorn, E., Saito, K.: Artin-gruppen und Coxeter-gruppen. Invent. Math. **17**, 245–271 (1972)
5. Coxeter, H.S.M.: The complete enumeration of finite groups of the form $R_i^2 = (R_i R_j)^{k_{ij}} = 1$. J. London Math. Soc. **10**, 21–25 (1935)
6. Deligne, P.: Les immeubles des groupes de tresses généralisés. Invent. Math. **17**, 273–302 (1972)
7. Garside, F.A.: The braid group and other groups. Quart. J. Math. Oxford, 2 Ser. **20**, 235–254 (1969)
8. Len, V.Y.: Artin's braids and their connections with groups and spaces. Etoge Nauke e Texneke. Algebra, Topologeya, Geometreya, Tom 17
9. Lyndon, R.C.: On Dehn's algorithm. Math. Ann. **166**, 208–228 (1966)
10. Lyndon, R.C., Schupp, P.E.: Combinatorial Group Theory. Ergebnisse der Mathematik und ihrer Grenzgebiete. Berlin: Springer 1977
11. Schupp, P.E.: A survey of small cancellation theory. Word Problems, pp. 569–589. Amsterdam: North Holland 1973
12. Tits, J.: Le problème des mots dans les groupes de Coxeter. Instituto Nazionale di Alta Mathematica, Symposia Mathematica 1, 175–185 (1968)
13. Tits, J.: Normalisateurs de tores I. Groupes de Coxeter étendus. J. Algebra 4, 96–116 (1966)

Oblatum 16-VIII-1981 & 29-IV-1982

Invent. math. 72, 221–239 (1983)

Inventiones mathematicae

The Second Homology Group of the Mapping Class Group of an Orientable Surface

John Harer*

Department of Mathematics, Columbia University, New York, NY 10027, USA

Let F be an oriented surface of genus g with r boundary components and n distinguished points. The mapping class group $\Gamma = \Gamma(F)$ is $\pi_0(\mathrm{Diff}^+ F)$ where $\mathrm{Diff}^+(F)$ is the topological group of orientation preserving diffeomorphisms of F which fix the n points and restrict to the identity on ∂F. It is known [9] that $H_1(\Gamma)=0$ for $g \geqq 3$. In this paper we shall prove the

Theorem.

$$H_2(\Gamma) \cong \begin{cases} \mathbb{Z}^{n+1} & g \geqq 5,\ r+n>0 \\ \mathbb{Z} \oplus \mathbb{Z}/(2g-2) & g \geqq 5,\ r=n=0. \end{cases}$$

As an immediate corollary we have the proof of a conjecture of Mumford [7]. To state this let $\mathfrak{M}_g^n$ be the moduli space for curves of genus g with n punctures ($r=0$). Γ acts properly discontinuously on the Teichmuller space $\mathcal{T}_g^n \cong \mathbb{C}^{3g-3+n}$ with quotient $\mathfrak{M}_g^n$. Furthermore, the stabilizer of any point is finite, so

$$H_i(\Gamma; \mathbb{Q}) \cong H_i(\mathfrak{M}_g^n; \mathbb{Q}) \quad \text{for all } i.$$

The codimension of the subset of $\mathcal{T}_g^n$ on which Γ fails to act freely (the curves with automorphisms) increases with g, so in fact

$$H_i(\Gamma; \mathbb{Z}) \cong H_i(\mathfrak{M}_g^n; \mathbb{Z}), \qquad g \gg i.$$

In [7] Mumford shows that the Picard group $\mathrm{Pic}(\mathcal{M})$ is isomorphic to $H^2(\Gamma; \mathbb{Z})$ and conjectures the latter is rank one, $g \geqq 3$. We prove this below for $g \geqq 5$.

Another interpretation of this theorem may be obtained by identifying $H_2(\Gamma)$ as bordism classes of fiber bundles $F \to W^4 \to T$ where T is a closed oriented surface (Sect. 0). When F is closed every such bundle is bordant to $F \to W' \to T'$, a bundle admitting a section $s\colon T' \to W'$. The theorem then says that

* This material is based upon work supported by the National Science Foundation under Grant No. NSF MCS 80-02325

$$\begin{pmatrix} F \to W \\ \downarrow \\ T \end{pmatrix} = \partial \begin{pmatrix} F \to M^5 \\ \downarrow \\ Y^3 \end{pmatrix}$$

if and only if W has signature 0 and the self-intersection number of $s(T')$ is divisible by $2g-2$.

The author would like to thank Karen Vogtmann for helpful discussions and the referee for invaluable suggestions.

Section 0. Outline of the Proof

When $g \geqq 3$ there are $n+1$ homomorphisms

$$S_0, \ldots, S_n \colon H_2(\Gamma) \to \mathbb{Z},$$

the components of a map φ to $\mathbb{Z}^{n+1}$ with image $\varphi = 4\mathbb{Z} \oplus \mathbb{Z}^n$. To describe these let Λ denote the topological group $\mathrm{Diff}^+(F)$, $B\Lambda$ its classifying space with $E\Lambda \to B\Lambda$ its universal covering. When $g \geqq 2$ each component of Λ is contractible [3] and $H_2(\Gamma) \cong H_2(B\Lambda)$. The latter is isomorphic to the bordism group $\Omega_2(B\Lambda)$. This means that every 2-cycle ξ on Γ may be represented by a map $T \xrightarrow{\xi} B\Lambda$, T some oriented closed surface, and $[\xi] = [\xi']$ in $H_2(\Gamma)$ iff $\xi \amalg -\xi'$ extends to $M^3 \to B\Lambda$, M^3 compact oriented with $\partial M = T \amalg -T'$.

The map ξ yields, by pulling back from $F \to E\Lambda \times_\Gamma F \to B\Lambda$, a fiber bundle $F \to W^4 \xrightarrow{\xi} T$. Set

$$S_0(\xi) = \text{signature}(W^4).$$

The monodromy group of ξ lies in Λ, whose elements fix the n distinguished points on F, so ξ has n canonical sections $s_i \colon T \to W$. Set

$$S_i(\xi) = [s_i(T)]^2,$$

the self-intersection of $s_i(T)$ in W. These are well defined homomorphisms by the above discussion. Meyer [6] proves that $\mathrm{im}(S_0) = 4\mathbb{Z}$. We will show surjectivity of the other S_i's and independence of all $n+1$ S_i's in Sect. 4. We may then state our theorem more explicitly by saying that φ is an isomorphism for $r+n>0$ and has kernel $\mathbb{Z}/|X(F)|$ for $r=n=0$.

To compute $H_2(\Gamma)$ completely it will be necessary to construct a cellular action of Γ on a simply-connected 3-complex Y_3. This complex has its origins in [5].

A well known spectral sequence technique then allows us to find $H_2(\Gamma)$ in terms of $H_2(Y_3/\Gamma)$ and the lower homology groups of the stabilizers of the cells of Y_3.

In Sect. 1 we show $H_1(\Gamma) \cong 0$ for $g \geqq 3$ and $H_1(\Gamma; H_1 F) = 0$, $g \geqq 4$. These facts will be needed in Sects. 3 and 4. In Sect. 2 we describe the complex Y_3 and prove it is simply connected. Section 3 computes $H_2(\Gamma)$ with $r \geqq 1$, $g \geqq 5$, $n=0$ directly from the action on Y_3. In Sect. 4 we use various short exact sequences to deal with the general cases and finish the proof.

Section 1. $H_1(\Gamma)$ and $H_1(\Gamma; H_1(F))$

During the course of our computation of $H_2(\Gamma)$ it will be necessary to know the following:

Lemma 1.1. $H_1(\Gamma)=0$ *for* $g\geqq 3$; r, n *arbitrary.*

Lemma 1.2. $H_1(\Gamma; \widetilde{H_1(F)})=0$ *for* $g\geqq 4$; r, n *arbitrary.*

Lemma 1.1 was first proven by Powell [9] for $r=n=0$.

Proof of 1.1. Let $C\subset F$ be a simple closed curve. The Dehn twist on C, τ_C, is the mapping class obtained by splitting F open at C and reglueing by a 360° twist to the right. Dehn [2] proved that mappings of this form generate Γ.

Let F_0 be a sphere with four disks removed, Γ_0 its mapping class group (recall ∂F_0 must remain fixed). Label its boundary components $C_0, \ldots, C_3$ and write τ_i for the Dehn twist on a circle in $F_0-\partial F_0$ parallel to C_i. Also write C_{ij} for the circle enclosing C_i and C_j, shown in Fig. 1, τ_{ij} for the twist on C_{ij}. The following relation in Γ_0 is easily verified

$$\tau_0\tau_1\tau_2\tau_3=\tau_{12}\tau_{13}\tau_{23}. \tag{$*$}$$

Since ∂F_0 remains fixed throughout, any embedding of F_0 in F will induce a relation in Γ.

As a first application of $(*)$ we wish to prove Γ is generated by Dehn twists on nonseparating curves in F. If $C\subset F$ is parallel to a component of ∂F and $g\geqq 2$ choose two disjoint simple closed curves $\alpha_1, \alpha_2\subset F$ with $F-\{\alpha_1, \alpha_2\}$ connected. It is easy to then find an embedding of F_0 in F with $C_1=\alpha_1$, $C_2=\alpha_2$ and $C_3=C$. All curves will be nonseparating in F except C, so τ_C is isotopic to a product of twists on nonseparating curves.

For arbitrary separating $C\subset F$, $g\geqq 3$, split F open on C and use the side of $F-C$ which has genus $\geqq 2$.

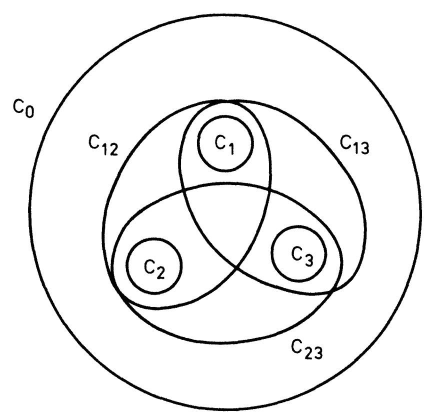


Fig. 1

Remark. For $g=2$, Γ is still generated by twists on nonseparating curves; the argument is only slightly more difficult.

We see immediately that $H_1(\Gamma)$ is cyclic because if C_2, C_1 are any two nonseparating curves, there exists $h\in \mathrm{Diff}^+(F)$ with $h(C_2)=C_1$. It follows that

$$\tau_{C_2}=h\tau_{C_1}h^{-1}$$

in Γ. To see $H_1(\Gamma)=0$ we notice that for $g\geqq 3$, there is an embedding of F_0 in F with all seven curves nonseparating. This completes the Proof of 1.1. □

Proof of 1.2. Construct a free $\mathbb{Z}\Gamma$ resolution $L_*\to\mathbb{Z}$ with $L_0=\mathbb{Z}\Gamma$, L_1 the free $Z\Gamma$ module on symbols $\hat{\gamma}$ for $\gamma\subset F$ a nonseparating curve, L_2 the free $\mathbb{Z}\Gamma$ module on symbols for relations, etc.; boundary maps are defined in the usual way. For each curve $C\subset F$ we choose an orientation to obtain $h_C\in H_1(F)$. If γ is nonseparating, there is a basis $\{C_i\}$ of $H_1(F)$ with $C_1=\gamma$, $\gamma\cap C_2=$ one point, and $\gamma\cap C_i=\emptyset$, $i\geqq 3$. Form F_i by splitting F along C_i, $i\neq 2$; since genus $(F_i)\geqq 3$, on F_i there is by Lemma 1.1 a relation R of the form

$$\tau_\gamma=\prod[\tau_{\gamma_j},\tau_{\delta_j}],\qquad \gamma_j,\delta_j\subset \mathrm{F}_i.$$

Since τ_{γ_j}, τ_{δ_j} and τ_γ act trivially on h_{C_i}

$$\partial_2(\hat{R}\otimes h_{C_i})=\hat{\gamma}\otimes h_{C_i},\qquad (i\neq 2).$$

To see that $\hat{\gamma}\otimes h_{C_2}$ bounds, choose orientations of C_1, C_2 so that $C_1\cdot C_2=1$; then, setting $\tau_i=\tau_{C_i}$,

$$\tau_{1*}(\mathrm{h}_{C_2})=\mathrm{h}_{C_1}+\mathrm{h}_{C_2}=\tau_{2*}^{-1}(\mathrm{h}_{C_1}).$$

Therefore

$$\begin{aligned}\hat{C}_1\otimes h_{C_2}&=\tau_1^{-1}(\hat{C}_1\otimes\tau_{1*}(h_{C_2}))\\&=\tau_1^{-1}\tau_2^{-1}(\hat{C}_1\otimes h_{C_1}).\end{aligned}$$

But $\hat{C}_1\otimes h_{C_1}=\partial_2(\hat{R}\otimes h_{C_1})$ by the above. □

Section 2. The Cut System Complex

Based on work of [5] we now construct the complex Y_3 mentioned in Sect. 0. Throughout this section we shall assume $n=0$, $r\geqq 1$.

A *cut system* on F is $\langle C_i\rangle$, the isotopy class of a collection of disjoint simple closed curves $C_1,\dots,C_g\subset F$ such that $F-(C_1\cup\dots\cup C_g)$ is connected and therefore planar. (We will often confuse a curve and its isotopy class.) There is no ordering of the circles and they are not oriented.

If C, C' are two isotopy classes set $I(C,C')$ equal to the minimum number of intersections (no signs) between any two representatives meeting transversely. When we replace $\langle C_i\rangle$ by $\langle C_i'\rangle$ where $I(C_i,C_i')=1$ and $C_j=C_j'$, $j\neq i$ we say $\langle C_i\rangle$ and $\langle C_i'\rangle$ differ by a *simple move*. We assume from now on that any circle omitted from the notation remains unchanged.

There are three main cycles of such moves

$$\begin{tikzcd}[column sep=small] \langle C_i\rangle \arrow[rr, dash] \arrow[dr, dash] & & \langle C_i'\rangle \arrow[dl, dash] \\ & \langle C_i''\rangle & \end{tikzcd} \tag{R$_1$}$$

$$\begin{tikzcd} \langle C_i, C_j\rangle \arrow[r, dash] \arrow[d, dash] & \langle C_i, C_j'\rangle \arrow[d, dash] \\ \langle C_i', C_j\rangle \arrow[r, dash] & \langle C_i', C_j'\rangle \end{tikzcd} \tag{R$_2$}$$

$$\begin{tikzcd}[column sep=small] & \langle C_i, C_j\rangle \arrow[r, dash] \arrow[dl, dash] & \langle C_j, C_j'\rangle \arrow[dr, dash] & \\ \langle C_i, C_i'\rangle \arrow[dr, dash] & & & \langle C_j, C\rangle. \arrow[dl, dash] \\ & \langle C_i, C\rangle \arrow[r, dash] & \langle C, C'\rangle & \end{tikzcd} \tag{R$_3$}$$

Each edge corresponds to a simple move. An illustration is shown in Fig. 2 although there is no restriction on $I(C', C_i')$, $I(C', C_j')$, or $I(C_i', C_j')$. Note that R_3 is different from [5] but part 2 of Lemma 1.7 of that paper shows that use of R_3 is equivalent.

Let X_0 be the 0-complex with one vertex for each cut system on F; X_1 the 1-complex obtained by attaching a 1-cell for each simple move; $X_2 = X_1$ with 2-cells attached for the cycles R_1, R_2 and R_3.

Theorem 2.1 [5]. *X_2 is connected and simply connected.*

From this they outline a presentation of Γ.

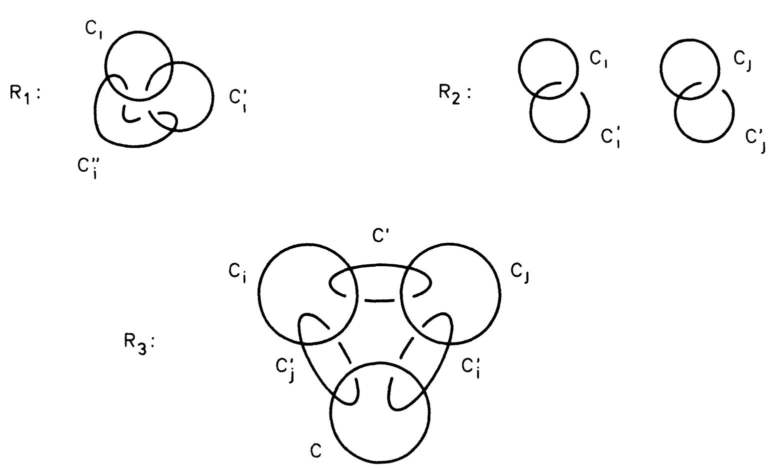


Fig. 2

Let $\varphi \cdot \langle C_i \rangle = \langle \varphi(C_i) \rangle$. Because all cells are determined by configurations of circles, extending linearly gives a natural action of Γ on X_2.

Our first task will be to describe a subcomplex $Y_2 \subset X_2$ with $Y_1 = X_1$, $\pi_1(Y_2) = 1$ and $\Gamma(Y_2) \subset Y_2$.

Consider type R_1 2-cells in X. Among such are those $\sigma(\gamma)$ corresponding to an R_1 cycle with $\{C_i, C_i', C_i''\} = \{\alpha_1, \beta_1, \gamma\}$ and $\{C_j, j \neq i\} = \{\alpha_j, j \neq 1\}$, where α_i and β_1 are the standard curves in Fig. 3. Under the action of Γ on X_2, every type R_1 2-cell of X_2 is identified with some $\sigma(\gamma)$, with γ ranging over a rather large but finite set of curves. For Y_2, include those 2-cells of X_2 in the Γ orbit of all $\sigma(\gamma)$ with γ among the $N = 2r + 2 + \binom{r-1}{2}$ curves of Fig. 4 (the parts of γ not drawn are straight arcs which do not link any handles).

Y_2 contains all 2-cells of X_2 of type R_2. For R_3, Y_2 contains all 2-cells in the Γ-orbit of a single 2-cell, the one corresponding to the R_3 cycle involving the standard cut system $\langle \alpha_i \rangle$ and the cycle of curves

$$(C_i, C', C_j, C_i', C, C_j') = (\alpha_1, \omega, \alpha_2, \beta_2, C_0, \beta_1)$$

shown in Fig. 3.

Clearly $\Gamma(Y_2) \subset Y_2$.

Theorem 2.2. *Y_2 is simply connected.*

Proof. Form X_3 by adding to X_2 the 3-cells of types Σ_1, Σ_2 and Σ_3 from Fig. 5.

(I) Consider first two 2-cells of type R_3 and the choice of the circle C_i'. $I(C_i', C) = I(C_i', C_j) = 1$, $I(C_i', C_k) = 0$, $k \neq j$.

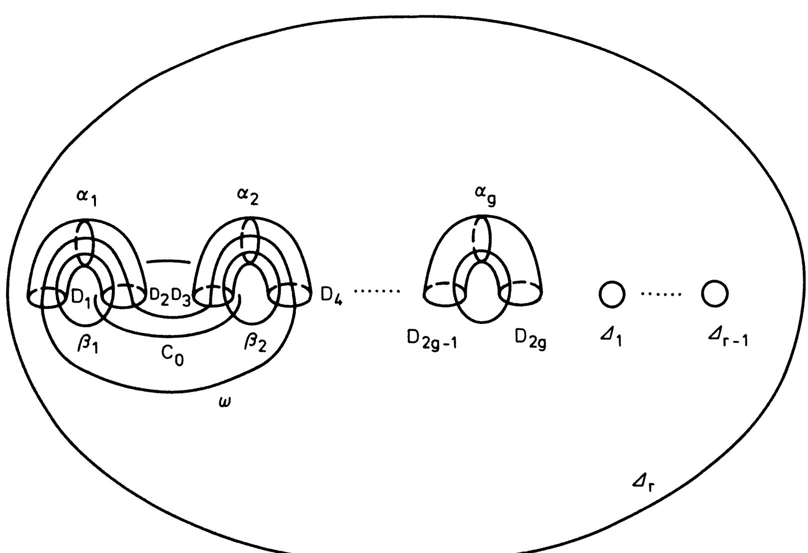


Fig. 3

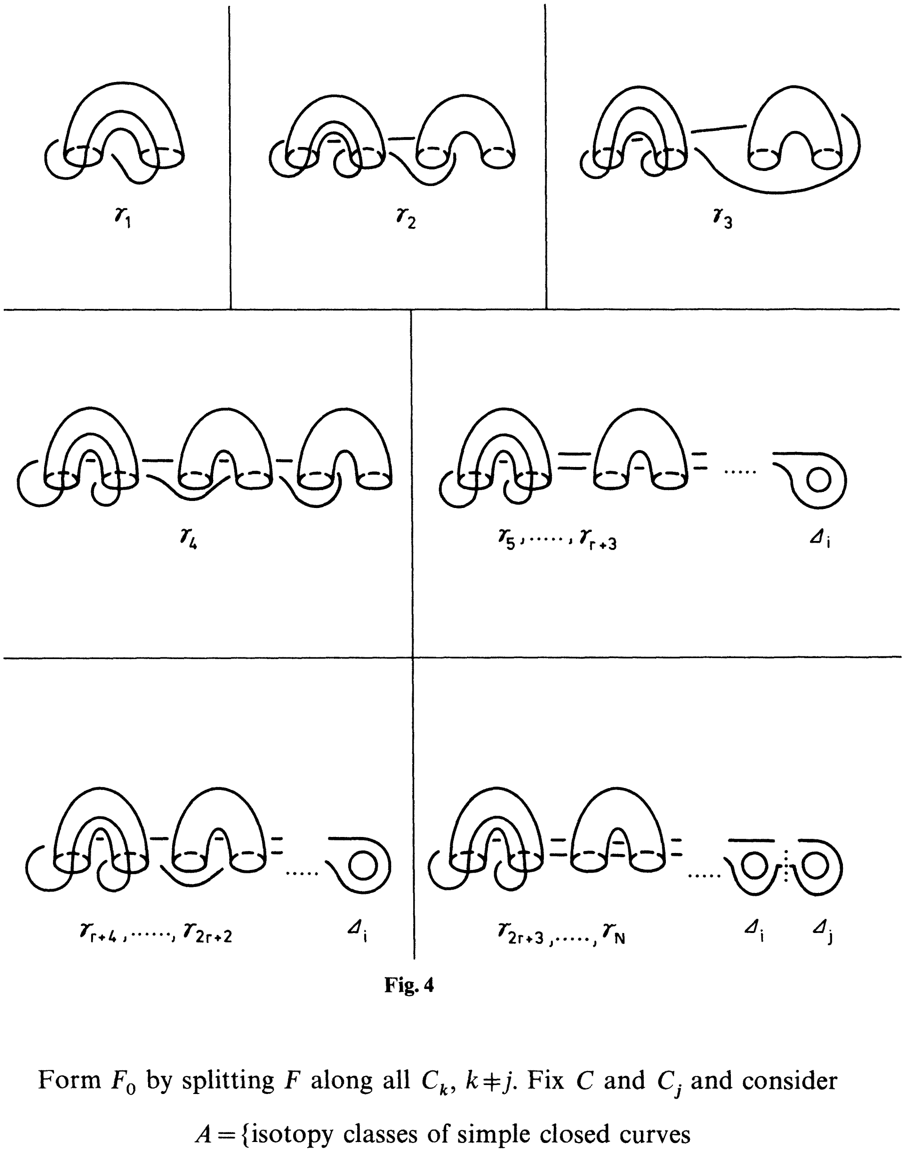


Fig. 4

Form F_0 by splitting F along all C_k, $k \neq j$. Fix C and C_j and consider

$$A = \{\text{isotopy classes of simple closed curves } \gamma \subset F_0 \text{ with } I(\gamma, C) = I(\gamma, C_j) = 1\}.$$

We may build a 1-complex Z by taking a vertex for each element of A and attaching an edge between any two elements γ, γ' for which $I(\gamma, \gamma') = 1$.

Lemma 2.3. *Z is connected.*

Proof. Fix a base point γ_0 in Z. F_0 is a torus with s holes. If s were 0 every element of A would be determined by its homology class. Furthermore, $C = C_j$ and each γ is $\gamma_0 + nC$, $n \in \mathbb{Z}$. Since $I(\gamma_0 + nC, \gamma_0 + mC) = |n - m|$ we see that Z may be identified with the real line, A the integer points.

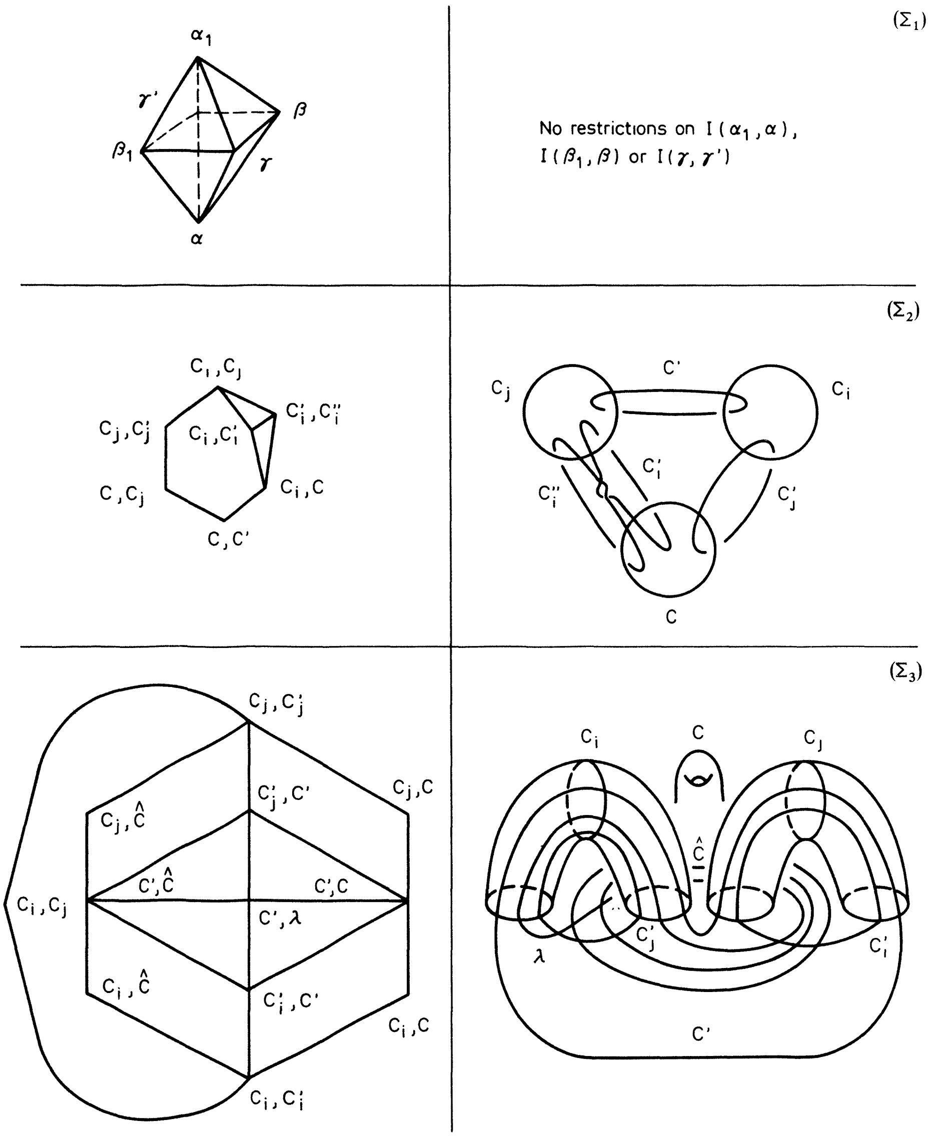


Fig. 5

For the general case proceed by induction on s. Attach a disk to one component of ∂F_0 to form F_1; $f\colon F_0 \to F_1$ the inclusion. Denote the set A for F_i by A_i; f induces a surjective map $f_*\colon A_0 \to A_1$. Suppose now that $\gamma_1 \in A_0$. By hypothesis there is a sequence $w_1, \ldots, w_r$ of elements of A_1 such that $I(w_i, w_{i+1}) = I(f_*(\gamma_0), w_1) = I(w_r, f_*(\gamma_1)) = 1$. We may choose $w_i^0, w_i^1 \in A_0$ with $f_*(w_i^j) = w_i$, $I(w_i^0, w_{i-1}^1) = 1$, $i > 1$, and $I(\gamma_0, w_1^0) = I(\gamma_1, w_r^1) = 1$.

Finally, an isotopy of $f(w_i^0)$ to $f(w_i^1)$ in F_1 gives rise to a sequence of curves $w_i^2, \ldots, w_i^j$ in F_0 each differing in F_1 by a single move across the disk $F_1 - F_0$. Choose $\eta_j \in A_0$ with $I(\eta_j, w_1^j) = I(\eta_j, w_1^{j+1}) = 1$ (for example by Dehn twisting a copy of w_i^j along C). These fill in the connection from γ_1 to γ_0 in Z. □

Lemma 2.3 also holds for C_j' and C'; therefore via type Σ_2 3-cells we see that we only need one 2-cell in Y_2 for each choice of $\langle C_k \rangle$ and C.

Consider C next. $F - \{C_k, C\}$ has two components. If either is a sphere with three holes, there is an element of Γ identifying $\{C_k\}$ with $\{\alpha_k\}$ and C with C_0 (Fig. 3). When both have more than 3 holes replace C by $\hat{C}$, a curve disjoint from C which cobounds with C_i and C_j a sphere with three holes (connect C_i to C_j by an arc disjoint from C and surger $C_i \amalg C_j$ to $\hat{C}$). It is not hard to find the extra curves C_i', C_j', C' and λ necessary to construct a 3-cell of type Σ_3. Therefore any two choices of C are equivalent and Y_2 has enough 2-cells of type R_3.

(II) No reduction is needed for type R_2. For R_1, we must reduce the collection of γ's needed for the 2-cells $\sigma(\gamma)$ to those of Fig. 4. Orient α_1, β_1, and γ so that $\alpha_1 \cdot \beta_1 = \alpha_1 \cdot \gamma = 1$. By switching α_1 and β_1 if necessary we may assume that $\beta_1 \cdot \gamma = -1$. Picture F as the boundary of a handlebody (with r 2-disks removed), label the attaching disks for the handles $D_1, \ldots, D_{2g}$ and write $\Delta_1, \ldots, \Delta_r$ for the curves of ∂F (Fig. 3). If necessary, isotope γ until the three points of intersection between α_1, β_1 and γ are distinct. Then $F - \{\alpha_i, \beta_1, \gamma\}$ has three components. If we orient α_1 once and for all, each γ is oriented by requiring $\alpha_1 \cdot \gamma = 1$; write F_1 for the region to the *right* of γ *after* it crosses α_1, F_2 for the region to the *left* of γ *before* it crosses α_1 and F_0 for the remaining region. Let ℓ_i be the number of ∂D_j $(j \geqq 3)$ and Δ_j lying in F_i. The γ curves for Y_2 occur when (assuming γ is oriented to the right as it crosses α_1 in Fig. 4) $\ell_1 = 0$, $\ell_2 \leqq 2$ and F_2 contains:

$\ell_2 = 0$	nothing,
$\ell_2 = 1$	$D_3, \Delta_1, \ldots, \Delta_{r-2}$ or Δ_{r-1},
$\ell_2 = 2$	$D_3, D_4; D_3, D_5; D_3, \Delta_i,\ 1 \leqq i \leqq r-1$ or $\Delta_i, \Delta_j\ 1 \leqq i < j \leqq r-1$.

The index (ℓ_1, ℓ_2) is not an invariant of $\sigma(\gamma)$: If we rotate the first handle and switch orientations of α_1 and γ we obtain an equivalence between curves of type (ℓ_1, ℓ_2) and type (ℓ_2, ℓ_1). Furthermore, types $(0, \ell_2)$ and $(2g-2+r-\ell_2, 0)$ are equivalent by an isotopy of γ with F_0, F_1 and F_2 permuted cyclically ((012)).

For the reduction, first suppose $\ell_1 > 2$. Build an octahedron (type Σ_1) with vertices α_1, α, β_1, β, γ, γ' where α_1, β_1, γ are as above, α links $\ell_1 - 2$ holes in F_1 while β and γ' link these and one more (Fig. 6). Each face is translated by Γ to a $\sigma(\gamma)$; if (ℓ_1', ℓ_2') is the index of any of the seven new faces of the octahedron, it is straightforward to verify that $\ell_1' + \ell_2' < \ell_1 + \ell_2$. Using symmetry we are therefore reduced to the cases where ℓ_1, $\ell_2 \leqq 2$.

For (2, 2) and (2, 1) use the curves of Fig. 7, γ' links both holes of F_1 while α and β each link a different one. If γ is type $(2, \ell_2)$ with $\ell_2 = 1$ or 2, the other triangles have types $(1, \ell_2)$, (1, 1), (2, 0), $(0, \ell_2)$, (0, 0) or (1, 0).

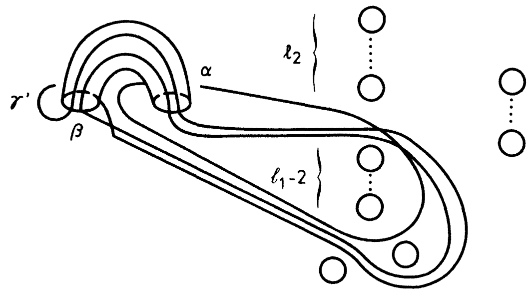


Fig. 6

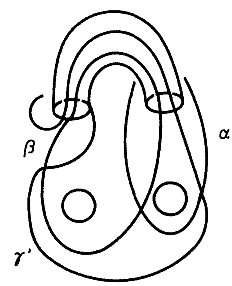


Fig. 7

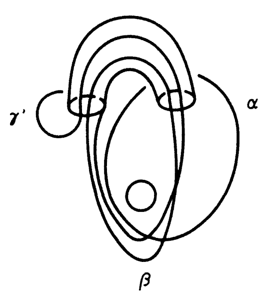


Fig. 8

With type (1, 1) use the curves of Fig. 8, where γ, α, β and γ' each link the hole of F_1. The other types are (0, 1), (0, 0). (1, 0) or (0, 2). Again by using symmetry we have reduced to types (0, 0), (0, 1) and (0, 2).

Finally, notice that in all three reductions any D_i or Δ_j lying in F_0 remains there. If Δ_r lies in F_2 for any of the γ of type $(0, \ell_2)$, $\ell_2 = 1, 2$ we use the equivalence $(0, \ell_2) \equiv (2g-2+r-\ell_2, 0)$ to put Δ_r in F_0. Then the reduction process is repeated to reach types $(0, \ell_2)$, $\ell = 0, 1, 2$ with $\Delta_r \subset F_0$. It is now easily verified that the action of Γ identifies the resulting $\sigma(\gamma)$ with those of Y_2. This finishes Theorem 2.2. $\square$

Section 3. Computation of $H_2\Gamma$ for $g \geqq 5$, $r \geqq 1$, $n=0$

Let Y_2 be the 2-complex of Sect. 2. We must add two types of 3-cells to form Y_3. For the first type notice that the curve γ_1 of Fig. 4 is disjoint from β_2. Therefore we may construct a 3-cell Σ_4 from the fact that the cycle of simple moves $\alpha_1-\beta_1-\gamma_1-\alpha_1$ commutes with the moves $\alpha_2-\beta_2-\alpha_2$. It is a triangular prism and is pictured in Fig. 9. Y_3 has a 3-cell of type Σ_3 for each configuration identified by Γ to this cell.

For the second type we add to Y_3 any Σ_3 3-cell equivalent under Γ to the one with $(C_i, C_j, C_i', \hat{C}, C') = (\alpha_1, \alpha_2, \beta_2, \beta_1, C_0, \omega)$, with λ identified with any fixed curve λ_0 which meets the other curves properly (for example, $\lambda_0 = \tau_{C_0}^{-1}(\beta_1)$, again τ_{C_0} is the right-handed Dehn twist on C_0; compare Figs. 3 and 5), and C the curve which encircles D_1 and D_4 and lies in front of D_2 and D_3, disjoint from ω ($\exists$ a map τ: $F \to F$ fixing ω', $\alpha_3, \ldots, \alpha_g$ which interchanges α_1 with α_2, β_1 with β_2 and C with C_0). Γ acts on Y_3.

Suppose next that $B\Gamma$ is a CW complex and a $K(\Gamma, 1)$, $E\Gamma \to B\Gamma$ its universal covering. From the fiber product $\Delta = E\Gamma \times_\Gamma Y_3$ there is a natural projection $f: \Delta \to B\Gamma$, a fibration with fiber Y_3. This means $\pi_1(\Delta) \cong \Gamma$ so a $K(\Gamma, 1)$ may be constructed by attaching cells to Δ of dimension $\geqq 3$. There is therefore a well defined surjection $\varphi: H_2(\Delta) \to H_2(\Gamma)$.

Theorem 3.1. *Image* $(\varphi) \cong \mathbb{Z}$.

Corollary 3.2. $H_2(\Gamma) \cong \mathbb{Z}$ *for* $g \geqq 5$, $r \geqq 1$, $n=0$.

Proof of 3.1. Write (C_*, ∂_*^C) for the cellular chain complex of Y_3, (K_*, ∂_*^K) for that of $E\Gamma$. If $R = \mathbb{Z}\Gamma$,

$$M_k = \bigoplus_{i+j=k} C_i \otimes_R K_j$$

and

$$\partial_k^M = [\bigoplus(\partial_i^C \otimes_R 1_{K_j})] + [\bigoplus(1_{C_i} \otimes_R (-1)^i \partial_j^K)],$$

then (M_*, ∂_*^M) is the chain complex of Δ. Define a filtration of M_* by setting

$$F_p(M_k) = \bigoplus_{\substack{i+j=k \\ i \leqq p}} C_i \otimes_R K_j.$$

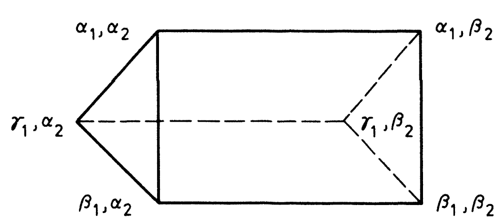


Fig. 9

The corresponding spectral sequence has $E^{\infty}_{p,q}=\dfrac{F_p(H_{p+q}(\Delta))}{F_{p-1}(H_{p+q}(\Delta))}$ where $F_p(H_{p+q}(\Delta))=\operatorname{Im}(H_{p+q}(\Delta_p)\to H_{p+q}(\Delta))$, $\Delta_p=E\Gamma\times_\Gamma Y_p$, $p\leqq 3$. Also

$$E^0_{p,q}=C_p\otimes_R K_q$$

and

$$d^0_{p,q}=1\otimes\partial^K_q.$$

Let σ_0 be the 0-cell of Y_2 corresponding to $\langle\alpha_i\rangle$, σ_1 the 1-cell for $\langle\alpha_1\rangle-\langle\beta_1\rangle$, $(\alpha_2,\ldots,\alpha_g$ understood), σ_2^i, $1\leqq i\leqq N$, the type R_1 2-cells for $\langle\alpha_1\rangle-\langle\beta_1\rangle-\langle\gamma_i\rangle-\langle\alpha_1\rangle$, σ_2^{N+1} the type R_2 2-cell for $\langle\alpha_1,\alpha_2\rangle-\langle\alpha_1,\beta_2\rangle-\langle\beta_1,\beta_2\rangle-\langle\beta_1,\alpha_2\rangle-\langle\alpha_1,\alpha_2\rangle$, σ_2^{N+2} the special type R_3 2-cell from the curves of Fig. 3, σ_3^1 the Σ_4 3-cell of Fig. 9 and σ_3^2 the Σ_3 3-cell described at the beginning of this section.

The action of Γ on C_p splits:

$$C_p=C_p^1\oplus\ldots\oplus C_p^{n_p}$$

with $\Gamma(C_p^i)\subset C_p^i$ and every generator of C_p^i identified by Γ with σ_p^i ($n_0=n_1=1$, $n_2=N+2$ and $n_3=2$). If Γ_p^i denotes the stabilizer of σ_p^i

$$C_p^i\cong R\otimes_{\mathbb{Z}[\Gamma_p^i]}\mathbb{Z}$$

via the correspondence $t\cdot g(\sigma_p^i)\leftrightarrow g\otimes t$. Hence

$$C_p\otimes_R K_q\cong\bigoplus_i(\langle\sigma_p^i\rangle\otimes_{\mathbb{Z}[\Gamma_p^i]}K_q).$$

Since $K_*\to\mathbb{Z}$ is a free $\mathbb{Z}[\Gamma_p^i]$ resolution

$$E^1_{p,q}\cong\bigoplus_i H_q(\Gamma_p^i;\langle\sigma_p^i\rangle).$$

Where $\langle\sigma_p^i\rangle\cong\mathbb{Z}\subset C_p$. Γ_0, Γ_2^i, $1\leqq i\leqq N$, Γ_2^{N+2} and Γ_3^1 act trivially on their respective $\langle\sigma_p^i\rangle$. Γ_1, Γ_2^{N+1} and Γ_3^2 however contain orientation reversing maps.

We begin the computations with

Lemma 3.3.

$$E^2_{p,0}\cong\begin{cases}\mathbb{Z} & p=0\\ 0 & p=1,3.\\ \mathbb{Z}^N\oplus\mathbb{Z}/2\mathbb{Z} & p=2\end{cases}$$

Proof. When Γ_p^i acts trivially on σ_p^i, $H_0(\Gamma_p^i;\langle\sigma_p^i\rangle)\cong\mathbb{Z}$. On the other hand for Γ_1, Γ_2^{N+1} and Γ_3^2 this group is $\mathbb{Z}/2\mathbb{Z}$. It is straight forward to then verify that

$$d^1: H_0(\Gamma_2^1;\langle\sigma_2^1\rangle)\to H_0(\Gamma_1;\langle\sigma_1\rangle)$$

is surjective, that

$$d^1: H_0(\Gamma_3^2;\langle\sigma_3^2\rangle)\to H_0(\Gamma_2^{N+1};\langle\sigma_2^{N+1}\rangle)\oplus H_0(\Gamma_2^1;\langle\sigma_2^1\rangle)$$

is injective and that

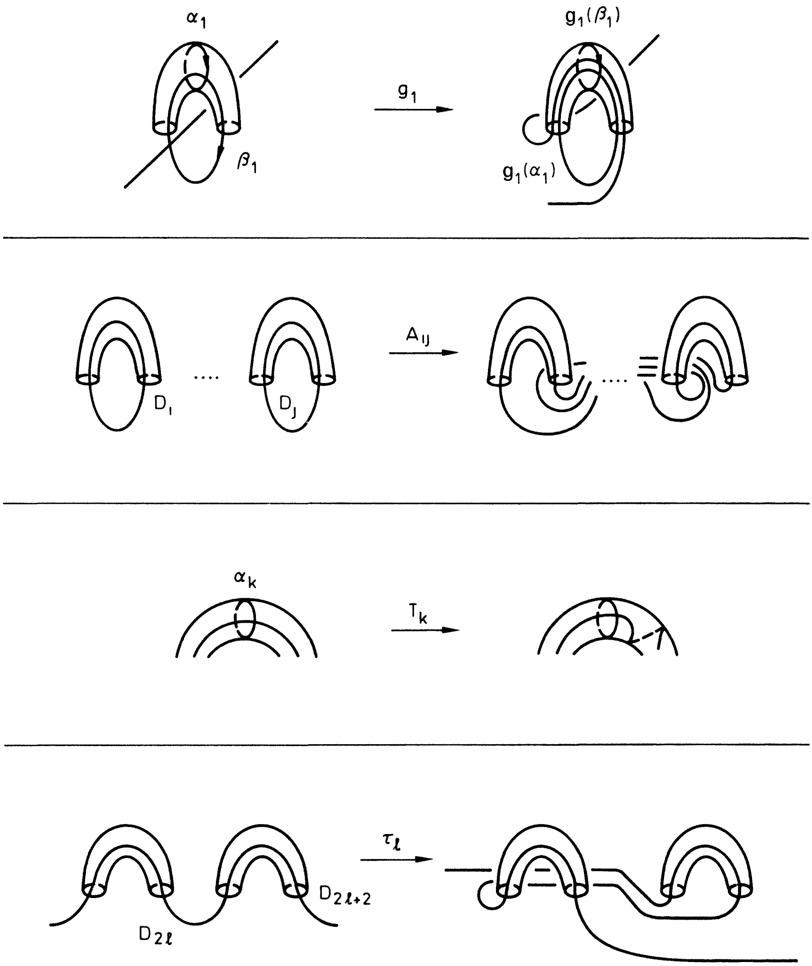


Fig. 10

$$d^1: H_0(\Gamma_3^1; \langle\sigma_3^1\rangle) \to H_0(\Gamma_2^{N+2}; \langle\sigma_2^{N+2}\rangle) \oplus \ldots$$

is multiplication by 2 in the first factor.

The lemma follows. □

Next we must analyze the stabilizer of each cell. Write

$$\hat{\Gamma}_0 = \{f \in \Gamma_0 : f \text{ fixes the curves which determine the cell } \sigma_0 \text{ pointwise}\}$$

with a similar definition of $\hat{\Gamma}_1$, $\hat{\Gamma}_2^i$ and $\hat{\Gamma}_3^i$.

Lemma 3.4. $\hat{\Gamma}_0 \cong P_{2g+r-1} \times \mathbb{Z}^{g+r-1}$ *where P_n is the pure braid group on n strands. As generators we have $A_{i,j}$, $1 \leqq i < j \leqq 2g+r-1$ and T_k, $1 \leqq k \leqq g+r-1$ with $A_{i,j}$ the diffeomorphisms of F obtained by sliding D_i around D_j (where $D_{2g+j} = \Delta_j$, Figs.* 3, 10) *and T_k the Dehn twist on α_k, $1 \leqq k \leqq g$ or Δ_{k-g}, $g+1 \leqq k \leqq g+r-1$.*

The proof from [5] for closed surfaces is easily adapted. $\square$

Let F_1 be obtained by splitting F along $\alpha_1 \cup \beta_1$; F_1 has genus $g-1$ with $r+1$ boundary components. One sees immediately that

$$\hat{\Gamma}_1 \cong \hat{\Gamma}_0(F_1)$$

with $\hat{\Gamma}_0(F_1) = \{f \in \Gamma(F_1): f|\{\alpha_2, \ldots, \alpha_g\} = 1\}$. We may likewise form F_2^i and F_3^i with

$$\hat{\Gamma}_2^i \cong \hat{\Gamma}_0(F_2^i)$$
$$\hat{\Gamma}_3^i \cong \hat{\Gamma}_0(F_3^i).$$

There are short exact sequences

$$1 \to \hat{\Gamma}_p^i \to \Gamma_p^i \to G_p^i \to 1 \qquad (*)$$

where G_p^i is the (finite) group of symmetries of σ_p^i. $G_0 \cong \pm\Sigma_g$, the group of signed permutations on g elements (the curves $\{\alpha_1, \ldots, \alpha_g\}$ may be permuted and have their orientations changed). Γ_0 is therefore generated by $\hat{\Gamma}_0$, x_1 and τ_ℓ, $1 \leqq \ell \leqq g-1$, where x_1 reverses α_1 (fixing $\alpha_2, \ldots, \alpha_g$) and τ_ℓ interchanges α_ℓ and $\alpha_{\ell+1}$ (Fig. 10; in Γ, $x_1 = g_1^2$, g_1 is the map σ from [5]). $G_1 \cong \mathbb{Z}/4\mathbb{Z} \times \pm\Sigma_{g-1}$ so Γ_1 is generated by $\hat{\Gamma}_1$, g_1, x_2 and τ_ℓ, $2 \leqq \ell \leqq g-1$. It will not be necessary (although it is not difficult) to compute G_2^i or G_3^i.

g_1 lies in the center of Γ_1 so $d^1: H_q(\Gamma_1; \langle\sigma_1\rangle) \to H_q(\Gamma_0)$ is zero for all q. Combining this with Lemma 3.3 we see that the E^2 term of our spectral sequence is

$$\begin{array}{|llll}
H_2\Gamma_0 & & & \\
H_1\Gamma_0 & E_{1,1}^2 & & \\
\mathbb{Z} & 0 & \mathbb{Z}^N \oplus \mathbb{Z}/2\mathbb{Z} & 0. \\
\hline
\end{array}$$

To complete the argument we shall show

Lemma 3.5. $H_1(\Gamma_0) \cong \mathbb{Z}^{N-1} \oplus \mathbb{Z}/2\mathbb{Z}$ *and* $E_{2,0}^3 \cong \mathbb{Z}$.

Lemma 3.6. *If* $F_p(H_2(\Gamma)) = \varphi(F_p(H_2(\Delta)))$ (φ *the map of* (3.1)),

$$F_0(H_2(\Gamma)) = F_1(H_2(\Gamma)) = 0.$$

Since $H_2(\Delta) = F_2(H_2(\Delta))$ *and* $E_{2,0}^3 = E_{2,0}^\infty = F_2(H_2(\Delta))/F_1(H_2(\Delta))$ *these lemmas complete the Proof of 3.1.*

Proof of 3.5. P_n has the following presentation ([1], see the errata, the presentation in the book in incorrect):

Generators are A_{ij}, $1 \leqq i < j \leqq n$.

Relations are $[A_{ij}, B_{i,j,r,s}]$, $1 \leqq i < j \leqq n$, $1 \leqq r < s \leqq n$ where

$$B_{i,j,r,s} = \begin{cases} A_{rs} & r<s<i<j \quad \text{or} \quad i<r<s<j, \\ A_{rs}A_{rj} & s=i, \\ A_{rs}A_{ij}A_{sj} & i=r<s<j \\ A_{rs}[A_{rj}, A_{sj}] & r<i<s<j. \end{cases}$$

Clearly then

$$H_1(P_n \times \mathbb{Z}^m) \cong \mathbb{Z}^{\binom{n}{2}+m}$$

generated by the classes of A_{ij} and T_k. In particular $H_1(\hat{\Gamma}_p^i)$ is torsion free for all p, i. If we analyze the spectral sequence associated with (*) we find (since $H_2(G_p^i)$ is torsion)

$$0 \to H_1(\hat{\Gamma}_p^i)/G_p^i \to H_1(\Gamma_p^i) \xrightarrow{\psi} H_1(G_p^i) \to 0.$$

For $p=0$, G_0 identifies

$$\begin{array}{lll} T_k \text{ with } T_1 & 1 \leqq k \leqq g, \\ A_{ij} \text{ with } A_{23} & 1 \leqq i < j \leqq 2g, \ (i,j) \neq (2s-1, 2s), \\ \phantom{A_{ij}} \text{ with } A_{12} & (i,j) = (2s-1; 2s), \ s \leqq g, \end{array}$$

or

$$\text{with } A_{1j} \quad 1 \leqq i \leqq 2g, \ j > 2g.$$

Thus $H_1(\hat{\Gamma}_0)/G_0 \cong \mathbb{Z}^{N-1}$, generators $T_1, T_{g+1}, \ldots, T_{g+r-1}, A_{12}, A_{23}, A_{1i}, A_{ij}$, $2g < i < j \leqq 2g + r - 1$.

$$H_1(G_0) \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$$

with $\psi(x_1)$ and $\psi(\tau_1)$ as generators. The τ_1 of Fig. 10 satisfies

$$\tau_1^2 = A_{23}^{-1} A_{13}^{-1} A_{34}^{-1} A_{24}^{-1} A_{14}^{-1} A_{34},$$

so $2(\tau_1 + 2A_{23})$ represents 0 in $H_1(\Gamma_0)$. x_1 on the other hand satisfies $x_1^2 = A_{12} T_1^2$. Putting this all together shows $H_1(\Gamma_0) \cong \mathbb{Z}^{n-1} \oplus \mathbb{Z}/2\mathbb{Z}$.

$H_1(\Delta) = 0$ means that

$$d^2 : \mathbb{Z}^N \oplus \mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}^{N-1} \oplus \mathbb{Z}/2\mathbb{Z}$$

must be surjective. Since $\alpha_2, \ldots, \alpha_g$ are fixed in type R_1 2-cells $d^2 | \mathbb{Z}^N$ misses the $\mathbb{Z}/2\mathbb{Z}$ generated by $\tau_1 + 2A_{23}$. Therefore $\operatorname{Ker}(d^2) \cong \mathbb{Z}$ and the lemma is proven. $\square$

Proof of 3.6. Consider again the sequence (*). Combining a presentation of $G_0 \cong \pm \Sigma_g$ with one for $\hat{\Gamma}_0$ gives us one for Γ_0. From this we may construct the 2-skeleton K_2 of a $K(\Gamma_0, 1)$. Γ_0 and g_1 generate Γ and g_1 commutes with Γ_1. Form $\hat{K}_2$ from K_2 by adding a 1-cell for g_1 and 2-cells for the relations $g_1^2 = x_1$ and $[g_1, \eta_i]$ with $\{\eta_i\}$ a generating set for Γ_1. $\hat{K}_2$ may be completed to K, a $K(\Gamma, 1)$, by adding 2-cells (for σ_2^i), 3-cells, etc.

Part 1. $F_0(H_2(\Gamma)) = \operatorname{Im}(H_2(\Gamma_0) \to H_2(\Gamma))$, for this we look at $K_2 \hookrightarrow K$. The spectral sequence associated to (*) includes the terms $E_{0,2}^2 \cong H_2(\hat{\Gamma}_0)/\pm\Sigma_g$, $E_{1,1}^2 \cong H_1(\pm\Sigma_g; H_1(\hat{\Gamma}_0))$ and $E_{2,0}^2 \cong H_2(\pm\Sigma_g)$. There is an exact sequence

$$1 \to (\mathbb{Z}/2\mathbb{Z})^g \to \pm\Sigma_g \to \Sigma_g \to 1$$

with Σ_g generated by $\tau_1, \ldots, \tau_{g-1}$ and $(\mathbb{Z}/2\mathbb{Z})^g$ generated by $x_1, \ldots, x_g$ (where x_i reverses α_i, fixing $\alpha_1, \ldots, \hat{\alpha}_i, \ldots, \alpha_g$).

$$\Sigma_g = \{\tau_1, \ldots, \tau_{g-1} : \tau_i \tau_{i+1} \tau_i = \tau_{i+1} \tau_i \tau_{i+1}, \quad \tau_1^2 = 1, \\ [\tau_i, \tau_j] = 1 \quad \text{if } |i-j| > 1\}.$$

Also, Σ_g acts on $(\mathbb{Z}/2\mathbb{Z})^g$ by permutations of the x_i. All together $E^2_{2,0}$ contributes

$$[\tau_1, \tau_3],$$
$$[x_1, x_2]$$

and

$$[x_1, \tau_2]$$

to K_2.

$E^2_{0,2}$ contributes

$$[T_i, T_j],$$
$$[T_k, A_{ij}]$$

and

$$[A_{ij}, B_{ijrs}],$$

for various choices of i, j, k, r, and s.

Finally, after allowing for the identification of certain generators of $\hat{\Gamma}_0$ by $\pm\Sigma_g$ it is straightforward to check that $E^2_{1,1}$ contributes

$$\begin{array}{ll} [\tau_2, T_k] & k=1,\ k>g \\ [x_1, T_k] & k=1,\ k>g \\ [A_{13}, \tau_1\bar{A}_{34}] & \\ [\tau_3, A_{ij}] & (i,j)=(1,2), (1,3), (1,s), (s,t), \qquad 2g<s<t \\ [x_3, A_{ij}] & \text{same } (i,j). \end{array}$$

What we finally see is that $H_2(\Gamma_0)$ is generated by commutators $[w_1, w_2]$ where either

a) w_1, w_2 are words in the generators which as elements of Γ are supported on disjoint subsurfaces one of which (say the one carrying w_2) is genus $\geqq 3$,

b) $(w_1, w_2)=(A_{13}, \tau_1\bar{A}_{34})$ or

c) $(w_1, w_2)=(x_1, T_1)$.

For (a), Γ is perfect for $g\geqq 1$ so

$$w_2 = \prod_{i=1}^{r} [y_i, z_i] \tag{$**$}$$

with $[w_1, y_i]=[w_1, z_i]=1$ for all i with y_i, z_i words in the generators of Γ. In building a $K(\Gamma, 1)$ from K_2 we will therefore have a 3-cell of the form $P\times I$ where P is a $(2r+1)$-gon giving the relation $(**)$ and the I factor is attached for w_1. Of course $\partial(P\times I)\equiv[w_1, w_2]$.

For (b) we again claim that $w_2=\tau_1\bar{A}_{34}$ has a commutator expansion $(**)$ with $[A_{13}, y_i]=[A_{13}, z_i]=1$. Equivalently we may work with $[T_1T_2A_{13}, \tau_1\bar{A}_{34}]$ since $[T_1T_2, \tau_1]$ and $[T_1T_2, \bar{A}_{34}]$ are known. Then $T_1T_2A_{13}$ is the twist T_0 on the curve C_0 which links D_1, D_3 in back of D_2. Furthermore $\tau_1\bar{A}_{34}$ fixes C_0. We may therefore find the y_i and z_i in $F-C_0$, fixing C_0 guarantees $[T_0, y_i]=[T_0, z_i]=1$ back in Γ.

Finally, for (c) the situation is slightly different since x_1 reverses the orientation of α_1. Consider the trefoil knot $K\subset S^3$.

$$\pi_1(S^3-K)\cong\{a, b\colon aba=bab\}.$$

The meridian m of K is a and the longitude ℓ is aba^2ba. Set $a=T_1$, $b=T_1^{-1}g_1T_1^{-1}$. Then $aba=bab$ is satisfied in Γ so we find $f: S^3-\dot{N}(K)\to K(\Gamma,1)$, $N(K)$ a tubular neighborhood of K, with $f(m)=T_1$ and $f(\ell)=aba^2ba=x_1$. $\operatorname{Im}(f)$ provides the nullhomology for $[T_1, x_1]$.

Part 2. $E^\infty_{1,1}\cong F_1(H_2(\Delta))/F_0(H_2(\Delta))$, $E^1_{1,1}=H_1(\Gamma_1;\langle\sigma_1\rangle)$. Since $d^r|E^r_{1,1}$ is 0 for every $r\geqq 1$, we have $E^1_{1,1}\to E^\infty_{1,1}\to H_2(\Delta)/F_0(H_2(\Delta))\to H_2(\Gamma)$; call the composition ψ. Image $(\psi)=F_1(H_2(\Gamma))$ is represented in K by the commutator 2-cells of $\hat{K}_2-K_2$. The argument of Part 1 then applies to show $F_1(H_2(\Gamma))=0$. $\square$

Section 4. Final Computations

Consider now the groups $\Gamma^n_{g,r}$ consisting of mapping classes of diffeomorphisms of F which fix $\Delta_1,\dots,\Delta_r$, the boundary curves of F, as well as distinguished points $p_1,\dots,p_n$. We will delete the indices r and/or n when equal to 0. Let π_g denote π_1 of the closed surface of genus g. We need two exact sequences:

$$1\to\mathbb{Z}\xrightarrow{f_1}\Gamma^n_{g,r}\xrightarrow{f_2}\Gamma^{n+1}_{g,r-1}\to 1 \tag{A}$$

$$1\to\pi_g\xrightarrow{f_3}\Gamma^1_g\xrightarrow{f_4}\Gamma_g\to 1. \tag{B}$$

Here f_2 is the map induced by adding a disk to Δ_r whose center becomes p_{n+1}, f_1 is the Dehn twist on a curve parallel to Δ_r, f_4 is obtained by forgetting p and f_3 comes by sliding p along a loop in F.

Analyzing the Hochschild-Serre-Lyndon spectral sequence for the *central extension* (A) gives E^2 term

$$\begin{array}{|llll}
0 & 0 & 0 & \\
\mathbb{Z} & H_1(\Gamma^{n+1}_{g,r-1}) & * & * \\
\mathbb{Z} & H_1(\Gamma^{n+1}_{g,r-1}) & H_2(\Gamma^{n+1}_{g,r-1}) & *. \\
\hline
\end{array}$$

Inductively we may assume $H_2(\Gamma^n_{g,r})\cong\mathbb{Z}^{n+1}$ and since $g>2$ $H_1(\Gamma^{n+1}_{g,r-1})\cong H_1(\Gamma^n_{g,r})\cong 0$. Therefore d^2: $H_2(\Gamma^{n+1}_{g,r-1})\to H_0(\Gamma^{n+1}_{g,r-1})$ is surjective and $H_2(\Gamma^{n+1}_{g,r-1})\cong\mathbb{Z}^{n+2}$ via $S_0,\dots,S_{n+1}$ as required.

For (B) the E^2 term is

$$\begin{array}{|ccc}
\mathbb{Z} & 0 & * \\
H_1(F_g)/\Gamma_g & H_1(\Gamma_g; H_1F_g) & * \\
\mathbb{Z} & 0 & H_2(\Gamma_g). \\
\hline
\end{array}$$

It is easy to check $H_1(F)/\Gamma=0$. Also Lemma 1.2 says $H_1(\Gamma; H_1F)=0$ so there is an exact sequence

$$0\to E^\infty_{0,2}\xrightarrow{\varphi}H_2(\Gamma^1_g)\xrightarrow{\psi}H_2(\Gamma_g)\to 0.$$

Recall

$$\left(\frac{S_0}{4}, S_1\right): H_2(\Gamma_g^1) \xrightarrow{\approx} \mathbb{Z} \oplus \mathbb{Z}.$$

Identifying $H_2(\Gamma_g^1)$ with $\Omega_2(B\,\mathrm{Diff}^+(F_g^1))$, image φ is generated by the class $[\eta]$ of the bundle over F induced by $f_3: \pi_g \to \Gamma_g^1$. Since S_0 factors through $H_2(\Gamma_g)$, $S_0[\eta]=0$. On the other hand $f_4 \circ f_3 = 0$ implies the total space of η is diffeomorphic to $F_g \times F_g$ with $s_1(F_g)$ the diagonal. Hence $S_1 \circ \varphi = 2-2g$ and the theorem follows. □

Section 5. Remarks

(1) Let f_1, f_2 be orientation preserving diffeomorphisms of F_g. Write $M_{f_i}^3$ for the mapping torus of f_i, $i=1,2$. Because the 3-dimensional bordism group is zero, $M_{f_i} = \partial W_i^4$ where W_i^4 is a compact oriented 4-manifold. By glueing W_1 and W_2 together along copies of $F \times I$ in M_{f_1} and M_{f_2} we obtain W^4 with $\partial W^4 = M_{f_2 f_1}$. Define

$$\Delta(f_1, f_2) = \mathrm{index}(W) - \mathrm{index}(W_1) - \mathrm{index}(W_2).$$

It is not difficult to check that Δ depends only on the isotopy classes of f_1 and f_2. Furthermore, Neumann [8] observes that Δ satisfies the cocycle condition

$$\Delta(f_2, f_3) - \Delta(f_1 f_2, f_3) + \Delta(f_1, f_2 f_3) - \Delta(f_1, f_2) = 0$$

for any $f_1, f_2, f_3 \in \Gamma_g$. Hence Δ represents an element of $H^2(\Gamma_g; \mathbb{Z})$. That Δ generates $H^2(\Gamma_g)$ may be seen directly from the isomorphism

$$\frac{S_0}{4}: H_2(M_g)/\mathrm{torsion} \to \mathbb{Z}.$$

(2) Γ is generated by Dehn twists $\tau(\gamma_i)$ on nonseparating circles γ_i (defined up to isotopy) in F. There is an easy relation among such Dehn twists, namely

$$\tau(\gamma_i)\tau(\gamma_j) = \tau(\gamma_j)\tau(\gamma_i^{\gamma_j}) \qquad (*)$$

where $\gamma_i^{\gamma_j}$ is the image of the curve γ_i under $\tau(\gamma_j)$. Let G be the group with generators $\tau(\gamma_i)$ taken over all nonseparating circles in F and relations of type (*) (for example when $g=1$, $G \cong \Gamma_{1,1}$). There is a natural surjection $G \xrightarrow{\varphi} \Gamma$. In [4] the kernel of φ is computed to be

$$H_2(\Gamma) \oplus \mathbb{Z}.$$

Because Γ is perfect $G' = [G, G]$ is also perfect and maps onto Γ. $H_1(G) = \mathbb{Z}$ so we have an exact sequence

$$0 \to H_2(\Gamma) \to G' \to \Gamma \to 1.$$

G' is the *universal central extension* of Γ. G' may also be obtained from G by adjoining the relation (*) from Sect. 1 for an embedding of F_0 in F with all curves nonseparating.

(3) Clearly Theorem 2.2 gives a simplification of Hatcher and Thurston's presentation of Γ. Wajnryb [10] has used this to give an incredibly simple presentation of Γ_g and $\Gamma_{g,1}$.

References

1. Birman, J.: Braids, links, and mapping class groups. Ann. of Math. Studies 82, Princeton Univ. Press, Princeton 1975
2. Dehn, M.: Die Gruppe der Abbildungsklassen. Acta Math. **69**, 135–206 (1938)
3. Earle, C.J., Eells, J.: The diffeomorphism group of a compact Riemann surface. Bull. AMS **73**, 557–559 (1967)
4. Harer, J.: Pencils of Curves on 4-Manifolds. Ph.D. Thesis, University of California, Berkeley 1979
5. Hatcher, A., Thurston, W.: A presentation for the mapping class group of a closed orientable surface. Top. **19**, 221–237 (1980)
6. Meyer, W.: Die Signatur von Flächenbündeln. Math. Ann. **201**, 239–264 (1973)
7. Mumford, D.: Abelian quotients of the Teichmuller modular group. Journal d'Analyse Mathematique **18**, 227–244 (1967)
8. Neumann, W.: Signature related invariants of manifolds. I. Monodromy and γ-Invariants. Top. **18**, 239–264 (1973)
9. Powell, J.: Two theorems on the mapping class group of a surface. Proc. Amer. Math. Soc. **68**, No. 3, 347–350 (1978)
10. Wajnryb, B.: A simple presentation for the mapping class group of an orientable surface. Preprint 1982

Oblatum 28-VIII-1982

Note Added in Proof

The author has been informed by Mumford that the computation of $H_2\Gamma$ gives a proof of the rational version of the Francetta conjecture.

Invent. math. 72, 241–265 (1983)

Inventiones mathematicae

On the Birch and Swinnerton-Dyer Conjecture

Ralph Greenberg*

Department of Mathematics, University of Washington, Seattle, Washington 98195/USA

Introduction

Let F be a finite extension of the rational field $\mathbb{Q}$. If E is an elliptic curve defined over F, then the Mordell-Weil group $E(F)$ of points on E with coordinates in F is a finitely generated abelian group. Let $L(E/F,s)$ be the Hasse-Weil zeta function of E over F which, at least in the case where E has complex multiplication, is known to be an analytic function on the entire complex plane. Birch and Swinnerton-Dyer have conjectured that $L(E/F,s)$ has a zero at $s=1$ of order precisely equal to the rank of $E(F)$ over $\mathbb{Z}$. The strongest result known in support of this conjecture is the following theorem of Coates and Wiles [4].

Theorem. *Assume E has complex multiplication by the ring of integers of an imaginary quadratic field K and that $F=\mathbb{Q}$ or K. If $E(F)$ has rank$\geqq 1$, then $L(E/F,s)$ vanishes at $s=1$.*

Generalizations of this theorem have been proved by Arthaud [1] and by Rubin [20].

In this paper, we will prove the following partial converse to the Coates-Wiles theorem.

Theorem 1. *Assume that E is an elliptic curve defined over $\mathbb{Q}$ with complex multiplication by the ring of integers of an imaginary quadratic field K. If $L(E/\mathbb{Q},s)$ has an odd order zero at $s=1$, then either $E(\mathbb{Q})$ has* rank$\geqq 1$ *or the p-primary subgroup of the Tate-Shafarevich group $III(E,\mathbb{Q})$ is infinite for all primes p where E has good, ordinary reduction (except possibly $p=2$ or 3).*

The function $L(E/\mathbb{Q},s)$ can be identified with a Hecke L-series $L(\Psi,s)$, where Ψ is a certain grossencharacter of K associated with the elliptic curve E, which we define in Sect. 1. From this, one can obtain the analytic continuation and also the following functional equation:

$$\Lambda(E/\mathbb{Q},2-s)=w\Lambda(E/\mathbb{Q},s),$$

* Supported in part by a National Science Foundation grant

where $\Lambda(E/\mathbb{Q}, s)=(2\pi/\sqrt{N})^{-s}\Gamma(s)\,L(E/\mathbb{Q}, s)$. Here N denotes the conductor of the elliptic curve E over $\mathbb{Q}$ and the "root number" w is ± 1. Obviously, $w=-1$ precisely when $L(E/\mathbb{Q}, s)$ has an odd order zero and so our theorem could be stated as follows: If $w=-1$, then the p^∞-Selmer group $S_{p^\infty}(E, \mathbb{Q})$ is infinite for all p where E has good, ordinary reduction ($p\neq 2, 3$). We will define the Selmer groups later, but will just mention here the fundamental exact sequence:

$$0\to E(F)\otimes_{\mathbb{Z}}(\mathbb{Q}_p/\mathbb{Z}_p)\to S_{p^\infty}(E, F)\to Ш(E, F)_{p\text{-primary}}\to 0.$$

(Here E is any elliptic curve defined over a number field F. Also $\mathbb{Q}_p$ is the field of p-adic numbers, $\mathbb{Z}_p$ the p-adic integers.)

The main reason we are able to prove a result of this kind is that the odd order zero reveals itself in a rather dramatic way. To explain what we mean, let K_∞^- denote the so-called anti-cyclotomic $\mathbb{Z}_p$-extension of K. Thus, $K_\infty^- = \bigcup_{n\geq 0} K_n^-$ where K_n^- is a cyclic extension of K of degree p^n and is Galois over $\mathbb{Q}$ with a dihedral Galois group. We assume (as we almost always will throughout this paper) that E is again defined over $\mathbb{Q}$, has complex multiplication by the ring of integers $\mathcal{O}$ of the imaginary quadratic field K, and that E has good, ordinary reduction at p ($p\neq 2, 3$). Then p must split in K, $p\mathcal{O}=\mathfrak{p}\mathfrak{p}^*$. Let $\mathfrak{p}=(\pi)$, $\pi\in\mathcal{O}$. (Under our assumptions, K must have class number one.) We will prove the following theorem.

Theorem 2. *If $w=-1$, then $S_{\pi^\infty}(E, K_n^-)$ contains a subgroup isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^{p^n}$ for $n=0, 1, 2, \ldots$.*

In particular, $S_{\pi^\infty}(E, K)$ is infinite. As we will explain in section 5, this implies that $S_{p^\infty}(E, \mathbb{Q})$ is also infinite and hence theorem 1 is a consequence of theorem 2. The π^∞-Selmer groups that occur here are contained in the exact sequences:

$$0\to E(K_n^-)\otimes_{\mathcal{O}}(K_\mathfrak{p}/\mathcal{O}_\mathfrak{p})\to S_{\pi^\infty}(E, K_n^-)\to Ш(E, K_n^-)_{\pi\text{-primary}}\to 0,$$

where $K_\mathfrak{p}$ and $\mathcal{O}_\mathfrak{p}$ denote the completions of K and $\mathcal{O}$ at $\mathfrak{p}$ (so that $K_\mathfrak{p}/\mathcal{O}_\mathfrak{p}\cong\mathbb{Q}_p/\mathbb{Z}_p$, but is also an $\mathcal{O}$-module). We have that $E(K_n^-)\otimes_{\mathcal{O}}(K_\mathfrak{p}/\mathcal{O}_\mathfrak{p})\cong(\mathbb{Q}_p/\mathbb{Z}_p)^{r_n}$, where r_n is the rank of $E(K_n^-)$ as a module over $\mathcal{O}$ (or equivalently, $r_n=\frac{1}{2}\operatorname{rank}_{\mathbb{Z}}(E(K_n^-))$). Now assume that $Ш(E, F)$ is finite (or at least its π-primary subgroup) for $F=K_n^-$, $n\geq 0$. It would then follow from Mazur's results in [17] that $r_n=ap^n+O(1)$ as $n\to\infty$ for some integer a, which from Theorem 2 must clearly be positive. We can actually prove however that $a=1$, if we make this assumption about $Ш$ for a somewhat larger collection of fields F (e.g. all subfields of $K(E_p)K_\infty^-$, where $K(E_p)$ is the field generated by points of order p on E).

The assertion that $a\geq 1$ is predicted by the Birch and Swinnerton-Dyer conjecture. The zeta function for E over K_n^- is given by $L(E/K_n^-, s) = \prod_\rho L(\Psi\rho, s)^2$, where ρ varies over all characters of $\mathrm{Gal}(K_n^-/K)$, considered (by class field theory) as Dirichlet characters for K. Now the functional equation for $L(\Psi\rho, s)$ will relate it to $L(\Psi(\bar\rho\circ c), s)$, where c is complex conjugation (in $\mathrm{Gal}(K/\mathbb{Q})$). But, since $\mathrm{Gal}(K_n^-/\mathbb{Q})$ is dihedral, we can see that $\bar\rho\circ c=\rho$. Thus the root number $w(\Psi\rho)$ will be ± 1, depending on the parity of

$\mathrm{ord}_{s=1}(L(\Psi\rho, s))$. However, as we explain in Sect. 1, $w(\Psi\rho)=w(\Psi)=w$ for all such ρ. Hence, if $w=-1$, $L(E/K_n^-, s)$ must have a zero of order at least $2p^n$ at $s=1$. Furthermore, if in fact $a=1$, then the Birch and Swinnerton-Dyer conjecture would imply that $L(\Psi\rho, s)$ has just a simple zero at $s=1$ for all but possibly finitely many complex characters ρ of $\mathrm{Gal}(K_\infty^-/K)$.

If $w=+1$, we will actually be able to prove that $L(\Psi\rho, 1)\neq 0$ for all but finitely many such ρ's. Using a theorem of K. Rubin, we can then immediately prove the following result (with the assumptions stated before Theorem 2).

Theorem 3. *If* $w=+1$, *then* $\mathrm{rank}_{\mathcal{O}}(E(K_n^-))$ *is bounded as* $n\to\infty$.

This result together with some similar but easier results concerning certain other $\mathbb{Z}_p$-extensions of K will be proved in Sect. 6. We also show that $E(K_\infty^-)$ is finitely generated as a consequence of the above result.

To prove our theorems, we consider the functions $L(\Psi^{2k+1}, s)$ for $k=0, 1, 2, \ldots$. These satisfy the functional equations:

$$\Lambda(\Psi^{2k+1}, 2k+2-s)=w_k\Lambda(\Psi^{2k+1}, s),$$

where

$$\Lambda(\Psi^{2k+1}, s)=(2\pi/\sqrt{N_k})^{-s}\Gamma(s)\,L(\Psi^{2k+1}, s).$$

Here

$$N_k=|\mathrm{disc}(K)|\,\mathcal{N}_{K/\mathbb{Q}}(f_{\Psi^{2k+1}}),$$

where $f_{\Psi^{2k+1}}$ is the conductor of the grossencharacter Ψ^{2k+1} of K. Let m denote the number of roots of unity in K. We will show that the root numbers w_k, which clearly must be ± 1, have the following properties: (i) w_k depends only on the residue class of k modulo m, and (ii) if $k_1+k_2\equiv -1 \pmod m$, then $w_{k_1}w_{k_2}=-1$. Since w_k determines the parity of the order of vanishing of $L(\Psi^{2k+1}, s)$ at $s=k+1$, we find that $L(\Psi^{2k+1}, k+1)=0$ for half of the integers $k\geq 0$ simply because of the sign in the functional equation. Our next theorem concerns the other k's.

Theorem 4. *There are only finitely many values of* k *with* $w_k=+1$ *such that* $L(\Psi^{2k+1}, k+1)$ *vanishes.*

This result was suggested to us by some calculations of Gross and Zagier [10]. They consider the above L-values where Ψ is the grossencharacter attached to several elliptic curves with complex multiplication by the ring of integers of $K=\mathbb{Q}(\sqrt{-7})$ and for a number of values of k. They find just one zero (when $w_k=+1$), namely $L(\Psi, 1)$ for a certain elliptic curve E such that $E(\mathbb{Q})$ has rank 2. In addition, among some unpublished calculations of Swinnerton-Dyer, Nelson Stephens found two cases where $L(\Psi^3, 2)$ vanishes with $+1$ as the corresponding root number. The elliptic curves were defined by $y^2=x^3-Dx$ for $D=2\cdot 73^2$ and 79^3, which have complex multiplication by $\mathbb{Z}[i]$. It would be very interesting to understand the meaning of such zeros, even conjecturally.

A crucial role in the proof of Theorem 4 is played by the two-variable p-adic L-functions constructed by Katz in [13] and by Manin and Vishik in [15]. The L-values being considered are essentially the values of the one-variable p-adic L-functions obtained by specializing Katz's functions to a certain "critical line". Theorem 4 is equivalent to the non-triviality of one of these specializations and this in turn implies that only finitely many of the $L(\Psi\rho, 1)$'s mentioned above can vanish when $w(\Psi)=+1$. As for the proof of Theorem 2, we will just say here that it involves a more subtle use of Theorem 4 together with a recent theorem of Yager [28] which gives an interpretation of the two-variable p-adic L-functions in terms of a certain Iwasawa module constructed from elliptic units.

The conclusion of Theorem 1 is undoubtedly true even when $L(E/\mathbb{Q}, s)$ has an even order zero at $s=1$. Although our approach here hasn't provided a proof, we will have some remarks to make about this case later. One can go further in other directions though. We can prove a more general nonvanishing result by the method used in proving Theorem 4. As a consequence, we can show that $\mathrm{rank}_{\mathbb{Z}}(E(K_n^-))$ will be bounded as $n \to \infty$, if E is any elliptic curve defined over K (and with complex multiplication by $\mathcal{O}$) but not isomorphic over K to an elliptic curve defined over $\mathbb{Q}$. We also want to remark that our results are valid for elliptic curves defined over $\mathbb{Q}$ with complex multiplication by a non-maximal order in K, because such elliptic curves are known to be $\mathbb{Q}$-isogenous to ones with $\mathcal{O}$ as an endomorphism ring.

In a subsequent paper [9], we will study a certain new type of root number, defined purely algebraically, and its relationship to the root number of complex L-functions. This topic is closely related and in fact led us to the results in this paper. The algebraic root number, although defined as the sign in a certain functional equation for a characteristic power series, turns out to be $(-1)^e$ where e is the power to which the critical divisor Θ defined in Sect. 2 divides this characteristic power series.

In conclusion, we want to thank Barry Mazur and Dick Gross for some very helpful discussions concerning the topics in this paper.

1. L-functions and Root Numbers

Let E be an elliptic curve defined over $\mathbb{Q}$. If E has good reduction at a prime l, the reduced elliptic curve has a zeta function of the form

$$\frac{P_l(t)}{(1-t)(1-lt)},$$

where $P_l(t)=(1-\alpha_l t)(1-\bar{\alpha}_l t)$ has integral coefficients and complex conjugate roots such that $\alpha_l\bar{\alpha}_l=l$. The zeta function of E over $\mathbb{Q}$ is defined by an Euler product $L(E/\mathbb{Q}, s)=\prod_l P_l(l^{-s})^{-1}$ for $\mathrm{Re}(s)>\frac{3}{2}$, where the polynomials $P_l(t)$ are defined as in Tate [25′] when E has bad reduction at l. Assume that E has complex multiplication by the ring of integers $\mathcal{O}$ of the imaginary quadratic field K. If E has good, ordinary reduction at l, then l splits in K. In this case, α_l and $\bar{\alpha}_l$ are in $\mathcal{O}$ and generate the prime ideals of $\mathcal{O}$ dividing l: $\mathfrak{l}=(\alpha_l)$, $\bar{\mathfrak{l}}=(\bar{\alpha}_l)$, $\mathfrak{l}\bar{\mathfrak{l}}=l\mathcal{O}$. We define $\Psi(\mathfrak{l})=\alpha_l$ and $\Psi(\bar{\mathfrak{l}})=\bar{\alpha}_l$. If E has good, supersingular reduction at l, then l remains prime in K, and $P_l(t)=1+lt^2$. We define $\Psi(\mathfrak{l})=-l$ for $\mathfrak{l}=l\mathcal{O}$. If E has bad reduction at l, then it must be of additive type (integrality of the j-invariant of E) and so $P_l(t)=1$. Define $\Psi(\mathfrak{l})=0$ for any prime ideal $\mathfrak{l}$ dividing such l. We can then define $\Psi(\mathfrak{a})$ for any ideal $\mathfrak{a}$ of $\mathcal{O}$ so that Ψ is multiplicative. Clearly, with these definitions,

$$L(E/\mathbb{Q},s)=\prod_{\mathfrak{l}}(1-\Psi(\mathfrak{l})\,\mathcal{N}(\mathfrak{l})^{-s})^{-1}=\sum_{\mathfrak{a}}\Psi(\mathfrak{a})\,\mathcal{N}(\mathfrak{a})^{-s}=L(\Psi,s)$$
$$\text{for } \operatorname{Re}(s)>\tfrac{3}{2}.$$

It is known that Ψ is a grossencharacter of type A_0 for K (Deuring [6]) of infinity type $(1,0)$. In general, a multiplicative complex-valued function $\mathscr{C}$ on the ideals of K is a grossencharacter of type A_0 with infinity type (a,b) if there is an ideal $f_{\mathscr{C}}$, the conductor of $\mathscr{C}$, such that $\mathscr{C}(\alpha\mathcal{O})=\alpha^a\bar{\alpha}^b$ for all $\alpha\in\mathcal{O}$ such that $\alpha\equiv 1\,(\mathrm{mod}\, f_{\mathscr{C}})$, where $a,b\in\mathbb{Z}$. There is a standard way (which is described in Weil [26] and which we will take for granted in what follows) to associate with such $\mathscr{C}$ a continuous character $\mathscr{C}\colon I_K\to\mathbb{C}$, where I_K denotes the idele group for K, such that $\mathscr{C}$ is trivial on the principal ideles $K^\times\subseteq I_K$. No confusion should result from using the same letter $\mathscr{C}$. In fact, it is for such continuous idele class group characters $\mathscr{C}$ that one can attach an L-function $L(\mathscr{C},s)$ which can be analytically continued and satisfies a functional equation (see Tate [25], for example). If $\mathscr{C}$ is of type A_0 and has infinity type (a,b), then the functional equation will relate $L(\mathscr{C},s)$ to $L(\bar{\mathscr{C}},d-s)$, where $d=a+b+1$. However, since complex conjugation c simply permutes the ideals of $\mathcal{O}$, $L(\bar{\mathscr{C}},s)$ is identical to $L(\bar{\mathscr{C}}\circ c,s)$ and, if $\bar{\mathscr{C}}\circ c=\mathscr{C}$, the functional equation relates the function $L(\mathscr{C},s)$ to itself and the root number $w(\mathscr{C})$ that occurs in this functional equation must be ± 1. Our definition of Ψ makes it clear that $\bar{\Psi}(\bar{\mathfrak{a}})=\Psi(\mathfrak{a})$ for all ideals $\mathfrak{a}$ of $\mathcal{O}$ and so Ψ and its powers have the property just mentioned. As in the introduction, let $w_k=w(\Psi^{2k+1})$ for $k=0,1,2,\ldots$. We should say here that we will only consider primitive grossencharacters. Thus, for example, Ψ^t is the primitive grossencharacter of K such that $\Psi^t(\mathfrak{a})=\Psi(\mathfrak{a})^t$ for all $\mathfrak{a}$ prime to f_Ψ, where $t\in\mathbb{Z}$.

The proof of the following proposition was essentially shown to us by B. Gross. A special case is mentioned in [10].

Proposition 1. *Let m denote the number of roots of unity of K. Then w_k depends only on the residue class of k modulo m. Also, if $k_1+k_2\equiv -1\,(\mathrm{mod}\, m)$, then $w_{k_1}w_{k_2}=-1$.*

Proof. The crucial observation is that, since $\Psi(\mathfrak{a})$ is in K for all ideals, Ψ^m is precisely the grossencharacter of K defined by $\mathfrak{a}=\alpha\mathcal{O}\to\alpha^m$ for any $\alpha\in\mathcal{O}$, which has trivial conductor. If $\mathscr{C}$ is a continuous idele class character, let $\mathscr{C}_{v_\infty}$ be its component at the infinite prime v_∞, a character of $\mathbb{C}^\times$ of the form $z\to z^n|z|^{s_0}$ for some $n=n(\mathscr{C})\in\mathbb{Z}$, $s_0\in\mathbb{C}$. We will also need the following useful fact (which can be found in Weil [27], page 161): Assume $\mathscr{C}$ and $\mathscr{C}'$ are idele class characters of absolute value 1 and relatively prime conductors such that $n(\mathscr{C})\,n(\mathscr{C}')\geqq 0$. Then

$$w(\mathscr{C}\mathscr{C}')=w(\mathscr{C})\,w(\mathscr{C}')\,\mathscr{C}(f_{\mathscr{C}'})\,\mathscr{C}'(f_{\mathscr{C}}). \tag{1}$$

If $j\equiv k\,(\mathrm{mod}\, m)$, $j\geqq k\geqq 0$, then $\Psi^{2j+1}=\Psi^{2k+1}\,\Psi^{tm}$ for some even integer $t\geqq 0$. Let $\mathscr{C}=\Psi^{2k+1}/|\Psi^{2k+1}|$ and $\mathscr{C}'=\Psi^{tm}/|\Psi^{tm}|$, where $|\Psi^v|(\mathfrak{a})=\mathcal{N}(\mathfrak{a})^{v/2}$ on ideals.

Then $w(\mathscr{C})=w(\Psi^{2k+1})=w_k$ and $\mathscr{C}(f_{\mathscr{C}'})=\mathscr{C}(\mathcal{O})=1$. To compute $w(\mathscr{C}')$, we write $\mathscr{C}'=\prod_v \mathscr{C}'_v$ as a product of local characters so that, as in Tate [25], $w(\mathscr{C}')$ equals the product of local root numbers $\prod_v w(\mathscr{C}'_v)$, the product running over all primes of K, finite and infinite. The formulas for these local root numbers are given in [25]. We have $w(\mathscr{C}'_{v_\infty})=(-i)^{tm}$ since $|n(\mathscr{C}')|=tm$. Also, let v_0 be the unique ramified prime in $K/\mathbb{Q}$. (Recall that K has class number 1.) Then $w(\mathscr{C}'_{v_0})=\mathscr{C}'_{v_0}(\delta_{v_0})$, where δ_{v_0} is a generator of the local different. If δ_K is the global different (an ideal), then $\mathscr{C}'(\delta_K)=\mathscr{C}'_{v_0}(\delta_{v_0})$. In all cases, δ_K has a generator δ such that $\delta/|\delta|=i$ and hence $\mathscr{C}'(\delta_K)=i^{tm}$. If v is a prime of K, $v \neq v_\infty$ or v_0, then $w(\mathscr{C}'_v)=1$. It follows that $w(\mathscr{C}')=1$ for any integer t. In order to compute $\mathscr{C}'(f_{\mathscr{C}})$, we observe that $\bar{f}_{\mathscr{C}}=f_{\mathscr{C}}$ since $\bar{\mathscr{C}}\circ c=\mathscr{C}$ and so $f_{\mathscr{C}}=\pi_{v_0}^e a\mathcal{O}$, where π_{v_0} is a generator of the ramified prime of K, $e \geqq 0$, and a is a rational integer prime to disc (K). We have $\mathscr{C}'(a\mathcal{O})=1$ obviously and so $\mathscr{C}'(f_{\mathscr{C}})=\mathscr{C}'(\pi_{v_0}\mathcal{O})=(\pi_{v_0}/|\pi_{v_0}|)^{tme}$. If $K \neq \mathbb{Q}(i)$, then π_{v_0} can be chosen so that $\pi_{v_0}/|\pi_{v_0}|=i$. If $K=\mathbb{Q}(i)$, we can take $\pi_{v_0}=1+i$ and we have $\pi_{v_0}/|\pi_{v_0}|=\eta$, where η is an 8-th root of unity. Now since $n(\mathscr{C})\, n(\mathscr{C}') \geqq 0$, we can apply (1) and we clearly get that $w(\Psi^{2j+1})=w(\Psi^{2k+1})$.

One can prove the remaining part of Proposition 1 in the cases where $m=2$ from the above discussion. The only additional result one needs is that e is odd in these cases. In fact, it is not too difficult to show that $e=1$ except when $K=\mathbb{Q}(\sqrt{-2})$ in which case $e=5$. It follows that $w_k=w_0(-1)^k$, which is the content of Proposition 1 when $m=2$. If $m>2$, e can be even or odd.

We could settle the general case by using a slightly stronger version of (1). However, we prefer to proceed as follows. Let $\mathscr{C}_1=\Psi^{2k_1+1}/|\Psi^{2k_1+1}|$ and $\mathscr{C}_2=\Psi^{2k_2+1}/|\Psi^{2k_2+1}|$ so that $\mathscr{C}=\mathscr{C}_1\mathscr{C}_2=\Psi^{2mt}/|\Psi^{2mt}|$ for some t. Then $w(\mathscr{C}_{1,v_\infty})=(-i)^{2k_1+1}$, $w(\mathscr{C}_{2,v_\infty})=(-i)^{2k_2+1}$ and the product of these local root numbers is $+1$. Now let v be any finite prime. For brevity, let $\theta_1=\mathscr{C}_{1,v}$, $\theta_2=\mathscr{C}_{2,v}$, and $\theta=\theta_1\theta_2=\mathscr{C}_v$, an unramified character of the completion K_v^x. Proposition 2.2 of [16] states that $w(\theta_1)\, w(\theta_1^{-1})=\theta_1(-1)$, essentially just the well-known fact about the product of Gaussian sums for θ_1 and θ_1^{-1}. Furthermore, $w(\theta_1^{-1})=w(\theta_2\theta^{-1})=w(\theta_2)\,\theta^{-1}(c_v)$, where c_v is a generator of $\delta_{K_v} f(\theta_2)$, δ_{K_v} is the local different, and $f(\theta_2)$ is the conductor of θ_2. This follows easily from the definition of local root numbers in [16], pages 29, 32 and the fact that θ^{-1} is unramified. (This is essentially the local version of (1) and can be used to prove it. See also [25″].) Thus $w(\theta_1)\, w(\theta_2)=\theta_1(-1)\,\theta(c_v)$. Now $\theta=\mathscr{C}_v$ and $\prod_{\substack{\text{finite}\\ v}} \mathscr{C}_v(c_v)=\mathscr{C}(\delta_K f_{\mathscr{C}_2})$ which is easily verified to be $+1$ because $\delta_K f_{\mathscr{C}_2}=\pi_{v_0}^b a\mathcal{O}$ for some $b, a \in \mathbb{Z}$ and one can argue as before. Also, $\theta_1=\mathscr{C}_{1,v}$ and, viewing $\mathscr{C}_1$ as a character of the idele group I_K,

$$\mathscr{C}_1(-1)=+1=\mathscr{C}_{1,v_\infty}(-1)\prod_{\substack{\text{finite}\\ v}} \mathscr{C}_{1,v}(-1).$$

From all of this we get that

$$w_{k_1} w_{k_2}=\prod_{\substack{\text{finite}\\ v}} w(\mathscr{C}_{1,v})\, w(\mathscr{C}_{2,v})=\mathscr{C}_{1,v_\infty}(-1)^{-1}=(-1)^{2k_1+1}=-1,$$

proving Proposition 1.

It will be useful to know how the N_k's vary. If v is any finite prime, let β_v denote the restriction of Ψ_v to the units in K_v. For any integer t, the conductor of Ψ^t is determined in a well-known way by the kernels of the β_v^t's. (See [16].) Since Ψ^m is unramified, β_v has order dividing m. If $m=2$ or 4, this shows that $f_{\Psi^{2k+1}}$ is independent of k and hence $N_k=N$ for all k. Therefore, in this case $\Psi^{2k+1}(\mathfrak{a})=\Psi(\mathfrak{a})^{2k+1}$ for all ideals $\mathfrak{a}$ of $\mathcal{O}$. If $K=\mathbb{Q}(\sqrt{-3})$, N_k can vary (which will cause some difficulty in our proof of Theorem 4), but clearly N_k depends only on k modulo m (allowing us to overcome the difficulty). We will also need the fact that $f_{\Psi^{2k+1}}$ is always divisible by the ramified prime of K. It is enough to show that $\Psi_{v_0}(-1)=-1$. Then $\beta_{v_0}^{2k+1}$ will be nontrivial. Let $\alpha=(\alpha_v)\in I_K$ be such that $\alpha_v=-1$ for all $v\neq v_0, v_\infty$ and $\alpha_{v_0}=\alpha_{v_\infty}=1$. It is easy to verify that there exists a $\gamma=(\gamma_v)\in I_K$ such that $\gamma c(\gamma)=\alpha$. We can assume γ_v is a unit for all finite v and $\gamma_{v_\infty}=1$. Now, as a character of I_K, Ψ still has the properties $\overline{\Psi}=\Psi\circ c$, $\Psi\overline{\Psi}=\mathcal{N}$ and so $\Psi(\alpha)=\Psi(\gamma)\,\overline{\Psi}(\gamma)=\mathcal{N}(\gamma)=1$. This implies that $\Psi_{v_\infty}(-1)\,\Psi_{v_0}(-1)=1$. Since $\Psi_{v_\infty}(-1)=-1$, we're done.

We want to now explain the remarks we made in the introduction about $w(\Psi\rho)$, where ρ is a Dirichlet character for K which corresponds by class field theory to a character of $\mathrm{Gal}(K_n^-/K)$. Here K_n^- is the n-th layer of the anti-cyclotomic $\mathbb{Z}_p$-extension of K for any prime p at which E has good, ordinary reduction. The conductor of ρ involves only the primes dividing p and so is certainly relatively prime to f_Ψ. We can use formula (1). First we observe that $L(\rho,s)=L(\rho^*,s)$, where ρ^* is the two-dimensional Artin character of $\mathrm{Gal}(K_n^-/\mathbb{Q})$ obtained by induction. Since this Galois group is dihedral, it is well-known that $w(\rho^*)=w(\rho)$ must be $+1$ (see [7]). Next since $\rho\circ c=\rho^{-1}$, we have $f_\rho=\bar{f}_\rho$ and so $f_\rho=(\mathfrak{p}\bar{\mathfrak{p}})^e$, where $\mathfrak{p}$, $\bar{\mathfrak{p}}$ are the two primes of K dividing p and $e\geqq 0$. Hence $\Psi(f_\rho)=(\Psi(\mathfrak{p})\,\Psi(\bar{\mathfrak{p}}))^e=(\Psi(\mathfrak{p})\,\overline{\Psi}(\mathfrak{p}))^e=p^e=|\Psi(f_\rho)|$. Finally, $\bar{f}_\Psi=f_\Psi$ and therefore $\rho(f_\Psi)=\rho(\bar{f}_\Psi)=\rho^{-1}(f_\Psi)$. But since ρ has p-power roots of unity as values, $\rho(f_\Psi)=1$ if p is odd. For $p=2$, a slightly more careful argument would show this. Hence (1) implies that $w(\Psi\rho)=w(\Psi)$ for such ρ.

It is interesting to notice what happens if E has good, supersingular reduction at p instead. Assume ρ has order p^n with $n\geqq 1$, p odd. One can verify that $f_\rho=p^{n+1}\mathcal{O}$. We then have $\Psi(f_\rho)=(-1)^{n+1}p^{n+1}=(-1)^{n+1}|\Psi(f_\rho)|$ and hence we get $w(\Psi\rho)=(-1)^{n+1}w(\Psi)$. The Birch and Swinnerton-Dyer conjecture implies that new points of infinite order appear in every other layer of the anti-cylotomic $\mathbb{Z}_p$-extension of K!

As a final remark in this section, we want to point out a similarity between the $L(\Psi\rho,1)$'s and the values $L(\Psi^{2k+1},k+1)$ discussed earlier. Namely, we can write $L(\Psi^{2k+1},k+1)$ as $L(\Psi^{2k+1}\mathcal{N}^{-k},1)$, where $\mathcal{N}$ is the norm grossencharacter of K. Since $\mathcal{N}=\Psi\overline{\Psi}$, we have $\Psi^{2k+1}\mathcal{N}^{-k}=\Psi\Phi^k$, where $\Phi=\Psi/\overline{\Psi}$. Note that $\Phi\circ c=\Phi^{-1}$ and similarly for all powers of Φ. Of course, the characters ρ of $\mathrm{Gal}(K_\infty^-/K)$ have the same property, as we've mentioned before.

2. Various Galois Groups and Modules

Let $\mathcal{F}_\infty=K(E_{p^\infty})$, the field generated by all p-power division points on E. Here E is defined over $\mathbb{Q}$ and $\mathrm{End}_K(E)=\mathcal{O}$, as before. Clearly, $\mathcal{F}_\infty$ is Galois over $\mathbb{Q}$.

The action of $G = \mathrm{Gal}(\mathscr{F}_\infty/K)$ on the Tate module for E gives an isomorphism $\imath$ of G into $(\mathcal{O}\otimes_{\mathbb{Z}}\mathbb{Z}_p)^x = \mathcal{O}_p^x$, say. Hence G is abelian. Therefore, complex conjugation c (in $\mathrm{Gal}(K/\mathbb{Q})$) induces a natural involution on G: $g \to c(g)$. There is also a natural action of $\mathrm{Gal}(K/\mathbb{Q})$ on $\mathcal{O}_p$ and it is easy to verify that $\imath$ is a $\mathrm{Gal}(K/\mathbb{Q})$-homomorphism. It is also not difficult to see that the image of $\imath$ is of finite index in $\mathcal{O}_p^x$. Hence $\mathscr{F}_\infty$ has a subfield $\mathscr{K}_\infty$ such that $\Gamma = \mathrm{Gal}(\mathscr{K}_\infty/K) \cong \mathbb{Z}_p^2$. In fact, $\mathscr{K}_\infty$ is just the composite of all $\mathbb{Z}_p$-extensions of K. Moreover, taking into account the action of $\mathrm{Gal}(K/\mathbb{Q})$ on $\mathcal{O}_p^x$, we find that, for odd p, $\Gamma = \Gamma^+ \times \Gamma^-$, where $\Gamma^+ \cong \Gamma^- \cong \mathbb{Z}_p$ and c acts on Γ^+ and Γ^- by $+1$ and -1 respectively. Corresponding to this decomposition of Γ, we have $\mathscr{K}_\infty = K_\infty^+ K_\infty^-$, where K_∞^+ is the cyclotomic $\mathbb{Z}_p$-extension of K, K_∞^- the anticyclotomic $\mathbb{Z}_p$-extension mentioned in the introduction. Fix a choice of topological generators σ (for Γ^+) and τ (for Γ^-).

Now assume from here on that p is odd, that $p \nmid N$, and that p splits in K: $p\mathcal{O} = \mathfrak{p}\mathfrak{p}^*$. The fields $K(E_{\mathfrak{p}})$, $K(E_{\mathfrak{p}^*})$ generated by the $\mathfrak{p}$ and $\mathfrak{p}^*$ division points on E are subfields of $\mathscr{F}_\infty$. As explained in [4], $\mathrm{Gal}(K(E_{\mathfrak{p}})/K) \cong (\mathbb{Z}/(p))^x$ and also $\mathfrak{p}$ is totally ramified in this extension. Since E has good reduction at p, we also have that $\mathfrak{p}^*$ is unramified in $K(E_{\mathfrak{p}})/K$. Similar remarks apply to $K(E_{\mathfrak{p}^*})/K$. Thus $K(E_{\mathfrak{p}}) \cap K(E_{\mathfrak{p}^*}) = K$, $K(E_{\mathfrak{p}})\,K(E_{\mathfrak{p}^*}) = K(E_p)$, and $\Delta = \mathrm{Gal}(K(E_p)/K) \cong (\mathbb{Z}/(p))^x \times (\mathbb{Z}/(p))^x$. Since $|\Delta| = (p-1)^2 = |\mathrm{tors}(\mathcal{O}_p^x)|$, we see that $G \cong \Delta \times \Gamma$, where Δ is identified with $\mathrm{Gal}(\mathscr{F}_\infty/\mathscr{K}_\infty)$ and Γ with $\mathrm{Gal}(\mathscr{F}_\infty/K(E_p))$. It should also be clear that the involution c acts on Δ by simply interchanging two copies of $(\mathbb{Z}/(p))^x$, i.e. $\Delta \cong \mathbb{Z}/(p-1)\,[\mathrm{Gal}(K/\mathbb{Q})]$.

There are two other subfields of $K(E_p)$ which we want to mention. One is $K(\mu_p)$, generated by the p-th roots of unity. We have $\mathrm{Gal}(K(\mu_p)/K) \cong (\mathbb{Z}/(p))^x$. Note that $K(\mu_{p^\infty}) = K(\mu_p)\,K_\infty^+$ which has its Galois group over K isomorphic to $\mathbb{Z}_p^x$. Also, let F_0^- denote the subfield of $K(E_p)$ fixed by Δ^{1+c} so that $\Delta^- = \mathrm{Gal}(F_0^-/K) \cong \Delta/\Delta^{1+c} \cong (\mathbb{Z}/(p))^x$ too. Later, we will consider the field $F_\infty^- = F_0^- K_\infty^-$. Note that $\mathrm{Gal}(F_\infty^-/K) \cong \mathbb{Z}_p^x$.

Let ψ and ψ^* denote the characters of Δ with values in $\mathbb{Z}_p^x$ which give the action of Δ on $E_{\mathfrak{p}}$ and $E_{\mathfrak{p}^*}$ respectively. Because E is defined over $\mathbb{Q}$ and complex conjugation maps $E_{\mathfrak{p}}$ to $E_{\mathfrak{p}^*}$, one can verify that $\psi^* = \psi \circ c$. Our remarks about Δ show that all the characters of Δ with values in $\mathbb{Z}_p^x$ are of the form $\psi^a(\psi^*)^b$, $0 \leqq a, b \leqq p-2$. Among these are $\omega = \psi\psi^*$, the determinant of the action of Δ on E_p (which of course is exactly the character giving the action of Δ on μ_p), and also $\phi = \psi/\psi^*$. This last character ϕ gives an isomorphism of $\mathrm{Gal}(F_0^-/K)$ to the $(p-1)st$ roots of unity in $\mathbb{Z}_p^x$. We will be especially interested in the characters χ of Δ with the property that $\chi(\chi \circ c) = \omega$; they are all of the form $\psi\phi^k$, $0 \leqq k \leqq p-2$.

Weil describes in [26] a way of forming a continuous Ω_p-valued character of $\mathrm{Gal}(\bar{K}/K)$ from any grossencharacter of type A_0 for K, given a fixed embedding of $\bar{\mathbb{Q}}$ into Ω_p. Here Ω_p denotes the completion of an algebraic closure of the p-adic numbers $\mathbb{Q}_p$. Fix such an embedding $\sigma_{\mathfrak{p}}$ so that it induces the $\mathfrak{p}$-adic valuation on K. In what follows, we will take for granted Weil's procedure and usually use the same notation for the grossencharacter and the Galois character. The simplest case is the norm grossencharacter $\mathscr{N}$ of K which, as a Galois character, gives the action on μ_{p^∞}. It factors through

$\mathrm{Gal}(K(\mu_{p^\infty})/K)$, giving a canonical isomorphism of this Galois group to $\mathbb{Z}_p^\times$. Since $\mu_{p^\infty} \subset \mathscr{F}_\infty$, we can regard $\mathscr{N}$ as a character of G.

If $\mathfrak{l}$ is any prime ideal of K not dividing N or p, then $\mathfrak{l}$ is unramified in $\mathscr{F}_\infty/K$ (by [23]) and Frob $(\mathfrak{l})$ (in G) acts on the $\mathfrak{p}$-power division points simply as multiplication by $\Psi(\mathfrak{l}) \in \mathcal{O}$, where Ψ is the grossencharacter attached to E. Furthermore, the action of $\mathcal{O}$ on $E_{\mathfrak{p}^\infty}$ obviously gives the isomorphism $\sigma_{\mathfrak{p}}\colon \mathcal{O} \to \mathbb{Z}_p$. Thus, the character of $\mathrm{Gal}(\bar{K}/K)$ obtained from Ψ (via $\sigma_{\mathfrak{p}}$) is just the character giving the Galois action on $E_{\mathfrak{p}^\infty}$ and so clearly factors through G. Hopefully, no confusion will result from letting Ψ denote this character of G also. Let Ψ^* denote the character of G giving the action on $E_{(\mathfrak{p}^*)^\infty}$. Then $\Psi\Psi^*$ is the determinant of the action of G on E_{p^∞} and hence must be $\mathscr{N}$. As grossencharacters of K, $\Psi\bar{\Psi} = \mathscr{N}$ and therefore Ψ^* is just the character of G obtained (via $\sigma_{\mathfrak{p}}$) from $\bar{\Psi}$. Also, since E is defined over $\mathbb{Q}$, we have $\Psi^* = \Psi \circ c$. The grossencharacter Φ defined at the end of Sect. 1 gives us a character $\Phi = \Psi/\Psi^*$ of G which induces an isomorphism of $\mathrm{Gal}(F_\infty^-/K)$ to $\mathbb{Z}_p^\times$.

Corresponding to the decomposition $G = \Delta \times \Gamma$, any character $\mathscr{C}$ of G can be written as a product $\mathscr{C} = \mathscr{C}_\Delta \mathscr{C}_\Gamma$ of characters of Δ and Γ. For the characters described above, we have $\Psi_\Delta = \psi$, $\Psi_\Delta^* = \psi^*$, $\mathscr{N}_\Delta = \omega$, and $\Phi_\Delta = \phi$. Also, the completed group ring $\Lambda_G = \mathbb{Z}_p[[G]]$ over $\mathbb{Z}_p$ is just $\Lambda_\Gamma[\Delta]$. Here $\Lambda_\Gamma = \mathbb{Z}_p[[\Gamma]]$ can be identified in the usual way with the power series ring $\mathbb{Z}_p[[S, T]]$, where $S = \sigma - \mathrm{id}_\Gamma$ and $T = \tau - \mathrm{id}_\Gamma$. If $\mathfrak{B}$ is any $\mathbb{Z}_p$-module on which G (or Λ_G) acts, then one can think of $\mathfrak{B}$ as a Γ (or Λ_Γ) module with a commuting action of Δ. If χ is any Ω_p-valued and therefore $\mathbb{Z}_p$-valued character of Δ, we can define a Γ (or Λ_Γ) submodule

$$\mathfrak{B}_\chi = \{b \in \mathfrak{B} \mid \delta(b) = \chi(\delta)\, b \text{ for all } \delta \in \Delta\},$$

the χ-component of $\mathfrak{B}$.

Let $\mathscr{M}_\infty$ denote the maximal abelian pro-p extension of $\mathscr{F}_\infty$ which is unramified outside the primes of $\mathscr{F}_\infty$ lying over $\mathfrak{p}$. Let $\mathfrak{X}_\infty = \mathrm{Gal}(\mathscr{M}_\infty/\mathscr{F}_\infty)$. Then $\mathfrak{X}_\infty$ is a $\mathbb{Z}_p$-module on which G acts by inner automorphisms (since $\mathscr{M}_\infty$ is clearly Galois over K). The action is $x \to gx = \bar{g}\, x\, \bar{g}^{-1}$ for $x \in \mathfrak{X}_\infty$, $g \in G$, where $\bar{g}$ is any extension of g to an automorphism of $\mathscr{M}_\infty$. Then $\mathfrak{X}_\infty$ can be considered as a Λ_G-module and turns out to be Noetherian and torsion as a Λ_Γ-module (see Coates [3]). We will denote a characteristic power series for the χ-component $(\mathfrak{X}_\infty)_\chi$, for any character χ of Δ, by $F_\chi(S, T)$, uniquely determined only up to an invertible factor in Λ_Γ.

We will also need to consider the maximal abelian pro-p-extension $\mathscr{L}_\infty$ of $\mathscr{F}_\infty$ which is unramified at all primes of $\mathscr{F}_\infty$. Then $\mathscr{L}_\infty \subseteq \mathscr{M}_\infty$ and the Λ_G-module $\mathfrak{H}_\infty = \mathrm{Gal}(\mathscr{L}_\infty/\mathscr{F}_\infty)$ is a quotient module of $\mathfrak{X}_\infty$. But $\mathscr{L}_\infty$ is actually Galois over $\mathbb{Q}$ and so we have an action of $\mathrm{Gal}(\mathscr{F}_\infty/\mathbb{Q})$ on $\mathfrak{H}_\infty$. This additional structure will be quite useful to us.

Let $\mathfrak{B}$ denote any $\mathbb{Z}_p$-module on which $\mathrm{Gal}(\mathscr{F}_\infty/\mathbb{Q})$ acts. Once and for all, choose some fixed extension of c (in $\mathrm{Gal}(K/\mathbb{Q})$) to an automorphism of $\bar{\mathbb{Q}}$ of order 2, which we also denote by c. We can regard c as an element of $\mathrm{Gal}(\mathscr{F}_\infty/\mathbb{Q})$ and so we have an involution $c\colon \mathfrak{B} \to \mathfrak{B}$, $b \to c(b) = cb$. If $g \in G$, $b \in \mathfrak{B}$, we have $c(gb) = c(g)\, c(b)$ as the following calculation shows: $c(gb) = cgb = cgc^{-1}cb = c(g)\, c(b)$. Let $\mathfrak{B}^c$ denote the G-module which is the same $\mathbb{Z}_p$-

module as $\mathfrak{B}$ but has a new action of G:

$$\text{new}\, gb = \text{old}\, c(g)\, b.$$

The mapping c then becomes a G-isomorphism $c\colon \mathfrak{B} \to \mathfrak{B}^c$. If χ is a character of Δ, we get a Γ-isomorphism $c\colon \mathfrak{B}_\chi \to (\mathfrak{B}^c)_\chi = (\mathfrak{B}_{\chi \circ c})^c$. Now assume that $\mathfrak{B}$ is a Noetherian, torsion Λ_Γ-module, and let $B_\chi(S, T)$ denote a characteristic power series for $\mathfrak{B}_\chi$. Then we have that

$$B_\chi(S, T) = B_{\chi \circ c}\left(S, -\frac{T}{1+T}\right)$$

up to an invertible factor in Λ_Γ. The reason is simply that $c(\sigma) = \sigma$, $c(\tau) = \tau^{-1}$ and so the involution c of Γ induces an involution of Λ_Γ which sends an arbitrary power series $P(S, T)$ to $P^c(S, T) = P\left(S, -\frac{T}{1+T}\right)$. These remarks apply to the characteristic power series $H_\chi(S, T)$ of the Λ_Γ-module $(\mathfrak{H}_\infty)_\chi$.

We want to define here a certain element of Λ_Γ which will play a crucial role in our later arguments. Let $\kappa = \mathcal{N}_\Gamma$, a $\mathbb{Z}_p^\times$-valued character of Γ with kernel Γ^-. Then $u = \kappa(\sigma)$ is a principal unit in $\mathbb{Z}_p$. We define

$$\Theta = \sigma - u\sigma^{-1},$$

which we will refer to as the "critical divisor". For odd p, it clearly differs by just an invertible factor in Λ_Γ from $\sigma - \sqrt{u} = S - (\sqrt{u} - 1)$, where $\sqrt{u}$ is the principal unit square root of u. Also note that $\Theta^c(S, T) = \Theta(S, T)$.

If Θ divides $H_\chi(S, T)$, then it would also divide $H_{\chi \circ c}(S, T)$. Since $H_\chi(S, T)$ divides $F_\chi(S, T)$, Θ would then divide both $F_\chi(S, T)$ and $F_{\chi \circ c}(S, T)$. For reasons that will be apparent later, we believe this cannot happen.

We now describe the kernel of the homomorphism $\mathfrak{X}_\infty \to \mathfrak{H}_\infty$. Let $\mathcal{U}_n = \prod_{v \mid \mathfrak{p}} U_{n,v}$, where $U_{n,v}$ is the group of principal units in the v-adic completion of $\mathcal{F}_n = K(E_{p^{n+1}})$ for $n \geqq 0$. Let E_n' be the group of global units of $\mathcal{F}_n$ which are $\equiv 1 \pmod{v}$ for all $v \mid \mathfrak{p}$. We can embed E_n' into $\mathcal{U}_n$ by the diagonal map. By class field theory, $\mathcal{U}_n/\bar{E}_n'$ is canonically isomorphic to the composite of inertia groups for all $v \mid \mathfrak{p}$ in $\mathrm{Gal}(\mathcal{M}_n/\mathcal{F}_n)$. Here $\mathcal{M}_n$ is the maximal abelian pro-p extension of $\mathcal{F}_n$ unramified outside of these v's and so $\mathcal{U}_n/\bar{E}_n' \cong \mathrm{Gal}(\mathcal{M}_n/\mathcal{L}_n)$, where $\mathcal{L}_n$ denotes the p-Hilbert class field of $\mathcal{F}_n$. Here $\bar{E}_n'$ is the closure of E_n' in $\mathcal{U}_n$. If E_n denotes the units of $\mathcal{F}_n \equiv 1 \pmod{v}$ for all v dividing p, then E_n is of prime to p index in E_n' and hence $\bar{E}_n = \bar{E}_n'$. Now we have the following commutative diagram for $m \geqq n \geqq 0$:

$$\begin{tikzcd}
\mathcal{U}_m/\bar{E}_m \arrow[r, "\text{norm}"] \arrow[d] & \mathcal{U}_n/\bar{E}_n \arrow[d] \\
\mathrm{Gal}(\mathcal{M}_m/\mathcal{L}_m) \arrow[r, "\text{rest.}"] & \mathrm{Gal}(\mathcal{M}_n/\mathcal{L}_n).
\end{tikzcd}$$

Since $\mathscr{M}_\infty = \bigcup \mathscr{M}_n$, $\mathscr{L}_\infty = \bigcup \mathscr{L}_n$ (as shown in a similar situation in [22]), we find that $\mathfrak{R}_\infty = \mathrm{Gal}(\mathscr{M}_\infty/\mathscr{L}_\infty)$ is isomorphic to $\varprojlim (\mathscr{U}_n/\bar{E}_n)$, the inverse limit defined by the norm maps. The isomorphism is compatible with the action of G and so is an isomorphism of Λ_G-modules. We have the exact sequence:

$$0 \to \mathfrak{R}_\infty \to \mathfrak{X}_\infty \to \mathfrak{H}_\infty \to 0. \tag{2}$$

The module $\mathfrak{R}_\infty$ is also related to a module considered by *R.* Yager in [28]. Namely, let $\mathscr{E}_n$ denote the group of elliptic units of $\mathscr{F}_n$ contained in E_n. Then $E_n/\mathscr{E}_n$ is finite. (We will say more about the definition of $\mathscr{E}_n$ in Sect. 3.) We let $\mathfrak{Y}_\infty = \varprojlim (\mathscr{U}_n/\bar{\mathscr{E}}_n)$, again defined by the norm maps. This time we have the exact sequence of Λ_G-modules

$$0 \to \mathfrak{J}_\infty \to \mathfrak{Y}_\infty \to \mathfrak{R}_\infty \to 0 \tag{3}$$

where $\mathfrak{J}_\infty = \varprojlim \bar{E}_n/\bar{\mathscr{E}}_n$. But, as we will explain, $\mathfrak{J}_\infty$ admits an action of $\mathrm{Gal}(\mathscr{F}_\infty/\mathbb{Q})$. Clearly, E_n is invariant under the action of $\mathrm{Gal}(\mathscr{F}_n/\mathbb{Q})$. So is $\mathscr{E}_n$, as we can see from the definition given in chapter 3 of [28], using the fact that the lattice L can be assumed to be invariant under complex conjugation in our case. Now, as Brumer has shown in [2],

$$\bar{E}_n \cong E_n \otimes_{\mathbb{Z}} \mathbb{Z}_p, \qquad \bar{\mathscr{E}}_n \cong \mathscr{E}_n \otimes_{\mathbb{Z}} \mathbb{Z}_p.$$

(The p-adic Leopoldt conjecture is valid for $\mathscr{F}_n$, since this field is abelian over the imaginary quadratic field K.) Hence $\bar{E}_n/\bar{\mathscr{E}}_n \cong (E_n/\mathscr{E}_n)_{p\text{-primary}}$, where the isomorphism is compatible with the action of $\mathrm{Gal}(\mathscr{F}_n/K)$. Therefore, $\mathfrak{J}_\infty$ is isomorphic to $\varprojlim (E_n/\mathscr{E}_n)_{p\text{-primary}}$ as G-modules. But the norm map $\mathscr{N}_{\mathscr{F}_m/\mathscr{F}_n}$ is compatible with the $\mathrm{Gal}(\mathscr{F}_m/\mathbb{Q})$ action on $E_m/\mathscr{E}_m$ since c simply permutes the elements of $\mathrm{Gal}(\mathscr{F}_m/\mathscr{F}_n)$ when it acts as an inner automorphism. This allows us to put a $\mathrm{Gal}(\mathscr{F}_\infty/\mathbb{Q})$ action on $\mathfrak{J}_\infty$.

For each character χ of Δ, let $G_\chi(S, T)$ denote a characteristic power series for $(\mathfrak{Y}_\infty)_\chi$, $J_\chi(S, T)$ one for $(\mathfrak{J}_\infty)_\chi$. The same considerations as before show that $J_{\chi\circ c}(S, T) = J_\chi^c(S, T)$, up to a factor in $\Lambda_\Gamma^\times$. Also $J_\chi(S, T)$ divides $G_\chi(S, T)$. We can summarize some of our observations in the following proposition.

Proposition 2. *Let Θ be the critical divisor. If Θ divides $H_\chi(S, T)$, then Θ divides both $F_\chi(S, T)$ and $F_{\chi\circ c}(S, T)$. If Θ divides $J_\chi(S, T)$ then Θ divides both $G_\chi(S, T)$ and $G_{\chi\circ c}(S, T)$.*

We will prove later (Proposition 6) that for χ of the form $\psi\phi^k$, Θ divides exactly one of the two power series $G_\chi(S, T)$ and $G_{\chi\circ c}(S, T)$, and therefore Θ cannot divide $J_\chi(S, T)$ for such χ. Note that, if $\chi = \psi\phi^{k_1}$, then $\chi\circ c = \psi^*\phi^{-k_1} = \psi\phi^{-1-k_1} = \psi\phi^{k_2}$ where $k_1 + k_2 \equiv -1 \pmod{p-1}$. If χ is not of the form $\psi\phi^k$, we can prove that Θ doesn't divide $G_\chi(S, T)$, but we will not go into this here.

3. *p*-Adic *L*-Functions and a Theorem of Yager

For primes p which split in the imaginary quadratic field K, Katz [13] has constructed a p-adic L-function $\mathfrak{L}_p(\mathscr{C})$ defined for all continuous characters

$\mathscr{C}: G \to \Omega_p^x$ and which takes values in Ω_p. It is constructed as an integral of $\mathscr{C}$ against a certain $\mathscr{A}$-valued measure on G, where $\mathscr{A}$ is the ring of integers in the completion of the maximal unramified extension of $\mathbb{Q}_p$. This implies that, for each (Ω_p-valued) character χ of Δ, there exists a power series $\mathscr{G}_\chi(S,T) \in \mathscr{A}[[S,T]]$ with the property that $\mathfrak{L}_p(\mathscr{C}) = \mathscr{C}_\Gamma(\mathscr{G}_\chi(S,T)) = \mathscr{G}_\chi(\mathscr{C}_\Gamma(\sigma)-1, \mathscr{C}_\Gamma(\tau)-1)$ for all $\mathscr{C}$ such that $\mathscr{C}_\Delta = \chi$. (Convergence is easily verified.) Thus $\mathfrak{L}_p(\mathscr{C})$ is an analytic function of two p-adic variables, because if κ_1, κ_2 are two independent characters of Γ with values in $1+p\mathbb{Z}_p$, then $\mathfrak{L}_p(\kappa_1^{s_1}\kappa_2^{s_2})$ will have a power series expansion in s_1 and s_2 convergent for $|s_1|, |s_2| \leqq 1$. The power series $\mathscr{G}_\chi(S,T)$ are uniquely determined by the interpolation property satisfied by $\mathfrak{L}_p(\mathscr{C})$, which we will now describe.

If $\mathscr{C}$ is a grossencharacter of type A_0 for K, then its values on ideals of K are in $\overline{\mathbb{Q}}$. We assume that we have chosen embeddings $\sigma_p: \overline{\mathbb{Q}} \to \Omega_p$ (which is p-adically continuous on K) and $\sigma_\infty: \overline{\mathbb{Q}} \to \mathbb{C}$. Then we obtain from $\mathscr{C}$ both a continuous Ω_p-valued character of $\mathrm{Gal}(\bar{K}/K)$, as we have already mentioned in Sect. 2, and a complex-valued grossencharacter. We denote both by $\mathscr{C}$. The defining property of $\mathfrak{L}_p(\mathscr{C})$ is:

$$\sigma_p^{-1}(A_p(\mathscr{C})\,\mathfrak{L}_p(\mathscr{C})) = \sigma_\infty^{-1}(A_\infty(\mathscr{C})\,L(\mathscr{C},1))$$

for characters $\mathscr{C}$ such that $\mathscr{C}$ is of type A_0, factors through G, and has infinity type (a,b) with $a \geqq 1$, $b \leqq 0$. Here $A_p(\mathscr{C})$, $A_\infty(\mathscr{C})$ are explicitly describable quantities (in Ω_p and $\mathbb{C}$ respectively) involving p-adic and complex periods, factorials, Euler factors, etc., which make the above expressions algebraic. We only need the fact that they are non-zero. For more details, see the explicit formulas in [13], page 546. Note however that we have applied the complex functional equation to his values. Also, $L(\mathscr{C},1)$ is the complex L-function for $\mathscr{C}$ evaluated at 1 (by analytic continuation if necessary). The properties of these functions are also described in [28] and [29]. In [29], Yager gives a new construction of the power series $\mathscr{G}_\chi(S,T)$.

In particular, if $\mathscr{C} = \Psi^a(\Psi^*)^b$ with integral a, b such that $a \geqq 1$, $b \leqq 0$, then $\mathfrak{L}_p(\mathscr{C})$ is "essentially" equal to $L(\Psi^a\bar{\Psi}^b, 1) = L(\Psi^{a-b}\mathscr{N}^b, 1) = L(\Psi^r, s)$ where $r = a-b$, $s = 1-b$ satisfy the inequalities $r \geqq 1$, $1 \leqq s \leqq r$. Damarell's theorem states that, up to a certain transcendental factor (which is part of $A_\infty(\mathscr{C})$), these L-values are algebraic [5]. For $k \geqq 0$, $L(\Psi^{2k+1}, k+1) = 0$ if and only if $\mathfrak{L}_p(\Psi\Phi^k) = 0$. This will allow us to prove the following result. We assume p is odd.

Proposition 3. *Let $0 \leqq k_0 \leqq p-2$. Then either all of the values $L(\Psi^{2k+1}, k+1)$ for $k \geqq 0$, $k \equiv k_0 \pmod{p-1}$ are zero, or only finitely many of them are zero. All will be zero if and only if the critical divisor Θ divides the power series $\mathscr{G}_{\psi\phi^{k_0}}(S,T)$.*

Proof. Up to an invertible factor in Λ_Γ, Θ is of the form $S-\alpha$ with α of absolute value <1. We can therefore write $\mathscr{G}_{\psi\phi^{k_0}}(S,T) = \mathscr{G}_{k_0}(T) + \Theta\mathscr{R}(S,T)$, where $\mathscr{G}_{k_0}(T) \in \mathscr{A}[[T]]$, $\mathscr{R}(S,T) \in \mathscr{A}[[S,T]]$. Now $\Psi^2(\sigma) = \Psi(\sigma)\Psi^*(\sigma) = \mathscr{N}(\sigma) = u$ and so $\Psi(\sigma) = \sqrt{u}$. Also, $\Phi(\sigma) = \Psi(\sigma)/\Psi^*(\sigma) = 1$. Hence $(\Psi\Phi^k)_\Gamma(\Theta) = \Psi\Phi^k(\Theta) = 0$ for all integers k. If $k \equiv k_0 \pmod{p-1}$, $(\Psi\Phi^k)_\Delta = \psi\phi^{k_0}$. It follows that

$$\mathfrak{L}_p(\Psi\Phi^k) = (\Psi\Phi^k)(\mathscr{G}_{k_0}(T)) = \mathscr{G}_{k_0}(v^{2k+1}-1)$$

where $v=\Psi(\tau)=(\Psi^*(\tau))^{-1}$. Note that v is a principal unit in $\mathbb{Z}_p$ and cannot be 1 since this would imply that τ is in the kernel of the two independent characters Ψ_Γ and Ψ_Γ^* of Γ, which is impossible. Now, by the Weierstrass Preparation Theorem (Bourbaki, Comm. Alg., Chapter 7) and the fact that p is a prime element in $\mathscr{A}$, either $\mathscr{G}_{k_0}(T)$ is the zero power series or we can write $\mathscr{G}_{k_0}(T)=p^e u(T)\mathscr{P}_{k_0}(T)$, where $e\geqq 0$, $u(T)$ is an invertible power series over $\mathscr{A}$, and $\mathscr{P}_{k_0}(T)$ is a so-called distinguished polynomial. Since a polynomial can have only finitely many zeros, we see that if $\mathfrak{L}_\mathfrak{p}(\Psi\Phi^k)=0$ for infinitely many $k\equiv k_0 \pmod{p-1}$, then $\mathscr{G}_{k_0}(T)$ is the zero power series. Proposition 3 follows.

Now if k_0 is such that $w_{k_0}=-1$, then we will have $w_k=-1$ for all $k\equiv k_0 \pmod{p-1}$, since m clearly divides $p-1$. (Recall that p is odd and splits in K.) Thus $L(\Psi^{2k+1}, k+1)$ vanishes trivially for such k, and hence we have the following result.

Corollary. *Let $0\leqq k_0\leqq p-2$. Assume $w_{k_0}=-1$. Then Θ divides $\mathscr{G}_{\psi\phi^{k_0}}(S,T)$.*

We also state the following very similar proposition. Here ϕ, which was defined as a certain character of Δ with values in the $(p-1)$st roots of unity in $\mathbb{Z}_p$, will also be regarded as a complex-valued Dirichlet character for K via $\sigma_\infty \sigma_\mathfrak{p}^{-1}$.

Proposition 4. *Let $0\leqq k_0\leqq p-2$. Then either all of the values $L(\Psi\phi^{k_0}\rho, 1)$ are zero or only finitely many are zero, where ρ varies over all complex-valued characters of* Gal(K_∞^-/K) *regarded as Dirichlet characters of K. All will be zero if and only if Θ divides $\mathscr{G}_{\psi\phi^{k_0}}(S,T)$.*

Proof. The proof is virtually identical to that of the previous proposition. Considering Ψ, ϕ, and ρ as Ω_p-valued via $\sigma_\mathfrak{p}\sigma_\infty^{-1}$, we have

$$\mathfrak{L}_\mathfrak{p}(\Psi\phi^{k_0}\rho)=(\Psi\phi^{k_0}\rho)_\Gamma(\mathscr{G}_{\psi\phi^{k_0}}(S,T))=\Psi_\Gamma\rho(\mathscr{G}_{k_0}(T))=\mathscr{G}_{k_0}(\rho(\tau)v-1),$$

since $\rho(\sigma)=1$ and so, as before, $\Psi_\Gamma\rho(\Theta)=0$. Since as ρ varies, $\rho(\tau)$ takes on all p-power roots of unity as values, the proof can be finished as before.

Yager's theorem relates the module $\mathfrak{Y}_\infty$ defined in Sect. 2 to Katz's function $\mathfrak{L}_\mathfrak{p}(\mathscr{C})$. Recall that, for each character χ of Δ, $G_\chi(S,T)$ is a characteristic power series of the Λ_Γ-module $(\mathfrak{Y}_\infty)_\chi$. We assume now, as Yager does, that $p\neq 2$ or 3.

Theorem (Yager). *For each character χ of Δ, $\mathscr{G}_\chi(S,T)=u_\chi(S,T)\,G_\chi(S,T)$ where $u_\chi(S,T)$ is invertible in $\mathscr{A}[[S,T]]$. Furthermore, there is an injective Λ_Γ-homomorphism of $(\mathfrak{Y}_\infty)_\chi$ to $\Lambda_\Gamma/(G_\chi(S,T))$ such that the cokernel is a finitely generated $\mathbb{Z}_p$-module.*

Thus $(\mathfrak{Y}_\infty)_\chi$ is pseudo-isomorphic to a cyclic Λ_Γ-module. Actually, for most χ, the Λ_Γ-isomorphism in Yager's theorem is surjective. (See Chap. 10 of [28] for more details.) We also remark that the critical divisor Θ divides $\mathscr{G}_\chi(S,T)$ (over $\mathscr{A}$) if and only if Θ divides $G_\chi(S,T)$ in Λ_Γ, as is easy to show. Hence our Propositions 3 and 4 could be stated in terms of the G's, if $p\neq 3$.

Actually, the above result is exactly what is proved in [28] only when the prime-to-p part of the conductor of χ is f_Ψ, which is sufficient for proving the results described in the introduction. Yager considers the powers of Ψ and $\overline{\Psi}$

as characters defined modulo f_Ψ and so his L-values may be non-primitive. As a result, the power series constructed in [28] may be only a multiple of $\mathcal{G}_\chi(S,T)$; the quotient will be a product of Iwasawa functions constructed from the missing Euler factors. Correspondingly, the elliptic units considered in [28] are defined by evaluating certain elliptic functions (of Robert) at division points whose annihilator is divisible by f_Ψ. By including the elliptic units constructed from "non-primitive" division points also, one gets a larger group of units $\mathscr{E}_n$. This group will be of finite index in E_n and will make the above theorem valid. We are grateful to J. Coates for pointing this difference out to us and to R. Yager for explaining how to modify his arguments to prove the result stated above.

Finally, we want to mention the so-called "two-variable main conjecture".

Conjecture. *For each character χ of Δ, $F_\chi(S,T)=G_\chi(S,T)$, up to an invertible factor in Λ_Γ.*

The exact sequences (2) and (3) show that this conjecture is equivalent to the statement that $H_\chi(S,T)$ and $J_\chi(S,T)$ differ just by a factor in $\Lambda_\Gamma^\times$. For a discussion of the above conjecture and the related one-variable conjectures, see [8].

4. The Non-Vanishing Theorem

Our approach to proving Theorem 4 is to show that, for $0\leqq k_0<m$, the numbers $L_k=L(\Psi^{2k+1},k+1)$ for $k\equiv k_0 \pmod m$ have Abel mean value equal to $(1+w_{k_0})\,L(\xi,1)$, where ξ is a certain non-principal Dirichlet character. Since this is nonzero for $w_{k_0}=+1$, we will then have that $L_k\neq 0$ for infinitely many $k\equiv k_0 \pmod m$. From this, we can prove that only finitely many of these L-values will be zero by a rather novel argument using p-adic L-functions for two different primes p.

First, we need an expression for L_k by a convergent series. Now $L(\Psi^{2k+1},s) = \sum_{n=1}^{\infty} a_n n^{-s}$, where the a_n's depend on k. Let $f(z)=\sum_{n=1}^{\infty} a_n q^n$ with $q=e^{2\pi i z}$ and let $g(t)=f(it/\sqrt{N_k})=\sum_{n=1}^{\infty} a_n e^{-Ant}$, where $A=A_k=2\pi/\sqrt{N_k}$. It is known that $f(z)$ is a cusp form for $\Gamma_0(N_k)$. (See Shimura [24] and Ogg [18].) As in Ogg, we obtain:

$$\Lambda(\Psi^{2k+1},s)=\int_1^\infty (t^{s-1}g(t)+w_k t^{d-s-1}g(t)\,dt.$$

Here $d=2k+2$. For $s=k+1$, this becomes

$$A^{-(k+1)}k!\,L_k=(1+w_k)\int_1^\infty t^k g(t)\,dt=(1+w_k)\sum_{n=1}^{\infty} a_n\int_1^\infty t^k e^{-Ant}\,dt.$$

Now $\int_1^\infty t^k e^{-Ant}\,dt=(An)^{-(k+1)}\int_{An}^\infty t^k e^{-t}\,dt$ and it is easy to show that

$$\int_x^\infty t^k e^{-t}\,dt=k!\,e^{-x}\sum_{j=0}^{k}\frac{x^j}{j!}.$$

This gives us the convergent series

$$\begin{aligned} L_k &= (1+w_k) \sum_{n=1}^{\infty} a_n n^{-(k+1)} e^{-An} \sum_{j=0}^{k} \frac{(An)^j}{j!} \\ &= (1+w_k) \sum_{\mathfrak{a}} \Psi^{2k+1}(\mathfrak{a}) \mathcal{N}(\mathfrak{a})^{-(k+1)} e^{-A\mathcal{N}(\mathfrak{a})} \sum_{j=0}^{k} \frac{(A\mathcal{N}(\mathfrak{a}))^j}{j!} \\ &= (1+w_k)\, G_k, \end{aligned}$$

with G_k denoting the sum over all ideals $\mathfrak{a}$ of $\mathcal{O}$.

We will assume throughout most of this section that $\Psi^{2k+1}(\mathfrak{a}) = \Psi(\mathfrak{a})^{2k+1}$ for all $k \geqq 0$, which is true except possibly if $K = \mathbb{Q}(\sqrt{-3})$. We will describe later how to handle this slightly awkward case. Assume $(\mathfrak{a}, f_\Psi) = 1$. Then $\Psi^{2k+1}(\mathfrak{a}) \mathcal{N}(\mathfrak{a})^{-(k+1)} = \bar{\Psi}(\mathfrak{a})^{-1} \Phi(\mathfrak{a})^k$, where $\Phi = \Psi/\bar{\Psi}$ as before. If $\mathfrak{a} = \bar{\mathfrak{a}}$, then $\mathfrak{a} = n\mathcal{O}$ for some integer $n \geqq 1$ (since, as we explained in Section 1, f_Ψ is divisible by the ramified prime in $K/\mathbb{Q}$) and we have $\Phi(\mathfrak{a}) = 1$, $\Psi(\mathfrak{a}) = \xi(n)\, n$, where ξ is the Dirichlet character with defining modulus N such that, for primes l not dividing N, $\xi(l) = +1$ if l splits in K and $\xi(l) = -1$ if l remains prime in K. Thus ξ is a (usually non-primitive) Dirichlet character for $\mathbb{Q}$ equivalent to the character associated with the extension $K/\mathbb{Q}$. If $\mathfrak{a} \neq \bar{\mathfrak{a}}$, $\Phi(\mathfrak{a})$ has absolute value 1, but cannot be equal to 1 (or even a root of unity). We can write $G_k = G_k' + G_k''$ where

$$G_k' = \sum_{n=1}^{\infty} \frac{\xi(n)}{n} e^{-An^2} \sum_{j=0}^{k} \frac{(An^2)^j}{j!} \tag{4}$$

$$G_k'' = \sum_{\mathfrak{a} \neq \bar{\mathfrak{a}}} \bar{\Psi}(\mathfrak{a})^{-1} \Phi(\mathfrak{a})^k e^{-A\mathcal{N}(\mathfrak{a})} \sum_{j=0}^{k} \frac{(A\mathcal{N}(\mathfrak{a}))^j}{j!} \tag{5}$$

The sum for G_k'' is over ideals $\mathfrak{a}$ of $\mathcal{O}$ prime to an integer $M (= N$ in the above case).

We will study the sums G_k', G_k'' in greater generality. We let A be any positive constant, ξ any non-principal Dirichlet character for $\mathbb{Q}$, and Ψ any function defined and nonzero for all ideals $\mathfrak{a} \neq \bar{\mathfrak{a}}$ relatively prime to some fixed integer $M \geqq 1$ and having the property that $\Psi(\mathfrak{a})$ is a generator of the ideal $\mathfrak{a}$. Consider the generating functions $F'(x) = \sum_{k=0}^{\infty} G_k' x^k$ and $F''(x) = \sum_{k=0}^{\infty} G_k'' x^k$. The calculation

$$\begin{aligned} \sum_{k=0}^{\infty} \left(\sum_{\mathfrak{a}} e^{-A\mathcal{N}(\mathfrak{a})} \sum_{j=0}^{k} \frac{(A\mathcal{N}(\mathfrak{a}))^j}{j!} \right) |x|^k &= \sum_{\mathfrak{a}} e^{-A\mathcal{N}(\mathfrak{a})} \sum_{j=0}^{\infty} \left(\frac{(A\mathcal{N}(\mathfrak{a}))^j}{j!} \sum_{k=j}^{\infty} |x|^k \right) \\ &= \frac{1}{1-|x|} \sum_{\mathfrak{a}} e^{-A\mathcal{N}(\mathfrak{a})} \sum_{j=0}^{\infty} \frac{(A\mathcal{N}(\mathfrak{a})\,|x|)^j}{j!} \\ &= \frac{1}{1-|x|} \sum_{\mathfrak{a}} e^{-(1-|x|)\, A\mathcal{N}(\mathfrak{a})} \end{aligned}$$

shows that the series defining $F'(x)$, $F''(x)$ are absolutely convergent for $|x| < 1$ and also justifies various changes in order of summation in similar sums in the rest of our argument.

The following lemma is the heart of our proof.

Lemma. *Let ζ be an m-th root of unity. Then*

(i) $\lim_{x\to 1}(1-x)F'(x)=L(\xi,1)$

(ii) $\lim_{x\to 1}(1-x)F'(\zeta x)=0$ *if* $\zeta\neq 1$

(iii) $\lim_{x\to 1}(1-x)F''(\zeta x)=0$ *for all* ζ

Proof. The series for $L(\xi,1)$ is convergent, since ξ is non-principal. If $\sum_{n=1}^{\infty} c_n$ is any convergent series, consider

$$C(x)=\sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty} c_n e^{-An}\sum_{j=0}^{k}\frac{(An)^j}{j!}\right)x^k=\frac{1}{1-x}\sum_{n=1}^{\infty} c_n u^n$$

where $u=u(x)=e^{-A(1-x)}$. As $x\to 1$ from below, so will $u(x)$, and a well-known theorem of Abel implies that $\lim_{x\to 1}(1-x)\,C(x)=\sum_{n=1}^{\infty} c_n$. If ζ is a root of unity, $\zeta\neq 1$, then $|u(\zeta x)|<1$ and is bounded away from 1 as $x\to 1$ from below (through real values). This makes it clear that $\lim_{x\to 1}(1-x)\,C(\zeta x)=0$ for such ζ. We obtain (i) and (ii) by letting $c_n=\xi(r)\,r^{-1}$ if $n=r^2$, $c_n=0$ otherwise.

To prove (iii), we have for $0\leqq x<1$,

$$\begin{aligned}
F''(\zeta x)&=\sum_{k=0}^{\infty}\left(\sum_{\mathfrak{a}\neq\bar{\mathfrak{a}}}\overline{\Psi}(\mathfrak{a})^{-1}\,\Phi(\mathfrak{a})^k\,e^{-A\mathcal{N}(\mathfrak{a})}\sum_{j=0}^{k}\frac{(A\mathcal{N}(\mathfrak{a}))^j}{j!}\right)(\zeta x)^k\\
&=\sum_{\mathfrak{a}\neq\bar{\mathfrak{a}}}\overline{\Psi}(\mathfrak{a})^{-1}\,e^{-A\mathcal{N}(\mathfrak{a})}\sum_{j=0}^{\infty}\frac{(A\mathcal{N}(\mathfrak{a}))^j}{j!}\sum_{k=j}^{\infty}(\Phi(\mathfrak{a})\,\zeta x)^k\\
&=\sum_{\mathfrak{a}\neq\bar{\mathfrak{a}}}\overline{\Psi}(\mathfrak{a})^{-1}\,e^{-A\mathcal{N}(\mathfrak{a})}\frac{e^{A\mathcal{N}(\mathfrak{a})\,\Phi(\mathfrak{a})\,\zeta x}}{1-\Phi(\mathfrak{a})\,\zeta x}\\
&=\sum_{\mathfrak{a}\neq\bar{\mathfrak{a}}}\frac{e^{-(1-\Phi(\mathfrak{a})\,\zeta x)\,A\mathcal{N}(\mathfrak{a})}}{\delta(\mathfrak{a},\zeta,x)},
\end{aligned}$$

where $\delta(\mathfrak{a},\zeta,x)=\overline{\Psi}(\mathfrak{a})-\Psi(\mathfrak{a})\,\zeta x$. We can write $\zeta=\eta/\bar{\eta}$ for some $\eta\in\mathcal{O}$ (by Hilbert's Theorem 90), and then

$$\delta(\mathfrak{a},\zeta,x)=\bar{\eta}^{-1}(\bar{\eta}\,\overline{\Psi}(\mathfrak{a})-\eta\,\Psi(\mathfrak{a})\,x)=\bar{\eta}^{-1}(\bar{\alpha}-\alpha x),$$

where $\alpha\in\mathcal{O}$, $\alpha\neq\bar{\alpha}$ (since they generate different ideals of $\mathcal{O}$). Since $|\mathrm{Im}(\bar{\alpha})|$ is bounded below for such α, we see that $|\delta(\mathfrak{a},\zeta,x)|>\delta>0$ for some fixed δ, independent of $\mathfrak{a}$ $(\neq\bar{\mathfrak{a}})$ and x $(0\leqq x<1)$. As for the exponent, we have

$$\mathrm{Re}((1-\Phi(\mathfrak{a})\,\zeta x)\,A\mathcal{N}(\mathfrak{a}))=\mathrm{Re}\left(1-\frac{\alpha}{\bar{\alpha}}x\right)B\alpha\bar{\alpha}=B\,\mathrm{Re}(\alpha\bar{\alpha}-\alpha^2 x),$$

where $B=A/\eta\bar{\eta}>0$. We may write $\alpha\,(\in\mathcal{O})$ as $\alpha=sa+tbi$, where $s,t\in\mathbb{Z}$, and a,b are fixed nonzero real constants depending only on $\mathcal{O}$. Then

$$\mathrm{Re}(\alpha\bar{\alpha}-\alpha^2 x)=s^2a^2(1-x)+t^2b^2(1+x).$$

Thus the final series we obtained for $F''(\zeta x)$ is dominated in absolute value by

$$\frac{1}{\delta}\sum_{s=-\infty}^{+\infty}\sum_{t=-\infty}^{+\infty} e^{-Bs^2a^2(1-x)}e^{-Bt^2b^2(1+x)}$$
$$<(\text{constant})\sum_{s=0}^{+\infty} e^{-Bs^2a^2(1-x)}$$
$$<(\text{constant})\int_0^\infty e^{-Bs^2a^2(1-x)}\,ds=0\left(\frac{1}{\sqrt{1-x}}\right)$$

as $x\to 1$. The final statement in our lemma follows immediately.

As a consequence of the above lemma, we see that for $0\leqq k_0<m$,

$$\lim_{x\to 1}(1-x)\sum_{\substack{k=0\\ k\equiv k_0(\operatorname{mod} m)}}^{\infty} G_k x^k=\frac{1}{m}L(\xi,1),$$

where $G_k=G'_k+G''_k$. Since $L(\xi,1)\neq 0$, $G_k\neq 0$ for infinitely many $k\equiv k_0(\operatorname{mod} m)$. This will allow us to prove the following.

Proposition 5. *Let* $0\leqq k_0<m$. *Assume* $w_{k_0}=+1$. *Then* $L(\Psi^{2k+1},k+1)\neq 0$ *for infinitely many* k's, $k\geqq 0$, $k\equiv k_0(\operatorname{mod} m)$.

Proof. We must just show that for an appropriate choice of A, ξ, Ψ, and M, we have $L_k=(1+w_k)G_k$. This is clear from what we said earlier if $K\neq\mathbb{Q}(\sqrt{-3})$, since N_k is independent of k. Assume $K=\mathbb{Q}(\sqrt{-3})$. If $k\not\equiv 1(\operatorname{mod} 3)$ (so that $2k+1,m)=1$), then we have $N_k=N$. We also have $\Psi^{2k+1}(\mathfrak{a})=\Psi(\mathfrak{a})^{2k+1}$ for all ideals $\mathfrak{a}$, where Ψ is the grossencharacter for E. If we take $M=N$ and ξ the Dirichlet character described earlier, then again $L_k=(1+w_k)G_k$ at least for $k\not\equiv 1(\operatorname{mod} 3)$. This leaves just $k_0=1$ and 4 to consider. We have $N_k=N_{k_0}$ if $k\equiv k_0(\operatorname{mod} m)$ and so we will choose $M=N_{k_0}$, $A=2\pi/\sqrt{N_{k_0}}$. For $n\in\mathbb{Z}$, we define $\xi(n)=\Psi^{2k_0+1}(n)/n^{2k_0+1}$, which must clearly be a Dirichlet character for $\mathbb{Q}$. (It is essentially the character described before but with N_{k_0} as defining modulus.) If N_{k_0} and N have the same prime factors we can choose Ψ exactly as above and will have $L_k=(1+w_k)G_k$ for $k\equiv k_0(\operatorname{mod} m)$. The other G_k's would not seem to be related to the L_k's necessarily. Finally, if there are primes l dividing N, but not N_{k_0}, let $\mathfrak{l}$ be a prime ideal factor of such l. Assume $\mathfrak{l}=(\beta)$. Let $\alpha\in\mathcal{O}$ be such that $(\alpha,N)=1$, $\alpha\beta\equiv 1\,(\operatorname{mod} f_{\Psi^{2k_0+1}})$. Then

$$\Psi^{2k_0+1}(\alpha\beta\mathcal{O})=(\alpha\beta)^{2k_0+1}=\Psi^{2k_0+1}(\alpha\mathcal{O})\,\Psi^{2k_0+1}(\mathfrak{l}),$$

and since $\Psi^{2k_0+1}(\alpha\mathcal{O})=(\Psi(\alpha\mathcal{O}))^{2k_0+1}$, we have $\Psi^{2k_0+1}(\mathfrak{l})=\beta_0^{2k_0+1}$ for some $\beta_0\in\mathcal{O}$ generating $\mathfrak{l}$. Define $\Psi_0(\mathfrak{l})=\beta_0$ for such $\mathfrak{l}$ and, for all prime ideals $\mathfrak{l}$ dividing N_{k_0} or prime to N, define $\Psi_0(\mathfrak{l})=\Psi(\mathfrak{l})$. Define $\Psi_0(\mathfrak{a})$ for all ideals of $\mathcal{O}$ by making Ψ_0 multiplicative. Then $\Psi^{2k+1}(\mathfrak{a})=\Psi_0(\mathfrak{a})^{2k+1}$ for all $k\equiv k_0(\operatorname{mod} m)$. We then take Ψ equal to the function Ψ_0 (and $\Phi=\Psi_0/\overline{\Psi}_0$) in (5) for all $k\geqq 0$. For $k\equiv k_0(\operatorname{mod} m)$, we will have $L_k=(1+w_k)G_k$, completing the proof of Proposition 5.

Remark. It follows from what we have proved that

$$\lim_{x\to 1}(1-x)\sum_{k=0}^{\infty} L(\Psi^{2k+1},k+1)\,x^k=L(\xi,1)$$

for some Dirichlet character ξ (except possibly if $K=\mathbb{Q}(\sqrt{-3})$ in which case the limit exists but might be slightly different). Now it is undoubtedly true that $L(\Psi^{2k+1}, k+1)\geqq 0$ for all k. This would follow from the Riemann Hypothesis for $L(\Psi^{2k+1}, s)$ for example, but is probably much more approachable. A similar result is proved by Zagier and Kohnen in [14] for Dirichlet series attached to modular forms for $SL(2,\mathbb{Z})$. If this could be proved, then a standard result in summability theory (Theorem 96 in Hardy [11]) would imply that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} L(\Psi^{2k+1}, k+1)$$

exists and is the same as the above Abelian limit. One consequence would be that $L(\Psi^{2k+1}, k+1)=o(k)$ as $k\to\infty$. It would be interesting to have a more precise result. Are these L-values bounded?

Now let $0\leqq k_0<m$ be such that $w_{k_0}=+1$. By the results proved above, there exists some $k_1\equiv k_0 \pmod{m}$ such that $L(\Psi^{2k_1+1}, k_1+1)\neq 0$. Let p be an odd prime for which E has good reduction and which splits in K (so that $m|p-1$). Then, by Proposition 3, $L(\Psi^{2k+1}, k+1)\neq 0$ for all but possibly finitely many $k\equiv k_1 \pmod{p-1}$. Choose an odd prime q such that $q\nmid N$, q splits in K, and $(p-1, q-1)=m$. The existence of such a prime q can be proved without too much difficulty using Dirichlet's theorem on primes in arithmetic progressions. Let k_2 be any integer such that $k_2\equiv k_0 \pmod{m}$. By the Chinese Remainder theorem, there are infinitely many k's such that $k\equiv k_1 \pmod{p-1}$ and $k\equiv k_2 \pmod{q-1}$. Therefore, again by Proposition 3, we see that $L(\Psi^{2k+1}, k+1)\neq 0$ for all but finitely many $k\equiv k_2 \pmod{q-1}$. It clearly follows that these L-values can vanish for at most finitely many $k\equiv k_0 \pmod{m}$. We have proved Theorem 4 and also the following result (using Yager's theorem).

Proposition 6. *Let* $0\leqq k_0\leqq p-2$, *where* p *is a prime* $(\neq 2,3)$ *for which* E *has good, ordinary reduction. Then* Θ *divides* $G_{\psi\phi^{k_0}}(S,T)$ *if and only if* $w_{k_0}=-1$.

5. The Selmer Group

We assume in this section that p is an odd prime for which E has good, ordinary reduction. Let F be any algebraic extension of K, not necessarily of finite degree. The Tate-Shafarevich group $Ш(E,F)$ is the subgroup of $H^1(\mathrm{Gal}(\bar{F}/F), E(\bar{F}))$ (or more briefly $H^1(F,E)$) defined by

$$Ш(E,F)=\bigcap_{v} \mathrm{Ker}(H^1(F,E)\to H^1(F_v,E)),$$

where the intersection is over all primes v of F and the maps between the cohomology groups are the natural restriction maps. We also define the modified Tate-Shafarevich group $Ш'(E,F)$ as above, but taking the intersection over all v not lying over $\mathfrak{p}$. If $\mathscr{A}$ is any $\mathcal{O}$-module and $n\geqq 1$, $\mathscr{A}_{\pi^n}$ denotes the kernel of π^n and $\mathscr{A}_{\pi^\infty}$ the π-primary submodule of $\mathscr{A}$. Then, for each $n\geqq 1$, the exact sequence of $\mathrm{Gal}(\bar{F}/F)$-modules

$$0\to E_{\pi^n}\to E(\bar{F}) \xrightarrow{\pi^n} E(\bar{F})\to 0$$

will give us an exact cohomology sequence

$$0 \to E(F)/\pi^n E(F) \to H^1(F, E_{\pi^n}) \to H^1(F, E)_{\pi^n} \to 0.$$

By taking the direct limits as $n \to \infty$, we obtain

$$0 \to E(F) \otimes_{\mathcal{O}} (K_{\mathfrak{p}}/\mathcal{O}_{\mathfrak{p}}) \to H^1(F, E_{\pi^\infty}) \to H^1(F, E)_{\pi^\infty} \to 0$$

The π^∞-Selmer group $S_{\pi^\infty}(E, F)$ is the subgroup of $H^1(F, E_{\pi^\infty})$ defined by the exact sequence

$$0 \to E(F) \otimes_{\mathcal{O}} (K_{\mathfrak{p}}/\mathcal{O}_{\mathfrak{p}}) \to S_{\pi^\infty}(E, F) \to Ш(E, F)_{\pi^\infty} \to 0.$$

We similarly define the modified Selmer group $S'_{\pi^\infty}(E, F)$ by using $Ш'$ in place of $Ш$ in this exact sequence.

We will need a number of results from Perrin-Riou [19], which we now briefly recall without proof. The references are to [19]. First of all, $S'_{\pi^\infty}(E, F) = S_{\pi^\infty}(E, F)$. If F is any subfield of $\mathscr{F}_\infty$ in which $\mathfrak{p}$ is "infinitely ramified". This condition is equivalent to saying that F contains a $\mathbb{Z}_p$-extension of K different than the unique one in which $\mathfrak{p}$ is unramified, denoted by N_∞^* in [19]. (Lemma 1.1) if F is a finite extension of K, then $S_{\pi^\infty}(E, F)$ is of finite index in $S'_{\pi^\infty}(E, F)$ (Proposition 3.1). Furthermore, the modified Selmer group behaves well Galois theoretically. Let F be a subfield of $\mathscr{F}_\infty$ and let $H = \mathrm{Gal}(\mathscr{F}_\infty/F)$. If F either contains $\mathscr{K}_\infty$ or at least a $\mathbb{Z}_p$-extension of K different than the one in which $\mathfrak{p}^*$ is unramified (denoted by N_∞ in [19]), then $S'_{\pi^\infty}(E, F) = S'_{\pi^\infty}(E, \mathscr{F}_\infty)^H$. Here the H indicates the subgroup of H-invariant elements. The isomorphism is the natural restriction map. (Proposition 1.3 together with the fact that the "prime to p" subgroup of H obviously acts well Galois theoretically.) If F is a finite extension of K, then the map $S'_{\pi^\infty}(E, F) \to S'_{\pi^\infty}(E, \mathscr{F}_\infty)^H$ has finite kernel and cokernel (Lemma 3.4).

We will also need the following result of Coates, which is derived from the Kummer theory for E corresponding to the π^∞-division points E_{π^∞}. (See [3′]. The proof is similar to the one variable version in [3].) The result is that there exists a perfect pairing of the compact G-module $\mathfrak{X}_\infty$ (defined in Sect. 2) and the discrete G-module $\mathscr{S}_\infty = S_{\pi^\infty}(E, \mathscr{F}_\infty)$

$$\mathfrak{X}_\infty \times \mathscr{S}_\infty \to E_{\pi^\infty}$$

such that

$$g(x, s) = (gx, gs) \tag{6}$$

for all $g \in G$, $x \in \mathfrak{X}_\infty$, and $s \in \mathscr{S}_\infty$. From this one gets a perfect pairing of Γ-modules

$$(\mathfrak{X}_\infty)_\chi \times (\mathscr{S}_\infty)_{\psi\chi^{-1}} \to E_{\pi^\infty}$$

for each character χ of Δ, where we recall that ψ is the character giving the action of Δ on E_π (and also on E_{π^∞} when Δ is regarded as a subgroup of G).

Before going further, we introduce the following terminology, taken from Mazur [17]. If $\mathscr{A}$ is a discrete $\mathbb{Z}_p$-module, we say $\mathscr{A}$ has $\mathbb{Z}_p$-corank r if its dual $\hat{\mathscr{A}}$ is a finitely generated $\mathbb{Z}_p$-module of rank r. This means $\mathscr{A}$ has a subgroup of finite index isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^r$. If $\mathscr{A}$ is a discrete Λ_Γ-module, such that $\hat{\mathscr{A}}$ is a torsion Λ_Γ-module, we say $\mathscr{A}$ is Λ_Γ-cotorsion. If $\hat{\mathscr{A}}$ is a finitely generated Λ_Γ-module of rank r, then we refer to r as the Λ_Γ-corank of $\mathscr{A}$. We use the same terminology for Λ_{Γ^-}-modules.

We will consider the π^∞-Selmer group for the field $F_\infty^- = \mathscr{F}_\infty^H$, where $H = \{gc(g) | g \in G\}$, described in Sect. 2. Recall that $\mathrm{Gal}(F_\infty^-/K) \cong \Delta^- \times \Gamma^-$ and that Φ gives an isomorphism of this Galois group to $\mathbb{Z}_p^\times$. It is interesting to note that, for $m=2$, F_∞^- as well as the isomorphism Φ does not depend on the elliptic curve $E/\mathbb{Q}$. One might consider F_∞^- as an "anticyclotomic" analogue of the field $K(\mu_{p^\infty})$.

Now the Galois group $(\mathfrak{X}_\infty)_\chi$ is a Noetherian, torsion Λ_Γ-module [3]. If the critical divisor Θ divides the characteristic power series $F_\chi(S,T)$ for this module, then $(\mathfrak{X}_\infty)_\chi$ will have a quotient pseudo-isomorphic to $\bigoplus_{j=1}^{a} \Lambda_\Gamma/(\Theta^{e_j})$ with the e_j's $\geqq 1$ for some positive integer $a = a_\chi$. We would then have that $(\mathfrak{X}_\infty)_\chi/\Theta(\mathfrak{X}_\infty)_\chi$ is pseudo-isomorphic to $(\Lambda_\Gamma/(\Theta))^a$. But by (6), $(\mathfrak{X}_\infty)_\chi/\Theta(\mathfrak{X}_\infty)_\chi$ is isomorphic to $\mathrm{Hom}((\mathscr{S}_\infty)_{\psi\chi^{-1}}^{\Gamma^+}, \mathbb{Q}_p/\mathbb{Z}_p)$ as a group and so this last module is pseudo-isomorphic to $(\Lambda_\Gamma/(S))^a$. If $\chi = \psi\phi^k$, we see that $\mathrm{Hom}(S_{\pi^\infty}(E, F_\infty^-)_{\phi^{-k}}, \mathbb{Q}_p/\mathbb{Z}_p)$ is pseudo-isomorphic as a Λ_{Γ^-}-module to $\Lambda_{\Gamma^-}^a \oplus M$, where M is a torsion Λ_{Γ^-}-module and $a = a_\chi$. This clearly implies that $S_{\pi^\infty}(E, F_n^-)_{\phi^{-k}}$ contains a subgroup isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^{ap^n}$ for $n = 0, 1, 2, \ldots$, where F_n^- is the cyclic extension of K of degree $(p-1)p^n$ in F_∞^-. These arguments are reversible. We summarize in the following proposition.

Proposition 7. *Let $0 \leqq k \leqq p-2$ and let $\chi = \psi\phi^k$. Then Θ divides $F_\chi(S,T)$ if and only if $S_{\pi^\infty}(E, F_\infty^-)_{\phi^{-k}}$ has positive Λ_{Γ^-}-corank. The corank is $a = a_\chi$, as defined above. Also, $S_{\pi^\infty}(E, F_n^-)_{\phi^{-k}}$ contains a subgroup isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^{ap^n}$ for $n \geqq 0$.*

In particular, let $k=0$. If Θ divides $F_\psi(S,T)$, we see that $S_{\pi^\infty}(E,K)$ contains a subgroup isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^a$, $a \geqq 1$. We want to prove the same statement about $S_{p^\infty}(E, \mathbb{Q})$. Since $E_{p^\infty} = E_{\pi^\infty} \times E_{(\pi^*)^\infty}$, it is clear that

$$S_{p^\infty}(E,K) \cong S_{\pi^\infty}(E,K) \times S_{(\pi^*)^\infty}(E,K) \cong B \times B^*,$$

say. Also, since E is defined over $\mathbb{Q}$, there is a natural action of c on $H^1(K, E_{p^\infty})$ (which is defined by sending a cocycle f to f^c, where $f^c(g) = c(f(c^{-1}gc))$ for $g \in G$). Since c acting on E_{p^∞} maps E_{π^∞} isomorphically to $E_{(\pi^*)^\infty}$, it is easy to see that $f \to f^c$ gives us an isomorphism of $H^1(K, E_{\pi^\infty})$ to $H^1(K, E_{(\pi^*)^\infty})$ and also of B to B^*, $b \to c(b)$. Therefore $S_{p^\infty}(E,K)$ contains a subgroup $\{(b, c(b)) | b \in B\}$ isomorphic to B and invariant under $\mathrm{Gal}(K/\mathbb{Q})$. But, because p is odd, $S_{p^\infty}(E, \mathbb{Q}) = S_{p^\infty}(E,K)^{\mathrm{Gal}(K/\mathbb{Q})}$, and the result we want follows.

It is also worth making the following remark. Assume that $\chi = \psi\phi^k$ and that $\Psi_\Gamma(F_\chi(S,T)) = 0$. Without too much difficulty, this can be seen to imply that $(\mathfrak{X}_\infty)_\chi$ has a quotient isomorphic to $\mathbb{Z}_p$ on which Γ acts by Ψ_Γ. Then $(\mathscr{S}_\infty)_{\phi^{-k}}$ will have a subgroup isomorphic to $\mathbb{Q}_p/\mathbb{Z}_p$ on which Γ acts trivially, by (6). It follows that $S_{\pi^\infty}(E, F_0^-)_{\phi^{-k}}$ is infinite. For $k=0$, we would get that $S_{p^\infty}(E, \mathbb{Q})$ is infinite.

We end this section with some remarks which would be of interest only if it turns out to be possible for an elliptic curve to have an infinite Tate-Shafarevich group. Let E be an elliptic curve defined over any number field F. It is known that $Ш(E,F)$ has only finitely many elements of order p, for any prime p, and hence the p-primary subgroup of $Ш(E,F)$ is isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^{t_p} \times$ (a finite group) for some $t_p \geqq 0$. Let $r = \operatorname{rank}_{\mathbb{Z}}(E(F))$. It follows that $S_{p^\infty}(E,F)$ contains a subgroup of finite index isomorphic to $(\mathbb{Q}_p/\mathbb{Z}_p)^{d_p}$, where $d_p = r + t_p$. It is tempting to state a variant on the conjecture of Birch and Swinnerton-Dyer (suggested by Coates): The order of vanishing of $L(E/F, s)$ at $s=1$ should be d_p. Obviously, this statement implies that d_p is independent of p, a fact which would be trivial if $Ш(E,F)$ is finite. We can give a conditional but rather convincing proof of the following result. If E is an elliptic curve defined over $F = \mathbb{Q}$ with complex multiplication and if w is the sign in the functional equation for $L(E/\mathbb{Q}, s)$, then $w = (-1)^{d_p}$ for all odd primes p where E has good, ordinary reduction. Thus at least the parity of d_p (and therefore t_p) would be independent of p for these primes. We will discuss this in [9].

6. The Proof of Theorem 3

We will prove a somewhat more general result. First we recall the following theorem of Rubin [20], for an elliptic curve defined over K and with complex multiplication by $\mathcal{O}$.

Theorem. *Let M be a finite, abelian extension of K and let χ be a complex valued character of* $\mathrm{Gal}(M/K)$. *If χ occurs in the representation space* $E(M) \otimes_{\mathcal{O}} \mathbb{C}$ *of* $\mathrm{Gal}(M/K)$, *then* $L(\bar{\Psi}\chi, 1) = 0$.

We remark that, by the functional equation, $L(\bar{\Psi}\chi, 1) = 0$ if and only if $L(\Psi\chi^{-1}, 1) = 0$.

In what follows, we will regard the character ϕ as either complex-valued or Ω_p-valued (via σ_∞ or σ_p). We assume that E is an elliptic curve defined over $\mathbb{Q}$ with complex multiplication by $\mathcal{O}$, and that p is an odd prime for which E has good, ordinary reduction. Let $0 \leqq k_0 \leqq p-2$. Denote by $T_K(\phi^{-k_0})$ the character of the irreducible representation of $\mathrm{Gal}(F_0^-/K)$ over K which has ϕ^{-k_0} as a constituent. The other constituents are ϕ^{-tk_0} where $(t, p-1) = 1$ and $t \equiv 1 \pmod{m}$. (We note in passing that $w_{k_0} = w_{tk_0}$ for such t.) We have the following result which includes Theorem 3.

Proposition 8. *Assume $w_{k_0} = +1$. Considering $E(F_n^-) \otimes_{\mathcal{O}} K$ as a representation space for $\Delta^- \cong \mathrm{Gal}(F_0^-/K)$ over K, the dimension of the $T_K(\phi^{-k_0})$-component is bounded as $n \to \infty$.*

Proof. If this were false, then there would exist infinitely many n's and complex-valued characters ρ of $\mathrm{Gal}(F_n^-/K)$ of order p^n such that $\phi^{-k_0}\rho$ occurs in $E(F_n^-) \otimes_{\mathcal{O}} \mathbb{C}$. Thus, by Rubin's theorem, $L(\Psi\phi^{k_0}\rho^{-1}, 1)$ must vanish for infinitely many characters ρ^{-1} of $\mathrm{Gal}(K_\infty^-/K)$. By Proposition 4, this implies that Θ divides $\mathscr{G}_{\psi\phi^{k_0}}(S,T)$, now regarding ϕ as Ω_p-valued. By Proposition 6 (or at least its proof if $p=3$) this contradicts our assumption that $w_{k_0} = +1$.

One can also apply Rubin's theorem to certain other $\mathbb{Z}_p$-extensions, where the necessary non-vanishing result is more trivial. To describe what we mean let $\mathscr{E}$ denote the set of all $\mathbb{Z}_p$-extensions of K. We can identify $\mathscr{E}$ with the projective line $\mathbb{P}_1(\mathbb{Q}_p)$ in the following way. Let $\lambda=\Psi_\Gamma$, $\kappa=\mathscr{N}_\Gamma$. If $\gamma\in\mathbb{P}_1(\mathbb{Q}_p)$ is represented by (a,b) with $a,b\in\mathbb{Z}_p$, not both zero, then $\kappa^{-a}\lambda^b\colon\Gamma\to 1+p\mathbb{Z}_p$ factors through a $\mathbb{Z}_p$-extension $K_\infty^{(\gamma)}$ of K, depending just on γ. If γ is the point at ∞, $K_\infty^{(\gamma)}$ is the cyclotomic $\mathbb{Z}_p$-extension K_∞^+. Otherwise (if $b\neq 0$) we can identify γ with $a/b\in\mathbb{Q}_p$. The anti-cyclotomic $\mathbb{Z}_p$-extension corresponds to $\gamma=\frac{1}{2}$. The extensions N_∞ and N_∞^* mentioned in Sect. 5 correspond to $\gamma=0$ and 1 respectively. We have a natural topology on $\mathscr{E}$ from the identification with $\mathbb{P}_1(\mathbb{Q}_p)$. The following result concerns a certain dense subset.

Proposition 9. *Let $K_n^{(\gamma)}$ denote the n-th layer in the $\mathbb{Z}_p$-extension $K_\infty^{(\gamma)}$. Assume γ is rational, $0\leqq\gamma\leqq 1$, and $\gamma\neq\frac{1}{2}$. Then* $\operatorname{rank}_{\mathcal{O}}(E(K_n^{(\gamma)}))$ *is bounded as $n\to\infty$.*

Proof. If $\operatorname{rank}_{\mathcal{O}}(E(K_n^{(\gamma)}))$ were unbounded, then we would have infinitely many complex-valued characters ρ of $\mathrm{Gal}(K_\infty^{(\gamma)}/K)$ such that $L(\Psi\rho,1)=0$. Viewing these characters as Ω_p-valued (via $\sigma_p\sigma_\infty^{-1}$), it would follow that $\mathfrak{L}_p(\Psi\rho)=0$ also. Therefore $\mathfrak{L}_p(\Psi\mathscr{C})$ must vanish for all Ω_p-valued characters $\mathscr{C}$ of $\mathrm{Gal}(K_\infty^{(\gamma)}/K)$. If $\gamma=a/b$ with a, b nonnegative integers, put $r=1+(p-1)bt$, $s=1+(p-1)at$ for integral $t\geqq 1$. We have $\mathfrak{L}_p(\Psi\Psi^{(p-1)bt}\mathscr{N}^{-(p-1)at})=0$ and, since $1\leqq s\leqq r$, it follows that $L(\Psi^r,s)=0$ also. But, because $\gamma\neq\frac{1}{2}$, either r and s or r and $r+1-s$ are in the region where the Euler products for these L-functions converge. Clearly, none of these L-values are zero.

It is shown in [19] that there are only finitely many $\mathbb{Z}_p$-extensions of K in which E has unbounded rank. We would tend to believe that in fact only the anti-cyclotomic $\mathbb{Z}_p$-extensions could have this property. For the cyclotomic $\mathbb{Z}_p$-extension, Rubin and Wiles [21] have proved that the rank must be bounded for a large class of elliptic curves defined over $\mathbb{Q}$ with complex multiplication by K.

It is not difficult to deduce from the results of this section a "Mordell-Weil theorem" for certain $\mathbb{Z}_p$-extensions. For example, we can prove that if K_∞ is any $\mathbb{Z}_p$-extension of K such that $\operatorname{rank}_{\mathcal{O}}(E(K_n))$ is bounded, then $E(K_\infty)$ is finitely generated. It is enough to prove that the torsion subgroup of $E(K_\infty)$ is finite, as the following argument (taken from [17]) shows. Assume $t=|\mathrm{tors}(E(K_\infty))|$ is finite. Let n be such that E achieves its maximal rank over K_n. Thus, if $m\geqq n$, $E(K_m)/E(K_n)$ is finite. Let g be a generator of $\mathrm{Gal}(K_\infty/K_n)$. If $P\in E(K_m)$, then dP is invariant under g for some $d\geqq 1$. Thus, $d(P^g-P)=0_E$. We must have $t(P^g-P)=0_E$. This implies that $[E(K_m)\colon E(K_n)]$ is bounded as $m\to\infty$. Obviously it follows that $E(K_m)=E(K_\infty)$ if m is large enough.

We first show that the prime-to-p torsion subgroup of $E(\mathscr{F}_\infty)$ is finite. Let l be a prime, $l\neq p$. The field $K(E_{l^\infty})$ is the composite of a finite extension of K and the extension of K with Galois group $\mathbb{Z}_l^2$, in which only primes over l can ramify (see [12]). It follows easily that any infinite subgroup of E_{l^∞} generates an extension of K in which at least one prime of K dividing l must be infinitely ramified. Since l is at most finitely ramified in F_∞, we get that the l-primary subgroup of $E(\mathscr{F}_\infty)$ is finite. If $l\nmid 2N$, the l-torsion must be trivial. To see this, note that such l either split or remain prime in K (since $\mathrm{disc}(K)$

divides N). Also $E(\mathscr{F}_\infty)_l$ is an $\mathcal{O}$-submodule of E_l. If l splits as $\mathfrak{l}\mathfrak{l}^*$, then the nontrivial $\mathcal{O}$-submodules of E_l are $E_{\mathfrak{l}}$, $E_{\mathfrak{l}^*}$, and E_l itself. As mentioned earlier, $K(E_{\mathfrak{l}})$ has degree $l-1$ over K and $\mathfrak{l}$ is totally ramified. Since only primes where E has bad reduction and p can ramify in $\mathscr{F}_\infty$, it is clear that $E(\mathscr{F}_\infty)$ has no $\mathfrak{l}$ (or $\mathfrak{l}^*$) torsion and hence no l-torsion. If l remains prime, then E_l has no proper $\mathcal{O}$-submodules. Since $K(E_l)$ contains $K(\mu_l)$ and l is certainly ramified nontrivially in $K(\mu_l)$ for $l \nmid 2\operatorname{disc}(K)$, we again see that $E(\mathscr{F}_\infty)$ has no l-torsion. All of these remarks obviously justify the statement made at the beginning of this paragraph.

Finally, it is simple to show that $E(K_\infty)$ has no p-torsion. The only nontrivial $\mathcal{O}$-submodules of E_p are $E_{\mathfrak{p}}$, $E_{\mathfrak{p}^*}$, and E_p. But $K(E_{\mathfrak{p}})$ has degree $p-1$ over K and therefore cannot be contained in K_∞. Similarly for $K(E_{\mathfrak{p}^*})$. Since $E(K_\infty)$ is an $\mathcal{O}$-module, $E(K_\infty)_p$ must be trivial. Combining our observations, it should be clear that $\operatorname{tors}(E(K_\infty))$ is in fact finite.

7. Concluding Arguments

Now we can give the final steps of the proofs of Theorems 1 and 2. By Proposition 7 and the discussion after it, it will be enough to prove the following result. We must assume $p \neq 2$ or 3, as well as the other usual assumptions.

Proposition 10. *If* $w_{k_0} = -1$, *then* Θ *divides* $F_{\psi\phi^{k_0}}(S,T)$.

Proof. Let $\chi_0 = \psi\phi^{k_0}$. Choose $k_1 \geqq 0$ so that $k_0 + k_1 \equiv -1 \pmod{p-1}$. By Proposition 1, $w_{k_0} w_{k_1} = -1$ and so $w_{k_1} = +1$. If $\chi_1 = \psi\phi^{k_1}$, then Proposition 6 implies that Θ divides $G_{\chi_0}(S,T)$, but not $G_{\chi_1}(S,T)$. Since $\chi_1 = \chi_0 \circ c$, it follows from Proposition 2 that Θ cannot divide $J_{\chi_0}(S,T)$. Now the exact sequence

$$0 \to (\mathfrak{J}_\infty)_{\chi_0} \to (\mathfrak{Y}_\infty)_{\chi_0} \to (\mathfrak{R}_\infty)_{\chi_0} \to 0$$

shows that Θ must divide the characteristic power series for $(\mathfrak{R}_\infty)_{\chi_0}$. Finally, the exact sequence

$$0 \to (\mathfrak{R}_\infty)_{\chi_0} \to (\mathfrak{X}_\infty)_{\chi_0} \to (\mathfrak{H}_\infty)_{\chi_0} \to 0$$

proves that Θ divides $F_{\chi_0}(S,T)$.

Of course, it seems almost certain that Θ divides $F_{\psi\phi^k}(S,T)$ if and only if $w_k = -1$. Obviously this would be a consequence of the two-variable main conjecture stated at the end of Sect. 3. It would also be a consequence of the following hypothesis: the π-primary subgroup of $\text{Ш}(E, F_n^-)$ is finite or at least of bounded $\mathbb{Z}_p$-corank for all $n \geqq 0$. For this hypothesis together with Proposition 8 implies $S_{\pi^\infty}(E, F_n^-)_{\phi^{-k}}$ has bounded $\mathbb{Z}_p$-rank as $n \to \infty$ (or, what amounts to the same $S_{\pi^\infty}(E, F_\infty^-)_{\phi^{-k}}$ is cotorsion as a Λ_{Γ^-}-module) when $w_k = +1$. By Proposition 7, this in turn is equivalent to the assertion that, if $w_k = +1$, then $F_{\psi\phi^k}(S,T)$ is not divisible by Θ. Finally, another equivalent statement is that Θ does not divide $H_{\psi\phi^k}(S,T)$ for any k. This follows from the exact sequences (2) and (3), Propositions 2 and 6.

If one assumes any of the statements in the last paragraph, then one could determine the a_χ's referred to in Proposition 7. This is because $(\mathfrak{R}_\infty)_\chi$ is always pseudo-isomorphic to a cyclic Λ_Γ-module. We would have the result that $a_\chi = 0$ or 1 for $\chi = \psi\phi^k$ depending just on the value of w_k. These comments should explain the remarks we made in the introduction about Mazur's invariant. We can also see that, if one assumes the hypothesis about $Ш$ stated above, then Mazur's invariant for the tower F_∞^-/F_0^- is exactly $(p-1)/2$.

Can one prove a result similar to Theorem 1 when $L(E/\mathbb{Q}, s)$ has an even order zero at $s=1$? We can only make the following remark. Assume $L(\Psi, 1) = 0$, but $w = +1$. Then $\mathfrak{L}_\mathfrak{p}(\Psi)$ vanishes and hence the power series $G_\psi(S, T)$ will have at least one irreducible factor $P(S, T)$ in Λ_Γ such that $\Psi_\Gamma(P(S, T)) = 0$. By Proposition 6, $P(S, T)$ is distinct from Θ. It would be enough to show that $P(S, T)$ divides $F_\psi(S, T)$. For then one could use the results from Sect. 5 to prove that $S_{p^\infty}(E, \mathbb{Q})$ is infinite. If $P(S, T)$ divides the characteristic power series for $(\mathfrak{R}_\infty)_\psi$, we would clearly have what we want. If not, then $P(S, T)$ must divide $J_\psi(S, T)$. Therefore $P^c(S, T)$ would divide $J_{\psi^*}(S, T)$ and hence $G_{\psi^*}(S, T)$. Since Θ also divides $G_{\psi^*}(S, T)$, and since both Θ and $P^c(S, T)$ vanish at Ψ_Γ^*, we would find that $\mathfrak{L}_\mathfrak{p}(\mathscr{C})$ has at least a second order zero at $\mathscr{C} = \Psi^*$ as an analytic function in two variables. That is, its power series expansion about Ψ^* would have to begin with at least quadratic terms. Thus we are led to a question about the behavior of $\mathfrak{L}_\mathfrak{p}(\mathscr{C})$ at a grossencharacter of type A_0 outside of the range of interpolation.

References

1. Arthaud, N.: On Birch and Swinnerton-Dyer's conjecture for elliptic curves with complex multiplication. I. Comp. Math. **37**, 209–232 (1978)
2. Brumer, A.: On the units of algebraic number fields. Mathematika **14**, 121–124 (1967)
3. Coates, J.: Infinite descent on elliptic curves with complex multiplication. To appear in Progress in Mathematics. Boston: Birkhauser

3'. Coates, J.: Elliptic curves with complex multiplication. Hermann Weyl lectures, 1979, Annals of Math. Studies, to appear

4. Coates, J., Wiles, A.: On the conjecture of Birch and Swinnerton-Dyer. Invent. Math. **39**, 223–251 (1977)
5. Damarell, R.: L-functions of elliptic curves with complex multiplication I. Acta Arith. **17**, 287–301 (1970)
6. Deuring, M.: Die Zetafunktionen einer algebraischen Kurve vom Geschlecht Eins I, II, III, IV. Nachr. Akad. Wiss. Göttingen, 85–94 (1953); 13–42 (1955); 37–76 (1956); 53–80 (1957)
7. Frohlich, A., Queyrut, J.: On the functional equation of the Artin L-function for characters of real representations. Invent. Math. **14**, 173–183 (1971)
8. Greenberg, R.: On p-adic L-functions and Iwasawa's theory for imaginary quadratic fields. Proceedings of conference on modern trends in Alg. No. theory related to Fermat's last theorem, pp. 215–286. Boston: Birkhäuser
9. Greenberg, R.: Functional equations, root numbers, and Iwasawa theory. In preparation
10. Gross, B., Zagier, D.: On the critical values of Hecke L-series. Soc. Math. de France, Memoire No. **2**, 49–54 (1980)
11. Hardy, G.H.: Divergent Series. Clarendon Press (1949)
12. Iwasawa, K.: On $\mathbb{Z}_l$-extensions of algebraic number fields. Ann. of Math. **98**, 246–326 (1973)
13. Katz, N.: p-adic interpolation of real analytic Eisenstein series. Ann. of Math. **104**, 459–571 (1976)

14. Kohnen, W., Zagier, D.: Values of L-series of modular forms at the center of the critical strip. Invent. Math. **64**, 175–198 (1981)
15. Manin, J., Vishik, S.: p-adic Hecke series for imaginary quadratic fields. Math. Sbornik **95** (137) (1974), No. 3
16. Martinet, J.: Character theory and Artin L-functions. Alg. No. fields. (Frohlich, A. ed.). New York: Academic Press (1977)
17. Mazur, B.: Rational points of abelian varieties with values in towers of number fields. Invent. Math. **18**, 183–266 (1972)
18. Ogg, A.: Modular forms and dirichlet series. New York: Benjamin 1969
19. Perrin-Riou, B.: Groupe de Selmer d'une courbe elliptique a multiplication complexe. Comp. Math. **43**, 387–417 (1981)
20. Rubin, K.: Elliptic curves with complex multiplication and the conjecture of Birch and Swinnerton-Dyer. Invent. Math. **64**, 455–470 (1981)
21. Rubin, K., Wiles, A.: Mordell-Weil groups of elliptic curves over cyclotomic fields. Proceedings of conference on modern trends in Alg. No. theory related to Fermat's last theorem, pp. 237–254. Boston: Birkhäuser
22. Serre, JP.: Classes de corps cyclotomique. Sem. Bourbaki **174** (1958)
23. Serre, JP., Tate, J.: Good reduction of abelian varieties. Ann. Math. **88**, 492–517 (1968)
24. Shimura, G.: On elliptic curves with complex multiplication as factors of the Jacobian of modular function fields. Nagoya Math. J. **43**, 199–208 (1971)
25. Tate, J.: Fourier analysis in number fields and Hecke's zeta functions. Alg. No. theory. (Cassels, Frohlich, eds.). London: Academic Press 305–347 (1967)
25'. Tate, J.: Arithmetic of elliptic curves. Invent. Math. **23**, 179–206 (1974)
25''. Tate, J.: Local constants. Alg. No. fields, (Frohlich, A., ed.). New York: Academic Press 1977
26. Weil, A.: On a certain type of character of the idele-class groups of an algebraic number field. Proc. Int'l. Symp. Tokyo-Nikko, 1–7 (1955)
27. Weil, A.: Dirichlet series and automorphic forms. Lecture notes in math. Vol. 189. Berlin-Heidelberg-New York: Springer 1971
28. Yager, R.: On two-variable p-adic L-functions. Ann. of Math. **115**, 411–449 (1982)
29. Yager, R.: p-adic measures on Galois groups. Preprint

Oblatum 13-IX-1982

Invent. math. 72, 267–284 (1983)

Inventiones mathematicae

Resolutions of Homology Manifolds, and the Topological Characterization of Manifolds

Frank Quinn

Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

The purpose of this paper is to present a proof of the following theorem:

Resolution Theorem. *Suppose X is an ENR homology manifold of dimension $\geqq 5$. Then there is a cell-like map $M \to X$ with domain a manifold (of the same dimension).*

Such a map is called a resolution, by analogy with the resolution of singularities of algebraic varieties. (Quick definitions: *ENR* = Euclidean neighborhood retract; a neighborhood of some closed embedding $X \subset \mathbb{R}^k$ retracts to X. For a finite dimensional homology manifold this is equivalent to locally 1-connected. *Homology manifold* means $H_*(X, X-p; \mathbb{Z}) \simeq H_*(\mathbb{R}^n, \mathbb{R}^n - 0; \mathbb{Z})$ for all $p \in X$. These satisfy all the homological properties of a manifold. A *cell-like map* has point inverses compact, nonempty, and contractible inside any neighborhood.)

This result has a long history and many special cases have been proved. The most advanced required X to already be mostly a manifold (Cannon et al. [5], Quinn [14] Theorem 3.5.2), or $X \times \mathbb{R}^k$ to be resolvable (Quinn [14], Theorem 3.3.2). For a history and discussion see J. Cannon [3, 4], and R.C. Lacher [12]. This has recently been extended to dimension 4, (Quinn [15]).

When combined with an approximation theorem of R.D. Edwards [1], this theorem implies:

Characterization of Manifolds. *A space is a manifold of dimension $\geqq 5$ if and only if it is an ENR homology manifold, and satisfies the disjoint disc property.*

A space X satisfies the *disjoint disc property* if any two maps $i, j\colon D^2 \to X$ can be approximated by maps with disjoint images. This property was first formulated by J. Cannon [3, 4], who conjectured the final result. Edwards' theorem asserts that if X satisfies the disjoint disc property then a resolution $M \to X$ can be approximated by a homeomorphism. A proof of this will be included in a forthcoming book by R. Davermann [8].

We remark that the disjoint disc property can fail very badly: Davermann and Walsh [9] have constructed *ENR* homology manifolds X such that any map $D^2 \to X$ which is injective on S^1, contains a nonempty open set in its image!

The methods used here are quite different from, and complementary to, those used by Edwards. In the decomposition theory one fixes a manifold and uses very geometric tools, engulfing and embedding theorems, to study the ways in which it can be decomposed. Since a manifold is required to even start, the basic structure of the theory would seem to preclude the construction of resolutions. By contrast here we fix X and consider manifolds mapping to it, in the aggregate. The tools are ε versions of methods from algebraic topology, homotopy theory, algebraic K-theory, and surgery. Finally, these methods do not seem to give a good enough hold on any one particular manifold to permit a proof of Edwards' theorem.

Section 1 contains a relative version of the theorem, some simple corollaries, and a few remarks on the proof. Sections 2–4 contain the proof of the theorem.

1. Statements and Applications

1.1. **Theorem.** *Suppose $(X, \partial X)$ is an ENR homology manifold pair, ∂X is a manifold, and $\dim X \geqq 5$. Then there is a cell-like map from a manifold $(M, \partial M) \to (X, \partial X)$ which is a homeomorphism on the boundary. Further, if M_1 and M_2 are two such resolutions, then for every $\varepsilon > 0$ there is a homeomorphism such that the diagram*

$$\begin{tikzcd}[column sep=small] M_1 \arrow[rr, phantom, "\simeq"] \arrow[dr] & & M_2 \arrow[dl] \\ & X & \end{tikzcd}$$

commutes up to ε, and on the boundary commutes exactly.

In fact this follows from the unbounded case: resolve the interior and then use the boundary collaring results of (Quinn [14]). The uniqueness is given in Quinn [14]. The theorem is extended to dimension 4 in Quinn [15].

It is simple to see, as was first observed by Davermann [7], that if X is an *ENR* homology manifold then $X \times \mathbb{R}^2$ has the disjoint disc property. Therefore it is a manifold, by the characterization theorem. More surprising is the fact that products $X \times Y$ have this property (C.D. Bass [1]).

1.2. **Corollary.** *Suppose X, Y have dimension $\geqq 2$. Then $X \times Y$ is a manifold if and only if X and Y are ENR homology manifolds.*

For all known examples, in fact $X \times \mathbb{R}$ is a manifold. The outstanding open question in the area (along with the 3 and 4 dimensional versions) is whether $X \times \mathbb{R}$ satisfies the disjoint disc property for all *ENR* homology manifolds X.

An important goal in decomposition theory for some time was the double suspension problem: characterize those manifolds whose second suspension is homeomorphic to a sphere. We can now identify the *spaces* with this property.

1.3. **Corollary.** *A space X satisfies $\Sigma^2 X \simeq S^{n+2}$ if and only if X is a closed ENR homology manifold, and $H_*(X: \mathbb{Z}) \simeq H_*(S^n; \mathbb{Z})$.*

$\Sigma^2 X$ is a manifold because it satisfies the disjoint disc property. (1.2 applies in the complement of the suspension circle. Near the circle we can deform the discs to cones intersecting the circle in discrete points, and separate these by pushing along the circle.) It is a simply connected homology sphere, so by the generalized Poincaré conjecture is homeomorphic to a sphere.

Remarks on the Proof. ε versions of a number of manifold and algebraic theorems are required. Versions of the h-corbordism and end theorems, and homotopy theory, were developed in Ends of Maps I (Quinn [14]). We will use this material heavily. The new material is given the minimum development required for this proof. The surgery theory for example, is done only in the simply connected $4k$ dimensional case. This is done in Sect. 2, along with an ε version of the theorem of Wall [17] relating chain complexes and CW complexes.

In Sect. 3 we begin a series of reductions. The surgery theorem gives obstructions to finding ε homotopy equivalences $M \to X$. Essentially if we can do this for all $\varepsilon > 0$, then we can use the end theorem of Ends of Maps I to "take the limit $\varepsilon \to 0$" and obtain a resolution. Therefore we must see that these obstructions vanish. Most of them are avoided by a naturality argument. They are essentially shown to form a cohomology class, so when the problem is restricted to a contractible subset only obstructions which occur on a point survive. I am indebted to R.D. Edwards for pointing out that this trick (developed for use elsewhere) could be used here. It considerably simplifies the proof.

The remaining obstruction, a single integer, is harder to avoid. It is locally defined, but globally constant. Therefore being non-zero would imply that no open set in X has a resolution. In Sect. 4 we transfer the problem to a torus (in the tradition of Novikov and Kirby!). There the last obstruction can be recognized as part of the ordinary surgery obstruction, and therefore seen to be zero.

2. Surgery Obstructions

In this section we show that the algebraic obstructions to surgery are ε versions of the ordinary obstructions. Only the "simply connected" $4k$ dimensional case is considered, since that is sufficient for the application. The development of Wall [18, Chap. 1, 5] adapts well to this setting, so we concentrate on the changes necessary for the ε estimates. There are also a few changes in notation (mainly the use of "normal map").

The usual surgery theorem (Wall [18, p. 37], Browder [2, p. 31] is roughly this:

(1) a degree 1 normal map $f: M^{4k} \to K$ (M a manifold, K Poincaré) has an invariant $\sigma(f)$ defined. This is an equivalence class of nonsingular even symmetric bilinear forms over $\mathbb{Z}$.

(2) If f is normally bordant to f': $M' \to K'$ (allowing both M and K to change) then $\sigma(f') = \sigma(f)$.

(3) If $\sigma(f) = 0$ and K is 1-connected, then f is normally bordant (holding K fixed) to a homotopy equivalence.

Theorem 2.1 is an ε version of this. The most awkward feature is that the ε version of the equivalence relation on forms is not an equivalence relation. ($A \underset{\varepsilon}{\sim} B \underset{\varepsilon}{\sim} C$ only implies $A \underset{2\varepsilon}{\sim} C$). Therefore instead of an equivalence class, the obstruction is a *set* of "associated forms".

New terms used in the statement (eg. ε Poincaré) will be defined below.

2.1. **Theorem.** *Suppose X is a locally compact metric ANR, $Y \subseteq X$ is compact, and $k > 1$.*

1) *Given $\varepsilon > 0$ there is $\delta > 0$ such that if f: $(M^{4k}, \partial M) \to (K, \partial K)$ is a proper degree 1 normal map and p: $K \to X$ is proper, satisfying over Y: K is $(\delta, 1)$ connected, ∂f is a δ homotopy equivalence, and $(K, \partial K)$ has a δ Poincaré structure of (total) dimension* $\leqq 100$ k then *there is a (nonempty) set of "associated" (see* 2.6*) even symmetric ε bilinear forms on geometric $\mathbb{Z}$ modules over X, which are ε nonsingular over $Y^{-\varepsilon}$.*

2) *Given $\gamma > 0$ there is $\gamma > \varepsilon > 0$ such that if there is an ε normal bordism g: $(N;\ \partial_0 N,\ \partial_1 N) \to (L;\ \partial_0 L, \partial_1 L)$ with L ε, 1-connected over Y and* $\dim(N) \leqq 100$ k *then forms ε associated to $\partial_0 g, \partial_1 g$ over $Y^{-\varepsilon}$ are γ bordant over $Y^{-\gamma}$.*

3) *Given $\alpha > 0$ there is $\varepsilon > 0$ such that if δ, f: $(M, \partial M) \to (K, \partial K)$ satisfy the conditions of* (1), *and a form associated to f is ε bordant to the trivial form over $Y^{-\varepsilon}$,* then *M is normally bordant* rel M *to f': $(M', \partial M) \to (K, \partial K)$ which is an α homotopy equivalence over $Y^{-\alpha}$.*

We recall and define the terms used, beginning with the algebra.

Geometric modules, and size conditions on homomorphisms, are defined in Ends I, p. 321. We will use *radius*, as in Ends II, rather than the *diameter* notion of Ends I, but the difference is unimportant. A bilinear form λ: $G \times G \to \mathbb{Z}$ has *radius* $< \varepsilon$ if for basis elements a, b of G, $\lambda(a, b) = 0$ if $d(a, b) \geqq \varepsilon$. Or equivalently, if the adjoint λ^a: $G \to G^*$ has radius $< \varepsilon$. λ is ε *nonsingular* over Y if the adjoint λ^a has an inverse of radius $< \varepsilon$ over Y. A *hyperbolic* form is the form $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ on a module $G \oplus G$. Finally, two forms (A_1, λ_1), (A_2, λ_2) are ε *bordant* over Y if there is a G so that $(A_1, \lambda_1) \oplus (A_2, -\lambda_2) \oplus \left(G \oplus G, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right)$ is ε isomorphic over Y to a hyperbolic form.

A degree 1 normal map is the standard item (Wall [18, p. 15]) except that degree 1 means with locally finite coefficients if M, K are not compact. The notion of an ε Poincaré space was introduced in Ends of Maps II. We repeat it here with slight modifications.

2.2. **Definition.** *Suppose $(K, \partial K)$ is a locally compact ANR pair, and p: $K \to X$ is proper. Then an ε Poincaré structure of (total) dimension $n + k$ for $(K, \partial K)$ over Y consists of the following:*

1) *A mapping cylinder neighborhood* $(U, \partial_0 U)$ *of a proper embedding* $(K, \partial K) \subset (\mathbb{R}^{n+k-1} \times [0, \infty), \mathbb{R}^{n+k-1} \times \{0\})$. *Let* $\partial_1 U$ *denote the closure of* $U - \partial_0 U$.

2) *A spherical fibration* $S^{k-1} \to S(\xi) \to K$.

3) *A map* $(U; \partial_0 U, \partial_1 U) \to (D(\xi); D(\xi | \partial K), S(\xi))$ *which is an* ε *homotopy equivalence over* Y.

Here if ξ is a spherical fibration then $S(\xi)$, $D(\xi)$ denote the total space and associated disc bundle respectively. Notice that the Poincaré duality dimension of $(K, \partial K)$ is n, different from the total dimension of the structure.

2.3. **Example.** *Suppose* $(X, \partial X)$ *is an ANR homology manifold pair of dimension* n. *Then for every* $\varepsilon > 0$ *there is an* ε *Poincaré structure for* $(X, \partial X)$, *over* X, *of total dimension* $2n+1$.

Proof. Since the dimension of X is n, there is a proper $1-LC$ embedding $(X, \partial X) \subset (\mathbb{R}^{2n} \times [0, \infty), \mathbb{R}^{2n} \times \{0\})$. Let $v\colon \partial_1 U \to X$ be the map in a mapping cylinder neighborhood of this embedding (one exists by Ends I, 3.1). Alexander duality shows that v is an approximate fibration with fiber $\simeq S^n$ (see Ends I, 3.3). The associated Hurewitz fibration ξ is therefore a S^n fibration, and the natural inclusions $(U; \partial_0 U, \partial_1 U) \subset (D\xi; D(\xi | \partial X), S(X))$ are ε homotopy equivalences for all $\varepsilon > 0$, over X.

This completes the proof of 2.3.

The next step is to extend Wall's connection [17] between chain and CW complexes to the ε situation. We will use this as a replacement for homology groups, which do not extend.

Suppose (K, L) is a relative CW complex, and $p\colon K \to X$ is such that the image of each cell has diameter $< \varepsilon$. Choose a basepoint in each cell, and let $C_n^c(K, L)$ be the geometric $\mathbb{Z}$ module generated by the images of basepoints of n-cells. These are the cellular chain groups. The boundary homomorphisms $\partial\colon C_n^c \to C_{n-1}^c$ can be defined by intersection numbers, so have radius $< 2\varepsilon$. The result we want is that changes in this chain complex (up to chain equivalence) can be realized by changes in the CW complex structure of (K, L).

2.4. **Proposition.** *Suppose* $X \supset Y$ *as usual* (X *locally compact metric ANR*, Y *compact*) *and* n, ε *are given. Then there exists* $\delta > 0$ *so that given the data*

1) *a* δ CW *pair* $(K, L) \to X$ *with* K, L *both* $(\delta, 1)$ *connected over* Y *and* $\dim(K - L) \leqq n$,

2) *a geometric* $\mathbb{Z}$ *chain complex* A_* *over* X *of radius* $< \delta$ *and with* $A_j = 0$ *for* $j = 0, 1$ *or* $j > n$,

3) *a chain map* $f\colon A_* \to C_*^c(K, L)$ *which is a* δ *chain equivalence over* Y,

then *there is an* ε CW *pair* (K', L), *a map* $g\colon (K', L) \to (K, L)$ *and a basis preserving* ε *isomorphism* $\theta\colon A_* \to C_*^c(K', L)$ *such that* $f = g_* \theta$ *over* $Y^{-\varepsilon}$.

Proof. (See the proof of Theorem 2 in Wall [17], Part II.) We show by induction that given δ_k there is δ small enough so there is $(K_k', L) \to (K, L)$ so that (K, K_k') is (δ_k, k) connected, and

$$C_*^c(K_k', L) = \begin{cases} A_*, & * \leqq k \\ 0, & * > k \end{cases}.$$

Note that when $k=n$ the theorem is complete, and that we can start with $k=1$ by the 1-connected hypotheses.

Assume that (K_k', L) satisfy the hypotheses above. Assume (by taking a mapping cylinder) that K_k' is a subcomplex of K. There is a natural δ_k chain equivalence $(A_*, *>k) \to C_*^c(K, K_k')$. By using the inverse we can represent the image of generators of A_{k+1} by small algebraic sums of $k+1$ cells. Adding copies gives maps $(D^{k+1}, S^k) \to (K, K_k' \cup K_k)$. But (K_k, K_k') is (δ_k, k) connected so these deform to maps to (K, K_k'). Use the boundaries to attach cells to K_k'; the result is K_{k+1}'. The maps $D^{k+1} \to K$ give a map $K_{k+1}' \to K$ which satisfies the chain complex hypotheses above.

To complete the induction step we must show that (K, K_{k+1}') is $(\delta_{k+1}, k+1)$ connected. The relative chains are equivalent to $(A_*, *>k+1)$, so it is homologically $(\delta_n, k+1)$ connected (Ends I, p. 302). Therefore by the eventual Hurewitz Theorem (Ends I, p. 302) with all $(A_i, B_i) = (K, K_{k+1}')$ there is δ_k small enough so that (K, K_{k+1}') is $(\delta_{n+1}, k+1)$ connected.

This completes the proof of 2.4. Notice that a corollary (which has a much shorter proof) is that if (K, L) is (δ, k) connected, then it is ε equivalent to (K', L) with no relative cells in dimensions $\leqq k$.

We begin the proof of 2.1 with an ε analog of Wall [2, Chap. 1]; surgery "below the middle dimension".

2.5. **Lemma.** *Suppose $X \supset Y$ as usual (locally compact metric ANR, Y compact), $k>0$, and $\varepsilon>0$. Then there is $\delta>0$ so that if $(K, \partial K) \to X$ is a δ CW pair of dimension $\leqq 100\,k$, K and ∂K are $(\delta, 1)$ connected over Y and $f\colon (M, \partial M) \to (K, \partial K)$ is a normal map, then*

1) *if* $\dim M = 2k$, *and ∂f is $(\delta, k-1)$ connected over Y then f is normally bordant* rel ∂M *to f' which is (ε, k) connected over $Y^{-\varepsilon}$.*

2) *if* $\dim M = 2k+1$, *and ∂f is (δ, k) connected over Y, then f is normally bordant to $f'\colon (M', \partial M') \to (K, \partial K)$ with f' (ε, k) connected over $Y^{-\varepsilon}$, $(K, M' \cup (\partial K))$ $(\varepsilon, k+1)$ connected over $Y^{-\varepsilon}$, and $\partial M'$ differs from ∂M by small trivial surgeries on $k-1$ spheres.*

Proof. This proceeds by induction, showing that if $j<k$ and $\delta_{j+1}>0$ then there is $\delta_j>0$ such that a map which is (δ_j, j) connected over $Y^{-\delta_j}$ can be made $(\delta_{j+1}, j+1)$ connected over $Y^{-\delta_{j+1}}$ (by surgery). We indicate modifications necessary in Wall's treatment.

First the maps used for surgery, $(D^{j+1}, S^j) \to (K, M)$ must be small. For this use 2.4 to represent (K, M) by a CW complex with no cells of dimension $\leqq j$, and use the $j+1$ cells of this complex. To represent $S^j \to M$ by a small immersion, use the ordinary immersion theorem in the inverse image of some small open set in X containing the image of D^{j+1}. Now general position gives small embeddings, and small surgeries can be performed. By comparing with the complex obtained above with no cells of dimension $\leqq j$, we can verify the connectivity conclusion.

This ends the discussion of 2.5. We begin the proof of 2.1 with the construction of "associated forms".

Suppose $f\colon (M, \partial M) \to (K, \partial K)$ is a map as in 2.1.1 (in particular $\dim M = 4k$). Then by 2.5 there is a normal bordism rel ∂M to a $(\delta', 2k)$ connected map $f'\colon M' \to K$.

At this point in the standard case, Poincaré duality is used to show that $H_*(K,M')=0$ if $*\neq 2k+1$. We obtain a similar conclusion on the chain level. First, there is a δ' chain equivalence $C^c_*(K,M')\to A_*$, where $A_*=0$ for $*\leqq 2k$. Next, the usual duality argument applied locally shows that if $Z\subset Y$ then for $j>2k+1$ $H_j(K(Z),M'(Z))\to H_j(K(Z^{2\delta}),M'(Z^{2\delta}))$ is zero. Here $K(Z)$ means the inverse image of Z in K. This implies a similar fact about A_*, which can be used to construct a small chain contraction of A_*, $*>2k+1$. The standard folding process uses the contraction to give a chain equivalence of A_* to a complex of the form $\partial\colon B_{2k+1}\to B_{2k}$, such that there is a right inverse $j\partial=1_{2k}$.

Next do surgery on M' on small $2k$ spheres corresponding to a bases of B_{2k}. This gives a map $f''\colon M''\to K$ with chains equivalent to $(\partial,0,0)\colon B_{2k+1}\oplus B_{2k}\oplus B_{2k}\to B_{2k}$. This complex is equivalent to $B_{2k+1}\oplus B_{2k}$ concentrated in dimension $2k+1$. Now we can apply 2.4 to find an δ'' homotopy equivalent CW pair $(K',M'')\simeq(K,M'')$ which has cells only in dimension $2k+1$. Denote the chain group by A $(=B_{2k+1}\oplus B_{2k}$ above).

There is an even symmetric $2\delta''$ bilinear form defined on A: if a, b are basis elements, they correspond to cells in K' attached to M'' by maps $S^{2k}\to M^{4k}$, and $\lambda(a,b)$ is the intersection number of these spheres in M. The standard arguments show that it is even and symmetric. It has radius $<2\delta''$ because the cells of K' have diameter $<\delta''$. Finally the standard Poincaré duality argument applied over subset $Z\subset Y$ as above, shows that it is $4\delta''$ nonsingular.

2.6. **Definition.** *An ε form (A,λ) over $Y^{-\varepsilon}$ is* associated *to the map f if it arises by the construction above: there is an ε normal bordism* rel ∂M *to $f''\colon M''\to K$, an ε equivalence over Y $(K',M'')\to(K,M'')$ such that (K',M'') has cells only in dimension $2k+1$, $C^c_{2k+1}(K',M'')=A$, and λ is given by intersection numbers in M''.*

According to the discussion above, such associated ε forms exist if the initial δ is small enough. Therefore 2.1.1 is complete.

Now suppose (as in 2.1.2) that f and g are normally bordant, and we are given associated ε forms. As part of the data for the forms we have bordism to highly connected maps f'', g''. Glue these bordisms together to obtain a normal map $F\colon(W,\partial W)\to(J,\partial J)$ with $\dim W=4k+1$. ∂F is the union of f'', $-g''$, and the bordism of ∂M to ∂N which is a δ homotopy equivalence. It is therefore $(\varepsilon,2k)$ connected. We can apply surgery below the middle dimension (2.5) to obtain $F'\colon(W',\partial W')\to(J,\partial J)$ which is $(\varepsilon',2k)$ connected, relatively $(\varepsilon',2k+1)$ connected, and whose boundary differs from ∂F by trivial surgeries in dimension $2k$.

We can use the previous data to get an ε' equivalence $((\partial J)',\partial W')\to(\partial J,\partial W')$ so that $\partial J'$ is $\partial W'$ union small $2k+1$ cells. These cells come from the form data for f'', g'', and the new surgeries, so the form induced on the chain group is (form of f'')+($-$form of g'')+(hyperbolic). The goal therefore is to show that the form for $\partial F'$ is isomorphic to a hyperbolic form.

At this point in Wall [18, p. 52] one considers the exact sequence

$$0\to H_{2k+2}(J,W'\cup\partial J)\to H_{2k+1}(\partial J,\partial W')\to H_{2k+1}(J,W')\to 0.$$

The end groups are dual, so this can be written as

$$0 \to B \xrightarrow{j} A \xrightarrow{h} B^* \to 0,$$

and $h = j^* \lambda^a$. Since B is free based (a geometric module) there is an isomorphism $B \simeq B^*$. Composing this with a splitting of h, v: $B^* \to A$, and adding to j gives an isomorphism $B \oplus B \to A$. The form λ composed with this isomorphism has matrix $\begin{bmatrix} 0 & I \\ I & H \end{bmatrix}$. Since H is even symmetric it can be written as $H_1 + H_1^t$ (t = transpose). Changing the splitting v to $(-H_1, 1)$ changes the isomorphism $B + B \to A$ to an isometry from $\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$ to λ. This is the required isomorphism with a hyperbolic form.

We must reformulate this argument on the chain level, with control.

First we can use local homology and duality to argue, as in the construction of associated forms, that $C_*^c(J, W')$ and $C_*^c(J, W' \cup \partial J)$ have chain contractions except in dimensions $2k+1$, $2k+2$ respectively. Boundary connected sums with small copies of $S^{2k} \times D^{2k+1}$ can be used to stabilize both complexes simultaneously, after which they are equivalent to complexes concentrated in a single dimension. Denote these by $C_*^c(J, W' \cup \partial J) \simeq B_{2k+2}$, $C_*^c(J, W') \simeq D_{2k+1}$.

Next note that because Poincaré duality can be defined using intersection numbers, the usual global duality homomorphism gives an ε'' isomorphism over $Y^{-\varepsilon''}$; $B_{2k+2} \simeq D_{2k+1}^*$. This satisfies the relation denoted by $h = j^* \lambda^a$ above.

For use in the last step we define ε exact sequences. A pair of homomorphisms of geometric modules over X, $A \xrightarrow{i} B \xrightarrow{j} C$ is *ε exact over Y* if they have radius $< \varepsilon$, $ji = 0$, and if for every $K \subset Y$, $(\ker j) \cap (B|K) \subseteq i(A|K^\varepsilon)$. Since the cells of J are small, the short exact sequence of chain complexes of the triad $(J, W' \cup \partial J, W')$ is actually ε short exact. The chain equivalence over Y of these chain complexes constructed above gives an ε'' short exact sequence over $Y^{-\varepsilon''}$

$$0 \to B_{2k+2} \xrightarrow{j} A_{2k+1} \xrightarrow{h} D_{2k+1} \to 0.$$

Now the argument outlined above for the $X =$ (point) case (considered by Wall) applies with ε estimates. This gives an ε'' isometry of the form on A_{2k+1} with a hyperbolic one. As observed above, this proves 2.1(2).

We now consider part 2.1(3), and suppose an associated form is ε bordant to 0. Part of this data is a normal bordism of f to an $(\varepsilon, 2k)$ connected map. Connected sums of M with small copies of $S^{2k} \times S^{2k}$ changes the form by addition of a hyperbolic form. By replacing M by this sum we may assume that the form of f is itself (rather than stably) isomorphic to a hyperbolic form. This is $(A, \lambda) \simeq \left(G \oplus G; \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$ in the notation above.

There is also a CW complex equivalent to (K, M) with only $(2k+1)$-cells. Use these to represent the image of the basis for $G + \{0\}$ in A by maps

$(D^{2k+1}, S^{2k}) \to (K, M)$. All intersections and selfintersection numbers of the boundaries $S^{2k} \to M$ are zero, and $\dim M \geqq 5$, so the Whitney trick given in Wall [18, Chap. 3] approximates these by disjoint framed embeddings. Since $K \to X$ is $(\delta, 1)$ connected, the 2-discs for the Whitney trick can be chosen of diameter $< \varepsilon$. Therefore the embeddings are within ε of the original maps. As in Wall [18, p. 51] use these embeddings and maps to do surgery on f. Then local homology calculations like those of Wall show that the result is an α homotopy equivalence, provided δ is chosen small enough to begin with.

This completes the proof of Theorem 2.1.

The section closes with a realization theorem for the obstructions. Theorem 2.1 as a special case of a more general situation in which the 1-connectedness of $K \to X$ is relaxed; and modules over group rings $\mathbb{Z}\pi$ are encountered. Proposition 2.7 is more specialized to the 1-connected situation, in the conclusion that it can be done with manifolds without boundary.

2.7. **Proposition.** *Fix $k > 1$, and consider locally compact metric ANRs of dimensions $\leqq k$. Then from a form λ on such an ANR X we construct a proper degree 1 normal map $g_\lambda: P_\lambda \to Q_\lambda \to X$ of dimension $4k$, 1-connected over X, with $\partial Q_\lambda = \emptyset$ and such that λ is an associated form of g_λ. The construction depends only on the bordism class of λ, is natural with respect to restriction to open sets, and natural up to normal bordism with respect to proper maps $X_1 \to X_2$. More precisely:*

1) *If $Y \subset X$ is compact and $\varepsilon > 0$, then there is $\delta > 0$ such that given a δ symmetric even nonsingular form (G, λ) on Y there is a canonical proper degree 1 normal map $g_\lambda: P_\lambda \to Q_\lambda \to Y^{-\varepsilon}$ with Q_λ $(\varepsilon, 1)$-connected over $Y^{-2\varepsilon}$, $\partial Q_\lambda = \emptyset$, and (G, λ) is associated to g_λ over $Y^{-2\varepsilon}$.*

2) *Given $\gamma > 0$ there is $\varepsilon > 0$ so that if (G, λ), (H, τ) are δ forms as above, and are ε bordant over $Y^{-\varepsilon}$, then over $Y^{-\gamma}$ there are γ homeomorphisms $P_\lambda \simeq P_\tau$, $Q_\lambda \simeq Q_\tau$ such that the diagram γ homotopy commutes.*

3) *Given $Y \supset Z$ and $\gamma > 0$, there is $\varepsilon > 0$ so that if δ is as in* (1) *for both Y and Z then the realization of the restriction $g_{\lambda|Z}$ is γ homeomorphic (as in* (2)*) to the restriction $(g_\lambda)|_Z$, over $Z^{-\gamma}$.*

4) *If $f: X_1 \to X_2$ is proper, and $f^{-1}(Y_2) \subset Y_1$, then given $\gamma > 0$ there is $\delta > 0$ so that if (G, λ) is a δ form on Y_1 then there is a $(\gamma, 1)$-connected normal bordism with $\partial = \emptyset$ between the realization of the image $g_{(f_* \lambda)}$ and the image of the realization $f_*(g_\lambda)$, over $Y_2^{-\gamma}$.*

Proof. Let $k = \dim X$, and let $X \subset W$ be a mapping cylinder neighborhood of 1 $-LC$ embedding of X in R^{4k-1}. Denote the projection by $p: W \to X$, then this is $(\delta, 1)$ connected for all δ. Since $X \subset W$ we can consider (G, λ) as δ form over W, nonsingular over $p^{-1}(Y)$. We apply the realization procedure of Wall [17, p. 53]: represent the basis of G by small embedded $2k$ discs in W. Then construct regular homotopes of the boundary spheres so that the tracks of these homotopies in $W \times I$ have intersection numbers given by λ. Finally add $2k$ handles to the resulting embeddings of S^{2k-1} in the top of $W \times I$. This gives a degree 1 normal map $M \to W \times I$.

By construction the normal map is a homeomorphism of the parts of ∂M lying over $(\partial M) \times I \cup M \times \{0\}$. We claim that it can be approximated to be a homeomorphism on $\partial M \times \{1\}$ also. More precisely, if $\alpha > 0$ there is $\delta > 0$ so that the result of the construction can be α approximated by a homeomorphism over $(\partial M \times \{1\}) \cap p^{-1}(Y^{-\alpha})$. To see this, first note that the homology calculations of Wall used locally show that since λ is nonsingular over $p^{-1}(Y)$, the boundary of M over $W \times \{1\}$ is homologically δ equivalent to $W \times \{1\}$ (δ measured in W). The eventual Hurewitz theorem (Ends I, 5.2) shows that if δ is small enough, it will be a γ homotopy equivalence over $p^{-1}(Y^{-\alpha})$, for given $\gamma > 0$. Finally Chapman and Ferry [6] have shown that if W, α are given, then there is $\gamma > 0$ so that a γ homotopy equivalence to W (measured in W) is α homotopic to a homeomorphism.

Since $f\colon M \to W \times I$ is a homeomorphism on the boundary (over $Y^{-\alpha}$), we can extend it by the identity map to obtain $g\colon M \cup_\partial W \times I \to (W \times I) \cup_\partial (W \times I)$. (The target space is the double of $W \times I$.) This is a degree 1 normal map without boundary (over $Y^{-\alpha}$ at least). Therefore g will satisfy the conclusion of the theorem if we show that (G, λ) is an ε associated form.

In fact g satisfies the conditions required to define the form: $M \cup_\partial W \times I$ is equivalent to the double $(W \times I) \cup_\partial (W \times I)$ wedge small $2k$ spheres corresponding to the handles added. Attaching $2k+1$ discs to these gives the equivalent CW complex with only $2k+1$ cells. The cellular chain group is exactly G. The boundaries in $M \cup_\partial W \times I$ are composed of the handles added to $W \times I$, union the track of the regular homotopies in $W \times I$, union the discs in $W \times \{0\}$. Therefore by construction of the homotopies the intersections are given by λ. This shows that (G, λ) is associated to g, and completes 2.7(1).

For statement (2), suppose forms λ, τ are bordant. This means that after stabilization by hyperbolic forms they are isomorphic. According to Ends I, 9.4, the isomorphism after further stabilization is equivalent to a deformation. Stabilization corresponds to adding cancelling pairs of handles, so does not change the underlying manifold. Similarly deformations can be realized by handle moves which do not change the manifold. We are therefore comparing two handlebodies formed by adding $(2k)$-handles to $W \times I$, on spheres which extend to discs in $W \times I$. There is a bijective correspondence between the two sets of discs, so that corresponding discs have the same intersection numbers.

We construct an h-cobordism between the handlebodies. Consider one of the sets of discs as lying in $(W \times I, W \times \{0\})$, the other in $(W \times I, W \times \{1\})$. Changing ends reverses orientation, so the collections now have opposite intersections. Taking connected sums of corresponding discs on each side gives framed embeddings $S^{2k-1} \times I \subset W \times I$ with zero intersections and selfintersections. Use the Whitney trick to remove these intersections, and obtain embeddings. Use these embeddings to attach (handles) $\times I$; $D^{2k} \times D^{2k} \times I$, $S^{2k-1} \times D^{2k} \times I$ to these embeddings in the top of $(W \times I) \times I$. The result is an (ε) h-cobordism, with the original handlebodies on the ends. The thin h-cobordism theorem therefore implies that they are homeomorphic.

The remaining choice was of a homeomorphism from the upper boundary to the lower. But since the homeomorphism group is locally contractible (Edwards and Kirby [11]), if ε is small enough these will be isotopic.

The statement (3) should be clear. For (4) one observes that the map f gives a codimension 0 embedding of the regular neighborhood of X_1 in the neighborhood of X_2. The bordism between W_1 and W_2 is given by $W_1 \times [0,1] \cup_{W_1 \times \{1\}} W_2 \times [1,2]$, and the handlebody and homeomorphism on W_1 are extended to W_2 by the identity.

This completes the proof of 2.7.

3. Reduction to the Single Obstruction

In this section the resolution problem is reduced to a single integer obstruction.

3.1. **First Reduction.** *To show that X of dimension $\geqq 5$ has a resolution, it is sufficient to show that for some k each point in $X \times \mathbb{R}^k$ has a neighborhood which has a resolution.*

Proof. Since $X \times \mathbb{R}^k$ $(k \geqq 2)$ satisfies the 2-disc condition (Daverman [7]) local resolvability implies that it is a manifold (Edwards [10]). But according to Ends I, 3.23, this implies that X itself is resolvable if $\dim X \geqq 5$.

3.2. **Definition of the Obstruction.** Suppose X is an ANR homology manifold. By crossing with some R^j we may assume that the dimension of X is $4k$, some k. Choose a point x_0 in X. The obstruction will be to the construction of a resolution of a neighborhood of x_0.

Suppose $X \subseteq \mathbb{R}^{n-4k}$ is a proper $1-LC$ embedding. Then there is a mapping cylinder neighborhood W, and the projection $v\colon W \to X$ is an approximate $n-1$ sphere fibration. Let X_1 be a neighborhood of x_0 which is contractible in X. Then the restriction of this approximate fibration to X_1 has a fiber homotopy trivialization. Considered as a smooth structure on the Spivak normal fibration, this defines in the traditional way a proper degree 1 normal map from a smooth manifold to X_1. Explicitly, let $\partial W_1 = v^{-1}(X_1)$, and W_1 the mapping cylinder of this. Then the fiber homotopy trivialization defines a map $(W_1, \partial W_1) \to (D^n, S^{n-1})$. Approximate this rel boundary to be transverse to $0 \in D^n$. The inverse image of 0 is then a smooth manifold $M \subset W$, and the composition $M \subset W \to X_1$ is a proper degree 1 normal map.

Now we apply the surgery Theorem 2.1. Choose $X_2 \subset X_1$ a compact neighborhood of x_0. Note $\partial M = \emptyset$, X_1 is $(\delta, 1)$ connected, and is δ Poincare over X_1, for every $\delta > 0$ (Example 2.3). Therefore for every $\alpha > 0$ there is $\varepsilon > 0$ and ε associated bilinear forms which are obstructions to obtaining an α homotopy equivalence over $X_2^{-\alpha}$. Let (G, λ) be one of these forms. Then by the realization Theorem 2.7 there is a proper degree 1 normal map $P \to Q \to X_2$ with $\dim P = 16k$, $\partial P = \partial Q = \emptyset$, and such that $P \to Q$ has (G, λ) as an associated form over $X_2^{-\varepsilon}$.

The fiber of $(W_2, \partial W_2) \to X_1$ is homotopy equivalent to $(D^n, S^{n-1}) \to (W_2, \partial W_2)$. The composition $Q \to X_2 \subset W_2$ is disjoint from ∂W_2, and therefore the image of S^{n-1}. Since these are all smooth manifolds we can closely approximate the maps to be transverse regular. Making P transverse gives a pullback diagram

$$\begin{array}{ccccc} P' & \longrightarrow & Q' & \longrightarrow & (D^n, S^{n-1}) \\ \downarrow & & \downarrow & & \downarrow \\ P & \longrightarrow & Q & \longrightarrow & (W_2, W_2). \end{array}$$

$P' \to Q'$ is a degree 1 normal map of closed smooth manifolds of dimension $12k$. Since this is a multiple of 4, the surgery obstruction is an integer ($=1/8$ (index P' - index Q')). This integer is the obstruction we associate to $x_0 \in X$.

This completes the Definition 3.2. Note that the point of going through forms and the realization theorem is to obtain a map with manifold range, so that transversality can be applied.

3.3. **Second Reduction.** *Suppose the obstruction defined in* 3.2 *is zero. Then* $x_0 \times \{0\}$ *has a neighborhood in* $X \times \mathbb{R}^3$ *which has a manifold resolution.*

Proof. The first step is to show that there is a neighborhood $X_4 \subset X_2$ such that for every $\varepsilon > 0$ the proper degree 1 normal map $M \to X_1$ constructed in 3.2 is normally bordant to $f_\varepsilon: M_\varepsilon \to X_1$ which is an ε homotopy equivalence over $X_4^{-\varepsilon}$.

Choose X_3 to be a neighborhood of x_0 in X_2 which is contractible in X_2. Let X_4 be a neighborhood whose closure is in X_3. As above let W_i be the part of W lying over X_i. The inclusion $(W_3, \partial W_3) \subset (W_2, \partial W_2)$ is homotopy equivalent to $X_3, \subset X_2$ crossed with (D^n, S^{n-1}), so the contraction of X_3 gives a factorization up to homotopy

$$\begin{array}{ccc} & (D^n, S^{n-1}) & \\ \nearrow & & \searrow \\ (W_3, \partial W_3) & \longrightarrow & (W_2, \partial W_2). \end{array}$$

We will use this homotopy to construct a normal bordism of $P \to Q$.

Choose a smooth triangulation of W_2 so that the simplices have images in X of diameter $< \varepsilon$, and so that there is a compact PL submanifold $W_3 \supset K \supset W_4$. Triangulate D^n so that $(D^n, S^{n-1}) \to (W_4, \partial W_4)$ is simplicial.

Next make the map $Q \to W_2$ transverse to this triangulation (transverse to each simplex), and make $P \to Q$ transverse to the resulting partition of Q. This breaks $P \to Q$ up into many small degree 1 normal maps. The inverse image of a simplex of W_2 in Q is a manifold, with boundary the inverse image of the boundary of the simplex.

The next step fits nicely into the geometric description of surgery deveioped in the author's thesis (see Quinn [13] or Wall [18, Chap. 17A]).

There is a simplicial complex NM (Δ-set actually; see Rourke and Sanderson [16]) whose simplices are degree 1 normal maps of manifolds, with boundary split up into pieces like the boundary of a simplex. The face operation ∂_j in the Δ-set corresponds to taking the part of the boundary of the normal map corresponding to $\partial_j \Delta$. The Δ-set NM was first described by C. Rourke in unpublished notes on Sullivan's work on surgery.

The association of a simplex in W_2 to the piece of the normal map $P \to Q \to W_2$ lying over the simplex, defines a map $W_2 \to NM$. Since ∂Q is empty, and the image of Q is disjoint from ∂W_2, ∂W_2 maps to the empty normal map $\emptyset \in NM$.

Next restrict this map to the neighborhood W_3. The homotopy factorization constructed above gives a homotopy rel ∂W_3 to a map which factors

$$(W_3, \partial W_3) \to (D^n, S^{n-1}) \to |NM, \emptyset|.$$

Using the simplicial approximation theorem and the Kan condition we can get simplicial maps

$$(W_3, \partial W_3) \to (\Delta^n, \partial \Delta^n) \to (NM, \emptyset).$$

The simplicial map $(\Delta^n, \partial \Delta^n) \to (NM, \emptyset)$ corresponds to a degree one normal map of closed manifolds, specifically the transverse pullback $P' \to Q'$ constructed in 3.2.

Now we assemble these maps. A simplicial map $K \to NM$ assigns to each simplex of K a normal map, with boundary divided into pieces corresponding to the faces of the simplex. If two simplices have a common face, then the corresponding parts of the normal maps are equal. Therefore we can fit them together. Take the disjoint union of the normal maps corresponding to nondegenerate simplices of K, and identify pieces of the boundaries corresponding to common faces. If K is a PL manifold then the result is a normal map of PL manifolds. (If K is a PL homology manifold the assembly is a topological normal map; see Ends I, 3.3.1 Part 2.)

The original map $W_2 \to NM$ was defined by splitting $P \to Q$ into pieces over simplices of W_2. Therefore the map assembles to give back $P \to Q$. The homotopy of $W_3 \to NM$ assembles to give a normal bordism of the restriction of $P \to Q$ to W_3. The result of the homotopy factors simplicially through Δ^n, so can be described as follows: a simplical map $(W_3, \partial W_3) \to (\Delta^n, \partial \Delta^n)$ is automatically transverse to the barycenter of Δ^n. Let $N \subset W_3$ be the inverse image manifold. Then the factored map assembles to the product $N \times (P' \to Q')$. We have therefore constructed a normal bordism from $P \to Q$ (over W_3) to $N \times (P' \to Q')$.

The next step is to apply the hypothesis of 3.3 that the surgery obstruction of $P' \to Q'$ is trivial. This means that $P' \to Q'$ is normally cobordant to a homotopy equivalence $P'' \to Q''$. Crossing with N and adding to the previous normal bordism gives a normal bordism of $P \to Q$ (over W_3) to a homotopy equivalence. Finally since it maps to X_3 by projection $N \times (P'' \to Q'') \to N \subset W_3 \to X_3$, it is an ε homotopy equivalence.

Actually because of simplicial technicalities the map is not quite the topological projection. Since the simplices were arranged to have diameter $< \varepsilon$ in X, it differs from the projection by only ε, so is still an ε homotopy equivalence.

Now we apply the uniqueness part of the surgery Theorem, 2.1(2). Since $P \to Q$ over X_3 is normally bordant to an ε homotopy equivalence, the form (G, λ) is δ bordant to the trivial form over $X_3^{-\gamma}$. The bordism constructed may not satisfy the $(\varepsilon, 1)$ connected condition of 2.1(2), but this can easily be arranged; before the maps to NM are assembled, do surgery on the image of each simplex to make it a normal map of 1-connected manifolds. On the space level this is a deformation into the subcomplex NM' of 1-connected normal maps. The dimensions of the manifolds involved are large enough (>4) for this surgery to be done.

Finally since the form (G, λ) is bordant to the trivial form, Theorem 2.1(3) implies (if δ is small enough) that the original degree 1 map $M_3 \to X_3$ is normally bordant to an α homotopy equivalence over $X_3^{-\alpha}$, for $\alpha > 0$.

This completes the first step in the proof of 3.3: normal bordisms of f to $f_\varepsilon \colon M_\varepsilon \to X_1$ which are ε homotopy equivalences over X_4. The next step is to construct h-cobordisms between these, when stabilized: we will show that there is $X_6 \subset X_4$ a neighborhood of x_0 such that if $\varepsilon > 0$ so that if $\alpha, \rho < \delta$ and f_α, f_β are the maps constricted above, then there is a proper normal bordism $F_\varepsilon \colon N_\varepsilon \to X_1 \times \mathbb{R}^3$ so that $\partial_0 F = f_\alpha \times 1_{R^3}$, $\partial_1 F = f_\beta \times 1_{R^3}$, and over $X_6 \times B^3$, F is an (ε, h) cobordism from $M_\alpha \times \mathbb{R}^3$ to $M_\beta \times \mathbb{R}^3$.

As above let $X_5 \subset X_4$ be a neighborhood which contracts in X_4, and X_6 a neighborhood whose closure is contained in X_5. A normal bordism from f_α to f_β can be considered as a normal map to $X \times I$. Crossing with $\mathbb{R}^3$ gives a normal map of dimension $4(k+1)$, $F \colon N \to X \times I \times \mathbb{R}^3$, whose boundary is a $\max(\alpha, \beta)$ equivalence over $X_4 \times \mathbb{R}^3$.

Applying the surgery theorem to this gives an associated form (G, λ) over $X_4 \times 2B^3$. As above we apply the realization theorem, and use the space NM to analyze the restriction to $X_5 \times 2B^3$. It is equivalent to a product $N \times (P' \to Q')$ as above. Now however, since $X \times \mathbb{R}^3$ has odd dimension, $P' \to Q'$ has odd dimension. Since odd dimensional simply connected normal maps are bordant to homotopy equivalences ($L_{2i+1}(\mathbb{Z}[1]) = 0$), the argument proceeds without obstruction to give an ε equivalence. This completes the second step.

The final step is an application of the end theorem of Ends I. Choose $\delta_i > 0$ so that $f_{\delta_i} \times 1_{\mathbb{R}^3}$ is $(1/2^i, h)$ cobordant to $f_{\delta_{i+1}} \times 1_{\mathbb{R}^3}$ over $X_6 \times B^3$, by step 2. Taking the union of these h-cobordisms gives a manifold $S \to X_1 \times \mathbb{R}^3$ with an end, which is tame and 1-connected over $X_6 \times B^3$. By Ends I, Theorem 1.4 there is a completion of S over $X_6 \times B^3$. The levels M_{δ_i} are approximate completions of this end, and are δ_i homotopy equivalent to $X_6 \times B^3$. It follows that the new boundary of the completion is a δ homotopy equivalence for all $\delta > 0$, hence a resolution of $X_6 \times B^3$.

This complete the proof of 3.3, and reduces the main theorem to the single obstruction.

4. The Last Obstruction

We show that the obstruction defined in 3.2 vanishes. The obstruction is defined by transversality on a manifold degree 1 normal map with the same form as $f \colon M \to X$. The analogous transversality construction on f itself would be: take a manifold point $p \in X$, make f transverse, and take the surgery obstruction of $f^{-1}(p) \to p$. This is a degree 1 map of a discrete set of points, so has obstruction zero. Therefore the obstruction will vanish if we can show that it depends only on the form, not the particular degree 1 normal map. This is done by transferring the problem to a torus, where it can be recognized as part of the ordinary surgery obstruction $L(\mathbb{Z}[\mathbb{Z}^n])$ (forget ε). The hard step will be extending an ε Poincaré space across the puncture in a punctured torus.

4.1. **Reduction to Poincaré Problems.** As in 3.2 we let X_1 be a neighborhood of x_0 whose mapping cylinder neighborhood in $\mathbb{R}^{4k+n}$ is fiber homotopically trivial. Fix $j: X_1^* \to (\mathbb{R}^{4k})^*$, a degree 1 map of 1-point compactifications, such that $g(\infty)=\infty$ and $g(x_0)=0$. Obstruction theory shows that there is one. Let B be an open ball about 0 in $\mathbb{R}^{4k}$, and identify it with a ball in the torus T^{4k}.

First Reduction. *Suppose that for every $\delta>0$ there is $h: Y\to T^{4k}$, a $(\delta, 1)$ connected degree 1 normal map, with Y δ Poincaré over T^{4k}, and $(Z: X_B, Y_B)\to B$ which over $B^{-\delta}$ is a $(\delta, 1)$ connected δ Poincaré normal bordism. Then the obstruction of* 3.2 *is zero. Here X_B, Y_B denote $j^{-1}(B)$, $h^{-1}(B)$ respectively.*

Proof. Note that normal map, normal bordism here just mean that the (relative) mapping cylinder neighborhoods in Euclidean space are δ equivalent to the product with (D^j, δ^{j-1}) for appropriate j. Proceeding as in 3.2 we make projection to D^j transverse to 0, and do surgery below the middle dimension. Over X_B this gives the ε form used in 3.2. The map $Y\to T^n$ defines an ε form over T^n. As in the uniqueness result for forms, 2.1(2), the δ bordism from X_B to Y_B over $B^{-\varepsilon}$ gives a bordism of the forms, over $B^{-\varepsilon}$.

Now we use the naturality properties of the realization construction 2.7. The construction gives $g_0: P_0\to Q_0\to X_B$ realizing the form for X_B, and $g_1: P_1\to Q_1\to T^n$ realizing the form for Y. Applying the naturality 2.7(4) gives a normal bordism from g_0 to the realization of the image form over B. The bordism of forms over B and naturality with respect to restriction 2.7(3) identify this image as the restriction of g_1. Therefore over the mapping cylinder $h: X_B\to B$ we have an ε normal bordism $G: P\to Q\to B_h$ with $\partial_0 G=g_0$, $\partial_1 G=g_1 | B^{-\varepsilon}$.

The next step is to make G transverse to something. As in the construction of the single obstruction (3.2) we let $(W; \partial_0 W, \partial_1 W)\supset(B_h, X_B, B)$ be a mapping cylinder neighborhood of a proper relative embedding into $\mathbb{R}^{4k+n}\times[0,1]$. As part of the data of δ Poincaré duality and normal maps, we have a δ homotopy equivalence $(W; \partial_0 W, \partial_1 W, \partial_2 W)\to(B_h\times D^n;\ X_B\times D^n,\ B\times D^n,\ B_h\times S^{n-1})$, over $B^{-\varepsilon}$. Consider the inverse applied to the disc crossed with the arc from x_0 to 0. This is a map $(D^n\times I;\ D^n\times\{0\},\ D^n\times\{1\},\ S^{n-1}\times I)\to(W; \partial_0 W, \partial_1 W, \partial_2 W)$. Make the normal bordism $G: P\to Q\to B_h\subset W$ transverse to this map. This gives a normal bordism of closed manifold normal maps. On the end over X is the normal map used to define the single invariant in 3.2. The invariant is therefore equal to the surgery obstruction at the other end, over $B\subset T^{4k}$.

Over T^{4k}, the mapping cylinder neighborhood is $T^{4k}\times D^n$. Transversality to the disc is therefore the same as the transverse inverse image of a point in $g_1: P_1\to Q_1\to T^{4k}$.

The surgery obstruction group $L_{4*}(\mathbb{Z}[\mathbb{Z}^{4k}])$ has a summand $L_{4*}(\mathbb{Z}[1])\simeq\mathbb{Z}$, which for closed manifold surgery problems is detected by the surgery obstruction of the inverse image of a point in T^{4k} (Wall [18]). Therefore the integer that we want is part of the total surgery obstruction $\sigma(g_1)\in L_{4*}(\mathbb{Z}[\mathbb{Z}^{4k}])$. But the surgery obstruction depends only on the form on the middle dimension after surgery below the middle dimension. Therefore it is the same as the degree 1 normal map $M\to Y$ used to find the form for g_1. Finally, as pointed out in the introduction of the section, the appropriate part of this obstruction vanishes; $Y\to T^{4k}$ is itself degree 1, so the transverse image of a point is 0-dimensional.

Second Reduction. We must show that the δ Poincaré space $X_B\to B\subset T^{4k}$ extends over all of T^{4k}. The objective here is to reduce this to a problem over an interval.

Let p be a point in $T^{4k}-B$, and let $\alpha\colon T^{4k}-p\to\mathbb{R}^{4k}$ be a smooth immersion which is the identity on B. Then the pullback

$$\begin{tikzcd} X' \arrow[r] \arrow[d] & T^{4k}-p \arrow[d, "\alpha"] \\ j^{-1}(\mathbb{R}^{4k}) \arrow[r, "j"] & \mathbb{R}^{4k} \end{tikzcd}$$

gives a proper degree 1 normal map, and over B, $X=X'$. Let $\delta>0$. We must find a $(\delta,1)$ connected normal bordism to a $(\delta,1)$ connected Poincaré space over $T^{4k}-p$, and then extend it over p.

For the first step, for any $\gamma>0$ there is a γ homotopy equivalence $X''\to X'$ (measured in X') such that X'' has a manifold neighborhood of a γ 1-skeleton. This is the ε version of Wall [19], Corollary 2.3.2, and is proved in the same way using the ε chain complex material of 2.4.

Now take a γ 2-skeleton of $T^{4k}-p$, deform the image of the 1-skeleton of X'' into it, and form the mapping cylinder. This gives a relative 2-complex which is universal for the $(\delta,1)$-connected lifting problem. Attach 0, 1 and 2 handles to the top of $X''\times[0,1]$ corresponding to the cells of this 2-complex. This gives a $(\delta,1)$ connected normal bordism of X'' to $Y_0\to T^{4k}-p$, which is also $(\delta,1)$ connected.

Define $T^{4k}\text{-}p\to T^{4k}\cup_p[0,1)$ by: choose an open collar $S^{4k-1}\times[-1,1)$. It is the projection to $[0,1)$ on $S^{4k-1}\times[0,1)$, is radial expansion on $S^{4k-1}\times[-1,0)\to S^{4k-1}\times[-1,1)$, and is the identity outside the collar. Composing gives $Y_0\to T^{4k}\cup_p[0,1)$ which is proper, degree 1 (on T^{4k}), normal, δ Poincaré, and $(\delta,1)$ connected over $T^{4k}\cup[0,1-\delta)$. Finally we can project this to $[0,1)$ by mapping T^{4k} to 0.

4.2. **Lemma.** *Given $k>0$ and $\frac{1}{4}>\varepsilon>0$ there is $\delta>0$ such that if $Y\to[0,1)$ is proper, over $[0,1-\delta)$ is a δ Poincaré CW complex of dimension $4k$ with fiber homotopically trivial normal fibration, and is $\delta,1$ connected over $(\delta,1-\delta)$, then there is $Y_1\to[0,1]$ which is a compact δ Poincaré space with trivial normal fibration which is $\varepsilon,1$ connected over $(\varepsilon,1]$ and the restrictions $Y_1|[0,\frac{1}{2}]$ and $Y|[0,\frac{1}{2}]$ are equal (homeomorphic commuting with the map to I).*

Completion of the Reduction. Apply the lemma to the situation preceeding it. There is a map $Y_1\to T^{4k}\cup_p[0,1]$ since $Y_0=Y_1$ near 0. Project to T^{4k} by collapsing $[0,1]$ to p, then the composition $Y_1 \to T^{4k}$ satisfies all the requirements of the reduction.

Proof of 4.2. We "double" Y to get something which is not Poincaré, but has a trivial normal bundle. Smooth transversality gives a manifold M mapping to this object. We use the surgery theory of Section 2 to make $M\,\varepsilon$ equivalent to Y near $\frac{3}{4}$, and then use the chain complex material 2.4 to patch together M and Y.

The homotopy and Poincaré data give the following: Let U be a regular neighborhood of Y in $\mathbb{R}^{4k+n}$, which is a mapping cylinder of $p: \partial U \to Y$. There is $V \subset \partial U$ containing $p^{-1}(Y[0, 1-2\delta])$, and a map $(U, V) \to (Y \times D^n, Y \times S^{n-1})$ which is projection on the Y factor and δ homotopy equivalence over $[0, 1-3\delta]$.

Choose t near $1-4\delta$ so that $Y \to [0,1]$ is transverse to t (we may assume p is PL). Let Y' be the inverse of $[0,t]$, Z the inverse of t. Then the double $p^{-1}(Y') \cup_{p-1(Z)} P^{-1}(Y') \to Y' \cup_Z Y'$ gives a regular neighborhood in $\mathbb{R}^{n+4k}$. Since $p^{-1}(Y') \subset V$, the map to D^n gives a map rel boundary of this neighborhood to D^n, S^{n-1}. Make this transverse to 0 to give a manifold $M \to Y' \cup_Z Y'$, which is closed since $Y' \cup_Z Y'$ is compact.

Now map $Y' \cup_Z Y' \to Y'/Z$ by collapsing the second piece to a point. Since Z is a point inverse, we get $Y'/Z \to [0,1]$. Over $[0, 1-5\delta]$, $Y'/Z = Y$. Since M was constructed using the Poincaré data of Y, $M \to Y'/Z$ is a degree 1 normal map over $[0, 1-5\delta]$. Since Y is $\delta, 1$ connected over $[\delta, 1-5\delta]$, and of dimension $4k$, we can apply the surgery Theorem 2.1. The obstruction is a form which we can analyse as in Sect. 3: realize it as a map of manifolds, make transverse to get a map $(\varepsilon, 1-\varepsilon) \to NM$. Since the interval is contractible the only obstruction is the inverse over a single point. But the dimension of this is $4k-1$, so has surgery obstruction 0. The form is therefore bordant to 0, and we can do surgery to obtain an ε equivalence over $(\varepsilon, 1-\varepsilon)$, if δ was small enough.

We now have $M \to Y'/Z$ an ε homotopy equivalence over $(\frac{1}{2}+\varepsilon,\ 1-\varepsilon)$. Define $X \to [0,1]$ to be the union of M over $[\frac{3}{4}, 1]$ and Y over $[0, \frac{3}{4}]$. This object fails to be Poincaré at $\frac{3}{4}$. However it maps to Y over $[\frac{1}{2}+\varepsilon, 1-\varepsilon]$ by an ε equivalence except at the ends and near $\frac{3}{4}$. We will add cells to X near $\frac{3}{4}$ to make it an ε' equivalence. The result will be Poincaré, and almost the object needed for the lemma.

Using the $\delta, 1$ connectedness of Y add 0, 1 and 2 cells to X near $\frac{3}{4}$ to make it ε, 1 connected over $(\frac{1}{2}+\varepsilon, 1-\varepsilon)$. The pair (Y, X) then satisfies the hypotheses of the chain complex Lemma 2.4. Therefore to add cells near $\frac{3}{4}$ to X to obtain a complex equivalent to Y, it is sufficient to show that $C_*(Y, X)$ is equivalent to a geometric chain complex concentrated near $\frac{3}{4}$. Since $X \to Y$ is an ε equivalence on regions on each side of $\frac{3}{4}$, $(\frac{1}{2}+\varepsilon, \frac{3}{4}-\varepsilon)$ and $(\frac{3}{4}+\varepsilon, 1-\varepsilon)$ the complex $C_*(Y, X)$ has an ε chain contraction over these regions. We use these to show C_* can be made zero on slightly smaller regions.

The standard folding argument concentrates C_* in two dimensions: if C_j is the lowest nonzero module over (a, b), let $C'_i = C_i$ for $i \neq j$, $j+2$, $C'_j = C_j \,|\, (a, a+\varepsilon) \cup (b-\varepsilon, b)$, $C'_{j+2} = C_{j+2} \oplus (C_j \,|\, [a+\varepsilon, b-\varepsilon])$, with boundary homeomorphism $\partial + s$ in dimension $j+2$. Here s is the chain contraction. Notice that the next step takes place on a smaller interval, and has larger radius. The loss is determined by the number of steps necessary, namely dim $Y=4k$. Since this is fixed in advance we can allow for it.

When C_* is concentrated in two dimensions the boundary homeomorphism is an isomorphism. The isomorphism Theorem 8.4 of Ends I shows that this can be stably deforned to be the identity on a smaller interval. Finally a subcomplex on which the boundary is the identity can be deleted from the complex. This eliminates C_* except near $\frac{3}{4}$.

As explained above this shows how to modify X to be equivalent to Y over $(\varepsilon', 1-\varepsilon')$, if δ was small enough. X is globally Poincaré, and satisfies all the conclusions of Lemma 4.3 except ε, 1 connectedness near 1. It is a manifold there, so low dimensional surgery there yields this condition.

This completes the proof of the lemma, and therefore the theorem.

References

1. Bass, C.D.: Some products of topological space which are manifolds. Proc. AMS **81**, 641–646 (1981)
2. Browder, W.: Surgery on simply-connected manifolds. Berlin-Heidelberg-New York: Springer 1972
3. Cannon, J.W.: The recognition problem: what is a topological manifold? Bull. AMS **84**, 832–866 (1978)
4. Cannon, J.W.: The characterization of topological manifolds of dimension $\geqq 5$. Proc. Int. Congess of Math. 1978, Acad. Sci. Fennica, Helsinki 1980, pp. 449–454
5. Cannon, J.W., Bryant, J., Lacher, R.C.: The structure of generalized manifolds having nonmanifold set of trivial dimension. In: Geometric Topology. Cantrell, J.C., (ed.). New York: Academic Press 1979, pp. 261–300
6. Chapman, T.A., Ferry, S.: Approximating homotopy equivalences by homeomorphisms. Amer. J. Math. **101**, 583–607 (1979)
7. Davermann, R.J.: Detecting the disjoint disks property. Pacific J. Math. **93**, 277–298 (1981)
8. Davermann, R.J.: Decompositions of manifolds. New York: Academic Press, in preparation
9. Davermann, R.J., Walsh, J.J.: A ghastly generalized manifold. Ill. J. Math.
10. Edwards, R.D.: The topology of manifolds and cell-like maps, Proc. Int. Congress of Math. 1978, Acad. Sci. Fennica, Helsinki 1980, pp. 111–127
11. Edwards, R.D., Kirby, R.C.: Deformations of spaces of embeddings. Ann. Math. **93**, 63–88 (1971)
12. Lacher, R.C.: Cell-like mappings and their generalizations. Bull. AMS **83**, 495–552 (1977)
13. Quinn, F.: A geometric formulation of surgery. In: Topology of Manifolds, Cantrell, J.C., Edwards, C.H., (eds.). Markham 1970, pp. 500–511
14. Quinn, F.: Ends of Maps, I, Ann. Math. **110**, 275–331 (1979)
15. Quinn, F.: Ends of Maps III; dimensions 4 and 5, J. Diff. Geom. 17, 503–521 (1982).
16. Rourke, C.P., Sanderson, B.J.: Δ-sets I, Quart. J. Math Oxford, **2**, (22) 321–328 (1971)
17. Wall, C.T.C.: Finiteness conditions for CW complexes. I. Ann. Math. **81**, 56–69 (1965); II, Proc. Royal Soc. A **295**, 129–139 (1966)
18. Wall, C.T.C.: Surgery on compact manifolds. New York: Academic press 1970
19. Wall, C.T.C.: Poincaré complexes. I. Ann. Math. **86**, 213–245 (1967)

Oblatum 22-IX-1982

Invent. math. 72, 285–298 (1983)

Inventiones mathematicae

Differentiability of Minima of Non-Differentiable Functionals

Mariano Giaquinta and Enrico Giusti

Istituto Matematico, Università di Firenze, Víale Morgagni 67/A, I-50134 Firenze, Italia

In this paper we shall consider the problem of the regularity of the derivatives of functions minimizing a variational integral

$$F(u;\Omega)=\int_{\Omega} f(x,u,Du)\,dx \tag{1.1}$$

where Ω is an open set in $\mathbb{R}^n$, $u\colon\Omega\to\mathbb{R}^N$, $Du=\{D_\alpha u^i\}$ $\alpha=1,\ldots,n$; $i=1,\ldots,N$, and $f\colon\Omega\times\mathbb{R}^N\times\mathbb{R}^{nN}\to\mathbb{R}$ is a Carathéodory function (i.e. measurable in x and continuous in u,p) satisfying

$$\lambda|p|^2-a\leqq f(x,u,p)\leqq\Lambda|p|^2+a,\qquad \lambda>0. \tag{1.2}$$

A local minimum for the functional F is a function $u\in W^{1,2}_{\mathrm{loc}}(\Omega,\mathbb{R}^N)$ such that for every $\phi\in W^{1,2}(\Omega,\mathbb{R}^N)$ with $\operatorname{supp}\phi\subset\subset\Omega$ we have

$$F(u;\operatorname{supp}\phi)\leqq F(u+\phi;\operatorname{supp}\phi).$$

In a recent article [6], we have proved basic regularity results for the local minima of the functional (1.1). In the scalar case ($N=1$) we have shown that every local minimum of F with condition (1.2) is Hölder-continuous in Ω.

In the general case $N\geqq 1$ such a result cannot hold; we proved however that $Du\in L^{2+\sigma}_{\mathrm{loc}}$ for some $\sigma>0$. More generally these results hold for Q-minima (see [7]).

In this paper we investigate the regularity of the first derivatives of the minima of F, under additional hypotheses on the function $f(x,u,p)$. Roughly speaking, we assume that f is twice differentiable and strictly convex in p, and Hölder-continuous in (x,u). We remark that we do not assume the existence of the derivative f_u, and therefore our functionals are in general non differentiable.

As usual, our results will take different form in the scalar and in the vector case. When $N=1$, we prove that every local minimum has Hölder-continuous derivatives in Ω. When $N\geqq 1$, we obtain that for every local minimum u there exists an open set $\Omega_0\subset\Omega$, with mean $s(\Omega-\Omega_0)=0$ such that $u\in C^{1,\alpha}(\Omega_0,\mathbb{R}^N)$

for some $\alpha>0$. This result has been obtained independently by P.A. Ivert [13]. In general $\Omega-\Omega_0$ is non-empty; however in some special case, e.g. when

$$F(u;\Omega)=\int_\Omega |Du|^2\,dx+\int_\Omega g(x,u)\,dx$$

we get regularity everywhere even if $N>1$.

2. Preliminary Results and Definitions

We shall consider local minima of the functional

$$F(u;\Omega)=\int_\Omega f(x,u,Du)\,dx \tag{2.1}$$

where *the function $f(x,u,p)$ satisfies*

$$\lambda|p|^2-a\leqq f(x,u,p)\leqq\Lambda|p|^2+a,\qquad \lambda>0 \tag{2.2}$$

and moreover

i) *for every $(x,u)\in\Omega\times\mathbb{R}^N$, $f(x,u,p)$ is twice differentiable in p, and we have*

$$|f_{pp}(x,u,p)|\leqq L, \tag{2.3}$$

$$f_{p^i_\alpha p^j_\beta}(x,u,p)\,\xi^\alpha_i\,\xi^\beta_j\geqq\nu|\xi|^2;\qquad \nu>0 \tag{2.4}$$

for every $\xi=\{\xi^\alpha_i\}$ in $\mathbb{R}^{nN}$, $i=1,\dots,N$, $\alpha=1,\dots,n$.

Here and in the following summation over repeated indices is understood.

(ii) *For every $p\in\mathbb{R}^{nN}$ the function $(1+|p|^2)^{-1}f(x,u,p)$ is continuous in $\Omega\times\mathbb{R}^N$ uniformly in p. In other words there exists a bounded continuous concave increasing function (t), with $(0)=0$, such that*

$$|f(x,u,p)-f(y,v,p)|\leqq(1+|p|^2)\,\omega(|x-y|^2+|u-v|^2). \tag{2.5}$$

Let $u\in W^{1,2}_{\text{loc}}(\Omega,\mathbb{R}^N)$ be a local minimum for the functional (2.1). We shall get estimates of the derivatives of u in the spaces $L^{2,\lambda}$ and $\mathscr{L}^{2,\lambda}$. We shall limit ourselves to a short description of these function spaces, referring to [1] for a detailed discussion. (See also [10], Chap. 4.)

A function $\phi\in L^2(\Omega)$ is said to belong to $L^{2,\lambda}(\Omega)$ if

$$\|\phi\|^2_{2,\lambda}=\sup\{\rho^{-\lambda}\int_{\Omega\cap B(x_0,\rho)}|\phi|^2\,dx;\ x_0\in\Omega,\ \rho<\operatorname{diam}\Omega\}<+\infty. \tag{2.6}$$

Similarly, $\mathscr{L}^{2,\lambda}(\Omega)$ denotes the space of all $\phi\in L^2(\Omega)$ such that

$$|||\phi|||^2_{2,\lambda}=\sup\{\rho^{-\lambda}\int_{\Omega\cap B(x_0,\rho)}|\phi-\phi_{x_0,\rho}|^2\,dx;\ x_0\in\Omega,\ \rho<\operatorname{diam}\Omega\}<+\infty, \tag{2.7}$$

where $\phi_{x_0,\rho}$ is the average of ϕ on $\Omega\cap B(x_0,\rho)$:

$$\phi_{x_0,\rho}=\fint_{\Omega\cap B(x,\rho)}\phi\,dx=|\Omega\cap B(x_0,\rho)|^{-1}\int_{\Omega\cap B(x_0,\rho)}\phi\,dx.$$

We recall that the two spaces $L^{2,\lambda}$ and $\mathscr{L}^{2,\lambda}$ coincide if $0<\lambda<n$. Moreover if $\lambda>n$ and if Ω satisfies the condition

$$|\Omega\cap B(x,\rho)|\geqq A\rho^n \qquad \forall x_0\in\Omega,\ \rho<\operatorname{diam}\Omega$$

then the functions in $\mathscr{L}^{2,\lambda}$ are Hölder-continuous in $\bar{\Omega}$ with exponent $\alpha=\frac{1}{2}(\lambda-n)$ (see [1]).

We note that if $\phi\in L^2_{\mathrm{loc}}(\Omega)$ and $D\phi\in L^{2,n-2+2\alpha}_{\mathrm{loc}}(\Omega)$, $0<\alpha<1$, we get immediately from Poincaré's inequality that $\phi\in\mathscr{L}^{2,n+2\alpha}_{\mathrm{loc}}(\Omega)$, and hence $\phi\in C^{0,\alpha}_{\mathrm{loc}}(\Omega)$. This is the classical Dirichlet growth theorem of Morrey ([16], Theorem 3.5.2).

Our technique consists in comparing u in a ball $B_R=B(x_0,R)$ with the function v minimizing the functional

$$F^0(v;B_R)=\int_{B_R} f(x_0,u_{x_0,R},Dv)\,dx \tag{2.8}$$

among all functions in $W^{1,2}(B_R,\mathbb{R}^N)$ taking the values u on ∂B_R; i.e. such that the function

$$w=u-v$$

belongs to $W^{1,2}_0(B_R,\mathbb{R}^N)$.

Lemma 2.1. *Let v minimize F^0 in B_R among all functions taking the values u on ∂B_R. Suppose that* (2.4) *hold, and let $w=u-v$. Then:*

$$\int_{R_R}|Dw|^2\,dx\leqq\frac{2}{\nu}\,[F^0(u;B_R)-F^0(v;B_R)] \tag{2.9}$$

Proof. Setting

$$f^0(p)=f(x_0,u_{x_0,R},p)$$

we have

$$\begin{aligned} &f^0(Du)-f^0(Dv)\\ &\quad=f^0_{p^i_\alpha}(Dv)\,D_\alpha w^i+\int_0^1(1-t)\,f^0_{p^i_\alpha p^j_\beta}(t\,Du+(1-t)\,Dv)\,D^\alpha w^i\,D^\beta w^j\,dt\\ &\quad\geqq f^0_{p^i_\alpha}(Dv)\,D_\alpha w^i+\tfrac{1}{2}\nu\,|Dw|^2. \end{aligned}$$

Integrating on B_R, we get immediately (2.9) since v satisfies the Euler equation

$$\int_{B_R} f^0_{p^i_\alpha}(Dv)\,D_\alpha\phi^i\,dx=0 \qquad \forall\phi\in W^{1,2}_0(B_R,\mathbb{R}^N).\quad \text{q.e.d.} \tag{2.10}$$

We conclude this section with a simple lemma that we shall use often. It essentially appears in [15] and [2].

Lemma 2.2. *Let $\phi(t)$ be a non-negative and non-decreasing function. Suppose that*

$$\phi(\rho)\leqq A\left[\left(\frac{\rho}{R}\right)^\alpha+\varepsilon\right]\phi(R)+BR^\beta$$

for all $\rho\leqq R\leqq R_0$, with A, α, β positive constants and B, ε non-negative constants, $\beta<\alpha$. Then there exists a constant $\varepsilon_0=\varepsilon_0(A,\alpha,\beta)$ such that if $\varepsilon<\varepsilon_0$, then for all

$\rho \leqq R \leqq R_0$ *we have*

$$\phi(\rho) \leqq c \left(\frac{\rho}{R}\right)^{\beta} [\phi(R) + BR^{\beta}] \tag{2.11}$$

with $c = c(\alpha, \beta, A)$.

Proof. For $0 < \tau < 1$ and $R \leqq R_0$, we have

$$\phi(\tau R) \leqq A\tau^{\alpha}[1 + \varepsilon\tau^{-\alpha}]\, \phi(R) + BR^{\beta}$$

Choose now $\tau < 1$ in such a way that $2A\tau^{\alpha} = \tau^{\gamma}$ with $\alpha > \gamma > \beta$ and assume that $\varepsilon_0 \tau^{-\alpha} < 1$. Then we get for every $R \leqq R_0$

$$\phi(\tau R) \leqq \tau^{\gamma}\, \phi(R) + BR^{\beta}$$

and therefore for all integers $k > 0$

$$\begin{aligned} \phi(\tau^{k+1} R) &\leqq \tau^{\gamma}\, \phi(\tau^k R) + B\tau^{k\beta} R^{\beta} \\ &\leqq \tau^{(k+1)\gamma}\, \phi(R) + B\tau^{k\beta} R^{\beta} \sum_{j=0}^{k} \tau^{j(\gamma-\beta)} \\ &\leqq c\tau^{(k+1)\beta}[\phi(R) + BR^{\beta}]. \end{aligned}$$

Choosing k such that $\tau^{k+1} R \leqq \rho \leqq \tau^k R$, the last inequality gives at once (2.11). q.e.d.

3. Local Regularity, $N = 1$

We shall begin with the scalar case $N = 1$. If the assumptions (1.2) are satisfied we can apply the results of [6], so that every local minimum $u(x)$ of $F(u; \Omega)$ is Hölder-continuous with some exponent $\alpha > 0$.

Theorem 3.1. *Let u be a local minimum of the functional*

$$F(u; \Omega) = \int_{\Omega} f(x, u, Du)\, dx \tag{3.1}$$

with f satisfying (2.2), (2.3), (2.4) *with* $N = 1$, *and* (2.5). *Then the gradient of u, Du, belongs to* $L^{2, n-\varepsilon}_{\mathrm{loc}}(\Omega)$ *for every* $\varepsilon > 0$.

Proof. Let $B_R \subset \subset \Omega$ and let v minimize the functional F^0 defined in (2.8) with boundary datum u on ∂B_R.

We estimate first the oscillation of v in B_R. If $k > k_0 = \sup_{B_R} u$, we have $F^0(v; B_R) \leqq F^0(\min(v, k); B_R)$ and therefore

$$\int_{A_k} |Dv|^2\, dx \leqq c_1 |A_k|$$

where $A_k = \{x \in B_R : v(x) > k\}$. As in [14], Lemmas 5.1 and 5.2, we conclude that

$$\sup_{B_R} v \leqq \sup_{B_R} u + c_2 R.$$

Similarly

$$\inf_{B_R} v \geqq \inf_{B_R} u - c_2 R$$

and therefore

$$\operatorname{osc}(v; B_R) \leqq \operatorname{osc}(u; B_R) + 2c_2 R \leqq c_3 R^\alpha. \tag{3.2}$$

The second step consists in estimating the supremum of $|Dv|$ in $B_{R/2}$. Starting from (2.10) it is easy to prove that $v \in W^{2,2}_{\mathrm{loc}}(B_R)$ and that its derivatives $D_\gamma v$ are solutions of the elliptic equation

$$\int_{B_R} f^0_{p_\alpha p_\beta}(Dv) D_\beta(D_\gamma v) D_\alpha \phi \, dx = 0 \qquad \forall \phi \in W^1_0(B_R). \tag{3.3}$$

(see e.g. [10], p. 166).

Taking $\phi = \eta D_\gamma v$, $\eta \in C^\infty_0(B_R)$, and summing over γ we conclude that $z = |Dv|^2$ satisfies the inequality

$$\int_{B_R} f^0_{p_\alpha p_\beta}(Dv) D_\beta z D_\alpha \eta \leqq 0 \qquad \forall \eta \in C^\infty_0(B_R),\ \eta \geqq 0.$$

From the standard elliptic estimates (see e.g. [9], p. 184) we get

$$\sup_{B_{R/2}} |Dv|^2 \leqq c_4 R^{-n} \int_{B_R} |Dv|^2 \, dx. \tag{3.4}$$

From (3.4) we obtain for every $\rho < R/2$

$$\int_{B_\rho} |Dv|^2 \, dx \leqq c_5 \left(\frac{\rho}{R}\right)^n \int_{B_R} |Dv|^2 \, dx; \tag{3.5}$$

changing possibly the constant c_5 the above inequality holds for every $\rho < R$.

Coming back to u we get

$$\begin{aligned} \int_{B_\rho} |Du|^2 \, dx &\leqq 2 \int_{B_\rho} |Dv|^2 \, dx + 2 \int_{B_\rho} |Dw|^2 \, dx \\ &\leqq c_6 \left\{ \left(\frac{\rho}{R}\right)^n \int_{B_R} |Du|^2 \, dx + \int_{B_R} |Dw|^2 \, dx \right\}. \end{aligned} \tag{3.6}$$

In order to estimate the last integral we use (2.9). We have

$$\begin{aligned} F^0(u; B_R) - F^0(v; B_R) = &\int_{B_R} [f(x_0, u_{x_0,R}, Du) - f(x, u, Du)] \, dx \\ &+ \int_{B_R} [f(x, v, Dv) - f(x_0, u_{x_0,R}, Dv)] \, dx + F(u; B_R) - F(v; B_R) \end{aligned}$$

and taking into account (2.5), (3.2) and the Hölder-continuity of u

$$\int_{B_R} |Dw|^2 \, dx \leqq c_7 \, \omega_1(R) \int_{B_R} (1 + |Du|^2) \, dx \tag{3.7}$$

with $\omega_1(t) \downarrow 0$ as $t \downarrow 0$. In conclusion we have

$$\int_{B_\rho} |Du|^2\,dx \leqq c_8 \left\{ \left[\left(\frac{\rho}{R}\right)^n + \omega_1(R) \right] \int_{B_R} |Du|^2\,dx + R^n \right\}$$

for every $\rho < R < \operatorname{dist}(x_0, \partial\Omega)$.

We can apply now Lemma 2.2. If $\Omega_1 \subset\subset \Omega$, for every $\varepsilon > 0$ there exists $R_0 < \operatorname{dist}(\Omega_1, \partial\Omega)$ and $c_9(\varepsilon)$ such that for every $x_0 \in \Omega_1$ and $\rho < R_0$

$$\int_{B_\rho(x_0)} |Du|^2\,dx \leqq c_9 \left\{ \left(\frac{\rho}{R}\right)^{n-\varepsilon} \int_{B_{R_0}(x_0)} |Du|^2\,dx + \rho^{n-\varepsilon} \right\}$$

and the conclusion follows at once.

Remark 3.2. From the Poincaré inequality we deduce immediately that $u \in \mathscr{L}^{2,n+2-\varepsilon}_{\mathrm{loc}}(\Omega)$, and therefore it is Hölder-continuous in Ω with any exponent less than 1.

Remark 3.3. The same conclusion holds if u minimizes the functional

$$\int_\Omega [f(x, u, Du) + g(x, u, Du)]\,dx$$

with f as above, and

$$|g(x, u, p)| \leqq A(1 + |p|^2)^{r/2}, \qquad r < 2.$$

In this case, instead of (3.7) we have

$$\int_{B_R} |Dw|^2\,dx \leqq c_7 \{\omega_1(R) \int_{B_R} |Du|^2\,dx + R^n + \int_{B_R} (|Du|^r + |Dv|^r)\,dx\}. \tag{3.8}$$

On the other hand

$$\begin{aligned} \int_{B_R} |Dv|^r\,dx &\leqq c_{10} \int_{B_R} (R^{\varepsilon\frac{2-r}{r}} |Dv|^2 + R^{-\varepsilon})\,dx \\ &\leqq c_{11} \{R^{\varepsilon\frac{2-r}{r}} \int_{B_R} |Du|^2\,dx + R^{n-\varepsilon}\}. \end{aligned}$$

A similar estimate holding for $\int_{B_R} |Du|^r\,dx$, we get in conclusion

$$\int_{B_R} |Dw|^2\,dx \leqq c_{12} \{\omega_2(R) \int_{B_R} |Du|^2\,dx + R^{n-\varepsilon}\}$$

$\omega_2(R) \downarrow 0$ as $R \downarrow 0$, and the result follows as above.

Besides its interest in itself, Theorem 3.1 is important as a first step in the proof of the regularity of the derivatives of u.

For that, we have to make a further assumption on the function f, namely that f is a Hölder-continuous in (x, u) with exponent 2σ. More precisely we shall assume that the function ω in (2.5) satisfies

$$\omega(t) \leqq A t^\sigma \tag{3.9}$$

for some $\sigma > 0$.

Theorem 3.4. *Let the assumptions of Theorem* 3.1 *be satisfied. Suppose moreover that* (3.9) *holds. Then u has Hölder-continuous first derivatives in Ω.*

Proof. Let $B_R \subset \Omega$ and let v be as above. Since Dv satisfies (3.3), we have from De Giorgi's theorem [3], the estimate

$$\int_{B_\rho} |Dv-(Dv)_{x_0,\rho}|^2\,dx \leqq c_{13}\left(\frac{\rho}{R}\right)^{n+2\delta} \int_{B_R} |Dv-(Dv)_{x_0,R}|^2\,dx$$

for some $\delta>0$ (compare also [4], Chap. VII).

Recalling that for every $z\in W^{1,2}(B_r;\mathbb{R}^N)$ and every $\lambda\in\mathbb{R}^{nN}$ we have

$$\int_{B_r} |Dz-(Dz)_{x_0,r}|^2\,dx \leqq \int_{B_r} |Dz-\lambda|^2\,dx,$$

we get easily

$$\int_{B_\rho} |Du-(Du)_{x_0,\rho}|^2\,dx \leqq c_{14}\left\{\left(\frac{\rho}{R}\right)^{n+2\delta} \int_{B_R} |Du-(Du)_{x_0,R}|^2\,dx + \int_{B_R} |Dw|^2\,dx\right\}.$$

The last integral can be estimated as above. Taking into account that $u\in C^{0,\alpha}$, we get

$$\int_{B_R} |Dw|^2\,dx \leqq c_{15} R^{2\alpha\sigma} \int_{B_R} (1+|Du|^2)\,dx$$

and using the conclusion of Theorem 3.1:

$$\int_{B_R} |Dw|^2\,dx \leqq c_{16} R^{n+2\alpha\sigma-\varepsilon}.$$

Taking $\varepsilon=\alpha\sigma$, we obtain

$$\int_{B_\rho} |Du-(Du)_{x_0,\rho}|^2\,dx \leqq c_{17}\left\{\left(\frac{\rho}{R}\right)^{n+2\delta} \int_{B_R} |Du-(Du)_{x,R}|^2\,dx + R^{n+\alpha\sigma}\right\}$$

for every $\rho<R$.

Using again Lemma 2.2 we conclude that

$$\int_{B_\rho} |Du-(Du)_{x,\rho}|^2\,dx \leqq c_{18}\,\rho^{n+2\tau}$$

with $\tau=\min\left(\frac{\alpha\sigma}{2},\frac{\delta}{2}\right)$, and hence u has Hölder-continuous first derivatives. q.e.d.

4. Local Regularity, $N\geqq 1$

It is well known that in the general case $N\geqq 1$ everywhere regularity is a very uncommon feature. Here we shall consider only functionals of the type:

$$F(u;\Omega)=\int_\Omega A_{ij}^{\alpha\beta}(x)\,D_\alpha u^i\,D_\beta u^j\,dx+\int_\Omega g(x,u,Du)\,dx \tag{4.1}$$

where the coefficients $A_{ij}^{\alpha\beta}$ are continuous in Ω and satisfy the Legendre-Hadamard condition:

$$A_{ij}^{\alpha\beta}(x)\,\xi_\alpha\,\xi_\beta\,\eta^i\,\eta^j \geqq \nu|\xi|^2|\eta|^2 \qquad \forall\,\xi\in\mathbb{R}^n,\ \forall\,\eta\in\mathbb{R}^N;\ \nu>0 \tag{4.2}$$

and

$$|g(x,u,p)| \leqq A(1+|p|^r) \qquad r<2. \tag{4.3}$$

We have

Theorem 4.1. *Under the hypotheses stated above, every local minimum u of* (4.1) *has first derivatives in $L_{\mathrm{loc}}^{2,n-\varepsilon}(\Omega)$ for every $\varepsilon>0$. In particular u is Hölder-continuous in Ω with every exponent less than* 1.

Proof. As in Theorem 3.1 we freeze the coefficients at $x_0\in\Omega$, and consider the function v minimizing the functional

$$F^0(v;B_R)=\int_{B_R} A_{ij}^{\alpha\beta}(x_0)\,D_\alpha v^i\,D_\beta v^j\,dx \tag{4.4}$$

among all functions taking the value $u(x)$ on ∂B_R. Such a function v is smooth in B_R and we have (see e.g. [2]; [4], p. 80; [10], p. 73):

$$\int_{B_\rho}|Dv|^2\,dx \leqq c_{19}\left(\frac{\rho}{R}\right)^n \int_{B_R}|Dv|^2\,dx$$

hence

$$\int_{B_\rho}|Du|^2\,dx \leqq c_{20}\left\{\left(\frac{\rho}{R}\right)^n \int_{B_R}|Dv|^2 + \int_{B_R}|Dw|^2\,dx\right\}.$$

Using Lemma 2.1 and arguing as in Remark 3.3 we get

$$\int_{B_R}|Dw|^2\,dx \leqq c_{21}\{\omega_3(R)\int_{B_R}|Du|^2\,dx + R^{n-\varepsilon}\}$$

with $\omega_3(R)\downarrow 0$ as $R\downarrow 0$. The conclusion now follows as in Theorem 3.1. q.e.d.

If we want $C^{1,\alpha}$-regularity we have to assume that the coefficients $A_{ij}^{\alpha\beta}$ are Hölder-continuous in Ω, and moreover that $g=g(x,u)$ is Hölder-continuous in (x,u).

Theorem 4.2. *Let u be a local minimum of the functional*

$$\int_\Omega A_{ij}^{\alpha\beta}(x)\,D_\alpha u^i\,D_\beta u^j\,dx + \int_\Omega g(x,u)\,dx$$

with Hölder-continuous coefficients $A_{ij}^{\alpha\beta}$ satisfying (4.2). *Suppose that g is Hölder-continuous in $\Omega\times\mathbb{R}^N$. Then u has Hölder-continuous first derivatives in Ω.*

Proof. Since our result is local, we can suppose without loss of generality that $Du\in L^{2,n-\varepsilon}(\Omega)$ for every $\varepsilon>0$ (Theorem 4.1). The function v which minimizes (4.4) among all functions coinciding with u on ∂B_R satisfies (see e.g. [2]; [4], p. 80; [10], p. 73):

$$\int_{B_\rho}|Dv-(Dv)_{x_0,\rho}|^2\,dx \leqq c_{22}\left(\frac{\rho}{R}\right)^{n+2}\int_{B_R}|Dv-(Dv)_{x_0,R}|^2\,dx. \tag{4.5}$$

We have $w=u-v\in W_0^{1,2}(B_R, \mathbb{R}^N)$ and

$$\int_{B_R} A_{ij}^{\alpha\beta}(x_0)\, D_\alpha w^i D_\beta \phi^j dx = \int_{B_R} A_{ij}^{\alpha\beta}(x_0)\, D_\alpha u^i D_\beta \phi^j dx$$
$$= \int_{B_R} h_j^\beta D_\beta \phi^j dx, \qquad \forall \phi\in W_0^{1,2}(B_R, \mathbb{R}^N)$$

with $h_j^\beta\in L^{2,n-\varepsilon}(\Omega)$ for every $\varepsilon>0$.

From the regularity theory for linear elliptic systems with constant coefficients (see e.g. [2]) we get $Dw\in L^{2,n-\varepsilon}(B_R)$ and

$$\|Dw\|_{L^{2,n-\varepsilon}(B_R)} \leqq c_{23}\|Du\|_{L^{2,n-\varepsilon}(B_R)} \leqq c_{23}\|Du\|_{L^{2,n-\varepsilon}(\Omega)}$$

with c_{23} independent of R. It follows that w, and hence v, are Hölder-continuous in B_R, and

$$\operatorname{osc}(v; B_R) \leqq c_{24} R^{1-\varepsilon/2}. \tag{4.6}$$

We can now proceed as in Theorem 3.4. We have

$$\int_{B_\rho} |Du-(Du)_{x,\rho}|^2 dx \leqq c_{25}\left\{\left(\frac{\rho}{R}\right)^{n+2} \int_{B_R} |Du-(Du)_{x,R}|^2 dx + \int_{B_R} |Dw|^2 dx\right\}$$

and taking into account Lemma 2.1 and (4.6) (compare with the proof of Theorem 3.1):

$$\int_{B_R} |Dw|^2 dx \leqq c_{26}\{R^{2\sigma} \int_{B_R} |Du|^2 dx + R^{n+2\alpha\sigma}\}$$

with $\alpha=1-\varepsilon/2$.

The result now follows as in Theorem 3.4. q.e.d.

Remark. We note that when g is Hölder-continuous with exponent 2σ, the minimum u belongs to $C^{1,\delta}$ for every $\delta<\sigma$.

We do not know whether this result is optimal; however the Hölder exponent of Du cannot reach 2σ, even if $N=1$ and the functional has the special form

$$\int (\tfrac{1}{2}|Du|^2 + |u|^{2\sigma})\, dx$$

(see [18]).

5. Partial Regularity

When $N>1$, we usually expect partial regularity, i.e. regularity in some open set Ω_0 with $\Omega-\Omega_0$ small. Actually we have

Theorem 5.1. *Let u be a local minimum of the functional* (1.1). *Suppose that $f(x,u,p)$ satisfies* (2.2), (2.3), (2.4) *and* (2.5) *with $\omega(t)\leqq At^\sigma$, $\sigma>0$.*

Then there exists an open set $\Omega_0\subset\Omega$ such that u has Hölder-continuous first derivatives in Ω_0.

Moreover we have $\Omega-\Omega_0=\Sigma_1\cup\Sigma_2$, where

$$\Sigma_1=\{x_0\in\Omega\colon \sup_{\rho>0} |(Du)_{x_0,\rho}| = +\infty\},$$
$$\Sigma_2=\{x_0\in\Omega\colon \liminf_{\rho\to 0} \rho^{-n} \int_{B_\rho} |Du-(Du)_{x_0,\rho}|^2 dx>0\}$$

and therefore

$$|\Omega-\Omega_0|=0.$$

Proof. We shall estimate

$$U(x_0,\rho)=\fint_{B(x_0,\rho)}|Du-(Du)_{x_0,\rho}|^2\,dx. \tag{5.1}$$

Let $x_0\in\Omega$, $R<\operatorname{dist}(x_0,\partial\Omega)$, and let $v\in W^{1,2}(B_R)$ be the minimum for $F^0(v;B_R)$ in (2.8) with boundary values $u(x)$ on ∂B_R. The function Dv is a solution of the system (3.3) which we rewrite in the form:

$$\int_{B_R}\delta^{\gamma\sigma}f^0_{p^i_\alpha p^j_\beta}(Dv)\,D_\beta(D_\sigma v^j)\,D_\alpha\phi^i_\gamma\,dx=0 \tag{5.2}$$

for every $\phi\in W^{1,2}_0(B_R;\mathbb{R}^{nN})$.

From the regularity theory of nonlinear elliptic systems [12, 17, 4] for every $L_0>0$ there exists $\eta_0(L_0)>0$ (depending only on L_0, v, n, N and on the modulus of continuity of f_{pp}) such that if

$$\fint_{B_R(x_0)}|Dv|^2\,dx\leqq L_0^2 \quad\text{and}\quad \fint_{B_R(x_0)}|Dv-(Dv)_{x_0,R}|^2\,dx<\eta_0^2 \tag{5.3}$$

then Dv is Hölder-continuous (with every exponent $\alpha<1$) in a neighborhood of x_0 and moreover

$$\int_{B_\rho(x_0)}|Dv-(Dv)_{x_0,\rho}|^2\,dx\leqq c_{28}\left(\frac{\rho}{R}\right)^{n+2\alpha}\int_{B_R(x_0)}|Dv-(Dv)_{x_0,R}|^2\,dx \tag{5.4}$$

for every $\rho<R$.

We want to show now that (5.3) can be replaced by similar inequalities involving the function u. For that we note first that from (2.2) we have

$$\fint_{B_R(x_0)}|Dv|^2\,dx\leqq c_{29}\fint_{B_R(x_0)}(1+|Du|^2)\,dx. \tag{5.5}$$

On the other hand from (2.10) that we can rewrite as

$$\int_{B_R(x_0)}[f^0_{p^i_\alpha}(Dv)-f^0_{p^i_\alpha}(\xi)]\,D_\alpha\phi^i\,dx=0 \qquad \forall\,\phi\in W^{1,2}_0(B_R,\mathbb{R}^N) \tag{5.6}$$

taking $\phi^i=v^i-u^i=v^i-\xi^i_\alpha x^\alpha-(u^i-\xi^i_\alpha x^\alpha)$, $\xi^i_\alpha=(D_\alpha u^i)_{x_0,R}$, we get

$$\fint_{B_R}|Dv-(Dv)_{x_0,R}|^2\,dx\leqq c_{30}\fint_{B_R}|Du-(Du)_{x_0,R}|^2\,dx. \tag{5.7}$$

From (5.5) and (5.7) we conclude that for every $M_0>0$ there exists $\varepsilon_0(M_0)$ such that the inequalities

$$\fint_{B_R(x)}|Du|^2\,dx\leqq M_0^2 \quad\text{and}\quad U(x_0,R)<\varepsilon_0^2 \tag{5.8}$$

imply (5.3) and therefore (5.4).

Suppose now that (5.8) is satisfied for some $R < \operatorname{dist}(x_0, \partial\Omega)$. In this case we have as in Theorem 3.4

$$\int_{B_\rho(x_0)} |Du-(Du)_{x_0,\rho}|^2\,dx \leqq \left\{c_{31}\left(\frac{\rho}{R}\right)^{n+2\alpha} \int_{B_R(x_0)} |Du-(Du)_{x_0,R}|^2\,dx + \int_{B_R} |Dw|^2\,dx\right\}. \tag{5.9}$$

The estimate of the last integral is made as in Theorem 3.4. We do not know now that u is Hölder-continuous and instead of (3.7) we get the inequality

$$\int_{B_R} |Dw|^2\,dx \leqq c_{32} \int_{B_R} \omega(R^2+|u-u_{x_0,R}|^2+|u-v|^2)(1+|Du|^2)\,dx.$$

We recall now that assumption (1.2) implies (see [6]) that $Du \in L^q$ for some $q>2$, and we have (compare with [5])

$$\left(\fint_{B_R} (1+|Du|^2)^{q/2}\,dx\right)^{2/q} \leqq c_{33} \fint_{B_{2R}} (1+|Du|^2)\,dx.$$

We can always suppose that q is so close to 2 that $\dfrac{q\sigma}{q-2} \geqq 1$. From the Hölder inequality:

$$\begin{aligned}\int_{B_R} |Dw|^2\,dx &\leqq c_{35}\left[\int_{B_R} (1+|Du|^2)^{q/2}\,dx\right]^{2/q}\left(\int_{B_R} \omega^{\frac{q}{q-2}}\,dx\right)^{1-2/q}\\ &\leqq c_{36} \int_{B_{2R}} (1+|Du|^2)\,dx \left(\fint_{B_R} \omega^{\frac{q}{q-2}}\,dx\right)^{1-2/q}.\end{aligned}$$

Since ω is bounded we have

$$\int_{B_R} |Dw|^2\,dx \leqq c_{37} \int_{B_R} (1+|Du|^2)\,dx. \tag{5.10}$$

Moreover, recalling that $\dfrac{q}{q-2} > \dfrac{1}{\sigma}$ and $\omega(t) \leqq A t^\sigma$, we get

$$\omega^{\frac{q}{q-2}} \leqq L\omega^{\frac{1}{\sigma}} \leqq c_{38}(R^2+|u-u_{x_0,R}|^2+|u-v|^2)$$

and therefore using the Poincaré inequality and (5.10) we obtain

$$\fint_{B_R} \omega^{q/q-2}\,dx \leqq c_{39} R^2 \fint_{B_R} (1+|Du|^2)\,dx.$$

In conclusion

$$\int_{B_R} |Dw|^2 \leqq c_{40} R^{n+2\left(1-\frac{2}{q}\right)} \left\{\fint_{B_{2R}} (1+|Du|^2)\,dx\right\}^{2-2/q}$$

and therefore from (5.9), writing R instead of $2R$:

$$U(x_0,\tau R) \leqq c_{41}\left\{\tau^{2\alpha} U(x_0,R) + R^{2\left(1-\frac{2}{q}\right)} \tau^{-n} M(x_0,R)^{4\left(1-\frac{1}{q}\right)}\right\} \tag{5.11}$$

where $0<\tau<\frac{1}{2}$ and

$$M(x_0, R)=1+|(Du)_{x_0, R}|+U(x_0, R)^{\frac{1}{2}}.$$

The proof now proceeds as in [8] and [11].

First we choose $\theta<2\alpha$ and $\tau<\frac{1}{2}$ in such a way that $c_{41}\tau^{2\alpha-\theta}\leqq 1$. Suppose further that for some M_0 and some $R<R_1(M_0)\leqq 1$ we have

$$M(x_0, R)\leqq\tfrac{1}{2}M_0, \qquad U(x_0, R)\leqq\varepsilon_1^2\leqq\tfrac{1}{2}\varepsilon_0^2$$

where R_1 and ε_1 will be specified later.

We have from (5.11)

$$U(x_0, \tau R)\leqq\tau^\theta U(x_0, R)+c_{42}M_0^\gamma R^\delta \tag{5.12}$$

$(c_{42}=c_{41}\tau^{-n})$ where we have set $\gamma=4\left(1-\frac{1}{q}\right)$ and $\delta=\min\left\{\frac{\theta}{2}, 2\left(1-\frac{2}{q}\right)\right\}$. We have

$$|(Du)_{x_0, \tau^k R}|\leqq|(Du)_{x_0, R}|+\sum_{j=0}^{k-1}U(x_0, \tau^j R)^{\frac{1}{2}}$$

and therefore

$$M(x_0, \tau^k R)\leqq M(x_0, R)+\sum_{j=0}^{k}U(x_0, \tau^j R)^{\frac{1}{2}}. \tag{5.13}$$

We shall prove by induction that

$$U(x_0, \tau^k R)\leqq\tau^{\theta k}U(x_0, R)+c_{42}M_0^\gamma(\tau^{k-1}R)^\delta\sum_{j=1}^{k}\tau^{(\theta-\delta)(k-j)}. \tag{5.14}$$

Actually (5.14) is true for $k=1$. Suppose it holds for $k\leqq s$. We have

$$U(x_0, \tau^k R)\leqq\tau^{\theta k}\varepsilon_1^2+c_{42}\frac{M_0^\gamma R^\delta}{\tau^\delta-\tau^\theta}\tau^{\delta k}\leqq\left(\varepsilon_1^2+c_{42}\frac{M_0^\gamma R^\delta}{\tau^\delta-\tau^\theta}\right)\tau^{\delta k} \tag{5.15}$$

and therefore from (5.13):

$$M(x_0, \tau^s R)\leqq\tfrac{1}{2}M_0+\left(\varepsilon_1^2+c_{42}\frac{M_0^\gamma R^\delta}{\tau^\delta-\tau^\theta}\right)\frac{1}{1-\tau^{\delta/2}}. \tag{5.16}$$

Let us choose now ε_1 and R_1 in such a way that

$$\frac{2\varepsilon_1^2}{1-\tau^{\delta/2}}\leqq\tfrac{1}{2}M_0$$

and

$$c_{42}\frac{M_0^\gamma R_1^\delta}{\tau^\delta-\tau^\theta}\leqq\varepsilon_1^2.$$

We have from (5.15) and (5.16)

$$M(x_0, \tau^s R)\leqq M_0, \qquad U(x_0, \tau^s R)\leqq 2\varepsilon_1^2\leqq\varepsilon_0^2$$

and therefore using (5.12)

$$U(x_0, \tau^{s+1}R)\leqq\tau^\theta U(x_0, \tau^s R)+c_{42}M_0^\gamma(\tau^s R)^\delta$$

so that (5.14) with $k \leqq s$ gives the same inequality for $k=s+1$. In particular we have for every integer k:

$$U(x_0, \tau^k R) \leqq 2\varepsilon_1^2 \tau^{\delta k}$$

and therefore with a simple argument

$$U(x_0, \rho) \leqq c_{43}\, \varepsilon_1^2 \left(\frac{\rho}{R}\right)^{\delta}. \tag{5.17}$$

Now (5.17) holds for every $x_0 \in \Omega$ such that for some $R < R_1$ we have

$$M(x_0, R) \leqq \tfrac{1}{2} M, \qquad U(x_0, R) \leqq \varepsilon_1^2. \tag{5.18}$$

Since $M(x, R)$ and $U(x, R)$ are continuous functions of x, the inequalities (5.18) are satisfied in a neighborhood of x_0 whenever they hold for x_0. We have therefore

$$U(x, \rho) \leqq c_{43}\, \varepsilon_1^2 \left(\frac{\rho}{R}\right)^{\delta}$$

for every x in a neighborhood B of x_0 and hence, by [1], the derivatives of u are Hölder-continuous in B.

Since (5.18) hold for every Lebesgue point of Du, we infer that the measure of the singular set $\Omega - \Omega_0$ is zero. This concludes the proof of the theorem. q.e.d.

References

1. Campanato, S.: Proprietà di hölderianità di alcune classi di funzioni. Ann. S.N.S. Pisa **17**, 137–160 (1963)
2. Campanato, S.: Equazioni ellittiche del secondo ordine e spazi $\mathscr{L}^{2,\lambda}$. Ann. Mat. Pura e Appl. **69**, 321–380 (1965)
3. De Giorgi, E.: Sulla differenziabilità e l'analiticità delle estremali degli integrali multipli regolari. Mem. Accad. Sci. Torino, Cl. Sci. Fis. Mat. Nat. (3) **3**, 25–43 (1957)
4. Giaquinta, M.: Multiple integrals in the Calculus of Variations and non linear elliptic systems, Vorlesungsreihe SFB 72 Bonn n.6 (1981)
5. Giaquinta, M., Giusti, E.: Non linear elliptic systems with quadratic growth. Manuscripta math. **24**, 323–349 (1978)
6. Giaquinta, M., Giusti, E.: On the regularity of the minima of variational integrals. Acta Math. **148**, 31–46 (1982)
7. Giaquinta, M., Giusti, E.: Quasi-minima, to appear
8. Giaquinta, M., Modica, G.: Almost-everywhere regularity results for solutions of nonlinear elliptic systems. Manuscripta math. **28**, 109–158 (1979)
9. Gilbarg, D., Trudinger, N.S.: Elliptic partial differential equations of second order. Berlin-Heidelberg-New York: Springer 1977
10. Giusti, E.: Equazioni ellittiche del secondo ordine. Quaderni Unione Mat. Italiana n.6 (1978); Ed. Pitagora, Bologna
11. Giusti, E.: Regolarità parziale delle soluzioni di sistemi ellittici quasi lineari di ordine arbitrario. Ann. Scuola Norm. Sup. Pisa **23**, 115–141 (1969)
12. Giusti, E., Miranda, M.: Sulla regolarità delle soluzioni deboli di una classe di sistemi ellittici quasilineari. Arch. Rat. Mech. Anal. **31**, 173–184 (1968)
13. Ivert, P.A.: Partial regularity of vector valued functions minimizing variational integrals, preprint

14. Ladyzhenskaya, O.A., Uraltseva, N.N.: Linear and quasilinear elliptic equations. New York: Acad. Press 1968
15. Morrey, C.B., Jr.: Second order elliptic systems of differential equations. Ann. of Math. Studies n. 33, Princeton Univ. Press 1954, pp. 101-159
16. Morrey, C.B., Jr.: Multiple integrals in the calculus of variations. Berlin-Heidelberg-New York: Springer 1966
17. Morrey, C.B., Jr.: Partial regularity results for non linear elliptic systems. J. Math. and Mech. **17**, 649-670 (1968)
18. Phillips, D.: A minimization problem and the regularity of solutions in the presence of a free boundary, pre-print

Oblatum 5-X-1982

Invent. math. 72, 299–314 (1983)

Inventiones mathematicae

Ergodic Measure Preserving Transformations with Arbitrary Finite Spectral Multiplicities

E.A. Robinson, Jr.

Department of Mathematics, University of Maryland, College Park, 20742, USA

Introduction

For an invertible measure preserving transformation T: $(X,\mu)\to(X,\mu)$ of a Lebesgue space (X,μ) we consider the induced unitary operator U_T: $L_2(X,\mu)\to L_2(X,\mu)$ given by $(U_T f)(x)=f(Tx)$. The simplest concept of spectral multiplicity (cf. [14]) is the maximal spectral multiplicity which we now define. Let U: $H\to H$ be a unitary operator on a Hilbert space H. The maximal spectral multiplicity

$$m(u)=\inf\{m\in\mathbb{Z}_+\cup\{\infty\}\colon \exists f_1\dots f_m\in H\colon \text{linear combinations of } U^i f_j,\ i\in\mathbb{Z},\ j=1,\dots,m \text{ are dense in } H\}.$$

When the multiplicity is 1 we say that the spectrum is simple. If for U_T, the eigenvalue 1 is simple and is the only eigenvalue we say T has continuous spectrum.

The question of spectral multiplicity in ergodic theory has a long and interesting history. Without trying to be complete we would like to mention its most important moments. The problem was first considered by Von Neuman [18] who observed that the spectrum is simple for ergodic flows with purely discrete spectrum and asked what is possible in the continuous spectrum case.

Von Neuman's results on flows were extended to the case of transformations with discrete spectrum by Halmos and Von Neumann in [5] where it was shown that such transformations are equivalent to rotations on compact abelian groups. As in the case of flows, these examples have simple spectrum. In contrast to the discrete spectrum case is another natural class of algebraic examples, the ergodic automorphisms of compact abelian groups. These were shown by Halmos [6] to have continuous spectrum with infinite multiplicity, namely the type of spectrum commonly called countable Lebesgue spectrum. Countable Lebesgue spectrum also occurs in a class of transformations which arise naturally in probability theory, the Bernoulli and Markov shifts. (Cf. e.g. [13] where there is also an interesting general discussion of the spectral multiplicity problem.) Rohlin [12] also described the connection between the spectrum and entropy. Any transformation which positive entropy has a countable Lebesgue component in the spectrum.

For a long time it was unknown whether mixed or continuous spectrum of finite multiplicity was possible. The first example was due to Girsanov [4] who constructed a transformation with simple continuous spectrum using the theory of Gaussian processes. Soon afterward Veršik proved in [16] and [17] that for any ergodic transformation generated by a Gaussian process, the spectral multiplicity is either 1 or infinity.

Examples with simple continuous spectrum of a geometric rather than probabilistic origin were first constructed by Oseledec [11] and Katok and Stepin [4]. These examples were found among interval exchange transformations and smooth flows on the two-dimensional torus. They are based on the idea of approximation by periodic transformations (cf. [5]). Using these ideas Yuzvinskii ([19] cf. also [10]) proved that simple continuous spectrum is typical in the weak topology on the space of all transformations.

In [17], Oseledec also proved that for an interval exchange transformation the maximal spectral multiplicity is bounded above by $p-1$ where p is the number of intervals exchanged. Moreover he constructed the first example of a transformation with continuous spectrum and finite multiplicity greater than 1. Since the example is an exchange of 30 intervals, the maximal spectral multiplicity m satisfies $2 \leq m < 30$.

Variations on the method of approximation by periodic transformations were subsequently introduced and applied to spectral problems by Chacon [2, 3] and Schwartzbauer [15]. In particular, the existence of a simple approximation with multiplicity M is given in [3] as a sufficient condition for an upper bound M on the spectral multiplicity, although no example is given where this bound is achieved for $M>1$. Baxter [1] noted that the transformations where $M=1$ are exactly those which may be constructed by cutting and stacking intervals in a single stack, the rank 1 transformations (cf. [8]). Thus rank 1 implies simple spectrum. Del Junco [8] showed that the converse is false. Recently, Thouvenot (unpublished) constructed a transformation with simple spectrum which is not loosely Bernoulli and consequently is not of finite rank. A sufficiently fast m-cyclic approximation by periodic transformations (cf. § 3) implies a simple approximation with multiplicity m and thus a rank of at most m.

Recently Katok (unpublished) showed that in the Oseledec example $m=2$ is attainable and is typical within the context of the construction. The upper bound is obtained by using the theory of approximation by periodic transformations. The construction described in this paper is a generalization of the Oseledec construction and Katok's upper bound. We show that for every $m>1$ there exists a measure preserving transformation T with maximal spectral multiplicity m. We show this to be typical for our construction. Furthermore we show that it is possible to realize such a transformation as an interval exchange.

We can give an equivalent definition of spectral multiplicity in terms of the spectral theorem for unitary operators. The operator U_T is described up to unitary equivalence by a sequence of spectral types [14]

$$\rho_1 < \rho_2 < \rho_3 < \dots$$

where each ρ_j is an equivalence class of measures on the circle $\mathbb{T}$, and $<$ denotes absolute continuity applied to these classes. The maximal spectral type $\rho_{\max}$ is the maximal element of the spectral sequence. The multiplicity function m is defined to be the essential number of ρ_j dominated by $\rho_{\max}$. It is $\rho_{\max}$ measurable and has an essential supremum equal to the maximal spectral multiplicity. In terms of this definition the construction described in this paper has a spectrum with a simple component and a component of multiplicity m.

The spectrum of a unitary operator is called homogeneous if the multiplicity function is essentially constant. A long standing unsolved problem is whether an ergodic measure preserving transformation can have homogeneous spectrum with multiplicity different from 1 or infinity in the orthogonal complement to the space of constants. More generally what essential ranges of the multiplicity function are possible? The examples constructed in this paper all have essential range $\{1, m\}$. A particularly interesting unsolved problem is whether Lebesgue spectrum of finite multiplicity is possible for an ergodic measure preserving transformation. In this case each ρ_j must be equivalent to Lebesgue measure. It follows from [10] that all the examples constructed here have singular spectrum.

In what follows we will assume all measure preserving transformations are invertible. We will often neglect to write the measures if no confusion will arise. We will denote by $\mathbb{Z}/m\mathbb{Z}$, the integers $\bmod m$, and by $GF(p)$ the finite field of order p. Where appropriate we will assume arithmetic is carried out $\bmod m$ or $\bmod p$ without so stating. The letter χ will be used both for group characters and characteristic functions.

I wish to express my sincere thanks to Professor A. Katok for suggesting this problem and for his continuing interest in this paper.

§ 1. Basic Construction

For a fixed prime p let us consider the finite field $GF(p)$ and recall that the nonzero elements $GF(p)^\times$ form a group under multiplication which is isomorphic to $\mathbb{Z}/(p-1)\mathbb{Z}$. We denote this isomorphism by $\phi_0\colon \mathbb{Z}/(p-1)\mathbb{Z} \to GF(p)^\times \subseteq GF(p)$.

Lemma 1.1. *For any $m>0$ there exists a finite field $GF(p)$ such that $\mathbb{Z}/m\mathbb{Z}$ is isomorphic to a subgroup of $GF(p)^\times$. We denote this isomorphism by ϕ.*

Proof. By the Dirichlet theorem on primes in an arithmetic progression, the sequence $mk+1$ contains a prime p. This implies that $m \mid p-1$. Consider the subgroup of $\mathbb{Z}/(p-1)\mathbb{Z}$ generated by $m'=(p-1)/m$. This subgroup is isomorphic to $\mathbb{Z}/m\mathbb{Z}$ via $\psi\colon \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/(p-1)\mathbb{Z}$, where $\psi(y)=m'y$. It follows that $\phi = \phi_0 \circ \psi$ is a one to one homomorphism from $\mathbb{Z}/m\mathbb{Z}$ into $GF(p)^\times$. $\square$

Let (X, μ_0) be a Lebesgue space, T_0: $X \to X$ an invertible ergodic measure preserving transformation and γ: $X \to \mathbb{Z}/m\mathbb{Z}$ a measurable function. We say that

$$\begin{aligned} &T_1\colon X \times \mathbb{Z}/m\mathbb{Z} \to X \times \mathbb{Z}/m\mathbb{Z} \\ &T_1(x, y) = (T_0 x, \gamma(x) + y) \end{aligned} \tag{1.1}$$

is the $\mathbb{Z}/m\mathbb{Z}$ extension of T_0 corresponding to γ. The natural product measure μ_1 on $X \times \mathbb{Z}/m\mathbb{Z}$, which is defined as the normalized product of μ_0 and the uniform (Haar) measure on $\mathbb{Z}/m\mathbb{Z}$, is clearly preserved by T_1.

All of our examples will be of the following type. Given T_1 as in (1.1) and a prime p which satisfies Lemma 1.1 for m, we construct the $GF(p)$ extension of T_1 corresponding to the homomorphism ϕ.

$$\begin{aligned} &T\colon X \times \mathbb{Z}/m\mathbb{Z} \times GF(p) \to X \times \mathbb{Z}/m\mathbb{Z} \times GF(p) \\ &T(x, y, z) = (T_0 x, \gamma(x) + y, \phi(y) + z). \end{aligned} \tag{1.2}$$

Note that the T in (1.2) is specified by a choice of T_0 and γ. The natural topology for the set of pairs (T_0, γ) is the product of the weak topology for measure preserving transformations and the L_1-topology for functions. (Cf. §6 for details).

In terms of this topology we can state our first main result which is proved in §§2–6:

Theorem 1.1. *For a generic set of (T_0, γ), T is ergodic and has continuous spectrum with maximal spectral multiplicity m. Moreover the spectrum consists of two components, one of which is simple and the other with multiplicity exactly m.*

The second main result, which deals with the realization of a given spectral multiplicity by an interval exchange, is Theorem 7.1 which is formulated and proved in §7.

§2. The Estimate of the Multiplicity from Below

Associated with a finite abelian group extension is a natural orthogonal decomposition of L_2 into U_T invariant subspaces corresponding to the characters of the group. The additive characters of $GF(p)$ are given by $\chi_w(z) = \exp 2\pi i z w/p$ where $w \in GF(p)$, so that if T is given by (1.2), we obtain the invariant decomposition

$$L_2(X \times \mathbb{Z}/m\mathbb{Z} \times GF(p)) = \bigoplus_{w \in GF(p)} H_w$$

where

$$H_w = \{X_w(z)\, f(x, y) \colon f \in L_2(X \times \mathbb{Z}/m\mathbb{Z})\}.$$

The decomposition is obtained by a discrete Fourier transform with respect to the third variable. Let us define the permutation σ: $GF(p) \to GF(p)$ by $\sigma(w) = \phi(1)\, w$ and for $w \neq 0$, the operator

$$S_w\colon H_w \to H_{\sigma(w)}$$

by

$$S_w(\chi_w(z)\, f(x, y)) = \chi_{\sigma(w)}(z)\, f(x, y+1).$$

Lemma 2.1.

$$U_T|_{H_{\sigma(w)}} \circ S_w = S_w \circ U_T|_{H_w}$$

Proof.

$$U_T|_{H_{\sigma(w)}} \circ S_w(\chi_w(z) f(x, y)) = \chi_{\sigma(w)}(z)\, \chi_{\sigma(w)}(\phi(y))\, f(T_0 x, \gamma(x)+y+1),$$
$$S_w \circ U_T|_{H_w}(\chi_w(z) f(x, y)) = \chi_{\sigma(w)}(z)\, \chi_w(\phi(y+1))\, f(T_0 x, \gamma(x)+y+1).$$

Equality follows from the observation that

$$\chi_{\sigma(w)}(\phi(y)) = \exp 2\pi i \phi(1)\, w \phi(y)/p = \chi_w(\phi(y+1)). \quad \square$$

The following lemma characterizes the action of the permutation σ.

Lemma 2.2. *σ has a fixed point 0 and $m' = (p-1)/m$ cycles of length m. Furthermore, each cycle is represented by exactly one element of the set $\phi_0(\{0, 1, \ldots, m' - 1\})$.*

Proof. Clearly $\sigma(0)=0$. If $z \neq 0$ then $z = \phi_0(y)$ for some $y \in \mathbb{Z}/(p-1)\mathbb{Z}$ and $\sigma(z) = \phi(1) z = \phi(1)\, \phi_0(y) = \phi_0(\psi(1))\, \phi_0(y) = (\phi_0(\psi(1)+y) = \phi_0(m'+y)$. It follows that σ on $GF(p)^\times$ is conjugate via ϕ_0 to the permutation $y \mapsto m'+y$ on $\mathbb{Z}/(p-1)\mathbb{Z}$. This permutation has m' cycles of length m represented by $\{0, 1, \ldots, m'-1\}$. $\square$

The previous lemma shows how the operators S_w permute the subspaces H_w. For each $j = 0, \ldots, m-1$ we define

$$H^j = H_{\sigma^j(\phi_0(0))} \oplus \cdots \oplus H_{\sigma^j(\phi_0(m'-1))} \tag{2.1}$$

and also

$$H^* = H_0.$$

Note that by Lemma 2.2 for each $w \neq 0$ $H_w \subseteq H^j$ for some j. We define the linear operator

$$S^j\colon H^j \to H^{j+1} \tag{2.2}$$

so that if $H_w \subseteq H^j$ then $S^j|_{H_w} = S_w$. Since $S^j|_{H_w} H_w = H_{\sigma(w)} \subseteq H^{j+1}$, the operator S^j is well defined.

Lemma 2.3.

(i) $S^j \circ U_T|_{H^j} = U_T|_{H^{j+1}} \circ S^j$

(ii) *The maximal spectral multiplicity for T in* (1.1) *is at least m.*

Proof. (i) is trivial by Lemmas 2.1 and 2.2. For (ii) we note that U_T is isomorphic on each H^j. Thus each H^j must have the same spectrum $j = 0, \ldots, m-1$. $\square$

§ 3. Approximation by Periodic Transformations

Let (X, μ) be a Lebesgue space and let $\{\xi_n\}$ be a sequence of partitions of X into sets $C_{n,j}, j = 1, \ldots, q_n$. We say that $\xi_n \to \xi$ if for any measurable $A \subseteq X$ there exists $\xi_n' \subseteq \xi_n$ and $A_n = \bigcup_{C_n \in \xi_n'} C_n$ such that $\mu(A_n \Delta A) \to 0$ as $n \to \infty$. Let $f(n) \to 0$ as $n \to \infty$. We say that the measure preserving transformation $T\colon X \to X$ admits an

m-cyclic approximation by periodic transformations with speed $f(n)$, (a.p.t. speed $f(n)$), if there exists $\xi_n \to \varepsilon$ and for every n a measure preserving permutation σ_n of the collection $\{C_{n,j}\}\ j=1, \ldots, q_n$ such that:

(i) $\sum_{j=1}^{q_n} \mu(T C_{n,j} \Delta \sigma_n C_{n,j}) < f(p_n)$,

(ii) σ_n has m cycles

where p_n is the length of the longest cycle in σ_n. In the case where $f(n)=o(1/n)$ as $n \to \infty$ we simply say the speed is $o(1/n)$.

Katok and Stepin [9] proved that if T admits a cyclic (1-cyclic) a.p.t. speed $o(1/n)$ then T is ergodic and has simple spectrum. The following result is also due to Katok and Stepin, (unpublished).

Theorem 3.1. *If T admits an m-cyclic a.p.t. with speed $o(1/n)$ then the multiplicity of the spectrum of U_T is at most m.*

The proof follows directly from a theorem proved in [3].

§4. Combinatorics of Extensions

The next few lemmas discuss the behavior of the extension of an approximation. Let $T_0\colon \mathbb{Z}/q\mathbb{Z} \to \mathbb{Z}/q\mathbb{Z}$ be a cyclic permutation. Any function $\gamma\colon \mathbb{Z}/q\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ may be written as $\gamma = \sum_{y \in \mathbb{Z}/m\mathbb{Z}} y \chi_{A_y}$ where $A_y = \gamma^{-1}(y)$. Letting $a_y = \operatorname{card}(A_y)$ we will define

$$t(\gamma) = \sum_{y \in \mathbb{Z}/m\mathbb{Z}} y a_y.$$

It is easy to see that $t(\gamma_1+\gamma_2)=t(\gamma_1)+t(\gamma_2)$. We will often write t for $t(\gamma)$.

Consider the extension

$$T_1\colon \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \tag{4.1}$$

given by

$$T_1(x,y) = (T_0 x, \gamma(x)+y).$$

Lemma 4.1. *T_1 has m cycles of length q if and only if $t \equiv 0 (\operatorname{mod} m)$. T_1 has one cycle of length mq if and only if $(t,m)=1$.*

Proof. $T_1^k(x,y) = (T_0^k x, y + \sum_{j=1}^{k-1} \gamma(T_0^j x))$ so that in particular $T_1^q(x,y) = (x, y + \sum_{j=1}^{q-1} \gamma(T_0^j x))$, where q is the least k so that $T_0^k x = x$. Thus it suffices to show that $t(\gamma) = \sum_{j=1}^{q-1} \gamma(T_0^j x)$. This is clear since $\gamma(T_0^j x) = y$ if and only if $T_0^j x \in A_y$. Because T_0 is a cycle this happens a_y times. $\square$

If T_1 does not have the behavior described in the last lemma, we can modify γ to obtain it. For the next lemma let

$$\Delta\gamma(x) = \begin{cases} k & \text{if } x = x_0 \\ 0 & \text{if } x \neq x_0 \end{cases} \quad \text{and} \quad \tilde{T}_1(x,y) = (T_0 x, (\gamma+\Delta\gamma)(x)+y).$$

Lemma 4.2. *For arbitrary γ and x_0 there are values of k so that $\tilde{T}_1$ has one cycle and so that it has m cycles.*

Proof. If we let $\tilde{a}_y = \operatorname{card}((\gamma+\Delta\gamma)^{-1}(y))$ and $\tilde{t} = \sum_{y=0}^{m} y\tilde{a}_y$ then $\tilde{t}-t = (\gamma+\Delta\gamma)(x_0) - \gamma(x_0) = k$. □

Let us now consider the double extension

$$T\colon \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \times GF(p) \to \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z} \times GF(p) \tag{4.2}$$

where

$$T(x, y, z) = (T_0 x, \gamma(x) + y, \phi(y) + z).$$

Proposition 4.1. *Suppose T_0: $\mathbb{Z}/q\mathbb{Z} \to \mathbb{Z}/q\mathbb{Z}$ is a cyclic permutation and γ is such that the extension T_1 given by* (4.1) *has m cycles of length q. Let p and ϕ be chosen according to Lemma* 1.1. *Then either the extension T given by* (4.2) *has m cycles of length pq, or for any $\Delta\gamma$ of the form*

$$\Delta\gamma = \chi_{\{x_0\}} - \chi_{\{T_0 x_0\}} \tag{4.3}$$

the extension

$$\tilde{T}(x, y, z) = (T_0 x, (\gamma+\Delta\gamma)(x) + y, \phi(y) + z)$$

has m cycles of length pq.

Proof. We adopt the following notation for the cycles of T_1

$$C_l = \bigcup_{j=0}^{q-1} T_1^j(0, l).$$

Since the extension of T_1 to T respects the decomposition into cycles we may consider the m extensions of $T_1|_{C_l}$ to $T|_{C_l x GF(p)}$ separately. To show that T consists of m cycles of length pq it suffices to show that for each l, $T|_{C_l x GF(p)}$ consists of one cycle of length pq. To formulate an equivalent statement we define $a_l(z) = \operatorname{card}(C_l \cap \phi^{-1}(z))$ and $t_l = \sum_{z=0}^{p-1} z a_l(z)$. It follows from Lemma 4.1 that it suffices to show that $(t_l, p) = 1$ for all l.

We first show that if this condition is satisfied for one l then it is satisfied for all l simultaneously. Observe that $(x, y) \in C_l$ if and only if $(x, y+k) \in C_{l+k}$ and thus

$$a_l(z) = a_{l+k}(\phi(k)\, z). \tag{4.4}$$

Making the substitutions $z = \phi(y)$ and $b_l(y) = a_l(\phi(y))$, we have by (4.4)

$$\begin{aligned} b_{l+k}(y) &= a_{l+k}(\phi(y)) \\ &= a_{l+k}(\phi(k)\, \phi(y-k)) \\ &= a_l(\phi(y-k)) \\ &= b_l(y-k). \end{aligned}$$

Since

$$t_l = \sum_{z \in GF(p)} z a_l(z) = \sum_{y \in \mathbb{Z}/m\mathbb{Z}} \phi(y)\, b_l(y)$$

and

$$t_{l+k}=\sum_{y\in\mathbb{Z}/m\mathbb{Z}}\phi(y)\,b_{l+k}(y)=\sum_{y\in\mathbb{Z}/m\mathbb{Z}}\phi(y+k)\,b_l(y)$$

it follows that

$$\begin{aligned} t_{l+k}-t_l&=\sum_{y\in\mathbb{Z}/m\mathbb{Z}}(\phi(y+k)-\phi(y))\,b_l(y)\\ &=(\phi(k)-1)\sum_{y\in\mathbb{Z}/m\mathbb{Z}}\phi(y)\,b_l(y)\\ &=(\phi(k)-1)\,t_l. \end{aligned}$$

We see that $t_{l+k}=\phi(k)\,t_l=\sigma^k t_l$ where σ is the permutation in Lemma 2.2. Since all the t_l are in the same orbit of σ either they are all zero or all nonzero.

We now assume that $t_l\equiv 0 \pmod p$ and consider the perturbation $\tilde{T}$ described in the statement, in particular we take $\Delta\gamma$ as in (4.3) and assume without loss of generality that $x_0\neq q-1$. Let $\tilde{T}_1(x,y)=T_0x,(\gamma+\Delta\gamma)(x)+y)$. It is easy to see that T_1 and $\tilde{T}_1$ are conjugate via the isomorphism

$$R(x,y)=(x,y+\chi_{\{x_0\}}(x))$$

so that these two maps have the same cyclic structure.

Let $\tilde{C}_l=\bigcup_{j=0}^{q-1}\tilde{T}_1^j(0,l)$ and $\tilde{a}_l(z)=\operatorname{card}(\tilde{C}_l\cap\phi^{-1}(z))$. Suppose $x_0=T_0^k(0)$ and $(x_0,y_0)=\tilde{T}_1^k(0,l)$. By the above we also have $(x_0,y_0)=T_1^k(0,l)$ and so $y_0=l+\sum_{j=0}^{k-1}\gamma(T_0^j0)$. Then

$$\tilde{a}_l(z)-a_l(z)=\begin{cases} 1 & \text{if } z=\phi(y_0+(\gamma+\Delta\gamma)(x_0))\\ -1 & \text{if } z=\phi(y_0+\gamma(x_0))\\ 0 & \text{otherwise.}\end{cases}$$

It follows that

$$\begin{aligned} \tilde{t}_l-t_l&=\sum_{\mathbb{Z}/p\mathbb{Z}}z(\tilde{a}_l(z)-a_l(z))\\ &=\phi(y_0+(\gamma+\Delta\gamma)(x_0))-\gamma(y_0+\gamma(x_0))\\ &=\phi(y_0+\gamma(x_0))\,(\phi(\Delta\gamma(x_0))-1). \end{aligned}$$

But $\phi(y)\neq 0$ and $\phi(\Delta\gamma(x_0))=\phi(1)\neq 1$ since ϕ is one to one. It follows that $t_l'\neq 0$. □

§ 5. The Estimate of the Multiplicity from Above

Let L be the set of measurable functions $\gamma\colon X\to\mathbb{Z}/m\mathbb{Z}$. To describe the topology on L we identify $\mathbb{Z}/m\mathbb{Z}$ with the set $\{0/m, 1/m, \ldots, (m-1)/m\}\subseteq[0,1)$. Let us define

$$\|\gamma_1-\gamma_2\|=\int_X|\gamma_1(x)-\gamma_2(x)|\,dx.$$

For a finite partition ξ let us define $L(\xi)=\{\gamma\in L\colon \gamma$ is constant a.e. on each $C\in\xi\}$. Let $\xi'=\xi\times\varepsilon_{\mathbb{Z}/m\mathbb{Z}}$ and $\xi''=\xi'\times\varepsilon_{GF(p)}$ be the natural extensions of the partition ξ to $X \times \mathbb{Z}/m\mathbb{Z}$ and $X \times \mathbb{Z}/m\mathbb{Z} \times GF(p)$ respectively. If $T_0: X \to X$ permutes the elements of ξ and $\gamma \in L(\xi)$ then the extension T_1 in (1.1) permutes the elements ξ', and the double extension T in (1.2) permutes the elements of ξ''.

Lemma 5.1. *Suppose T_0: $X \to X$ is a measure preserving transformation which admits an a.p.t. $T_{0,n}$ with speed $f(n)$ such that $\xi_n \to \varepsilon$ is the sequence of partitions permuted by $T_{0,n}$. Suppose $\gamma \in L$, $\gamma_n \in L(\xi_n)$ and $\|\gamma_n - \gamma\| < g(n)$. Then the extension T_1 in* (1.1) *corresponding to γ has an a.p.t. $T_{1,n}$ with speed $f(n)+g(n)$, where $T_{1,n}$ is given by*

$$T_{1,n}(x, y) = (T_{0,n} x, \gamma_n(x) + y). \tag{5.1}$$

In addition, the double extension T in (1.2) has an a.p.t. $T_{2,n}$ with speed $f(n)+g(n)$, where $T_{2,n}$ is given by

$$T_{2,n}(x, y, z) = (T_{0,n} x, \gamma_n(x) + y, \phi(y) + z). \tag{5.2}$$

Proof. Let us define

$$T'_{1,n}(x, y) = (T_{0,n} x, \gamma(x) + y)$$

and let μ' be the normalized product measure on $X \times \mathbb{Z}/m\mathbb{Z}$. Then

$$\begin{aligned}
&\sum_{C \in \xi'_n} \mu'(T_1 C \mathbin{\Delta} T_{1,n} C) \\
&\qquad \leqq \sum_{C \in \xi'_n} \mu'(T_1 C \mathbin{\Delta} T'_{1,n} C) + \sum_{C \in \xi'_n} \mu'(T'_{1,n} C \mathbin{\Delta} T_{1,n} C) \\
&\qquad \leqq \sum_{C \in \xi_n} \mu(T_0 C \mathbin{\Delta} T_{0,n} C) + \|\gamma_n - \gamma\| \\
&\qquad \leqq f(n) + g(n).
\end{aligned}$$

The second statement follows from the first. □

Let T_0 be a measure preserving transformation which admits a cyclic a.p.t $T_{0,n}$ with speed $o(1/n)$.

Definition 5.1. We will say that γ_n *is of type 1* if the extension $T_{1,n}$ of $T_{0,n}$ corresponding to γ_n is cyclic. We will say that γ_n *is of type 2* if both $T_{1,n}$ and $T_{2,n}$ have m cycles.

Definition 5.2. Let us define $\Gamma(T_0)$ to be the set of all $\gamma \in L$ such that there exists $\gamma_n \to \gamma$ where γ_n is of type 1 when n is even, γ_n is of type 2 when n is odd and $\|\gamma_n - \gamma\| = o(1/n)$.

Definition 5.3. We define $\mathscr{W}$ to be the set of all pairs (T_0, γ) such that:

(i) T_0 has continuous spectrum and admits a cyclic a.p.t T_0, n with speed $o(1/n)$ such that $(q_n, m) = 1$, where $q_n = \operatorname{card}(\xi_n)$ and

(ii) $\gamma \in \Gamma(T_0)$.

The following is a general lemma on the continuity of the spectrum of a finite cyclic group extension:

Lemma 5.2. *Suppose T_0 and γ are chosen so that $(T_0, \gamma) \in \mathscr{W}$. Then the extension T_1 of T_0 corresponding to γ, given by* (1.1), *has continuous spectrum.*

The proof depends on the following lemma.

Lemma 5.3. *Suppose $(T_0, \gamma)\in\mathscr{W}$. For $k\in\mathbb{Z}/m\mathbb{Z}$ let $\tilde{\gamma}=k\gamma-1$ and let $\tilde{T}_1$ be the $\mathbb{Z}/m\mathbb{Z}$ extension of T_0 corresponding to $\tilde{\gamma}$. Then $\tilde{T}_1$ is ergodic.*

Proof. Since $\gamma\in\Gamma(T_0)$ there exists a sequence of type 2 γ_n such that $\|c_n-\gamma\| =o(1/n)$. If we define $\tilde{\gamma}_n=k\gamma_n-1$, then by Lemma 4.1

$$t(\tilde{\gamma}_n)=kt(\gamma_n)-t(1)=-q_n.$$

We will write $\tilde{T}_{1,n}$ for the $\mathbb{Z}/m\mathbb{Z}$ extension of $T_{0,n}$ corresponding to $\tilde{\gamma}_n$. Since $(q_n, m)=1$, $\tilde{\gamma}_n$ is type 1 and so $\tilde{T}_{1,n}$ is a cyclic a.p.t speed $o(1/n)$ for $\tilde{T}_1$. It follows from Theorem 3.1 that $\tilde{T}_1$ is ergodic. □

Proof of Lemma 5.2. We consider the U_{T_1} invariant orthogonal decomposition

$$L_2(X\times\mathbb{Z}/m\mathbb{Z})=\bigoplus_{k\in\mathbb{Z}/m\mathbb{Z}} H_k$$

where

$$H_k=\{\chi_k(y)\,f(x)\colon f\in L_2(X)\}.$$

Each eigenvalue ζ corresponds to an eigenfunction of the form $\chi_k(y)\,f(x)\in H_k$ for some k. Since

$$\zeta\chi_k(y)\,f(x)=U_{T_1}\chi_k(y)\,f(x)=\chi_k(\gamma(x)+y)\,f(T_0x)$$

and since $\chi_k^m\equiv 1$,

$$f^m(T_0x)=\zeta^m f^m(x).$$

The continuity of the spectrum of T_0 implies $\zeta^m=1$.

We next establish the ergodicity of T_1. Since $\gamma\in\Gamma(T_0)$ there exists a sequence of type 1 γ_n such that $\|\gamma_n-\gamma\|=o(1/n)$. By Lemma 5.1 T_1 admits a cyclic a.p.t. speed $o(1/n)$ and it follows from Theorem 3.1 that T_1 is ergodic.

Ergodicity implies that the eigenvalues ζ_j of U_{T_1} form a subgroup of the m'th roots of unity and thus

$$\zeta_j=\exp 2\pi ijl_1/m,\qquad j=0,\ldots,m_1-1$$

where $m_1|m$ and $l_1=m/m_1$. Since each eigenvalue ζ_j is simple, the associated eigenfunction $F_j\in H_k$ for some $k\in\mathbb{Z}/m\mathbb{Z}$. The relation between j and k is that $j=\psi(k)$ for some function ψ: $\mathbb{Z}/m_1\mathbb{Z}\to\mathbb{Z}/m\mathbb{Z}$. We may assume without loss of generality [7] that $F_jF_{j'}=F_{j+j'}$, from which it follows that $\psi(j+j')=\psi(j)+\psi(j')$. In fact, ψ is one to one since otherwise there are two eigenvalues ζ and ζ' corresponding to different eigenfunctions f and f' in H_0, contradicting the continuity of the spectrum of T_0. Thus $\psi(j)=kl_1j$ for some $k\in\mathbb{Z}/m\mathbb{Z}^\times$. It follows that

$$\chi_{\psi(j)}(\gamma(x))\,f(T_0x)=\zeta_i\,f(x)$$

and so

$$\chi_{l_1j}(k\gamma(x)-1)\,f(T_0x)=f(x). \tag{5.3}$$

We note however that (5.3) implies that the function $\chi_{l_1j}(y)\,f(x)$ is an invariant function for the transformation $\tilde{T}_1$ constructed in Lemma 5.2. This contradicts the ergodicity of $\tilde{T}_1$. □

Proposition 5.1. *Let $(T_0, \gamma)\in\mathscr{W}$. Then the double extension T in (1.2) has continuous spectrum with maximal spectral multiplicity m. Moreover the spectrum in $\bigoplus_{\omega\in GF(p)^\times} H_w$ is homogeneous with multiplicity m.*

Proof. Since $\gamma\in\Gamma(t_0)$ there exists $\gamma_n\to\gamma$ such that γ_n is of type 2, and by Lemma 5.1 an m-cyclic a.p.t. speed $o(1/n)$, $T_{2,n}$ of T. By Theorem 2.1 the maximal spectral multiplicity of T is at most m and by Lemma 2.3 it is at least m.

By Lemma 2.3 the spectra in each H^j of the decomposition (2.1) are identical. It follows that the spectra in each H^j are simple, since if not the multiplicity would be at least $2m$ contradicting the upper bound.

Now suppose that f is an eigenfunction for $U_{TH^0\oplus\ldots\oplus H^{m-1}}$. Then at least one projection f_j into H^j is nonzero. Let $f_0, \ldots, f_{m-1}$ be the images under the operators S^k in (2.2) of f_j' in the spaces $H^0, \ldots, H^{m-1}$. Then $f_0, \ldots, f_{m-1}$ are a set of m with the same eigenvalue. If we let $f_j'=f_j/f_0$ then the functions f_j' are a set of invariant functions, which for $j>0$ are not constant.

We now show that for $j>0$, $f_j'\perp H^*$ where H^* is given by (2.1). It is clear that $U_T|_{H^*}$ is equivalent to U_{T_1}. By Lemma 5.2 T_1 is ergodic and thus has no invariant functions besides constants.

For some $j>0$ let $f_0'', \ldots, f_{m-1}''$ be the images of f_j' under the operators S^k with $f_k''\in H^k$. The functions $f_0'', \ldots, f_{m-1}''$ are a set of m non-constant invariant functions. Together with the constants this implies that the multiplicity of 1 in the spectrum is $m+1$ contradicting the established upper bound. It follows that the spectra in each H^j are continuous. This fact combined with the continuity of the spectrum of T_1 implies T has continuous spectrum. It is clear that the spectrum in H^* is disjoint from the spectra in the H^j. Otherwise the multiplicity at some point would be $m+1$ contradicting the upper bound. □

§6. Genericity

To complete the proof of Theorem 1.1 we must show that $(T_0, \gamma)\in\mathscr{W}$ generically. In §5 we defined a topology on the set of $\gamma\in L$. We now recall the definition of the weak topology on the set $\mathscr{U}$ of invertible measure preserving transformations of X. For $T\in\mathscr{U}$ a subbase for the neighborhoods of T is given by sets of the form

$$N(T, \xi, \varepsilon)=\{S\colon \sum_{C\in\xi} \mu(T\,C\,\Delta\,S\,C)<\varepsilon\}$$

where ξ is an arbitrary finite partition and $\varepsilon>0$. We give the set of pairs $(T_0, \gamma)\in\mathscr{U}\times L$ the product topology.

Proposition 6.1. *The set $\mathscr{W}$ has a subset which is an everywhere dense G_δ subset of $\mathscr{U}\times L$ in the product topology.*

Proof. Let $\xi_n\to\varepsilon$ be a fixed sequence of partitions and $f(n)=o(1/n)$ a fixed speed. Define Z_n the set of cyclic measure preserving permutations of the elements of ξ_n. For $\sigma\in Z_n$ we define

$$U_\sigma(\xi_n)=N(\sigma, \xi_n, f(n))$$

and

$$\mathscr{U}_1 = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \bigcup_{\sigma \in Z_n} U_\sigma(\xi_n).$$

Let $\mathscr{U}_2$ be the set of those elements of $\mathscr{U}_1$ which have continuous spectrum. In [5] it is proved that $\mathscr{U}_1$ contains a subset which is everywhere dense and G_δ in $\mathscr{U}$. It follows from the fact that the T_0 with continuous spectra are an everywhere dense G_δ set [3], that $\mathscr{U}_2$ has an everywhere dense G_δ subset in $\mathscr{U}$.

We now make some definitions: Let

$$B(\gamma, f(n)) = \{\gamma' \in L \colon \|\gamma - \gamma'\| < f(n)\}$$

and for $j = 1, 2$

$$L_j(\sigma, \xi_n) = \{\gamma \in L(\xi_n) \colon \gamma \text{ is type } j \text{ for } \sigma\}.$$

We also define

$$V_\sigma(\xi_n) = \bigcup_{\sigma \in L_j(\sigma, \xi_n)} B(\gamma, f(n))$$

where

$$j = \begin{cases} 1 & \text{if } 2 \nmid n \\ 2 & \text{if } 2 \mid n. \end{cases}$$

Let

$$\mathscr{W}_1 = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \bigcup_{\sigma \in Z_n} (U_\sigma(\xi_n) \times V_\sigma(\xi_n))$$

and $\mathscr{W}_2$ be the set of those $(T_0, \gamma) \in \mathscr{W}_1$ such that T_0 has continuous spectrum. If $(T_0, \gamma) \in \mathscr{W}_2$ then $T_0 \in \mathscr{U}_2$ and $\gamma \in \Gamma(T_0)$. Thus $\mathscr{W}_2 \subseteq \mathscr{W}_1 \subseteq \mathscr{W}$. Since $\mathscr{W}_2$ is G_δ it remains to show that $\mathscr{W}_2$ is everywhere dense.

By Lemma 4.2 and Proposition 4.1, for any $\gamma \in L(\xi_n)$ there exists $\gamma' \in L_j(\sigma, \xi_n)$ such that $\|\gamma - \gamma'\| \leqq \dfrac{m+1}{m q_n}$. Since for $\gamma_0 \in L$ there is a $\gamma \in L(\xi_n)$ such that $\|\gamma_0 - \gamma'\| \leqq \dfrac{1}{m q_n}$, we can find $\gamma' \in L_j(\sigma, \xi_n)$ such that $\|\gamma_0 - \gamma'\| \leqq \dfrac{m+2}{m q_n}$, for $j = 1, 2$.

Let $T_{0,n}$ be a cyclic a.p.t. speed $f(n)$ of T_0 on a subsequence ξ_{n_k}. Refine this subsequence further so that

$$\frac{m+2}{m q_{n_{k+1}}} < f(q_{n_k}) \tag{6.1}$$

Let us define

$$\mathscr{V}(T_0) = \bigcap_{N=1}^{\infty} \bigcup_{k=N}^{\infty} V_{T_{0,n_k}}(\xi_{n_k})$$

and

$$\mathscr{W}_3 = \{(T_0, \gamma) \colon T_0 \in \mathscr{U}_2, \gamma \in \mathscr{V}(T_0)\}.$$

Clearly $\mathscr{W}_3 \subseteq \mathscr{W}_2$.

To complete the proof we will show that $\mathscr{V}(T_0)$ is dense in L. Given $\gamma_0 \in L$ and $\varepsilon > 0$ we choose n_1 large enough so that $D_1 \neq \emptyset$ where

$$D_1 = L_1(T_{0,n_1}, \xi_{n_1}) \cap B(\gamma_0, \varepsilon).$$

Given $\gamma_k \in D_k$ let us define D_{k+1} inductively by

$$D_{k+1}=L_{k+1}(T_{0,n_{k+1}},\xi_{n_{k+1}})\cap B(\gamma_k,f(q_{n_k})).$$

It follows from (6.1) that $D_{k+1}\neq\emptyset$. Furthermore $\gamma_k\to\gamma\in\mathscr{V}(T_0)$, and $\|\gamma_0-\gamma\|<\varepsilon$. □

§ 7. Interval Exchange Transformation

In this section we show that our construction may be realized as an interval exchange transformation. Recall that an invertible transformation $T\colon [0,1)\to[0,1)$ is called an interval exchange if it is piecewise continuous, Lebesgue measure preserving and orientation preserving. The simplest non-trivial case is when three intervals are exchanged according to the permutation $(1,2,3)\to(3,2,1)$. Katok and Stepin [5] have shown that for almost every pair (α,β) with $\alpha<\beta$ the transitive three interval exchange with discontinuities α and β has simple continuous spectrum.

The proof relies on the following fact. $T^{(\alpha,\beta)}$ is equivalent to the mapping induced on the interval $[0,B)$ by the rotation $T(A)\colon [0,1)\to[0,1)$ where $T^A x=(x+A)\,(\mathrm{mod}\,1)$. The relation between α, β, A and B is given by

$$A=\frac{1-\alpha}{1+\beta-\alpha}\qquad 1-B=\frac{\beta-\alpha}{1+\beta-\alpha}. \tag{7.1}$$

Note that a finite cyclic group extension of an interval exchange is an interval exchange provided γ is piecewise constant. The following theorem is the second main result of this paper.

Theorem 7.1. *For any $m>1$ there exists an interval exchange transformation T which is ergodic and has continuous spectrum of multiplicity m. In fact T has the form* (1.2) *where T_0 an exchange of three intervals and γ is piecewise constant with three points of discontinuity.*

Proof. We choose (α,β) such that $T_0=T^{(\alpha,\beta)}$ has simple continuous spectrum and such that the following conditions hold.

There exists a sequence $q_n\to\infty$ such that for some p_n and r_n

$$|A-P_n/q_n|=o(1/q_n^2)\quad\text{and}\quad |B-r_n/q_n|=o(1/q_n). \tag{7.2}$$

$$(q_n,m)=1, \tag{7.3}$$

$$\alpha>1/2. \tag{7.4}$$

These conditions clearly hold for a set of (α,β) of positive measure. Condition (7.4) is equivalent to the condition $1-A>B/2$.

It suffices to show that there is a piecewise constant $\gamma\in\Gamma(T_0)$. We write $T_0^*=T^{(A)}$ and consider first an extension of T_0^* by a piecewise constant $\gamma\in\Gamma(T_0^*)$ such that $\operatorname{supp}\gamma\subseteq[0,B)$, where α, β, A and B are related as in (7.1). Let $\xi_n^*=\{[k/q_n,k+1/q_n)\colon k=0,\ldots,q_n-1\}$ and $T_{0,n}^*x=(x+p_n/q_n)\,(\mathrm{mod}\,1)$. Let ξ_n be the restriction of ξ_n^* to the interval $[0,r_n/q_n)$ and $T_{0,n}$ the mapping induced by $T_{0,n}^*$

on $[0, r_n/q_n)$. It is clear from (7.2) and (7.3) that $T^*_{0,n}$ and $T_{0,n}$ determine cyclic a.p.t.'s speed $o(1/n)$ of T^*_0 and T_0 respectively.

We will define $\gamma_n \to \gamma$ inductively such that at each step

$$\gamma_n = \chi_{Q_{1,n} \cup Q_{2,n}}$$

where

$$Q_{1,n} = [g_n, h_n)$$

and

$$Q_{2,n} = [h_n + p_n, r_n).$$

Also define γ'_n corresponding to g'_n and h'_n. Choose a subsequence of $q_n \to \infty$ such that $q_{n+1} > q_n^2$. Assume that q_1 is large enough that for some arbitrary $0 < g_1 < h_1$ the inequality

$$0 < g_n < h_n < h_n + p_n < q_n - p_n < r_n < q_n \tag{7.5}$$

holds for $n=1$. This determines γ_1.

We now show how to construct γ'_{n+1} given γ_n. Let us define

$$g''_{n+1} = \min_{j \in \mathbb{N}} |g_n/q_n - j/q_{n+1}|,$$
$$h''_{n+1} = \min_{j \in \mathbb{N}} |h_n/q_n - j/q_{n+1}|.$$

The construction breaks into two cases depending on whether γ_n is of type 1 or type 2. If we let $s_n = r_n - p_n - q_n$ then $t(\gamma_n) = s_n (\operatorname{mod} m)$. By Lemma 4.1, if γ_n is of type 1 then $(s_n, m) = 1$ and if γ_n is of type 2 then $s_n \equiv 0 (\operatorname{mod} m)$. Let us define $s''_{n+1} = r_{n+1} - p_{n+1} - q''_{n+1}$.

Case 1. γ_n is of type 2.

We wish to construct γ'_{n+1} of type 1. If $(s''_{n+1}, m) = 1$ then let $g'_{n+1} = g''_{n+1}$ and $h'_{n+1} = h''_{n+1}$. It follows that γ'_{n+1} is of type 1. If $(s''_{n+1}, m) \neq 1$ then there exists $k < m$ such that $(s''_{n+1} + k, m) = 1$. Let $q'_{n+1} = q''_{n+1} + k$ and $h'_{n+1} = h''_{n+1}$. γ'_{n+1} is of type 1.

Case 2. γ_n is of type 1.

In this case we do not require that γ'_{n+1} be of type 2 but only that $t(\gamma'_{n+1}) = 0$. To do this we repeat the procedure of Case 1, making $g'_{n+1} = g''_{n+1} + k$ for some $k < m$ if necessary to insure that $t(\gamma'_{n+1}) = r_{n+1} - p_{n+1} - q'_{n+1} \equiv 0 (\operatorname{mod} m)$.

We note the following fact. If $T_{1,n+1}(x, y) = (T_{0,n+1} x, \gamma'_{n+1}(x) + y)$ then the cyclic structure of $T_{1,n}$ depends only on the type of γ_n. The corresponding fact for $T^*_{0,n+1}$ is the consequence of the construction of γ'_{n+1}. The statement for $T_{0,n+1}$ follows from the remark that $T_{0,n+1}$ is equivalent to the transformation induced on the interval $[0, r_n + 1/q_{n+1})$ by $T^*_{0,n+1}$, and that $\operatorname{supp}(\gamma'_{n+1}) \subseteq [0, r_{n+1}/q_{n+1})$. Consequently we need only consider extensions of the three interval exchange T_0.

The final step is to show that we can construct γ_{n+1} from γ'_{n+1} such that $\gamma_n \to \gamma$ where $\gamma \in \Gamma(T_0)$. In the case where γ_n is type 2, γ'_{n+1} is type 1 and so we define $\gamma_{n+1} = \gamma'_{n+1}$ making $g_{n+1} = g'_{n+1}$ and $h_{n+1} = h'_{n+1}$. In the case where γ_n is a type

1 we must insure that γ_{n+1} is of type 2. We consider the double extension

$$T_{2,n+1}(x,y,z)=(T_{0,n+1},\gamma'_{n+1}(x)+y,\phi(y)+z).$$

If $T_{2,n+1}$ has m cycles then we again define $\gamma_{n+1}=\gamma'_{n+1}$. If not then we apply Proposition 4.1 with

$$\Delta\gamma_{n+1}=\chi_{R^1_{n+1}}-\chi_{R^{-1}_{n+1}}$$

where

$$R^1_{n+1}=[h'_{n+1}/q_{n+1},(h'_{n+1}+1)/q_{n+1}),$$
$$R^{-1}_{n+1}=[(h'_{n+1}+p_{n+1})/q_{n+1},(h'_{n+1}+p_{n+1}+1)/q_{n+1}).$$

It follows from condition (7.4) that $\Delta\gamma_{n+1}$ is of the form (4.3). Thus we define $\gamma_{n+1}=\gamma'_{n+1}+\Delta\gamma_{n+1}$, making $g_{n+1}=g'_{n+1}$ and $h_{n+1}=h'_{n+1}+1$.

Inequality (7.5) follows from the inequality $q_{n+1}>q_n^2$. In addition we have the following

$$|g_{n+1}/q_{n+1}-g_n/q_n|<m/q_{n+1},$$
$$|h_{n+1}/q_{n+1}-h_n/q_n|<2/q_{n+1},$$
$$|p_{n+1}/q_{n+1}-p_n/q_n|=o(1/q_n^2),$$
$$|r_{n+1}/q_{n+1}-r_n/q_n|=o(1/q_n)$$

and so

$$\|\gamma_{n+1}-\gamma_n\|<\frac{m+2}{q_{n+1}}+o(1/q_n^2)+o(1/q_n)=o(1/q_n).$$

It follows that

$$\|\gamma_n-\gamma\|=o(1/q_n).$$

Letting $G=\lim_{n\to\infty} g_n$ and $H=\lim_{n\to\infty} h_n$ we have

$$\gamma=\chi_{[G,H)\cup[H+A,B)}.$$

Thus $\gamma\in\Gamma(T_0)$ and γ is piecewise constant. □

References

1. Baxter, J.R.: A class of ergodic transformations having simple spectrum. Proc. Amer. Math. Soc. **27**, 275–279 (1971)
2. Chacon, R.V.: Approximation of transformations with continuous spectrum. Pacific J. Math. **2**, 293–302 (1969)
3. Chacon, R.V.: Approximation and spectral multiplicity. Lecture Notes in Math. Vol. 160, pp. 18–22. Berlin-Heidelberg-New York: Springer 1970
4. Girsanov, I.V.: Spectra of dynamical systems generated by stationary Gaussian processes, (in Russian) Dokl. Akad. Nauk SSSR **119**, 851–853 (1958)
5. Halmos, P.R., von Neumann, J.: Operator methods in classical mechanics II. Ann. Math. **43**, 332–350 (1942)
6. Halmos, P.R.: On automorphisms of compact groups. Bull. AMS, **49**, 619–624 (1943)
7. Halmos, P.R.: Lectures on ergodic theory. The Math. Society of Japan, Tokyo 1956
8. Del Junco, A.: A transformation with simple spectrum which is not rank one. Can J. Math. **29**, 655–663 (1977)

9. Katok, A.B., Stepin, A.M.: Approximation of ergodic dynamic systems by periodic transformations. Dokl. Akad. Nauk SSSR **171**, 1268–1271 (1966); Soviet Math. Dokl. **7**, 1638–1641 (1966)
10. Katok, A.B., Stepin, A.M.: Approximations in ergodic theory. Uspekhi Mat. Nauk **22**, (1967); Russian Math. Surveys **22**, 77–102 (1967)
11. Oseledec, V.I.: The spectrum of ergodic automorphisms. Dokl. Akad. Nauk SSSR **168**, 776–779 (1966)
12. Rohlin, V.A.: Entropy of a metric automorphism (in Russian). Dokl. Akad. Nauk. SSSR **124**, 980–983 (1959)
13. Rohlin, V.A.: New progress in the theory of transformations with invariant measure. Uspekhi Mat. Nauk **15**, 2–36 (1960); Russian Math. Surveys **15**, 1–22 (1960)
14. Rohlin, V.A.: Lectures on the entropy theory of measure preserving transformations. Uspekhi Mat. Nauk **22**, (1967; Russian Math. Surveys **22**, (1967)
15. Schwartzbauer, T.: A general method for approximating measure preserving transformations. Proc. Amer. Math. Soc. **24**, 643–648 (1970)
16. Veršik, A.M.: On the theory of normal dynamic systems. Dokl. Akad. Nauk SSSR **114**, 9–12 (1962); Soviet Math. Dokl. **3**, 625–628 (1962)
17. Veršik, A.M.: Spectral and metric isomorphism of some normal dynamical systems. Dokl. Akad. Nauk SSSR **114**, 255–257 (1962); Soviet Math. Dokl. **3**, 693–696 (1962)
18. Von Neumann, J.: Zur Operatorenmethode in der Klassischen Mechanik. Ann. Math. **33**, 587–642 (1932)
19. Yuzvinskii, S.A.: Metric automorphisms with a simple spectrum. Dokl. Akad. Nauk SSSR **172**, 1036–1038 (1967); Soviet Math. Dokl. **8**, 243–245 (1967)

Oblatum 27-VIII-1982

Invent. math. 72, 315–322

Inventiones mathematicae

A Cohomological Characterization of $\mathbb{P}^n$

J.M. Wahl*

Department of Mathematics, University of North Carolina, Chapel Hill, NC 27514, USA

Introduction

Our main goal is the proof of the following

Theorem 1. *Let X be a complex projective non-singular variety, L an ample line bundle, Θ_X the tangent bundle. If $H^0(\Theta_X \otimes L^{-1}) \neq 0$, then $(X, L) = (\mathbb{P}^n, \mathcal{O}(1))$ or $(\mathbb{P}^1, \mathcal{O}(2))$.*

This had been conjectured in [6], p. 354, where it was verified for curves and surfaces (using classification). A substantial improvement was made by Mori and Sumihiro [3], who proved the Theorem assuming $L = \mathcal{O}(E)$, with E effective; in this case, $H^0(\Theta_X(-E)) \neq 0$ implies there is a one-parameter group of automorphisms of X which fixes E pointwise. On the other hand, an effectivity assumption is irrelevant for the applications we had in mind (see below).

In fact, we prove a stronger result than Theorem 1:

Theorem 2. *Let X be a complex projective normal variety, L an ample line bundle, with* $\dim X > 1$. *If $H^0(\Theta_X \otimes L^{-1}) \neq 0$, then*

a) *$L \simeq \mathcal{O}(E)$, where the effective divisor E is a normal variety*

b) *$X \simeq \operatorname{Proj} A[t]$, where $A = \bigoplus_{n=0}^{\infty} H^0(E, \mathcal{O}_E(nE))$ is the cone over E, t has weight* 1, *and E is the divisor at $\infty (t=0)$.*

In b), X may be viewed as follows: form $\mathbb{P}(\mathcal{O}_E \oplus \mathcal{O}_E(-E)) \to E$, a $\mathbb{P}^1$-bundle. Then X is the blow-down of the section with normal bundle $\mathcal{O}(-E)$. For example, let $\mathbb{P}^3$ have coordinates x, y, z, t, and let $X = V(xy - z^2)$; if $E = X \cap V(t)$ (a non-singular plane quadric), then $t\dfrac{\partial}{\partial t} \in H^0(\Theta_X(-E))$. Here, X is the ruled surface F_2 blown down.

Note that Theorem 2 implies Theorem 1. For, X non-singular implies $\operatorname{Spec} A$ (an open subset of $\operatorname{Proj} A[t]$, as weight $t = 1$) is non-singular, hence A is

* Supported partially by the National Science Foundation, Grant #MCS-8100750

a graded polynomial ring, as is $B=A[t]$. But $\operatorname{Proj} B$ non-singular implies $\operatorname{Proj} B \simeq \mathbb{P}^n$ (e.g., [2], 1.3.3); that is, $\mathbb{P}^n$ is the only non-singular weighted projective space. It is easy to see that if $X=\mathbb{P}^n$, then $L=\mathcal{O}(1)$ (except in dimension 1, where the entire result is a triviality via Riemann-Roch).

Here are the main ideas of the proof of Theorem 2. Let $R=\bigoplus_0^\infty \Gamma(X, L^{\otimes i}) =\bigoplus_0^\infty R_i$. First, the elements of $H^0(\Theta_X \otimes L^{-1})$ may be viewed as derivations of weight -1 of R, i.e., which map R_n to R_{n-1} (Corollary 1.5 below - but this is essentially proved in [6], §3). Second, if $R_1 \neq 0$ (i.e., $H^0(L) \neq 0$), and $D \neq 0$ is a derivation of weight -1, then there is a $t \in R_1$ with $Dt=1$ (this is essentially [3], Lemma 12); by Zariski's lemma, $R=A[t]$, for some subring A. In fact, it is necessary to prove a stronger version of this result (where the graded ring R is only reduced - see Lemma 2.7 below). Third, for general R (posibly with $R_1 =0$), we use a graded version of the "branched covering trick" in algebraic geometry; this suggests that if $H^0(L^i)=0$, $1 \leqq i<d$, and $L^d \simeq \mathcal{O}(D)$, take a d-fold cyclic covering branched along D. This yields the desired graded ring result (Theorem 2.2 below). The verification of Theorem 2 then follows easily.

Our original motivation for the conjecture was our work with D. Burns [1] on the contribution that rational -2 curves make to the deformations of a smooth projective surface. If Y is a projective non-singular variety, X a non-singular divisor with ample conormal bundle L, there is a natural map $H_X^1(\Theta_Y) \rightarrow H^1(\Theta_Y)$. If $(X, L)=(\mathbb{P}^1, \mathcal{O}(2))$, then the map is injective, and the first space has dimension 1. Theorem 1 above implies $H_X^1(\Theta_Y)=0$ in every other case, except $(X, L)=(\mathbb{P}^n, \mathcal{O}(1))$, in which case the map above is 0([6], p. 354).

Another motivation is the following: if X is projective, L ample, then $\operatorname{Spec} R=\operatorname{Spec} \oplus \Gamma(X, L^{\otimes n})$ has an isolated normal singularity, plus a natural resolution $Y=V(L^{-1}) \rightarrow \operatorname{Spec} R$ (where $V(L^{-1}) \rightarrow X$ is the geometric line bundle). According to Hironaka, $\operatorname{Spec} R$ has an equivariant resolution $Y' \rightarrow \operatorname{Spec} R$ (i.e., one for which $H^0(\Theta_{Y'})=\Theta_R$). Theorem 1 implies that the resolution $Y \rightarrow \operatorname{Spec} R$ is equivariant. For, $H_X^1(\Theta_Y)=0$ as above ($X \subset Y$ is the zero-section), so $H^0(\Theta_Y)=H^0(Y-X, \Theta_Y)=\Theta_R$ (cf. [5], p. 21).

Finally, we have been interested in derivations of negative weight of graded normal singularities [7]. The moral is that there shouldn't be any, if the grading is carefully chosen (we proved this in [7] for dimension 2). Theorem 1 gives this result for cones (i.e., $R=\oplus \Gamma(X, L^{\otimes n})$), since $H^0(\Theta_X \otimes L^{-i})=0$, $i>0$, if R is not regular.

The reader will notice that all our arguments are valid over any algebraically closed field of characteristic zero; we have stuck to $\mathbb{C}$ for convenience. On the other hand, Theorem 1 is definitely false in characteristic 2: consider the smooth hypersurface $X=\left\{x_0^2=\sum_1^n x_i x_{n+i}\right\}$ in $\mathbb{P}^{2n}$, and the derivation "$\partial/\partial x_0$" in $H^0(\Theta_X(-1))$. However, we know no counterexamples in characteristic $\neq 2$.

Our current work was stimulated by reading the aforementioned paper of Mori and Sumihiro.

§1. Cones, Graded Rings, and Derivations

(1.1) Let X be a complete scheme $/\mathbb{C}$ (hence of finite type), L an ample line bundle on X. Let $R=\bigoplus_{i=0}^{\infty}\Gamma(X, L^{\otimes i})=\bigoplus_{i=0}^{\infty}R_i$ be the "cone of (X, L)". The following facts are well-known or easy to verify:

(1.1.1) R is finitely generated/$\mathbb{C}$, R_0 is an artinian $\mathbb{C}$-algebra

(1.1.2) $X\simeq \operatorname{Proj} R$

(1.1.3) If X is a variety (=integral scheme), then for all $n\gg 0$,

$$\sqrt{R_n R}=m_R(=\bigoplus_{i>0}R_i)$$

(1.1.4) X is a normal variety iff R is a normal domain.

(1.1.5) The first chern class $c_1(L)\in H^1(\Omega_X^1)$ gives a short exact sequence

$$0\to\Omega_X^1\to F\to\mathcal{O}_X\to 0$$

(1.2) Let $\pi\colon Y=V(L^{-1})\to X$ be the geometric line bundle associated to L^{-1}; the only section is the 0-section, allowing one to view X as a Cartier divisor on Y. Denoting by Θ the sheaf of $\mathbb{C}$-derivations on a scheme, define ([6], §3)

(1.2.1) $S_Y=\ker(\Theta_Y\to\Theta_Y\otimes\mathcal{O}_X\to N_{X/Y})$.

These "logarithmic derivations of Y along X" send the ideal sheaf of X into itself. If $P\in X$ is a non-singular point of Y, with local analytic coordinates $t, x_1, \ldots, x_n$, and $X=\{t=0\}$, then $S_Y\subset\Theta_Y$ is spanned by $t\dfrac{\partial}{\partial t}, \dfrac{\partial}{\partial x_1}, \ldots, \dfrac{\partial}{\partial x_n}$.

Theorem 1.3 (cf. [6], 3.3). *Let X be a scheme of finite type $/\mathbb{C}$, L a line bundle, $\pi\colon Y=V(L^{-1})\to X$. Then there is an exact sequence*

(1.3.1) $$0\to\mathcal{O}_Y\to S_Y\to\pi^*\Theta_X\to 0,$$

obtained by taking π^* of the dual of (1.1.5) *on X:*

(1.3.2) $$0\to\mathcal{O}_X\to M\to\Theta_X\to 0.$$

Proof. This was proved in [6], 3.3, assuming X is non-singular; essentially the same proof applies in general. Since $\pi\colon Y\to X$ is smooth, there is a natural surjection

$$\Theta_Y\to\pi^*\Theta_X.$$

We claim the induced map $S_Y\to\pi^*\Theta_X$ is surjective, with kernel $\mathcal{O}_Y$. If $\operatorname{Spec} A$ is an affine open in X, $L|\operatorname{Spec} A$ is free (with generator T), then $\pi^{-1}(\operatorname{Spec} A)$ $=\operatorname{Spec} A[T]$. S_Y is then generated by Θ_A and $T\dfrac{\partial}{\partial T}$, as seen by examining (1.2.1). Since the $T\dfrac{\partial}{\partial T}$'s patch to form a global section (same as [6], 3.3), we deduce an exact sequence

$$0 \to \mathcal{O}_Y \to S_Y \to \pi^* \Theta_X \to 0,$$

and it is clear this sequence is obtained as π^* of some sequence on X. Since X is a Cartier divisor on Y, tensoring with $\mathcal{O}_X$ yields an exact sequence

$$0 \to \mathcal{O}_X \to S_Y \otimes \mathcal{O}_X \to \Theta_X \to 0.$$

The verification that this sequence is the dual of the one in (1.1.5) is similar to [6] and is left to the reader (as we shall not use this fact later). We note it is easiest to proceed with $\Omega^1_Y(\log X)$, defined via

$$0 \to \Omega^1_Y(\log X) \to \Omega^1_Y(X) \to \Omega^1_X(X) \to 0,$$

then find

$$0 \to \pi^* \Omega^1_X \to \Omega^1_Y(\log X) \to \mathcal{O}_Y \to 0,$$

and identify the restriction of this sequence to X with the one in (1.1.5).

(1.4) Let us now assume X is a projective variety, $\dim X > 0$. If $U = Y - X$, $\pi: U \to X$ the $\mathbb{C}^*$-bundle, then for any coherent F on X

$$H^0(U, \pi^* F) \simeq \bigoplus_{-\infty}^{\infty} H^0(X, F \otimes L^i).$$

Therefore, $\Gamma(Y, \mathcal{O}_Y) = \Gamma\left(X, \bigoplus_0^\infty L^i\right) \equiv R$ is a graded integral domain of depth $\geqq 2$ at the maximal ideal (since $\Gamma(X, L^{-i}) = 0$, $i > 0$). Also, $N_{X/Y}$ is supported off U, so

$$H^0(U, S_Y \mid U) = H^0(U, \Theta_U).$$

There is an isomorphism of graded modules

$$\Theta_R \simeq H^0(U, \Theta_U)$$

([5], p. 21), by depth considerations. Therefore, writing $\Theta_R = \bigoplus \Theta_R(i)$, Theorem 1.3 implies

Corollary 1.5. *Let X be a projective variety, L an ample line bundle, $R = \bigoplus_{n=0}^{\infty} \Gamma(X, L^{\otimes n})$. Then the dual exact sequence of* (1.1.5),

$$0 \to \mathcal{O}_X \to M \to \Theta_X \to 0,$$

is such that

(1.5.1) $$\Theta_R(i) = H^0(X, M \otimes L^i), \quad i \in \mathbb{Z}.$$

If $\dim X > 1$, *then*

(1.5.2) $$\Theta_R(-i) = H^0(X, \Theta_X \otimes L^{-i}), \quad i > 0.$$

Proof. It remains only to verify (1.5.2). But this follows from (1.5.1) and the exact sequence for M, once we note $H^0(X, L^{-i}) = H^1(X, L^{-i}) = 0$, $i > 0$, by the Kodaira vanishing theorem [4].

Remarks (1.6.1) The graded pieces $\Theta_R(i)$ are easily described as the derivations of R sending R_n into R_{n+i}, for all n; if i is negative, we speak of derivations of negative weight.

(1.6.2) Using the identification of the extension class of (1.1.5) in (1.5.1), it is easy to show that if $\dim X = 1$, then $\Theta_R(-i) = 0$, $i < 0$, unless $(X, L) = (\mathbb{P}^1, \mathcal{O}(1))$. In particular, $(\mathbb{P}^1, \mathcal{O}(2))$ gives no derivation of negative weight, even though $H^0(\Theta_{\mathbb{P}^1}(-2)) \neq 0$; what one does get are "local contributions to global deformations" [1].

§ 2. Derivations of Weight -1 on Graded $\mathbb{C}$-Algebras

(2.1) The goal of this section is the following result.

Theorem 2.2. *Let* $R = \bigoplus_{i=0}^{\infty} R_i$ *be a finitely generated graded domain, with* $R_0 = \mathbb{C}$, *and* $\sqrt{R_N R} = m_R$, *all* $N \gg 0$. *Suppose* $D \neq 0$ *is a derivation of weight* -1. *Then there is an* $x \in R_1$ *such that* $Dx = 1$, *and* $R = A[x]$, *where* $A = \{r \in R \mid Dr = 0\}$ *is a graded subring, and* x *is transcendental over* A.

(2.3) Once we find $x \in R_1$ with $Dx = 1$, the rest follows by the well-known

Proposition 2.4 = Zariski's Lemma ([8], Lemma 4). *Let* $R = \bigoplus_{i=0}^{\infty} R_i$ *be a graded* $\mathbb{C}$-*algebra. Suppose* D *is a derivation of weight* -1, *and* $Dx = 1$ *for some* $x \in R_1$. *Then*

$$R = A[x],$$

where $A = \{r \in R \mid Dr = 0\}$ *is a graded subring, and* x *is transcendental over* A; *further,* $D = \partial/\partial x$ *on* $A[x]$.

Proof. A is certainly a graded $\mathbb{C}$-subalgebra. Consider the ring homomorphism $A[T] \to R$, sending $T \to x$; the map is clearly graded. If $\sum a_i x^i$ is in R, with $a_i \in A$, then $D(\sum a_i x^i) = \sum a_i i x^{i-1}$; so, $A[T] \to R$ is equivariant with respect to the derivations $\partial/\partial T$ and D. We show $A[T] \to R$ is an isomorphism.

Suppose $\sum_{i=0}^{n} a_i T^i$ is a graded element of $A[T]$, in the kernel, with degree $n > 0$ minimal (note $a_i \in R_{n-i}$). Then $\partial/\partial T$ of it, which has smaller degree, is still in the kernel, hence has degree 0. But $A \subset R$, so $\sum a_i T^i = 0$.

We prove surjectivity by induction on R_n, $n = 0$ being clear. If $y \in R_n$, then $Dy \in R_{n-1}$, whence $Dy = \sum_{0}^{n-1} a_i x^i$, $a_i \in R_{n-1-i}$. Therefore, $y - \sum a_i x^{i+1}/(i+1) \in A$, and the result follows.

Remark 2.5. Zariski's lemma actually says if R is a complete local k-algebra, with $\operatorname{char} k = 0$, and D is a k-derivation with $Dx = 1$, then $R = R_1[[x]]$; this is proved by setting $R_1 =$ image of the endomorphism e^{xD} of R.

(2.6) We first prove Theorem 2.2 under somewhat different hypotheses. The key step in the following result is due to Mori and Sumihiro.

Lemma 2.7 (cf. [3], Lemma 12). *Let* $R=\bigoplus_{i=0}^{\infty} R_i$ *be a reduced, finitely generated, graded* $\mathbb{C}$*-algebra, with* $R_0=\mathbb{C}$, *and* $\sqrt{R_N R}=m_R$, *all* $N \gg 0$. *Suppose* R_1 *contains a non-0 divisor. If* $D \neq 0$ *is a derivation of weight* -1, *then there is an* $x \in R_1$ *with* $Dx=1$.

Proof. Let $x_1 \in R_1$ be a non-0 divisor, and let $\{x_1, \ldots, x_s\}$ be a set of homogeneous generators of R. Let $n_i=\deg x_i$, and $r_i=$largest integer r such that $D^r x_i \neq 0$. Since D has weight -1, we have $r_i \leqq n_i$.

We claim that for some i, $r_i=n_i$. Then $x'=D^{n_i-1} x_i \in R_1$ will have $Dx' = a \in R_0=\mathbb{C}$, with $a \neq 0$, so $x=\frac{1}{a} x'$ will be the desired element.

Following [3], suppose $\max(r_i/n_i)=b/a<1$, with $(a,b)=1$ (note that some $r_i>0$, else $D \equiv 0$). Thus, $Dx_1=0$. Therefore, writing $x=x_1$, the derivation xD, of weight 0, acts nilpotently on each R_n; in fact, if $F \in R_n$, $(xD)^i F=x^i D^i F$. For each $t \in \mathbb{C}$, there is a graded automorphism e^{txD} of R, defined on R_n by

$$e^{txD} F=F+txDF+, \ldots, +\frac{t^n}{n!} x^n D^n F;$$

this defines a graded $\mathbb{G}_a$-action on R, hence on $\operatorname{Proj} R$. Using the functions $\{x_i\}$, a closed point in $\operatorname{Proj} R$ is described by a certain weighted homogeneous s-tuple $[\alpha_1, \alpha_2, \ldots, \alpha_s]$, with

$$\text{(2.7.1)} \qquad [\alpha_1, \alpha_2, \ldots, \alpha_s]=[\lambda^{n_1}\alpha_1, \lambda^{n_2}\alpha_2, \ldots, \lambda^{n_s}\alpha_s], \quad \text{each } \lambda \in \mathbb{C}^*.$$

The $\mathbb{G}_a$-action on $\operatorname{Proj} R$ gives a family of isomorphisms $\tau(t)$, where

$$\tau(t)[\alpha_1, \ldots, \alpha_s]=[e^{txD} x_1(\underline{\alpha}), \ldots, e^{txD} x_s(\underline{\alpha})].$$

Suppose now $r_s/n_s=b/a$; then $D^{r_s} x_s \neq 0$, hence $xD^{r_s} x_s \neq 0$ (x is not a 0-divisor). Since R has no nilpotents, there is a point $\underline{\alpha} \in \operatorname{Proj} R$ with $xD^{r_s} x_s(\underline{\alpha}) \neq 0$. For this $\underline{\alpha}$, using the condition (2.7.1) to divide by powers of t, the point $\tau(t)(\underline{\alpha})$ has i^{th} coordinate (if $t \neq 0$)

$$t^{-\frac{b}{a} n_i}\left(\frac{t^{r_i}}{(r_i)!} x^{r_i} D^{r_i} x_i(\underline{\alpha})+\text{terms of lower } t\text{-degree}\right).$$

The leading term has t-exponent $-\frac{b}{a} n_i+r_i \leqq 0$, with $=$ iff $r_i/n_i=b/a$. Therefore, $\lim_{t \to \infty} \tau(t)(\underline{\alpha})=P$ (a point in $\operatorname{Proj} R$) has coordinate 0 in i^{th} entry if $r_i/n_i<b/a$, whence $P \in V(x_i \mid r_i/n_i<b/a)$. Since $(a,b)=1$, $r_i/n_i=b/a$ implies n_i is a multiple of a. Therefore, the ideal generated by x_i with $r_i/n_i<b/a$ contains R_{Na+1}. Since $\sqrt{R_{Na+1} R}=m_R$, all $N \gg 0$, $P \in V(m_R)$, a contradiction.

(2.8) To eliminate the hypothesis $R_1 \neq 0$, we use the graded version of the "branched covering trick" ([4], p. 97). The verification of the following lemma will complete the proof of Theorem 2.2.

Lemma 2.9. *Let $R = \bigoplus_{i=0}^{\infty} R_i$ be a finitely generated domain, $R_0 = \mathbb{C}$, and $\sqrt{R_N R} = m_R$, all $N \gg 0$. Let $D \neq 0$ be a derivation of weight -1. Then $R_1 \neq 0$ (hence contains x with $Dx = 1$).*

Proof. Suppose $R_1 = \ldots = R_{d-1} = 0$, $R_d \neq 0$, $z \in R_d$. Let $R' = R[t]/t^d - z$, where t has weight 1. It is easy to check that every element of R' may be written uniquely as $a_0 + a_1 t + \ldots + a_{d-1} t^{d-1}$, with $a_i \in R$. R' is graded, $R'_0 = \mathbb{C}$, $R'_1 = \mathbb{C}t$ (as $d > 1$), and

$$R'_a = R_a + tR_{a-1} + \ldots + t^{d-1} R_{a-d+1}. \tag{2.9.1}$$

An exercise shows t is not a 0-divisor and R' is reduced (if $a_0 + a_1 t + \ldots + a_{d-1} t^{d-1}$ is nilpotent, then so is the trace $= d a_0$, whence $a_0 = 0$). Since $Dz \in R_{d-1}$, $Dz = 0$; thus D extends uniquely to a derivation of weight -1 on R', by setting $Dt = 0$. Since $\sqrt{R_N R} = m_R$, $N \gg 0$, (2.9.1) implies $\sqrt{R'_N R'} \supset (m_R R', t) = m_{R'}$. The hypotheses of Lemma 2.7 are satisfied, so there is an $x \in R'_1$ with $Dx = 1$. Since $Dt = 0$ and $R'_1 = \mathbb{C}t$, this is a contradiction.

(2.10) One could easily improve Theorem 2.2 by showing there are no derivations of weight $-k$ on R, $k \geqq 2$. Use induction on k. If $R_i \neq 0$, some i between 1 and $k-1$, multiply the derivation by an element of R_i, and use the inductive hypothesis (and 2.2). If $R_k \neq 0$, $Dz \neq 0$ for $z \in R_k$, modify Zariski's lemma to show $R = A[z]$; but then $\sqrt{R_N R} = m_R$, $N \gg 0$, fails if $k > 1$. Otherwise, use the branched covering trick.

§ 3. Proof of Theorem 2

(3.1) Let X be a complete normal complex variety of dimension > 1; then $R = \bigoplus \Gamma(X, L^{\otimes n}) = \bigoplus R_n$ is a finitely generated graded normal domain, with $R_0 = \mathbb{C}$, and $\sqrt{R_N R} = m_R$, all $N \gg 0$ (1.1). If $H^0(\Theta_X \otimes L^{-1}) \neq 0$, there is a derivation $D \neq 0$ on R, of weight -1 (1.5.2). By Theorem 2.2, there is a $t \in R_1$ with $Dt = 1$, and $R = A[t]$, where $A = \{r \in R \mid Dr = 0\}$ is a graded subring. Now $t \in H^0(X, L)$ gives an isomorphism $L \simeq \mathcal{O}_X(E)$, where E is an effective Cartier divisor. Consider the natural exact sequences

$$0 \to \mathcal{O}_X((n-1)E) \to \mathcal{O}_X(nE) \to \mathcal{O}_E(nE) \to 0.$$

Taking cohomology, one obtains a graded exact sequence

$$0 \to R \xrightarrow{t} R \to \bigoplus_0^{\infty} H^0(E, \mathcal{O}_E(nE)) \to \bigoplus_0^{\infty} H^1(X, \mathcal{O}_X(n-1)E). \tag{3.1.1}$$

Lemma 3.2. *Under the above hypotheses, $H^1(\mathcal{O}_X(iE)) = 0$, all $i \in \mathbb{Z}$.*

Proof. R is normal, of dimension $\geqq 3$; as $R=A[t]$, A is normal, of dimension $\geqq 2$. Therefore, depth $R \geqq 3$. Let $U=\operatorname{Spec} R-\{m_R\}$; then $H^1(U,\mathcal{O}_U) \xrightarrow{\sim} H^2_{\{m\}}(R) = 0$. Recalling (1.4), the $\mathbb{C}^*$-bundle $\pi\colon U \to X$ gives

$$H^1(U,\mathcal{O}_U)=\bigoplus_{-\infty}^{\infty} H^1(X, L^{\otimes i}),$$

whence the result $(L^{\otimes i} \simeq \mathcal{O}_X(iE))$.

(3.3) Since $A \xrightarrow{\sim} R/tR$, (3.1.1) collapses to a (graded) isomorphism

$$A \xrightarrow{\sim} \bigoplus_{n=0}^{\infty} H^0(E, \mathcal{O}_E(nE)).$$

Since A is normal, so is $\operatorname{Proj} A \simeq E$ (1.1). As $X=\operatorname{Proj} R=\operatorname{Proj} A[t]$, Theorem 2.2 now follows.

References

1. Burns, D., Wahl, J.: Local contributions to global deformations of surfaces. Invent. Math. **26**, 67–88 (1974)
2. Dolgachev, I.: Weighted projective spaces. In: Group Actions and Vector Fields, Lecture Notes in Mathematics, vol. 956. Berlin-Heidelberg-New York: Springer 1982
3. Mori, S., Sumihiro, H.: On Hartshorne's conjecture. J. Math. Kyoto Univ. **18**, 523–533 (1978)
4. Mumford, D.: Pathologies III. Amer. J. Math. **89**, 94–104 (1967)
5. Schlessinger, M.: Rigidity of quotient singularities. Invent. Math. **14**, 17–26 (1971)
6. Wahl, J.: Equisingular deformations of normal surface singularities, I. Annals of Math. **104**, 325–356 (1976)
7. Wahl, J.: Derivations of negative weight and non-smoothability of certain singularities. Math. Ann. **258**, 383–398 (1982)
8. Zariski, O.: Studies in equisingularity, I. Amer. J. Math. **87**, 507–536 (1965)

Oblatum 8-VII-1981 & 29-IV & 10-XI-1982

Invent. math. 72, 323-334 (1983)

Inventiones mathematicae

A Classical Diophantine Problem and Modular Forms of Weight 3/2

J.B. Tunnell

Department of Mathematics, Fine Hall, Box 37, Princeton University, Princeton, NJ 08540, USA

to S. Chowla

Introduction

It is a classical Diophantine problem to determine which integers are the area of some right triangle with rational sides. The main result of this paper is the following.

Theorem. *Let formal power series in the variable q be given by $g=q\prod_{1}^{\infty}(1-q^{8n})(1-q^{16n})$ and, for each positive integer t, $\theta_t=\sum_{-\infty}^{\infty} q^{tn^2}$. Set $g\,\theta_2=\sum_{1}^{\infty} a(n)\,q^n$ and $g\,\theta_4=\sum_{1}^{\infty} b(n)\,q^n$.*

(a) *If $a(n)\neq 0$, then n is not the area of any right triangle with rational sides.*

(b) *If $b(n)\neq 0$, then $2n$ is not the area of any right triangle with rational sides.*

The power series $g\,\theta_2$ and $g\,\theta_4$ are the q-expansions of certain modular forms of weight 3/2. It would follow from some current conjectures in the theory of elliptic curves that the converses of statements (a) and (b) are true for square-free positive integers n (see Sect. 3).

In this introduction we will briefly recall the history of the problem and the connections to elliptic curves, as well as giving a description of the methods of the paper.

Let $\mathscr{C}$ be the set of areas of right triangles with rational sides. This is a subset of $(\mathbf{Q}^*)^+$, and consideration of similar triangles shows that it is a union of cosets of $(\mathbf{Q}^*)^2$. Classically, an integer in $\mathscr{C}$ was called a congruent number, and Dickson [9, Chap. XVI] traces the question of whether a given number is congruent back to Arab manuscripts and the Greeks prior to that. The positive integers not in $\mathscr{C}$ are called noncongruent numbers, a terminology we will use unless confusion with ideal theoretic congruence might result.

Partially supported by a grant from the National Science Foundation

From the Pythagorean formula it is clear that the rational number D is the area of a rational right triangle with hypotenuse h if and only if $(h/2)^2 \pm D$ are both rational squares. Hence D is in $\mathscr{C}$ if and only if the simultaneous Diophantine system

$$\begin{aligned} u^2 + D v^2 &= w^2 \\ u^2 - D v^2 &= z^2 \end{aligned} \tag{1}$$

has a rational solution (u, v, w, z) with $v \neq 0$. Geometrically, the two quadrics in $\mathbf{P}^3$ given by (1) intersect in a smooth quartic in $\mathbf{P}^3$ which contains the point $(1, 0, 1, 1)$. The intersection is thus an elliptic curve over $\mathbf{Q}$ and projection from $(u, v, w, z) = (1, 0, 1, 1)$ to the plane $z = 0$ gives a birational isomorphism with a plane cubic curve E^D having Weierstrass form $y^2 = x^3 - D^2 x$. The points on the space curve with $v = 0$ correspond to the points where $y = 0$ and the point at infinity on E^D. It is easy to see by reducing modulo primes that these points on $E^D(\mathbf{Q})$ are precisely those of finite order. Thus we arrive at the well-known result that D is the area of a rational right triangle if and only if the group $E^D(\mathbf{Q})$ of rational points on $y^2 = x^3 - D^2 x$ is infinite.

In making explicit the relation of the elliptic curve E^D to rational right triangles, D. Zagier has pointed out to me that it is more convenient to use the natural system of quadrics

$$\begin{aligned} a^2 + b^2 &= c^2 \\ a b &= 2 D t^2 \end{aligned} \tag{1'}$$

in place of (1). This leads directly to a plane cubic when the first equation is parameterized by $(a, b, c) = (c(1 - \lambda^2)/(1 + \lambda^2),\ 2c\lambda/(1 + \lambda^2),\ c)$ and these values are used in the second quadric. This yields $D(t(1 + \lambda^2)/c)^2 = \lambda - \lambda^3$, which upon multiplication by D^3 and setting $x = -D\lambda$ and $y = D^2 t(1 + \lambda^2)/c$ gives $y^2 = x^3 - D^2 x$. A rational solution (x, y) of this equation with $y \neq 0$, corresponds to a right triangle with sides $|(D^2 - x^2)/y|$, $|2Dx/y|$, $|(D^2 + x^2)/y|$.

There are several known criteria for an integer D to be noncongruent, all of which seem to be equivalent to proving by means of a 2-descent on E^D that the group of rational points is finite. References [1, 9, 12, and 15] contain samples of these results. For example, 1 and primes congruent to 3 modulo 8 are not the area of any rational right triangle. The smallest integer in $\mathscr{C}$ is 5; it is the area of the right triangle with sides 9/6, 40/6, 41/6 discovered by Fibonacci, among others. Investigations have been undertaken by making a computer search for solutions to the original Diophantine system and tabulating the results [1, 2]. The most recent tabulations and references can be found in [26]. Numerical evidence from such calculations suggested to the authors of [1] that all positive integers congruent to 5, 6, or 7 modulo 8 should be congruent numbers. Stephens [21] observed that this would follow from a weak form of the conjecture of Birch and Swinnerton-Dyer and asserted that the method of Heegner points could be applied to prove that primes congruent to 5 or 7 modulo 8 or twice primes congruent to 3 modulo 8 are in fact the areas of rational right triangles. B. Gross has informed me that refinements of these methods show that a positive integer with at most two prime factors which is congruent to 5, 6, or 7 modulo 8 is the area of a rational right triangle.

The fact that there exist modular forms of weight 3/2 such that the non-vanishing of the d^{th} Fourier coefficient implies that $E^d(\mathbf{Q})$ is finite follows from several recent theorems. First, the L-series of the elliptic curve E: $y^2=x^3-x$ is the Mellin transform of the image φ of some form of weight 3/2 (and in fact of several) under the correspondence of Shimura [18]. Second, the main theorem of Waldspurger [25] shows that the square of the n^{th} coefficient of a suitable form of this type is a multiple of $L(E^d,1)$ for d equals n or $2n$. Finally, the result of Coates-Wiles [7] shows that if $L(E^d,1)\neq 0$, then $E^d(\mathbf{Q})$ is finite.

When forms having the above properties are found, they provide an efficient way to prove that certain numbers are not areas of any rational right triangle. Conjecturally, it reduces the problem of determining if D is in $\mathscr{C}$ to an algebraic computation involving $O(D^{3/2})$ steps. For example, the coefficient $a(n)$ is the number of triples of integers (x,y,z) such that $2x^2+y^2+32z^2=n$ minus one-half the number of triples such that $2x^2+y^2+8z^2=n$.

The first section of the paper considers the Shimura correspondence and determines forms of weight 3/2 giving rise to the modular form ϕ with L-series $L(E,s)$. In the second section the calculations necessary to apply Waldspurger's results are carried out, and the values of $L(E^d,1)$ are computed. This may be compared with the calculations of [3]. The third section applies these results to the problem of congruent numbers. A table of square free noncongruent integers less than 1000 is given. Conjecturally, any square free integer less than 1000 not in that table is the area of some rational right triangle. The comparison of the results here with the conjecture of Birch and Swinnerton-Dyer gives a formula for the conjectural order of the Tate-Shafarevitch group of E^d. The final section discusses the proof of some classical criteria for noncongruent numbers from the results of previous sections.

1. The Curve $y^2=x^3-D^2x$ and Forms of Weight 3/2

The elliptic curve E^D: $y^2=x^3-D^2x$ has complex multiplication by $\mathbf{Z}[i]$, and the L-function $L(E^1,s)$ is the Mellin transform of the unique normalized newform ϕ of weight 2, level 32 and trivial character. Thus the L-series of E^D is the Mellin transform of the form $\phi\otimes\chi_D$, where χ_D is the quadratic Dirichlet character corresponding to $\mathbf{Q}(\sqrt{D})$. The curve $E=E^1$ is the curve $32A$ of Table 1 of [4]. It is isogenous to $X_0(32)$, and $L(E,s)=\sum\chi(\mathfrak{a})N\mathfrak{a}^{-s}$ for a weight 1 Hecke character χ of $\mathbf{Q}(i)$. Some coefficients of the q-expansion of ϕ are tabulated in [4, Page 117]; they are easily computed from χ or by counting points over $\mathbf{F}_p$. The expansion begins $\phi=q-2q^5-3q^9+6q^{13}+2q^{17}+\dots$.

Shimura has shown in [18] that if f is a cusp form of weight $k/2$, for $k>1$ odd, which is an eigenform for Hecke operators $T(p^2)$ with eigenvalue λ_p, for all primes p, then there exists a form of weight $k-1$ which is an eigenform with eigenvalue λ_p for $T(p)$ for all p. This is called the Shimura map from cusp forms of half integer weight $k/2$ to forms of weight $k-1$. The effect of this map is to square the corresponding characters. From [10, § 5.3] we see that the form ϕ of weight 2 giving the L-series $L(E,s)$ is the image of at least one form of weight 3/2 with quadratic character χ. Further, it is established there that if

f is of weight 3/2 and is orthogonal to the forms of the type $\sum_{m=1}^{\infty} \psi(m)\, m\, q^{tm^2}$ which have the same level and character as f, then the image of f under the Shimura map is a cusp form. Contrary to the assertion of [10, Page 120], the modular form ϕ of level 32 is not the image of a form of weight 3/2, level 64 with quadratic character. For, from the dimension formulas of [8], we see that the space of such forms with trivial character is spanned by $\sum_{m=1}^{\infty} \psi(m)\, m\, q^{m^2}$ where ψ is of conductor 4, while the space of such forms with nontrivial quadratic character is zero. The situation is more favorable for forms of weight 3/2 and level 128. From [8] the dimension of the space of modular forms of weight 3/2, level 128 and fixed quadratic character is 3. This is the same as the dimension of the space of forms of weight 1/2 with level 128 and quadratic character, which suggests constructing such weight 3/2 forms by multiplying forms of weight 1/2 by a weight 1 form g. Let $\theta_t = \sum_{-\infty}^{\infty} q^{tm^2}$. This is a modular form of weight 1/2, level $4t$ and character χ_t. By the results of Serre-Stark [17], $\{\theta_2, \theta_8, \theta_{32}\}$ is a basis for forms of weight 1/2, level 128 and character χ_2. The set $\{\theta_1, \theta_4, \theta_{16}\}$ is a basis for the analogous space with trivial character. The next theorem gives a weight one form g of level 128 and character χ_{-2} which enables the weight 3/2, level 128 spaces to be analyzed completely. This method is computationally simpler than constructing the weight 3/2 forms via theta-functions of ternary quadratic forms.

Theorem 1. *There exists a unique normalized newform g of weight* 1, *level* 128 *and character χ_{-2}. The q-expansion of this form is*

$$g = \sum (-1)^{m+n} q^{(4m+1)^2+16n^2} = \sum (-1)^n q^{(4m+1)^2+8n^2}, \qquad \text{where } (m,n) \text{ is in } \mathbf{Z}\times\mathbf{Z}.$$

Proof. Suppose that such a form g exists. Then g, $g\otimes\chi_2$, $g\otimes\chi_{-1}$, $g\otimes\chi_{-2}$ will also be normalized newforms of level 128 with the same character [19]. They are not all independent, for multiplication by θ_1 gives forms of weight 3/2 and level 128 with character χ_2, which lie in a space of dimension 3. Since the 4 newforms above are dependent, it must be true that $g = g\otimes\chi_t$ for some nontrivial quadratic character χ_t of conductor dividing 8. We wish to show that the Dirichlet series associated to g is the Artin-L-series of a two-dimensional irreducible Artin representation which is induced from a character of an index two subgroup. Then the problem of finding all such modular forms g will be reduced to a problem of Galois theory. It is a special case of a general result of Labesse and Langlands (L-indistinguishability for $SL(2)$, Canad. J. Math. XXXI (1979), 726–785; Proposition 6.5) that if $g = g\otimes\chi_t$, then g is as described above. Alternately, the theorem of Deligne and Serre (Formes modulaires de poids 1, Ann. Sc. de l'Ec. Norm. Sup., t 7 (1974), 507–530) shows that the L-series of g is the Artin L-series of some irreducible two-dimensional Artin representation σ of Artin conductor 128 and determinant χ_{-2}. Since $g = g\otimes\chi_t$, we have that $\sigma = \sigma\otimes\chi_t$, which implies by Frobenius reciprocity that σ is induced from an index two subgroup. It is easy to check that there is up to isomorphism only one such Artin representation with Artin conductor 128 and determinant the character χ_{-2}. It is in fact induced from any of the 3 quadratic extensions inside the field of 8^{th} roots of unity. The L-series of g may be expressed in three ways as the Dirichlet L-series of a character of a quadratic extension K. When $K=\mathbf{Q}(i)$, η may be taken to be the character of the $(1+i)^5$-ideal classes which is trivial on $(1+2i)$ and -1 on (5) (these ideals generate the ideal class group in question, which has order 4). It is easy to see that $g=\sum \eta(\mathfrak{a})\, q^{N\mathfrak{a}}=\sum (-1)^{m+n} q^{(4m+1)^2+16n^2}$, the sum taken over all (m,n) in $\mathbf{Z}\times\mathbf{Z}$. A similar computation shows that when $K=\mathbf{Q}(\sqrt{-2})$, the Dirichlet character of this field may be taken to be the character η' of the 4-ideal class group which is trivial on (3) and takes value i on $(1+\sqrt{-2})$. Then $g=\sum \eta'(\mathfrak{b})\, q^{N\mathfrak{b}}=\sum (-1)^n q^{(4m+1)^2+8n^2}$. The expression coming from the field $\mathbf{Q}(\sqrt{2})$ will not be used in the sequel.

Remark. The form g has a long history. Jacobi noticed that $g=q\,\Pi(1-q^{8n})(1-q^{16n})$ and remarked on the two representations given in Theorem 1. A recent reference to this form is [13], a serendipitous one is H.J. Smith's Report on Number Theory [20], where he treats Jacobi's example in his article 128!

By Theorem 1 and the previous remarks, a basis for the space of cusp forms of weight 3/2, level 128 and trivial character is $\{g\theta_2, g\theta_8, g\theta_{32}\}$. Similarly, $\{g\theta_1, g\theta_4, g\theta_{16}\}$ is a basis for the weight 3/2 cusp forms of level 128 and character χ_8.

Theorem 2. *The modular forms $g\theta_2$, $g\theta_4$, $g\theta_8$ and $g\theta_{16}$ correspond to the weight two form ϕ (of level 32, trivial character) under Shimura's map from forms of weight 3/2 to forms of weight 2.*

Proof. The first few terms of the q-expansions of the forms of weight 3/2 with level 128 and trivial character are as follows:

$$g\theta_2=q+2q^3+q^9-2q^{11}-4q^{17}-2q^{19}-3q^{25}+4q^{33}-4q^{35}+\dots$$
$$g\theta_8=q+q^9-4q^{17}-3q^{25}+4q^{33}+\dots$$
$$g\theta_{32}=q-q^9-2q^{17}+q^{25}+2q^{33}+\dots.$$

The Hecke operators $T(p^2)$ preserve this space. Consideration of $T(3^2)$ and $T(5^2)$ shows that $g\theta_2, g\theta_8$ and $2g\theta_{32}-g\theta_8$ are eigenforms. The first two have eigenvalues $\lambda_3=0$, $\lambda_5=-2$, while it is clear that $2g\theta_{32}-g\theta_8=\sum_{-\infty}^{\infty}\psi(m)\, m q^{m^2}$, where ψ is the nontrivial quadratic character of conductor 4.

To derive that $g\theta_2$ and $g\theta_8$ are eigenforms for all $T(p^2)$, notice that they are orthogonal to $\sum\psi(m)\, m q^{m^2}$ since the eigenvalues of $T(3^2)$, $T(5^2)$ are different than on the latter form. Hence the span of $g\theta_2$ and $g\theta_8$ is $T(p^2)$ stable. The form $g(\theta_2-\theta_8)$ has q^n appearing with nonzero coefficient only when $n\equiv 3$ (8) (since $g=\sum c(n)\, q^n$ and $c(n)=0$ unless $n\equiv 1$ (8)). Similarly $g\theta_8$ has exponents in the q-expansion all congruent to 1 modulo 8. It is clear from the formula for the action of $T(p^2)$ [18, Theorem 1.7] that $T(p^2)(g(\theta_2-\theta_8))$ and $T(p^2)(g\theta_8)$ have the same properties with respect to exponents modulo 8 appearing in the

q-expansion. Since $T(p^2)$ acts on the span of $g(\theta_2-\theta_8)$ and $g\theta_8$, it must be that they are individually eigenforms for all $T(p^2)$.

Finally, from Shimura's theory there exist weight two forms ϕ_1, ϕ_2 of level at most 128 having $T(p)$ eigenvalues in agreement with those of $T(p^2)$ on $g(\theta_2-\theta_8)$ and $g\theta_8$ respectively. Since the eigenvalues for $p=3, 5$ are known, it is easy to compare these values with the forms appearing in Table 3 of [4] of level dividing 128 to see that $\phi_1=\phi_2=\phi$ is the only possibility.

The case of the forms $g\theta_4$ and $g\theta_{14}$ with character χ_8 is analogous to the preceeding; we will not carry it out here.

For future reference write $g\theta_2=\sum a(n)q^n$ and $g\theta_4=\sum b(n)q^n$. Let the form of Theorem 1 be given by $g=\sum c(n)q^n$, with $c(n)=0$ for $n\leqq 0$. Then $a(n) = \sum_{m=-\infty}^{\infty} c(n-2m^2)$ and $b(n)=\sum_{m=-\infty}^{\infty} c(n-4m^2)$. Formulas for $a(n)$ and $b(n)$ may be given in terms of the characters η, η' of the proof of Theorem 1 and expressions of n by ternary quadratic forms. It is easy to see by using Theorem 1 that for n odd, $a(n)$ equals the number of triples of integers (x, y, z) such that $2x^2+y^2+32z^2=n$ minus one-half the number of (u, v, w) with $2u^2+v^2+8w^2=n$.

2. Waldspurger's Theorem on L-Series Values

We will specialize to our situation the following theorem of Waldspurger.

Theorem (Waldspurger [25, Theorem 1]). *Let ϕ be a newform of weight $k-1$ and character χ^2 which is the image of a form f of weight $k/2$ under Shimura's map. Assume further that* 16 *divides the level of ϕ. Then there exists a function $A(t)$ from square free integers to* $\mathbf{C}$ *such that*

(i) $A(t)^2\,\varepsilon(\chi^{-1}\chi_{-1}^{(k-1)/2}\chi_t, 1/2)=2(2\pi)^{(1-k)/2}\,\Gamma((k-1)/2)\,L(\phi\chi^{-1}\chi_{-1}^{(k-1)/2}\chi_t, (k-1)/2)$.

(ii) *For each positive integer N, there exists a finite set of explicitly described functions $c(n)$ such that $\sum A(n^{sf})\,c(n)\,q^n$ for $c(n)$ in this set spans the forms of weight $k/2$, level N, and character χ which correspond to ϕ via Shimura's map.*

Remark. The statement here is a special case of that of [25], which is sufficient for our purposes. The factor $\varepsilon(\eta, 1/2)$ for a Hecke character η is the one in [24]; Waldspurger uses the inverse in his statement. In particular, when η is quadratic $\varepsilon(\eta, 1/2)=1$ (since ε is inductive and $\varepsilon(1)=1$ [24]). The sets of functions $c(n)$ are given in the 11 equations of Sect. VIII.4 of [25]. They simplify immensely in the case of interest here.

Theorem 3. *Let $g\theta_2=\sum a(n)q^n$ and $g\theta_4=\sum b(n)q^n$ be modular forms of weight* 3/2 *and level* 128 *corresponding to the unique weight two normalized newform of level* 32 *and trivial character. For d a square-free odd positive integer we have*

$$L(E^d, 1)=a(d)^2\,\beta\,d^{-1/2}/4$$

$$L(E^{2d}, 1)=b(d)^2\,\beta(2d)^{-1/2}/2.$$

where

$$\beta = \int_1^\infty dx/(x^3 - x)^{1/2} = 2.62205\ldots \quad \textit{is the real period of } E.$$

Proof. We compare the forms constructed in Waldspurger's theorem with those constructed in Theorem 2. In this case, the possible functions are all of the form $c(n) = n^{1/4} \Pi\, c_p(n)$, where for p odd and n square-free, $c_p(n) = 1$ [25, VIII.4.3]. The possible nonzero choices for $c_2(n)$ are the characteristic function of an odd residue class modulo 8 [25, VIII.4.1].

When χ is trivial, we apply Waldspurger's theorem to find that $\sum A(n^{sf})\, c(n)\, q^n$ for the four choices of the function $c(n)$ above span the same space as $g\theta_2, g\theta_8$. Since $g\theta_2$ and $g\theta_8$ have no terms q^n appearing with nonzero coefficient when n is not congruent to 1 or 3 modulo 8, it must be true that $A(n) = 0$ for $n \equiv 5, 7$ (modulo 8). Choosing $c(n)$ to be respectively the characteristic functions of 1 and 3 modulo 8 shows that for n square-free $a(n) = \beta_1 A(n)\, n^{1/4}$ and $a(n) = \beta_3 A(n)\, n^{1/4}$. Thus we have

$$\begin{aligned} a(n)^2 &= \beta_1^2 L(E^n, 1)\, n^{1/2} & n &\equiv 1 \quad (8), \\ a(n)^2 &= \beta_3^2 L(E^n, 1)\, n^{1/2} & n &\equiv 3 \quad (8), \\ a(n)^2 &= 0 = A(n)^2 = L(E^n, 1) & n &\equiv 5, 7 \ (8). \end{aligned}$$

To compute β_1, β_3 we need to explicitly evaluate some L-series values. Let β be the period of E. It is shown in [3, Table 1] that $L(E^n, 1)\, n^{1/2}/\beta$ is rational, and a table is given for certain n. In particular $L(E, 1)/\beta = 1/4$ and $L(E^3, 1)\, 3^{1/2}/\beta = 1$. Since $a(1) = 1$ and $a(3) = 2$ this shows that $4/\beta = \beta_1^2 = \beta_3^2$.

When $\chi = \chi_2$, Waldspurger's theorem again applies. In this case, $A(t)^2 = L(\phi\, \chi_2\, \chi_t, 1) = L(E^{2t}, 1)$. Comparing the Waldspurger basis with the basis $(g\theta_4 - g\theta_{16})$, $g\theta_{16}$ gives in this case that $A(n) = 0$ when $n \equiv 3, 7$ modulo 8 and that

$$\begin{aligned} b(n) &= \gamma_1 A(n)\, n^{1/4} & n &\equiv 1 \ (8), \\ b(n) &= \gamma_5 A(n)\, n^{1/4} & n &\equiv 5 \ (8). \end{aligned}$$

Using this to compute $L(E^{2n}, 1)$, and comparing with the tables of [3] to find $L(E^2, 1)$ and $L(E^{10}, 1)$ verifies the theorem.

3. Applications

The previous results can be applied to prove that certain numbers are not the areas of rational right triangles by invoking the following theorem.

Theorem (Coates-Wiles [7]). *Let E be an elliptic curve over $\mathbf{Q}$ with complex multiplication by the ring of integers in a quadratic field of class number* 1. *If $L(E, 1) \neq 0$, then $E(\mathbf{Q})$ is finite.*

From this and Theorem 3 we obtain immediately our main result. Recall that $g\theta_2 = \sum a(n)\, q^n$ and $g\theta_4 = \sum b(n)\, q^n$ are forms of weight 3/2, that $a(n) = 0$ unless $n \equiv 1$ or 3 modulo 8 and $b(n) = 0$ unless $n \equiv 1$ or 5 modulo 8. As a notational device, let $b(n/2)$ be zero if $n/2$ is not integral. Notice that $a(n)+b(n/2)$ is one of $a(n)$ or $b(n/2)$.

Theorem 4. *If $a(n)+b(n/2)\neq 0$ then n is not the area of a rational right triangle.*

Comparison with the conjecture of Birch and Swinnerton-Dyer [3], [22, § 8] leads to a sharp conjecture about congruent numbers. To frame their conjecture we need the basic conjecture that the L-series of an elliptic curve over Q has a meromorphic continuation to a function on the complex plane. Attached to such a curve is a cohomologically defined group $Ш$, the Tate-Shafarevitch group, which is conjectured to be finite. The conjecture of Birch and Swinnerton-Dyer relates the L-series value at 1 to the group of rational points of the elliptic curve and the conjectural order of $Ш$.

Conjecture (Birch and Swinnerton-Dyer). *Let A be an elliptic curve over $\mathbf{Q}$ and let $A(\mathbf{Q})$ be the group of rational points of A. Then*

(i) *$L(A,1)\neq 0$ if and only if $A(\mathbf{Q})$ is finite.*

(ii) *When $A(\mathbf{Q})$ is finite, $L(A,1)=\alpha|Ш|\,\Pi\, c_p/|A(\mathbf{Q})|^2$, where α is the integral of a minimal Néron differential over $A(\mathbf{R})$, $c_p=[A(Q_p)\colon A_0(Q_p)]$ and $|G|$ denotes the order of a group G.*

If this conjecture is valid for the curves E^d, then d is the area of a rational right triangle if and only if $L(E^d,1)=0$. Combining the above with Theorem 3 gives the following conjectural description of congruent numbers. Let $\sigma_0(n)$ be the number of positive divisors of n.

Conjecture. *Let d be a square-free positive integer. Then d is a congruent number if and only if $a(d)+b(d/2)=0$. If d is not congruent, the order $|Ш(E^d)|$ of the Tate-Shafarevitch group is $(a(d)/\sigma_0(d))^2$ when d is odd and $(b(d/2)/\sigma_0(d/2))^2$ when d is even.*

Remark. In order to derive the conjecture from that of Birch-Swinnerton-Dyer and Theorem 3, it is only necessary to apply the algorithm of [23] to check that $c_p=4$ when p is odd, $p|d$ and that $c_2=2$ or 4 according to d odd or even. The remaining $c_p=1$, and $E^d(\mathbf{Q})$ has order 4 in all cases.

Among other things, the conjecture predicts that $a(n)/\sigma_0(n)$ and $b(n/2)/\sigma_0(n/2)$ are integers, and gives an efficient algebraic criterion for deciding when a number is the area of a rational right triangle. Table 1 contains a list of all square-free positive integers n less than 1000 such that $a(n)+b(n/2)\neq 0$, By Theorem 4, these are all noncongruent, and conjecturally this list contains all noncongruent square-free positive integers less than 1000. The table contains several numbers left undecided in previous works [1, 2, 26]. The numbers are tabulated according to the value of $a(n)/\sigma_0(n)$ for n odd and $b(n/2)/\sigma_0(n/2)$ for n even; that is by the signed square root of the conjectural order of $Ш(E^n)$.

Table 1. Noncongruent square-free integers <1000

$a(N)/\sigma_0(N)$	N
1	1 3 33 51 57 59 83 139 177 187 209 211 267 321 339 345 379 385 411 451 489 499 515 555 587 595 649 659 665 681 707 803 811 827 835 899 921 969
−1	11 19 35 67 91 105 115 123 129 179 195 201 227 235 249 273 347 393 403 419 427 435 473 483 563 611 635 683 691 705 715 739 753 779 787 795 817 843 851 993
2	73 155 185 203 241 281 329 355 545 553 579 601 627 641 697 755 763 785 865 937
−2	17 89 97 193 217 233 259 305 377 401 449 481 497 617 667 713 745 769 897 929 955 977 979
3	43 131 163 417 491 537 571 619 849 913 923
−3	107 251 283 331 547 633 643 699 737 771 883
4	113 337 577 593 809 857 881 953
−4	409 521 569 939
5	307 859 971
−5	443 523 947
−6	433 673
−7	467
9	907

$b(N/2)/\sigma_0(N/2)$	N
1	2 10 58 74 114 122 130 170 258 290 314 346 354 362 370 402 474 506 586 610 618 642 714 730 746 786 826 906 922 946 962 970 986
−1	26 42 66 106 186 202 266 418 498. 530 554 570 634 682 690 754 762 770 834 858 874 930
2	82 282 562 626 818 914
−2	146 178 274 322 466 938 994
3	298 778
−3	218 394 458 538 794 842 978
4	706 802
−4	482 898
5	698

4. Criteria for Noncongruent Numbers

There are several classical criteria which yield noncongruent numbers [1, 9, 15]. For example, it is known that if p and q are primes congruent to 5 modulo 8, then pq and $2p$ are not congruent numbers. This section will explain how such classical criteria may be derived from Theorems 1 and 4. I am indebted to K. Kramer for pointing out to me that all the classical criteria can be obtained by making a 2-descent on the curve E^d, or on an isogenous curve, to prove that $E^d(\mathbf{Q})$ is finite. He also provided an extensive list of criteria obtained in this fashion of which a few simple examples are treated here.

Proposition 5. *Let p be a prime congruent to* 3 *modulo* 8. *Then* $a(p) \equiv 2 \pmod 4$, *so that p is not a congruent number.*

Proof. It is easy to see from the proof of Theorem 1 and the definition that $a(n) = \sum \eta(x+iy)$ over all triples of integers (x, y, z) such that $x>0$ is odd, y is even and such that $n=x^2+y^2+2z^2$. When $n \equiv 3 \pmod 8$, z must be nonzero. There are two expressions $p=x^2+2z^2$, one with z the negative of the other. Since $\eta(x+iy)+\eta(x-iy)$ is even, the sum of $\eta(x+iy)$ over $p=x^2+y^2+2z^2$ with $x>0$ and odd and $y \neq 0$ is divisible by 4. Hence $a(p) \equiv 2\eta(a) \equiv 2 \pmod 4$ if $p=a^2+2b^2$.

For further applications it will necessary to count the number of representations of an integer as $2a^2+b^2+c^2$. This ternary quadratic form is in a genus with one class. It is well known that the number of representations of n by quadratic ternary forms in a genus is related to class numbers of quadratic fields. The following result, taken from [11, Page 194] will be sufficient for applications here. Let d be an odd square-free integer greater than 1. Let $N(d)$ be the number of triples of integers, modulo the action of unimodular integral matrices stabilizing the form $q(x, y, z)=x^2+y^2+2z^2$, such that $q(x, y, z)=d$. Then $N(d)=h(-2d)/2$, where $h(-2d)$ is the class number of $\mathbf{Q}(\sqrt{-2d})$.

Proposition 6. *Let* $p \equiv 1 \pmod 8$ *be a prime. Write* $p=a^2+4b^2$ *and suppose that* 16 *does not divide* $p-1+4b$. *Then* $a(p) \equiv 4 \pmod 8$, *and p is not congruent.*

Proof. There are unique expressions $p=a^2+4b^2=c^2+2d^2$ in positive integers. Hence, of the expressions of p as $x^2+y^2+2z^2$, exactly two are such that $xyz=0$. Of the $h(-2p)/2-2$ remaining expressions (x, y, z) and $(x, -y, -z)$ are counted together, since multiplying y by -1, z by -1 is a unimodular transformation preserving $x^2+y^2+2z^2$. In the sum these contribute $\eta(x+iy)+\eta(x-iy)=\pm 2$. The contributions of (x, y, z) and $(x, y, -z)$ are the same, showing that $a(p) \equiv 2(\eta(a+2bi)+\eta(c))+h(-2p)+4 \pmod 8$.

The hypothesis is equivalent to $h(-p) \equiv 4 \pmod 8$, since by [6], $h(-p) \equiv \frac{p-1}{2}+2b$. From Proposition 2 of [14], $h(-p)+h(-2p) \equiv \frac{p-1}{2} \pmod 8$. From the fact [5] that $p=r^2+2\cdot 16s^2$ implies that $h(-p) \equiv 0 \pmod 8$ we see that $c \equiv \pm 3$ when $p \equiv 1 \pmod{16}$ and $c \equiv \pm 1$ when $p \equiv 9 \pmod{16}$. It is easy to see that $\eta(a+2bi)=-1$ in all cases. Thus, $a(p) \equiv 2(-1+\eta(c))+\frac{p-1}{2} \equiv 4 \pmod 8$.

The hypothesis of the previous theorem may be stated in several ways, which are equivalent to those considered by Razar in [16, Theorem 2].

Proposition 7. *Let p and q be primes congruent to* 5 *modulo* 8. *Then* $b(pq) \equiv 4 \pmod 8$, *so that* $2pq$ *is not the area of a rational right triangle.*

Proof. The alternate expression for g in terms of an ideal class character η' of $\mathbf{Q}(\sqrt{-2})$ in Theorem 1 may be used. Then $b(pq)=\sum \eta'(x+y\sqrt{-2})$ over all triples of integers (x, y, z) such that $pq=x^2+2y^2+z^2$ with $x>0$ and z even. Of the $h(-2pq)/2$ expressions of pq (up to unimodular automorphism stabilizing $x^2+2y^2+z^2$) there are 2 with $y=0$. In the remaining $(h(-2pq)/2-2)$, (x, y, z) and $(x, -y, -z)$ are counted once, but contribute $\eta(x+y\sqrt{-2})+\eta(x-y\sqrt{-2}) = \pm 2$ to the sum. Hence, $b(pq)\equiv 4+2(h(-2p)/2-2)\equiv h(-2p) \pmod 8$. From [14; Cor. 1, Prop. 5] $h(-2p)\equiv 4 \pmod 8$. This establishes the result.

Similar criteria are obtainable in other cases by expressing $a(n)$ or $b(n)$ in terms of a quadratic class number and computing modulo 8.

References

1. Alter, R., Curtz, T.B., Kubota, K.K.: Remarks and results on congruent numbers. Proc. Third Southeastern Conf. on Combinatorics, Graph Theory and Computing 1972, pp. 27–35
2. Alter, R.: The congruent number problem. Amer. Math. Monthly **87**, 43–45 (1980)
3. Birch, B.J., Swinnerton-Dyer, H.P.F.: Notes on elliptic curves II. J. reine angewandte Math. **218**, 79–108 (1965)
4. Birch, B.J., Kuyk, W.: Tables on elliptic curves. In: Modular functions of one variable IV. Lecture Notes in Mathematics, vol. 476, pp. 81–144. Berlin-Heidelberg-New York: Springer 1979
5. Barrucand, P., Cohn, H.: Note on primes of type x^2+32y^2, class number, and residuacity. J. reine angewandte Math. **238**, 67–70 (1969)
6. Brown, E.: The class number of $\mathbf{Q}(\sqrt{-p})$, for $p\equiv 1 \pmod 8$ a prime. Proc. Amer. Math. Soc. **31**, 381–383 (1972)
7. Coates, J., Wiles, A.: On the conjecture of Birch and Swinnerton-Dyer. Invent. Math. **39**, 223–251 (1977)
8. Cohen, H., Oesterlé, J.: Dimension des espaces de formes modulaires. In: Modular functions of one variable VI. Lecture Notes in Mathematics, vol. 627, pp. 69–78. Berlin-Heidelberg-New York: Springer 1977
9. Dickson, L.E.: History of the theory of numbers II. Carnegie Institution, Washington, DC (1920) (reprinted by Chelsea, 1966)
10. Flicker, Y.: Automorphic forms on covering groups of $GL(2)$. Invent. Math. **57**, 119–182 (1980)
11. Jones, B.W.: The arithmetic theory of quadratic forms. Math. Assoc. of Amer., Baltimore, MD 1950
12. Lagrange, J.: Thèse d'Etat de l'Université de Reims, 1976
13. Moreno, C.J.: The higher reciprocity laws: an example. J. Number Theory **12**, 57–70 (1980)
14. Pizer, A.: On the 2-part of the class number of imaginary quadratic number fields. J. Number Theory **8**, 184–192 (1976)
15. Razar, M.: The nonvanishing of $L(1)$ for certain elliptic curves with no first descents. Amer. J. Math. **96**, 104–126 (1974)
16. Razar, M.: A relation between the two-component of the Tate-Shafarevitch group and $L(1)$ for certain elliptic curves. Amer. J. Math. **96**, 127–144 (1974)
17. Serre, J-P., Stark, H.M.: Modular forms of weight 1/2. In: Modular functions of one variable VI. Lecture Notes in Mathematics, vol. 627, pp. 27–68. Berlin-Heidelberg-New York: Springer 1977
18. Shimura, G.: On modular forms of half-integral weight. Ann. of Math. **97**, 440–481 (1973)
19. Shimura, G.: Introduction to the arithmetic theory of automorphic functions. Iwanami Shoten and Princeton University Press 1971
20. Smith, H.J.: Collected Mathematical Papers, Volume 1, Oxford (1894). (reprinted by Chelsea, 1965)

21. Stephens, N.M.: Congruence Properties of Congruent numbers. Bull. London Math. Soc. pp. 182–184 (1975)
22. Tate, J.: The arithmetic of elliptic curves. Invent. Math. **23**, 179–206 (1974)
23. Tate, J.: Algorithm for determining the type of a singular fiber in an elliptic pencil. In: Modular functions of one variable IV. Lecture Notes in Mathematics, vol. 476, pp. 33–52. Berlin-Heidelberg-New York: Springer 1975
24. Tate, J.: Number theoretic background. In: Automorphic forms, representations, and *L*-functions. Proc. Symp. in Pure Math. XXXIII, Part 2, pp. 3–26 (1979)
25. Waldspurger, J.-L.: Sur les coefficients de Fourier des formes modulaires de poids demi-entier. J. de Math. pures et appliquées **60**, (4) 375–484 (1981)
26. Guy, R.K.: Unsolved problems. Amer. Math. Monthly **88**, 758–761 (1981)

Oblatum 22-I-1983

Invent. math. 72, 335-358 (1983)

Inventiones mathematicae

Topological Triviality of Deformations of Functions and Newton Filtrations

James Damon[1]* and Terence Gaffney[2]*

[1] Department of Mathematics, University of North Carolina, Chapel Hill, NC 27514, USA
[2] Department of Mathematics, Northeastern University, Boston, MA 02115, USA

Table of Contents

Introduction

We consider a deformation f of a germ $f_0 = k^n, 0 \to k, 0$ which has an (algebraically) isolated singularity at 0 (here $k = \mathbb{R}$ or $\mathbb{C}$ and the germs are correspondingly smooth, analytic or holomorphic). A fundamental problem is to determine when such a deformation is topologically trivial (i.e. topologically right equivalent as a deformation to the constant deformation). When $k = \mathbb{C}$, the basic approach to this problem is to prove the Milnor number is constant and use the results of Lê-Ramanujam and Timourian [10], [14] to conclude that the deformation is topologically trivial. However, this method does not apply when $k = \mathbb{R}$ (nor when f_0 defines a complex surface singularity). Also, the computation of the Milnor number can be difficult (see e.g. Kouchnirenko [8]).

* Partially supported by grants from the National Science Foundation and the British Science Research Council

An alternate approach is to use Teisser's μ_*-constant condition [13] to prove that the Whitney conditions hold and then obtain topological triviality as a corollary.

Here we describe an approach based on the Newton filtration of f_0. Kouchnirenko [8] made use of the Newton filtration in his computation of the Milnor number for non-degenerate functions. Our method uses certain filtration preserving properties of the deformation f to prove directly that f is topologically trivial without requiring non-degeneracy. This method gives a reasonably practical algorithm for proving topological triviality which also works for real germs and the germs defining complex surface singularities.

The general form of this result is given by a filtration condition involving the deformed terms $\partial f/\partial u_i$ and the deformation f. We also describe how this general form can be reduced to a useable criterion involving the deformed terms and the original germ f_0. For this, we consider the polyhedral structure of the Newton polyhedron of f_0. We construct local patching data for the faces of the polyhedron. When the filtration properties of the deformation behave well with respect to the local data, then the local conditions "patch together" to give the "global" filtration properties. Then, topological triviality is proven along the lines of [5] or [11] by solving the localized equation for triviality (with respect to right equivalence) using controlled vector fields.

In the case of non-degenerate functions, the filtration criteria reduces to the requirement that the deformation does not reduce the Newton filtration. Thus, we obtain as a corollary, a topological proof of the result of Kouchnirenko that for non-degenerate functions, the Milnor number only depends on the Newton polyhedron. Since our original announcement [6], we have learned of several other geometric proofs of this same fact; both Merle and Briançon have described distinct ways of obtaining Teisser's μ_*-constant condition from non-degeneracy (in fact, our method also implies the analytic version of Teissier's condition (c) which is equivalent to μ_*-constant). Also, Oka [12] has used the non-degeneracy condition to bound the Milnor radius and obtain topological triviality directly.

We wish to express our gratitude to the British Science Research Council for its support and to the Department of Mathematics of the University of Liverpool for its warm hospitality during the completion of this research.

§ 0. Notation

Throughout we consider germs $f\colon k^n, 0 \to k^p$ which are either C^∞ or real analytic if $k = \mathbb{R}$ or holomorphic if $k = \mathbb{C}$. We will use local coordinates $\mathbf{x}$ for k^n and denote the ring of germs $k^n, 0 \to k$ (in the appropriate category) by $\mathscr{C}_x$ with maximal ideal $\mathscr{m}_x$. For a germ $f\colon k^n, 0 \to k^p, 0$ we let $\theta(f)$ denote the module of germs of vector fields (in the appropriate category) $\zeta\colon k^n, 0 \to Tk^p$ such that $\pi_p \circ \zeta = f$ for $\pi_p\colon Tk^p \to k^p$. We denote $\theta(id_{k^n})$ by θ_n; this is a free $\mathscr{C}_x$-module generated by $\partial/\partial x_i$ $i = 1, \ldots, n$. For a deformation $f\colon k^{n+r}, 0 \to k, 0$ of a germ $f_0\colon k^n, 0 \to k, 0$ we let $\mathbf{u}$ denote local coordinates for k^r, the space of deformation parameters. Also, we let $\mathscr{C}_{x,u}$ denote the ring of germs $k^{n+r}, 0 \to k$. If

$\pi_{n,r}: k^{n+r},0 \to k^r,0$ is the projection, then $\theta(\pi_{n,r})$ is also a free $\mathscr{C}_{x,u}$-module generated by the $\partial/\partial x_i$.

Also, we denote the polynomial ring $k[x_1,\ldots,x_n]$ and the formal power series ring $k[[x_1,\ldots,x_n]]$ by $k[\mathbf{x}_n]$ and $k[[\mathbf{x}_n]]$ respectively. For any ring R and elements $u_1,\ldots,u_r$, we let $R\{u_i\}$ denote the R-module generated by the u_i (the number r will be clear from the context).

§ 1. The Newton Filtration

In this section, we recall [8] the definition of the Newton polyhedron and Newton filtration associated to a germ $f_0: k^n,0 \to k,0$. In fact, we give a slightly more general definition which is defined even when the usual Newton filtration may not be.

Let $\sum c_\alpha x^\alpha$ denote the Taylor series of f_0. Also, let $\mathbb{R}^n_+ = \{(x_1,\ldots,x_n) \in \mathbb{R}^n$, each $x_i \geqq 0\}$. By a *defining* half-space for f_0, we mean a closed half-space H of $\mathbb{R}^n$ such that i) if $c_\alpha \neq 0$, then $\alpha \in H$; ii) there are $n+1$ independent α with $c_\alpha \neq 0$ which lie in the hyperplane boundary ∂H for H, and iii) ∂H intersects each axis of $\mathbb{R}^n_+$ (at some point other than the origin).

If f_0 has at least one defining half-space, we will call f_0 *semi-fit*. For a semi-fit f_0, we let $\Gamma_+(f_0)$ denote the intersection in $\mathbb{R}^n_+$ of all defining half-spaces for f_0. It is called the *Newton polyhedron* for f_0. It is a convex polyhedron in $\mathbb{R}^n_+$ and so has a polyhedral decomposition. We let $\Gamma(f_0)$ denote the union of the compact faces of $\Gamma_+(f_0)$ and call it the *Newton boundary*. It also has a polyhedral decomposition.

We also say that f_0 is *fit* (commode in the terminology of Kouchnirenko [8]) if for each j there is an $x_j^{m_j}$ with non-zero coefficient in the Taylor expansion of f_0. Note that Kouchnirenko only carries out his analysis for fit germs.

It is easy to see that if f_0 is fit, it is semi-fit; and then the Newton polyhedron defined here agrees with that defined in [8]. However, the converse is not true if f_0 is only semi-fit.

Example. (Briançon-Speder example [4]).

$$f_0(x,y,z) = x^5 + y^7 z + z^{15}.$$

Then, f_0 is semi-fit but not fit. It is possible to change coordinates $(z \to z+y)$ to make f_0 fit; however, the complicates the problem of determining the topologically trivial deformations (see § 7).

From the Newton polyhedron, we construct the Newton filtration. For each face $\Delta \subset \Gamma(f_0)$ (or closed face $\bar{\Delta}$), $C(\Delta)$ (respectively $C(\bar{\Delta})$) denotes the cone of half-rays emanating from 0 and passing through Δ (or $\bar{\Delta}$). The $\{C(\Delta)\}$ gives a polyhedral decomposition of $\mathbb{R}^n_+$ which we call the Newton decomposition. The Newton filtration is then defined via a piecewise-linear map $\phi: \mathbb{R}^n_+ \to \mathbb{R}$ satisfying:

i) ϕ is linear on each $C(\bar{\Delta})$.

ii) ϕ takes positive integer values on the lattice points of $\mathbb{R}^n_+ \backslash \{0\}$.

iii) $\phi | \Gamma(f_0) \equiv m$ (for some positive integer m).

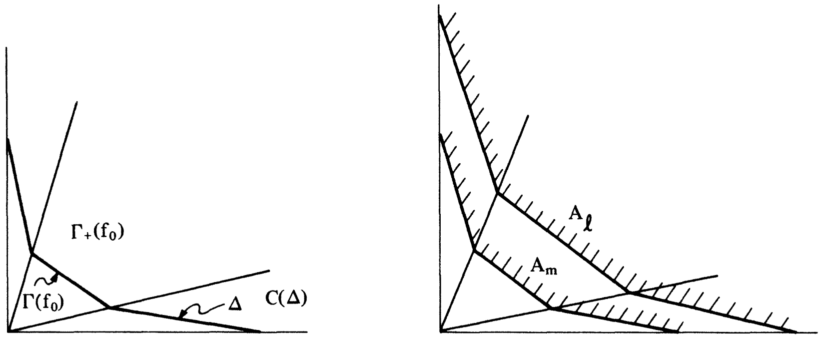


Fig. 1

For any monomial x^{α}, we define $\mathrm{fil}(x^{\alpha}) = \phi(\alpha)$. This extends to a filtration on $k[\mathbf{x}_n]$, $k[[\mathbf{x}_n]]$, and $\mathscr{C}_x$ (via Taylor expansions) by defining

$$\mathrm{fil}(\textstyle\sum c_{\alpha} x^{\alpha}) = \min\{\phi(\alpha)\colon c_{\alpha} \neq 0\}.$$

It also extends to $\mathscr{C}_{x,u}$ by defining $\mathrm{fil}(x^{\alpha} u^{\beta}) = \mathrm{fil}(x^{\alpha})$. We denote the set of g with $\mathrm{fil}(g) \geqq l$ in $k[\mathbf{x}_n]$, $k[[\mathbf{x}_n]]$, $\mathscr{C}_x$, or $\mathscr{C}_{x,u}$ by A_l, $\mathscr{A}_l$, $\mathscr{A}_{l,x}$, *or* $\mathscr{A}_{l,x,u}$, respectively.

By its definition, this filtration has the property that on any cone $C(\Delta)$, it is given by assigning appropriate weights to the x_i. Furthermore, it has the additional property (see [8]) that $\mathrm{fil}(x^{\alpha} x^{\beta}) = \mathrm{fil}(x^{\alpha}) + \mathrm{fil}(x^{\beta})$ if and only if α and β are contained in a common closed cone.

Because of the identification of a monomial x^{α} with $\alpha \in \mathbb{R}^n_+$, we will frequently abuse notation and speak of x^{α} belonging to $\bar{C}(\Delta)$ when we mean that the associated $\alpha \in C(\Delta)$. Also, we can define rings $k[\bar{\Delta}]$, $k[[\bar{\Delta}]]$, etc., to consist of polynomials, power series, etc., with non-zero monomials in $C(\bar{\Delta})$. For these rings, the associated Newton filtration is the restriction of $\{A_l\}$, $\{\mathscr{A}_l\}$, etc. For these rings, it corresponds exactly to assigning appropriate weights to the x_i.

The rings $k[[x_n]]$, $k[[\bar{\Delta}]]$, $\mathscr{C}_x$, and $\mathscr{C}_{x,u}$ are all local rings and also have filtrations associated to the powers of the ideals m_n, m_{Δ}, m_x, and $m_x \mathscr{C}_{x,u}$. In each case, the Newton filtration is equivalent to the filtration by powers of these ideals (i.e. for example in $k[[\mathbf{x}_n]]$, for each l there are k_i so that

$$\mathscr{A}_l \supset m_n^{k_1} \quad \text{and} \quad m_n^l \supset \mathscr{A}_{k_2}).$$

A useful tool in working with these filtered rings is the following simple extension of Nakayama's lemma.

Lemma 1.1 (Filtered version of Nakayama's lemma). *Let $(R, \{I_j\})$ be a filtered local ring (with say $I_1 \subsetneqq R$) and let $(M, \{M_j\})$ be a finitely generated filtered R-module ($I_j \cdot M_k \subset M_{j+k}$) with $\{M_j\}$ equivalent to the filtration $\{I_1^j \cdot M\}$. If $(N, \{N_j\})$ is a submodule of $(M, \{M_j\})$ (so $N_j \subseteq M_j$ all j), then*

$$M_j = N_j + M_{j+1} \quad \textit{all} \quad j \geqq 0 \quad \textit{implies} \quad M = N.$$

Proof.

$$M = M_0 = N_0 + M_1 = N + M_1$$
$$= N + N_1 + M_2 = N + M_2.$$

Thus, by induction

$$M = N + M_k \quad \text{for all } k \geqq 0.$$

By the equivalence of $\{M_k\}$ and $\{I_1^k \cdot M\}$ there is a k so that $I_1 \cdot M \supset M_k$. Hence

$$M = N + I_1 M.$$

Since $I_1 \subset \mathscr{m}$, the maximal ideal of R, Nakayama's lemma implies $M = N$. $\square$

As with any filtration, we can define for a germ g in any of these rings, its initial part, in(g), to be that part of the Taylor expansion of g whose terms have filtration = fil(g) and so that fil(g − in(g)) > fil(g). For a face $\Delta \in \Gamma(f_0)$, we let in$(g)|_{\bar{\Delta}}$ denote the terms remaining after further removing from in(g) terms not in $C(\bar{\Delta})$.

Finally, we also mention the special case of semi-weighted homogeneous germs. For such germs $f_0: k^n, 0 \to k, 0$ we can assign weights $\mathrm{wt}(x_i) = a_i$ and obtain the associated weight filtration where $\mathrm{fil}(x^{\alpha}) = \sum \alpha_i a_i$. Then, f_0 is semi-weighted homogeneous if for some such weighting in(f_0) still defines an isolated singularity. For such semi-weighted homogeneous germs, we can replace the Newton filtration by the weight filtration and the results we shall describe will also follow for it.

§ 2. A Sufficient Filtration Condition for Topological Triviality

In preparation for the specific conditions which we will describe in later sections, we begin by giving general sufficient conditions that a deformation be topologically trivial. We consider a germ $f_0: k^n, 0 \to k, 0$ which is semi-fit. Then, the vertices of the Newton boundary $\Gamma(f_0)$ will in general only have rational coordinates. We say that the level $\mathscr{A}_l$ of the Newton filtration is *fit* if all of the vertices of $\phi^{-1}(l)$ are lattice points of $\mathbb{R}^n_+$ (here ϕ is the function used to define the Newton filtration). If $\mathscr{A}_l$ is fit, then by the linearity of the Newton filtration on cones, $\mathscr{A}_{lr}$ is also fit for integers $r \geqq 1$. Also, if f_0 is already fit, then $\mathscr{A}_{lm}$ is fit for all integers $l \geqq 1$. For $\mathscr{A}_l$ which is fit, we let

$$\mathrm{ver}(\mathscr{A}_l) = \{x^{\boldsymbol{\alpha}}: \boldsymbol{\alpha} \text{ is a vertex of } \phi^{-1}(l)\}.$$

We also define a filtration on θ_n associated to that defined on $\mathscr{C}_x$. Specifically (with $\mathscr{A}_{l+1}\{\partial/\partial x_i\}$ denoting the $\mathscr{A}_{l+1}$-module generated by the $\partial/\partial x_i$, $i = 1, \ldots, n$), we let

$$\mathscr{V}_{l,x} = \left\{\zeta \in \mathscr{A}_{l+1}\left\{\frac{\partial}{\partial x_i}\right\}: \zeta(\mathscr{A}_k) \subset \mathscr{A}_{l+k}\right\}.$$

Then $\{\mathscr{V}_l\}$ is a filtration on θ_n which makes θ_n a filtered $\mathscr{C}_x$-module. We can analogously define a filtration $\{\mathscr{V}_{l,x,u}\}$ on $\theta(\pi_{n,r})$ which makes it a filtered $\mathscr{C}_{x,u}$ module.

For example, we have for the Newton filtration

$$\mathscr{A}_{l,x}\left\{x_i\frac{\partial}{\partial x_i}\right\}\subset\mathscr{V}_{l,x}.$$

If $\{\mathscr{A}_{l,x}\}$ is the weight filtration for a semi-weighted homogeneous germ f_0, then

$$\sum_{i=1}^{n}\mathscr{A}_{l+\mathrm{wt}(x_i)}\frac{\partial}{\partial x_i}\subset\mathscr{V}_{l,x}.$$

We let $\mathscr{V}(f)=\{\zeta(f)\colon \zeta\in\mathscr{V}\}$.

By a deformation $f\colon k^{n+r},0\to k,0$ of $f_0\colon k^n,0\to k,0$ being topologically trivial, we mean that there is a germ of a homeomorphism $\psi\colon k^{n+r},0\to k^{n+r},0$ of the form $\psi(x,u)=(\bar{\psi}(x,u),u)$ such that

$$f\circ\psi(x,u)=f_0(x).$$

Then, we have the sufficient condition for topological triviality.

Theorem 1. *Suppose that $f\colon k^{n+r},0\to k,0$ is a deformation of a semi-fit germ $f_0\colon k^n,0\to k,0$. Then a sufficient condition that f be a topologically trivial deformation is that there exists a fit $\mathscr{A}_l$ so that*

$$\mathrm{ver}(\mathscr{A}_l)\cdot\frac{\partial f}{\partial u_i}\subset\mathscr{V}_{l,x,u}(f),\qquad 1\leqq i\leqq r \tag{*}$$

Remark. Even if f_0 is not semi-fit but is semi-weighted homogeneous, then the result holds using the weight filtration.

An important special case is given by the following result.

Theorem 2. *Suppose that $f\colon k^{n+r},0\to k,0$ is a deformation of the fit germ $f_0\colon k^n,0\to k,0$. Then, a sufficient condition that f is a topologically trivial deformation of f_0 is that there exists an l so that*

$$\mathrm{ver}(\mathscr{A}_{lm})\cdot\frac{\partial f}{\partial u_i}\subset\mathscr{A}_{l\,m,x,u}\left\{x_j\frac{\partial f}{\partial x_j}\right\},\qquad 1\leqq i\leqq r \tag{**}$$

$\Big($again recall that $\mathscr{A}_{l\,m,x,u}\left\{x_j\frac{\partial f}{\partial x_j}\right\}$ denotes the $\mathscr{A}_{l\,m,x,u}$-module generated by the $x_j\frac{\partial f}{\partial x_j}$, $1\leqq j\leqq n\Big)$.

Remark. We will also show that condition (**), in fact, implies Teissier's condition (c) [13]; hence, these deformations will, in fact, satisfy the Whitney conditions.

We will call the conditions (*) or (**) the *filtration conditions*. In the sections that follow, one of our principal goals will be to derive simple conditions in terms of the deformed terms $\partial f/\partial u_i$ and the Newton filtration of f_0 which will imply the filtration conditions. However, before doing so, we first give the proofs of these results in the next section.

§ 3. Topological Triviality via Controlled Vector Fields

Here we use the method of solving the localized equations for triviality using controlled vector fields to obtain topological triviality (see [5] and [11]). Specifically, we use the terminology and notation of [5]. Although in these papers triviality of unfoldings was considered, the method likewise applies to deformations $f\colon k^{n+r},0\to k,0$ of a germ $f_0\colon k^n,0\to k,0$.

We seek to solve the localized equations

$$-\rho\frac{\partial f}{\partial u_i}=\xi_i'(f) \qquad 1\leqq i\leqq r \tag{3.1}$$

where $\rho\colon k^{n+r},0\to\mathbb{R}_+$ is a (C^∞) control function, so that $\rho^{-1}(0)=\{0\}\times k^r$ and $\xi_i'\colon k^{n+r},0\to Tk^n$ is a smooth germ of a vector field. Also, we require that $\rho^{-1}\xi_i'$ be controlled by ρ along $\partial/\partial u_i$. This requires that:

i) $\rho^{-1}\xi_i'$ extends continuously in a neigborhood of 0 to be zero on $\{0\}\times k^r$

ii) in a neighborhood of 0 there are constants C_i so that

$$|\rho^{-1}\xi_i'(\rho)|\leqq C_1\cdot\rho \quad\text{and}\quad \left|\frac{\partial\rho}{\partial u_i}\right|\leqq C_2\cdot\rho.$$

If ρ is independent of u, then ii) is equivalent to $|\xi_i'(\rho)|\leqq C_1\cdot\rho^2$.

If we can solve (3.1) with vector fields ξ_i' controlled by ρ, then f is a topologically trivial deformation. Thus, to prove theorem 1, we will show that the filtration condition implies that we can solve the localized equations (3.1) using controlled vector fields. If $\mathrm{ver}(\mathscr{A}_l)=\{x^\alpha\}$ then we may write

$$-x^\alpha\frac{\partial f}{\partial u_i}=\sum_{j=1}^{n}\xi_{ij}^{(\alpha)}\frac{\partial f}{\partial x_i}=\xi_i^{(\alpha)}(f).$$

Multiplying by x^α, or its conjugate if $k=\mathbb{C}$, and summing over $x^\alpha\in\mathrm{ver}(\mathscr{A}_l)$ we obtain

$$-(\textstyle\sum|x^\alpha|^2)\dfrac{\partial f}{\partial u_i}=\textstyle\sum\xi_{ij}'\dfrac{\partial f}{\partial x_i}. \tag{3.2}$$

We extend the filtration so that $\mathrm{fil}(\bar{x}^\alpha)=\mathrm{fil}(x^\alpha)$ in the complex case. Then, we let

$$\rho=\sum|x^\alpha|^2 \quad\text{and}\quad \xi_i'=\sum\xi_{ij}'\frac{\partial}{\partial x_j}.$$

From the filtration properties of the $\xi_{ij}^{(\alpha)}$, we obtain

$$\mathrm{fil}(\xi_{ij}')\geqq\mathrm{fil}(\xi_{ij}^{(\alpha)})+\mathrm{fil}(x^\alpha)\geqq 2l+1.$$

Also, since $\xi_i'=\sum\bar{x}^\alpha\,\xi_i^{(\alpha)}$

$$\mathrm{fil}(\xi_i'(\rho))\geqq\min_{x^\alpha\in\mathrm{ver}(\mathscr{A}_l)}\{\mathrm{fil}(\bar{x}^\alpha\cdot\xi_i^{(\alpha)}(\rho))\}. \tag{3.3}$$

Then

$$\mathrm{fil}(\xi_i^{(\alpha)}(\rho))\geqq\min_{x^\beta\in\mathrm{ver}(\mathscr{A}_l)}\{\mathrm{fil}(\bar{x}^\beta\cdot\xi_i^{(\alpha)}(x^\beta)\}.$$

By the properties of $\xi_i^{(\alpha)}$, $\operatorname{fil}(\xi_i^{(\alpha)}(x^{\beta}))\geqq 2l$. Thus,

$$\operatorname{fil}(\xi_i^{(\alpha)}(\rho))\geqq l+2l=3l.$$

From (3.3) we obtain

$$\operatorname{fil}(\xi_i'(\rho))\geqq l+3l=4l.$$

As ρ is independent of u, to establish i) and ii) it is sufficient to show:

i) $\|\xi_i'\|\leqq C\rho^{1+\delta}$ in a neighborhood of 0 for some $\delta>0$. Then $\|\rho^{-1}\xi_i'\|\leqq C\rho^{\delta}$ on the neighborhood minus $\{0\}\times k^r$. This implies that it extends continuously to be zero on $\{0\}\times k^r$.

ii) $|\xi_i'(\rho)|\leqq C_1\cdot\rho^2$.

Both conditions will follow from the following lemmas guaranteeing the boundedness of functions satisfying filtration conditions.

For the first lemma, we consider the case where $\rho\colon k^n,0\to\mathbb{R}_+$ is any continuous function. Let $\mathbb{Q}_+^n=\mathbb{Q}^n\cap\mathbb{R}_+^n$. We define

$B=\{\boldsymbol{\alpha}\in\mathbb{Q}_+^n$: $\exists$ a constant $C>0$ and a neighborhood U of 0 (which depends on $\boldsymbol{\alpha}$) such that $|x^{\boldsymbol{\alpha}}|\leqq C\cdot\rho$ on $U\}$.

Note that $x^{\boldsymbol{\alpha}}$ may only be defined as a complex number even though $k=\mathbb{R}$; however, this is allowed for the definition of B.

The first lemma describes the structure of B.

Lemma 3.4. *Given ρ and B as above, then B is a convex set (in $\mathbb{Q}_+^n$) closed under addition by elements of $\mathbb{Q}_+^n$.*

Proof. For convexity, consider $\boldsymbol{\alpha},\boldsymbol{\beta}\in B$ so that $|x^{\boldsymbol{\alpha}}|\leqq C_1\cdot\rho$ on U_1 and $|x^{\boldsymbol{\beta}}|\leqq C_2\cdot\rho$ on U_2. Let $\lambda\in\mathbb{Q}$, with $0\leqq\lambda\leqq 1$, and define $\gamma=\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}$. Then, on $U=U_1\cap U_2$ we have

$$|x^{\gamma}|=|x^{\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}}|=|x^{\boldsymbol{\alpha}}|^{\lambda}\,|x^{\boldsymbol{\beta}}|^{1-\lambda}\leqq(C_1\rho)^{\lambda}(C_2\cdot\rho)^{1-\lambda}.$$

Hence, on U, $|x^{\gamma}|\leqq C\cdot\rho$ where $C=C_1^{\lambda}C_2^{1-\lambda}$.

Secondly, if $\boldsymbol{\alpha}\in B$ and $\boldsymbol{\beta}\in\mathbb{Q}_+^n$, then $|x^{\boldsymbol{\alpha}}|\leqq C_1\cdot\rho$ on U_1 and $|x^{\boldsymbol{\beta}}|\leqq C_2$ on some neighborhood U_2 of 0. Thus

$$|x^{\boldsymbol{\alpha}+\boldsymbol{\beta}}|=|x^{\boldsymbol{\alpha}}|\,|x^{\boldsymbol{\beta}}|\leqq C_1C_2\cdot\rho \quad\text{on}\quad U_1\cap U_2.\ \square$$

Next, we return to the situation of interest to us, namely $\rho=\sum|x^{\boldsymbol{\alpha}}|^2$ where $x^{\boldsymbol{\alpha}}\in\operatorname{ver}(\mathscr{A}_l)$ for a fit $\mathscr{A}_l$.

Lemma 3.5. *If $g\in\mathscr{A}_{m,x,u}$, there is a neighborhood U of 0 in k^{n+r} and a constant $C>0$ so that*

$$|g|\leqq C\cdot\rho^{m/2l}\quad\textit{on}\quad U.$$

Proof. Raising the desired inequality to the $2l$-th power, we see it is sufficient to assume $g\in\mathscr{A}_{2lm,x,u}$ and show

$$|g|\leqq C\cdot\rho^{m}\quad\text{on}\quad U.$$

Since $\mathscr{A}_{2lm}$ as an ideal is generated by a finite number of monomials x^γ in $k[[x_n]]$, it follows by Taylor's theorem that $\mathscr{A}_{2lm,x,u}$ is also generated by a finite number of monomials x^γ in $\mathscr{C}_{x,u}$. It is enough to verify the result for the monomials x^γ for then it is also true for any finite linear combination $\sum g_\gamma x^\gamma$ with $g_\gamma \in \mathscr{C}_{x,u}$ by just bounding each g_γ on some neighborhood by a constant.

Applying Lemma 3.4 to ρ^m, we see that B contains $(x^\alpha)^{2m}$ for $x^\alpha \in \mathrm{ver}(\mathscr{A}_l)$. By the linearity of the Newton filtration on cones, it follows that

$$\mathrm{ver}(\mathscr{A}_{2lm}) = \{(x^\alpha)^{2m}\colon\ x^\alpha \in \mathrm{ver}(\mathscr{A}_l)\}.$$

Hence, the $x^\gamma \in B$. The result follows. $\square$

Lastly, we complete the verification of the conditions i) and ii). We have $\mathrm{fil}(\xi'_{ij}) \geqq 2l+1$. Thus, by Lemma 3.5,

$$|\xi'_{ij}| \leqq C_i \cdot \rho^{\frac{2l+1}{2l}} = C_i \rho^{1+\delta} \quad \text{with} \quad \delta = (2l)^{-1} \quad \text{on} \quad U_i.$$

Hence

$$\|\xi'_i\| \leqq \sum_j |\xi'_{ij}| \leqq C \cdot \rho^{1+\delta} \quad \text{on} \quad U$$

where $C = \sum C_i$ and $U = \bigcap U_i$. Also, $\mathrm{fil}(\xi'_i(\rho)) \geqq 4l$, so

$$|\xi_i(\rho)| \leqq C_1 \cdot \rho^{\frac{4l}{2l}} = C_1 \cdot \rho^2 \qquad \text{on some } U.$$

This completes the proof of the topological triviality. Theorem 2 is then a special case of Theorem 1. However, conversations with Andrew Duplessis led us to realize that the form the localized equations take for Theorem 2, in fact, implies that the deformation satisfies Teissier's condition (c). We obtain the localized equation

$$-\rho \frac{\partial f}{\partial u_i} = \sum g_{ij} x_j \frac{\partial f}{\partial x_j}$$

where $\xi'_i = \sum g_{ij} x_j \dfrac{\partial}{\partial x_j}$ and $\mathrm{fil}(g_{ij}) \geqq 2l$. Thus

$$\rho \cdot \left|\frac{\partial f}{\partial u_i}\right| \leqq n \cdot \max_j |g_{ij}| \cdot \max_j |x_j| \cdot \max_j \left|\frac{\partial f}{\partial x_j}\right|.$$

By Lemma 3.5, there is a common constant $C_1 > 0$ so that

$$|g_{ij}| \leqq C_1 \rho \quad \text{on some } U \qquad 1 \leqq j \leqq n.$$

Also, $\max_j |x_j| \leqq \|x\|$, $\max_j \left|\dfrac{\partial f}{\partial x_j}\right| \leqq \|\mathrm{grad}_x f\|$. Thus

$$\rho \cdot \left|\frac{\partial f}{\partial u_i}\right| \leqq C \cdot \rho \cdot \|x\| \ \|\mathrm{grad}_x f\|$$

dividing by ρ for $x \neq 0$, we obtain the inequality which is also valid for $x=0$

$$\left|\frac{\partial f}{\partial u_i}\right| \leqq C\,\|x\|\,\|\mathrm{grad}_x f\| \quad \text{on} \quad U.$$

This is exactly the analytic version of Teissier's condition (c).

§ 4. Patching Data, and Non-Degeneracy

The local patching data which we define give us a method for determining the effects of a deformation on the various cones in the Newton polyhedron. Let $f_0: k^n, 0 \to k, 0$ be a semi-fit germ.

Definition 4.1. A set of *local patching data* for f_0 consists of: for each face $\Delta \subset \Gamma(f_0)$, a set $\{\zeta_{\Delta,i}\}$ of germs of polynomial vector fields in θ_n such that

i) $\zeta_{\Delta,i} = \sum g_{\Delta,i,j} x_j \frac{\partial}{\partial x_j}$

ii) $\{\mathrm{in}\,(\zeta_{\Delta,i}(f_0))|_{\bar{\Delta}}\}$ generates an ideal of finite codimension in $k[[\bar{\Delta}]]$.

Remark 1. The $\{C(\Delta)\}$ give a polyhedral decomposition of $\mathbb{R}^n_+$ which is the cone on the polyhedral decomposition of $\Gamma(f_0)$. The local patching data can be viewed as being defined via $\{C(\Delta)\}$ or $\{\Delta\}$.

Remark 2. It is sufficient to construct local patching data for the top dimensional faces and restrict these to obtain data for the lower dimensional faces.

The existence of local patching data is guaranteed by the following

Proposition 4.2. *If* $\left\{x_i \frac{\partial f_0}{\partial x_i}\right\}$ *generates an ideal of finite codimension in* $\mathscr{C}_x$, *then there exists a set of local patching data for* f_0.

Proof. If $\left\{x_i \frac{\partial f_0}{\partial x_i}\right\}$ generates an ideal of finite codimension in $\mathscr{C}_x$, then $\left\{x_i \frac{\partial \tilde{f}_0}{\partial x_i}\right\}$ generates an ideal of finite codimension in $k[[\mathbf{x}_n]]$, where $\tilde{f}_0$ denotes the Taylor expansion of f_0. If $\mathscr{m}_n$ denotes the maximal ideal of $k[[\mathbf{x}_n]]$, then by assumption there is an r so that $\mathscr{m}_n^r \subset \left(x_i \frac{\partial \tilde{f}_0}{\partial x_i}\right)$, $\left(\text{where } \left(x_i \frac{\partial \tilde{f}_0}{\partial x_i}\right)\right.$ denotes the ideal generated by the $\left. x_i \frac{\partial \tilde{f}_0}{\partial x_i}\right)$.

Then, $\mathscr{m}_n^r \cap k[[\bar{\Delta}]]$ is an ideal of finite codimension in $k[[\bar{\Delta}]]$. As $k[[\bar{\Delta}]]$ is Noetherian, there exists a finite set of monomials $h_{\Delta,i} \in \left(x_i \frac{\partial \tilde{f}_0}{\partial x_i}\right)$ which generate $\mathscr{m}_n^r \cap k[[\bar{\Delta}]]$. We can write

$$h_{\Delta,i} = \sum \tilde{g}_{\Delta,i,j} x_j \frac{\partial \tilde{f}_0}{\partial x_j}.$$

Let $g_{\Delta,i,j}$ denote $\tilde{g}_{\Delta,i,j}$ truncated above filtration level $= \mathrm{fil}(h_{\Delta,i})$, and define

$$\zeta_{\Delta,i} = \sum_j g_{\Delta,i,j} x_j \frac{\partial}{\partial x_j}.$$

Then,

$$h_{\Delta,i}=\mathrm{in}(h_{\Delta,i})|_{\bar{\Delta}}=\mathrm{in}(\zeta_{\Delta,i}(f_0))|_{\bar{\Delta}}. \quad \square$$

If f_0 has an (algebraically) isolated singularity, then $\{\partial f_0/\partial x_i\}$ generate an ideal of finite codimension. Hence, for $\left\{x_i\frac{\partial f_0}{\partial x_i}\right\}$ to fail to generate an ideal of finite codimension, the restriction of f_0 to some coordinate hyperplane $x_i=0$ must fail to have an algebraically isolated singularity. However, for almost all hyperplanes $0\in\Pi\subset k^n, f_0|\Pi$ has an algebraically isolated singularity. Thus, for almost all choices of coordinates $\mathbf{x}$, $\left\{x_i\frac{\partial f_0}{\partial x_i}\right\}$ will generate an ideal of finite codimension.

Thus, by a generic choice of coordinates, we can ensure that f_0 has local patching data. At the same time we point out that a generic choice of coordinates gives the least interesting form for the Newton polyhedron. The most desirable course is to choose coordinates so that $\left\{x_i\frac{\partial f_0}{\partial x_i}\right\}$ generates an ideal of finite codimension, yet the Newton polyhedron "presents the richest picture of the singularity".

When $\left\{x_i\frac{\partial f_0}{\partial x_i}\right\}$ generates an ideal of finite codimension, we can algebraically construct a canonical choice for local patching data. If we construct the graded algebra associated to $\{k[[\bar{\Delta}]], \mathscr{A}_{l|\bar{\Delta}}\}$, we obtain $k[\bar{\Delta}]\xrightarrow{\sim}gr(k[[\bar{\Delta}]])$.

Given the ideal $I=\left(x_i\frac{\partial f_0}{\partial x_i}\right)$, we associate the ideal I_Δ in $k[\bar{\Delta}]$ consisting of the images of $g\in I$ in $gr(k[[\bar{\Delta}]])$ by the map $g\mapsto\mathrm{in}(g)|_{\bar{\Delta}}$. As $k[\bar{\Delta}]$ is Noetherian, I_Δ is generated by a finite number of images $\{g_i=\zeta_{\Delta,i}(f_0)\}$. These $\{\zeta_{\Delta,i}\}$ are the canonical patching data associated to I_Δ. It is not necessary to find the canonical patching data to apply our results.

We recall the definition of a non-degenerate germ f_0 [8] as one which is fit and for which $\left\{\mathrm{in}\left(x_i\frac{\partial f_0}{\partial x_i}\right)\Big|_{\bar{\Delta}}\right\}$ generate an ideal of finite codimension in $k[[\bar{\Delta}]]$ for each closed face $\bar{\Delta}\subset\Gamma(f_0)$. It is immediate that $\left\{x_i\frac{\partial}{\partial x_i}\right\}$ is a set of local patching data (in fact, canonical data) for each $\Delta\subset\Gamma(f_0)$. We shall also say that a germ f_0 is *non-degenerate on a face* $\Delta\subset\Gamma(f_0)$ if $\left\{\mathrm{in}\left(x_i\frac{\partial f_0}{\partial x_i}\right)\Big|_{\bar{\Delta}}\right\}$ generates an ideal of finite codimension in $k[[\bar{\Delta}]]$. More generally we shall see how local patching data also allow us to handle the situation where f_0 is degenerate. In fact, the data $\{\zeta_{\Delta,i}\}$ arise from the relations between the $\mathrm{in}\left(x_i\frac{\partial f_0}{\partial x_i}\right)\Big|_{\bar{\Delta}}$.

§ 5. Jump Conditions

In this section, we fix a set of local patching data. Then, to measure the effect of local patching data applied to the filtration, we use the notion of jumps. For local patching data $\{\zeta_{\Delta,i}\}$ for f_0, we define

$$\operatorname{jump}(\zeta_{\Delta,i})=\operatorname{fil}(\zeta_{\Delta,i}(f_0))-\min_j\left\{\operatorname{fil}(g_{\Delta,i,j}x_j\frac{\partial f_0}{\partial x_j}\right\}.$$

Also, for each face $\Delta\subset\Gamma(f_0)$ we define

$$\operatorname{jump}(\Delta)=\max_i\{\operatorname{jump}(\zeta_{\Delta,i})\}.$$

Lastly, for each vertex $v\in\Gamma(f_0)$ we define

$$\operatorname{jump}(v)=\max\{\operatorname{jump}(\Delta)\colon\ \Delta\subset\operatorname{star}(v)\}.$$

Here recall that for any face $\Delta\in\Gamma(f_0)$, star(Δ) is defined as for any polyhedral decomposition $\operatorname{star}(\Delta)=\bigcup\{\Delta'\in\Gamma(f_0)\colon\ \Delta\subset\bar{\Delta}'\}$. In particular, star$(\Delta)$ is an open subset of $\Gamma(f_0)$. In the non-degenerate case, with local patching data $\left\{x_i\frac{\partial}{\partial x_i}\right\}$, all jumps are zero.

Lastly, for each vertex $v\in\Gamma(f_0)$, we wish to be able to "pull" any polynomial germ in x, $\psi\in\mathscr{C}_u[\mathbf{x}_n]$ into star(v) and measure the filtration of its image. For this we let $\boldsymbol{\alpha}\in C(v)$. Given any $\phi\in\mathscr{C}_{x,u}$ we write its Taylor expansion with respect to $\mathbf{x}$ as $\sum b_{\boldsymbol{\beta}}(u)x^{\boldsymbol{\beta}}$ with $b_{\boldsymbol{\beta}}(u)\in\mathscr{C}_u$. We define

$$\operatorname{supp}(\phi)=\{\Delta\in\Gamma(f_0)\colon\quad \text{there is a } \boldsymbol{\beta}\in C(\Delta) \text{ with } b_{\boldsymbol{\beta}}\neq 0\}.$$

Then, we have the following lemma which allows us to contract polynomials into stars of faces.

Lemma 5.1. (contraction lemma). *Given any polynomial germ $\psi\in\mathscr{C}_u[\mathbf{x}_n]$, there is an l so that for $l'\geqq l$*

i) $\operatorname{supp}((x^{\boldsymbol{\alpha}})^{l'}\cdot\psi)\subset\operatorname{star}(v)$

ii) $\operatorname{fil}((x^{\boldsymbol{\alpha}})^{l'}\cdot\psi)-\operatorname{fil}((x^{\boldsymbol{\alpha}})^{l'})$ *is independent of $l'\geqq l$.*

We then define for the vertex v of $\Gamma(f_0)$,

$$\operatorname{fil}_v(\psi)=\operatorname{fil}((x^{\boldsymbol{\alpha}})^{l'}\cdot\psi)-\operatorname{fil}((x^{\boldsymbol{\alpha}})^{l'})\qquad l'\geqq l.$$

Proof. It is clearly enough to establish the result for a monomial in x, say $x^{\boldsymbol{\beta}}$.

For i), if we consider the plane P spanned by $C(v)$ and $x^{\boldsymbol{\beta}}$, then star(v) will intersect P in an open cone C containing $C(v)$ in its interior. See Fig. 2. Then, by the "parallelogram law for multiplication", $(x^{\boldsymbol{\alpha}})^l\cdot x^{\boldsymbol{\beta}}$ lies on the ray parallel to $C(v)$ through $x^{\boldsymbol{\beta}}$. Thus, for all large l if will lie in C.

For ii), note that once $\operatorname{supp}((x^{\alpha})^l\cdot\psi)\subset\operatorname{star}(v)$, then by the property of Newton filtrations described in § 1

$$\operatorname{fil}((x^{\boldsymbol{\alpha}})^k\cdot(x^{\boldsymbol{\alpha}})^l\cdot\psi)=\operatorname{fil}((x^{\boldsymbol{\alpha}})^k)+\operatorname{fil}((x^{\boldsymbol{\alpha}})^l\cdot\psi).$$

Then, ii) follows immediately. □

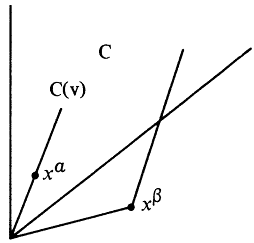


Fig. 2

For any germ $\psi \in \mathscr{C}_{x,u}$, we can also define $\mathrm{fil}_v(\psi)$. If $\psi_{(L)}$ denotes the Taylor expansion of ψ in x up through filtration $\leqq L$, then $\mathrm{fil}_v(\psi_{(L)})$ is independent of L for all sufficiently large L. Then, $\mathrm{fil}_v(\psi)$ is defined to be this common value.

Remark. $\mathrm{fil}_v(\ )$ can also be defined as the piecewise linear extension to $\mathbb{R}^n_+$ of $\mathrm{fil}_v | \mathrm{star}(v)$.

We are now ready to define the jump conditions which will be sufficient to establish the filtration conditions of § 2.

Definition 5.3. A germ ψ satisfies a *simple jump condition* for f_0 if

$$\mathrm{fil}_v(\psi) \geqq m + \mathrm{jump}(v) \qquad \text{for all vertices } v \in \Gamma(f_0).$$

To illustrate the simple jump condition, consider a semi-fit germ f_0 with $\Gamma(f_0)$ as shown in Fig. 3 so that f_0 is non-degenerate on each face except Δ and $\mathrm{jump}(\Delta) = l$. Then, monomials satisfying the simple jump condition are those in the shaded region.

This is a purely numerical condition. It must be supplemented, in general, with a more specific higher order jump condition.

Definition 5.5. A germ $\psi \in \mathscr{C}_{x,u}$ satisfies a *general jump condition* for a deformation f of f_0 if for each vertex $v \in \Gamma(f_0)$, there is an $x^\alpha \in C(v)$ of filtration $= m_v$

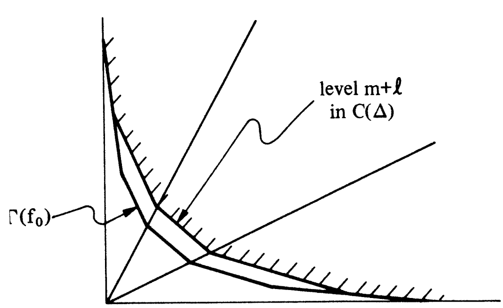


Fig. 3

and a polynomial germ $\psi' \in \mathscr{A}_{m_v, x, u}\left\{x_i \dfrac{\partial f}{\partial x_i}\right\}$ so that

$$\operatorname{fil}_v(x^{\boldsymbol{\alpha}} \cdot \psi - \psi') \geqq m_v + m + \operatorname{jump}(v).$$

There is an inductive procedure for establishing the general jump condition. If $\operatorname{fil}_v(\psi) < m + \operatorname{jump}(v)$, then we inductively seek to "stably" reduce the inequality. First seek a $x^{\boldsymbol{\alpha}_1} \in C(v)$ and a $\xi_1 = \sum \xi_{1,i} x_i \dfrac{\partial}{\partial x_i}$ with $\xi_{1,i} \in \mathscr{A}_{m_1, x, u}$ where $m_1 = \operatorname{fil}(x^{\boldsymbol{\alpha}_1})$ so that

$$\operatorname{fil}_v(x^{\boldsymbol{\alpha}_1} \cdot \psi - \xi_1(f_0)) > \operatorname{fil}_v(\psi) + m_1.$$

If the remainder also satisfies

$$\operatorname{fil}_v(\xi_1(f) - \xi_1(f_0)) > \operatorname{fil}_v(\psi) + m_1,$$

then we replace ψ by $\psi_1 = x^{\boldsymbol{\alpha}_1} \psi - \xi_1(f)$. Then,

$$\operatorname{fil}_v(\psi_1) - m_1 > \operatorname{fil}_v(\psi).$$

Inductively, we repeat the construction finding $x^{\boldsymbol{\alpha}_i}$, ξ_i, and ψ_i so

$$\operatorname{fil}_v(\psi_i) - m_i > \operatorname{fil}_v(\psi_{i-1}).$$

Eventually,

$$\operatorname{fil}_v(\psi_k) - m_k > m + \operatorname{jump}(v).$$

Then, we let $\boldsymbol{\alpha} = \sum \boldsymbol{\alpha}_i$, $m_v = \sum m_i$, and $\xi = \sum x^{\boldsymbol{\beta}_i} \xi_i$ where $\boldsymbol{\beta}_i = \sum_{j=i+1}^{k} \boldsymbol{\alpha}_j$; $\xi(f)$ is our desired ψ'.

In what follows, we will generically refer to either form of jump condition as simply a *jump condition.*

To place the ideas of local patching data and jumps into perspective, we indicate how they naturally arise in a spectral sequence which is a slight variation of the one used by Arnold [1]. Let $U = \mathscr{A}_0\left\{x_i \dfrac{\partial}{\partial x_i}\right\}$ and I be the ideal $\left(x_i \dfrac{\partial \tilde{f}_0}{\partial x_i}\right)$ as defined earlier. Define $d\colon U \to I$ by $d(\xi) = \xi(f_0)$. Then, U has a filtration with $U_p = \mathscr{A}_p\left\{x_i \dfrac{\partial}{\partial x_i}\right\}$, and I has a natural filtration which can be renumbered so that d preserves filtration. We consider the spectral sequence of the complex

$$0 \to U \xrightarrow{d} I \to 0, \qquad \{d_p^r\colon S_p^r \to I_{p+r}^r\}$$

$$(S_p^0 = U_p/U_{p+1}, I_p^0 = (I \cap \mathscr{A}_{p+m})/(I \cap \mathscr{A}_{p+m+1}), \text{ and } d^0 = d).$$

Then, S_p^r is generated by vector fields of filtration p, which when applied to f_0 jump filtration r. Thus, if $d^r = \sum_p d_p^r$, $\operatorname{Im}(d^r)$ is generated by $\xi(f_0)$ with $\operatorname{jump}(\xi) = r$. Also, if r_0 is the minimum integer such that $I_p^{r_0}$ is non-zero for only finitely many p, then there exists a set of local patching data $\{\zeta_{\Delta,i}\}$ with $\max\{\operatorname{jump}(\zeta_{\Delta,i})\} = r_0$.

§ 6. Sufficient Conditions via Local Patching Data

We are now ready to describe the principal result which states how the jump conditions for the local patching data for a semifit germ f_0 allow us to derive the global filtration condition.

Theorem 3. *Let $f: k^{n+r}, 0 \to k, 0$ be a deformation of f_0. Suppose that f_0 has local patching data $\{\zeta_{\Delta,i}\}$ so that:*

i) $\operatorname{fil}(\zeta_{\Delta,i}(f)) = \operatorname{fil}(\zeta_{\Delta,i}(f_0))$ *for all* Δ, i

ii) *the $\partial f/\partial u_i$ satisfy a jump condition.*

Then, f satisfies the filtration condition (**) *and hence is a topologically trivial deformation.*

As an immediate corollary we have

Corollary 1. *Suppose that $f_0: k^n, 0 \to k, 0$ is a non-degenerate germ with an isolated singularity. Any deformation of f_0 of non-decreasing Newton filtration* ($\operatorname{fil}(f) = \operatorname{fil}(f_0)$) *is a topologically trivial deformation.*

As a corollary of this, we have a topological proof of a result of Kouchnirenko.

Corollary 2. *If $f_0: \mathbb{C}^n, 0 \to \mathbb{C}, 0$ is a non-degenerate function with isolated singularity then the topological type of f_0 and hence its Milnor number only depends on its Newton polyhedron.*

Over the reals we can still say

Corollary 3. *For non-degenerate germs $f_0: \mathbb{R}^n, 0 \to \mathbb{R}, 0$ the topological type of f_0 only depends on its connected component in the subspace of all such germs with the same Newton polyhedron.*

Also, as a corollary, we have an estimate of the order of topological determinacy in terms of the filtration.

Corollary 4. *Let f_0 be a germ with local patching data $\{\zeta_{\Delta,i}\}$. If $s = m + \max\{\text{jump}(v): v \in \Gamma(f_0)\}$, then f_0 is s-topologically right determined with respect to the filtration $\{\mathscr{A}\}$. Specifically, if $f_1 \equiv f_0 \bmod \mathscr{A}_{s+1}$, then f_1 and f_0 are topologically right equivalent.*

In the semi-weighted homogeneous case, we obtain

Corollary 5. *If f_0 is a semi-weighted homogeneous germ, then any deformation which does not decrease weight is topologically trivial.*

Before turning to the proofs of the results, we first illustrate these results with several examples.

§ 7. Examples

By Kouchnirenko, we expect almost all germs with a given Newton polyhedron to be non-degenerate. For example, if we examine Laufer's list of minimally elliptic surface singularities [9] (or Arnold's list of unimodal and bimodal singularities [2]), we see that with the exception of eight families, all other singularities are either weighted homogeneous or non-degenerate (possibly after a change of coordinates to make them fit). Thus, we can conclude that, excluding these eight families, deformations of the other minimally elliptic surface singularities of non-decreasing Newton Filtration (using new coordinates where necessary) are topologically trivial.

For the eight families, we must construct local patching data. We consider the first such family where this occurs.

Example 7.1. $f_0(x, y, z)=(x^2+y^3)^2-z^2+x^a y^b$ where $3a+2b=n+11$ ($n\geqq 2$). The Newton polyhedron has just one face.

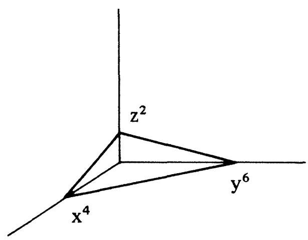


Fig. 4

This is degenerate on the top dimensional face and on the edge $\{x^4, y^6\}$. We see that on these faces, in $\left(x\,\dfrac{\partial f_0}{\partial x}\right)$ and in $\left(y\,\dfrac{\partial f_0}{\partial y}\right)$ have the common factor x^2+y^3. There is a relation between these initial parts obtained from

$$\zeta=3y^3\cdot x\,\frac{\partial}{\partial x}-2x^2\cdot y\,\frac{\partial}{\partial y}.$$

Then,

$$\zeta(f_0)=(3ay^3-2bx^2)\,x^a y^b.$$

This ζ together with $x\,\dfrac{\partial}{\partial x}$, $y\,\dfrac{\partial}{\partial y}$, and $z\,\dfrac{\partial}{\partial z}$ forms local patching data for these two faces. Then

$$\operatorname{fil}(\zeta(f_0))=n+11+6=n+17 \quad\text{and}\quad \operatorname{jump}(\zeta)=(n+17)-18=n-1.$$

Hence, a deformation f must satisfy

(1) $\operatorname{fil}(\zeta(f))=n+17$ and

(2) $\operatorname{fil}\left(\dfrac{\partial f}{\partial u_i}\right)\geqq n-1+12=n+11.$

In fact, in this case (2) implies (1); hence, we conclude that any deformation of f_0 by terms of filtration $\geqq n+11$ is topologically trivial.

As a second example, we consider

Example 7.2. $f_0(x, y)=y^5+2x^2y^3+x^4y+x^6.$

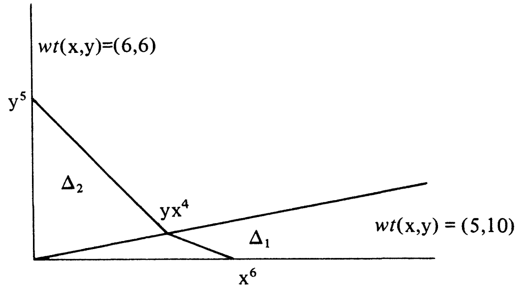


Fig. 5

This has the Newton polygon shown in Fig. 5 where the Newton filtration on each cone can be described using the indicated weights for x and y. Then $m=30$.

Also,

$$x\frac{\partial f_0}{\partial x}=4x^2y(x^2+y^2)+6x^6$$

$$y\frac{\partial f_0}{\partial y}=y(5y^2+x^2)(y^2+x^2).$$

On Δ_1 and the vertices, f_0 is non-degenerate; thus, for these faces we may use $\left\{x\frac{\partial}{\partial x}, y\frac{\partial}{\partial y}\right\}$ for our patching data. On Δ_2, f_0 is degenerate; in fact, $\operatorname{in}\left(x\frac{\partial f_0}{\partial x}\right)\Big|_{\Delta_2}$ and $\operatorname{in}\left(y\frac{\partial f_0}{\partial y}\right)\Big|_{\Delta_2}$ have the common factor y^2+x^2. We obtain a relation between these using

$$\zeta=y^2(5y^2+x^2)\cdot x\frac{\partial}{\partial x}-4y^2x^2\cdot y\frac{\partial}{\partial y}.$$

Then,

$$\zeta(f_0)=30y^4x^6+6y^2x^8.$$

In this case, $\operatorname{in}(\zeta(f_0))=\zeta(f_0)$, $\operatorname{fil}(\zeta(f_0))=60$ and $\operatorname{jump}(\zeta)=60-54=6$. Thus, for Δ_2 we use local patching data $\left\{x\frac{\partial}{\partial x}, y\frac{\partial}{\partial y}, \zeta\right\}$. We must have $\operatorname{fil}(\zeta(f))\geqq 60$. This only requires in $C(\Delta_1)$ that we must use terms of filtration $\geqq 30$; and in $C(\Delta_2)$, we must use terms of filtration $\geqq 36$. Also, it is readily checked that such terms satisfy $\operatorname{fil}_{y^5}(\ )\geqq 36$. It remains to check the jump condition at x^4y. Since x^4y is a vertex of both faces, $\operatorname{fil}_{x^4y}(\phi)=\operatorname{fil}(\phi)$. Thus, we must have $\operatorname{fil}(\phi)\geqq 36$. Thus, deformations involving terms except x^6 and x^7 satisfy the hypothesis of Theorem 3. More generally, for a deformation

$$f=f_0+u_1x^6+u_2x^7+h(x,y,u_3,\ldots)$$

with $\mathrm{fil}(h) \geqq 36$, we claim that both x^6 and x^7 satisfy a general jump condition. As $y\frac{\partial f}{\partial y} \equiv y\frac{\partial f_0}{\partial y} \bmod \mathscr{A}_{36,x,u}$, both $x^4 y \cdot x^6 - x^6 \cdot y\frac{\partial f}{\partial y}$ and $x^4 y \cdot x^7 - x^7 \cdot y\frac{\partial f}{\partial y} \in \mathscr{A}_{66,x,u}$. Thus, the general deformation f, and hence any deformation involving x^6, x^7 and terms of filtration $\geqq 36$ is topologically trivial.

Example 7.3. (Modified Briançon-Speder example [4]).

$$f_0(x, y, x) = x^5 + y^7 z + y^8 + z^{15}.$$

This f_0 is non-degenerate with Newton polyhedron shown in Fig. 6.

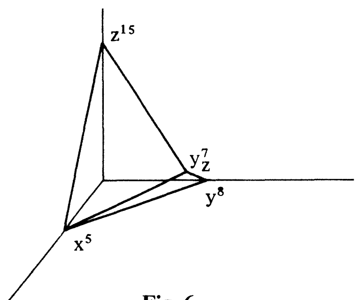


Fig. 6

The term xy^6 *lies* below the Newton polyhedron. However, we still claim the deformation $f_0 + uxy^6$ is topologically trivial. In fact, f_0 is semi-weighted homogeneous, and with respect to the weighting $wt(x, y, z) = (3, 2, 1)$, $wt(xy^6) = 15$. Thus, by Corollary 5, the deformation is topologically trivial. A similar situation can arise for other deformations under the Newton polyhedron, where it is necessary to replace a non-degenerate function by a semi-fit part which still has an isolated singularity and apply Theorem 1 to prove topological triviality.

§ 8. The Push-off Lemma and a First Consequence

Using the polyhedral structure of $\Gamma(f_0)$, we use an inductive procedure to establish the filtration conditions from the jump conditions. We inductively show that, modulo higher filtration, we can push the support of germs for which the filtration condition fails onto higher dimensional faces. After n-steps, we have pushed it into a higher filtration and then we can use the filtered version of Nakayama's lemma (Lemma 1.1). In the most general form, we have to be able to do this "stably", i.e., after first multiplying by sufficiently high powers of vertices.

Lemma 8.1 (push-off lemma). *Let* $f\colon k^{n+r}, 0 \to k, 0$ *be a deformation of the germ* $f_0\colon k^n, 0 \to k, 0$ *so that* f_0 *has local patching data* $\{\zeta_{\Delta,i}\}$ *and* $\mathrm{fil}(\zeta_{\Delta,i}(f)) = \mathrm{fil}(\zeta_{\Delta,i}(f_0))$ *for all* $\Delta \in \Gamma(f_0)$ *and all i. Then, there exists an N so that if* $\psi \in \mathscr{C}_{x,u}$

has $\mathrm{fil}(\psi)=q>N$ *and* $\mathrm{supp}(\psi)\subset\Delta$ *then there is a* $\psi_1\in\mathscr{A}_{q,x,u}$ *with* $\mathrm{supp}(\psi_1)\subset\mathrm{star}(\Delta)-\Delta$ *and*

$$\psi-\psi_1\in\sum_i \mathscr{A}_{q_i,x,u}\,\zeta_{\Delta,i}(f)+\mathscr{A}_{q+1,x,u}$$

where $q_i=q-\mathrm{fil}(\zeta_{\Delta,i}(f_0))$.

Proof. Let $\mathscr{A}_l$ be a fit level for the Newton filtration of f_0, then $\mathscr{A}_{lr}$ is fit for all integers $r\geqq 1$. From the definition of local patching data, it follows that there is r so that the ideal I_Δ generated by $\{\mathrm{in}(\zeta_{\Delta,i}(f_0))|_{\bar{\Delta}}\}$ contains $\mathscr{A}_{rl}\cap k[[\bar{\Delta}]]$ for all faces $\Delta\in\Gamma(f_0)$. Then, there are two main steps.

Step 1. We let $\{\mathscr{A}_{l,u}\}$ denote the Newton Filtration on $\mathscr{C}_u[[\mathbf{x}_n]]$ analogous to that defined on $\mathscr{C}_{x,u}$. We first wish to give for each face $\Delta\in\Gamma(f_0)$ a set of generators for $\mathscr{A}_{rl,u}\cap\mathscr{C}_u[[\bar{\Delta}]]$ which restrict to a set of monomial generators for $\mathscr{A}_{rl}$ when $u=0$.

We note that $\mathscr{A}_{rl}\cap k[[\bar{\Delta}]]$ is generated as an ideal in $k[[\bar{\Delta}]]$ by a finite number of monomials. Then, it is easily seen that $\mathscr{A}_{rl,u}\cap\mathscr{C}_u[[\bar{\Delta}]]$ is also generated as an ideal in $\mathscr{C}_u[[\bar{\Delta}]]$ by the same set of monomials. Thus, by Nakayama's lemma it is sufficient to give a set of elements in $\mathscr{A}_{rl,u}\cap\mathscr{C}_u[[\bar{\Delta}]]$ which restrict to the set of monomial generators of $\mathscr{A}_{rl}\cap k[[\bar{\Delta}]]$ when $u=0$.

We may represent the monomial generators of $\mathscr{A}_{rl}\cap k[[\bar{\Delta}]]$ in the form

$$x^{\boldsymbol{\alpha}_i}=\sum_j h_{ij}\,\mathrm{in}(\zeta_{\Delta,j}(f_0))|_{\bar{\Delta}}.$$

As $\mathrm{fil}(\ )|k[[\bar{\Delta}]]$ comes from a weighting and $\mathrm{in}(\ )|_{\bar{\Delta}}$ is weighted homogeneous with respect to this weighting, we may also assume the h_{ij} are weighted homogeneous with respect to this weighting. Thus, $\mathrm{in}(h_{ij})=h_{ij}$ and $\mathrm{fil}(h_{ij})=\mathrm{fil}(x^{\boldsymbol{\alpha}_i})-\mathrm{fil}(\zeta_{\Delta,i}(f_0))$. Then, the set of generators for $\mathscr{A}_{rl,u}\cap\mathscr{C}_u[[\bar{\Delta}]]$ is given by $\{\psi_i\}$ where

$$\psi_i=\sum_j h_{ij}\,\mathrm{in}(\zeta_{\Delta,i}(f))|_{\bar{\Delta}}.$$

Note the assumption $\mathrm{fil}(\zeta_{\Delta,i}(f))=\mathrm{fil}(\zeta_{\Delta,i}(f_0))$ ensures that $\psi_i\in\mathscr{A}_{rl,u}$.

Step 2. By Step 1, $x^{\boldsymbol{\alpha}}\in\mathrm{ver}(\mathscr{A}_{rl})\cap k[[\bar{\Delta}]]$ has a representation

$$x^{\alpha}=\sum h_i^{(\boldsymbol{\alpha})}\,\mathrm{in}(\zeta_{\Delta,i}(f))|_{\bar{\Delta}}(\mathrm{mod}\,\mathscr{A}_{rl+1,u}) \tag{8.2}$$

with $h_i^{(\boldsymbol{\alpha})}\in k[\bar{\Delta}]$. Then, by the contraction lemma, we may suppose

$$x^{\boldsymbol{\alpha}}-\sum h_i^{(\alpha)}\,\zeta_{\Delta,i}(f)\equiv g_{\boldsymbol{\alpha}}(\mathrm{mod}\,\mathscr{A}_{rl+1,u})$$

with $\mathrm{supp}(g_{\boldsymbol{\alpha}})\subset\mathrm{star}(x^{\boldsymbol{\alpha}})-\bar{\Delta}$ (otherwise, we could multiply (8.2) by $(x^{\alpha})^L$ for some L and replace rl by a larger Rl). Denote $\sum h_i^{(\boldsymbol{\alpha})}\zeta_{\Delta,i}(f)$ by $\phi_{\Delta}^{(\boldsymbol{\alpha})}$.

Then, $\mathrm{ver}(\mathscr{A}_{rl})\cap C(\bar{\Delta})$ generates an ideal of finite codimension in $k[[\bar{\Delta}]]$ for each face $\Delta\subset\Gamma(f_0)$. We suppose that N is chosen so that $\mathscr{A}_N\cap k[[\bar{\Delta}]]$ is contained in the ideal generated by $m_\Delta\cdot(\mathrm{ver}(\mathscr{A}_{rl})\cap C(\bar{\Delta}))$ in $k[[\bar{\Delta}]]$.

To prove the result, it is enough to verify it for monomials x^γ with $\gamma\in C(\Delta)$ and $\mathrm{fil}(x^\gamma)=q>N$. We may write $x^\gamma=x^{\boldsymbol{\beta}}\cdot x^{\boldsymbol{\alpha}}$ with $x^{\boldsymbol{\alpha}}\in\mathrm{ver}(\mathscr{A}_{rl})\cap C(\bar{\Delta})$, and $\boldsymbol{\beta}\in C(\Delta)$. As $\gamma\in C(\Delta)$, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are not both contained in a closed subface of $\bar{\Delta}$

(hence $\operatorname{star}(x^{\beta}) \cap \operatorname{star}(x^{\alpha}) = \operatorname{star}(\Delta)$, where for x^{β} not a vertex, $\operatorname{star}(x^{\beta})$ denotes $\operatorname{star}(\Delta')$ for Δ' the smallest face with $C(\Delta')$ containing β). Thus,

$$x^{\beta}(x^{\alpha} - \phi_{\Delta}^{(\alpha)}) \equiv x^{\beta} g'_{\alpha} (\operatorname{mod} \mathscr{A}_{q+1,u})$$

and

$$\operatorname{supp}(x^{\beta} g'_{\alpha}) \subset \operatorname{star}(x^{\beta}) \cap \operatorname{star}(x^{\alpha}) - \Delta = \operatorname{star}(\Delta) - \Delta.$$

Here g'_{α} consists of the terms of g_{α} of filtration q and $\operatorname{supp} \subset \operatorname{star}(x^{\beta})$, otherwise by the properties of the Newton filtration, the filtration of x^{β} times such a term will have filtration $> q$.

Lastly,

$$x^{\beta} \phi_{\Delta}^{(\alpha)} = \sum x^{\beta} h_i^{(\alpha)} \zeta_{\Delta,i}(f);$$

and by the properties of $\operatorname{fil}(\ ) \mid C(\bar{\Delta})$

$$\operatorname{fil}(h_i^{(\alpha)}) = \operatorname{fil}(x^{\alpha}) - \operatorname{fil}(\zeta_{\Delta,i}(f))$$

so

$$\begin{aligned} \operatorname{fil}(x^{\beta} h_i^{(\alpha)}) &= \operatorname{fil}(x^{\beta}) + \operatorname{fil}(x^{\alpha}) - \operatorname{fil}(\zeta_{\Delta,i}(f)) \\ &= \operatorname{fil}(x^{\gamma}) - \operatorname{fil}(\zeta_{\Delta,i}(f)). \end{aligned}$$

Hence,

$$x^{\gamma} \equiv x^{\beta} g'_{\alpha} (\operatorname{mod} \sum \mathscr{A}_{q_i,u} \zeta_{\Delta,i}(f) + \mathscr{A}_{q+1,u}).$$

Since $x^{\beta} h_i^{(\alpha)} \in \mathscr{C}_{x,u}$, this also yields

$$x^{\gamma} \equiv x^{\beta} g'_{\alpha} (\operatorname{mod} \sum \mathscr{A}_{q_i,x,u} \zeta_{\Delta,i}(f) + \mathscr{A}_{q+1,x,u}). \quad \square$$

As a consequence of the push-off lemma we have

Proposition 8.3. *Let f_0 have local patching data $\{\zeta_{\Delta,i}\}$ and let f be a deformation of f_0 so that $\operatorname{fil}(\zeta_{\Delta,i}(f)) = \operatorname{fil}(\zeta_{\Delta,i}(f_0))$ for all $\Delta \subset \Gamma(f_0)$ and all i. Then, there is an N so that if $\psi \in \mathscr{C}_{x,u}$ with $\operatorname{fil}(\psi) = q \geqq N$ then*

$$\psi \in \sum_{\Delta,i} \mathscr{A}_{q_{\Delta,i},x,u} \cdot \zeta_{\Delta,i}(f)$$

where $q_{\Delta,i} = q - \operatorname{fil}(\zeta_{\Delta,i}(f)) > 0$.

Proof. Let N be as in the push-off lemma and so that $N > \operatorname{fil}(\zeta_{\Delta,i}(f))$ all Δ, i. For $q \geqq N$, we have the $\mathscr{C}_{x,u}$-modules $\mathscr{A}_{q+j,x,u}$ and $\sum \mathscr{A}_{j_{\Delta,i},x,u} \cdot \zeta_{\Delta,i}(f)$ (summed over all Δ, j) where $j_{\Delta,i} = q + j - \operatorname{fil}(\zeta_{\Delta,i}(f))$.

By inductively applying the push-off lemma to faces of $\Gamma(f_0)$ of increasing dimensions, we obtain

$$\mathscr{A}_{q+j,x,u} \subset \sum \mathscr{A}_{j_{\Delta,i},x,u} \cdot \zeta_{\Delta,i}(f) + \mathscr{A}_{q+j+1,x,u} \quad j \geqq 0. \tag{8.4}$$

Since the filtration $\{\mathscr{A}_{q+j,x,u}\}$ is equivalent to that of $\{\mathscr{A}^{q+j}_{1,x,u}\} = \{m_x^{q+j} \cdot \mathscr{C}_{x,u}\}$, we can apply the filtered version of Nakayama's lemma to (8.4) to conclude

$$\mathscr{A}_{q+j,x,u} = \sum \mathscr{A}_{j_{\Delta,i},x,u} \cdot \zeta_{\Delta,i}(f). \quad \square$$

§ 9. Deriving the Filtration Conditions from Jump Conditions

We need a "stable version" of the push-off lemma.

Lemma 9.1 (stable push-off lemma). *Let f_0 be a germ with local patching data $\{\zeta_{\Delta,i}\}$ and let f be a deformation with $\mathrm{fil}(\zeta_{\Delta,i}(f))=\mathrm{fil}(\zeta_{\Delta,i}(f_0))$ for all $\Delta\subset\Gamma(f_0)$ and all i. Let $\psi\in\mathscr{C}_u[\mathbf{x}_n]$ have $\mathrm{fil}(\psi)=q$ and $\mathrm{supp}(\psi)\subset\mathrm{star}(v)$ for v a vertex of $\Gamma(f_0)$. Also, let $x^\alpha\in C(v)$ with $\mathrm{fil}(x^\alpha)=l$. Then, there is a $r>0$ and a polynomial*

$$\psi_1\in\sum\mathscr{A}_{q_{\Delta,i},x,u}\cdot\zeta_{\Delta,i}(f)\qquad\textit{where}\quad q_{\Delta,i}=q+rl-\mathrm{fil}(\zeta_{\Delta,i}(f_0))$$

(*summed over $\Delta\subset\mathrm{star}(v)$*) *so that*

i) $(x^\alpha)^r\psi\equiv\psi_1(\mathrm{mod}\,\mathscr{A}_{q+rl+1,x,u})$
ii) $\mathrm{supp}((x^\alpha)^r\cdot\psi-\psi_1)\subset\mathrm{star}(v)$.

Proof. We will inductively apply the push-off lemma for a r_i and $\psi_1^{(i)}$ so that

$$\mathrm{supp}((x^\alpha)^{r_i}\cdot\psi-\psi_1^{(i)})_{(q+r_il)})\subset\mathrm{star}(v)-\bigcup\{\Delta\text{ of dim }<i\}.$$

Here $g_{(r)}$ denotes the terms of g of filtration $\leq r$. Then, r_n and $\psi_1^{(n)}$ will be our desired r and ψ_1.

We begin the induction with $r_0=0$ and $\psi_1^{(0)}=0$. Suppose the result is true for $i<j$ (where $j\leq n$). Let $\psi'=(x^\alpha)^{r_{j-1}}\psi-\psi_1^{(j-1)}$. There is an s so that $\mathrm{fil}(\psi')+sl>N$, where N is that obtained from the push-off lemma. Then, by the push-off lemma, combined with Proposition 8.3, there is, in fact, a polynomial

$$\psi''\in\sum\mathscr{A}_{p_{\Delta,i},x,u}\cdot\zeta_{\Delta,i}(f)$$

with

$$p_{\Delta,i}=\mathrm{fil}(\psi')+sl-\mathrm{fil}(\zeta_{\Delta,i}(f))$$

so that

$$\mathrm{supp}((x^\alpha)^s\cdot\psi'-\psi'')_{(\mathrm{fil}(\psi')+sl)})\subset\mathrm{star}(x^\alpha)-\bigcup\{\Delta\text{ of dim }\leq j\}.$$

However, $\mathrm{supp}((x^\alpha)^s\cdot\psi'-\psi'')$ may not be in $\mathrm{star}(v)$. Then, by the contraction lemma there is an s' so that

$$\mathrm{supp}((x^\alpha)^{s'}((x^\alpha)^s\cdot\psi'-\psi''))\subset\mathrm{star}(v).$$

We let $r_j=r_{j-1}+s+s'$ and

$$\psi_1^{(j)}=(x^\alpha)^{s+s'}\psi_1^{(j-1)}+(x^\alpha)^{s'}\cdot\psi''.\quad\square$$

Corollary 9.2. *Let f_0 be a germ with local patching data $\{\zeta_{\Delta,i}\}$ and let f be a deformation with $\mathrm{fil}(\zeta_{\Delta,i}(f))=\mathrm{fil}(\zeta_{\Delta,i}(f_0))$ for all $\Delta\subset\Gamma(f_0)$ and all i. Let $\psi\in\mathscr{C}_u[\mathbf{x}_n]$ with $q_v=\mathrm{fil}_v(\psi)$ for all $v\in\Gamma(f_0)$. Then, for $x^\alpha\in C(v)$ with $\mathrm{fil}(x^\alpha)=l$ and $r>0$, there exists an $s>0$ so that*

$$(x^\alpha)^s\cdot\psi\in\sum\mathscr{A}_{q_{\Delta,i},x,u}\cdot\zeta_{\Delta,i}(f)+\mathscr{A}_{q_v+sl+r,x,u}$$

with $q_{\Delta,i}=q_v+sl-\mathrm{fil}(\zeta_{\Delta,i}(f))$ and the sum is over $\Delta\subset\mathrm{star}(v)$ and the i's for these Δ's.

Proof. We repeatedly apply the stable push-off lemma beginning with $(x^{\alpha})^{s_1}\cdot\psi$ so that $\operatorname{supp}((x^{\alpha})^{s_1}\cdot\psi)\subset\operatorname{star}(v)$ (by the contraction lemma). After r applications, we have the desired result. □

We are now in a position to prove Theorem 3.

Proof of Theorem 3

Let $\mathscr{A}_l$ be fit and let $\operatorname{ver}(\mathscr{A}_l)=\{x^{\alpha}\}$. Next, let N be as in proposition 8.3. We consider a germ $\psi\in\mathscr{C}_{x,u}$ which satisfies a jump condition. We let ψ_1 denote the terms of filtration $<N$ in the Taylor expansion of ψ (in the x-coordinates). Then $\psi_2=\psi-\psi_1$ has filtration $\operatorname{fil}(\psi_2)\geqq N$. First, consider the case of the simple jump condition.

Let $x^{\alpha}\in\operatorname{ver}(\mathscr{A}_l)$ with say $x^{\alpha}\in C(v)$ for $v\in\Gamma(f_0)$. By Corollary 9.2, there is an $s>0$ so that

$$(x^{\alpha})^s\cdot\psi_1\in\sum\mathscr{A}_{q_{\Delta,i},x,u}\cdot\zeta_{\Delta,i}(f)+\mathscr{A}_{sl+q_v+N} \tag{9.3}$$

where the sum is over $\Delta\subset\operatorname{star}(v)$ and all i for each such Δ and $q_{\Delta,i}=q_v+sl-\operatorname{fil}(\zeta_{\Delta,i}(f))$. Then

$$\begin{aligned}\operatorname{fil}(\zeta_{\Delta,i}(f))&=\operatorname{fil}(\zeta_{\Delta,i}(f_0))\\&=\min\left\{\operatorname{fil}\left(g_{\Delta,i,j}x_j\frac{\partial f_0}{\partial x_j}\right)\right\}+\operatorname{jump}(\zeta_{\Delta,i})\\&=\min\left\{\operatorname{fil}(g_{\Delta,i,j})+\operatorname{fil}\left(x_j\frac{\partial f_0}{\partial x_j}\right)\right\}+\operatorname{jump}(\zeta_{\Delta,i})\end{aligned}$$

because f_0 being semi-fit guarantees that $x_j\dfrac{\partial f_0}{\partial x_j}$ will have a non-zero term of filtration $=m$ in every maximal dimensional closed face. Thus,

$$\operatorname{fil}(\zeta_{\Delta,i}(f))\leqq\min\{\operatorname{fil}(g_{\Delta,i,j})\}+m+\operatorname{jump}(v).$$

Hence,

$$q_{\Delta,i}\geqq q_v+sl-m-\operatorname{jump}(v)-\min\{\operatorname{fil}(g_{\Delta,i,j})\}. \tag{9.4}$$

If from 9.3, we write

$$(x^{\alpha})^s\cdot\psi_1=\sum h_{\Delta,i}\zeta_{\Delta,i}(f)+\psi'$$

then,

$$(x^{\alpha})^s\cdot\psi_1=\sum_j\Big(\sum_{\Delta,i}h_{\Delta,i}g_{\Delta,i,j}\Big)x_j\frac{\partial f}{\partial x_j}+\psi'.$$

Also,

$$\begin{aligned}\operatorname{fil}(h_{\Delta,i}g_{\Delta,i,j})&\geqq\operatorname{fil}(h_{\Delta,i})+\operatorname{fil}(g_{\Delta,i,j})\\&\geqq q_{\Delta,i}+\operatorname{fil}(g_{\Delta,i,j})\\&\geqq sl+q_v-m-\operatorname{jump}(v)\\&\geqq sl\end{aligned}$$

by (9.4) and the simple jump condition on ψ (recall $q_v=\operatorname{fil}_v(\psi_1)$). Thus,

$$(x^\alpha)^s \cdot \psi_1 \equiv \psi' \bmod \mathscr{A}_{sl,x,u}\left\{x_j \frac{\partial f}{\partial x_j}\right\}.$$

Then,

$$(x^\alpha)^s \psi = (x^\alpha)^s \psi_1 + (x^\alpha)^s \cdot \psi_2.$$

Hence,

$$(x^\alpha)^s \cdot \psi \equiv \psi' + (x^\alpha)^s \cdot \psi_2 \bmod \mathscr{A}_{sl,x,u}\left\{x_j \frac{\partial f}{\partial x_j}\right\}. \tag{9.5}$$

Also,

$$\mathrm{fil}\,(\psi' + (x^\alpha)^s \psi_2) \geqq sl + L > sl + N.$$

Thus, by Proposition 8.3

$$\psi' + (x^\alpha)^s \cdot \psi_2 \in \sum \mathscr{A}_{r_{\Delta,i},x,u}\, \zeta_{\Delta,i}(f)$$

where $r_{\Delta,i} \geqq sl + N - \mathrm{fil}(\zeta_{\Delta,i}(f))$. By Proposition 8.3 $N \geqq \mathrm{fil}(\zeta_{\Delta,i}(f))$, thus, $r_{\Delta,i} \geqq sl$. Hence,

$$\psi' + (x^\alpha)^s \psi_2 \in \mathscr{A}_{sl,x,u}\left\{x_j \frac{\partial f}{\partial x_j}\right\}.$$

By (9.5), we conclude

$$(x^\alpha)^s \cdot \psi \in \mathscr{A}_{sl,x,u}\left\{x_j \frac{\partial f}{\partial x_j}\right\}.$$

For the general jump condition, the argument is similar. We let $(x^\alpha) \cdot \psi_1 - \psi' = \psi''$, where $\psi' \in \mathscr{A}_{m_v,x,u}\left\{x_j \frac{\partial f}{\partial x_j}\right\}$ for $m_v = \mathrm{fil}(x^\alpha)$ and $\mathrm{fil}_v(\psi'') \geqq m_v + m + \mathrm{jump}(v)$. By an argument similar to that for the simple jump condition, there is an s so that

$$(x^\alpha)^s \cdot \psi'' \in \mathscr{A}_{(s+1)m_v,x,u}\left\{x_j \frac{\partial f}{\partial x_j}\right\}. \tag{9.6}$$

Then,

$$\begin{aligned}(x^\alpha)^{s+1} \cdot \psi &= (x^\alpha)^{s+1} \psi_1 + (x^\alpha)^{s+1} \psi_2 \\ &= (x^\alpha)^s \psi' + (x^\alpha)^s \cdot \psi'' + (x^\alpha)^{s+1} \psi_2.\end{aligned}$$

Again

$$(x^\alpha)^s \psi' + (x^\alpha)^{s+1} \psi_2 \in \mathscr{A}_{(s+1)m_v,x,u}\left\{x_j \frac{\partial f}{\partial x_j}\right\}.$$

This together with (9.6) gives the result. □

§ 10. μ-constant Deformations of Real Singularities

The structure of real isolated singularities $f_0: \mathbb{R}^n, 0 \to \mathbb{R}, 0$ barely begins to approach the rich structure for complex isolated singularities which results from the Milnor fibration. Nonetheless, we have seen that even without the richer topological structure, deformations of real germs are topologically trivial under conditions analogous to when complex germs are (however, not conversely, for example, any Pham-Brieskorn polynomial $e_1 x_1^{a_1} + \ldots + e_n x_n^{a_n}$ over $\mathbb{R}$ defines a germ topologically equivalent to either the germ of a submersion or a

Morse singularity). This suggests that there may be an analogue of the Lê-Ramanujam result for real germs. We end our discussion by indicating how results of King [7] yield analogues for real plane curve singularities and surface singularities in $\mathbb{R}^3$.

Theorem 10.1. *Let* $f(x,u)\colon \mathbb{R}^{s+r}, 0 \to \mathbb{R}, 0$ *be a deformation of* $f_0\colon \mathbb{R}^s, 0 \to \mathbb{R}, 0$, $s=2$ *or* 3, *so that the Milnor number* $\mu(f(\cdot,u))$ *is constant in a neighborhood of* 0. *Then,* f *is a topologically trivial deformation.*

Proof. Since $\mu(f(\cdot,u)) = \dim_{\mathbb{R}} \mathscr{C}_x \Big/ \left(\frac{\partial f(\cdot,u)}{\partial x_1}, \ldots, \frac{\partial f(\cdot,u)}{\partial x_s}\right)$, which is the dimension of the local algebra of $\operatorname{grad}_x f\colon \mathbb{R}^s, 0 \to \mathbb{R}^s, 0$, it follows by the upper semicontinuity of the algebraic multiplicity that in a neighborhood of 0 in $\mathbb{R}^{s+r}$, the singular set of f is $\{0\} \times \mathbb{R}^r$. Thus, $\operatorname{grad}_x f\colon \mathbb{R}^s - \{0\} \to \mathbb{R}^s - \{0\}$ is a homotopy for small $u \in \mathbb{R}^r$. Hence, $d(f(\cdot,u)) = \deg(\operatorname{grad}_x f(\cdot,u))$ is independent of u for small u. In the case of plane curve singularities, this is sufficient to determine the structure of the curve singularity. For by a result of Arnold [3], the Euler class of $f(\cdot,u)^{-1}(0) \cap (S^1_\varepsilon \times \{u\})$ (which is a finite set of points) is given by $2(1-d(f(\cdot,u)))$. Thus, locally $f(\cdot,u)^{-1}(0)$ is a cone on this number of points, which is independent of u for u near 0.

However, more generally since the singular set of $f(x,u) = \{0\} \times \mathbb{R}^r$ in a neighborhood of $(0,0)$, by the results of Henry King (Theorem 2 and Corollary 1 of [7]) these deformations for the case of curves and surfaces are topologically trivial. □

References

1. Arnold, V.I.: Spectral sequence for reduction of functions to normal form. Funct. Anal. and Appl. **9**, 251–253 (1975)
2. Arnold, V.I.: Local normal forms of functions. Invent. Math. **35**, 87–109 (1976)
3. Arnold, V.I.: Index of a singular point of a vector field, the Petrovski-Oleinik inequality, and mixed hodge structures. Funct. Anal. and Appl. **12**, 1–11 (1978)
4. Briançon, J., Speder, J.P.: La trivialite topologigue n'implique pas les conditions de Whitney. Comp. Rendus t **280**, (series A) 365 (1975)
5. Damon, J.: Finite determinacy and topological triviality I. Invent. Math. **62**, pp. 299–324 (1980)
6. Damon, J., Gaffney, T.: Topological triviality of deformations of functions and newton filtrations. preliminary announcement
7. King, H.: Topological type in families of singularities. Invent. Math. **62**, 1–13 (1980)
8. Kouchnirenko, A.G.: Polyedres de Newton et Nombres de Milnor. Invent. Math. **32**, 1–31 (1976)
9. Laufer, H.: On minimally elliptic singularities. Amer. Jour. Math. **99**, 1257–1295 (1977)
10. Lê Dũng Tráng, Ramanujam, C.P.: The invariance of Milnor's number implies the invariance of topological type. Amer. Jour. Math. **98**, 67–78 (1976)
11. Looijenga, E.: Semi-universal deformation of a simple elliptic singularity: Part I unimodularity. Topology **16**, 257–262 (1977)
12. Oka, M.: On the bifurcation of the multiplicity and topology of the newton boundary. J. Math. Soc. Japan **31**, 435–450 (1979)
13. Teissier, B.: Cycles evanescents, sections planes, et conditions de Whitney. Singularities a Cargese. Asterisque **7** and **8**, 285–362 (1973)
14. Timourian, J.G.: Invariance of Milnor's number implies topological triviality. Amer. Jour. Math. **99**, 437–446 (1977)

Oblatum 39-I & 25-X-1982

Invent. math. 72, 359-362 (1983)

Inventiones mathematicae

Automorphisms of Representation Finite Algebras

R. Martínez-Villa and J.A. de la Peña

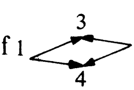
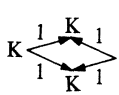
Instituto de Matematicas U.N.A.M., Ciudad Universitaria,
Delegàción Coyoácan, 04510 México, D.F., Mexico

Coverings in representation theory have been successfully used by: Bongartz-Gabriel [2], Gabriel [3], Green [4], Riedtman [6, 7].

Of particular importance is the study of the relations between the coverings of the Auslander-Reiten quiver and the ordinary quiver. In this direction P. Gabriel posed, in the Third International Conference on Representations of Algebras, Puebla México, 1980, the following conjecture:

"Let g be an automorphism of an algebra Λ of finite representation type. For each right module M, denote by M^g the Λ-module with underlying group M and scalar multiplication $(m, \lambda) \to mg^{-1}(\lambda)$. If $S^g \not\simeq S$ for each simple S, then $M^g \not\simeq M$ for each indecomposable M".

The aim of this note is to give a short proof of this statement.

The conjecture is not true for algebras of infinite representation type. Consider, for example, the quiver algebra of  with automorphism g acting on vertices as $g=(12)\,(34)$. $S^g \simeq S$ for each simple but, if M is the module corresponding to the representation , then $M^g \simeq M$.

We will use freely the definitions and notation introduced in [2].

K will denote an algebraically closed field.

For a locally bounded K-category Λ, $\operatorname{mod}_\Lambda$ denotes the category of all finitely generated modules and $\operatorname{ind}_\Lambda$ the full subcategory formed by choosing representatives of the indecomposable modules.

The category $\operatorname{mod}_\Lambda$ has Auslander-Reiten sequences [1], hence we can construct the Auslander-Reiten quiver Γ_Λ.

Let M be an Auslander category, $M \cong \operatorname{ind}_\Lambda$ with Λ locally representation finite. We know by [2] that the mesh category $K(\Gamma_\Lambda)$ is an Auslander category, $K(\Gamma_\Lambda) \cong \operatorname{ind} \Lambda_s$, where Λ_s denotes the full subcategory of $K(\Gamma_\Lambda)$ formed by the projective vertices of Γ_Λ. The Auslander-Reiten quiver of Λ_s is Γ_Λ.

We say that an automorphism g of Λ "moves" the family of indecomposables Λ-modules $\mathscr{C}$ if $M^g \not\simeq M$ for each M in $\mathscr{C}$.

Lemma 1. *Let g be an automorphism of the Auslander category $M \cong \mathrm{ind}_\Lambda$. Let Γ_Λ be the Auslander-Reiten quiver of Λ, and $K(\Gamma_\Lambda) \cong \mathrm{ind}_{\Lambda_s}$. Then:*

a) *g induces an automorphism g_s of $K(\Gamma_\Lambda)$.*

b) *If g moves projectives then so does g_s.*

c) *If M has an indecomposable non projective X with $X^g \cong X$, then $K(\Gamma_\Lambda)$ has a non projective object Y with $Y^{g_s} \cong Y$.*

Proof. $g: \mathrm{ind}_\Lambda \to \mathrm{ind}_\Lambda$ can be extended to an equivalence $g: \mathrm{mod}_\Lambda \to \mathrm{mod}_\Lambda$; as any equivalence g preserves Auslander-Reiten sequences, projective modules and irreducible morphisms, g induces an automorphism of translation quivers $g_s: \Gamma_\Lambda \to \Gamma_\Lambda$. So we have the following facts:

i) g moves indecomposable projectives if and only if g_s does so.

ii) $X^g \cong X$ if and only if for the corresponding vertex $[X]$, $[X]^{g_s} = [X]$.

Hence g_s gives an automorphism of the Riedtmann category $K(\Gamma_\Lambda)$ satisfying b) and c). //

Lemma 2. *Let g: $\Gamma \to \Gamma$ be an automorphism of an Auslander-Reiten quiver such that $g(x) = x$ for some $x \in \Gamma_0$. Let π: $\tilde{\Gamma} \to \Gamma$ be the universal cover of Γ. Then there is a point $y \in \tilde{\Gamma}_0$ and an automorphism $h: \tilde{\Gamma} \to \tilde{\Gamma}$ of translation quives, satisfying:*

a) $h(y) = y$.

b) $\pi h = g \pi$.

c) *If g acts freely on projectives then so does h.*

Proof. This is a trivial consequence of the definition of $\tilde{\Gamma}$ in [2]. //

Lemma 3. *Let M be an Auslander category, $M \cong \mathrm{ind}_\Lambda$ and assume Λ is simply connected (in the sense of* [2], *i.e. Λ is locally representation finite and Γ_Λ is its own universal cover). Then an automorphism $g: M \to M$ acting freely on indecomposable projectives acts freely on all indecomposable modules.*

Proof. As $\Pi_1(\Gamma_\Lambda, x) = 1$, Γ_Λ has no oriented cycles and $M \cong K(\Gamma_\Lambda)$. Suppose $X \in \mathrm{ind}_\Lambda$ is a non projective module such that $X^g \cong X$. We know that g can be extended to an automorphism $g: \mathrm{mod}_\Lambda \to \mathrm{mod}_\Lambda$.

Let $P_1 \xrightarrow{f_1} P_0 \xrightarrow{f_0} X \to 0$ be a minimal projective presentation. Then, $P_1^g \xrightarrow{f_1^g} P_0^g \xrightarrow{f_0^g} X^g \to 0$ is also a minimal projective presentation of $X^g \cong X$. Therefore, $P_1^g \cong P_1$, $P_0^g \cong P_0$. Define P as the full subcategory of ind_Λ formed by representatives of the direct summands of $P_0 \oplus P_1$, so $P^g \cong P$.

There is a full inclusion $\Lambda \otimes_P$: $\mathrm{mod}_P \to \mathrm{mod}_\Lambda$ whose image $\mathscr{C}$ consists of all the Λ-modules Y with minimal projective presentation $Q_1 \to Q_0 \to Y \to 0$ such that $Q_1 \oplus Q_0$ is a summand of P^n for some $n \in \mathbb{N}$. Then $\mathscr{C}$ is stable under the action of g.

Any object in $\mathscr{C}$ is a factor of P^n for some $n \in \mathbb{N}$; therefore we have $\mathrm{supp}\, Y \subset \mathrm{supp}\, P$ for $Y \in \mathrm{Ob}\mathscr{C}$. As Λ is locally representation finite, $\mathscr{C}$ has only a finite number of non isomorphic indecomposable Λ-modules.

Let $\Lambda_0 := \mathrm{End}_\Lambda(P)^{\mathrm{op}}$ be the finite dimensional K-algebra with $\mathrm{mod}_{\Lambda_0} \cong \mathrm{mod}_P$; then Λ_0 is of finite representation type and g "induces" an equivalence g_0: $\mathrm{mod}_{\Lambda_0} \to \mathrm{mod}_{\Lambda_0}$ such that:

a) g_0 acts freely on indecomposable projectives.

b) If φ: $\mathrm{mod}_{\Lambda_0} \to \mathscr{C}$ is the equivalence given above and $\varphi(X') = X$, then X' is indecomposable non projective and $X'^g = X'$.

The Auslander-Reiten quiver Γ_{Λ_0} of Λ_0 has no oriented cycles because otherwise the full inclusion $\mathrm{mod}_{\Lambda_0} \hookrightarrow \mathrm{mod}_\Lambda$ would induce cycles in Γ_Λ. Therefore, for every non projective Λ_0-module Y, there exists an integer n with $\tau^n Y$ projective. In particular $\tau^n X' = Q$ is projective for some integer n. As an automorphism preserves Auslander-Reiten sequences, $Q^{g_0} \cong Q$ which is a contradiction. //

We are now in position to prove the mentioned conjecture of Gabriel.

Theorem 4. *Let $M \cong \mathrm{ind}_\Lambda$ be an Auslander category. Then an automorphism g of M moving the indecomposable projectives, moves all the indecomposable Λ-modules.*

Proof. Assume there is an indecomposable X such that $X^g \cong X$. Let Γ_Λ be the Auslander-Reiten quiver of Λ. By Lemma 1 $K(\Gamma_\Lambda) \cong \mathrm{ind}_{\Lambda_s}$ has an automorphism g_s which acts freely on indecomposable projectives and fixes a non projective vertex $x \in (\Gamma_\Lambda)_0$.

By Lemma 2, there is an automorphism $\tilde{g}_s$: $K(\tilde{\Gamma}_\Lambda) \to K(\tilde{\Gamma}_\Lambda)$, where $\tilde{\Gamma}_\Lambda$ is the universal cover of Γ_Λ, $\tilde{g}_s$ acts freely on projective vertices and fixes a non projective one y.

As $K(\Gamma_\Lambda)$ is an Auslander category, so is $K(\tilde{\Gamma}_\Lambda)$ ([2]). Assume $K(\tilde{\Gamma}_\Lambda) \cong \mathrm{ind}_{\tilde{\Lambda}}$ with $\tilde{\Lambda}$ locally representation finite. The Auslander-Reiten quiver of $\tilde{\Lambda}$ is $\tilde{\Gamma}_\Lambda$ which is its own universal cover; so $\tilde{\Lambda}$ is a simply connected category. But this contradicts Lemma 3. //

This result has some consequences that we will use in a forthcoming publication [5]. Also, it simplifies and generalizes some proofs in [3]. We give an example below.

We recall some definitions appearing in [3]: If M is a locally finite dimensional K-category and G is a group of K-automorphisms of M, we say that the action of G on M is free if for every $1 \neq g \in G$ and any $x \in \mathrm{Ob}\, M$ $gx \neq x$; and it is locally bounded if for every pair (x, y) there are only a finite number of $g \in G$ with $M(x, gy) \neq 0$. When G satisfies both conditions, the category M/G can be defined as a locally finite dimensional category with a covering functor F: $M \to M/G$ such that $Fg = F$ for $g \in G$; it is universal for this property. In this situation we get the following:

Theorem 5. *If M is a locally representation-finite category then so is M/G.*

Proof. As G acts freely on M, it acts freely on the indecomposable projectives of $\mathrm{ind}\, M$ which is an Auslander category. By Theorem 4, G acts freely on $\mathrm{ind}\, M$ and the result follows from Theorem 3.6 of [3]. //

References

1. Auslander, M., Reiten, I.: Representation theory of Artin Algebras IV Comm. in Algebra, 5, (5) (1977) 443–518
2. Bongartz, K., Gabriel, P.: Covering spaces in representation theory. Invent. Math. **65**, 331–378 (1982)
3. Gabriel, P.: The universal cover of a representation-finite algebra. Representations of Algebras. Lecture Notes, Vol. 903, pp. 68–105. Berlin-Heidelberg-NewYork: Springer 1981
4. Green, E.L.: Graded Artin algebras. Preprint (1980)
5. Martínez, R., de la Peña, J.A.: The universal cover of a quiver with relations. Preprint (1982)
6. Riedtmann, Ch.: Algebren. Darstellungsköcher, Überlagerungen und zurück. Comm. Math. Helv. **55**, 199–224 (1980)
7. Riedtmann, Ch.: Representation finite selfinjective algebras of type. An. in Representation Theory II. Lecture Notes, Vol. 832, pp. 449–520. Berlin-Heidelberg-NewYork: Springer 1980

Oblatum 1-VII-1982 & 1-XII-1982

Invent. math. 72, 363-388 (1983)

Inventiones mathematicae

Über das Abelsche Analogon des Lindemannschen Satzes I

G. Wüstholz

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

1. Einleitung und Ergebnisse

Genau vor hundert Jahren, im Jahre 1882, bewies F. Lindemann [7] ein bemerkenswertes Resultat über lineare Unabhängigkeit von Werten der Exponentialfunktion. Insbesondere bewies er das folgende Resultat. Es seien $\alpha_1, \ldots, \alpha_n$ algebraische Zahlen, welche linear unabhängig über den rationalen Zahlen seien. Dann sind die Zahlen

$$e^{\alpha_1}, \ldots, e^{\alpha_n}$$

algebraisch unabhängig über dem Körper der algebraischen Zahlen. Beinahe fünfzig Jahre lang war dies das einzige ernsthafte Resultat über algebraische Unabhängigkeit, bevor Siegel Lindemann's Methode auf sogenannte E-Funktionen ausdehnte und so zum Beispiel algebraische Unabhängigkeit von Werten der Besselfunktion bewies. Siegels Methode wurde später beträchtlich von Shidlovskij erweitert.

Erst kürzlich entwickelten nun D.W. Masser und der Autor [10, 11] eine neue Methode, um die algebraische Unabhängigkeit von Werten einer weiteren Klasse von Funktionen zu beweisen, welche zum Beispiel abelsche Funktionen enthält. Diese Funktionen treten auf als Komponenten der Exponentialabbildung einer algebraischen Gruppe. Die Methode gründet sich auf die sogenannten Nullstellenabschätzungen auf Gruppenvarietäten, welche von D.W. Masser und dem Autor in [9] eingeführt wurden. Zum Beispiel gilt das folgende Resultat. Es sei $\wp$ eine Weierstraßsche elliptische Funktion, welche über $\overline{\mathbb{Q}}$ definiert sei, und N und k seien positive ganze Zahlen mit $N \geqq 3 \cdot 2^{k+2}(k+7)$. Dann sind wenigstens k der Zahlen

$$\wp(\pi), \ldots, \wp(\pi^N)$$

algebraisch unabhängig über $\overline{\mathbb{Q}}$.

In einem Spezialfall verschärfte kürzlich P. Philippon [14, 15] diese Methode und erzielte das folgende Resultat. Sei $\wp$ wie oben die Weierstraßsche elliptische Funktion und u eine komplexe Zahl sowie α eine algebraische Zahl vom Grade $d \geqq 2$. Ist $\wp$ definiert für die Zahlen $u, \alpha u, \ldots, \alpha^{d-1} u$, so sind wenigstens $d/6$ der Zahlen

$$\wp(u), \wp(\alpha u), \ldots, \wp(\alpha^{d-1} u)$$

algebraisch unabhängig über $\mathbb{Q}$. Ist u ein algebraischer Punkt unendlicher Ordnung und besitzt $\wp$ komplexe Multiplikation, so ist Philippon unter Benutzung einer Idee des Autors, die wesentlich für die vorliegende Arbeit ist, in der Lage, die Zahl $d/6$ durch $\max(1, (d-2)/2)$ zu ersetzen.

In dieser Arbeit werden wir ein Resultat über die algebraische Unabhängigkeit von Werten abelscher Funktionen beweisen, das unter gewissen zusätzlichen Voraussetzungen das abelsche Analogon des Lindemannschen Ergebnisses liefert. Dieses Ergebnis wurde in [19] vom Autor angekündigt. Das erste Resultat in dieser Richtung und Schärfe stammt von G.V. Choodnovsky [4], der den elliptischen Fall unseres Resultats für $n \leqq 3$ bewies.

Wir werden nun einige Bezeichnungen einführen, um unser Ergebnis dann anzugeben. Dazu sei A eine einfache abelsche Varietät, die über $\overline{\mathbb{Q}}$ definiert sei und die Dimension d besitze. Der Ring der Endomorphismen $\operatorname{End} A$ von A ist eine Divisionsalgebra und wir setzen $k = (\operatorname{End} A) \otimes \mathbb{Q}$ und $m = [k : \mathbb{Q}]$. Sei $\mathscr{A}$ eine integre Unteralgebra über k von $M_n(k)$ der Dimension n über k. Schließlich sei $\mathcal{O}$ ein Untermodul von $\mathscr{A}$ über $\operatorname{End} A$, der gleichzeitig eine Ordnung in $\mathscr{A}$ sei.

Bezeichnen wir mit $T(A^n)$ den Tangentialraum der abelschen Varietät im neutralen Element der Gruppe, so operiert die Algebra $\mathcal{O}$ in natürlicher Weise auf diesem Vektorraum. Diese Operation kann in der folgenden Weise beschrieben werden. Es sei γ ein Element von $\mathcal{O}$. Da $\mathcal{O}$ eine Ordnung in $\mathscr{A}$ und eine Unteralgebra von $M_n(k)$ ist, kann γ als eine Matrix $\gamma = (\gamma_{ij})$ mit Elementen γ_{ij} aus $\operatorname{End} A$ geschrieben werden. Diese können wir als Endomorphismen von $T(A)$, dem Tangentialraum von A im neutralen Element, auffassen, welche das Periodengitter stabilisieren. Sei $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ in $T(A^n)$ für $\mathbf{z}_i$ in $T(A)$ und $1 \leqq i \leqq n$. Dann setzen wir

$$\gamma \cdot \mathbf{z} = \mathbf{w},$$

wobei $\mathbf{w} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ definiert ist durch

$$\mathbf{w}_i = \gamma_{i,1} \mathbf{z}_1 + \ldots + \gamma_{i,n} \mathbf{z}_n \qquad (1 \leqq i \leqq n).$$

Hierdurch wird eine Operation von $\mathcal{O}$ auf A^n vermittels der Exponentialabbildung definiert. Wir bezeichnen für Elemente γ in $\mathcal{O}$ und g in A^n mit $\gamma \cdot g$ das Bild von g unter dieser Operation. Ferner sei $\rho = \rho(g; \mathcal{O})$ die Dimension der kleinsten analytischen Untergruppe B (s. [21]) von A^n, welche über $\overline{\mathbb{Q}}$ definiert sei und die Bahn $\mathcal{O} g$ von g enthält. $\mathcal{O}$ stabilisiert dann B und man erhält damit eine Darstellung von $\mathcal{O}$ in der allgemeinen linearen Gruppe $GL(T(B), \overline{\mathbb{Q}})$. Diese Darstellung ordnet einem Element aus $\mathcal{O}$ sein Differential zu. Da die Elemente aus $\mathcal{O}$ wie auch B über $\overline{\mathbb{Q}}$ definiert sind, erhalten wir ein Element aus $GL(T(B), \overline{\mathbb{Q}})$ (siehe hierzu auch [14]). Ist $\rho = 1$, so ist $GL(T(B), \overline{\mathbb{Q}}) \cong \overline{\mathbb{Q}}$.

Satz 1. *Es sei* $\alpha \neq 0$ *ein Element in* $T(A^n)(\bar{\mathbb{Q}})$ *und* $g = \exp_{A^n}(\alpha)$ *sein Bild in* A^n *unter der Exponentialabbildung* $\exp_{A^n}: T(A^n) \to A^n$ *der Lieschen Gruppe* A^n. *Ist* $X(g)$ *der Abschluß von* g *über* $\bar{\mathbb{Q}}$ *in der Zariski-Topologie von* A^n *und gilt* $\rho(g; \mathcal{O}) = 1$, *so gilt*

$$2 \dim X(g) \geqq nm.$$

Wir sollten an dieser Stelle bemerken, daß die Einschränkung $\rho(g; \mathcal{O}) = 1$ nicht wirklich notwendig ist. Im Beweis von Satz 1 werden wir sie an zwei Stellen benutzen. Zum einen werden wir sie im analytischen Teil benutzen, wo wir das Schwarzsche Lemma benötigen. Diese Einschränkung erlaubt es, im wesentlichen ein Schwarzsches Lemma in einer Variablen zu benützen. Dies ist völlig klassisch. Lassen wir diese Einschränkung fallen, so müssen wir eine schwierigere Version eines nicht effektiven Schwarzschen Lemmas für mehrere Variablen anwenden. Diese wurde von J.C. Moreau bewiesen (s. [18], Corollaire 7.6.2). Das zweite Hindernis ist etwas ernsthafter. Die Bedingung $\rho(g; \mathcal{O}) = 1$ zieht nach sich, daß wir mit einer analytischen Untergruppe der Dimension eins arbeiten und daher nur eine Derivation zur Verfügung haben. Nun benötigen wir im Beweis des Satzes eine Nullstellenabschätzung mit Multiplizitäten. Da wir nur mit einer Derivation arbeiten, können wir ein Ergebnis von D.W. Masser und dem Autor [12] benutzen, wo die dafür notwendige Nullstellenabschätzung bewiesen wird. Vor kurzem hat nun der Autor diese Nullstellenabschätzungen mit Multiplizitäten erweitert auf Nullstellenabschätzungen mit Multiplizitäten, in die eine beliebige Anzahl von Derivationen eingehen können. Benutzt man diese Ergebnisse, so kann der Satz leicht auch ohne die Einschränkung $\rho(g; \mathcal{O}) = 1$ bewiesen werden. Mit den Bezeichnungen in Satz 1 erhalten wir dann

$$2\rho(g; \mathcal{O}) \dim X(g) \geqq mn.$$

Da die Dimension der abelschen Varietät gleich nd ist, gibt uns Satz 1 die Abschätzung

$$nd \leqq 2 \dim X(g) \leqq 2nd.$$

Es ist bekannt, daß die Dimension m des Körpers k die Zahl $2d$ teilt. Damit erhalten wir offenbar die besten Ergebnisse, wenn $m = 2d$ ist. Dies ist der Fall, wenn die abelsche Varietät A vom Typ CM ist. Dann erhalten wir das folgende Resultat.

Korollar 1. *Mit denselben Bezeichnungen wie in Satz* 1 *erhalten wir für abelsche Varietäten vom Typ CM, daß* $X(g) = A^n$ *ist, falls* $\rho(g; \mathcal{O}) = 1$ *ist.*

Wir geben nun eine Anwendung dieses Resultats auf elliptische Kurven mit komplexer Multiplikation. Es sei E eine elliptische Kurve definiert über $\bar{\mathbb{Q}}$ mit komplexer Multiplikation und zugehörigem Körper k vom Grade 2 über den rationalen Zahlen. Sei $\wp$ die zugehörige Weierstraßsche elliptische Funktion. Ist F ein algebraischer Zahlkörper vom Grade n über k, so kann F dargestellt werden als eine Unteralgebra über k von $M_n(k)$ der Dimension n. Eine Ordnung von F hat dann die Dimension $2n$ über $\bar{\mathbb{Q}}$. Benutzt man dies, so erhält man aus dem Korollar 1 leicht das folgende elliptische Analogon des Satzes von Lindemann.

Korollar 2. *Sei E eine elliptische Kurve definiert über $\bar{\mathbb{Q}}$ mit komplexer Multiplikation. Die algebraischen Zahlen $\alpha_1, \ldots, \alpha_n$ seien linear unabhängig über k. Dann sind die Zahlen*

$$\wp(\alpha_1), \ldots, \wp(\alpha_n)$$

algebraisch unabhängig über den rationalen Zahlen.

Wir beenden diesen Abschnitt mit einigen Bemerkungen zum Beweis von Satz 1. Der Beweis ist sehr kompliziert, doch haben wir versucht, ihn so transparent wie möglich zu machen, indem wir die verschiedenen Ingredienzien separiert haben. Wir hoffen, daß sie dadurch so in sich abgeschlossen wie möglich geworden sind. Die grobe Idee des Beweises ist sehr ähnlich zu der, welche der Lindemann-Siegelschen Methode zugrunde liegt. Diese ist sehr klar zum Beispiel in dem Buch von Th. Schneider [16] dargestellt. Der Unterschied besteht sehr grob und vereinfacht gesprochen darin, daß die lineare Algebra durch die kommutative Algebra ersetzt werden muß. Der Beweis von Satz 1 kann in fünf Hauptschritte unterteilt werden. Der erste Schritt besteht in der Konstruktion des Hilfsideals. Dies wird ein gewisses Ideal in einem Polynomring über einem Zahlkörper sein und die bisher gebräuchliche Hilfsfunktion ersetzen. Hier benützen wir unter anderem eine Idee von P. Philippon [13]. Sie besteht in der Verwendung von Polynomen in sehr vielen Variablen, um dadurch Polynome mit ganzen Koeffizienten zu erhalten. Dies wurde auch von D.W. Masser und dem Autor in [11] benutzt, sowie später auch von Philippon in [15]. Der zweite Schritt besteht in einer Verallgemeinerung einer Idee von D.W. Masser, die von M. Anderson in seiner Thesis [1] ausgearbeitet wurde, siehe auch Choodnovsky [4]. Wir erweitern diese ingeniöse Idee auf beliebige algebraische Gruppen. Der nächste Schritt ist eine Nullstellenabschätzung mit hohen Multiplizitäten und nur wenigen Derivationen. Diese Hilfsmittel wurden von D.W. Masser und dem Autor in [9, 11, 12 und 20] entwickelt in dem Falle, daß die Multiplizität nicht zu hoch ist. Da wir aber mit hohen Multiplizitäten arbeiten müssen, müssen wir zusätzliche Betrachtungen anstellen, um diese Schwierigkeiten zu bewältigen. Diese Multiplizitätsabschätzungen erlauben es, den Spezialisierungstrick von Lindemann zu immitieren, um ein System von homogenen Polynomen zu gewinnen, welches keine nichttrivialen Nullstellen besitzt. Dies ist der vierte Schritt. Im letzten Schritt haben wir die Elimination durchzuführen. Hier wenden wir alte Techniken der Resultantenbildung an, wie sie zum Beispiel in dem Buch von Macaulay [8] dargestellt sind. Die Benutzung der Resultanten ersetzt das Lösen von linearen Gleichungssystemen, wie dies in der Lindemann-Siegelschen Methode gebräuchlich ist. Unsere Elimination spiegelt die Tatsache wider, daß die Projektion von einem Produkt $\mathbb{P}^n \times \mathbb{A}^m$ eines projektiven Raumes mit einem affinen Raum auf $\mathbb{A}^m$ eine abgeschlossene Abbildung ist. Dies ist nicht mehr länger der Fall, wenn man den projektiven Raum $\mathbb{P}^n$ durch einen affinen Raum $\mathbb{A}^n$ ersetzt. Dies ist auch der Grund für die Schwierigkeiten, die auftauchen, wenn man versucht, diese Methoden auf nicht-vollständige algebraische Gruppen auszudehnen. So sind wir zum Beispiel nicht in der Lage, mit Hilfe von unserer Methode einen neuen Beweis des Lindemannschen Satzes zu geben.

Alle von nun an auftretenden algebraischen Objekte seien über einem algebraischen Zahlkörper K definiert, sofern wir dies nicht ausdrücklich anders vermerken.

Diese Arbeit ist eine der Folgen einer sehr fruchtbaren Zusammenarbeit mit D.W. Masser, dem wir an dieser Stelle herzlichst danken möchten.

2. Additions- und Multiplikationsformeln

In diesem Abschnitt werden wir explizit einige Morphismen einer algebraischen Gruppe in eine zweite studieren. Wir werden hauptsächlich interessiert sein in dem Morphismus, der durch die Addition mehrerer Punkte der algebraischen Gruppe gegeben wird sowie in dem Morphismus, der durch die Multiplikation eines Punktes der Gruppe mit einer ganzen Zahl gegeben ist.

Es sei G eine kommutative algebraische Gruppe definiert über K und $\bar{G}$ ihre Kompaktifizierung (s. [17]). Wir können die Varietät $\bar{G}$ in einem projektiven Raum $\mathbb{P}^N$ für ein geeignetes N wie in [17] einbetten. Für eine positive ganze Zahl n sei G^n das n-fache Produkt von G mit sich selbst. Dies ist dann eingebettet in dem Produkt $\mathbb{P}^N \times \ldots \times \mathbb{P}^N$ von n Exemplaren des projektiven Raums $\mathbb{P}^N$. Seien

$$\mathbf{X}_1 := (X_{0,1}, \ldots, X_{N,1}), \ldots, \mathbf{X}_n := (X_{0,n}, \ldots, X_{N,n})$$

die homogenen Koordinaten der verschiedenen Faktoren. Wir nennen ein Polynom $P(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ multihomogen vom Multigrad $(D_1, \ldots, D_n)$, falls es homogen vom Grade D_i in $\mathbf{X}_i$ für $1 \leqq i \leqq n$ ist. Weiter bezeichnen wir für $(P_1, \ldots, P_n)$ in G^n mit $S(P_1, \ldots, P_n)$ die Summe $P_1 + \ldots + P_n$. Es seien $e_1, \ldots, e_n$ fest gewählte Elemente aus dem Endomorphismenring von G. Die Summenbildung ist ein Morphismus von G^n nach G. Dann zieht die Definition eines Morphismus (s. [5]) sofort das folgende Resultat nach sich.

Lemma 1. *Es gibt eine endliche offene Überdeckung $\{U_i\}$ von G^n mit der folgenden Eigenschaft. Für jedes i existiert ein System von multihomogenen Polynomen mit Koeffizienten in K*

$$\mathscr{A}_i(\mathbf{X}_1, \ldots, \mathbf{X}_n) = (A_{i,0}(\mathbf{X}_1, \ldots, \mathbf{X}_n), \ldots, A_{i,N}(\mathbf{X}_1, \ldots, \mathbf{X}_n))$$

vom Multigrad $(d_i, \ldots, d_i)$ für positives ganzes d_i, so daß für $(P_1, \ldots, P_n)$ in U_i die homogenen Koordinaten von $e_1 P_1 + \ldots + e_n P_n$ durch $\mathscr{A}_i(\mathbf{X}_1(P_1), \ldots, \mathbf{X}_n(P_n))$ gegeben werden.

Beweis. Dies ist nichts anderes als eine lokale Beschreibung des Morphismus

$$S \circ (e_1, \ldots, e_n): G^n \to G.$$

Im nächsten Lemma werden wir eine Beschreibung des Morphismus geben, der durch die Multiplikation mit einer festen ganzen Zahl gegeben wird. Dazu sei m eine solche ganze Zahl und $[m]_G: G \to G$ der Morphismus $P \mapsto mP$. Ferner definieren wir die Höhe eines Polynoms mit algebraischen Koeffizienten als das Maximum der Höhen dieser Koeffizienten (s. hierzu [18]).

Lemma 2. *Es gibt $N+1$ homogene Polynome mit Koeffizienten in K*

$$\phi_0^{(m)}(X_0, \ldots, X_N), \ldots, \phi_N^{(m)}(X_0, \ldots, X_N)$$

in den homogenen Koordinaten $\mathbf{X} := (X_0, \ldots, X_N)$ von $\mathbb{P}^N$ mit den folgenden Eigenschaften. Der Grad dieser Polynome ist m^2 und die Höhe höchstens gleich $c_1^{m^2}$ für eine positive Konstante c_1. Sie verschwinden nicht gleichzeitig in irgend einem Punkte von G. Ist P ein Punkt auf G, so werden die homogenen Koordinaten von mP gegeben durch

$$(\phi_0^{(m)}(\mathbf{X}), \ldots, \phi_N^{(m)}(\mathbf{X})) = \phi^{(m)}(\mathbf{X}).$$

Beweis. Dies ist eine einfache Folgerung aus Corollaire 2 der Proposition 3 und Proposition 5 aus [17], sowie Lèmme 7 aus [3].

Wir kombinieren nun diese zwei Lemmata in der folgenden Weise. Dazu setzen wir für ganze Zahlen $m_1, \ldots, m_n$ $m = m_1 e_1 + \ldots + m_n e_n$ und betrachten den Morphismus von G^n nach G, welcher durch

$$(P_1, \ldots, P_n) \longmapsto m_1 e_1 P_1 + \ldots + m_n e_n P_n$$

gegeben wird. Dann haben wir die folgende algebraische Beschreibung dieses Morphismus.

Proposition 1. *Es gibt eine endliche offene Überdeckung $\{U_i\}$ von G^n mit der folgenden Eigenschaft. Für jedes i und für alle ganzen Zahlen $m_1, \ldots, m_n$ gibt es $N+1$ multihomogene Polynome*

$$\psi_0^{(i,m)}(\mathbf{X}_1, \ldots, \mathbf{X}_n), \ldots, \psi_N^{(i,m)}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$$

mit Koeffizienten in K, mit dem Multigrad $(m_1 d_i, \ldots, m_n d_i)$ und mit Höhen höchstens $c_2^{m_1^2 + \cdots + m_n^2}$, so daß für $(P_1, \ldots, P_n)$ in U_i die homogenen Koordinaten von $m_1 e_1 P_1 + \ldots + m_n e_n P_n$ durch

$$(\psi_0^{(i,m)}(\mathbf{X}_1(P_1), \ldots, \mathbf{X}_n(P_n)), \ldots, \psi_N^{(i,m)}(\mathbf{X}_1(P_1), \ldots, \mathbf{X}_n(P_n)))$$

gegeben werden.

Beweis. Wir wählen die Überdeckung $\{U_i\}$ wie in Lemma 1 und definieren die Polynome $\psi_j^{(i,m)}$ für $0 \leqq j \leqq N$ durch

$$\psi_j^{(i,m)} = A_{i,j}(\phi^{(m_1)}(\mathbf{X}_1), \ldots, \phi^{(m_n)}(\mathbf{X}_n)).$$

Dann prüft man die in der Proposition angegebenen Eigenschaften leicht unter Benutzung von Lemma 1 und Lemma 2 nach.

Seien $\mathbf{X} = (X_0, \ldots, X_N)$ projektive Koordinaten von G. Dann können wir diese Proposition in dem Spezialfall $P_1 = \ldots = P_n =: P$ anwenden. Es gilt dann $mP = m_1 e_1 P_1 + \ldots + m_n e_n P_n$, und m ist ein Endomorphismus von G wie in Proposition 1 mit $\mathbf{X}_1 = \ldots = \mathbf{X}_n = \mathbf{X}$.

3. Der Trick von Masser-Anderson für beliebige algebraische Gruppen

In seiner Thesis [1] arbeitete Anderson eine sehr ingeniöse Idee von Masser aus, um sehr gute Schranken für den Grad und die Höhe von abgeleiteten elliptischen Polynomen zu erhalten. Dieses Hilfsmittel wurde darauf von einer großen Zahl von Autoren benutzt und wir werden es in diesem Abschnitt auf beliebige kommutative algebraische Gruppen ausdehnen. Um dies zu entwickeln, werden wir weiterhin mit der algebraischen Gruppe G wie im vorangegangenen Abschnitt arbeiten. Die Gruppe G ist in einem projektiven Raum $\mathbb{P}^N$ vermöge eines sehr amplen Divisors eingebettet. Dies ermöglicht es uns, ein System von Koordinaten $X_0, \ldots, X_N$ so zu wählen, daß für eine endlich erzeugte Untergruppe Γ von G für alle γ in Γ gilt $X_0(\gamma) \neq 0$. Denn sind die projektiven Koordinaten der Elemente von Γ so gewählt, daß sie in einer endlich erzeugten Erweiterung von $\mathbb{Q}$ liegen, so gibt es eine Linearform X_0' in $X_0, \ldots, X_N$ mit algebraischen Koeffizienten, so daß $X_0'(\gamma) \neq 0$ gilt für alle γ aus Γ. Wir brauchen nun nur noch X_0 durch X_0' zu ersetzen und erhalten die gewünschte Eigenschaft. Insbesondere gilt dann für das neutrale Element 0 der Gruppe $X_0(0) \neq 0$. Wir nehmen nun ein zweites Exemplar G' von G mit projektiven Koordinaten $X_0', \ldots, X_N'$ und betrachten das Produkt $G \times G'$ eingebettet in dem Raum $\mathbb{P}^N \times \mathbb{P}^N$. Auf der algebraischen Gruppe G der Dimension n können wir die Derivationen $\partial_1, \ldots, \partial_n$ wie in [20] definieren und wir bezeichnen die zugehörigen Derivationen auf G' mit $\partial_1', \ldots, \partial_n'$. Wir wählen nun für eine fest vorgegebene positive ganze Zahl d in dem Vektorraum, der durch $\partial_1, \ldots, \partial_n$ über dem Körper K erzeugt wird, d linear unabhängige Derivationen $\Delta_1, \ldots, \Delta_d$ und bezeichnen die entsprechenden Derivationen auf G' mit $\Delta_1', \ldots, \Delta_d'$. Diese Wahl entspricht in eindeutiger Weise der Wahl einer analytischen Untergruppe von G (s. ebenfalls [20]).

Wir setzen nun $x_i = X_i/X_0$ für ganze Zahlen i mit $1 \leqq i \leqq N$ und $\mathbf{x} = (1, x_1, \ldots, x_N)$. Dies sind affine Koordinaten für die affin offene Menge $X_0 \neq 0$ in $\mathbb{P}^N$. Wie üblich bezeichnen wir mit $x_1', \ldots, x_N'$ und $\mathbf{x}'$ die entsprechenden Koordinaten auf G'. Sei $\mathscr{A}$ eine Additionsformel, so daß $A_0(\mathbf{X}, \mathbf{x}'(0))$ nicht identisch verschwindet. Dann gilt auf der offenen Teilmenge $X_0 \neq 0$, $A_0(\mathbf{X}, \mathbf{x}'(0)) \neq 0$ von G

$$\Delta_1^{\tau_1} \ldots \Delta_d^{\tau_d}(P(\mathbf{X})/X_0^D) = \Delta_1'^{\tau_1} \ldots \Delta_d'^{\tau_d}(P \circ \mathscr{A}/A_0^D)(\mathbf{X}, \mathbf{x}'(0)) \tag{1}$$

für homogene Polynome $P(\mathbf{X})$ vom Gerade D. Wir sollten bemerken, daß beide Seiten von (1) definiert sind, da sowohl $P(\mathbf{X})/X_0^D$ als auch $(P \circ \mathscr{A})/A_0^D$ rationale Funktionen auf G und $G \times G_0'$ sind, wenn wir mit G_0' die offene Teilmenge von G' bezeichnen, wo $X_0' \neq 0$ ist.

Für nichtnegative ganze Zahlen $\sigma_1, \ldots, \sigma_d$, $\sigma_1', \ldots, \sigma_d'$ und $\tau_1, \ldots, \tau_d$ setzen wir $\sigma = (\sigma_1, \ldots, \sigma_d)$, $\sigma' = (\sigma_1', \ldots, \sigma_d')$ und $\tau = (\tau_1, \ldots, \tau_d)$. Dann definieren wir $\Delta^\sigma = \Delta_1^{\sigma_1} \ldots \Delta_d^{\sigma_d}$ und entsprechend $\Delta'^{\sigma'}$, sowie $\sigma! = \sigma_1! \ldots \sigma_d!$. Wir schreiben $\sigma \leqq \tau$ dann und nur dann, wenn $\sigma_1 \leqq \tau_1, \ldots, \sigma_d \leqq \tau_d$ gilt. Damit können wir die rechte Seite von (1) schreiben als

$$\Delta'^\tau(P \circ \mathscr{A}/A_0^D)(\mathbf{X}, \mathbf{x}'(0)) = \sum_{\sigma + \sigma' = \tau} \frac{\tau!}{\sigma!\,\sigma'!} \Delta'^\sigma(A_0^{-D})\,\Delta'^{\sigma'}(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}'(0)), \tag{2}$$

wobei wir die Leibnitzsche Regel benützt haben. Bezeichnen wir mit M die multiplikative Menge der Potenzprodukte von X_0 und $A_0(\mathbf{X}, \mathbf{x}'(0))$ mit nichtnegativen Exponenten und mit $I(G)$ das definierende Ideal von G, welches dasselbe ist wie das von $\bar{G}$, dem Abschluß von G, so folgt aus (1) und (2)

$$\Delta^{\tau}(P(\mathbf{X})/X_0^D) \equiv \sum_{\sigma+\sigma'=\tau} \frac{\tau!}{\sigma!\,\sigma'!} \Delta'^{\sigma}(A_0^{-D})\, \Delta'^{\sigma'}(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}'(0))$$

modulo $I(G)\,M^{-1}$. Deswegen können wir die Ausdrücke $\Delta^{\tau}(P(\mathbf{X})/X_0^D)$ modulo $I(G)\,M^{-1}$ ausdrücken durch die Ausdrücke $\Delta'^{\sigma'}(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}'(0))$. Es kann leicht verifiziert werden, daß das Gleichungssystem (3) sukzessive modulo $I(G)\,M^{-1}$ gelöst werden kann, da es Diagonalgestalt besitzt. Wir gehen etwas anders vor, um dieses Gleichungssystem zu lösen. Unser Vorgehen hat den Vorteil, daß wir einen expliziten Ausdruck erhalten. Wir schreiben

$$(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}') = A_0^D \cdot \frac{P \circ \mathscr{A}}{A_0^D}(\mathbf{X}, \mathbf{x}')$$

und wenden darauf den Differentialoperatoren Δ'^{τ} an. Dann erhalten wir mit Hilfe der Leibnitzschen Regel

$$\Delta^{\tau}(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}'(0)) = \sum_{\sigma+\sigma'=\tau} \frac{\tau!}{\sigma!\,\sigma'!} \Delta'^{\sigma}(A_0^D)\, \Delta'^{\sigma'}(P \circ \mathscr{A}/A_0^D)(\mathbf{X}, \mathbf{x}'(0)). \tag{4}$$

Darin ersetzen wir den Ausdruck $\Delta'^{\sigma'}(P \circ \mathscr{A}/A_0^D)(\mathbf{X}, \mathbf{x}'(0))$ durch $\Delta^{\sigma'}(P(\mathbf{X})/X_0^D)$ und erhalten dann schließlich

$$\Delta'^{\tau}(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}'(0)) \equiv \sum_{\sigma+\sigma'=\tau} \frac{\tau!}{\sigma!\,\sigma'!} \Delta'^{\sigma} A_0^D(\mathbf{X}, \mathbf{x}'(0))\, \Delta^{\sigma'}(P(\mathbf{X})/X_0^D) \tag{5}$$

modulo $I(G)\,M^{-1}$. Dies fassen wir in der folgenden Proposition zusammen.

Proposition 2. *Es sei $P(\mathbf{X})$ ein homogenes Polynom vom Grade D und $\mathscr{A} = (A_0, \ldots, A_N)$ eine Additionsformel, so daß $A_0(\mathbf{X}, \mathbf{x}'(0)) \neq 0$. Dann gilt für beliebiges $\tau \geqq (0, \ldots, 0)$*

$$\Delta^{\tau}(P(\mathbf{X})/X_0^D) \equiv \sum_{\sigma+\sigma'=\tau} \frac{\tau!}{\sigma!\,\sigma'!} \Delta'^{\sigma}(A_0^{-D})\, \Delta'^{\sigma'}(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}'(0))$$

und

$$\Delta'(P \circ \mathscr{A})(\mathbf{X}, \mathbf{x}'(0)) \equiv \sum_{\sigma+\sigma'=\tau} \frac{\tau!}{\sigma!\,\sigma'!} \Delta'^{\sigma}(A_0^D)(\mathbf{X}, \mathbf{x}'(0))\, \Delta^{\sigma'}(P(\mathbf{X})/X_0^D)$$

modulo dem Ideal $I(G)\,M^{-1}$.

4. Wurzelziehen

Ein sehr nützliches Hilfsmittel in der algebraischen Geometrie ist das Wurzelziehen. Diese Technik wurde in der Theorie der transzendenten Zahlen von D.W. Masser wiederentdeckt. In diesem Abschnitt geben wir eine Modifikation davon in einer multiplikativen Version an. In diesem Abschnitt sei $\mathbb{C}$ der Grundkörper.

Es sei wiederum G eine kommutative algebraische Gruppe, die wie üblich projektiv eingebettet sei. Für eine positive ganze Zahl l sei $[l]_G$ der Endomorphismus $g \mapsto lg$ auf G. Da wir die Gruppe G festhalten wollen, schreiben wir dafür kurz $[l]$. Für genauere Einzelheiten verweisen wir auf [17]. Auf dem Tangentialraum $T(G)$ induziert dieser Morphismus eine lineare Abbildung, das Differential $d[l]$, und auf dem Körper der rationalen Funktionen $\mathbb{C}(G)$ einen Endomorphismus $[l]^*$. Er wird gegeben durch $F \mapsto [l]^*(F) = F \circ [l]$.

Wir können nun die Basis $\partial_1, \ldots, \partial_n$ von $T(G)$ so wählen, daß

$$d[l](\partial_i) = l\partial_i \qquad (1 \leqq i \leqq n),$$

wobei $n = \dim G$. Wir setzen $\partial_i' := d[l](\partial_i)$ für $1 \leqq i \leqq n$. Sei nun ∂ ein Element des Vektorraumes, der durch $\partial_1, \ldots, \partial_n$ über dem Körper der komplexen Zahlen erzeugt wird. Dann gilt

$$\partial = \alpha_1 \partial_1 + \ldots + \alpha_n \partial_n$$

für $\alpha_1, \ldots, \alpha_n$ in $\mathbb{C}$. Wir erhalten dann

$$\partial' := d[l](\partial) = \alpha_1 d[l](\partial_1) + \ldots + \alpha_n d[l](\partial_n) = l\partial.$$

Wir definieren nun die Verschwindungsordnung einer rationalen Funktion auf G und bestimmen ihr Verhalten unter dem Morphismus $[l]$.

Sei F ein Element aus dem Körper der rationalen Funktionen $\mathbb{C}(G)$ auf G. Weiter seien für eine ganze Zahl d mit $1 \leqq d \leqq n$ Derivationen $\Delta_1, \ldots, \Delta_d$ aus dem Vektorraum fest gewählt, der von $\partial_1, \ldots, \partial_n$ über dem Körper der komplexen Zahlen erzeugt wird. Wir definieren dann die Verschwindungsordnung von F im Punkte g aus G längs $\Delta_1, \ldots, \Delta_d$ in der folgenden Weise. Zunächst schreiben wir $\Delta = (\Delta_1, \ldots, \Delta_d)$. Dann sei ohne Beschränkung der Allgemeinheit $X_0(g) \neq 0$. Dann können wir F schreiben als

$$F(\mathbf{x}) = G(\mathbf{x})/H(\mathbf{x})$$

für Polynome G und H. Für ein Polynom $P(\mathbf{x})$ definieren wir nun die Verschwindungsordnung in g als die kleinste ganze Zahl $T \geqq 0$ mit der Eigenschaft, daß es nichtnegative ganze Zahlen $t_1, \ldots, t_d$ mit $t_1 + \ldots + t_d = T$ gibt, so daß

$$\Delta_1^{t_1} \ldots \Delta_d^{t_d}(P)(\mathbf{x}(g)) \neq 0$$

gilt. Wir schreiben dann $T = \operatorname{ord}_{\Delta, g}(P)$ und definieren

$$\operatorname{ord}_{\Delta, g}(F) = \operatorname{ord}_{\Delta, g}(G) - \operatorname{ord}_{\Delta, g}(H).$$

Es kann leicht gezeigt werden, daß diese Definition nicht von der Wahl von G und H abhängt. Sind weiter von Null verschiedene komplexe Zahlen $\alpha_1, \ldots, \alpha_d$ gegeben und setzen wir $\alpha\Delta = (\alpha_1 \Delta_1, \ldots, \alpha_d \Delta_d)$, so gilt

$$\operatorname{ord}_{\alpha\Delta, g}(F) = \operatorname{ord}_{\Delta, g}(F).$$

Sei nun g' in G so gewählt, daß $g' = lg$ gilt für ein g aus G. Dan erhalten wir für F aus $\mathbb{C}(G)$ und ganze Zahlen $0 \leqq t_1, \ldots, t_d$ mit $t = (t_1, \ldots, t_d)$ und den Bezeichnungen aus Abschnitt 3 mit $\Delta' = l\Delta$

$$\Delta'^t(F)(\mathbf{x}(g')) = (l\Delta)^t(F)(\mathbf{x}(lg)) = (l\Delta)^t([l]^*(F))(\mathbf{x}(g)).$$

Daraus erhalten wir sofort das folgende Lemma.

Lemma 3. *Für eine positive ganze Zahl l sei $[l]$ der Endomorphismus von G, der gegeben wird durch $g \mapsto lg$. Dann gilt für alle nichttrivialen Derivationen $\Delta_1, \ldots, \Delta_d$, welche in dem Punkte g aus G definiert sind,*

$$\operatorname{ord}_{\Delta, lg}(F) = \operatorname{ord}_{\Delta, g}([l]^*(F))$$

für alle rationalen Funktionen F auf G.

Wir werden dieses Lemma im nächsten Abschnitt benutzen, um eine Nullstellenabschätzung mit hohen Vielfachheiten zu beweisen. In [20] und [12] haben wir Einschränkungen an die Größe der Ordnung der Differentialoperatoren. Diese Restriktionen können mit Hilfe dieses Lemmas eliminiert werden. Weiter können wir mit Hilfe dieses Lemmas die Grad-Theorie für Produkte projektiver Räume in vielen Fällen vermeiden einschließlich des Falles, den wir später betrachten werden.

5. Eine neue Nullstellenabschätzung

In diesem Abschnitt werden wir die Theorie der Nullstellenabschätzungen weiter entwickeln, die von D.W. Masser und dem Autor in [9] begründet wurde und in [12] und [20] auf Multiplizitätsabschätzungen erweitert wurde. In all diesen Arbeiten hatten wir große Einschränkungen an die Größenordnung der Nullstellenordnungen zu machen, um bestmögliche Resultate zu erzielen. Dies war notwendig wegen der Nichtlinearität der Differentialgleichungen, die von den Funktionen erfüllt werden, die in der Exponentialabbildung der algebraischen Gruppe auftauchen. In viele Problemen können diese Einschränkungen leicht in Kauf genommen werden. Ein gutes Beispiel hierfür ist Bakers Methode. Jedoch hier in dieser Arbeit haben wir ein Beispiel vorliegen, wo diese Einschränkungen ernsthafter Natur sind. Deswegen sind wir gezwungen, sie zu eliminieren. Dies kann mit Hilfe der Resultate in Abschnitt 4 zusammen mit einigen Erweiterungen der Nullstellenabschätzungen, die in [12] bewiesen wurden, sowie einigen sehr einfachen Ideen durchgeführt werden. Jedoch können diese Restriktionen wegen einiger technischer Schwierigkeiten nicht im allgemeinsten Falle elimiert werden, sondern nur in einem Spezialfall, welcher den Fall umfaßt, an dem wir hauptsächlich interessiert sind.

Wie bereits bemerkt benützen wir dazu ein Resultat, das in [12] bewiesen wurde. Der besseren Lesbarkeit halber wollen wir diese Resultate hier kurz referieren. Auch in diesem Abschnitt wählen wir $\mathbb{C}$ als Grundkörper.

Es seien $G_1, \ldots, G_n$ kommutative algebraische Gruppen und $G = G_1 \times \ldots \times G_n$ ihr Produkt. Für $1 \leqq i \leqq n$ bezeichnen wir mit d_i die Dimension von G_i und setzen $d = d_1 + \ldots + d_n$. Wir stellen dann an die Gruppen $G_1, \ldots, G_n$ die folgende Bedingung. Ist H eine zusammenhängende Untergruppe von G, dann gibt es zusammenhängende Untergruppen H_i von G_i für $1 \leqq i \leqq n$, so daß $H = H_1 \times \ldots \times H_n$.

Sei Γ eine endliche Menge in G, die das neutrale Element 0 der Gruppe G enthalte. Dann definieren wir für positive ganze Zahlen r die Menge $r\Gamma$ als

$$r\Gamma = \Gamma + \ldots + \Gamma,$$

die Menge aller Linearkombinationen von r Elementen aus Γ. Da Γ das neutrale Element der Gruppe enthält, gilt $\Gamma \subseteq r\Gamma$. Für nicht-negative ganze Zahlen $r_1, \ldots, r_n$ mit $0 \leqq r_i \leqq d_i$ für $1 \leqq i \leqq n$ definieren wir die ganzen Zahlen $Q_{r_1, \ldots, r_n}(\Gamma)$ in der folgenden Weise. Gibt es keine algebraische Untergruppe H von G der Form $H = H_1 \times \ldots \times H_n$ mit $\operatorname{cod} H_i = r_i$ für $1 \leqq i \leqq n$, so setzen wir

$$Q_{r_1, \ldots, r_n}(\Gamma) = |\Gamma|.$$

Andernfalls setzen wir

$$Q_{r_1, \ldots, r_n}(\Gamma) = \min(|(\Gamma + H)/H|),$$

wobei das Minimum über alle zusammenhängenden Untergruppen H von G genommen wird, für die $\operatorname{cod} H_i = r_i$ für $1 \leqq i \leqq n$.

Die Gruppen G_i seien nun in einen projektiven Raum $\mathbb{P}^{N_i}$ und G deswegen im Raum $\mathbb{P}^{N_1} \times \ldots \times \mathbb{P}^{N_n}$ eingebettet. Dann seien $\mathbf{X}_i = (X_{i,0}, \ldots, X_{i,N_i})$ für $1 \leqq i \leqq n$ projektive Koordinaten von $\mathbb{P}^{N_i}$. Schließlich sei Δ eine Derivation auf G, die einer eindimensionalen Unteralgebra der Lie-Algebra von G entspreche. Sei weiter B die zugehörige eindimensionale analytische Untergruppe von G. Dann wurde in [12] das folgende Ergebnis bewiesen. Es verallgemeinert in einer Richtung das Theorem ABC aus [9].

Satz. *Es gibt eine positive Konstante c, die nur von der Gruppe G abhängt, mit der folgenden Eigenschaft. Sei für ganze Zahlen $0 \leqq D_1, \ldots, D_n$ und $T > 0$ ein multihomogenes Polynom $P(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ vom Multigrad $(D_1, \ldots, D_n)$ gegeben, welches auf $d\Gamma$ mindestens von der Ordnung T längs Δ verschwinde. Gilt dann*

$$T Q_{r_1, \ldots, r_n}(\Gamma) \geqq (c D_1)^{r_1} \ldots (c D_n)^{r_n} \tag{6}$$

für $0 \leqq r_i \leqq d_i$ und $1 \leqq i \leqq n$ mit $1 \leqq r = r_1 + \ldots + r_n$ und

$$Q_{r_1, \ldots, r_n}(\Gamma) \geqq (c D_1)^{r_1} \ldots (c D_n)^{r_n} \delta^{-1} \tag{7}$$

für $1 \leqq r < d$ mit $\delta = \max(1, \min(c D_i))$, wobei das innere Minimum über alle i genommen wird, für die die Gruppe G_i nichtlinear ist und das Minimum über einer leeren Menge als Null gesetzt wird, so verschwindet das Polynom P auf ganz $B + \gamma$ für ein γ in Γ.

Wir werden nun den Spezialfall betrachten, wo $n = 2$ und $G_1 = A$ ist für eine abelsche Varietät A und G_2 das d_2-fache Produkt der additiven Gruppe $\mathbb{G}_a$.

Wir werden dann zeigen, daß der Satz auch ohne die Bedingung (7) richtig ist. Um genau zu sein, seien $\mathbf{Z}$ die homogenen Koordinaten, welche der Gruppe G_2 entsprechen mögen und $\mathbf{X}$ die zur Gruppe A gehörigen homogenen Koordinaten.

Satz 2. *Sei in dem Satz $G_1=A$ für eine abelsche Varietät A, $G_2=\mathbb{G}_a^{d_2}$ und P ein bihomogenes Polynom wie im Satz. Ist dann B dicht in G und gilt*

$$T Q_{r_1,r_2}(\Gamma) \geqq (c D_1)^{r_1} (c D_2)^{r_2} \tag{8}$$

für $0 \leqq r_i \leqq d_i$ und $i=1,2$ mit $r=r_1+r_2 \geqq 1$. Dann verschwindet P auf ganz G.

Bevor wir den Satz 2 beweisen, machen wir noch eine Bemerkung. Erstens ist es kein Problem, den Satz 2 auf den Fall auszudehnen, wo $G_1=A_1, \ldots, G_{n-1}=A_{n-1}$ und $G_n=\mathbb{G}_a^{d_n}$ ist, wobei $A_1, \ldots, A_{n-1}$ abelsche Varietäten sind, sofern nur die Voraussetzungen vom Satz erfüllt sind. Dies erreicht man dadurch, daß man die Ergebnisse aus Abschnitt 4 auf diese etwas allgemeinere Situation ausdehnt. Zweitens ist es möglich, ohne größere Anstrengung ein Produkt von multiplikativen Gruppen mit einzubeziehen. Dafür sind keine neuen Ideen notwendig.

Beweis von Satz 2. Wir werden den Satz 2 aus dem Satz herleiten. Dazu wählen wir eine positive ganze Zahl l mit $c l^2 D_1 \geqq T$ und bezeichnen wie in Abschnitt 2 und Abschnitt 4 mit $[l]$ den Morphismus, der durch die Multiplikation $g \mapsto lg$ auf der Gruppe $G=A \times \mathbb{G}_a^{d_2}$ gegeben wird. Wir können dann den Satz auf das bihomogene Polynom $[l]^*(P)$ anwenden. Dieses hat den Bigrad $(l^2 D_1, D_2)$. Dies folgt aus Lemma 2 aus Abschnitt 2 und der Tatsache, daß auf der Gruppe $\mathbb{G}_a^d$ die Multiplikation mit l durch Multiplikation der Koordinaten mit l gegeben ist. Für Γ nehmen wir die Menge $[l]^{-1}(\Gamma)$ und müssen nur die Zahlen $Q_{r_1,r_2}([l]^{-1}(\Gamma))$ berechnen. Um dies zu tun, sei H eine zusammenhängende algebraische Untergruppe der Form $H=H_1 \times H_2$ mit $\operatorname{cod} H_1=r_1$ und $\operatorname{cod} H_2=r_2$. Dann erhalten wir

$$([l]^{-1}(\Gamma)+H)/H=[\bar{l}]^{-1}((\Gamma+H)/H),$$

wobei $[\bar{l}]$ der Morphismus ist, der durch $[l]$ auf G/H induziert wird. Dessen Kern ist eine Untergruppe von G/H der Ordnung l^{2r_1}, so daß wir

$$|([l]^{-1}(\Gamma)+H)/H| = l^{2r_1} |(\Gamma+H)/H|$$

erhalten. Dies zieht

$$Q_{r_1,r_2}([l]^{-1}(\Gamma)) = l^{2r_1} Q_{r_1,r_2}(\Gamma)$$

nach sich. Weiter folgt aus Lemma 3, daß das Polynom $[l]^*(P)$ in $[l]^{-1}(\Gamma)$ wenigstens mit der Ordnung T längs Δ verschwindet. Daher gilt

$$T Q_{r_1,r_2}([l]^{-1}(\Gamma)) = T l^{2r_1} Q_{r_1,r_2}(\Gamma).$$

Die rechte Seite ist wenigstens gleich $l^{2r_1}(c D_1)^{r_1}(c D_2)^{r_2}$ wegen (8). Deswegen erhalten wir

$$T Q_{r_1,r_2}([l]^{-1}(\Gamma)) \geqq (c l^2 D_1)^{r_1} (c D_2)^{r_2}, \tag{9}$$

und deswegen ist (6) für $[l]^*(P)$ erfüllt. Aufgrund der Wahl von l erhalten wir aus (9)

$$c l^2 D_1 Q_{r_1,r_2}([l]^{-1}(\Gamma)) \geqq (c l^2 D_1)^{r_1} (c D_2)^{r_2},$$

und dies zieht (7) nach sich, da $G_2 = \mathbb{G}_a^{d_2}$ linear ist und somit $\delta = \max(1, c l^2 D_1)$ ist. Der Satz impliziert deshalb, daß $[l]^*(P)$ auf ganz G verschwindet und somit auch P, wie dies behauptet war. Damit ist der Satz 2 bewiesen.

Wir können nun Satz 2 auf die folgende Situation anwenden, wie sie später im Beweis von Satz 1 auftritt. Sei G die Gruppe $(\mathbb{G}_a \times A^n)^\kappa$ für ein $\kappa \geqq 1$ und Δ eine Derivation auf $\mathbb{G}_a \times A^n$, welche einer dichten analytischen Untergruppe B von $\mathbb{G}_a \times A^n$ der Dimension 1 entspricht. Seien $\Delta_1, \ldots, \Delta_\kappa$ κ Exemplare von Δ, die den verschiedenen Faktoren vom Typ $\mathbb{G}_a \times A^n$ von G entsprechen. Sei $g = (g', g'')$ ein Element von $\mathbb{G}_a \times A^n$. Dann definieren wir mit der Bezeichnung von Abschnitt 1 für γ in $\mathcal{O}$ das Element $\gamma g = (\gamma g', \gamma g'')$ und für Elemente $g_1, \ldots, g_\kappa$ in $\mathbb{G}_a \times A^n$ die Mengen $\Gamma_i = \mathcal{O} g_i$ für $1 \leqq i \leqq \kappa$. Wie üblich definieren wir weiter $\Gamma_i(S)$ für reelle Zahlen $S \geqq 0$ (s. [9]) und $\Gamma(S) = \Gamma_1(S) \times \ldots \times \Gamma_\kappa(S)$ in $G = (\mathbb{G}_a \times A^n)^\kappa = \mathbb{G}_a^\kappa \times A^{n\kappa}$. Da die Operation von Γ auf $\mathbb{G}_a \times A^n$ auf dem ersten Faktor komplexe Multiplikation ist, erhalten wir

$$|\Gamma_i(S)| \geqq S^{nm} \qquad (1 \leqq i \leqq \kappa),$$

falls $g'_1 \neq 0, \ldots, g'_\kappa \neq 0$ ist.

Korollar. *Es gibt eine positive Konstante c mit der folgenden Eigenschaft. Für ganze Zahlen $S \geqq 0$, $D \geqq 0$, $L \geqq 0$ und $T \geqq 1$ sei $P(\mathbf{Z}_1, \mathbf{X}_1, \ldots, \mathbf{Z}_\kappa, \mathbf{X}_\kappa)$ ein multibihomogenes Polynom vom Multi-Bigrad (L, D), welches auf $(nd+1)\Gamma(S)$ mindestens von der Ordnung T längs B^κ verschwindet, aber nicht identisch auf G. Gilt dann*

$$T \geqq (cL)^{nd} \tag{10}$$

und

$$T S^{mn} \geqq c D (cL)^{nd}, \tag{11}$$

so ist wenigstens eines der Elemente $g'_1, \ldots, g'_\kappa$ das neutrale Element der Gruppe $\mathbb{G}_a$.

Beweis. Wir sind in der Produkt-Situation von Satz 2. Ein einfaches Argument, das wir dem Leser überlassen können, zeigt, daß es genügt, das Korollar für den Fall $\kappa = 1$ zu beweisen. Dazu müssen wir nur noch die Zahlen $Q_{r_1,r_2}(\Gamma(S))$ berechnen. Sei dazu $H = H_1 \times H_2$ eine algebraische Untergruppe von G, wobei H_1 in $\mathbb{G}_a$ und H_2 in A^n liegt und $\operatorname{cod} H_i = r_i$ für $i = 1, 2$ sei. Bezeichnen wir mit Γ' die Projektion von $\Gamma \cap H$ auf H_1, so erzeugt Γ' einen Untervektorraum des komplexen Vektorraums H_1 dessen Kodimension wenigstens gleich r_1 ist. Deswegen ist der Korang von Γ' wenigstens gleich $r_1 mn$ und dies zieht

$$Q_{r_1,r_2}(\Gamma(S)) \geqq S^{r_1 mn}$$

für alle r_1, r_2 mit $0 \leqq r_1 \leqq 1$, $0 \leqq r_2 \leqq dn$ und $r = r_1 + r_2 \geqq 1$ nach sich. Gilt $r_1 = 0$, so folgt in Verbindung mit (10) die Ungleichung

$$T Q_{0,r_2}(\Gamma(S)) \geqq (cL)^{nd} \geqq (cL)^{r_2}.$$

Gilt $r_1=1$, so erhalten wir wegen (11) die Ungleichung

$$TQ_{1,r_2}(\Gamma(S)) \geqq TS^{mn} \geqq cD(cL)^{nd}.$$

In beiden Fällen erhalten wir schließlich

$$TQ_{r_1,r_2}(\Gamma(S)) \geqq (cL)^{r_2}(cD)^{r_1}.$$

Wegen der Ungleichung (12) sind alle Voraussetzungen von Satz 2 erfüllt, und wir erhalten, daß das Polynom P auf ganz G verschwindet oder aber das Element g_1' das neutrale Element der additiven Gruppe ist. Da das Polynom P nicht identisch auf G verschwindet, muß das letztere eintreten. Damit ist aber das Korollar bewiesen.

Wir beenden diesen Abschnitt mit einer zusätzlicher Bemerkung zum Beweis des Korollars von Satz 2. In unserer Anwendung dieses Korollars im Beweis von Satz 1 werden wir die Zahlen S, T, D und L so wählen, daß $D \geqq cL$ gilt. Dann können wir das Wurzelziehen mit einer Zahl l so vornehmen, daß $cl^2L = D$ gilt. Dann wird das Polynom $[l]^*(P)$ multihomogen vom Multigrad $(D, \ldots, D)$. Wir erhalten dadurch ein homogenes Polynom vom Grade $2\kappa D$. In diesem Falle benötigen wir nicht den Satz für verschiedene Grade, sondern es genügt ein Resultat im Stile von Theorem A in [9].

6. Das Hilfsideal

In diesem Abschnitt werden wir in allgemeiner Form die Konstruktion des Hilfsideals durchführen. Dies werden wir später im Transzendenzbeweis benötigen. Von nun an sei der Grundkörper wieder der in Abschnitt 1 eingeführte algebraische Zahlkörper K.

Um dies durchzuführen sei $\mathcal{O}$ die Ordnung, die wir in Abschnitt 1 eingeführt haben. Wir halten dann ein für alle Mal eine Menge von freien Erzeugenden $e_1, \ldots, e_{nm}$ von $\mathcal{O}$ über $\mathbb{Z}$ fest, die wir auch zur Definition von $\mathcal{O}^\kappa(S)$ benutzen. Diese bilden dann eine Basis von $\mathcal{O} \otimes \mathbb{Q}$ über $\mathbb{Q}$. In Abschnitt 1 haben wir weiter eine Operation von $\mathcal{O}$ als Endomorphismen von $\mathbb{G}_a$ und A^n definiert. Auf $\mathbb{G}_a$, wir erinnern nochmals daran, ist dies einfach die Multiplikation mit einer komplexen Zahl. Ihr algebraischer Grad ist daher gleich 1. Wir betrachten nun die Gruppe $G = \mathbb{G}_a \times A^n$ und für eine positive ganze Zahl κ die Gruppe G^κ. Wir betten G in das Produkt von projektiven Räumen $\mathbb{P} \times \mathbb{P}^N$ wie üblich ein. Dadurch wird die Gruppe G^κ in den Raum $(\mathbb{P} \times \mathbb{P}^N)^\kappa$ eingebettet. Seien $\mathbf{Z}_j = (Z_{j,0}, Z_{j,1})$ für $1 \leqq j \leqq \kappa$ projektive Koordinaten von $\mathbb{P}$ und $\mathbf{X}_j = (X_{j,0}, \ldots, X_{j,N})$ projektive Koordinaten von $\mathbb{P}^N$. Weiter seien $\mathscr{A}_1, \ldots, \mathscr{A}_u$ für eine positive ganze Zahl u ein vollständiges System von Additionsformeln, die den Morphismus der Addition vollständig beschreiben. Sei $P(\mathbf{Z}_1, \mathbf{X}_1, \ldots, \mathbf{Z}_\kappa, \mathbf{X}_\kappa)$ ein multihomogenes Polynom und für ein weiteres System von affinen Variablen $\mathbf{z}_j'$ und $\mathbf{x}_j'$ wie in Abschnitt 2 $P \circ \mathscr{A}_i$ das Polynom

$$(P \circ \mathscr{A}_i)(\mathbf{Z}_1, \mathbf{X}_1, \ldots, \mathbf{Z}_\kappa, \mathbf{X}_\kappa, \mathbf{z}_1', \mathbf{x}_1', \ldots, \mathbf{z}_\kappa', \mathbf{x}_\kappa').$$

Dies ist wieder ein multihomogenes Polynom in denselben Variablen. Die Differentialoperatoren $\Delta'_1, \ldots, \Delta'_\kappa$ aus Abschnitt 5 mit den Bezeichnungen aus Abschnitt 3 operieren auf den inhomogenen Variablen $\mathbf{z}'_1, \ldots, \mathbf{x}'_\kappa$. Wie üblich setzen wir für nichtnegative ganze Zahlen $t_1, \ldots, t_\kappa$

$$\Delta'^t = \Delta_1'^{t_1} \ldots \Delta_\kappa'^{t_\kappa}$$

mit $t=(t_1, \ldots, t_\kappa)$. Schließlich definieren wir für $\gamma_1, \ldots, \gamma_\kappa$ in $\mathcal{O}$ den Endomorphismus $\gamma=(\gamma_1, \ldots, \gamma_\kappa)$ von G^κ durch $g=(g_1, \ldots, g_\kappa) \mapsto \gamma g=(\gamma_1 g_1, \ldots, \gamma_\kappa g_\kappa)$. Dann folgt aus der Proposition 1 zusammen mit der Bemerkung nach ihrem Beweis und der Tatsache, daß auf dem linearen Faktor von G^κ diese Operation Multiplikation mit einer komplexen Zahl ist, daß es eine endliche offene Überdeckung $\{U_i\}$ von G gibt, so daß der Morphismus $[\gamma]$, der durch γ auf G^κ induziert wird, auf U_i eine überall definierte rationale Abbildung $[\gamma]_i$ ist. Diese wird gegeben durch homogene Polynome wie in der Proposition 1 in den Variablen $\mathbf{X}_j$ und durch lineare Polynome in den Variablen $\mathbf{Z}_j$. Ist $\gamma_j = s_{j,1} e_1 + \ldots + s_{j,nm} e_{nm}$ für $1 \leqq j \leqq \kappa$ und P ein multihomogenes Polynom vom Multigrad $(D_1, L_1, \ldots, D_\kappa, L_\kappa)$, so ist das Polynom $[\gamma]_i^* (P)$ multihomogen vom Multigrad

$$(D_1, (s_{1,1}^2 + \ldots + s_{1,nm}^2)\, d_i L_1, \ldots, D_\kappa, (s_{\kappa,1}^2 + \ldots + s_{\kappa,nm}^2)\, d_i L_\kappa).$$

Nun definieren wir für ganze Zahlen $S \geqq 0$ und $T \geqq 1$ das Ideal $I(S, T)$ durch

$$I(S, T) = (\Delta'^t [\gamma]_i^* P \circ \mathscr{A}_j(\mathbf{Z}_1, \mathbf{X}_1, \ldots, \mathbf{Z}_\kappa, \mathbf{X}_\kappa; \mathbf{z}'_1(0), \ldots, \mathbf{x}'_\kappa(0)))$$

für $t=(t_1, \ldots, t_\kappa)$ mit $t_1 + \ldots + t_\kappa = |t| \leqq T$, γ in $\mathcal{O}^\kappa(S)$, $1 \leqq j \leqq u$. Dabei durchläuft die Zahl i die endliche Indexmenge der offenen Überdeckung $\{U_i\}$.

Dies ist ein multihomogenes Ideal und wir bezeichnen mit $I(S, T)(1)$ das Ideal, das in $\mathbb{C}[\mathbf{X}_1, \ldots, \mathbf{X}_\kappa]$ von denselben Polynomen wie $I(S, T)$ erzeugt wird, wobei nur die Variablen $\mathbf{Z}_1, \ldots, \mathbf{Z}_\kappa$ durch $(1, 1), \ldots, (1, 1)$ ersetzt werden.

Proposition 3. *Es sei h die Höhe des multihomogenen Polynoms $P(\mathbf{Z}_1, \mathbf{X}_1, \ldots, \mathbf{Z}_\kappa, \mathbf{X}_\kappa)$, dessen Multigrad in den $\mathbf{Z}_i$ höchstens D und in den $\mathbf{X}_i$ höchstens L für $1 \leqq i \leqq \kappa$ sei. Dann wird das Ideal $I(S, T)$ erzeugt von multihomogenen Polynomen mit Höhen höchstens gleich*

$$T!\, c_1^{S^2 L + T + D \log S}\, h$$

für eine Konstante c_1, die nur von der Gruppe G^κ abhängt, deren Multigrad höchstens D in den Variablen $\mathbf{Z}_i$ und $c_2 L S^2$ in den Variablen $\mathbf{X}_i$ ist für $1 \leqq i \leqq \kappa$. Dies gilt insbesondere für das Ideal $I(S, T)(1)$, das multihomogen in den Variablen $\mathbf{X}_i$ vom Multigrad höchstens $(c_2 L S^2, \ldots, c_2 L S^2)$ ist, wo c_2 wiederum nur von der Gruppe G^κ abhängt.

Beweis. Die Abschätzungen für den Grad der Erzeugenden des Ideals $I(S, T)$ und a fortiori des Ideals $I(S, T)(1)$ folgen unmittelbar aus der Proposition 1. Die Abschätzungen für die Höhen dieser Erzeugenden folgen ebenfalls aus dieser Proposition zusammen mit der wohlbekannten Tatsache, daß der zusätzliche Beitrag, welcher von den Differentialoperatoren kommt, höchstens gleich $T!\, c_3^{T + L + D \log S}$ ist, wo c_3 ebenfalls nur von der Gruppe abhängt.

Bis jetzt hängen die Ideale $I(S, T)$ und $I(S, T)(1)$ noch von dem Polynom P ab. Wir werden es nun festlegen. Um dies zu tun, müssen wir die Parameter κ, S, T, D und L, die in dem Ideal $I(S, T)$ auftauchen, einschränken. Wir wählen eine positive Zahl $1/4 > \varepsilon > 0$ und definieren die positive ganze Zahl κ durch

$$\kappa = [N/\varepsilon] + 1. \tag{13}$$

Weiter wählen wir hinreichend große ganze Zahlen L und S mit $L^{nd/4\varepsilon} \geqq S \geqq L^{d/m}$. Es sei D ein Parameter, welcher nach unendlich geht. Dann setzen wir

$$T = [DL^{nd-\varepsilon}/S^{mn+2\varepsilon}]. \tag{14}$$

Wir bemerken hier, daß die Wahl von κ und T nach sich zieht, daß für eine beliebige Konstante $c_4 > 0$ gilt

$$T^{\kappa} S^{mn\kappa} (LS^2)^N \leqq c_4 (DL^{nd})^{\kappa} \tag{15}$$

sobald nur L oder S hinreichend groß bezüglich dieser Konstanten sind.

Es sei $y = (y_0, \ldots, y_N)$ ein weiteres System von unabhängigen Variablen. Damit definieren wir das multihomogene Ideal $\mathfrak{m}(y)$ durch

$$\mathfrak{m}(y) = (I(G^{\kappa}), Z_{i,0} - Z_{i,1}, X_{i,j} y_k - X_{i,k} y_j;\ 1 \leqq i \leqq \kappa, 0 \leqq k, j \leqq N).$$

Dies ist ein Ideal der Dimension Null in dem Polynomring $K[y][\mathbf{Z}_1, \ldots, \mathbf{X}_{\kappa}]$ über $K[y]$. Das folgende Lemma beendet die Konstruktion des Hilfsideals.

Lemma 4. *Es gibt ein multihomogenes Polynom mit ganzzahligen Koeffizienten vom Multigrad höchstens D in den $\mathbf{Z}_i$ und L in den $\mathbf{X}_i$ für $1 \leqq i \leqq \kappa$, das nicht in $I(G^{\kappa})$ liegt und dessen Höhe höchstens gleich*

$$T!\, c_5^{S^2 L + D \log S}$$

ist, so daß

$$(I(G^{\kappa}), I(S, T)) \subset \mathfrak{m}(y). \tag{16}$$

Bemerkung. Ist h ein Punkt aus A^n und spezialisieren wir y zu $(X_0(h), \ldots, X_N(h))$, so bedeutet (16) nichts anderes, als daß das fragliche Polynom in $(\mathcal{O}(S)(1, 1, h))^{\kappa}$ mindestens mit der Ordnung T verschwindet.

Beweis. Da die Dimension von A^n gleich nd ist, gibt es $nd+1$ unter den Variablen $\mathbf{X}_i$ für jedes i mit $1 \leqq i \leqq \kappa$, die algebraisch unabhängig modulo dem Ideal von G sind. Ohne Beschränkung der Allgemeinheit seien dies die ersten $nd+1$ Variablen, d.h.

$$\mathbf{X}'_i := (X_{i,0}, \ldots, X_{i,dn+1}) \qquad (1 \leqq i \leqq \kappa).$$

Dann gilt für jedes von Null verschiedene multihomogene Polynom $P(\mathbf{Z}_1, \mathbf{X}'_1, \ldots, \mathbf{Z}_{\kappa}, \mathbf{X}'_{\kappa})$

$$P \not\equiv 0 \qquad (\operatorname{mod} I(G^{\kappa})).$$

Wir nehmen nun ein multihomogenes Polynom solch einer Gestalt mit Multigrad D in den Variablen $\mathbf{Z}_i$ und L in den Variablen $\mathbf{X}'_i$ für $1 \leqq i \leqq \kappa$ mit unbestimmten Koeffizienten. Dei Bedingung (16) ist äquivalent zu einem System von homogenen Gleichungen für diese Koeffizienten. Die Zahl dieser Gleichungen ist höchstens gleich $c_6 T^\kappa S^{nm\kappa}(LS^2)^N$. Die Koeffizienten dieses Systems liegen in einem algebraischen Zahlkörper und werden in der Proposition 1 abgeschätzt. Wegen (15) können wir das Siegelsche Lemma anwenden und erhalten ein von Null verschiedenes Polynom $P(\mathbf{Z}_1, \ldots, \mathbf{X}'_\kappa)$, dessen zugehöriges Ideal die Bedingung (16) erfüllt. Dies beweist das Lemma.

7. Die Nullstellenabschätzung

In diesem Abschnitt werden wir den Hauptschritt im Beweis von Satz 1 tun. Wir werden zeigen, daß für geeignetes $S' \geqq S$ das Ideal $I(S', T)(1)$ keine Nullstelle auf $A^{n\kappa}$ besitzt. Wir werden dies mit Hilfe des Korollars von Satz 2 beweisen. Um dieses Ergebnis anzuwenden, müssen wir eine weitere Bedingung an die Wahl unserer Parameter stellen. Wir wählen die Parameter, so daß

$$T \geqq L^{nd+\varepsilon} \tag{17}$$

gilt. Da L groß genug gewählt war, impliziert diese Bedingung die Bedingung (10). Wir können nun die Nullstellenabschätzung angeben.

Proposition 4. *Es sei $S' = S(LS^2)^{2\varepsilon/mn}$. Dann hat unter der Voraussetzung* (17) *das Ideal $I(S', T)(1)$ keine Nullstelle auf $A^{n\kappa}$.*

Beweis. Wir nehmen an, daß das Ideal $I(S', T)(1)$ eine Nullstelle g'' besitzt. Dann können wir g'' schreiben in der Form $(g''_1, \ldots, g''_\kappa)$ für $g''_1, \ldots, g''_\kappa$ in A^n. Nun setzen wir $g'_i = (1, 1)$ für $1 \leqq i \leqq \kappa$ und $g_i = (g'_i, g''_i)$. Dann liegt g_i in G für $1 \leqq i \leqq \kappa$. Da $I(S', T)(1)$ in g'' verschwindet, verschwindet das Ideal $I(S', T)$ in $g = (g_1, \ldots, g_\kappa)$. Wir können nun erreichen, daß der Orbit $\mathcal{O} g$ von g ganz in der offenen Menge $X_{0,1} \ldots X_{0,\kappa} \neq 0$ liegt. Dies erreichen wir mit einer Änderung der Einbettung von A^n in $\mathbb{P}^N$ in der folgenden Weise. Sind $X_0, \ldots, X_N$ die projektiven Koordinaten in $\mathbb{P}^N$, so wählen wir eine Linearform $L_0(X_0, \ldots, X_N)$ mit algebraischen Koeffizienten mit $\frac{\partial}{\partial X_0} L_0 \neq 0$ und $L_0(X_0, \ldots, X_N)\,(\gamma g) \neq 0$ für alle γ aus $\mathcal{O}$. Dies ist möglich, da wir die projektiven Koordinaten der Elemente aus $\mathcal{O} g$ so wählen können, daß sie in einem über $\mathbb{Q}$ endlich erzeugten Körper liegen. Wir definieren nun die lineare Abbildung L als $(Y_0, \ldots, Y_N) = (L_0, X_1, \ldots, X_N)$ und dies sind die neuen Koordinaten. Dieser Koordinatenwechsel induziert einen Wechsel der homogenen Koordinaten auf jedem Faktor der Gestalt $\mathbb{P} \times \mathbb{P}^N$. Wir bezeichnen diese Koordinatentransformation von $(\mathbb{P} \times \mathbb{P}^N)^\kappa$ mit $\mathscr{L}$ und mit $\mathscr{L}^{-1}$ die inverse. Dann definieren wir für multihomogene Polynome F das Polynom $\mathbf{F}_L$ durch $F \circ \mathscr{L}^{-1}$. Die neuen Additionsformeln $\mathscr{A}_{i,L}$ sind dann definiert durch

$$\mathscr{A}_{i,L} = (\mathscr{L} \times \mathscr{L}) \circ \mathscr{A} \circ (\mathscr{L}^{-1} \times \mathscr{L}^{-1}).$$

Da das Ideal $I(S', T)$ in g verschwindet, verschwinden die Polynome

$$(P \circ \mathscr{A}_j)_\gamma := P \circ \mathscr{A}_j(\mathbf{Z}_1(\gamma g), \ldots, \mathbf{X}_\kappa(\gamma g); \mathbf{Z}'_1, \ldots, \mathbf{X}'_\kappa)$$

mindestens mit der Ordnung $T+1$ in 0 für alle γ aus $\mathcal{O}^\kappa(S')$. Es folgt, daß das Polynom

$$(P_L \circ \mathcal{A}_{j,L})_\gamma = ((P \circ \mathcal{A}_j)_L)_\gamma$$

mindestens mit der Ordnung $T+1$ in 0 verschwindet. Nun können wir die Proposition 2 benützen, um die Derivationen $\Delta_1, \ldots, \Delta_\kappa$ und $\Delta'_1, \ldots, \Delta'_\kappa$ bezüglich der neuen Variablen in Verbindung zu bringen. Zunächst ergibt sich

$$\Delta'^t (P \circ \mathcal{A}_j)_L(\gamma g, 0) = 0$$

für alle $t=(t_1, \ldots, t_\kappa)$ mit $t_1 + \ldots + t_\kappa \leqq T$ und alle γ aus $\mathcal{O}^\kappa(S')$. Wegen der Proposition 2 und wegen $Y_{0,i}(\gamma g) \neq 0$ für alle i und alle γ folgt, daß das Polynom

$$P_L / Z_{0,1}^{D_1} \ldots Z_{0,\kappa}^{D_\kappa} Y_{0,1}^{L_1} \ldots Y_{0,\kappa}^{L_\kappa}$$

mindestens mit der Ordnung $T+1$ in $\mathcal{O}^\kappa(S')g$ verschwindet, wobei D_i der Grad von P in den Variablen Z_i für $1 \leqq i \leqq \kappa$ ist und L_i der Grad von P in den Variablen X_i ebenfalls für $1 \leqq i \leqq \kappa$.

Mit anderen Worten haben wir hergeleitet, daß das Polynom P_L auf der Menge $\mathcal{O}^\kappa(S')$ mindestens mit der Ordnung T verschwindet. Dies ist genau die Situation in dem Korollar von Satz 2. Die Wahl der Parameter S', T, D und L implizieren unmittelbar die Bedingungen (10) und (11) und deswegen hat mindestens eines der Elemente g'_i für $1 \leqq i \leqq \kappa$ das neutrale Element der Gruppe $\mathbb{G}_a$ zu sein. Aber dies heißt nichts anderes als $g'_i = (1, 0)$ und dies ist ein Widerspruch zur Wahl von g'_i. Deswegen hat das multihomogene Ideal $I(S', T)$ keine Nullstelle auf der Gruppe $A^{n\kappa}$. Damit ist die Proposition beweisen.

8. Das Approximationsideal

In Abschnitt 6 haben wir das Hilfsideal $I(S, T)$ und in Abschnitt 7 das Ideal $I(S', T)(1)$ eingeführt, von dem wir gezeigt haben, daß es auf $A^{n\kappa}$ keine Nullstelle besitzt. Wir nennen es das Approximationsideal und werden diese Bezeichnung später rechtfertigen. Wir beginnen nun damit, den Punkt α in $T(A^n)$ in Erinnerung zu rufen und setzen

$$g_i = (1, 1, \exp_{A^n}(\alpha))$$

für $1 \leqq i \leqq \kappa$ und weiter $g = (g_1, \ldots, g_\kappa)$. Dann definieren wir wie gewöhnlich $\gamma \cdot g$ für γ in $\mathcal{O}^\kappa$. Das Ideal $I(S, T)$ war so konstruiert, daß es auf $\mathcal{O}^\kappa(S) \cdot g$ verschwindet.

Proposition 5. *Es gibt eine positive Konstante c_6, die nur von G und α abhängt, so daß für die ausgezeichneten Erzeugenden $\Delta'^t[\gamma]_i^* P \circ \mathcal{A}_j$ von $I(S', T)$ gilt*

$$\log |\Delta'^t[\gamma]_i^* \mathcal{P} \circ \mathcal{A}_j(g, 0)| \leqq -c_7 T S^{mn} \log T \tag{18}$$

für alle i, j, für alle nichtnegativen ganzen Zahlen $t_1, \ldots, t_\kappa$ mit $t_1 + \ldots + t_\kappa \leqq T$ und für alle γ in $\mathcal{O}^\kappa(S')$. Insbesondere genügen die ausgezeichneten Erzeugenden von $I(S', T)(1)$ dieser Abschätzung.

Beweis. Da nach Konstruktion das Ideal $I(S,T)$ auf $\mathcal{O}^\kappa(S)\cdot g$ verschwindet, erhalten wir für alle nichtnegativen ganzen Zahlen $t_1,\ldots,t_\kappa$ mit $t_1+\ldots+t_\kappa\leqq T$ und für γ in $\mathcal{O}^\kappa(S)$

$$\Delta'^t[\gamma]_i^* P\circ\mathcal{A}_j(g,0)=0.$$

Da wir annehmen können, daß für jedes γ in $\mathcal{O}^\kappa$ gilt $Z_{0,1}\ldots Z_{0,\kappa}X_{0,1}\ldots X_{0,\kappa}(\gamma g)\neq 0$, so können wir mit Hilfe von Proposition 2 dies auf die Derivationen $\Delta_i,\ldots,\Delta_\kappa$ übertragen und erhalten

$$\Delta^t(P/Z_{0,1}^{D_1}X_{0,1}^{L_1}\ldots Z_{0,\kappa}^{D_\kappa}X_{0,\kappa}^{L_\kappa})(\gamma g)=0 \tag{19}$$

für γ in $\mathcal{O}^\kappa(S)$ und alle nichtnegativen ganzen Zahlen $t_1,\ldots,t_\kappa$ mit $t_1+\ldots+t_\kappa\leqq T$. Daraus folgt, daß das Polynom P auf $\mathcal{O}^\kappa(S)$ mindestens mit der Ordnung T verschwindet. Nun können wir das Schwarzsche Lemma für Produkte anwenden (s. [18], Proposition 7.2.1).

Es sei $\phi:\mathbb{C}\to A^n$ die analytische Darstellung der analytischen Untergruppe B, die zu Beginn eingeführt wurde. Wir können ohne Beschränkung der Allgemeinheit annehmen, daß B bereits $\mathcal{O}g$ enthält. Dann ist $\phi=(\phi_0,\ldots,\phi_N)$, und wir können die Darstellung des Produktes B^κ schreiben als

$$\Phi=(\phi^1,\ldots,\phi^\kappa)\colon\mathbb{C}^\kappa\to A^{n\kappa},$$

wobei $\phi^i=\phi$ ist für $1\leqq i\leqq\kappa$. Seien $z_1,\ldots,z_\kappa$ die Koordinaten von B^κ. Dies ist nach Voraussetzung eine κ-dimensionale Untergruppe von $A^{n\kappa}$. Nun ersetzen wir in dem Polynom P die Variablen $\mathbf{Z}_i$ durch $(1,z_i)$ und $\mathbf{X}_i$ durch $\phi^i(z_i)$. Dadurch erhalten wir eine analytische Funktion in den Variablen $z_1,\ldots,z_\kappa$. Die algebraischen Differentialoperatoren $\Delta_1,\ldots,\Delta_\kappa$ entsprechen den analytischen Differentialoperatoren $\partial/\partial z_1,\ldots,\partial/\partial z_\kappa$. Dann folgt aus (19), daß die daraus resultierende Funktion $\Psi(z_1,\ldots,z_\kappa)$ eine Nullstelle mindestens der Ordnung T in den Punkten $\gamma(1,\ldots,1)=(\gamma_1,\ldots,\gamma_\kappa)$ für Elemente $\gamma=(\gamma_1,\ldots,\gamma_\kappa)$ in $\mathcal{O}^\kappa(S)$ hat. Die Funktionen, die in Φ auftauchen, haben alle die Wachstumsordnung zwei. Setzen wir $R=T^{1/4}$, so finden wir mit Hilfe des Schwarzschen Lemmas für Produkte und der Cauchyschen Ungleichung, daß für alle nichtnegativen ganzen Zahlen $t_1,\ldots,t_\kappa$ mit $t_1+\ldots+t_\kappa\leqq T$

$$\log\left|\left(\frac{\partial}{\partial z_1}\right)^{t_1}\ldots\left(\frac{\partial}{\partial z_\kappa}\right)^{t_\kappa}(\Psi/(\phi_0^1)^{L_1}\ldots(\phi_0^\kappa)^{L_\kappa})\,\gamma(1,\ldots,1)\right|\leqq -c_8\,TS^{mn}\log T.$$

gilt für γ in $\mathcal{O}^\kappa(S')$. Dies können wir auch in der Form

$$\log|\Delta_1^{t_1}\ldots\Delta_\kappa^{t_\kappa}(P/Z_{0,1}^{D_1}\ldots X_{0,\kappa}^{L_\kappa})(\gamma g)|\leqq -c_8\,TS^{mn}\log T \tag{20}$$

schreiben. Um (18) zu erhalten, müssen wir nochmals die Proposition 2 anwenden. Eine einfache Rechnung zeigt dann, daß

$$\log|\Delta_1'^{t_1}\ldots\Delta_\kappa'^{t_\kappa}[\gamma]_i^* P\circ\mathcal{A}_j(g,0)|\leqq -c_9\,TS^{mn}\log T$$

gilt. Daraus folgt die behauptete Ungleichung in Proposition 5. Die zweite Behauptung ergibt sich unmittelbar hieraus und der Definition von $I(S',T)(1,1)$.

9. Vollständige Durchschnitte

Ein wichtiges Hilfsmittel, um am Ende den gewünschten Widerspruch zu erhalten, wird die Resultante sein. Dazu ist es nützlich, das folgende Resultat zu haben. Wir betrachten für eine nichtnegative ganze Zahl M den Polynomring $K[Y_0, \ldots, Y_M]$ und darin ein Ideal, welches von einer großen Zahl von homogenen Polynomen erzeugt wird. Wir möchten gerne die Anzahl der Erzeugenden möglichst weit reduzieren, ohne aber den Rang des Ideals zu verändern. Dies ist der Inhalt der folgenden Proposition.

Proposition 6. *Es sei v eine nichtnegative ganze Zahl und für eine positive ganze Zahl l seien $I_1, \ldots, I_l$ homogene Ideale in dem Ring $K[Y_0, \ldots, Y_M]$ mit den folgenden Eigenschaften. Für $1 \leqq \lambda \leqq l$ sei das Ideal I_λ durch homogene Polynome $P_{\lambda,1}, \ldots, P_{\lambda,v}$ vom Grade D_λ erzeugt. Seien $\rho_1, \ldots, \rho_l$ ganze Zahlen, so daß $r_\lambda = \rho_1 + \ldots + \rho_\lambda$ der Rang des Ideals $(I_1, \ldots, I_\lambda)$ ist für $1 \leqq \lambda \leqq l$. Dann gibt es für $1 \leqq \lambda \leqq l$, $1 \leqq \mu \leqq \rho_\lambda$ ganze Zahlen $u_{\lambda,\mu,\nu}$ mit $1 \leqq \nu \leqq v$ und den folgenden Eigenschaften. Der Absolutbetrag von $u_{\lambda,\mu,\nu}$ ist höchstens gleich $D_1^{\rho_1} \ldots D_\lambda^{\rho_\lambda}$ und für $1 \leqq \lambda \leqq l$, $1 \leqq \mu \leqq \rho_\lambda$ sind höchstens $D_1^{\rho_1} \ldots D_\lambda^{\rho_\lambda}$ der Zahlen $u_{\lambda,\mu,1}, \ldots, u_{\lambda,\mu,v}$ von Null verschieden. Setzen wir*

$$F_{\lambda,\mu} = \sum_{\nu=1}^{v} u_{\lambda,\mu,\nu} P_{\lambda,\nu}$$

für $1 \leqq \lambda \leqq l$, $1 \leqq \mu \leqq \rho_\lambda$, dann hat das Ideal

$$(F_{1,1}, \ldots, F_{1,\rho_1}, \ldots, F_{\lambda,1}, \ldots, F_{\lambda,\rho_\lambda})$$

den Rang r_λ für $1 \leqq \lambda \leqq l$.

Beweis. Wir beweisen diese Proposition durch Induktion. Wir können zunächst ohne Beschränkung der Allgemeinheit annehmen, daß alle ρ_λ von Null verschieden sind für $1 \leqq \lambda \leqq l$. Dann betrachten wir zuerst den Fall $\lambda = 1$. Hier können wir annehmen, daß wir für ein ρ mit $1 \leqq \rho \leqq \rho_1$ wir bereits die Polynome $F_{1,1}, \ldots, F_{1,\rho}$ so konstruiert haben, daß das von ihnen erzeugte Ideal I den Rang ρ besitzt. Gilt $\rho = \rho_1$, so sind wir fertig. Andernfalls können wir die Anzahl der Primärkomponenten von I durch D_1^ρ abschätzen. Da I_1 den Rang ρ_1 besitzt, finden wir für jede Primkomponente von I eines der I_1 erzeugenden Polynome, welches nicht in dieser Primkomponente liegt. Den Fall $\lambda = 1$ erledigt man nun mit dem Lemma 1 aus Kap. IV in [11].

Wir können nun annehmen, daß die Proposition bereits für $1 \leqq \lambda < l$ bewiesen ist. Dann betrachten wir den Fall $\lambda + 1$. Wir bezeichnen mit I das von $F_{1,1}, \ldots, F_{\lambda,\rho_\lambda}$ erzeugte Ideal. Die Anzahl der Primärkomponenten ist höchstens gleich dem Grad von I, welcher seinerseits durch $D_1^{\rho_1} \ldots D_\lambda^{\rho_\lambda}$ beschränkt ist. Da I in $(I_1, \ldots, I_\lambda)$ enthalten ist, hat es den Rang r_λ, und es gibt für jede Primkomponente von I eines der Erzeugenden von $I_{\lambda+1}$, welches nicht in dieser Primkomponente enthalten ist. Wiederum erledigt das Lemma 1 aus dem Kap. IV in [11] den vorliegenden Fall. Wir können daher annehmen, daß für $1 \leqq \rho \leqq \rho_{\lambda+1}$ die Polynome $F_{1,1}, \ldots, F_{\lambda+1,\rho}$ bereits so konstruiert sind, daß sie ein Ideal J vom Rang $r_\lambda + \rho$ erzeugen und die in der Proposition geforderten Eigenschaften besitzen. Wir wollen dann das Polynom $F_{\lambda+1,\rho+1}$ konstruieren. Wiederum ist die Anzahl der Primärkomponenten von J höchstens gleich dem Grad, welcher durch $D_1^{\rho_1} \dots D_\lambda^{\rho_\lambda} D_{\lambda+1}^{\rho}$ abgeschätzt werden kann. Wieder können wir annehmen, daß $\rho < \rho_{\lambda+1}$ gilt. Da der Rang von J gleich $r_\lambda + \rho$ ist, finden wir wieder in dem Ideal $I_{\lambda+1}$ für jede Primärkomponente von J eines der Erzeugenden von $I_{\lambda+1}$, welches nicht in der entsprechenden Primkomponente liegt. Wiederum gibt uns das Lemma 1 aus Kap. IV in [11] das gewünschte Polynom $F_{\lambda+1,\rho+1}$. Damit ist der Induktionsschritt getan und die Proposition bewiesen.

Wir werden die Proposition in der folgenden Situation anwenden. Sei $I(A^{n\kappa})$ das Ideal von $A^{n\kappa}$ in dem Ring $K[\mathbf{X}_1, \dots, \mathbf{X}_\kappa]$ und J das Ideal, welches zur Diagonalen in $(\mathbb{P}^N)^\kappa$ gehört. Es wird erzeugt von den Polynomen $X_{i,j} X_{l,k} - X_{l,j} X_{i,k}$ für $1 \leqq i,\ l \leqq \kappa,\ 0 \leqq j,\ k \leqq N$. Weiter sei $I(S', T)(1)$ das Ideal, das in Abschnitt 8 konstruiert worden war.

Das Produkt von projektiven Räumen $(\mathbb{P}^N)^\kappa$ kann in dem projektiven Raum $\mathbb{P}^M$ für $M = (N+1)^\kappa - 1$ mit Hilfe der Segre-Einbettung eingebettet werden (siehe [6]). Die entsprechenden Koordinaten bezeichnen wir mit $Y_0, \dots, Y_M$. Die Varietät V in dem Raum $(\mathbb{P}^N)^\kappa$ sei durch die Polynome $G_i(\mathbf{X}_1, \dots, \mathbf{X}_\kappa)$ für $1 \leqq i \leqq m$ definiert. Dann können wir ohne Einschränkung der Allgemeinheit annehmen, daß für $1 \leqq i \leqq m$ der Grad von G_i in jedem Block von Variablen derselbe ist. Denn ist $D_{i,j}$ der Grad in den Variablen $X_{j,0}, \dots, X_{j,N}$, so setzen wir $D_i = \max_{1 \leqq j \leqq \kappa} (D_{i,j}) + 1$. Dann können wir G_i mit allen möglichen Monomen vom Multigrad $D_i - D_{i,j}$ in den Variablen $X_{j,0}, \dots, X_{j,N}$ multiplizieren. Diese haben keine gemeinsame Nullstelle in $(\mathbb{P}^N)^\kappa$ und so erhalten wir neue Polynome $G'_1, \dots, G'_{m'}$ für ein neues m'. Jedes G'_ν ist ein Produkt eines G_μ mit einem Monom. Die Segre-Abbildung wird gegeben durch

$$X_{1,i_1} \dots X_{\kappa,i_\kappa} \mapsto Y_{i_1,\dots,i_\kappa} =: Y_j$$

für $0 \leqq i_1, \dots, i_\kappa \leqq N$ und eine beliebige Anordnung $Y_0, \dots, Y_M$ der obigen Variablen. Da der Grad von G_i für $1 \leqq i \leqq m$ in jedem Block von Variablen derselbe ist, können wir die Polynome $G_1, \dots, G_m$ schreiben als

$$G_i(\mathbf{X}_1, \dots, \mathbf{X}_\kappa) = H_i(Y_0, \dots, Y_M)$$

für $1 \leqq i \leqq m$ und homogene Polynome H_i vom selben Grad wie die Polynome G_i in $X_{j,0}, \dots, X_{j,N}$. Die Polynome $H_1, \dots, H_m$ sind natürlich nicht eindeutig bestimmt. Jedoch sind sie eindeutig modulo des definierenden Ideals des Bildes von $(\mathbb{P}^N)^\kappa$ in $\mathbb{P}^M$ unter der Segre-Abbildung. Dieses Ideal wird erzeugt in $K[Y_0, \dots, Y_M]$ durch Polynome der Form $Y_i Y_j - Y_{i'} Y_{j'}$ für gewisse i, j, i', j' (für Details s. [6]). Wir bezeichnen dieses Ideal mit I_1. Auf diese Weise erhalten wir aus dem Ideal $I(A^{n\kappa})$ ein Ideal I_2 und aus den Idealen J und $I(S', T)(1)$ die Ideale I_3 und I_5 in dem Polynomring $K[Y_0, \dots, Y_M]$. Wir wählen nun ein beliebiges multihomogene Ideal E in $K[\mathbf{X}_1, \dots, \mathbf{X}_\kappa]$ mit assoziiertem Ideal I_4 in $K[Y_0, \dots, Y_M]$. Gilt mit denselben Bezeichnungen wie in der Proposition 6 $\rho := \rho_4 = r_4 - r_3 > 0$, so erhalten wir die folgende Konsequenz aus der Proposition 6.

Lemma 5. *Wird das Ideal E durch multihomogene Polynome vom Multigrad höchstens D_0 und mit Höhe höchstens H_0 erzeugt, so gibt es homogene Polynome*

$F_1, \ldots, F_{M-nd}$, $G_1, \ldots, G_\rho$, $H_1, \ldots, H_{nd+1-\rho}$ in $K[Y_0, \ldots, Y_M]$ mit den folgenden Eigenschaften. Die Polynome $F_1, \ldots, F_{M-nd}$ sind in dem von I_1, I_2 und I_3 erzeugten Ideal und ihre Höhe und ihr Grad sind durch eine Konstante beschränkt. Die Polynome $G_1, \ldots, G_\rho$ sind in I_4 und haben höchstens den Grad D_0+1 und höchstens die Höhe $D_0^{2\rho} H_0$. Die Polynome $H_1, \ldots, H_{nd+1-\rho}$ liegen in I_5 und ihr Grad ist höchstens $c_{13} LS'^2$ und ihre Höhe höchstens $D_0^{2\rho} c_{12}^{T \log T + LS'^2 + D \log S}$ für eine positive Konstante c_{13}. Schließlich ist der Rang des Ideals, welches von all diesen Polynomen erzeugt wird, gleich $M+1$.

Beweis. Das Ideal, das von $I(S', T)(1)$ und $I(A^{n\kappa})$ erzeugt wird hat den Rang $M+1$. Dies folgt aus der Proposition 4. Aufgrund der Eigenschaften der Segre-Einbettung zieht dies nach sich, daß der Rang des von den Idealen $I_1, \ldots, I_5$ erzeugten Ideals ebenfalls gleich $M+1$ ist. Nun folgt das Lemma aus der Proposition 6 zusammen mit den Abschätzungen in Proposition 3 and Lemma 4.

10. Die Resultante

In diesem Abschnitt werden wir das folgende Problem diskutieren. Gegeben seien $N+1$ homogene Polynome $F_0, \ldots, F_N$ in den Variablen $X_0, \ldots, X_N$. Ihre Höhe sei h_i und ihr Grad D_i für $0 \leqq i \leqq N$. Wir nehmen weiter an, daß sie keine nichttriviale gemeinsame Nullstelle besitzen mögen. Dann ist es wohlbekannt, daß das Radikal des von ihnen erzeugten Ideals das Ideal $(X_0, \ldots, X_N)$ ist. Daher gibt es eine positive ganze Zahl $e \geqq 1$ mit der Eigenschaft, daß

$$(X_0, \ldots, X_N)^e \subseteq (F_0, \ldots, F_N)$$

gilt. Insbesondere können wir X_0^e in der Form

$$X_0^e = H_0 F_0 + \ldots + H_N F_N$$

darstellen. Hierin sind $H_0, \ldots, H_N$ homogene Polynome. Wir sind nun interessiert in möglichst guten oberen Schranken für die Höhe und den Grad der Polynome $H_0, \ldots, H_N$.

Wir beginnen mit einem Resultat, welches im wesentlichen von Macaulay [8] stammt. Um es anzugeben, müssen wir einige Bezeichnungen einführen. Es seien $G_0, \ldots, G_N$ homogene Polynome in $X_0, \ldots, X_N$ mit den Graden $D_0, \ldots, D_N$ und unabhängige Variablen als Koeffizienten. Für $0 \leqq i \leqq N$ definieren wir die Zahlen $l_i + 1 = \binom{D_i + N}{N}$ und bezeichnen mit $\mathbf{g}_i = (g_{i,0}, \ldots, g_{i,l_i})$ die Koeffizienten von G_i. Schließlich setzen wir

$$e = D_0 + \ldots + D_N - N + 1$$

und $D = D_0 \ldots D_N$. Dann gilt das folgende.

Proposition 7. *Es gibt ein multihomogenes Polynom*

$$R = R(\mathbf{g}_0, \ldots, \mathbf{g}_N)$$

mit ganzzahligen Koeffizienten vom Grade D/D_i in den Variablen $g_{i,0}, \ldots, g_{i,l_i}$ für $0 \leqq i \leqq N$ mit den folgenden Eigenschaften. Das Polynom R verschwindet für spezielle Werte $\mathbf{g}'_0, \ldots, \mathbf{g}'_N$ von $\mathbf{g}_0, \ldots, \mathbf{g}_N$ dann und nur dann, wenn die spezialisierten Polynome $G'_0, \ldots, G'_N$, die man dadurch erhält, daß man die Koeffizienten $g_{i,j}$ von G_i durch die speziellen Werte $g'_{i,j}$ für $0 \leqq i \leqq N$, $0 \leqq j \leqq l_i$ ersetzt, eine gemeinsame Nullstelle $(\xi_0, \ldots, \xi_N)$ besitzen. Weiter ist die Höhe $h(R)$ von R höchstens gleich $\exp\left(\binom{e+N}{N}\right) \cdot \binom{e+N}{N}!$.

Beweis. Wir betrachten das Ideal in dem Polynomring $Z[\mathbf{g}_0, \ldots, \mathbf{g}_N][X_0, \ldots, X_N]$, welches durch die Polynome $G_0, \ldots, G_N$ erzeugt wird. Es ist ein multihomogenes Ideal vom Rang $N+1$. Es folgt dann aus [8], S. 8, daß es Polynome A und R in $Z[\mathbf{g}_0, \ldots, \mathbf{g}_N]$ mit den folgenden Eigenschaften. Beide Polynome sind koprim und multihomogen. Das Polynom R ist prim und verschwindet dann und nur dann für $\mathbf{g}'_0, \ldots, \mathbf{g}'_N$, wenn die Polynome $G'_0, \ldots, G'_N$, die man dadurch erhält, daß man die $\mathbf{g}_i$ durch die $\mathbf{g}'_i$ ersetzt, eine gemeinsame Nullstelle besitzen. Der Grad von R in den Variablen $g_{i,0}, \ldots, g_{i,l_i}$ ist D/D_i für $0 \leqq i \leqq N$. Das Produkt AR ist eine Determinante einer $U \times U$-Matrix mit $U \leqq \binom{e+N}{N}$, deren Koeffizienten entweder gleich Null oder eine der Variablen ist. Es folgt, daß die Höhe $h(AR)$ durch die Anzahl der Glieder in der Determinante beschränkt ist, welche höchstens gleich $\binom{e+N}{N}!$ ist. Da aufgrund eines wohlbekannten Lemmas (s. z.B. [2], Chap. 12, Lemma 2) gilt

$$h(AR)\, e^{\deg(AR)} \geqq h(R),$$

so erhalten wir für die Höhe $h(R)$ von R die gewünschte Schranke.

Wir werden nun die Proposition 7 dazu benutzen, das folgende Resultat herzuleiten, welches eine große Rolle im Beweis von Satz 1 spielt.

Proposition 8. *Es seien $F_0, \ldots, F_N$ homogene Polynome der Grade $D_0, \ldots, D_N$ und mit Höhen $h_0, \ldots, h_N \geqq 4$, deren Koeffizienten in einem algebraischen Zahlkörper K liegen. Diese Polynome seien ohne gemeinsame projektive Nullstelle. Dann gibt es eine nur von K abhängige positive Konstante $c_{14} \geqq 1$ mit der folgenden Eigenschaft: Es sei X_l für $0 \leqq l \leqq N$ eine homogene Koordinate, ξ ein Punkt in $P^N(\mathbb{C})$ mit $X_l(\xi) \neq 0$ und $D = D_0 \ldots D_N$ gesetzt. Dann gilt*

$$-c_{14}\left(N^2 e^N \log e + \sum_{i=0}^{N} (D/D_i) \log h_i\right) \leqq \max_{0 \leqq i \leqq N} (\log |(F_i/X_l^{D_i})(\xi)|). \tag{21}$$

Beweis. Wir können ohne Beschränkung der Allgemeinheit annehmen, daß $X_l(\xi) = 1$ gilt. Damit identifizieren wir ξ mit einem Punkt in $A^{N+1}(\mathbb{C})$. Ferner können wir ohne Beschränkung der Allgemeinheit annehmen, daß $l = 0$ gilt und setzen $X_0(\xi) = \xi_0$. Schließlich können wir noch annehmen, daß die rechte Seite von (21) negativ ist. Denn andernfalls folgt die Behauptung der Proposition trivialerweise.

Wir ersetzen nun die Koeffizienten von $F_0, \ldots, F_N$ durch Variablen $\mathbf{g}_0, \ldots, \mathbf{g}_N$ und bezeichnen mit $g_0, \ldots, g_N$ die Koeffizienten der höchsten reinen

Potenzen von X_0 in $G_0, \ldots, G_N$. Da für $g_0 = \ldots = g_N = 0$ die Polynome $G_0, \ldots, G_N$ die gemeinsame Nullstelle $(1, 0, \ldots, 0)$ in $\mathbb{P}^N$ besitzen, verschwindet die Resultante R. Wir haben deshalb

$$R = R(\mathbf{g}_0, \ldots, \mathbf{g}_N) \in (g_0, \ldots, g_N).$$

Wir schreiben nun

$$R = \sum_{i_0=0}^{D/D_0} \cdots \sum_{i_N=0}^{D/D_N} r_{i_0, \ldots, i_N} g_0^{i_0} \cdots g_N^{i_N},$$

wobei die Koeffizienten $r_{i_0, \ldots, i_N}$ in $\mathbb{Z}[\mathbf{g}_0, \ldots, \mathbf{g}_N]$ liegen und nicht von $g_0, \ldots, g_N$ abhängen. Ihre Höhen sind beschränkt durch die Höhe $h(R)$ von R.

Ersetzen wir in $G_0, \ldots, G_N$ die Koeffizienten $g_0, \ldots, g_N$ durch $g_0 - G_0(\xi), \ldots, g_N - G_N(\xi)$, so haben die daraus resultierenden Polynome $G_{j,\xi} = G_j(X) - G_j(\xi) X_0^{D_j}$ $(0 \leqq j \leqq N)$ eine gemeinsame Nullstelle, nämlich ξ. Daher verschwindet ihre Resultante. Wir erhalten dann

$$0 = \sum_{i_0=0}^{D/D_0} \cdots \sum_{i_N=0}^{D/D_N} r_{i_0, \ldots, i_N} (g_0 - G_0(\xi))^{i_0} \ldots (g_N - G_N(\xi))^{i_N}.$$

Wir können die rechte Seite als ein Polynom in $G_0(\xi), \ldots, G_N(\xi)$ schreiben. Der konstante Term dieses Polynoms ist gerade die Resultante $R(\mathbf{g}_0, \ldots, \mathbf{g}_N)$ und nun folgt die Proposition aus der Tatsache, daß die obige Relation gültig bleibt, wenn wir die Spezialisierung $\mathbf{g}_i \mapsto \mathbf{f}_i$ für $0 \leqq i \leqq N$ durchführen, wobei die $\mathbf{f}_i = (f_{i,0}, \ldots, f_{i,l_i})$ die Koeffizienten von F_i sind. Der konstante Term $R(\mathbf{f}_0, \ldots, \mathbf{f}_N)$ ist die Spezialisierung der Resultanten $R(\mathbf{g}_0, \ldots, \mathbf{g}_N)$ und verschwindet nicht, da $F_0, \ldots, F_N$ keine gemeinsame Nullstelle besitzen. Die Behauptung in der Proposition 8 folgt nun aus leicht durchzuführende Abschätzungen.

11. Beweis von Satz 1

Wir sind nun in der Lage, den Beweis von Satz 1 zu beenden. Wir betrachten in $(A^n)^\kappa$ den Punkt $\xi = (\exp_{A^n}(\alpha), \ldots, \exp_{A^n}(\alpha))$. Sei W sein Zariski-Abschluß über $\bar{\mathbb{Q}}$. Dies ist eine projektive Varietät in $(\mathbb{P}^N)^\kappa$. Sei E' das entsprechende Ideal in $\bar{\mathbb{Q}}[\mathbf{X}_1, \ldots, \mathbf{X}_\kappa]$. Dann besitzt E' eine Basis von Polynomen, die über einem algebraischen Zahlkörper definiert sind. Wir wollen in Zukunft daher annehmen, daß alle auftretende Objekte über einem algebraischen Zahlkörper K definiert sind. Wir setzen $E = E' \cap K[\mathbf{X}_1, \ldots, \mathbf{X}_\kappa]$. Mit den Bezeichnungen aus Abschnitt 9 erhalten wir die Ideale $I_1, \ldots, I_5$ in $K[Y_0, \ldots, Y_M]$. Wir nehmen nun an, daß der Satz 1 falsch ist und erhalten insbesondere, daß $\rho = \rho_4 > 0$. Nach Lemma 5 erhalten wir Polynome $F_1, \ldots, F_{M-nd}$, $G_1, \ldots, G_\rho$ und $H_1, \ldots, H_{nd+1-\rho}$ in $K[Y_0, \ldots, Y_M]$ mit den folgenden Eigenschaften. Die Polynome $F_1, \ldots, F_{M-nd}$, $G_1, \ldots, G_\rho$ liegen in dem Ideal, welches von den Idealen I_1, I_2, I_3 und I_4 erzeugt wird und ihre Höhe und ihr Grad sind durch eine Konstante c_{15} beschränkt. Sie verschwinden alle auf dem Bild von W in $\mathbb{P}^M$

unter der Segre-Abbildung. Die Polynome $H_1, \ldots, H_{nd-\rho+1}$ schließlich liegen in I_5 und ihr Grad ist höchstens gleich $c_{16} LS'^2$ und ihre Höhe höchstens gleich $c_{17}^{T\log T + LS'^2 + D\log S}$ für positive Konstanten c_{16} und c_{17}. Alle diese Polynome haben keine gemeinsame Nullstelle. Daher können wir die Proposition 8 anwenden und erhalten

$$\max(\log|F_1(\eta)|, \ldots, \log|H_{nd+1-\rho}(\eta)|) \geqq -V, \tag{22}$$

wo $V = c_{18}[(LS'^2)^N \log(LS') + (T\log T + D\log S)(LS'^2)^{nd-\rho}]$ und η das Bild von ξ unter der Segre-Abbildung ist. Da $F_1(\eta) = \ldots = G_\rho(\eta) = 0$ ist, genügt es, die Absolutbeträge von $H_1(\eta), \ldots, H_{nd-\rho+1}(\eta)$ abzuschätzen. Die Polynome H_i sind für $1 \leqq i \leqq nd-\rho+1$ Linearkombinationen von höchstens $c_{19}(LS'^2)^{nd}$ der Erzeugenden von I_5 mit ganzzahligen Koeffizienten, deren Absolutbeträge die Zahl $c_{20}(LS'^2)^{nd}$ nicht übersteigen. Da weiter das Ideal I_5 das Bild von $I(S', T)(1)$ unter der Segre-Abbildung ist, erhalten wir schließlich aus der Proposition 5 die Abschätzung

$$\log|H_i(\eta)| \leqq -c_{21}\, TS^{mn} \log T. \tag{23}$$

Nun vergleichen wir die Abschätzungen (22) und (23). Dazu wählen wir T so groß, daß $D\log S \leqq T\log T$ ist. Dann sind (22) und (23) äquivalent zu der Ungleichung

$$2(nd-\rho) + 8\varepsilon(nd-\rho)/mn \geqq mn.$$

Da dies für jedes positive ε richtig sein muß, muß

$$2(nd-\rho) \geqq mn$$

gelten. Da aber ρ gleich der Kodimension des Zariski-Abschlusses von $\exp_{A^n}(\alpha)$ über $\bar{\mathbb{Q}}$ ist und da die Dimension von A^n gleich nd ist, erhalten wir die gewünschte Abschätzung für die Dimension dieses Zariski-Abschlusses. Damit ist der Satz 1 vollständig bewiesen.

Literatur

1. Anderson, M.: Linear forms in algebraic points of an elliptic function, Ph.D. dissertation (Nottingham, 1978)
2. Baker, A.: Transcendental Number Theory, Cambridge University Press 1975
3. Bertrand, D.: Problèmes locaux. Soc. Math. de France, Astérisque **69–70**, 163–189 (1979)
4. Choodnovsky, G.V.: Algebraic independence of the values of elliptic function at algebraic points. Invent. math. **61**, 267–290 (1980)
5. Hartshorne, R.: Algebraic geometry. Berlin-Heidelberg-New York: Springer 1979
6. Hodge, W.V.D., Pedoe, D.: Methods of algebraic geometry, vol. 1, 2, Cambridge (1952)
7. Lindemann, F.: Über die Zahl π. Math. Ann. **20**, 213–225
8. Macaulay, F.S.: The algebraic theory of modular systems. Stechert-Hafner Service Agency, New York and London 1964 (reprinted)
9. Masser, D.W., Wüstholz, G.: Zero estimates on group varieties I. Invent. math. **64**, 489–516 (1981)
10. Masser, D.W., Wüstholz, G.: Algebraic independence properties of values of elliptic functions, Journées Arith. Exeter (1980), London M.S. Lecture Notes
11. Masser, D.W., Wüstholz, G.: Fields of large transcendence degree generated by values of elliptic functions. Invent. math.

12. Masser, D.W., Wüstholz, G.: Zero Estimates on Group Varieties II. in preparation
13. Philippon, P.: Indépendance algébrique de valeurs de fonctions p-adiques. J. reine u. angew. Math. **329**, 42–51 (1981)
14. Philippon, P.: Indépendance Algébrique et Variétés Abéliennes. C.R. Acad. Sci. Paris, Ser. I, t. 294, 257–259 (1982)
15. Philippon, P.: Variétés Abéliennes et Indépendance Algébrique I. (preprint)
16. Schneider, Th.: Einführung in die transzendenten Zahlen. Berlin-Heidelberg-New York: Springer 1957
17. Serre, JP.: Quelques propriétés des groupes algébriques commutatifs. Soc. Math. de France, Astérisque **69–70**, 191–202 (1979)
18. Waldschmidt, M.: Nombres transcendants et groupes algébriques. Soc. Math. de France Astérisque **69–70**, (1979)
19. Wüstholz, G.: Sur l'analogue abélien du Théorème de Lindemann. C.R. Acad. Sci. Paris Ser. I, t. 295, 35–37 (1982).
20. Wüstholz, G.: Multiplizitätsabschätzungen auf algebraischen Gruppen. Ann. of Math. (in press)
21. Hochschild, G.: The Structure of Lie Groups. San Francisco, London, Amsterdam: Holden-Day 1965

Oblatum 11-VIII & 22-XI-1982

Nachtrag bei der Korrektur

Der Beweis von Satz 1 wurde so geführt, daß man sofort ein Maß für die algebraische Unabhängigkeit erhält. Dazu seien die Voraussetzungen dieselben wie vom Korollar 1.

A^n sei wie üblich in einen projektiven Raum $\mathbb{P}^N$ eingebettet. Dann sei $Y \subset A^n$ eine Hyperfläche vom Grad $d(Y)$ definiert über einem algebraischen Zahlkörper K. Wir definieren dann die Höhe $H(Y)$ von Y als das Minimum über die Höhen aller homogener Polynome mit Koeffizienten in K, welche auf Y verschwinden und deren Grad durch $d(Y)$ beschränkt ist. Ist h ein Punkt auf A^n, so bezeichnen wir mit $\operatorname{dist}(h; Y)$ den Abstand von h zu Y. Dann gilt der folgende Satz.

Satz 3. *Es gibt eine effektiv berechenbare Konstante $c_1 > 0$, welche nicht von $H(Y)$ abhängt, so daß unter den Voraussetzungen von Korollar 1 gilt*

$$\log(\operatorname{dist}(g; Y)) \geqq -c_1(1 + \log H(Y)).$$

Der Beweis von Satz 3 motiviert insbesondere das Lemma 5, welches wir nicht in voller Schärfe im Beweis von Satz 1 ausgenutzt haben. Die Konstante c_1 hängt natürlich von dem Grad $d(Y)$ von Y ab. In einer nachfolgenden Arbeit soll unter anderem diese Abhängigkeit genauer untersucht werden. Dies führt dann zu dem folgenden Ergebnis.

Satz 3'. *Es sei $\varepsilon > 0$ eine positive reelle Zahl. Dann gibt es eine effektive berechenbare Konstante $c_2 > 0$, die nicht von $d(Y)$ und $H(Y)$ abhängt, so daß unter den Voraussetzungen von Korollar 1 gilt*

$$\log(\operatorname{dist}(g; Y)) \geqq -c_2\, d(Y)^{nd+\varepsilon}(d(Y) + \log H(Y)).$$

Invent. math. 72, 389-405 (1983)

Inventiones mathematicae

Variétés abéliennes et indépendance algébrique II: Un analogue abélien du théorème de Lindemann-Weierstraß

P. Philippon

Centre de Mathématiques de l'Ecole Polytechnique, Plateau de Palaiseau, F-91128 Palaiseau Cedex, France

§ 0. Introduction – Énoncé des résultats

Le présent article est la suite de [5] et [6] et s'appuie notamment sur certains des résultats qui y sont démontrés. Nous nous intéressons ici à un analogue du théorème de Lindemann-Weierstraß (cf. [7]) sur les variétés abéliennes. Dans le cas des variétés abéliennes à multiplications complexes le résultat principal de ce travail fournit l'indépendance algébrique de valeurs de fonctions abéliennes en des points algébriques, voir le corollaire 0.2 et surtout le corollaire 0.3 qui répond à une conjecture bien connue. On retrouve ainsi tous les résultats de [2] que l'on étend au cas des variétés abéliennes de dimensions quelconques.

Signalons que dans [9] G. Wüstholz a annoncé qu'il avait également obtenu une démonstration du théorème que nous nous proposons d'établir.

La différence essentielle entre ce qui va suivre et [6] réside d'une part dans l'introduction d'opérateurs de dérivations et d'autre part dans le fait que l'on travaille sur un groupe algébrique produit (c'est en fait une extension triviale, par une puissance du groupe additif, de la variété abélienne considérée).

Soit F un corps de nombres que l'on considère plongé dans $\mathbb{C}$. Soit A une variété abélienne simple de dimension g définie sur F. On pose $\mathbf{k}=\mathbb{Q}\otimes_{\mathbb{Z}}\operatorname{End} A$ et $G=A^d$ pour un entier $d\geqq 1$. Soit K un corps contenant $\mathbf{k}$ de dimension d sur $\mathbf{k}$. On dispose alors d'une représentation fidèle de K de degré d sur $\mathbf{k}$, c'est-à-dire un homomorphisme ρ injectif de K dans l'algèbre $\mathcal{M}(d,\mathbf{k})$ des matrices $d\times d$ à coefficients dans $\mathbf{k}$, tel que $\rho(1)$ soit la matrice identité. Soit $\mathcal{O}$ un sous-anneau unitaire de $\operatorname{End} A$ de rang r sur $\mathbb{Z}$ et Γ un sous-$\mathcal{O}$-module de K de rang d et engendrant K sur $\mathbf{k}$. L'action naturelle de $\mathcal{M}(d,\mathcal{O})$ sur G permet d'identifier Γ à une sous-algèbre de $\operatorname{End} G$, et $\mathcal{M}(q,\Gamma)$ à une sous-algèbre de $\operatorname{End} G^q$.

On dira qu'un point p de $G(\mathbb{C})$ est fortement propre pour Γ s'il existe un élément $\mathbf{z}$ de $TG(\mathbb{C})\simeq\mathbb{C}^{gd}$ tel que l'application exponentielle $\exp_G$ du groupe de Lie $G(\mathbb{C})$ vérifie $\exp_G(\mathbf{z})=p$ et que $\mathbf{z}$ soit un vecteur propre pour l'action naturelle de $d_0\gamma$ sur $TG(\mathbb{C})$ pour tout γ dans Γ.

Laboratoire associé au C.N.R.S. n° 169

Le résultat principal s'énonce alors:

Théorème 0.1. *Soit p un point non nul de $G(\mathbb{C})$ fortement propre pour Γ appartenant à* $\exp_G((\bar{\mathbb{Q}})^{gd})$. *Si X désigne la plus petite sous variété algébrique de G définie sur $\bar{\mathbb{Q}}$ et contenant p, on a:*

$$\dim X \geqq rd/2.$$

Le corollaire le plus frappant est le suivant:

Corollaire 0.2. *Si A est une variété abélienne à multiplications complexes et si l'on choisit $\mathcal{O} = \operatorname{End} A$ alors sous les hypothèses du théorème* 0.1 *on a: $X = G$.*

Et enfin dans le cas où la variété abélienne est de dimension 1 le corollaire 0.2 entraîne:

Corollaire 0.3. *Soit $\wp$ une fonction elliptique de Weierstraß d'invariants algébriques admettant des multiplications complexes, et $\alpha_1, \dots, \alpha_n$ des nombres algébriques linéairement indépendants sur le corps des multiplications complexes de $\wp$. Alors les nombres*

$$\wp(\alpha_1), \dots, \wp(\alpha_n)$$

sont algébriquement indépendants sur $\bar{\mathbb{Q}}$.

Le corollaire 0.2 découle immédiatement du théorème 0.1. Pour le corollaire 0.3 il convient de remarquer qu'il suffit de démontrer que sous les mêmes hypothèses sur la fonction $\wp$, et si β est un nombre algébrique de degré d sur le corps des multiplications complexes de $\wp$, les nombres

$$\wp(1),\ \wp(\beta), \dots,\ \wp(\beta^{d-1})$$

sont algébriquement indépendants sur $\bar{\mathbb{Q}}$. Or ce dernier énoncé se déduit du corollaire 0.2 de la même façon que le corollaire 0.1 de [6] se déduit du théorème 3.2 de [6], on se reportera donc au paragraphe 3 de [6] pour les détails.

Cet article se présente comme suit. Dans le paragraphe 1 on rappelle, dans le contexte du théorème 0.1, quelques faits classiques sur les groupes algébriques et l'on prépare les notations pour la suite. Le paragraphe 2 est sans doute le point le plus important qui permet d'améliorer la méthode de [6]. Il s'agit de formaliser pour les opérateurs de dérivations sur le groupe algébrique considéré une idée de Baker-Coates-Anderson déjà utilisée par G.V. Choodnovsky dans [2], et dans une situation identique. Nous travaillerons avec des dérivations sur un groupe algébrique produit d'un groupe linéaire et d'une variété abélienne en séparant systématiquement les opérations sur l'un et l'autre des deux facteurs, les résultats de [3] ne suffisant pas à nos besoins nous démontrons dans le paragraphe 3 un lemme de zéro ad hoc pour la suite. Nous n'avons pas cherché à démontrer le lemme de zéros le meilleur possible ni le plus général, mais avons essayé d'être relativement brefs. Le paragraphe 4 donne la démonstration du théorème 0.1 et quelques commentaires.

Pour conclure ce paragraphe fixons une fois pour toute la notion de hauteur d'un polynôme à coefficient dans un corps de nombres F, que nous utiliserons. Si $p_1, \dots, p_v$ désignent les coefficients dans F d'un polynôme P, sa hauteur, notée $h(P)$, sera la hauteur logarithmique, telle que définie dans [8] page 19, du point $(1, p_1, \ldots, p_\nu)$ de l'espace projectif $\mathbb{P}_\nu(F)$.

Dans tout cet article nous ferons librement référence à [6].

C'est lors d'un séjour à Wuppertal en février 1982 que j'ai pu boucler la démonstration du théorème 0.1. Je tiens à ce propos à remercier G. Wüstholz qui m'a alors aimablement expliqué les détails qui forment maintenant l'important paragraphe 2.

§ 1. Préliminaires

Soit q un entier $\geqq 1$.

On choisit un plongement de G^q dans un espace projectif $\mathbb{P}_N$:

$$\Psi: G^q \to \mathbb{P}_N(\mathbb{C}).$$

Identifions l'espace tangent TG^q de G^q à l'origine avec $\mathbb{C}^{gdq}$, l'application

$$\Psi \circ \exp_{G^q}: \mathbb{C}^{gdq} \to \mathbb{P}_N(\mathbb{C})$$

est représentée par des fonctions analytiques $(\Theta_0(\mathbf{u}), \ldots, \Theta_N(\mathbf{u}))$ ne s'annulant simultanément en aucun point de $\mathbb{C}^{gdq}$ et d'ordre strict $\leqq 2$.

Nous allons décrire trois types d'opérations sur $\Psi(G^q)$ (que l'on identifie à G^q). Après le choix d'une base $\gamma_1, \ldots, \gamma_{rd}$ du $\mathbb{Z}$-module Γ on définit Γ_H pour $H \in \mathbb{N}$ comme dans [8] page 16.

α) Les lois d'additions (cf. [3] pages 492–3).

Lemme 1.1. *Il existe un recouvrement fini de $G^q \times G^q$ par des ouverts de Zariski $(U_\alpha)_{\alpha\in\mathscr{A}}$ et des familles de polynômes $((A_0^{(\alpha)}, \ldots, A_N^{(\alpha)}))_{\alpha\in\mathscr{A}}$ bihomogènes en $(X_0, \ldots, X_N;\ X'_0, \ldots, X'_N)$ tels que pour tout couple $(p, p') \in G^q \times G^q$, si $(x_0, \ldots, x_N)$* (resp. $(x'_0, \ldots, x'_N)$) *sont des coordonnées projectives de p(resp. p'), chaque famille $(A_0^{(\alpha)}(\underline{x}, \underline{x}'), \ldots, A_N^{(\alpha)}(\underline{x}, \underline{x}'))$ est soit identiquement nulle, soit forme un système de coordonnées projectives du point $p+p'$. De plus la famille d'indice α est non nulle si $(p, p') \in U_\alpha$.*

Démonstration. C'est une conséquence directe de la définition d'un groupe algébrique.

β) Les dérivations (cf. [1] page 289).

Soit $\beta \in \mathscr{B} = \{0, \ldots, N\}$ et $V_\beta = G^q \cap \{X_\beta \neq 0\}$. La famille $(V_\beta)_{\beta\in\mathscr{B}}$ est un recouvrement de G^q par des ouverts de Zariski. Sur chaque ouvert $\exp_{G^q}^{-1}(V_\beta)$ et pour tout $i = 1, \ldots, gdq$ et tout $j \neq \beta$ on a (cf. [8] page 26):

$$\frac{\partial}{\partial u_i}\left(\frac{\Theta_j}{\Theta_\beta}\right) = \Phi_{j,i}^{(\beta)}\left(\frac{\Theta_0}{\Theta_\beta}, \ldots, \frac{\Theta_N}{\Theta_\beta}\right),$$

où $\Phi_{j,i}^{(\beta)}$ est un polynôme homogène de $F[X_0, \ldots, X_N]$ de degré $\delta+1$, où δ est indépendant de i et j. On peut donc énoncer le lemme:

Lemme 1.2. *L'opérateur*

$$\Delta_{i,\beta}\colon F[X_0,\dots,X_N]\to F[X_0,\dots,X_N]$$
$$P\to\sum_{j\neq\beta}\Phi_{j,i}^{(\beta)}\cdot P'_{X_j}$$

qui transforme le polynôme homogène P de degré D en un polynôme homogène de degré $D+\delta$, vérifie:

$$\left.\frac{\Delta_{i,\beta}P\circ\Psi\circ\exp_{G^q}}{\Theta_\beta^{D+\delta}}\right|_{\exp_{G^q}^{-1}(V_\beta)}=\frac{\partial}{\partial u_i}\left.\left(\frac{P\circ\Psi\circ\exp_{G^q}}{\Theta_\beta^D}\right)\right|_{\exp_{G^q}^{-1}(V_\beta)}.$$

γ) Les endomorphismes (cf. [6] lemme 2.4).

Lemme 1.3. *Il existe un nombre réel c_1 ne dépendant que de Ψ, G^q, $\gamma_1,\dots,\gamma_{rd}$ tel que pour tout $H\in\mathbb{N}$ il existe un recouvrement fini de G^q par des ouverts de Zariski $(W_\xi)_{\xi\in\mathscr{C}_H}$ et pour tout $\gamma\in\mathscr{M}(q,\Gamma_H)$ des familles $\underline{F}_\gamma^{(\xi)}=(F_{0,\gamma}^{(\xi)},\dots,F_{N,\gamma}^{(\xi)})$, indexées par $\xi\in\mathscr{C}_H$, de polynômes homogènes de même degré $\leqq c_1H^2$ en $X_0,\dots,X_N$ à coefficients dans $\mathcal{O}_F$ de hauteurs $\leqq c_1(H^2+1)$ et tels que pour tout point p de G^q de coordonnées projectives $(X_0,\dots,X_N)$ la famille $\underline{F}_\gamma^{(\xi)}(X_0,\dots,X_N)$ ou soit nulle ou forme un système de coordonnées projectives du point $\gamma\cdot p$. De plus pour au moins un ξ la famille d'indice ξ est non nulle.*

Démonstration. C'est le lemme 2.4 de [6].

δ) Normalisation des coordonnées projectives.

Nous désignerons dorénavant par p le point fortement propre pour Γ de $\exp_G(TG(\bar{\mathbb{Q}}))$ considéré dans le théorème 0.1. Et $(p,\dots,p)$ est un point de G^q. On définit les idéaux de $F[X_0,\dots,X_N]$:

$$\mathscr{G}=(P\in F[X_0,\dots,X_N];\ P|_{G^q}\equiv 0)$$

et

$$\mathscr{E}=(P\in F[X_0,\dots,X_N];\ P(p,\dots,p)=0).$$

La dimension κ de l'idéal $\mathscr{E}$ est égale à la dimension de la sous-variété algébrique X de G du théorème 0.1. On a bien sûr $\mathscr{E}\supset\mathscr{G}$.

Lemme 1.4. *Il existe des formes linéaires $L_0,\dots,L_N$ en $X_0,\dots,X_N$ linéairement indépendantes sur F telles que $L_0,\dots,L_\kappa$ (resp. $L_0,\dots,L_{gdq}$) forment une base de transcendance sur F de l'anneau $F[X_0,\dots,X_N]/\mathscr{E}$ (resp. $F[X_0,\dots,X_N]/\mathscr{G}$) et que de plus L_i pour $i=\kappa+1,\dots,N$ (resp. L_j pour $j=gdq+1,\dots,N$) vérifie une relation de dépendance intégrale sur $F[L_0,\dots,L_{i-1}]$ modulo $\mathscr{E}$ (resp. $F[L_0,\dots,L_{j-1}]$ modulo $\mathscr{G}$).*

Démonstration. Le lemme 3.5 de [6] nous fournit des formes linéaires $L'_0,\dots,L'_{gdq},L_{gdq+1},\dots,L_N$ en $X_0,\dots,X_N$ linéairement indépendantes telles que $L'_0,\dots,L'_{gdq}$ forment une base de transcendence sur F de $F[X_0,\dots,X_N]/\mathscr{G}$ et que L_j pour $j=gdq+1,\dots,N$ vérifie une relation de dépendance intégrale sur $F[L_0,\dots,L_{j-1}]$ modulo $\mathscr{G}$. Réutilisant le lemme 3.5 de [6] on trouve des formes linéaires $L_0, \ldots, L_{gdq}$ en $L'_0, \ldots, L'_{gdq}$ donc en $X_0, \ldots, X_N$, linéairement indépendantes telles que $L_0, \ldots, L_\kappa$ forment une base de transcendance sur F de $F[L'_0, \ldots, L'_{gdq}]/\mathscr{E} \cap F[L'_0, \ldots, L'_{gdq}]$ et que L_i pour $i=\kappa+1, \ldots, gdq$ vérifie une relation de dépendance intégrale sur $F[L_0, \ldots, L_{i-1}]$ modulo $\mathscr{E}$. Et comme

$$F[L_0, \ldots, L_{gdq}] = F[L'_0, \ldots, L'_{gdq}],$$

d'une part $L_0, \ldots, L_{gdq}$ forment une base de transcendance sur F de $F[X_0, \ldots, X_N]/\mathscr{G}$, d'autre part L_i pour $i=gdq+1, \ldots, N$ vérifie une relation de dépendance intégrale sur $F[L_0, \ldots, L_{i-1}]$ modulo $\mathscr{G}$ et donc aussi modulo $\mathscr{E}$ car $\mathscr{E} \supset \mathscr{G}$. Enfin $L_0, \ldots, L_\kappa$ forment clairement une base de transcendance sur F de $F[X_0, \ldots, X_N]/\mathscr{E}$, ce qui achève la preuve du lemme 1.4.

Désormais nous choisissons une fois pour toutes $(N+1)$ formes linéaires que nous notons encore $X_0, \ldots, X_N$ et qui satisfont aux assertions du lemme 1.4. On en déduit que pour $i=\kappa+1, \ldots, N$ il existe des relations homogènes de dépendance intégrale de X_i sur $F[X_0, \ldots, X_\kappa]$ modulo $\mathscr{E}$,

$$R_i(X_0, \ldots, X_i) = X_i^{n_i} + a_1^i(X_0, \ldots, X_\kappa) X_i^{n_i-1} + \ldots + a_{n_i}^i(X_0, \ldots, X_\kappa) \in \mathscr{E}.$$

Et pour $j=gdq+1, \ldots, N$ il existe des relations de dépendance intégrale de X_j sur $F[X_0, \ldots, X_{gdq}]$ modulo $\mathscr{G}$,

$$R_j^*(X_0, \ldots, X_j) = X_j^{n_j^*} + a_1^{*j}(X_0, \ldots, X_{gdq}) X_j^{n_j^*-1} + \ldots + a_{n_j}^{*j}(X_0, \ldots, X_{gdq}) \in \mathscr{G}.$$

Soit $\mathbf{z}=(z_1, \ldots, z_{gd})$ un élément de $\bar{\mathbb{Q}}^{gd}$ propre pour l'action de $d_0\gamma$ lorsque γ parcourt Γ et tel que $p=\exp_G(\mathbf{z})$.

Nous allons considérer le sous-groupe à q-paramètres de G^q suivant:

$$\varepsilon\colon \mathbb{C}^q \xrightarrow{\mathscr{L}} \mathbb{C}^{gdq} \xrightarrow{\Psi \circ \exp_{G^q}} \mathbb{P}_N$$

où $\mathscr{L}$ est l'application linéaire de matrice $q \times gdq$, dont la transposée s'écrit:

$$\begin{pmatrix} z_1 \ldots z_{gd} & 0 \ldots & & & \\ 0 & 0 & z_1 \ldots z_{gd} & 0 \ldots & \\ \vdots & \vdots & & & \vdots \quad \vdots \\ & & & & 0 \ldots 0 \\ & & & \ldots 0 & z_1 \ldots z_{gd} \end{pmatrix}.$$

ε est représenté par les fonctions analytiques $\varphi_0, \ldots, \varphi_N$ vérifiant:

$$\varphi_i(t_1, \ldots, t_q) = \Theta_i(\mathscr{L}(t_1, \ldots, t_q)).$$

Remarquons que l'image $\varepsilon(\mathbb{C}^q)$ du sous-groupe à q-paramètres de G^q ci-dessus contenant l'orbite sous Γ du point p, qui n'est pas de torsion dans G^q, est Zariski dense dans G^q. En effet s'il n'en était pas ainsi on devrait avoir $\mu(\Gamma \cdot p, G)=0$ contredisant le lemme 2.1 de [6].

Le lemme 1.3 nous a permis de décrire l'action sur G^q des endomorphismes $\gamma \in \mathscr{M}(q, \Gamma)$, terminons ce paragraphe en expliquant comment nous ferons agir ces mêmes endomorphismes sur $\mathbb{G}_a^q$ plongé dans $\mathbb{P}_q \cdot A$ tout élément $\gamma^0 \in \Gamma$ et au vecteur propre $\mathbf{z}$ de $d_0 \gamma^0$ est associé la valeur propre $\lambda(\gamma^0) \in \bar{\mathbb{Q}}$. Pour $(Z_0, \ldots, Z_q)$ un $(q+1)$-uplet on pose:

$$\begin{aligned} \gamma \cdot \underline{Z} &= \gamma \cdot (Z_0, \ldots, Z_q) \\ &= \left(Z_0, \sum_{i=1}^{q} \lambda(\gamma^{i,1}) Z_i, \ldots, \sum_{i=1}^{q} \lambda(\gamma^{i,q}) Z_i\right) \quad \text{où } \gamma = (\gamma^{i,j})_{\substack{1 \leqq i \leqq q \\ 1 \leqq j \leqq q}}. \end{aligned}$$

Pour un q-uplet $\mathbf{t} = (t_1, \ldots, t_q)$ on notera aussi $\gamma \cdot \mathbf{t}$, le q-uplet défini par:

$$\gamma \cdot (1, t_1, \ldots, t_q) = (1, \gamma \cdot \mathbf{t}),$$

ce qui décrit l'action de γ sur $\mathbb{G}_a^q$.

Dans toute la suite on supposera sans perte de généralité que les composantes $z_1, \ldots, z_{gd}$ de $\mathbf{z}$ ainsi que les $(\lambda(\gamma^0))_{\gamma^0 \in \Gamma}$ sont dans F. Remarquons encore que si $\mathbf{1}$ est le q-uplet $(1, \ldots, 1)$, pour tout γ dans $\mathcal{M}(q, \Gamma)$ le $(N+1)$-uplet $(\varphi_0(\gamma \cdot \mathbf{1}), \ldots, \varphi_N(\gamma \cdot \mathbf{1}))$ est un système de coordonnées projectives du point $\gamma \cdot (p, \ldots, p)$, $((p, \ldots, p) \in G^q)$.

§ 2. La remarque fondamentale

Nous considérerons dans la suite le groupe algébrique $\mathbb{G}_a^q \times G^q$. A l'aide du plongement $\Psi: G^q \to \mathbb{P}_N$ on identifie $\mathbb{G}_a^q \times G^q$ à un ouvert de Zariski d'une sous-variété de $\mathbb{P}_q \times \mathbb{P}_N$.

Soit $\tilde{\mathcal{G}}$ l'idéal bihomogène de $F[Z_0, \ldots, Z_q, X_0, \ldots, X_N] = F[\underline{Z}, \underline{X}]$ engendré par l'idéal homogène $\mathcal{G}$ de $F[X_0, \ldots, X_N]$ introduit précédemment au paragraphe 1.δ. Ainsi $\tilde{\mathcal{G}}$ est un idéal premier de codimension $N - gdq$, et sa variété des zéros $\mathcal{Z}(\mathcal{G})$ dans $\mathbb{P}_q \times \mathbb{P}_N$ est $\mathbb{P}_q \times G^q$. A un polynôme $P \in F[\underline{Z}, \underline{X}]$ bihomogène en $\underline{Z}$ et $\underline{X}$ on associe la fonction analytique des q-variables $\mathbf{t} = (t_1, \ldots, t_q)$:

$$\Phi(t_1, \ldots, t_q) = P(1, t_1, \ldots, t_q, \varphi_0(\mathbf{t}), \ldots, \varphi_N(\mathbf{t})) = P(1, \mathbf{t}, \underline{\varphi}(\mathbf{t})),$$

et l'on déduit aisément du paragraphe 1.β le lemme suivant qui décrit l'action des opérateurs $\dfrac{\partial^{|\sigma|}}{\partial t_1^{\sigma_1} \ldots \partial t_q^{\sigma_q}}$ (*où* $\boldsymbol{\sigma} \in \mathbb{N}^q$ et $|\boldsymbol{\sigma}| = \sigma_1 + \ldots + \sigma_q$) sur l'algèbre des polynômes bihomogènes en $\underline{Z}$ et $\underline{X}$.

Lemme 2.1. *Pour tout multi-indice $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_q)$ de $\mathbb{N}^q$ et tout $\beta \in \mathcal{B}$ il existe un opérateur*

$$\begin{aligned} \Delta_{\boldsymbol{\sigma}, \beta}: F[Z_0, \ldots, Z_q, X_0, \ldots, X_N] &\to F[Z_0, \ldots, Z_q, X_0, \ldots, X_N] \\ P &\mapsto \Delta_{\boldsymbol{\sigma}, \beta} P \end{aligned}$$

transformant le polynôme bihomogène P de degrés L en $\underline{Z}$ et D en $\underline{X}$ en un polynôme bihomogène de degrés L en $\underline{Z}$ et $D + \delta|\boldsymbol{\sigma}|$ en $\underline{X}$, vérifiant

$$\left.\frac{\Delta_{\boldsymbol{\sigma}, \beta} P(1, \mathbf{t}, \underline{\varphi}(\mathbf{t}))}{(\varphi_\beta(\mathbf{t}))^{D + \delta|\boldsymbol{\sigma}|}}\right|_{\varepsilon^{-1}(V_\beta)} = \frac{\partial^{|\boldsymbol{\sigma}|}}{\partial t_1^{\sigma_1} \ldots \partial t_q^{\sigma_q}} \left.\left(\frac{P(1, \mathbf{t}, \underline{\varphi}(\mathbf{t}))}{(\varphi_\beta(\mathbf{t}))^D}\right)\right|_{\varepsilon^{-1}(V_\beta)}.$$

On considère l'ensemble $\{U_\alpha;\ \alpha\in\mathscr{A}/U_\alpha\cap\{\varepsilon(\mathbf{0})\}\times G^q\neq\emptyset\}=\{U_\alpha;\ \alpha\in\mathscr{A}'\}$, cette famille forme un recouvrement fini de $\{\varepsilon(\mathbf{0})\}\times G^q$. Soit $\alpha\in\mathscr{A}'$, $\beta\in\mathscr{B}$ et $\mathbf{t}$ vérifiant $\varepsilon(\mathbf{t})\in V_\beta$ et $(\varepsilon(\mathbf{0}),\ \varepsilon(\mathbf{t}))\in U_\alpha$, alors pour tout $\mathbf{w}$ suffisamment proche de $\mathbf{0}$ on a encore $\varepsilon(\mathbf{t}+\mathbf{w})\in V_\beta$ et $(\varepsilon(\mathbf{w}),\ \varepsilon(\mathbf{t}))\in U_\alpha$. Et l'on peut écrire:

$$\frac{\Phi(\mathbf{w}+\mathbf{t})}{[\varphi_\beta(\mathbf{w}+\mathbf{t})]^D}=\frac{P(1,\mathbf{w}+\mathbf{t},A_0^{(\alpha)}(\underline{\varphi}(\mathbf{w});\ \underline{\varphi}(\mathbf{t})),\dots,A_N^{(\alpha)}(\underline{\varphi}(\mathbf{w});\ \underline{\varphi}(\mathbf{t})))}{[A_\beta^{(\alpha)}(\underline{\varphi}(\mathbf{w});\ \underline{\varphi}(\mathbf{t}))]^D}.$$

Appliquant l'opérateur $\dfrac{\partial^{|\boldsymbol{\sigma}|}}{\partial\mathbf{w}^{\boldsymbol{\sigma}}}$ à cette formule (après multiplication des deux membres par $[A_\beta^{(\alpha)}(\underline{\varphi}(\mathbf{w});\ \underline{\varphi}(\mathbf{t}))]^D$) on trouve:

$$\begin{aligned}&\frac{\partial^{|\boldsymbol{\sigma}|}}{\partial\mathbf{w}^{\boldsymbol{\sigma}}}P(1,\mathbf{w}+\mathbf{t},\underline{A}^{(\alpha)}(\underline{\varphi}(\mathbf{w});\ \underline{\varphi}(\mathbf{t})))\\&\quad=\sum_{\substack{\boldsymbol{\sigma}'\leqq\boldsymbol{\sigma}\\ \boldsymbol{\sigma}'\in\mathbb{N}^q}}\binom{\boldsymbol{\sigma}}{\boldsymbol{\sigma}'}\frac{\partial^{|\boldsymbol{\sigma}'|}}{\partial\mathbf{w}^{\boldsymbol{\sigma}'}}\left(\frac{\Phi(\mathbf{w}+\mathbf{t})}{\varphi_\beta(\mathbf{w}+\mathbf{t})^D}\right)\cdot\frac{\partial^{|\boldsymbol{\sigma}|-|\boldsymbol{\sigma}'|}}{\partial\mathbf{w}^{(\boldsymbol{\sigma}-\boldsymbol{\sigma}')}}([A_\beta^{(\alpha)}(\underline{\varphi}(\mathbf{w});\ \underline{\varphi}(\mathbf{t}))]^D).\end{aligned}$$

On remarque alors que:

$$\frac{\partial^{|\boldsymbol{\sigma}|}}{\partial\mathbf{t}^{\boldsymbol{\sigma}}}\left(\frac{\Phi(\mathbf{t})}{\varphi_\beta(\mathbf{t})^D}\right)=\frac{\partial^{|\boldsymbol{\sigma}|}}{\partial\mathbf{w}^{\boldsymbol{\sigma}}}\left(\frac{\Phi(\mathbf{w}+\mathbf{t})}{\varphi_\beta(\mathbf{w}+\mathbf{t})^D}\right)\bigg|_{\mathbf{w}=\mathbf{0}}.$$

Prenant la limite lorsque $\mathbf{w}\to\mathbf{0}$ dans la formule ci-dessus on obtient;

$$Q_{\boldsymbol{\sigma},\alpha}(1,\mathbf{t},\underline{\varphi}(\mathbf{t}))=\sum_{\boldsymbol{\sigma}'\leqq\boldsymbol{\sigma}}\binom{\boldsymbol{\sigma}}{\boldsymbol{\sigma}'}\frac{\partial^{|\boldsymbol{\sigma}'|}}{\partial\mathbf{t}^{\boldsymbol{\sigma}'}}\left(\frac{\Phi(\mathbf{t})}{\varphi_\beta(\mathbf{t})^D}\right)\cdot B_{\boldsymbol{\sigma}-\boldsymbol{\sigma}',\alpha,\beta}(\underline{\varphi}(\mathbf{t}))$$

où

$$\begin{aligned}&Q_{\boldsymbol{\sigma},\alpha}(\underline{Z},\underline{X})=\\&\quad\frac{\partial^{|\boldsymbol{\sigma}|}}{\partial\mathbf{w}^{\boldsymbol{\sigma}}}P(Z_0,w_1+Z_1,\dots,w_q+Z_q,A_0^{(\alpha)}(\underline{\varphi}(\mathbf{w}),\underline{X}),\dots,A_N^{(\alpha)}(\underline{\varphi}(\mathbf{w}),\underline{X}))|_{\mathbf{w}=\mathbf{0}}\end{aligned}$$

est un polynôme bihomogène en $\underline{Z}$ et $\underline{X}$ et $B_{\boldsymbol{\sigma}-\boldsymbol{\sigma}',\alpha,\beta}(\underline{X})$ un polynôme homogène en $\underline{X}$ qui vérifie pour $|\boldsymbol{\sigma}|-|\boldsymbol{\sigma}'|\leqq D$:

$$[A_\beta^{(\alpha)}(\underline{\varphi}(\mathbf{0});\ \underline{X})]^{D+|\boldsymbol{\sigma}'|-|\boldsymbol{\sigma}|}\quad\text{divise } B_{\boldsymbol{\sigma}-\boldsymbol{\sigma}',\alpha,\beta}(\underline{X}).$$

Mais on a,

$$\frac{\partial^{|\boldsymbol{\sigma}'|}}{\partial\mathbf{t}^{\boldsymbol{\sigma}'}}\left(\frac{\Phi(\mathbf{t})}{\varphi_\beta(\mathbf{t})^D}\right)=\frac{\Delta_{\boldsymbol{\sigma}',\beta}P(1,\mathbf{t},\underline{\varphi}(\mathbf{t})}{(\varphi_\beta(\mathbf{t}))^{D+\delta|\boldsymbol{\sigma}'|}}$$

sur $\varepsilon^{-1}(V_\beta)$, et donc, en utilisant le fait que $\varepsilon(\mathbb{C}^q)$ est Zariski dense dans G^q, on a pour tout $\alpha\in\mathscr{A}'$ et $\beta\in\mathscr{B}$:

$$X_\beta^{D+\delta|\boldsymbol{\sigma}|}Q_{\boldsymbol{\sigma},\alpha}\in(\tilde{\mathscr{G}},\Delta_{\boldsymbol{\sigma}',\beta}P;\ |\boldsymbol{\sigma}'|\leqq|\boldsymbol{\sigma}|).$$

De même on voit facilement que, du fait que $B_{\mathbf{0},\alpha,\beta}(\underline{X})=[A_\beta^{(\alpha)}(\underline{\varphi}(\mathbf{0}),\underline{X})]^D$,

$$[A_\beta^{(\alpha)}(\underline{\varphi}(\mathbf{0}),\underline{X})]^{D+|\boldsymbol{\sigma}|}\Delta_{\boldsymbol{\sigma},\beta}P\in(\tilde{\mathscr{G}},Q_{\boldsymbol{\sigma}',\alpha};\ |\boldsymbol{\sigma}'|\leqq|\boldsymbol{\sigma}|).$$

On pose $A_\beta'^{(\alpha)}(\underline{X})=A_\beta^{(\alpha)}(\underline{\varphi}(\mathbf{0}),\underline{X})$.

Et des estimations aisées permettent d'énoncer le lemme:

Lemme 2.2. *Soit T un entier $\geqq 0$, L et D des entiers $\geqq 1$, et P un polynôme bihomogène de $F[Z_0, \ldots, Z_q, X_0, \ldots, X_N]$ de degrés L en $\underline{Z}$ et D en $\underline{X}$ et de hauteur $\leqq h$. Il existe des familles $(\Delta_{\boldsymbol{\sigma},\beta} P)_{\substack{|\boldsymbol{\sigma}|\leqq T\\ \beta\in\mathscr{B}}}$ et $(Q_{\boldsymbol{\sigma},\alpha})_{\substack{|\boldsymbol{\sigma}|\leqq T\\ \alpha\in\mathscr{A}}}$ de polynômes bihomogènes, et une constante c_2 ne dépendant pas de T et P, telles que:*

(i) $d_Z^0 Q_{\boldsymbol{\sigma},\alpha} \leqq L$, $d_X^0 Q_{\boldsymbol{\sigma},\alpha} \leqq c_2 D$ *et* $h(Q_{\boldsymbol{\sigma},\alpha}) \leqq h + c_2(D + T\log(T+L+D))$ *pour tout* $|\boldsymbol{\sigma}| \leqq T$ *et* $\alpha \in \mathscr{A}'$

(ii) $([A_\beta'^{(\alpha)}(\underline{X})]^{D+T} \cdot \Delta_{\boldsymbol{\sigma},\beta} P;\ |\boldsymbol{\sigma}| \leqq T,\ \beta \in \mathscr{B}) \subset (\tilde{\mathscr{G}}, Q_{\boldsymbol{\sigma},\alpha};\ |\boldsymbol{\sigma}| \leqq T)$

(iii) *pour tout* $|\boldsymbol{\sigma}| \leqq T$ *et* $\alpha \in \mathscr{A}'$ *et* $\beta \in \mathscr{B}$ *on a:*

$$X_\beta^{D+\delta T} \cdot Q_{\boldsymbol{\sigma},\alpha} \equiv \sum_{\boldsymbol{\sigma}' \leqq \boldsymbol{\sigma}} C_{\boldsymbol{\sigma}',\alpha,\beta} \cdot \Delta_{\boldsymbol{\sigma}',\beta} P (\bmod \tilde{\mathscr{G}}),$$

où $C_{\boldsymbol{\sigma}',\alpha,\beta}$ est un polynôme homogène de degré $\leqq c_2 D + \delta T$ en $\underline{X}$ et de hauteur $h(C_{\boldsymbol{\sigma}',\alpha,\beta}) \leqq c_2(D + T\log(T+L+D))$

(iv) *pour tout* $|\boldsymbol{\sigma}| \leqq T$ *et* $\beta \in \mathscr{B}$ *on a:*

$$\left.\frac{\Delta_{\boldsymbol{\sigma},\beta} P(1, \mathbf{t}, \underline{\varphi}(\mathbf{t}))}{(\varphi_\beta(\mathbf{t}))^{D+\delta|\boldsymbol{\sigma}|}}\right|_{\varepsilon^{-1}(V_\beta)} = \frac{\partial^{|\boldsymbol{\sigma}|}}{\partial \mathbf{t}^{\boldsymbol{\sigma}}} \left.\left(\frac{P(1, \mathbf{t}, \underline{\varphi}(\mathbf{t}))}{(\varphi_\beta(\mathbf{t}))^D}\right)\right|_{\varepsilon^{-1}(V_\beta)}.$$

§ 3. Le lemme de zéros

Ce paragraphe est consacré à la preuve du lemme 3.2. Je remercie Y.V. Nesterenko pour ses remarques sur une première version de ce paragraphe.

Lemme 3.1. *Il existe une constante $c_3 \geqq 1$ ne dépendant que de $\mathscr{G}$ et vérifiant ce qui suit. Soit $P \in F[\underline{Z}, \underline{X}]$ un polynôme, n'appartenant pas à $\tilde{\mathscr{G}}$, bihomogène de degrés L en $\underline{Z}$ et D en $\underline{X}$ avec $1 \leqq D \leqq L$. Soit T un entier $> c_3 \cdot gdq \cdot D^{gdq}$ et $(Q_{\boldsymbol{\sigma},\alpha})$ la famille de polynômes introduite dans le lemme 2.2. On pose;*

$$I_T = (\tilde{\mathscr{G}}, Q_{\boldsymbol{\sigma},\alpha};\ \alpha \in \mathscr{A}', |\boldsymbol{\sigma}| \leqq T)$$

alors I_T engendre dans $F(\underline{Z})\ [X_0, \ldots, X_N]$ un idéal de codimension $N+1$. (càd la variété des zéros $\mathscr{Z}(I_T)$ de I_T dans $\mathbb{P}_q \times \mathbb{P}_N$ est contenue dans $H \times \mathbb{P}_N$ où H est un sous-ensemble algébrique de codimension 1 de $\mathbb{P}_q$)

Démonstration. On va faire la démonstration par récurrence en montrant pour tout $r = 1, \ldots, gdq+1$ l'assertion $(\mathscr{H})_r$ suivante:

$(\mathscr{H})_r$: pour $T > c_3 \cdot (r-1) \cdot D^{r-1}$ il existe une décomposition primaire normale

$$I_T = \mathfrak{Q}_1 \cap \ldots \cap \mathfrak{Q}_{h'},$$

et un entier $h \in \{1, \ldots, h'\}$ tels que:

- $\forall i = 1, \ldots, h$ $\mathfrak{Q}_i \cap F[\underline{Z}] = (0)$ et $\mathfrak{P}_i = \sqrt{\mathfrak{Q}_i}$ est un idéal premier de codimension $\geqq N - gdq + r$,
- $\forall i = h+1, \ldots, h'$ $\mathfrak{Q}_i \cap F[\underline{Z}] \neq (0)$.

1°) Pour $r = 1$, comme on sait que codim $\tilde{\mathscr{G}} = N - gdq$, et $P \notin \tilde{\mathscr{G}}$ on a codim $I_T = N - gdq + 1$ pour $T > 0$ et donc l'assertion $(\mathscr{H})_1$ est immédiate.

2°) Montrons que pour $r \in \{1, \ldots, gdq\}$ l'hypothèse $(\mathscr{H})_r$ entraine $(\mathscr{H})_{r+1}$. Soit T un entier tel que I_T vérifie l'assertion $(\mathscr{H})_r$.

Soit i un indice entre 1 et h tel que l'idéal premier $\mathfrak{P}_i$ soit de codimension exactement $N-gdq+r$ (s'il n'y en a pas alors l'assertion $(\mathscr{K})_{r+1}$ est vérifiée). On continue la preuve en trois temps:

i) D'abord on montre que l'exposant e_i de l'idéal primaire $\mathfrak{Q}_i$ associé à $\mathfrak{P}_i$ est $\leqq [c_3 \cdot D^r]$. Pour cela on va localiser par rapport à $F[\underline{Z}] \setminus \{0\}$. Désignons par $\bar{I}_T, \bar{\mathfrak{Q}}_j, \bar{\mathfrak{P}}_j, \ldots$ les idéaux engendrés dans l'anneau $F(\underline{Z})[\underline{X}]$ par $I_T, \mathfrak{Q}_j, \mathfrak{P}_j, \ldots$. On démontre alors comme dans [1] (voir le paragraphe 2) que, lorsque $\mathfrak{P}_j \cap F[\underline{Z}] = (0)$ l'idéal $\bar{\mathfrak{P}}_j$ est premier dans $F(\underline{Z})[\underline{X}]$ de codimension égale à celle de $\mathfrak{P}_j$ dans $F[\underline{Z}, \underline{X}]$ et l'idéal $\bar{\mathfrak{Q}}_j$ est primaire associé à $\bar{\mathfrak{P}}_j$ de même exposant e_j que $\mathfrak{Q}_j$. Enfin

$$\bar{I}_T = \bar{\mathfrak{Q}}_1 \cap \ldots \cap \bar{\mathfrak{Q}}_h$$

est une décomposition primaire normale de $\bar{I}_T$.

A l'aide de combinaisons linéaires génériques des générateurs de I_T on construit une suite régulière dans $\bar{I}_T$ de la forme:

$$Q_1, \ldots, Q_{N-gdq}, Q_{N-gdq+1}, \ldots, Q_{N-gdq+r}$$

où $Q_1, \ldots, Q_{N-gdq}$ ne dépendent que de $\mathscr{G}$ et $Q_{N-gdq+1}, \ldots, Q_{N-gdq+r}$ sont des polynômes homogènes de $F(\underline{Z})[\underline{X}]$ de degré D en $\underline{X}$.

Comme $\bar{\mathfrak{Q}}_i$ est une composante primaire de $\bar{I}_T$ on sait que son exposant e_i est majoré par le degré de l'idéal $\bar{I}_T$. Utilisant la suite régulière ci-dessus et le lemme 3 de [1] on conclut que le degré de l'idéal $\bar{I}_T$ est majoré par $c_3 \cdot D^r$ où c_3 ne dépend que de $Q_1, \ldots, Q_{N-gdq}$, donc de $\mathscr{G}$.

Ceci fournit la majoration de l'exposant e_i de l'idéal primaire $\mathfrak{Q}_i$ cherchée.

ii) On pose $\tilde{V}_\beta = \{Z_0 \neq 0\} \times V_\beta$ pour $\beta \in \mathscr{B}$.

Soit $\beta_0 \in \mathscr{B}$ tel que $\tilde{V}_{\beta_0} \cap \mathscr{Z}(\mathfrak{P}_i) \neq \emptyset$. Il y a nécessairement un tel β_0 car $(\tilde{V}_\beta)_{\beta \in \mathscr{B}}$ est un recouvrement de $\{Z_0 \neq 0\} \times \mathbb{P}_N$ et $\mathscr{Z}(\mathfrak{P}_i) \not\subset \{Z_0 = 0\}$. On montre alors que $\Delta_{\boldsymbol{\sigma}, \beta_0} I_T \not\subseteq \mathfrak{P}_i$ pour un multi-indice $\boldsymbol{\sigma}$ tel que $|\boldsymbol{\sigma}| = e_i$.

Pour cela, suivant [1] et [4], on définit pour $Q \in F[\underline{Z}, \underline{X}]$, et $p' \in \tilde{V}_{\beta_0} \cap \mathscr{Z}(\mathfrak{P}_i)$,

$$ord_{p'} Q = \max\{l / \forall\, |\boldsymbol{\sigma}| < l, \Delta_{\boldsymbol{\sigma}, \beta_0} Q(p') = 0\}.$$

Et l'on pose

$$\operatorname{ord}_{p'} \mathfrak{P}_i = \min\{\operatorname{ord}_{p'} Q;\, Q \in \mathfrak{P}_i\} > 0.$$

Si P_i est un polynôme de $\mathfrak{P}_i$ réalisant ce minimum il existe un multi-indice $\boldsymbol{\sigma}_0$ de la forme $\boldsymbol{\sigma}_0 = (0, \ldots, 0, 1, 0, \ldots, 0)$ tel que $\operatorname{ord}_{p'} \Delta_{\boldsymbol{\sigma}_0, \beta_0} P_i < \operatorname{ord}_{p'} P_i$ ce qui entraine que $\Delta_{\boldsymbol{\sigma}_0, \beta_0} P_i \notin \mathfrak{P}_i$.

Choisissons pour tout $j \neq i$ un élément $P_j \in \mathfrak{Q}_j \setminus \mathfrak{P}_i$. On peut même, pour $j \geqq h+1$, prendre $P_j \in \mathfrak{Q}_j \cap F[\underline{Z}]$. Alors si $\boldsymbol{\sigma} = e_i \cdot \boldsymbol{\sigma}_0$ on a:

$$\Delta_{\boldsymbol{\sigma}, \beta_0}(P_i^{e_i} \prod_{j \neq i} P_j) \equiv e_i! (\Delta_{\boldsymbol{\sigma}_0, \beta_0} P_i)^{e_i} \cdot \prod_{j \neq i} P_j \not\equiv 0 \pmod{\mathfrak{P}_i},$$

et comme $P_i^{e_i} \prod_{j \neq i} P_j \in I_T$ on a bien montré que $\Delta_{\boldsymbol{\sigma}, \beta_0} I_T \not\subseteq \mathfrak{P}_i$ pour un multi-indice $\boldsymbol{\sigma}$ vérifiant $|\boldsymbol{\sigma}| = e_i$.

iii) D'après le lemme 2.2 on a pour $\beta \in \mathscr{B}$, $\alpha \in \mathscr{A}'$ et $|\boldsymbol{\sigma}'| \leqq T$;

$$X_\beta^{D+\delta T} \cdot Q_{\boldsymbol{\sigma}', \alpha} \in (\tilde{\mathscr{G}}, \Delta_{\boldsymbol{\sigma}'', \beta} P;\, |\boldsymbol{\sigma}''| \leqq T)$$

d'où pour le $\boldsymbol{\sigma}_0$ du (ii) et pour tout $\beta\in\mathscr{B}$,

$$\Delta_{\boldsymbol{\sigma}_0,\beta_0}(X_\beta^{D+\delta T}\cdot Q_{\boldsymbol{\sigma}',\alpha})=\Delta_{\boldsymbol{\sigma}_0,\beta_0}(X_\beta^{D+\delta T})\cdot Q_{\boldsymbol{\sigma}',\alpha}+X_\beta^{D+\delta T}\cdot\Delta_{\boldsymbol{\sigma}_0,\beta_0}Q_{\boldsymbol{\sigma}',\alpha},$$

avec

$$X_\beta^\delta\cdot\Delta_{\boldsymbol{\sigma}_0,\beta_0}(X_\beta^{D+\delta T}\cdot Q_{\boldsymbol{\sigma}',\alpha})\in(\tilde{\mathscr{G}},\Delta_{\boldsymbol{\sigma}'',\beta}P;|\boldsymbol{\sigma}''|\leqq T+1).$$

Et enfin pour $\boldsymbol{\sigma}=e_i\boldsymbol{\sigma}_0$;

$$X_\beta^{D+\delta(T+e_i)}\cdot\Delta_{\boldsymbol{\sigma},\beta_0}Q_{\boldsymbol{\sigma}',\alpha}\in(\tilde{\mathscr{G}},\Delta_{\boldsymbol{\sigma}'',\beta}P;|\boldsymbol{\sigma}''|\leqq T+e_i).$$

Soit $T'\geqq T+[c_3D^r]\geqq T+e_i$, on déduit toujours du lemme 2.2 que pour tout $\beta\in\mathscr{B}$ et $\alpha'\in\mathscr{A}'$

$$[A_\beta'^{(\alpha')}(\underline{X})\cdot X_\beta]^{D+\delta T'}\cdot\Delta_{\boldsymbol{\sigma},\beta_0}Q_{\boldsymbol{\sigma}',\alpha}\in I_{T'}.$$

Comme lorsque α' parcourt $\mathscr{A}'$ et β parcourt $\mathscr{B}$ les polynômes $X_\beta\cdot A_\beta'^{(\alpha')}(\underline{X})$ n'ont pas de zéro commun on obtient que, pour $\alpha\in\mathscr{A}'$ et $|\boldsymbol{\sigma}'|\leqq T$,

$$\Delta_{\boldsymbol{\sigma},\beta_0}Q_{\boldsymbol{\sigma}',\alpha}\in I_{T'}^*,$$

où le symbole * signifie que l'on omet les composantes primaires de radical contenant $(X_0,\ldots,X_N)$.

D'où il découle que:

$$\Delta_{\boldsymbol{\sigma},\beta_0}I_T\subset I_{T'}^*.$$

On en déduit que pour $T'\geqq T+[c_3D^r]$ on a $I_{T'}^*\nsubseteq\mathfrak{P}_i$ pour tous les $i\in\{1,\ldots,h\}$ tels que la composante $\mathfrak{P}_i$ de I_T soit de codimension $N-gdq+r$. Et par suite, vu que $I_{T'}\supseteq I_T$, les seules composantes $\mathfrak{P}_i'$ de $I_{T'}$ de codimension $\leqq N-gdq+r$ sont parmi celles qui proviennent des $\mathfrak{P}_i$ avec $i=h+1,\ldots,h'$ et vérifient donc $\mathfrak{P}_i'\cap F[\underline{Z}]\neq(0)$. Mais pour tout $T'>c_3\cdot r\cdot D^r$ on peut trouver $T>c_3(r-1)D^{r-1}$ tel que $T'\geqq T+[c_3D^r]$, l'assertion $(\mathscr{H})_{r+1}$ est donc vérifiée.

Pour achever la preuve du lemme 3.1 il suffit de remarquer que les idéaux $\mathfrak{Q}$ primaires bihomogènes de $F[\underline{Z},\underline{X}]$ de codimension $N+1$ vérifient nécessairement $\mathfrak{Q}\cap F[\underline{Z}]\neq(0)$ ou $\sqrt{\mathfrak{Q}}=(X_0,\ldots,X_N)$, le lemme découle de l'assertion $(\mathscr{H})_{gdq+1}$.

Le lemme de zéro que nous avons en vue s'énonce:

Lemma 3.2. *Il existe une constante $c_4>0$ ne dépendant que de G^q et des relations (R_j^*) choisies au paragraphe 1 telle que la proposition suivante soit vraie.*

Soit P un élément de $F[\underline{Z},\underline{X}]$ bihomogène de degrés L en $\underline{Z}$ et D en $\underline{X}$, n'appartenant pas à $\tilde{\mathscr{G}}$. Soit T et H des entiers vérifiant:

$$(T-c_3gdD^{gd})\cdot H^{rd}>c_4LD^{gd}.$$

Alors l'idéal

$$J=(\tilde{\mathscr{G}},Q_{\boldsymbol{\sigma},\alpha}(\gamma\cdot\underline{Z},\underline{F}_\gamma^{(\xi)}(\underline{X}));\ \alpha\in\mathscr{A}',\ \xi\in\mathscr{C}_H,\ |\boldsymbol{\sigma}|\leqq T \text{ et } \gamma\in\mathscr{M}(q,\Gamma_H))$$

vérifie:

$$\mathscr{Z}(J)\subset\mathscr{Z}((Z_0))\cup\mathscr{Z}((Z_1,\ldots,Z_q)).$$

Démonstration. On procède par récurrence sur q.

(i) *le cas* $q=1$: Le lemme 3.1 avec $q=1$ nous assure que dans l'anneau $F(\underline{Z})[X_0,\ldots,X_N]$ l'idéal engendré par $I_{T'}$ pour $T'>c_3 gdD^{gd}$ est de codimension $N+1$. Les relations $R^*_{gd+1},\ldots,R^*_N$ introduites à la fin du paragraphe 1 étant dans $\mathcal{G}$ sont dans $I_{T'}$. Notant $P_1,\ldots,P_m$ des générateurs de $\mathcal{G}$ et les $(Q_{\sigma,\alpha})$ on peut alors appliquer le lemme 1.2 de [6] qui fournit un polynôme Δ de $F[\underline{Z}]$, non nul, homogène de degré $\leqq c'_4 L\cdot D^{gd}$ où c'_4 est une constante ne dépendant que des (R^*_j).

Comme $X_\beta^{c_2(N+1)D}\cdot\Delta\in I_{T'}$ on déduit du lemme 2.2 que:

$$X_\beta^{c'_2 D+\delta T'}\cdot\Delta\in(\tilde{\mathcal{G}},\Delta_{\sigma',\beta}P;\,|\sigma'|\leqq T')$$

où $c'_2=1+c_2(N+1)$. D'où encore,

$$X_\beta^{c'_2 D+\delta T'+|\sigma|}\cdot\Delta_{\sigma,\beta}(\Delta)\in(\tilde{\mathcal{G}},\Delta_{\sigma',\beta}P;\,|\sigma'|\leqq T'+|\sigma|)$$

et enfin,

$$[A'^{(\alpha')}_\beta(\underline{X})\cdot X_\beta]^{c'_2 D+\delta(T'+|\sigma|)}\cdot\Delta_{\sigma,\beta}(\Delta)\in(\tilde{\mathcal{G}},Q_{\sigma',\alpha};\,\alpha\in\mathcal{A}' \text{ et } |\sigma'|\leqq T'+|\sigma|)$$

pour tout $\alpha'\in\mathcal{A}'$ et tout $\beta\in\mathcal{B}$.

On remarque alors que $\Delta_{\sigma,\beta}(\Delta)=Z_0^{\sigma_1}\cdot\dfrac{\partial^{\sigma_1}}{\partial Z_1^{\sigma_1}}\Delta$ est indépendant de β. On peut donc comme à la fin de la preuve du lemme 3.1 déduire que:

$$\Delta_{\sigma,\beta}(\Delta)\in(\tilde{\mathcal{G}},Q_{\sigma',\alpha};\,\alpha\in\mathcal{A}' \text{ et } |\sigma'|\leqq T'+|\sigma|)^*,$$

d'où:

$$\forall|\sigma|\leqq T-T',\quad \forall\gamma\in\Gamma_H\quad \Delta_{\sigma,\beta}(\Delta)(\gamma\cdot\underline{Z})\in J^*.$$

Si $(z_0,z_1,x_0,\ldots,x_N)$ est un système de coordonnées projectives d'un zéro de J dans $\mathbb{P}_1\times\mathbb{P}_N$ avec $z_0 z_1\neq 0$ on doit avoir:

$$\forall\sigma_1\leqq T-T',\quad \forall\gamma\in\Gamma_H\quad \frac{\partial^{\sigma_1}}{\partial Z_1^{\sigma_1}}\Delta(\gamma\cdot\underline{z})=0$$

ce qui est clairement impossible pour $(T-T')H^{rd}>c'_4 LD^{gd}$ car le degré de Δ est borné par $c'_4 LD^{gd}$.

(ii) *la récurrence* $q-1\to q$: On suppose que contrairement à la conclusion du lemme 3.2 il existe un zéro $(z_0,\ldots,z_q)\times(p_1,\ldots,p_q)$ de J dans $\mathbb{P}_q\times G^q$ tel que $z_0 z_i\neq 0$ pour un certain i entre 1 et q. Quitte à remplacer ce zéro par son image par un élément de $\mathcal{M}(q,\Gamma_1)$ convenable et à remplacer H par $[H/q]$ on peut faire l'hypothèse que $z_0\ldots z_q\neq 0$. On peut également supposer ce point rationnel sur une extension finie de F, que l'on notera encore F. Soit $z\times p$ un point générique de $\mathbb{G}_a\times G$ et J' l'idéal de $F[\underline{Z},\underline{X}]$ suivant,

$$J'=(\mathcal{G},Z_q-zZ_0,Q_{\sigma,\alpha}(\gamma\cdot\underline{Z},\underline{F}_\gamma^{(\xi)}(\underline{X}));\quad \alpha\in\mathcal{A}',\ \xi\in\mathcal{C}_H,\ \sigma\in\Sigma_{q,T} \text{ et } \gamma\in\mathcal{M}(\mathrm{q}-1,\Gamma_H))$$

où $\Sigma_{q,T}=\{\sigma=(\sigma_1,\ldots,\sigma_{q-1},0);\,|\sigma|<(1-1/q)T\}$ et où un élément $\gamma\in\mathcal{M}(q-1,\Gamma_H)$ est identifié à élément $\begin{pmatrix}\gamma & 0\\ 0 & 1\end{pmatrix}$ de $\mathcal{M}(q,\Gamma_H)$.

Comme $\mathscr{G}=(Q\in F[\underline{X}];\, Q|_{G^{q-1}\times\{p\}}\equiv 0)$ le lemme admis à l'ordre $q-1$ entraine que l'idéal J' n'a pas de zéro de la forme $(z_0,\dots,z_{q-1},zz_0)\times(p_1,\dots,p_{q-1},p)$. (notons que l'on utilise en fait le lemme 3.2 à l'ordre $q-1$ avec F remplacé par $F(z)$, et que ceci est bien licite). Il existe donc un polynôme $P^{(1)}=Q_{\boldsymbol{\sigma}_0,\alpha_0}$ n'appartenant pas à l'idéal

$$\mathscr{G}_1=(Q\in F[\underline{X}];\, Q|_{\gamma_0((p_1,\dots,p_{q-1})\times G)}\equiv 0).$$

Ecrivons $\gamma_0\cdot(z_0,\dots,z_q)=(z_0,z_1',\dots,z_{q-1}',z_q)$, un des z_i' est non nul. Soient T' un entier $>c_3 gdD^{gd}$ et $Q^{(1)}_{\boldsymbol{\sigma},\alpha}$ les polynômes donnés par le lemme 2.2 à partir de $P^{(1)}$. Utilisant le lemme 3.1 pour $q=1$ on déduit que l'idéal

$$I_1=(\mathscr{G}_1,Z_0z_i'-Z_iz_0,Q^{(1)}_{\boldsymbol{\sigma},\alpha};\qquad i=1,\dots,q-1,\alpha\in\mathscr{A}',\boldsymbol{\sigma}=(0,\dots,0,\sigma_q)\text{ avec }\sigma_q\leqq T'/q)$$

engendre dans $F(\underline{Z})[\underline{X}]$ un idéal de codimension $N+1$.

Réduisant les polynômes $Q^{(1)}_{\boldsymbol{\sigma},\alpha}$ modulo les relations $Z_0z_i'-Z_iz_0$ ($i=1,\dots,q-1$) on obtient des éléments $\tilde{Q}^{(1)}_{\boldsymbol{\sigma},\alpha}$ de $F[Z_0,Z_q,X_0,\dots,X_N]$ bihomogènes de degré L en (Z_0,Z_q) et $\leqq c_2'D$ en $\underline{X}$. Posons:

$$J_1=(\mathscr{G}_1,\tilde{Q}^{(1)}_{\boldsymbol{\sigma},\alpha}(\gamma\cdot\underline{Z},\underline{F}^{(\xi)}_{\gamma}(\underline{X}));\qquad \alpha\in\mathscr{A}',\ \xi\in\mathscr{C}_H,\ \boldsymbol{\sigma}=(0,\dots,0,\sigma_q)\text{ et }\sigma_q<T/q,\text{ et }\gamma\in\Gamma_H)$$

où l'on a identifié $\gamma\in\Gamma_H$ et l'élément $\begin{pmatrix}Id & 0\\ 0 & \gamma\end{pmatrix}$ *de* $\mathscr{M}(q,\Gamma_H)$.

On reprend alors la construction du pas $q=1$ qui fournit un polynôme Δ_1 de $F[Z_0,Z_q]$ homogène de degré $\leqq c_4'LD^{gd}$ et tel que pour tout $\sigma_q\leqq(T-T')/q$ et tout $\gamma\in\Gamma_H$ on ait,

$$Z_0^{\sigma_q}\cdot\frac{\partial^{\sigma_q}}{\partial Z_q^{\sigma_q}}\Delta_1(Z_0,\gamma Z_q)\in J_1^*.$$

On vérifie alors par des arguments similaires à ceux utilisés dans les démonstrations des lemmes 2.2 et 3.1 que pour tout les $\alpha,\xi,\boldsymbol{\sigma}$ et γ intervenant dans la définition de l'idéal J_1 et tout $\xi'\in\mathscr{C}_H$ on a:

$$Q^{(1)}_{\boldsymbol{\sigma},\alpha}(\gamma\gamma_0\cdot\underline{Z},\underline{F}^{(\xi)}_{\gamma}(\underline{F}^{(\xi')}_{\gamma_0}(\underline{X})))\in J^*.$$

Mais $(z_0,\dots,z_q)\times(p_1,\dots,p_q)$ étant un zéro de J (donc de J^*) on en déduit que $(z_0,z_q)\times\gamma_0(p_1,\dots,p_q)$ est un zéro de J_1 et donc, comme $z_0\neq 0$, que,

$$\forall\gamma\in\Gamma_H,\qquad\forall\sigma_q<(T-T')/q\quad\frac{\partial^{\sigma_q}}{\partial Z_q^{\sigma_q}}\Delta_1(z_0,\gamma z_q)=0.$$

Ceci contredit la majoration du degré de Δ_1, car $(T-T')H^{rd}>c_4LD^{gd}$ et $z_q\neq 0$, et établit le lemme 3.2 au pas q.

§ 4. Preuve du Théorème 0.1

Soit ε un nombre réel >0 arbitrairement petit et q un entier $\geqq 2(gd+1/\varepsilon)$. Soit L un paramètre entier arbitrairement grand, on pose $D=[(\log L)^{\varepsilon}]+1$. Dans ce qui suit $c_5, c_6, \ldots$ sont des constantes réelles strictement positives ne dépendant pas de L.

Lemme 4.1. *Il existe un polynôme $P\in\mathbb{Z}[\underline{Z},\underline{X}]$ bihomogène de degrés L en $\underline{Z}$ et D en $\underline{X}$ et de hauteur $\leqq c_5\, L$ tel que:*

(i) $P\notin\tilde{\mathscr{G}}$

(ii) *la fonction* $\Phi(t_1,\ldots,t_q)=P(1,t_1,\ldots,t_q,\varphi_0(\mathbf{t}),\ldots,\varphi_N(\mathbf{t}))$ *vérifie pour tout réel* $R_0\leqq \operatorname{Log} L$ *et tout* $|\boldsymbol{\sigma}|<L/\log L$,

$$\sup_{\mathbf{t}\in B(\mathbf{0},R_0)}\left\{\left|\frac{\partial^{|\boldsymbol{\sigma}|}}{\partial t_1^{\sigma_1}\ldots\partial t_q^{\sigma_q}}\Phi(\mathbf{t})\right|\right\}\leqq \exp(-c_6 L(\log L)^{1+(\varepsilon/2)}).$$

Démonstration. Soit Ξ une base des monômes de $F[\underline{Z},\underline{X}]$ bihomogènes de degrés L en $\underline{Z}$ et D en $\underline{X}$ linéairement indépendants sur F modulo $\tilde{\mathscr{G}}$. On sait que $\mathscr{Z}(\tilde{\mathscr{G}})$ étant $\mathbb{P}_q\times G^q\subset\mathbb{P}_q\times\mathbb{P}_N$, le cardinal de Ξ est $\geqq c_7 L^q D^{gdq}$. Ecrivons

$$\Phi(\mathbf{t})=\sum_{\underline{Z}^{\underline{\alpha}}\underline{X}^{\underline{\beta}}\in\Xi} p_{\underline{\alpha},\underline{\beta}}\, t_1^{\alpha_1}\ldots t_q^{\alpha_q}\cdot\varphi_0^{\beta_0}(\mathbf{t})\ldots\varphi_N^{\beta_N}(\mathbf{t}),$$

la non nullité de la famille $(p_{\underline{\alpha},\underline{\beta}})$ entrainera bien que $P\notin\tilde{\mathscr{G}}$.

En développant les fonctions $\varphi_0,\ldots,\varphi_N$ en séries entières au voisinage de $\mathbf{0}$, on obtient:

$$\Phi(\mathbf{t})=\sum_{\mathbf{v}\in\mathbb{N}^q}\Big(\sum_{\underline{Z}^{\underline{\alpha}}\underline{X}^{\underline{\beta}}\in\Xi} a_{\underline{\alpha},\underline{\beta},\mathbf{v}}\cdot p_{\underline{\alpha},\underline{\beta}}\Big)t_1^{v_1}\ldots t_q^{v_q},$$

où les $a_{\underline{\alpha},\underline{\beta},\mathbf{v}}$ sont des éléments du corps de nombres F dont un dénominateur commun et les tailles sont majorés par $c_8(|\underline{\alpha}|+|\underline{\beta}|+|\mathbf{v}|\log|\mathbf{v}|)$.

Soit $T=[L\cdot D^{gd-(1/2)}]$, les équations $\Big(\sum_{\underline{Z}^{\underline{\alpha}}\underline{X}^{\underline{\beta}}\in\Xi} a_{\underline{\alpha},\underline{\beta},\mathbf{v}}\cdot p_{\underline{\alpha},\underline{\beta}}=0\Big)_{|\mathbf{v}|<T}$ donnent au plus $c_9 T^q$ équations linéaires en les $(p_{\underline{\alpha},\underline{\beta}})$ à coefficients dans $\mathbb{Z}$ et de tailles majorées par $c_{10} T\log T$. Le principe des tiroirs (de Dirichlet) nous perment donc de trouver une famille $(p_{\underline{\alpha},\underline{\beta}})$ d'éléments, non tous nuls, de $\mathbb{Z}$ vérifiant d'une part

$$\sum_{\underline{Z}^{\underline{\alpha}}\underline{X}^{\underline{\beta}}\in\Xi} a_{\underline{\alpha},\underline{\beta},\mathbf{v}}\cdot p_{\underline{\alpha},\underline{\beta}}=0 \qquad \text{pour tout } |\mathbf{v}|<T,$$

et d'autre part

$$\max_{\underline{Z}^{\underline{\alpha}}\underline{X}^{\underline{\beta}}\in\Xi}\{|p_{\underline{\alpha},\underline{\beta}}|\}\leqq \exp\left(c_{11} LD^{gd-(1/2)}\log L\times\frac{L^q D^{gdq-(q/2)}}{L^q D^{gdq}}\right),$$

mais on a $D^{q/2-gd+(1/2)}\geqq(\log L)^{\varepsilon\left(\frac{q+1}{2}-gd\right)}>\log L(\text{car } q\geqq 2(gd+(1/\varepsilon))$ et donc la majoration ci-dessus se réécrit:

$$\max_{\underline{Z}^{\underline{\alpha}}\underline{X}^{\underline{\beta}}\in\Xi}\{|p_{\underline{\alpha},\underline{\beta}}|\}\leqq \exp(c_5 L).$$

De plus le polynôme $P=\sum_{\underline{Z}^{\alpha}\underline{X}^{\beta}\in\Xi} p_{\alpha,\beta}\underline{Z}^{\alpha}\underline{X}^{\beta}$, ainsi déterminé, vérifie bien (i) et la fonction Φ associée à P s'annule à l'ordre T à l'origine. Un "lemme de Schwarz" très classique (par exemple le théorème 7.1.5 de [8] avec $r=R_0\leqq\log L$ et $R=L^{1/4}$) nous permet de conclure, étant donné que les fonctions $\varphi_0,\ldots,\varphi_N$ sont d'ordre strict $\leqq 2$, à la validité de (ii) pour la fonction Φ. Ceci achève donc la preuve du lemme 4.1.

Démonstration du théorème 0.1: On sait d'après le théorème de Schneider-Lang (cf. [8] page 59) que dans la situation du théorème 0.1 on a toujours $\dim X=\kappa\geqq 1$, on peut donc sans perte de généralité supposer pour la démonstration que $rd>2$.

La démonstration se fait par un raisonnement par l'absurde en supposant au départ que $1\leqq\kappa<rd/2$, on choisit alors un nombre réel $\varepsilon>0$ tel que

$$\varepsilon<\min\left\{\frac{1}{2g},\frac{rd-2\kappa}{d\kappa(2g+r)}\right\}.$$

q et D sont alors choisis en fonction de L et ε comme précédemment, et on considère le polynôme P fourni par le lemme 4.1.

Nous décomposons la suite de l'argument en trois pas.

i) Les estimations analytiques

On sait d'après le lemme 2.2 (iv) que pour β et $\mathbf{t}$ tels que $\varepsilon(\mathbf{t})\in V_\beta$ on a:

$$\begin{aligned}\Delta_{\boldsymbol{\sigma},\beta}P(1,\mathbf{t},\underline{\varphi}(\mathbf{t}))&=(\varphi_\beta(\mathbf{t}))^{D+\delta|\boldsymbol{\sigma}|}\cdot\frac{\partial^{|\boldsymbol{\sigma}|}}{\partial\mathbf{t}^{\boldsymbol{\sigma}}}\left(\frac{\Phi(\mathbf{t})}{(\varphi_\beta(\mathbf{t}))^D}\right)\\&=\sum_{\boldsymbol{\sigma}'\leqq\boldsymbol{\sigma}}\binom{\boldsymbol{\sigma}}{\boldsymbol{\sigma}'}(\varphi_\beta(\mathbf{t}))^{D+\delta|\boldsymbol{\sigma}|}\cdot\frac{\partial^{|\boldsymbol{\sigma}-\boldsymbol{\sigma}'|}}{\partial\mathbf{t}^{\boldsymbol{\sigma}-\boldsymbol{\sigma}'}}\left(\frac{1}{(\varphi_\beta(\mathbf{t}))^D}\right)\frac{\partial^{|\boldsymbol{\sigma}'|}}{\partial\mathbf{t}^{\boldsymbol{\sigma}'}}(\Phi(\mathbf{t})).\end{aligned}$$

Posons

$$M_\beta(\boldsymbol{\sigma},\mathbf{t})=\max_{\boldsymbol{\sigma}'\leqq\boldsymbol{\sigma}}\left\{(1+|\varphi_\beta(\mathbf{t})|)^{D+\delta|\boldsymbol{\sigma}|}\cdot\left|\frac{\partial^{|\boldsymbol{\sigma}-\boldsymbol{\sigma}'|}}{\partial\mathbf{t}^{\boldsymbol{\sigma}-\boldsymbol{\sigma}'}}\left(\frac{1}{(\varphi_\beta(\mathbf{t}))^D}\right)\right|\right\}.$$

En faisant les hypothèses que $|\boldsymbol{\sigma}|\leqq L/\log L$ et $R_0\leqq\log L$ on déduit du lemme 4.1 (ii) et des égalités ci-dessus que pour tout $\beta\in\mathscr{B}$ et tout $\mathbf{t}\in B(\mathbf{0},R_0)$

$$|\Delta_{\boldsymbol{\sigma},\beta}P(1,\mathbf{t},\underline{\varphi}(\mathbf{t}))|\leqq M_\beta(\boldsymbol{\sigma},\mathbf{t})\exp(-c_{12}L(\log L)^{1+(\varepsilon/2)}).$$

et combinant avec le lemme 2.2 (iii) on trouve que pour tout $\alpha\in\mathscr{A}'$, tout $\beta\in\mathscr{B}$ et tout $\mathbf{t}\in B(\mathbf{0},R_0)$ on a

$$\begin{aligned}&|\varphi_\beta(\mathbf{t})|^{D+\delta|\boldsymbol{\sigma}|}\cdot|Q_{\boldsymbol{\sigma},\alpha}(1,\mathbf{t},\underline{\varphi}(\mathbf{t}))|\\&\qquad\leqq\Big(\sum_{\boldsymbol{\sigma}'\leqq\boldsymbol{\sigma}}|C_{\boldsymbol{\sigma}',\alpha,\beta}(1,\mathbf{t},\underline{\varphi}(\mathbf{t}))|\Big)M_\beta(\boldsymbol{\sigma},\mathbf{t})\times\exp(-c_{12}L(\log L)^{1+(\varepsilon/2)}).\end{aligned}$$

Mais un argument classique utilisant les équations fonctionnelles des fonctions thêta montre qu'il existe pour tout $\mathbf{t}\in B(\mathbf{0},R_0)$ un $\beta\in\mathscr{B}$ tel que $|\varphi_\beta(\mathbf{t})|\geqq\exp(-c_{13}R_0^2)$ où c_{13} ne dépend pas de β, $\mathbf{t}$ et R_0.

Une conséquence immédiate est que pour tout $\mathbf{t}\in B(\mathbf{0}, R_0)$ il existe un $\beta\in\mathscr{B}$ tel que

$$M_\beta(\boldsymbol{\sigma}, \mathbf{t}) \leqq \exp(c_{14}(D+|\boldsymbol{\sigma}|)(R_0^2+\log|\boldsymbol{\sigma}|))$$

Le lemme 2.2 (iii) montre que pour ce même β et pour tout $\alpha\in\mathscr{A}'$ et tout $\boldsymbol{\sigma}'\leqq\boldsymbol{\sigma}$ on a

$$|C_{\boldsymbol{\sigma}',\alpha,\beta}(1, \mathbf{t}, \underline{\varphi}(\mathbf{t}))| \leqq \exp(c_{14}|\boldsymbol{\sigma}|(R_0^2+\log(|\boldsymbol{\sigma}|+L))).$$

D'où enfin si $|\boldsymbol{\sigma}|<L/\log L$ et $R_0\leqq\log L$

$$\max_{\alpha\in\mathscr{A}'} \sup_{\mathbf{t}\in B(\mathbf{0}, R_0)} \{|Q_{\boldsymbol{\sigma},\alpha}(1, \mathbf{t}, \underline{\varphi}(\mathbf{t}))|\} \leqq \exp(-c_{15}L(\log L)^{1+(\varepsilon/2)}).$$

Soit $\mathbf{1}=(1, \ldots, 1)\in\bar{\mathbb{Q}}^q$. On a remarqué à la fin du paragraphe 1 que $(\varphi_0(\gamma\cdot\mathbf{1}), \ldots, \varphi_N(\gamma\cdot\mathbf{1}))$ est un système de coordonnées projective du point $\gamma\cdot(p, \ldots, p)\in G^q$, et comme $(F_{0,\gamma}^{(\xi)}(\underline{\varphi}(\mathbf{1})), \ldots, F_{N,\gamma}^{(\xi)}(\underline{\varphi}(\mathbf{1})))$ est un autre système de coordonnées projectives de ce même point on peut écrire si $\varphi_\beta(\gamma\cdot\mathbf{1})\neq 0$:

$$Q_{\boldsymbol{\sigma},\alpha}(1, \gamma\cdot\mathbf{1}, \underline{F}_\gamma^{(\xi)}(\underline{\varphi}(\mathbf{1}))) = \left(\frac{F_{\beta,\gamma}^{(\xi)}(\underline{\varphi}(\mathbf{1}))}{\varphi_\beta(\gamma\cdot\mathbf{1})}\right)^{d_X^0 Q_{\boldsymbol{\sigma}\alpha}} \cdot Q_{\boldsymbol{\sigma},\alpha}(1, \gamma\cdot\mathbf{1}, \underline{\varphi}(\gamma\cdot\mathbf{1})).$$

On note alors que pour $\gamma\in\mathscr{M}(q, \Gamma_H)$ avec $H\leqq c_{16}R_0$ on a $\gamma\cdot\mathbf{1}\in B(\mathbf{0}, R_0)$ et donc il existe un $\beta\in\mathscr{B}$ tel que $|\varphi_\beta(\gamma\cdot\mathbf{1})|\geqq\exp(-c_{13}R_0^2)$ d'où l'on déduit que pour tout $\alpha\in\mathscr{A}'$, tout $\gamma\in\mathscr{M}(q, \Gamma_H)$ et tout $\xi\in\mathscr{C}_H$;

$$|Q_{\boldsymbol{\sigma},\alpha}(1, \gamma\cdot\mathbf{1}, \underline{F}_\gamma^{(\xi)}(\underline{\varphi}(\mathbf{1})))| \leqq \sup_{\mathbf{t}\in B(\mathbf{0}, R_0)} \{|Q_{\boldsymbol{\sigma},\alpha}(1, \mathbf{t}, \underline{\varphi}(\mathbf{t}))|\} \times \exp(c_{17}D(H^2+R_0^2)).$$

On peut donc écrire si $|\boldsymbol{\sigma}|<L/\log L$, et $H\leqq c_{16}R_0\leqq c_{16}\log L$,

$$|Q_{\boldsymbol{\sigma},\alpha}(1, \gamma\cdot\mathbf{1}, \underline{F}_\gamma^{(\xi)}(\underline{\varphi}(\mathbf{1})))| \leqq \exp(-c_{18}L(\log L)^{1+(\varepsilon/2)})$$

pour tout $\alpha\in\mathscr{A}'$, $\xi\in\mathscr{C}_H$ et $\gamma\in\mathscr{M}(q, \Gamma_H)$.

ii) L'élimination

Posons $T=[L/\log L]$, $H=c_{19}(\log L)^{(1/rd)+(\varepsilon g/r)}$ et $R_0=H/c_{16}$ de sorte que l'on ait pour L assez grand $(T-c_3 gdD^{gd})H^{rd}>c_4LD^{gd}$. Remarquons que l'on a aussi $R_0\leqq\log L$ car $rd>2$ et $\varepsilon<1/2g$.

Reprenons l'idéal J introduit dans le lemme 3.2

$$J=(\tilde{\mathscr{G}}, Q_{\boldsymbol{\sigma},\alpha}(\gamma\cdot\underline{Z}, \underline{F}_\gamma^{(\xi)}(\underline{X})); \alpha\in\mathscr{A}', \xi\in\mathscr{C}_H, |\boldsymbol{\sigma}|\leqq T \text{ et } \gamma\in\mathscr{M}(q, \Gamma_H)).$$

Etant donné que T et H satisfont à l'hypothèse du lemme 3.2 et que $P\notin\tilde{\mathscr{G}}$ on sait que, dans $\mathbb{P}_q\times\mathbb{P}_N$;

$$\mathscr{Z}(J)\subseteq\mathscr{Z}((Z_0))\cup\mathscr{Z}((Z_1, \ldots, Z_q)).$$

En reprenant l'idéal $\mathscr{E}$ de la fin du paragraphe 1, on a $\mathscr{E}\supset\mathscr{G}$. L'idéal de $F[\underline{X}]$ suivant

$$\tilde{J}=(\mathscr{E}, Q_{\boldsymbol{\sigma},\alpha}(1, \gamma\cdot\mathbf{1}, \underline{F}_\gamma^{(\xi)}(\underline{X})); \alpha\in\mathscr{A}', \xi\in\mathscr{C}_H, |\boldsymbol{\sigma}|\leqq T \text{ et } \gamma\in\mathscr{M}(q, \Gamma_H))$$

vérifie donc qu'un point p' de $\mathscr{Z}(\tilde{J})$ dans $\mathbb{P}_N$ donne un point $(1,\mathbf{1})\times p'$ de $\mathscr{Z}(J)$ dans $\mathbb{P}_q\times\mathbb{P}_N$, comme $\mathscr{Z}(J)$ ne contient pas de point de la forme $(1,\mathbf{t})\times p'$ dans $\mathbb{P}_q\times\mathbb{P}_N$ avec $\mathbf{t}\neq\mathbf{0}$, on en conclut que la sous-variété $\mathscr{Z}(\tilde{J})$ de $\mathbb{P}_N$ est vide.

Les relations $R_{\kappa+1},\ldots,R_N$ introduites avec l'idéal $\mathscr{E}$ à la fin du paragraphe 1 sont dans cet idéal et donc dans $\tilde{J}$. Notons $\{P_1,\ldots,P_m\}$ un choix de générateurs de $\mathscr{E}$ et les $(Q_{\boldsymbol{\sigma},\alpha}(1,\gamma\cdot\mathbf{1},\underline{F}_\gamma^{(\xi)}(\underline{X})))$. On peut alors appliquer le lemme 1.2 de [6] qui nous fournit un élément δ, non nul, dans F.

iii) Les estimations arithmétiques

Rappelons que $\kappa=\dim X$ et que $N-\kappa$ est la codimension de l'idéal $\mathscr{E}$ de $F[\underline{X}]$. D'après les estimations des lemmes 1.3, 2.2 et 4.1 on peut affirmer que les polynômes $P_1,\ldots,P_m$ introduit ci-dessus sont des éléments homogènes de $F[\underline{X}]$ de degré $\leqq c_{20}DH^2$ et de hauteur $\leqq c_{20}(L\log H+T\log L+DH^2)$. Reportant ces estimations dans l'énoncé du lemme 1.2 de [6] on trouve que d'une part δ est un élément de $F\setminus\{0\}$ de taille majorée par

$$\begin{aligned}&c_{21}(L\log H+T\log L+DH^2)(DH^2)^\kappa\\&\leqq c_{22}\cdot L\cdot(\log L)^{\frac{2\kappa}{rd}+\varepsilon\kappa\frac{2g+r}{r}}\log\log L,\end{aligned}$$

et d'autre part que pour tout $\beta\in\mathscr{B}$ on a;

$$X_\beta^{c_{23}DH^2}\cdot\delta=\sum_{i=1}^m A_i(\underline{X})\cdot P_i(\underline{X})$$

où A_i est un polynôme homogène de $F[\underline{X}]$ de degré $\leqq c_{23}DH^2$ et de hauteur $\leqq c_{24}(L\log H+T\log L+DH^2)(DH^2)^\kappa$. On en déduit que pour $i=1,\ldots,m$

$$|A_i(\underline{\varphi}(\mathbf{1}))|\leqq\exp\left(c_{25}L(\log L)^{\frac{2\kappa}{rd}+\varepsilon\kappa\frac{2g+r}{r}}\cdot\log\log L\right),$$

mais on sait à la suite du (i) que pour $i=1,\ldots,m$

$$|P_i(\underline{\varphi}(\mathbf{1}))|\leqq\exp\left(-c_{18}L(\log L)^{1+(\varepsilon/2)}\right)$$

et comme on a fait l'hypothèse que $\dfrac{2\kappa}{rd}+\varepsilon\kappa\dfrac{2g+r}{r}<1$ il s'ensuit que

$$|\delta|\leqq\exp\left(-c_{26}L\cdot(\log L)^{1+\varepsilon/2}\right).$$

D'un autre côté l'inégalité de la taille nous assure que

$$|\delta|\geqq\exp\left(-c_{27}L(\log L)^{\frac{2\kappa}{rd}+\varepsilon\kappa\frac{2g+r}{r}}\cdot\log\log L\right).$$

Mais alors l'hypothèse $\dfrac{2\kappa}{rd}+\varepsilon\kappa\dfrac{2g+r}{r}<1$ entraine une contradiction entre la majoration et la minoration de $|\delta|$ obtenues, c'est donc que l'on ne peut choisir

$0<\varepsilon<\left(1-\frac{2\kappa}{rd}\right)\cdot\frac{r}{\kappa(2g+r)}$ d'où l'on déduit que $\frac{2\kappa}{rd}\geqq 1$. On a donc bien montré que

$$\kappa=\dim X\geqq rd/2$$

dans tous les cas du théorème 0.1, et ceci en achève la preuve.

En conclusion faisons quelques remarques:

La démonstration ci-dessus s'apparente naturellement à la démonstration du théorème de Lindemann-Weierstraß (cf. [7]), elle en diffère néanmoins en plusieurs endroits. D'abord la construction de la fonction auxiliaire se fait de façon beaucoup plus grossière que dans le cas exponentiel, ensuite le lemme de zéro que l'on utilise est nettement moins précis que celui que l'on démontre dans le cas exponentiel. Ces deux défauts sont palliés par l'utilisation d'un grand nombre d'endomorphismes et ce n'est que lorsque l'on a suffisamment d'endomorphismes (i.e. dans le cas de multiplications complexes) que l'on obtient le meilleur résultat espérable.

Enfin on se sert d'un procédé d'élimination projective inutilisable tel quel dans le cas exponentiel, où le fait que le groupe algébrique considéré n'est pas fermé pour la topologie de Zariski nécessite l'emploi de techniques d'élimination affine.

Références

1. Brownawell, W.D., Masser, D.W.: Multiplicity estimates for analytic functions II. Duke Mathematical Journal **47**, 273–295 (1980)
2. Choodnovsky, G.V.: Algebraic independence of the values of elliptic functions at algebraic points. Invent. Math. **61**, 267–290 (1981)
3. Masser, D.W., Wüstholz, G.: Zero estimates on group varieties I. Invent. Math. **64**, 489–516 (1981)
4. Nesterenko, Y.V.: Estimates for the orders of zeros of functions of a certain class and their applications in the theory of transcendental numbers. Izv. Akad. Nauk. SSSR. Math. **41**, 253–284 (1977); Math. USSR Izv. **11**, 239–270 (1977)
5. Philippon, P.: Indépendance algébrique et variétés abéliennes. C.R. Acad. Sci. Paris, ser I **294**, 257–259 (1982)
6. Philippon, P.: Variétés abéliennes et indépendance algébrique I. Invent. Math. **70**, 289–318 (1983)
7. Siegel, C.L.: Transcendental numbers. Ann. of Math. Studies, Princeton, n° 16, 102 pp. (1949)
8. Waldschmidt, M.: Nombres transcendants et groupes algébriques, Astérisque 69–70 (1979)
9. Wüstholz, G.: Sur l'analogue abélien du théorème de Lindemann. C.R. Acad. Sci. Paris, ser I **295**, 35–37 (1982)

Oblatum 10-VII & 25-X-1982

Invent. math. 72, 407–464 (1983)

Inventiones mathematicae

Fields of Large Transcendence Degree Generated by Values of Elliptic Functions

D.W. Masser[1] and G. Wüstholz[2]

[1] Department of Mathematics, University of Nottingham, University Park, Nottingham, UK
[2] Gesamthochschule Wuppertal, Fachbereich 7 – Mathematik, Gaußstraße 20, D-5600 Wuppertal, Federal Republic of Germany

Contents

General Introduction

The purpose of this article is to prove the following result.

Main Theorem. *Let $\wp(z)$ be a Weierstrass elliptic function with algebraic invariants and no complex multiplication. For $n \geqq 1$ let $u_1, \ldots, u_n$ be complex numbers linearly independent over $\mathbb{Q}$, and for $m \geqq 1$ let also $v_1, \ldots, v_m$ be complex numbers linearly independent over $\mathbb{Q}$. Assume further that there exists a real number $\kappa \geqq 0$ such that*

$$|t_1 u_1 + \ldots + t_n u_n| \geqq \exp(-T^\kappa), \quad |s_1 v_1 + \ldots + s_m v_m| \geqq \exp(-S^\kappa) \tag{0.1}$$

for all integers $t_1, \ldots, t_n, s_1, \ldots, s_m$ with

$$T = \max(|t_1|, \ldots, |t_n|), \quad S = \max(|s_1|, \ldots, |s_m|)$$

sufficiently large. Finally suppose that for some integer $k \geqq 1$ we have

$$mn \geqq \{2^{k+1}(k+7) + 4\kappa\}\,(m+2n). \tag{0.2}$$

Offprint requests to: G. Wüstholz

Then at least k of the numbers

$$\wp(u_i v_j) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m) \tag{0.3}$$

are defined and are algebraically independent over $\mathbb{Q}$.

This yields the first known examples of fields of arbitrarily large transcendence degree generated by values of elliptic functions. As a very special case we see easily that if $N \geqq 3 \cdot 2^{k+2}(k+7)$ then at least k of the numbers

$$\wp(e), \ldots, \wp(e^N)$$

are defined and are algebraically independent over $\mathbb{Q}$. The same conclusion holds also for

$$\wp(\pi), \ldots, \wp(\pi^N),$$

or indeed when e and π are replaced by any transcendental number x satisfying a rather weak condition. In fact it suffices that for each $d \geqq 1$ and $\varepsilon > 0$ we have the inequality

$$|a_0 + a_1 x + \ldots + a_d x^d| \geqq \exp(-A^{\varepsilon})$$

for all integers $a_0, \ldots, a_d$ with

$$A = \max(|a_0|, \ldots, |a_d|)$$

sufficiently large.

But before we discuss our Theorem further, let us recall what is known at present about the analogous situation when $\wp(z)$ is replaced by the exponential function. The first results were obtained in 1974 by Chudnovsky [6]; among other things he proved that if $u_1, \ldots, u_n$, $v_1, \ldots, v_m$ are complex numbers as above, then, provided the hypotheses (0.1) are strengthened somewhat, the slightly weaker condition (with $k \geqq 2$)

$$mn \geqq 2^{k-1}(m+n) \tag{0.4}$$

is enough to ensure that at least k of the numbers

$$\exp(u_i v_j) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m) \tag{0.5}$$

are algebraically independent over $\mathbb{Q}$. He also proved two companion results: firstly, if the numbers $u_1, \ldots, u_n$ are adjoined to (0.5), then the condition (0.4) can be relaxed to $mn + n \geqq 2^{k-1}(m+n)$, and secondly, if the numbers $v_1, \ldots, v_m$ are further adjoined, then the strict inequality

$$mn + m + n > 2^{k-1}(m+n)$$

suffices.

We should explain here that the proofs of these theorems depend in an essential way on Tijdeman's well-known estimates [24] for small values of exponential polynomials, together with a criterion for algebraic independence that generalises Gelfond's classical criterion [9] for transcendence. In this context we note that relatively recently this criterion has been extended further by Reyssat [21], and with its aid Philippon [19] has given proofs of marginally weaker versions of Chudnovsky's results that are valid even in the p-adic case.

Also, one expects that the above theorems should remain true when 2^{k-1} is replaced by k, and even when the conditions analogous to (0.1) are omitted altogether. Some results of this type have been announced by Chudnovsky (see [7] and [8]), but as yet no proofs have appeared.

Now let us return to the elliptic case. It will be seen that the proof of our Main Theorem is rather long and that it differs significantly from the proofs in [6], [21] and [19] mentioned above. In fact it involves a number of fairly independent components, which we have separated into different chapters. So it may be useful here to give an overall summary of the contents of these chapters. Actually all our difficulties arise from one basic fact: there is no known elliptic analogue of Tijdeman's results [24] for small values of exponential polynomials. Even for the underlying zero estimates [23] the only satisfactory substitute at the moment comes from the general work of the present authors [17] on group varieties.

Thus in Chapter 1 of this paper we give a refinement of the main result of [17], in the same general context. We need to know in some detail what sort of degeneracy can occur when the coefficient $\mu(\Gamma, G)$ of [17] is not maximal. Our result (Theorem I), together with an additional remark (the Proposition), has already been found useful in other areas related to transcendence theory (see for example [14] on heights on abelian varieties).

Then in Chapter 2 we give some technical estimates for primary ideal components in a polynomial ring. These are fairly straightforward deductions from classical degree theory, but we were not able to find a suitable reference in the literature. The estimates (Theorem II and the Corollary) are needed at two different places in the overall argument.

Now the general result of Chapter 1 involves in a fundamental way the algebraic subgroups H of the group variety G. Consequently it is necessary to make a detailed study of such subgroups. This would be rather difficult in the general context, but in the sequel we shall be considering only the example $G=E^n$, where E is the elliptic curve associated with $\wp(z)$. Thus in Chapter 3 we prove some results about algebraic subgroups of E^n, taking as a starting point a well-known theorem of Kolchin [12]. Our Theorem III is in fact a quantitative refinement of Kolchin's Theorem, and the Corollary can be viewed as an elliptic analogue of the Vandermonde determinant. Thus it has perhaps an independent interest.

Next, in Chapter 4 we work out an explicit version (Theorem IV) of Hilbert's Nullstellensatz for polynomials with algebraic coefficients, which we need for the main proof. We use standard techniques from the classical paper [11] of Hermann, and very little here is new, but once again we were unable to find a satisfactory reference in the literature.

Finally in Chapter 5 we give the proof of the Main Theorem. The results of Chapters 1 and 3 yield a special zero estimate for the group variety $G=E^n$. We need to convert this into an estimate for small values. In principle, since E^n is a complete variety, this could be done by means of generalised resultants, but in practice the results obtained in this way appear to be too weak for application to the present problem. Instead we deduce a potential small value estimate by reformulating the zero estimate in purely ideal-theoretic language using the Nullstellensatz of Chapter 4. If we now assume our Main Theorem is false, we can work over a polynomial ring in comparatively few variables corresponding to a transcendence basis. This means that the coefficients and degrees involved do not become too large. Then, right at the end, we construct a suitable auxiliary function in the usual way, with a large number of values which are sufficiently small to yield the final contradiction.

Note that our proof does not use any general criterion of algebraic independence as in [6] or [21]. Another interesting feature is that the purely algebraic estimates of Chapters 2 and 4 are reflected so directly in the equality (0.2) of the Main Theorem. In fact our version of the Nullstellensatz is rather weak and it involves an exponent 2^k which appears as a factor in (0.2). One might expect that this Nullstellensatz remains valid with an exponent of order k, but it seems very difficult to give any improvement whatsoever. In any case, there still remains another factor of order k in (0.2) which arises from the estimates of Chapter 2. As these latter estimates are essentially best possible, it seems that we are still some way from a sharp version of the Main Theorem in which $mn/(m+2n)$ is only of order k. However, it should be noted that such best possible results (and much else besides) follow from the claims of Chudnovsky in [8].

Finally let us discuss possible improvements and extensions of the Main Theorem. As our own results are relatively weak, we have taken no particular pains to find the sharpest estimates that our method will give. It seems likely, for example, that k could be replaced by $k-1$ in (0.2), which immediately halves the estimate. Furthermore the constant 7 in (0.2) can be reduced without too much trouble. But, as we have indicated above, it appears difficult to improve upon the order of magnitude $2^k k$. Let us note in passing, however, that for $k=1$ and $k=2$ much better results are known. In these cases the hypotheses (0.1) can be dropped entirely, and the corresponding inequalities for m and n, due to Ramachandra [20] and the present authors (see [18]), are

$$mn \geqq m+2n, \qquad mn \geqq 2(m+2n)$$

respectively. Similar results hold for the analogous companion theorems mentioned at the beginning of this introduction.

As far as these companion theorems for large k are concerned, there is no difficulty in principle in writing down the appropriate generalisations of Theorem I (see Theorem A of [17]), and these would lead to versions of the Main Theorem with the numbers $u_1, \ldots, u_n$, $v_1, \ldots, v_m$ adjoined to (0.3). But we have not thought it worthwhile to carry out the details, as results in this direction are already trivially implied by the Main Theorem as it stands.

Neither have we thought it worthwhile to consider elliptic functions with complex multiplication, although in principle it would suffice to rewrite Chapter 3 replacing $\mathbb{Z}$ by the ring of endomorphisms of the elliptic curve. In any case the results would be only slightly stronger, and still rather weak.

We end by remarking that the Main Theorem can be extended to arbitrary abelian varieties of dimension $g \geqq 1$; but it then perhaps loses some of its appeal. We now have suitably normalised algebraically independent abelian functions $f_1(z), \ldots, f_g(z)$ meromorphic on $\mathbb{C}^g$, together with elements $u_1, \ldots, u_n$ of $\mathbb{C}^g$ and elements $v_1, \ldots, v_m$ of $\mathbb{C}$. Assume for simplicity that the ring of endomorphisms of the abelian variety is $\mathbb{Z}$. Then with the analogous linear independence measures on the elements $u_1, \ldots, u_n$, $v_1, \ldots, v_m$, the inequality (0.2) becomes

$$g\,m\,n \geqq \{2^{k+1}(k+7g^2)+4g^3\kappa\}\,(m+2g\,n),$$

and this is enough to ensure that at least k of the numbers

$$f_h(u_i v_j) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m,\ 1 \leqq h \leqq g)$$

are defined and are algebraically independent over $\mathbb{Q}$. Once again the only essential modifications to the argument are needed in the results of Chapter 3.

In this paper we have tried to make Chapters 1, 2, 3 and 4 as independent of each other as possible. They all contain results proved in somewhat greater generality than is needed for the proof of the Main Theorem in Chapter 5. We have also taken some care with the calculation of the numerical constants, even though these are unlikely to be important for applications. Throughout we take for granted some basic facts about group varieties, and also some elementary concepts from commutative algebra; see [17] for suitable references. Finally we ought to mention that for convenience we usually work over the complex field $\mathbb{C}$ or its p-adic analogue $\mathbb{C}_p$, although many of our results remain valid for arbitrary fields of zero characteristic.

Chapter 1. A Zero Estimate for General Group Varieties

1. Introduction

In this chapter we give a refinement of the Main Theorem of [17]. This refinement has already been found useful in other applications. But before we state the result we recall the basic definitions in [17].

Let K be either the complex field $\mathbb{C}$ or the analogous p-adic field $\mathbb{C}_p$ for some rational prime p. Let G be a quasiprojective commutative group variety of dimension $n \geqq 1$ defined over K. Suppose G is embedded in projective space $\mathbb{P}_N$ of dimension $N \geqq 1$. We define constants a, b, depending only on G and its embedding, as follows. Firstly, let $a \geqq 1$ be any integer satisfying the conditions of Lemma 1 (p. 492) of [17]. In other words, for any finitely generated subgroup Γ of G and any γ in Γ there are homogeneous polynomials of degrees at most a representing translation by γ on an open set of G containing Γ. Secondly, if $N=n$, let $b=1$; otherwise we let $b \geqq 1$ be any integer satisfying the conditions of Lemma 5 (p. 499) of [17]. In other words, the Zariski closure of G in $\mathbb{P}_N$ is defined by the vanishing of homogeneous polynomials of degrees at most b. Now we put

$$c = a^{-n} b^{-(N-n)} \leqq 1;$$

this is the constant appearing in the Main Theorem of [17].

Next, we shall be considering algebraic subsets S of G, and for brevity we shall say that S can be defined in G by the homogeneous polynomials $P_1, \ldots, P_k$ if the set of common zeroes in $\mathbb{P}_N$ of $P_1, \ldots, P_k$ meets G precisely on S. Thus S is itself a quasiprojective algebraic set, and we recall that its dimension is the maximum of the dimensions of those irreducible components of its Zariski closure in $\mathbb{P}_N$ that meet G.

Finally, for $m \geqq 1$ let $\mathbb{Z}^m$ denote the group of all elements $\sigma = (s_1, \ldots, s_m)$ for integers $s_1, \ldots, s_m$, and write

$$|\sigma| = \max(|s_1|, \ldots, |s_m|).$$

For real $S \geqq 0$ let $\mathbb{Z}^m(S)$ denote as usual the subset of all such σ with $0 \leqq s_1, \ldots, s_m \leqq S$. We fix elements $\gamma_1, \ldots, \gamma_m$ of G and we define a homomorphism Ψ from $\mathbb{Z}^m$ to G by

$$\Psi(\sigma) = s_1 \gamma_1 + \ldots + s_m \gamma_m.$$

Our main result can now be stated as follows.

Theorem I. *Suppose for some integer $D \geqq 1$ and some real $\theta \geqq n/m$ there is a homogeneous polynomial P, of degree at most D, that vanishes on the set $\Psi(\mathbb{Z}^m(n(D/c)^\theta))$ but not on all of G. Then there are integers k, r with*

$$1 \leqq k \leqq m, \quad 1 \leqq r \leqq n, \quad k + r\theta^{-1} > m,$$

together with a subgroup Z of $\mathbb{Z}^m$ of rank at least k and an algebraic subgroup H of G of dimension at most $n-r$, such that $\Psi(Z) \subseteq H$. Furthermore Z contains elements $\sigma_1, \ldots, \sigma_k$, linearly independent over $\mathbb{Z}$, with

$$|\sigma_j| \leqq (D/c)^{r/(m+1-j)} \qquad (1 \leqq j \leqq k),$$

and H is contained in an algebraic subset S of G, of dimension at most $n-r$, that is defined in G by homogeneous polynomials of degrees at most D/c.

We prove this result in sections 2 and 3 of this chapter. However, in applications (see for example Chapter 3), the Theorem has to be combined with some sort of classification of the possible algebraic subgroups of G, using for example the well-known result of Kolchin [12]. To this end it is often useful to know the equations defining H itself in Theorem I, rather than merely those defining the larger set S. The extra information needed can be deduced from a Proposition perhaps of some interest in its own right. We give this in section 4.

We conclude this introduction by observing that Theorem I does indeed imply the Main Theorem of [17]. For suppose Γ is a subgroup of G generated by elements $\gamma_1, \ldots, \gamma_m$, and let $\mu = \mu(\Gamma, G)$ be the index defined in [17]. For real $D \geqq 0$ and $S \geqq 0$ let P be a homogeneous polynomial of degree at most D vanishing on $\Gamma(S) = \Psi(\mathbb{Z}^m(S))$. If $D < c(S/n)^\mu$, we have to show that P vanishes on all of G. As in [17] we may suppose that in fact D is a positive integer and $\mu > 0$. In this case $S > n(D/c)^{1/\mu}$, so if P does not vanish on all of G we can apply Theorem I with $\theta = 1/\mu$. From the definition of μ we have $\mu \leqq m/n$, so $\theta \geqq n/m$ as required. We deduce that there are integers k, r with

$$1 \leqq k \leqq m, \quad 1 \leqq r \leqq n, \quad k + r\mu > m,$$

together with a subgroup Z of $\mathbb{Z}^m$ of rank at least k and an algebraic subgroup H of G of dimension at most $n-r$, such that $\Psi(Z) \subseteq H$. Now if l is the rank of Γ, the kernel of the map Ψ from $\mathbb{Z}^m$ to Γ has rank $m-l$. Hence the rank of $\Psi(Z)$ is at least $k-(m-l)=l-(m-k)$, and so its corank is at most $m-k$. Thus if $s \geqq r$ is the codimension of H in G, the integer p_s of [17] satisfies $p_s \leqq m-k$. It follows that

$$p_s/s \leqq (m-k)/s \leqq (m-k)/r < \mu,$$

which contradicts the definition of μ. Hence P does indeed vanish on all of G, and this gives the Main Theorem of [17].

2. A Local Result

Let Γ be an arbitrary finitely generated subgroup of G containing the group $\Psi(\mathbb{Z}^m)$. We show here that a relatively straightforward modification of the arguments of [17] yield a version of Theorem I localised with respect to Γ. In section 3 we shall extend the localisation to all of G, giving the full Theorem.

Thus if $\mathfrak{J}$ is an ideal of $K[X_0, \ldots, X_N]$ we denote the corresponding contracted extension of $\mathfrak{J}$ by $\mathfrak{J}^*$, as in [17] (p. 494). It is the ideal of $K[X_0, \ldots, X_N]$ generated by all polynomials P such that QP is in $\mathfrak{J}$ for some polynomial Q not vanishing at any point of Γ. Finally, we let $\mathfrak{G}$ be the prime ideal corresponding to G, and we put $h = N - n$.

Technical Lemma. *Suppose for some integer $D \geqq 1$ and some real $\theta \geqq n/m$ there is a homogeneous polynomial P, of degree at most D, that vanishes on $\Psi(\mathbb{Z}^m(n(D/c)^\theta))$ but not on all of G. Then there are integers k, r with*

$$1 \leqq k \leqq m, \quad 1 \leqq r \leqq n, \quad k + r\theta^{-1} > m, \tag{1.1}$$

together with a subgroup Z of $\mathbb{Z}^m$ of rank k and homogeneous polynomials $Q_1, \ldots, Q_r$, of degrees at most $a^n D$, vanishing on $\Psi(Z)$, such that if $\mathfrak{J} = (Q_1, \ldots, Q_r)$ the ideal $(\mathfrak{J}, \mathfrak{G})^*$ has rank $h+r$. Furthermore Z contains elements $\sigma_1, \ldots, \sigma_k$, linearly independent over $\mathbb{Z}$, with

$$|\sigma_j| \leqq (D/c)^{r/(m+1-j)} \quad (1 \leqq j \leqq k). \tag{1.2}$$

Proof. We argue much as in the proof of the Inductive Lemma (p. 504) of [17], and we shall be relatively brief where the arguments are especially similar.

We start by determining the integer r. Write

$$D_{h+r} = a^{r-1} D \quad (1 \leqq r \leqq n+1)$$

and

$$B_{h+r} = b^h D_{h+1} \ldots D_{h+r} \quad (1 \leqq r \leqq n).$$

If $h \geqq 1$ let $P_1, \ldots, P_h$ be the polynomials constructed in Lemma 6 (p. 499) of [17]. Consider the following statement for $1 \leqq r \leqq n+1$, which we shall denote by (I_r).

(I_r): There exist homogeneous polynomials $P_{h+1}, \ldots, P_{h+r}$, of degrees at most $D_{h+1}, \ldots, D_{h+r}$ respectively, such that the ideal $\mathfrak{J}_r = (P_1, \ldots, P_{h+r})$ vanishes on

$$\Gamma_r = \Psi(\mathbb{Z}^m((n-r+1)(D/c)^\theta))$$

and

(i) the rank of $\mathfrak{J}_r^*$ is $h+r$,

(ii) if $r \neq n+1$ the degree of $\mathfrak{J}_r^*$ is at most B_{h+r}.

We note that (I_1) is true while (I_{n+1}) is false. To verify the former, we take $P_{h+1} = P$. Then $\mathfrak{J}_1$ vanishes on Γ_1 by hypothesis, so it suffices to check (i) and (ii). But since $\mathfrak{J}_1, \mathfrak{J}_1^*$ vanish at the same points of Γ, we can repeat the arguments of the second and third paragraphs of the proof of the Inductive Lemma of [17], with $\Gamma(S)$ replaced by the non-empty set Γ_1. Likewise, if (I_{n+1}) were true, the ideal $\mathfrak{J}_{n+1}^*$ would have rank $h+n+1 = N+1$, so it would be primary with associated prime ideal $(X_0, \ldots, X_N)$. But in that case neither $\mathfrak{J}_{n+1}$ nor $\mathfrak{J}_{n+1}^*$ could vanish at the origin of G, which is just the set Γ_{n+1}. Hence indeed (I_1) is true while (I_{n+1}) is false.

Therefore there exists an integer r, with $1 \leqq r \leqq n$, such that (I_r) is true while (I_{r+1}) is false. We proceed to construct the desired subgroup Z of $\mathbb{Z}^m$, and for this we shall need the operators $E(\gamma)$, $\mathscr{E}(\gamma)$ of [17] (pp. 494–5) defined for each γ in Γ.

Firstly, by Lemma 7 (p. 501) of [17], the ideal $\mathfrak{J}_r^*$ is unmixed. We claim that there exists some prime component $\mathfrak{P}$ of $\mathfrak{J}_r^*$ such that

$$\mathscr{E}(\gamma)\,\mathfrak{J}_r^* \subseteq \mathfrak{P} \tag{1.3}$$

for all γ in $\Gamma_n = \Psi(\mathbb{Z}^m((D/c)^\theta))$. Otherwise, for each such $\mathfrak{P}$ there would exist γ in Γ_n with $\mathscr{E}(\gamma)\,\mathfrak{J}_r^* \nsubseteq \mathfrak{P}$. As in the arguments following equation (30) of [17] (pp. 505–6), we deduce that $E(\gamma)\,\mathfrak{J}_r^* \nsubseteq \mathfrak{P}$; and similarly there exists j with $h+1 \leqq j \leqq h+r$ such that $Q = E(\gamma) P_j$ is not in $\mathfrak{P}$. Since $\mathfrak{J}_r$ vanishes on Γ_r, it follows that Q vanishes on Γ_{r+1}. Now by taking P_{h+r+1} as a suitable combination of such Q in the standard way we obtain a proof of the statement (I_{r+1}), contrary to assumption. Hence indeed there is a prime component $\mathfrak{P}$ of $\mathfrak{J}_r^*$ such that (1.3) holds for all γ in Γ_n.

It follows that for each γ in Γ_n the ideal $\mathscr{E}(-\gamma)\,\mathfrak{P}$ is a prime component of $\mathfrak{J}_r^*$. We define an equivalence relation on $\mathbb{Z}^m((D/c)^\theta)$ by saying that τ is equivalent to τ' if the images $\gamma = \Psi(\tau)$, $\gamma' = \Psi(\tau')$ satisfy $\mathscr{E}(-\gamma)\,\mathfrak{P} = \mathscr{E}(-\gamma')\,\mathfrak{P}$. In this case the difference $\tau' - \tau$ lies in the set Σ of all σ in $\mathbb{Z}^m$ with $|\sigma| \leqq (D/c)^\theta$ such that $\mathscr{E}(-\Psi(\sigma))\,\mathfrak{P} = \mathfrak{P}$. Also the number B of such equivalence classes does not exceed the degree of $\mathfrak{J}_r^*$, which is at most B_{h+r}.

We now put

$$k = 1 + [m - r\theta^{-1}], \tag{1.4}$$

and we construct by induction elements $\sigma_1, \ldots, \sigma_k$ of Σ, linearly independent over $\mathbb{Z}$, for which (1.2) holds. To start with, we use Lemma 8 (p. 501) of [17] with $S = (D/c)^{r/m}$ and $q = 1$, after noting that

$$B \leqq B_{h+r} \leqq (D/c)^r = S^m.$$

This gives two distinct elements of $\mathbb{Z}^m(S)$; and since $S \leqq (D/c)^\theta$ their difference σ_1 is a non-zero element of Σ satisfying (1.2). So if $k=1$ the inductive construction is finished. Otherwise, if $k \geqq 2$, we suppose $\sigma_1, \ldots, \sigma_{j-1}$ to have been constructed for some j with $2 \leqq j \leqq k$. We can use Lemma 8 of [17] with $S=(D/c)^{r/(m+1-j)}$ and $q=j$, as now

$$B \leqq B_{h+r} \leqq (D/c)^r = S^{m+1-j}.$$

This gives elements $\tau_1, \tau_1', \ldots, \tau_j, \tau_j'$ of $\mathbb{Z}^m(S)$ such that τ_i' is equivalent to $\tau_i (1 \leqq i \leqq j)$ and the differences $\tau_1' - \tau_1, \ldots, \tau_j' - \tau_j$ are linearly independent over $\mathbb{Z}$. Since $j \leqq k$ we see from (1.4) that $m+1-j \geqq r\theta^{-1}$; so once more $S \leqq (D/c)^\theta$ and hence all these differences lie in Σ. At least one must be linearly independent of $\sigma_1, \ldots, \sigma_{j-1}$, and this one we call σ_j. Clearly (1.2) holds, and the inductive construction is now complete.

Finally, as $\mathfrak{P}^* = \mathfrak{P}$, we see that $\mathfrak{P}$ vanishes at some point γ_0 of Γ. We proceed now to verify that the Technical Lemma holds with

$$Q_i = E(\gamma_0) P_{h+i} \qquad (1 \leqq i \leqq r), \tag{1.5}$$

and with Z as the subgroup of $\mathbb{Z}^m$ generated by $\sigma_1, \ldots, \sigma_k$.

To begin with, the basic inequalities (1.1) are immediate from (1.4). Furthermore Z is of rank exactly k and we have the desired estimates (1.2) for its basis elements. Next, the polynomials $Q_1, \ldots, Q_r$ have degrees at most $aD_{h+r} \leqq a^r D \leqq a^n D$, and they lie in the ideal $\mathscr{E}(\gamma_0)\mathfrak{J}_r$, by (1.5). Hence they lie in $\mathscr{E}(\gamma_0)\mathfrak{J}_r^* \subseteq \mathscr{E}(\gamma_0)\mathfrak{P}^* = \mathscr{E}(\gamma_0)\mathfrak{P}$. Now if δ is any element of $\Psi(Z)$, then $\mathscr{E}(-\delta)\mathfrak{P} = \mathfrak{P}$, as this already holds for the generators $\Psi(\sigma_1), \ldots, \Psi(\sigma_k)$ of $\Psi(Z)$. Therefore $Q_1, \ldots, Q_r$ lie in $\mathscr{E}(\gamma_0)\mathscr{E}(-\delta)\mathfrak{P} = \mathscr{E}(\gamma_0 - \delta)\mathfrak{P}$. As $\mathfrak{P}$ vanishes at γ_0, it follows from Lemma 4 (p. 497) of [17] that $\mathscr{E}(\gamma_0 - \delta)\mathfrak{P}$ vanishes at $\gamma_0 - (\gamma_0 - \delta) = \delta$. Thus indeed $Q_1, \ldots, Q_r$ vanish on all of $\Psi(Z)$.

It remains only to prove that if $\mathfrak{J} = (Q_1, \ldots, Q_r)$ the ideal $(\mathfrak{J}, \mathfrak{G})^*$ has rank $h+r$. But $(\mathfrak{J}, \mathfrak{G})^* = \mathscr{E}(\gamma_0)\mathfrak{J}_r$, for if $h \geqq 1$ the polynomials $E(\gamma_0)P_1, \ldots, E(\gamma_0)P_h$ already lie in $\mathfrak{G}$, by Lemma 2 (p. 494) of [17]. Since $\mathfrak{J}_r^*$ is special of rank $h+r$, so is $\mathscr{E}(\gamma_0)\mathfrak{J}_r^*$ again by Lemma 4 of [17], and so it suffices now to check that $\mathscr{E}(\gamma_0)\mathfrak{J}_r = \mathscr{E}(\gamma_0)\mathfrak{J}_r^*$. In fact this holds for any homogeneous ideal $\mathfrak{J}_r$ whatsoever, as we see from the following simple argument.

Let Q be an arbitrary homogeneous polynomial of the set $E(\gamma_0)\mathfrak{J}_r^*$, so that $Q = E(\gamma_0)P$ for some P in $\mathfrak{J}_r^*$. Then MP is in $\mathfrak{J}_r$ for some M in the multiplicative set $\mathscr{M}$ of [17] (p. 494), and on applying $E(\gamma_0)$ we deduce that $M'Q$ lies in $E(\gamma_0)\mathfrak{J}_r$, where $M' = E(\gamma_0)M$. By Lemma 2 of [17] we see that M' is in $\mathscr{M}$, and therefore Q lies in $(E(\gamma_0)\mathfrak{J}_r, \mathfrak{G})^* = \mathscr{E}(\gamma_0)\mathfrak{J}_r$. As Q was arbitrary, this shows that $E(\gamma_0)\mathfrak{J}_r^* \subseteq \mathscr{E}(\gamma_0)\mathfrak{J}_r$. Thus

$$\mathscr{E}(\gamma_0)\mathfrak{J}_r^* = (E(\gamma_0)\mathfrak{J}_r^*, \mathfrak{G})^* \subseteq (\mathscr{E}(\gamma_0)\mathfrak{J}_r, \mathfrak{G})^* = \mathscr{E}(\gamma_0)\mathfrak{J}_r.$$

As the opposite inclusion is trivial, we conclude that indeed $\mathscr{E}(\gamma_0)\mathfrak{J}_r^* = \mathscr{E}(\gamma_0)\mathfrak{J}_r$. As noted above, this completes the proof of the Technical Lemma.

3. Proof of Theorem 1

Here we get rid of the local nature of the Technical Lemma by taking larger and larger groups Γ. We start with an integer $D \geqq 1$ and a real number $\theta \geqq n/m$, and a homogeneous polynomial P of degree at most D vanishing on the set $\Psi(\mathbb{Z}^m(n(D/c)^\theta))$ but not on all of G. For each integer $i \geqq 0$ we shall define a finitely generated subgroup Γ_i of G containing $\Psi(\mathbb{Z}^m)$, and integers k_i, r_i with

$$1 \leqq k_i \leqq m, \quad 1 \leqq r_i \leqq n, \; k_i + r_i\,\theta^{-1} > m, \tag{1.6}$$

together with a subgroup Z_i of $\mathbb{Z}^m$ of rank k_i generated by elements satisfying (1.2) with $k = k_i$, and an ideal $\mathfrak{J}_i$ generated by homogeneous polynomials of degrees at most $a^n D$ vanishing on $\Psi(Z_i)$, such that the ideal $(\mathfrak{J}_i, \mathfrak{G})^{(i)}$ has rank $h + r_i$. Here $\mathfrak{J}^{(i)}$ denotes the contracted extension with respect to Γ_i of the ideal $\mathfrak{J}$ of $K[X_0, \ldots, X_N]$.

We put $\Gamma_0 = \Psi(\mathbb{Z}^m)$ and we give an inductive construction consisting of the following two steps.

(I): Having defined Γ_i for some $i \geqq 0$, to define k_i, r_i, Z_i and $\mathfrak{J}_i$.

(II): Having defined Γ_{i-1} and $\mathfrak{J}_0, \ldots, \mathfrak{J}_{i-1}$ for some $i \geqq 1$, to define Γ_i.

We start with (I). As Γ_i contains $\Psi(\mathbb{Z}^m)$, we may apply the Technical Lemma with $\Gamma = \Gamma_i$. This immediately gives us k_i, r_i, Z_i and $\mathfrak{J}_i$ satisfying the required conditions.

As for (II), we pick any positive integer $l \leqq i$ and then any integers $i_1, \ldots, i_l$ satisfying $0 \leqq i_1 < \ldots < i_l \leqq i-1$, and we define a finite set $\Omega(i_1, \ldots, i_l)$ of points of G in the following way. Consider the isolated prime components of the ideal $(\mathfrak{J}_{i_1}, \ldots, \mathfrak{J}_{i_l}, \mathfrak{G})$ whose associated irreducible variety V meets G, and for each one of these choose arbitrarily a point on the intersection $V \cap G$. These points make up $\Omega(i_1, \ldots, i_l)$. We then let Γ_i be the group generated over Γ_{i-1} by the points in all these finite sets as l and $i_1, \ldots, i_l$ run over all possibilities.

This completes the inductive construction, and it is clear that Γ_i contains $\Psi(\mathbb{Z}^m)$ for each $i \geqq 0$.

Now there are only finitely many possibilities for the integers k_i, r_i, and also for the generators of Z_i, as they satisfy (1.2). Hence by (1.6) we can find integers k, r satisfying (1.1) together with a subgroup Z of $\mathbb{Z}^m$ of rank k whose generators satisfy (1.2) such that $k_i = k$, $r_i = r$ and $Z_i = Z$ for an infinite increasing sequence of non-negative integers $i = i_1, i_2, \ldots$.

By the ascending chain condition for ideals, there exists a positive integer l such that if $\mathfrak{J} = (\mathfrak{J}_{i_1}, \ldots, \mathfrak{J}_{i_l})$ we have $\mathfrak{J}_{i_{l+1}} \subseteq \mathfrak{J}$. The ideal $\mathfrak{J}$ is generated by homogeneous polynomials of degrees at most $a^n D$; let S be the algebraic subset of G defined in G by these polynomials. We proceed to prove that the dimension of S is at most $n-r$.

It suffices to show that every isolated prime component $\mathfrak{P}$ of $(\mathfrak{J}, \mathfrak{G})$ whose associated irreducible variety V meets G has rank at least $N-(n-r) = h+r$. By the inductive step (II) with $i = i_{l+1}$, the subgroup $\Gamma_{i_{l+1}}$ meets V. Hence $\mathfrak{P}^{(i_{l+1})} = \mathfrak{P}$. But now as $(\mathfrak{J}_{i_{l+1}}, \mathfrak{G}) \subseteq (\mathfrak{J}, \mathfrak{G}) \subseteq \mathfrak{P}$ this gives

$$(\mathfrak{J}_{i_{l+1}}, \mathfrak{G})^{(i_{l+1})} \subseteq \mathfrak{P}^{(i_{l+1})} = \mathfrak{P}.$$

Hence the rank of $\mathfrak{P}$ is at least $h + r_{i_{l+1}} = h + r$, which is what we want.

Finally, as all the ideals $\mathfrak{J}_{i_1}, \ldots, \mathfrak{J}_{i_t}$ vanish on $\Psi(Z)$, it follows that S contains $\Psi(Z)$. Therefore S contains the relative Zariski closure H of $\Psi(Z)$ in G, which is an algebraic subgroup of G. As S is defined in G by homogeneous polynomials of degrees at most $a^n D \leqq D/c$, this completes the proof of Theorem I.

4. Interposing the Subgroup

We shall need the following result in Chapter 3 when we study in more detail the algebraic subgroups of E^n. We give it here because the proof works for any group variety.

Proposition. *Let H be an algebraic subgroup of G contained in an algebraic subset S of G. For an integer $D \geqq 1$ suppose that S is defined in G by homogeneous polynomials of degrees at most D. Then there exists an algebraic subgroup H' of G, itself defined in G by homogeneous polynomials of degrees at most aD, such that*

$$H \subseteq H' \subseteq S. \tag{1.7}$$

Proof. We define H' as follows. Firstly we let V be the intersection of the translates $S-h$ as h runs over all elements of H. Then we let H' be the stabiliser of V; that is, the set of all h' in G such that $h'+V=V$. Evidently H' is a subgroup of G. We start by verifying (1.7).

It is clear from group theory that

$$h+V=V \tag{1.8}$$

for all h in H; thus immediately $H \subseteq H'$. Next, as h is in S for all h in H, it follows that V contains the origin 0 of G. Hence if h' is any element of H' we see in particular that $h'=h'+0$ is in V; in other words, $H' \subseteq V$. As $V \subseteq S$, this shows that $H' \subseteq S$ and so establishes (1.7).

It remains only to prove that H' is defined in G by homogeneous polynomials of degrees at most aD (in particular H' is then an algebraic subgroup of G). We first observe that if I temporarily denotes the intersection of the translates $S-v$ as v runs over all points of V, then $H'=I$. To see this, let g be any element of G. Then g is in H' if and only if $g+V=V$. However, it is easy to see by applying the descending chain condition to the sequence

$$V, g+V, 2g+V, \ldots$$

that this holds if and only if $g+V \subseteq V$. From the definition of V, this in turn holds if and only if

$$g+V \subseteq S-h \tag{1.9}$$

for all h in H. Finally by adding h to both sides of (1.9) and using (1.8) we see that g is in H' if and only if $g+V \subseteq S$. But this latter inclusion says simply that g is in I. Hence indeed $H'=I$ as asserted.

Now we can obtain the polynomials defining H' from those defining S by translating by points of V. Fix γ and δ in G. By applying Lemma 1 of [17] with Γ as the group generated by γ and δ we see that there exist homogeneous polynomials $E_i^{(\gamma,\delta)}(X_0,\ldots,X_N)$ $(0 \leqq i \leqq N)$, of degrees at most a, and an open set $\mathcal{O}^{(\gamma,\delta)}$ of G containing δ, such that for any g in $\mathcal{O}^{(\gamma,\delta)}$ with projective coordinates $x_0,\ldots,x_N$ the numbers

$$E_i^{(\gamma,\delta)}(x_0,\ldots,x_N) \quad (0 \leqq i \leqq N) \tag{1.10}$$

are projective coordinates of $g+\gamma$. Furthermore the usual argument of Zariski continuity ([17] p. 492) shows that, if necessary by enlarging $\mathcal{O}^{(\gamma,\delta)}$, we can assume that if g is in G but not in $\mathcal{O}^{(\gamma,\delta)}$ then the above numbers (1.10) are all zero.

Now suppose that S is defined in G by the homogeneous polynomials $P_1,\ldots,P_k$ of degrees at most D. We proceed to verify that the set H' is defined in G by the homogeneous polynomials

$$P_j(E_0^{(v,\delta)}(X_0,\ldots,X_N),\ldots,E_N^{(v,\delta)}(X_0,\ldots,X_N)) \quad (1 \leqq j \leqq k) \tag{1.11}$$

of degrees at most aD, as v runs over all points of V and δ runs over all points of G.

In one direction, suppose g in H' has projective coordinates $x_0,\ldots,x_N$. To show that all the polynomials (1.11) vanish at g, fix v in V and δ in G. If g does not lie in $\mathcal{O}^{(v,\delta)}$, then the numbers

$$y_i = E_i^{(v,\delta)}(x_0,\ldots,x_N) \quad (0 \leqq i \leqq N) \tag{1.12}$$

are all zero. Now none of $P_1,\ldots,P_k$ can be non-zero constants, else S would be empty, contradicting $H \subseteq S$. Hence in this case we have trivially

$$P_j(y_0,\ldots,y_N)=0 \quad (1 \leqq j \leqq k), \tag{1.13}$$

so that the polynomials (1.11) vanish at g. Otherwise, g does lie in $\mathcal{O}^{(v,\delta)}$, and so $y_0,\ldots,y_N$ are projective coordinates of $g+v$. But as $H'=I$ we see that $g+v$ lies in S. Hence again (1.13) holds, showing as before that the polynomials (1.11) vanish at g.

Conversely, let g be an element of G with projective coordinates $x_0,\ldots,x_N$ at which all the polynomials (1.11) vanish. To show that g lies in H', fix an arbitrary v in V. Then g lies in $\mathcal{O}^{(v,\delta)}$ for $\delta=g$, and so the numbers (1.12) with $\delta=g$ are projective coordinates of $g+v$. By hypothesis (1.11) and (1.12) show that (1.13) holds, whence $g+v$ lies in S. As v was arbitrary, we conclude that g lies in $I=H'$. This completes the proof of the Proposition.

Chapter 2. Counting Primary Ideal Components in Polynomial Rings

1. Introduction

Let K be either $\mathbb{C}$ or $\mathbb{C}_p$, and for $n \geqq 1$ denote by $\mathfrak{R}=K[X_0,\ldots,X_n]$ the corresponding polynomial ring in the independent variables $X_0,\ldots,X_n$. In this chapter we give a general estimate, that seems to be not quite well-known, for primary components of a homogeneous ideal $\mathfrak{J}$ of $\mathfrak{R}$. While the result is more or less implicit in the arguments of [4] (section 5), it seems clearer to give an explicit statement for future applications. Recall that a non-zero proper homogeneous ideal $\mathfrak{J}$ of $\mathfrak{R}$ has rank r satisfying $1 \leqq r \leqq n+1$. It is convenient here to define the rank of the zero ideal (0) as zero. Also, if $1 \leqq r \leqq n$, we denote the degree of $\mathfrak{J}$ by $\deg \mathfrak{J}$. In addition, for integers k, t with $1 \leqq t \leqq k$ and integers $D_1 \geqq 1, \ldots, D_k \geqq 1$ we write $M_t(D_1, \ldots, D_k)$ for the maximum of the products of $D_1, \ldots, D_k$ taken t at a time. If $t=0$ we interpret this expression as unity. Our main result can then be stated as follows.

Theorem II. *For integers r, D with $0 \leqq r \leqq n$ and $D \geqq 1$ let $\mathfrak{J}_0$ be an ideal of $\mathfrak{R}$ of rank r generated by homogeneous polynomials of degrees at most D. For integers $k \geqq 1$ and $D_1 \geqq D, \ldots, D_k \geqq D$ let $P_1, \ldots, P_k$ be homogeneous polynomials of degrees at most $D_1, \ldots, D_k$ respectively. Suppose that $\mathfrak{J}=(\mathfrak{J}_0, P_1, \ldots, P_k)$ is a non-zero ideal of $\mathfrak{R}$ and that for some s with $1 \leqq s \leqq n$ it has at least one isolated primary component of rank s. Then $s \geqq r$, and as $\mathfrak{Q}$ runs over all isolated primary components of $\mathfrak{J}$ of rank s we have*

$$\sum \deg \mathfrak{Q} \leqq D^{s-t} M_t(D_1, \ldots, D_k)$$

where

$$t = \min(s-r, k).$$

This result is essentially best possible, even with regard to the constants. It can be shown that it remains true even if $s=r=n+1$ as long as $\deg \mathfrak{Q}$ is defined in this case as the length $l(\mathfrak{Q})$ of $\mathfrak{Q}$. In fact a homogeneous ideal of rank $n+1$ is necessarily primary with associated prime $(X_0, \ldots, X_n)$, so the ideal $\mathfrak{J}$ cannot have more than a single such primary component.

We note, however, that the result becomes false if the conditions $D_1 \geqq D, \ldots, D_k \geqq D$ are dropped. An example is given by $n=2$, $r=1$, $\mathfrak{J}_0=(X_0 X_1^2, X_0 X_2^2)$, $D=3$, $k=1$, $D_1=1$, $P_1=0$. For then

$$\mathfrak{J} = \mathfrak{J}_0 = (X_0) \cap \mathfrak{Q}_2$$

has an isolated primary component $\mathfrak{Q}_2=(X_1^2, X_2^2)$ of rank $s=2$ and degree $4>3$. Less trivially we could take $n=3$ with $\mathfrak{J}_0$ as above, but now $k=2$, $D_1=D_2=1$, $P_1=P_2=X_3$. In this case we find that

$$\mathfrak{J} = (\mathfrak{J}_0, X_3) = (X_0, X_3) \cap \mathfrak{Q}_3$$

has an isolated primary component $\mathfrak{Q}_3=(X_1^2, X_2^2, X_3)$ of rank $s=3$ and degree $4>3$.

We shall deduce the following consequence for polynomials in the affine ring ${}^a\mathfrak{R}=K[x_1, \ldots, x_n]$ that are not necessarily homogeneous. Recall that the length $l(\mathfrak{Q})$ of a proper non-zero primary ideal $\mathfrak{Q}$ of ${}^a\mathfrak{R}$ is still well-defined.

Corollary. *For integers $k \geqq 1$ and $D_1 \geqq 1, \ldots, D_k \geqq 1$ let $Q_1, \ldots, Q_k$ be polynomials of ${}^a\mathfrak{R}$ of total degrees at most $D_1, \ldots, D_k$ respectively. Suppose that $\mathfrak{J}=(Q_1, \ldots, Q_k)$ is a proper non-zero ideal of ${}^a\mathfrak{R}$ and that for some s with $1 \leqq s \leqq n$ it has an isolated primary component of rank s. Then as $\mathfrak{Q}$ runs over all isolated primary components of $\mathfrak{J}$ of rank s, we have*

$$\sum l(\mathfrak{Q}) \leqq M_t(D_1, \ldots, D_k)$$

where

$$t = \min(s, k).$$

In fact by using a theory of degree in ${}^a\mathfrak{R}$ (see Brownawell [3]) one can establish a precise analogue of Theorem II for affine ideals, but we content ourselves here with a simplified situation.

2. Preliminaries

In the inductive construction of the next section we shall make frequent use of a standard argument from [4] and [17]. Thus it is convenient to formulate it here in a general context.

For $m \geqq 1$ let $\mathfrak{M}_1, \ldots, \mathfrak{M}_m$ be non-homogeneous maximal ideals of $\mathfrak{R}$, and let $\mathfrak{N}$ be their intersection. For an ideal $\mathfrak{J}$ of $\mathfrak{R}$ we write $\mathfrak{J}^*$ for the associated contracted extension; it is the ideal of all P in $\mathfrak{R}$ such that QP is in $\mathfrak{J}$ for some Q not in any of $\mathfrak{M}_1, \ldots, \mathfrak{M}_m$. Clearly $\mathfrak{N}^* = \mathfrak{N}$. As K is algebraically closed, the ideals $\mathfrak{M}_1, \ldots, \mathfrak{M}_m$ correspond to non-zero points of K^{n+1}, and so we can select a linear form L in $\mathfrak{R}$ not in any of $\mathfrak{M}_1, \ldots, \mathfrak{M}_m$.

Lemma. *For integers $p \geqq 1$, $E \geqq 1$ let $S_1, \ldots, S_p$ be homogeneous polynomials of $\mathfrak{R}$ of degrees at most E, and let $\mathscr{S}$ be the set of homogeneous polynomials of the form*

$$\lambda_1 L^{a_1} S_1 + \ldots + \lambda_p L^{a_p} S_p$$

for non-negative integers $a_1, \ldots, a_p$ and elements $\lambda_1, \ldots, \lambda_p$ of K with $\lambda_1 = 1$. Let i be an integer with $0 \leqq i < n$. There are two cases.

(A): $i = 0$. *Suppose in this case that not all of $S_1, \ldots, S_p$ are zero. Then there exists R_1 in $\mathscr{S}$ such that if $\mathfrak{J}_1 = (R_1)$ the ideal $\mathfrak{J}_1^*$ is of rank* 1 *and of degree at most E.*

(B): $1 \leqq i < n$. *In this case let $R_1, \ldots, R_i$ be homogeneous polynomials of $\mathfrak{R}$ such that if $\mathfrak{J}_i = (R_1, \ldots, R_i)$ the ideal $\mathfrak{J}_i^*$ has rank i. Suppose for each prime component $\mathfrak{P}$ of $\mathfrak{J}_i^*$ of rank i that not all of $S_1, \ldots, S_p$ lie in $\mathfrak{P}$. Then there exists R_{i+1} in $\mathscr{S}$ such that if $\mathfrak{J}_{i+1} = (\mathfrak{J}_i, R_{i+1})$ the ideal $\mathfrak{J}_{i+1}^*$ is of rank $i+1$ and of degree at most $E \deg \mathfrak{J}_i^*$.*

Proof. Let e_j be the exact degree of S_j, and put $a_j = E - e_j$ $(1 \leqq j \leqq p)$. We do (A) first. We apply Lemma 5 (p. 285) of [4] (which works just as well for $K = \mathbb{C}_p$) to the polynomials

$$L^{a_1} S_1, \ldots, L^{a_p} S_p \tag{2.1}$$

and the zero ideal. This yields a non-zero polynomial R_1 in $\mathscr{S}$ of degree E. Then $\mathfrak{J}_1 = (R_1)$ is contained in $\mathfrak{N}$, so $\mathfrak{J}_1^* \subseteq \mathfrak{N}^* = \mathfrak{N}$. In particular $\mathfrak{J}_1^* \neq \mathfrak{R}$ and so by Lemma 1 (p. 84) of [15] the rank of $\mathfrak{J}_1^*$ is at most 1. As $\mathfrak{J}_1 \subseteq \mathfrak{J}_1^*$, these two ideals must both have rank 1 and

$$\deg \mathfrak{J}_1^* \leqq \deg \mathfrak{J}_1 = E.$$

This proves (A).

To establish (B), we apply Lemma 5 of [4] to the polynomials (2.1) and the prime components $\mathfrak{P}$ of $\mathfrak{J}_i^*$ of rank i. As each such $\mathfrak{P}$ is contained in at least one of $\mathfrak{M}_1, \ldots, \mathfrak{M}_m$, the linear form L is not in $\mathfrak{P}$, and so not all the polynomials (2.1) lie in $\mathfrak{P}$. Thus we get R_{i+1} in $\mathscr{S}$ of degree E not lying in any prime component of $\mathfrak{J}_i^*$ of rank i. However, by Lemma 1 of [15] the ideal $\mathfrak{J}_i^*$ is unmixed of rank i. Hence R_{i+1} does not lie in any prime component of $\mathfrak{J}_i^*$ whatsoever. So by Lemma 3 (p. 281) of [4] the ideal $(\mathfrak{J}_i^*, R_{i+1})$ has rank $i+1$ and degree $E \deg \mathfrak{J}_i^*$. But now $\mathfrak{J}_{i+1} = (\mathfrak{J}_i, R_{i+1}) \subseteq \mathfrak{N}$, so $\mathfrak{J}_{i+1}^* \subseteq \mathfrak{N}^* = \mathfrak{N}$; whence $\mathfrak{J}_{i+1}^* \neq \mathfrak{R}$ and so from Lemma 1 of [15] the rank of $\mathfrak{J}_{i+1}^*$ is at most $i+1$. As $\mathfrak{J}_{i+1}^* \supseteq (\mathfrak{J}_i^*, R_{i+1})$ these two ideals must both have the same rank $i+1$ and

$$\deg \mathfrak{J}_{i+1}^* \leqq \deg (\mathfrak{J}_i^*, R_{i+1}) = E \deg \mathfrak{J}_i^*.$$

This establishes (B) and completes the proof of the Lemma.

3. The Inductive Steps

We now start on the proof of Theorem II. Thus for integers r, D with $0 \leqq r \leqq n$ and $D \geqq 1$ let $\mathfrak{J}_0$ be an ideal of $\mathfrak{R}$ of rank r generated by homogeneous polynomials of degrees at most D, and for integers $k \geqq 1$ and $D_1 \geqq D, \ldots, D_k \geqq D$ let $P_1, \ldots, P_k$ be homogeneous polynomials of degrees at most $D_1, \ldots, D_k$ respectively. Suppose $\mathfrak{J} = (\mathfrak{J}_0, P_1, \ldots, P_k)$ is a proper ideal of $\mathfrak{R}$ and that for some s with $1 \leqq s \leqq n$ it has at least one isolated primary component $\mathfrak{Q}$ of rank s. As $\mathfrak{J}_0 \subseteq \mathfrak{J} \subseteq \mathfrak{Q}$, it is clear that $s \geqq r$. Also we may assume $s \leqq r+k$. For the result with $s > r+k$ then follows simply by adjoining $s-r-k$ zero polynomials to $P_1, \ldots, P_k$ and estimating their degrees by D. Thus we have

$$t = s - r \leqq k.$$

We may evidently also suppose without loss of generality that

$$D_1 \geqq \ldots \geqq D_k \geqq D.$$

Let $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_m$ be the isolated primary components of $\mathfrak{J}$ of rank s, so that $m \geqq 1$, and let $\mathfrak{P}_1, \ldots, \mathfrak{P}_m$ be the associated prime ideals. As these primes are minimal, we can choose for each i with $1 \leqq i \leqq m$ a maximal ideal $\mathfrak{M}_i$ of $\mathfrak{R}$ containing $\mathfrak{P}_i$ but not any other isolated prime component of $\mathfrak{J}$. This corresponds to choosing a point on the variety of $\mathfrak{P}_i$ in projective space $\mathbb{P}_n$ not lying on any other isolated component of the variety of $\mathfrak{J}$, and then fixing coordinates on the underlying affine space K^{n+1}. Thus $\mathfrak{M}_1, \ldots, \mathfrak{M}_m$ are non-homogeneous. As in the previous section let $\mathfrak{N}$ denote their intersection and for an ideal $\mathfrak{J}$ let $\mathfrak{J}^*$ denote the corresponding contracted extension; also fix a linear form L in $\mathfrak{R}$ not in any of $\mathfrak{M}_1, \ldots, \mathfrak{M}_m$. Note that

$$\mathfrak{J}_0 \subseteq \mathfrak{J} \subseteq \mathfrak{J}^* = \mathfrak{Q}_1 \cap \ldots \cap \mathfrak{Q}_m \subseteq \mathfrak{N} \tag{2.2}$$

and in particular $\mathfrak{J}^*$ is unmixed of rank s.

Next, we can assume that for some $h \geqq 1$ we have $\mathfrak{J}_0 = (Q_1, \ldots, Q_h)$ for homogeneous polynomials $Q_1, \ldots, Q_h$ of degrees at most D. We shall carry out two rather similar inductive procedures. The ultimate aim of these is to find homogeneous polynomials $R_1, \ldots, R_s$ of $\mathfrak{J}$ such that the ideal $(R_1, \ldots, R_s)^*$ has rank s and degree at most $D^r M_{s-r}(D_1, \ldots, D_k)$. The Theorem will then follow rapidly in section 4.

Our first inductive procedure takes place only if $r \geqq 1$. We shall choose homogeneous polynomials $R_1, \ldots, R_r$ of the form

$$R_i = C_{i1} Q_1 + \ldots + C_{ih} Q_h \qquad (1 \leqq i \leqq r)$$

such that if $\mathfrak{J}_i = (R_1, \ldots, R_i)$ the ideal $\mathfrak{J}_i^*$ has rank i and degree at most D^i $(1 \leqq i \leqq r)$. To do this we shall use the Lemma with $S_1, \ldots, S_p$ as

$$Q_1, \ldots, Q_h \tag{2.3}$$

and $E = D$. Note that by (2.2) these all lie in $\mathfrak{N}$. For $i = 1$ we use part (A). Note that as $\mathfrak{J}_0$ has rank $r \geqq 1$ these polynomials are not all zero. Hence if $r = 1$ the first inductive procedure is already completed. Otherwise, suppose $r > 1$ and that $R_1, \ldots, R_i$ have been constructed for some i with $1 \leqq i < r$. Then we apply part (B) of the Lemma with $S_1, \ldots, S_p$ as in (2.3) again with $E = D$. Note that these polynomials cannot lie in any prime component $\mathfrak{P}$ of $\mathfrak{J}_i^*$ of rank i, for otherwise $\mathfrak{J}_0 \subseteq \mathfrak{P}$ and this would give a contradiction on considering ranks. The resulting ideal $\mathfrak{J}_{i+1}^*$ therefore has rank $i+1$ and degree at most $ED^i = D^{i+1}$. This completes the first inductive construction.

Thus if $r \geqq 1$ and $s = r$ we have also finished the construction of $R_1, \ldots, R_s$, as $M_{s-r}(D_1, \ldots, D_k) = 1$. So henceforth we shall assume either $r \geqq 1$ and $s > r$, or $r = 0$ (and so necessarily $s > r$ in this case too).

We shall now choose homogeneous polynomials $R_{r+1}, \ldots, R_s$ of the form

$$R_{r+j} = A_{jj} P_j + \ldots + A_{jk} P_k + B_{j1} Q_1 + \ldots + B_{jh} Q_h \qquad (1 \leqq j \leqq s-r), \tag{2.4}$$

with A_{jj} a power of L, such that if $\mathfrak{J}_{r+j} = (R_1, \ldots, R_{r+j})$ the ideal $\mathfrak{J}_{r+j}^*$ has rank $r+j$ and degree at most $D^r D_1 \ldots D_j$. We start with R_{r+1}. We use the Lemma with $S_1, \ldots, S_p$ as

$$P_1, \ldots, P_k, \quad Q_1, \ldots, Q_h \tag{2.5}$$

and $E = D_1$. Again these polynomials lie in $\mathfrak{N}$, by (2.2). If $r = 0$ we use part (A). This gives R_{r+1} of the required form (2.4) immediately, with $\mathfrak{J}_{r+1}^*$ of rank $r+1$ and of degree at most $E = D^r D_1$. If $r \geqq 1$, however, we use part (B) of the Lemma for $i = r$, with $S_1, \ldots, S_p$ as in (2.5) again and $E = D_1$. Now we must verify that not all of the polynomials (2.5) can lie in any prime component $\mathfrak{P}$ of $\mathfrak{J}_r^*$ of rank r. But if they did, we should have $\mathfrak{J} \subseteq \mathfrak{P}$, so $\mathfrak{J}^* \subseteq \mathfrak{P}^* = \mathfrak{P}$, which is again impossible because of ranks. So we get R_{r+1} of the required form (2.4) with $\mathfrak{J}_{r+1}^*$ of rank $r+1$ and of degree at most $ED^r = D^r D_1$.

Thus if $s = r+1$ we have once more found the required polynomials $R_1, \ldots, R_s$. Otherwise suppose $s > r+1$ and that $R_1, \ldots, R_{r+j}$ have been found for some j with $1 \leqq j < s-r$. We apply part (B) of the Lemma for $i = r+j$ with $S_1, \ldots, S_p$ as

$$P_{j+1}, \ldots, P_k, \quad Q_1, \ldots, Q_h \tag{2.6}$$

and $E = D_{j+1}$. We have seen that these polynomials are in $\mathfrak{R}$, by (2.2); thus it suffices to check that they cannot all lie in any prime component $\mathfrak{P}$ of $\mathfrak{J}^*_{r+j}$ of rank $r+j$. But if they did, we could use the triangular nature of (2.4) to deduce by induction on l that the polynomials

$$P_{j+1-l}, \ldots, P_k, \quad Q_1, \ldots, Q_h \tag{2.7}$$

would lie in $\mathfrak{P}$ for $0 \leqq l \leqq j$. For this is already known for $l=0$; so suppose it has been shown that the polynomials (2.7) lie in $\mathfrak{P}$ for some l with $0 \leqq l < j$. It follows from (2.4) that

$$R_{r+j-l} - A_{j-l,\, j-l} P_{j-l}$$

lies in $\mathfrak{P}$; but also R_{r+j-l} lies in $(R_1, \ldots, R_{r+j})^* = \mathfrak{J}^*_{r+j} \subseteq \mathfrak{P}$. As $A_{j-l,\, j-l}$ is a power of L, we deduce that P_{j-l} lies in $\mathfrak{P}$. This does the induction step. Thus the polynomials (2.7) lie in $\mathfrak{P}$ for $l=j$, so once again we get the contradiction from $\mathfrak{J}^* \subseteq \mathfrak{P}^* = \mathfrak{P}$ on comparing ranks.

Thus indeed the Lemma yields R_{r+j+1} of the required form (2.4) with $\mathfrak{J}^*_{r+j+1}$ of rank $r+j+1$ and of degree at most $ED^r D_1 \ldots D_j = D^r D_1 \ldots D_{j+1}$. This completes the second inductive procedure. Hence in the case $s > r$ as well we have found the required polynomials $R_1, \ldots, E_s$, as $M_{s-r}(D_1, \ldots, D_k) = D_1 \ldots D_{s-r}$.

4. Proof of Theorem 2

Thus in all cases we have succeeded in finding homogeneous polynomials $R_1, \ldots, R_s$ of $\mathfrak{J}$ such that if $\mathfrak{J}_s = (R_1, \ldots, R_s)$ then $\mathfrak{J}^*_s$ is of rank s and of degree at most $D^r D_1 \ldots D_{s-r} = D^r M_{s-r}(D_1, \ldots, D_k)$. Now recall the isolated primary components $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_m$ of $\mathfrak{J}$ and their associated prime ideals $\mathfrak{P}_1, \ldots, \mathfrak{P}_m$ of rank s. Then for each i with $1 \leqq i \leqq m$ we have $\mathfrak{J}_s \subseteq \mathfrak{J} \subseteq \mathfrak{P}_i$ and so $\mathfrak{J}^*_s \subseteq \mathfrak{P}^*_i = \mathfrak{P}_i$. Hence $\mathfrak{P}_i$ is a prime component of $\mathfrak{J}^*_s$, which must be isolated because $\mathfrak{J}^*_s$ is unmixed. Let $\mathfrak{Q}'_i$ be the corresponding primary component of $\mathfrak{J}^*_s$. Then as $\mathfrak{J}^*_s \subseteq \mathfrak{J}^*$ and both these ideals are unmixed of rank s, we deduce that $\mathfrak{Q}'_i \subseteq \mathfrak{Q}_i$ $(1 \leqq i \leqq m)$ on taking the contracted extension with respect to the complement of $\mathfrak{P}_i$ in $\mathfrak{R}$.

Finally degree theory (see for example [4], Section 3) for $\mathfrak{J}^*_s$ shows that

$$\sum_{i=1}^{m} \deg \mathfrak{Q}'_i \leqq \deg \mathfrak{J}^*_s \leqq D^r M_{s-r}(D_1, \ldots, D_k),$$

and since

$$\deg \mathfrak{Q}_i \leqq \deg \mathfrak{Q}'_i \quad (1 \leqq i \leqq m)$$

the estimate of Theorem II follows immediately.

To deduce the Corollary we use the standard operators of homogenisation and de-homogenisation [26] (p. 182). If $\mathfrak{J} = ({}^hQ_1, \ldots, {}^hQ_k)$ in the notation of the Corollary, then $(0) \neq \mathfrak{J} \neq \mathfrak{R}$, otherwise ${}^a\mathfrak{J} = (Q_1, \ldots, Q_k)$ would be either (0) or ${}^a\mathfrak{R}$, contrary to hypothesis. Thus let

$$\mathfrak{I} = \mathfrak{Q}_1 \cap \ldots \cap \mathfrak{Q}_m \tag{2.8}$$

be an irredundant decomposition into homogeneous primary ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_m$ for some $m \geqq 1$, with associated prime ideals $\mathfrak{P}_1, \ldots, \mathfrak{P}_m$. If each of these prime ideals contains X_0 then $\mathfrak{I}$ would contain a power of X_0 and so again we should obtain the contradiction ${}^a\mathfrak{I} = {}^a\mathfrak{R}$. Hence we can assume that there exists l with $1 \leqq l \leqq m$ such that $\mathfrak{P}_1, \ldots, \mathfrak{P}_l$ do not contain X_0 while the remaining primes (if any) do contain X_0. Now

$${}^a\mathfrak{I} = {}^a\mathfrak{Q}_1 \cap \ldots \cap {}^a\mathfrak{Q}_l \tag{2.9}$$

is an irredundant decomposition into primary ideals ${}^a\mathfrak{Q}_1, \ldots, {}^a\mathfrak{Q}_l$. Hence we can assume that there exists h with $1 \leqq h \leqq l$ such that only the ideals ${}^a\mathfrak{Q}_1, \ldots, {}^a\mathfrak{Q}_h$ are isolated in (2.9) of rank s, while the remaining primary ideals in (2.9) (if any) are either embedded or of rank different from s. Since for each i with $1 \leqq i \leqq l$ the ideals $\mathfrak{Q}_i$, ${}^a\mathfrak{Q}_i$ have the same ranks, it follows that the ideals $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_h$ have rank s, and also, since $\mathfrak{P}_1, \ldots, \mathfrak{P}_h$ cannot contain primes already containing X_0, that $\mathfrak{Q}_1, \ldots, \mathfrak{Q}_h$ are isolated in (2.8).

Thus Theorem II applied to ${}^hQ_1, \ldots, {}^hQ_k$ with $\mathfrak{I}_0 = (0)$, $D = 1$, $r = 0$ gives

$$\sum_{i=1}^{h} \deg \mathfrak{Q}_i \leqq M_t(D_1, \ldots, D_k).$$

Now the Corollary follows on noting that

$$l({}^a\mathfrak{Q}_i) = l(\mathfrak{Q}_i) \leqq \deg \mathfrak{Q}_i \qquad (1 \leqq i \leqq l).$$

Chapter 3. On Algebraic Subgroups of Products of Elliptic Curves

1. Introduction

Let E be an elliptic curve defined over $\mathbb{C}$, and suppose that E has no complex multiplication. For $n \geqq 1$ denote by E^n the group variety obtained by taking the n-fold product of E with itself, and for $1 \leqq i \leqq n$ let π_i be the projection from E^n to its i-th factor. A well-known theorem of Kolchin [12] implies that if H is a proper algebraic subgroup of E^n, then there are integers $t_1, \ldots t_n$, not all zero, such that

$$t_1 \pi_1(h) + \ldots + t_n \pi_n(h) = 0$$

for all h in H.

Now as it stands, Kolchin's proof gives no way of estimating the integers $t_1, \ldots, t_n$; indeed, there is no information about H in terms of which such estimates could be expressed. In this chapter we shall establish a quantitative version of Kolchin's theorem for E^n in the following sense. We suppose E^n embedded in projective space and that there exists a homogeneous polynomial P vanishing on H but not on all of E^n. Then we shall estimate $t_1, \ldots, t_n$ solely in terms of the degree of P. This kind of result does not seem to be a straightforward extension of Kolchin's theorem, and it will be seen that our proof differs from that of Kolchin in several respects.

Now we describe the embedding of E^n in projective space. To avoid problems with poles, we follow the construction of [17] (p. 511). Let $\mathscr{L}$ be a lattice in $\mathbb{C}$ such that E is analytically isomorphic to the quotient $\mathbb{C}/\mathscr{L}$, and let $\wp(w)$ be the Weierstrass elliptic function associated with $\mathscr{L}$. If $\sigma(w)$ is the corresponding Weierstrass sigma function we pick any complex number w_0 and we write

$$h(w)=(\sigma(w+w_0))^3, \quad f(w)=h(w)\,\wp(w+w_0), \quad g(w)=h(w)\,\wp'(w+w_0).$$

We define a map ψ from $\mathbb{C}/\mathscr{L}$ to projective space $\mathbb{P}_2$ by associating to w the point $\psi(w)$ with projective coordinates

$$f(w), \quad g(w), \quad h(w). \tag{3.1}$$

This gives an embedding of E into $\mathbb{P}_2$. In addition we assume that ψ is a homomorphism of groups; in other words, that the group law on E is obtained from the group law on $\mathbb{C}/\mathscr{L}$ via the map ψ. Note that this is not the standard projective realisation of E, unless w_0 lies in the lattice $\mathscr{L}$.

Next, for $1 \leqq i \leqq n$ let x_i, y_i, z_i be projective coordinates for the i-th copy of E, and write $N=3^n-1$. By using projective coordinates $X_0, \ldots, X_N$ given by the products $l_1 \ldots l_n$, where each l_i is one of x_i, y_i, z_i, $(1 \leqq i \leqq n)$, we obtain the standard Segre embedding of $(\mathbb{P}_2)^n$ into $\mathbb{P}_N$. This gives the required embedding of E^n into $\mathbb{P}_N$.

As in Chapter 1, we shall say that an algebraic subset S of E^n can be defined in E^n by the homogeneous polynomials $P_1, \ldots, P_k$ in $\mathfrak{R}=\mathbb{C}[X_0, \ldots, X_N]$ if the set of common zeroes of $P_1, \ldots, P_k$ in $\mathbb{P}_N$ meets E^n precisely on S. Also for $\tau=(t_1, \ldots, t_n)$ in $\mathbb{Z}^n$ we write

$$|\tau|=\max(|t_1|, \ldots, |t_n|),$$

and for g in E^n we put

$$\tau(g)=t_1\pi_1(g)+\ldots+t_n\pi_n(g).$$

We can now state the main result of this chapter.

Theorem III. *For integers r, D with $1 \leqq r \leqq n$ and $D \geqq 1$ let S be an algebraic subset of E^n, of dimension at most $n-r$, that is defined by homogeneous polynomials of degrees at most D. Suppose S contains an algebraic subgroup H of G. Then there exist elements $\tau_1, \ldots, \tau_r$ of $\mathbb{Z}^n$, linearly independent over $\mathbb{Z}$, with*

$$|\tau_i| \leqq 2^{5N} D^{2r/(r+1-i)} \quad (1 \leqq i \leqq r)$$

such that

$$\tau_i(h)=0 \quad (1 \leqq i \leqq r)$$

for all h in H.

The following corollary avoids the language of algebraic groups; for simplicity we give only the special case corresponding to the choice $r=1$ in Theorem III.

Corollary. *Let W be an additive subgroup of $\mathbb{C}^n$ such that for each $(w_1, \ldots, w_n)$ in W none of the numbers $w_i + w_0$ $(1 \leqq i \leqq n)$ lies in $\mathscr{L}$. Suppose for some integer $D \geqq 1$ there is a non-zero polynomial P, of degree at most D in each variable, such that*

$$P(\wp(w_1+w_0), \ldots, \wp(w_n+w_0)) = 0$$

for all $(w_1, \ldots, w_n)$ in W. Then there are integers $t_1, \ldots, t_n$, not all zero, of absolute values at most $2^{5N} D^2$, such that

$$t_1 w_1 + \ldots + t_n w_n \equiv 0 (\bmod \mathscr{L})$$

for all $(w_1, \ldots, w_n)$ in W.

At first sight it might be supposed that the true order of magnitude of $t_1, \ldots, t_n$, at least in the Corollary, should be $D^{1/2}$, since the multiplication formula for $\wp(tw)$ is of degree t^2 in $\wp(w)$. However, the order of magnitude D^2 given by the Corollary cannot be improved without further hypotheses. For example, for $n=2$ take a positive integer T, not a square, and let W_2 be the additive group of $\mathbb{C}^2$ generated by $(\omega_1/T, \omega_2/T)$, where ω_1 and ω_2 are fixed primitive generators of $\mathscr{L}$. Since $D=[T^{1/2}]$ satisfies $(D+1)^2 > T$, we can find a non-zero polynomial P, of degree at most $D < T^{1/2}$ in each variable, such that

$$P(\wp(w_0 + t\omega_1/T), \wp(w_0 + t\omega_2/T)) = 0 \quad (0 \leqq t < T).$$

The hypotheses of the Corollary are now satisfied for $W = W_2$, by periodicity. On the other hand, if t_1 and t_2 are integers with

$$t_1 \omega_1/T + t_2 \omega_2/T \equiv 0 (\bmod \mathscr{L}),$$

then clearly T divides both t_1 and t_2. Hence if $\max(|t_1|, |t_2|)$ is non-zero, it is at least $T > D^2$. So for $n=2$ the Corollary would be false with $2^{5N} D^2$ replaced by D^2.

In fact this last statement is true for any $n > 2$, as we can easily see by considering the direct sum of W_2 with $n-2$ copies of $\mathbb{C}$. However, for $n=1$ a simple direct argument shows that the upper bound $2^{5N} D^2$ in Theorem III (for $i=1$) can be replaced by $3D$.

We note here that similar results can be obtained when the elliptic curve E has complex multiplication, although in this case completely different proofs are available using diophantine approximation. It suffices to replace $t_1, \ldots, t_n$ by elements of the ring of endomorphisms of E, as in [12]. The expression D^2 occurring in the estimates for their absolute values can then be improved to one of order D, and as above this is essentially best possible. Furthermore, the methods of this chapter work in the abelian case, when the underlying group G is A^n for a simple abelian variety A of dimension $g \geqq 1$. For example, if the ring of endomorphisms of A is trivial, then D^2 should be replaced by a term of order D^{2g^2}, and this too, perhaps surprisingly, is best possible.

Finally one can treat in an analogous way the exponential case, when $G = \mathbb{G}_m^n$ for the multiplicative group $\mathbb{G}_m$ of non-zero complex numbers; here, however, the results are much easier and they may be proved directly by means of repeated use of the Vandermonde determinant (see for example [15] p. 92).

2. Preliminaries

Here we determine the constants a, b of Chapter 1 associated with the group variety $G=E^n$ embedded in $\mathbb{P}_N$ as above. To start with, it is easy to write down the equations defining E^n in $\mathbb{P}_N$. First of all, it is well-known that $(\mathbb{P}_2)^n$ is defined in $\mathbb{P}_N$ by homogeneous polynomials of degree 2. Secondly, if g_2, g_3 are the invariants of the lattice $\mathscr{L}$, the equations

$$y_i^2 z_i = 4x_i^3 - g_2 x_i z_i^2 - g_3 z_i^3 \qquad (1 \leqq i \leqq n)$$

give rise to homogeneous polynomials of degree 3 which define E^n in $(\mathbb{P}_2)^n$. Hence E^n is defined in $\mathbb{P}_N$ by homogeneous polynomials of degrees at most 3, and we can take $b=3$ in Chapter 1.

It is not so easy to write down a complete set of addition laws on E^n (even if $w_0=0$). Fortunately we need only consider translation formulae with respect to finitely generated subgroups Γ of E^n. We shall show that we can take $a=2$. Note that the calculations of Section 8 of [17] do not quite suffice for this, as they apply only when Γ lies in the affine part of E^n defined by $z_1 \dots z_n \neq 0$. The following result provides the necessary extension of Lemma 10 (p. 511) of [17].

Lemma 1. *Let p be a point on E in $\mathbb{P}_2$, and let A be a countable subset of E in $\mathbb{P}_2$. Then there is a Zariski open subset $\mathcal{O}$ of E containing A, and homogeneous polynomials F, G, H of degree 2, such that for all g in $\mathcal{O}$ with projective coordinates ξ, η, ζ the numbers*

$$F(\xi, \eta, \zeta), \qquad G(\xi, \eta, \zeta), \qquad H(\xi, \eta, \zeta)$$

are projective coordinates of $g+p$.

Proof. We can write $p=\psi(w_1)$ and $A=\psi(U)$ for a complex number w_1 and a countable subset U of $\mathbb{C}$. If w_1 is in $\mathscr{L}$ then $p=0$ and the lemma is obvious with

$$F(x, y, z)=x, \qquad G(x, y, z)=y, \qquad H(x, y, z)=z.$$

Henceforth we shall assume that w_1 is not in $\mathscr{L}$. We fix a complex number w_2 such that none of the expressions

$$\begin{gathered} w_2, \quad w_2+w_1, \quad w_2-w_0+u, \quad 2w_2-3w_1, \quad 2w_2-4w_1, \\ 2w_2+w_0-3w_1+u, \quad 3w_2-3w_1 \end{gathered} \tag{3.2}$$

lie in $\mathscr{L}$ for any u in U. Then we determine complex coefficients $\alpha_1, \dots, \alpha_6$, not all zero, such that the function

$$\begin{aligned} \phi(w) = {} & \alpha_1 + \alpha_2 \wp(w) + \alpha_3 \wp'(w) + \alpha_4 (\wp(w))^2 \\ & + \alpha_5 \wp(w) \wp'(w) + \alpha_6 (\wp'(w))^2 \end{aligned} \tag{3.3}$$

has at least a triple zero at $w=-w_1$ and at least a double zero at $w=w_2$. Since w_2+w_1 is not in $\mathscr{L}$, the order of $\phi(w)$ in the usual sense is either 5 or 6. But as $2w_2-3w_1$ is not in $\mathscr{L}$, well-known results on the sums of zeroes and poles of elliptic functions show that the order cannot be 5; it is therefore 6, and in particular $\alpha_6 \neq 0$.

Now the functions $\phi(w)\,\wp(w+w_1)$, $\phi(w)\,\wp'(w+w_1)$ are analytic at $w=-w_1$, and it is easily seen that they must also have the form (3.3). Thus, on replacing w by $w+w_0$ and multiplying throughout by $(\sigma(w+w_0))^6$, we obtain homogeneous polynomials F, G, H such that

$$\begin{aligned} F(f(w), g(w), h(w)) &= (\sigma(w+w_0))^6\, \phi(w+w_0)\, \wp(w+w_0+w_1) \\ G(f(w), g(w), h(w)) &= (\sigma(w+w_0))^6\, \phi(w+w_0)\, \wp'(w+w_0+w_1) \\ H(f(w), g(w), h(w)) &= (\sigma(w+w_0))^6\, \phi(w+w_0). \end{aligned} \tag{3.4}$$

Since $\alpha_6 \neq 0$, these polynomials have degrees exactly 2.

Now if $g=\psi(w)$ for some complex number w for which not all the expressions on the left-hand side of (3.4) are zero, it is clear that these expressions are projective coordinates of the sum $g+p=\psi(w+w_1)$ on E. Therefore to complete the proof of Lemma 1 it suffices to verify that for any u in U not all these expressions vanish at $w=u$. For this we have to distinguish three cases.

Firstly, suppose that $u+w_0$ is in $\mathscr{L}$. Then by making $w \to u$ in the third equation of (3.4) we find that

$$H(f(u), g(u), h(u)) = 4\alpha_6 \neq 0.$$

Hence we may from now on assume that $u+w_0$ is not in $\mathscr{L}$. Consider the zeroes of $\phi(w)$. There are 6 zeroes $(\mathrm{mod}\,\mathscr{L})$ counted with multiplicity, whose sum $(\mathrm{mod}\,\mathscr{L})$ must be zero. Since $3w_2-3w_1$ and $2w_2-4w_1$ are not in $\mathscr{L}$, it follows that there is exactly one simple zero w_3, not congruent to w_2 or $-w_1$ $(\mathrm{mod}\,\mathscr{L})$, and that

$$w_3+2w_2-3w_1 \equiv 0 \pmod{\mathscr{L}}.$$

In particular $\phi(w)$ has an exact triple zero at $w=-w_1$, so that

$$\beta = \phi^{(3)}(-w_1) \neq 0.$$

For the second case we suppose that $u+w_0+w_1$ is in $\mathscr{L}$. Then by making $w \to u$ in the second equation of (3.4) we deduce after a short calculation that

$$G(f(u), g(u), h(u)) = -(\sigma(u+w_0))^6\, \phi^{(3)}(u+w_0) = -\beta(\sigma(u+w_0))^6 \neq 0,$$

since $u+w_0 \equiv -w_1$ $(\mathrm{mod}\,\mathscr{L})$ and $-w_1$ is not in $\mathscr{L}$.

Lastly suppose that neither $u+w_0$ nor $u+w_0+w_1$ are in $\mathscr{L}$. Then we have

$$H(f(u), g(u), h(u)) = (\sigma(u+w_0))^6\, \phi(u+w_0) \neq 0,$$

because $u+w_0$ is not congruent $(\mathrm{mod}\,\mathscr{L})$ to any of the zeroes of $\phi(w)$ determined above. This completes the proof of Lemma 1.

Now let Γ be any finitely generated subgroup of E^n, and let γ be any element of Γ. For each i with $1 \leqq i \leqq n$ we apply Lemma 1 to the factor $\pi_i(E^n)$ with $p=\pi_i(\gamma)$ and $A=\pi_i(\Gamma)$. The resulting formulae can then be combined in the usual way to give a collection of homogeneous polynomials of degree 2 defining translation by γ on an open set of E^n containing Γ. Thus indeed we can take $a=2$ in Chapter 1.

3. Subgroups of E^n

It seems necessary for the proof of Theorem III to give an exact description of all algebraic subgroups of E^n, in particular taking into account questions of connectivity. This we do in the present section. A basic role is played by the following two dual definitions. For a subset H of E^n let $\mathscr{A}(H)$ denote the subgroup of $\mathbb{Z}^n$ consisting of all τ such that $\tau(h)=0$ for all h in H. Similarly for a subset A of $\mathbb{Z}^n$ let $\mathscr{H}(A)$ be the algebraic subgroup of E^n consisting of all h such that $\tau(h)=0$ for all τ in A.

Lemma 2. *Let r be an integer with $1 \leqq r \leqq n$. If H is an algebraic subgroup of E^n of dimension $n-r$, then $\mathscr{A}(H)$ is a subgroup of $\mathbb{Z}^n$ of rank r. Conversely, if A is a subgroup of $\mathbb{Z}^n$ of rank r, then $\mathscr{H}(A)$ is an algebraic subgroup of E^n of dimension $n-r$.*

Proof. Suppose first that A is a subgroup of $\mathbb{Z}^n$ of rank r. If $r=n$ clearly $\mathscr{H}(A)$ is a finite set and therefore of zero dimension. Otherwise, if $r<n$, we can suppose without loss of generality that $\tau_i=(t_{i1}, \ldots, t_{in})$ $(1 \leqq i \leqq r)$ are elements of a basis of A with

$$\det_{1 \leqq i,\, j \leqq r} (t_{ij}) \neq 0.$$

Fix a point h_0 of $\mathscr{H}(A)$ with $\pi_i(h_0)=\psi(w_{i0})$ $(1 \leqq i \leqq n)$ in the notation (3.1). Then for any other point h of $\mathscr{H}(A)$ with $\pi_i(h)=\psi(w_i)$ $(1 \leqq i \leqq n)$ the numbers

$$t_{i1}(w_1-w_{10})+\ldots+t_{in}(w_n-w_{n0}) \qquad (1 \leqq i \leqq r) \tag{3.5}$$

are in the lattice $\mathscr{L}$. Since $\mathscr{L}$ is discrete in $\mathbb{C}$, it follows that the expressions (3.5) vanish if h is sufficiently close to h_0. In this case all the differences w_i-w_{i0} $(1 \leqq i \leqq n)$ can be written as linear forms in $w_{r+1}-w_{r+1,0}, \ldots, w_n-w_{n0}$. Also, for all sufficiently small $\varepsilon_{r+1}, \ldots, \varepsilon_n$ there is a point h on $\mathscr{H}(A)$ with $\pi_i(h)=\psi(w_i)$ $(1 \leqq i \leqq n)$ and $w_i-w_{i0}=\varepsilon_i$ $(r+1 \leqq i \leqq n)$. It follows that $w_{r+1}-w_{r+1,0}, \ldots, w_n-w_{n0}$ are a system of local parameters at h_0 on $\mathscr{H}(A)$. Consequently $\mathscr{H}(A)$ has dimension $n-r$. This proves the second claim of the lemma.

To establish the first claim, let H be an algebraic subgroup of E^n of dimension $n-r$. By Lemma 11 (p. 512) of [17] the group $\mathscr{A}(H)$ has rank at least r. Thus by what has just been proved the algebraic subgroup $\mathscr{H}(\mathscr{A}(H))$ has dimension at most $n-r$. But this clearly contains H, so its dimension must be exactly $n-r$. Again by what we proved above, we conclude that $\mathscr{A}(H)$ has rank exactly r. This completes the proof of Lemma 2.

Since the lattice $\mathscr{L}$ is of rank 2 over $\mathbb{Z}$, it is easily seen that in general we do not have $H=\mathscr{H}(\mathscr{A}(H))$ (even though the analogous assertion in the exponential case is true). In order to get a precise description of an arbitrary algebraic subgroup H of E^n, it is necessary to introduce certain translations of $\mathscr{H}(A)$. Let E_∞ denote the group of torsion points of E, and for a subgroup A of $\mathbb{Z}^n$ denote by $\operatorname{Hom}(A, E_\infty)$ the additive group of all homomorphisms from A to E_∞. For an element ρ in $\operatorname{Hom}(A, E_\infty)$ let $\mathscr{H}_\rho(A)$ be the set of h in E^n such that

$$\tau(h)=\rho(\tau)$$

for all τ in A. This is a subgroup of E^n if and only if ρ is the zero element of $\mathrm{Hom}(A, E_\infty)$, when it coincides with $\mathscr{H}(A)$; otherwise, if $\rho \neq 0$, it is a translation of $\mathscr{H}(A)$ by an element not in $\mathscr{H}(A)$.

We say that a subgroup A of $\mathbb{Z}^n$ is primitive if it is maximal among all subgroups of $\mathbb{Z}^n$ with the same rank.

Lemma 3. *For an integer r with $1 \leqq r \leqq n$ suppose H is an algebraic subgroup of E^n of dimension $n-r$. Then there exists a primitive subgroup A_0 of $\mathbb{Z}^n$ of rank r, together with a finite subgroup R of* $\mathrm{Hom}(A, E_\infty)$, *such that*

$$H=\bigcup \mathscr{H}_\rho(A_0), \tag{3.6}$$

where the union is taken over all ρ in R. Furthermore these properties uniquely define A_0 and R in terms of H.

Proof. Let H_0 be the connected component of H containing the origin 0 of E^n (see for example [2] p. 88). Then H_0 is an algebraic subgroup of E^n also of dimension $n-r$. We start by proving that $A_0=\mathscr{A}(H_0)$ is primitive. It suffices to verify that if $k\tau$ is in A_0 for some positive integer k, then τ is in A_0. But we then have $0=(k\tau)(H_0)=k(\tau(H_0))$, so $\tau(H_0)$ is contained in E_k the group of torsion points on E whose order divides k. But $\tau(H_0)$ is connected, and therefore $\tau(H_0)=0$. Hence τ is in A_0, and so A_0 must indeed be primitive.

Next, as A_0 has rank r, the algebraic subgroup $H'_0=\mathscr{H}(A_0)$ has dimension $n-r$. We show now that H'_0 is connected. If $r=n$ then $A_0=\mathbb{Z}^n$ so $H'_0=0$; thus we may assume that $1 \leqq r < n$. Let $\tau_1, \ldots, \tau_r$ be elements of a basis of A_0. Since A_0 is primitive, it is well-known that we can find elements $\tau_{r+1}, \ldots, \tau_n$ of $\mathbb{Z}^n$ such that $\tau_1, \ldots, \tau_n$ together generate all of $\mathbb{Z}^n$ (see [5] p. 15). Consider the map Φ from H'_0 to E^{n-r} given by

$$\Phi(g)=(\tau_{r+1}(g), \ldots, \tau_n(g)).$$

Since already

$$\tau_1(g)=\ldots=\tau_r(g)=0$$

for all g in $H'_0=\mathscr{H}(A_0)$, it follows that Φ is injective. Consider now the image $\Phi(H'_0) \subseteq E^{n-r}$. This is an algebraic subgroup of E^{n-r} (see [2] p. 88). Assume for the moment that this is a proper subgroup of E^{n-r}; then Kolchin's Theorem [12] would imply the existence of integers $s_{r+1}, \ldots, s_n$, not all zero, such that

$$s_{r+1}\tau_{r+1}(g)+\ldots+s_n\tau_n(g)=0$$

for all g in H'_0. Thus $s_{r+1}\tau_{r+1}+\ldots+s_n\tau_n$ would lie in $\mathscr{A}(H'_0)$. But as $\tau_1, \ldots, \tau_r$ already lie in $A_0 \subseteq \mathscr{A}(\mathscr{H}(A_0))=\mathscr{A}(H'_0)$, this contradicts the fact that $\mathscr{A}(H'_0)$ has rank r. Hence indeed Φ is surjective, and so it defines an isomorphism between H'_0 and E^{n-r}. Since E^{n-r} is connected, it follows that H'_0 is also connected, as claimed.

Now we have

$$H'_0=\mathscr{H}(A_0)=\mathscr{H}(\mathscr{A}(H_0)) \supseteq H_0,$$

whence we conclude that $H'_0=H_0$.

The rest of the proof is easy. Let $k=[H:H_0]$ be the index of H_0 in H. For each τ in A_0 we have $\tau(H_0)=0$, and it follows that $\tau(H)\subseteq E_k$. Denote the quotient group H/H_0 by R. We define a map from R to $\operatorname{Hom}(A_0, E_k)$ by sending the coset $h+H_0$ to the element ρ of $\operatorname{Hom}(A_0, E_k)$ defined by

$$\rho(\tau)=\tau(h+h_0)$$

for all τ in A_0 and any h_0 in H_0. This is well-defined because $A_0=\mathscr{A}(H_0)$. It is also injective, because if $\tau(h+H_0)=0$ for all τ in A_0, then

$$h+H_0\subseteq\mathscr{H}(A_0)=H_0'=H_0,$$

and so $h+H_0$ is the zero coset of H/H_0.

Hence in this manner R may be identified with a finite subgroup of $\operatorname{Hom}(A_0, E_\infty)$. Finally for ρ in R let H_ρ be the corresponding coset of H_0 in H. Then clearly $H_\rho=\mathscr{H}_\rho(A_0)$, and this establishes the decomposition (3.6).

To prove uniqueness (which strictly speaking we could do without), assume that (3.6) holds for some primitive subgroup A_0 of $\mathbb{Z}^n$ of rank r and some finite subgroup R of $\operatorname{Hom}(A_0, E_\infty)$. Then it follows easily that $A=\mathscr{A}(H)$ contains kA_0 for some positive integer k. By Lemma 2 the rank of A is r, and this means that A_0 is determined as the unique primitive subgroup of $\mathbb{Z}^n$ of rank r which contains A. Finally it is clear that R is determined uniquely, as the sets $\mathscr{H}_\rho(A_0)$ are disjoint for different ρ in $\operatorname{Hom}(A_0, E_\infty)$. This finishes the proof of Lemma 3.

4. Estimates for Subgroups

We introduce here two numerical functions $\delta(H)$, $\kappa(H)$ for an algebraic subgroup H of E^n, and we establish some simple inequalities we shall need later.

Let A be a subgroup of $\mathbb{Z}^n$ of rank r, with $1\leqq r\leqq n$. We can define $\det(A)$ by choosing a basis of row vectors of A, writing down the corresponding integer matrix with r rows and n columns, and taking the maximum of the absolute values of all the determinants of the minors of order r. Since any two bases of A are related by a unimodular matrix of order r, this definition does not depend on the particular basis chosen. Furthermore, if B is a subgroup of A also of rank r, it is easily seen that

$$\det(B)=[A:B]\det(A). \tag{3.7}$$

Given an algebraic subgroup H of E^n of dimension $n-r$ for some r with $1\leqq r\leqq n$, we define

$$\delta(H)=\det(A_0), \qquad \kappa(H)=\operatorname{card}(R),$$

where A_0 and R are as in Lemma 3, and $\operatorname{card}(R)$ denotes the cardinality of R. The first of these invariants is related to the geometric degree of the connected component H_0 of H through the origin, and the second is exactly the total number of connected components of H. But we shall not make explicit use of these geometric interpretations in what follows.

Lemma 4. *For an integer r with $1 \leqq r \leqq n$ let A be a subgroup of $\mathbb{Z}^n$ of rank r. Then A contains elements $\tau_1, \ldots, \tau_r$, linearly independent over $\mathbb{Z}$, with*

$$|\tau_i| \leqq r(\det(A))^{1/(r+1-i)} \qquad (1 \leqq i \leqq r).$$

Proof. It is not difficult to deduce this from Lemma 8 (p. 501) of [17]. But to show the connection with the geometry of numbers we give here a less elementary proof depending on the theory of successive minima. Without loss of generality we can assume that A is generated by elements $\tau_i' = (t_{i1}', \ldots, t_{in}')$ $(1 \leqq i \leqq r)$ with

$$\det(A) = |\det_{1 \leqq i, j \leqq r}(t_{ij}')|. \tag{3.8}$$

Then the lattice Λ in $\mathbb{R}^r$ generated by the truncated vectors $\tau_i'' = (t_{i1}', \ldots, t_{ir}')$ $(1 \leqq i \leqq r)$ has determinant $\det(A)$. Let $\lambda_1, \ldots, \lambda_r$ be the successive minima of Λ with respect to the distance function

$$F(x) = \max(|x_1|, \ldots, |x_r|)$$

for $x = (x_1, \ldots, x_r)$ in $\mathbb{R}^r$ (see [5] p. 201). We deduce from Minkowski's Theorem V (p. 218) of [5] that

$$\lambda_1 \ldots \lambda_r \leqq \det(A).$$

But since Λ consists only of integral vectors, we have $\lambda_1 \geqq 1$. As $\lambda_1 \leqq \ldots \leqq \lambda_r$, it follows that

$$\lambda_i^{r+1-i} \leqq \lambda_1 \ldots \lambda_r \leqq \det(A) \qquad (1 \leqq i \leqq r),$$

and consequently

$$\lambda_i \leqq (\det(A))^{1/(r+1-i)} \qquad (1 \leqq i \leqq r). \tag{3.9}$$

Thus by Lemma 1 (p. 204) of [5] there exist linearly independent elements $\tau_1^*, \ldots, \tau_r^*$ of Λ with

$$F(\tau_i^*) = \lambda_i \qquad (1 \leqq i \leqq r),$$

and so we have

$$\tau_i^* = s_{i1}\tau_1'' + \ldots + s_{ir}\tau_r'' \qquad (1 \leqq i \leqq r)$$

for integers $s_{i1}, \ldots, s_{ir}$. We now proceed to verify that the elements

$$\tau_i = s_{i1}\tau_1' + \ldots + s_{ir}\tau_r' \qquad (1 \leqq i \leqq r)$$

satisfy the conditions of the present lemma.

Firstly, it is clear that $\tau_1, \ldots, \tau_r$ are linearly independent over $\mathbb{Z}$. Also, if we write $\tau_i = (t_{i1}, \ldots, t_{in})$ $(1 \leqq i \leqq r)$ we have from (3.9) and the above remarks

$$|t_{ij}| \leqq \lambda_i \leqq (\det(A))^{1/(r+1-i)} \qquad (1 \leqq i, j \leqq r) \tag{3.10}$$

If $r = n$ we are finished. Otherwise, if $r < n$, we have identities of the form

$$\Delta t_{ik} = \Delta_1 t_{i1} + \ldots + \Delta_r t_{ir} \qquad (1 \leqq i \leqq r,\ r+1 \leqq k \leqq n), \tag{3.11}$$

where $\Delta_1, \ldots, \Delta_r$ are signed determinants of certain minors of order r in the original matrix representing A, and $\Delta = \det(A)$ is the determinant (3.8). By hypothesis Δ is the largest in absolute value of all these determinants, and we deduce from (3.10) and (3.11) that

$$|t_{ik}| \leqq r(\det(A))^{1/(r+1-i)} \qquad (1 \leqq i \leqq r,\ r+1 \leqq k \leqq n).$$

This, together with (3.10), completes the proof of Lemma 4.

Lemma 5. *For an integer r with $1 \leqq r \leqq n$ let H be an algebraic subgroup of E^n of dimension $n-r$. Then*

$$\det(\mathscr{A}(H)) \leqq \delta(H)(\kappa(H))^2.$$

Proof. Let A_0 and R be as in Lemma 3, and choose elements $\tau_1, \ldots, \tau_r$ of a basis of A_0. For ρ in R let H_ρ be the set of h in E^n satisfying

$$\tau_i(h) = \rho(\tau_i) \qquad (1 \leqq i \leqq r). \tag{3.12}$$

Then $H_\rho = \mathscr{H}_\rho(A_0)$ in the notation of Lemma 3, and so H is the union of the H_ρ as ρ runs over all elements of R. Next choose elements of a basis of the period lattice $\mathscr{L}$ over $\mathbb{Z}$. These give in the obvious way an isomorphism ω from E_∞ to $(\mathbb{Q}/\mathbb{Z})^2$. Let $\omega = (\omega^{(1)}, \omega^{(2)})$, so that each $\omega^{(e)} (e=1,2)$ is a map from E_∞ to $\mathbb{Q}/\mathbb{Z}$. Then, for each e, as ρ runs over all elements of R, the elements

$$(\omega^{(e)}(\rho(\tau_1)), \ldots, \omega^{(e)}(\rho(\tau_r))) \tag{3.13}$$

run over a certain subgroup $S^{(e)}$ of $(\mathbb{Q}/\mathbb{Z})^r$, and the cardinality of $S^{(e)}$ is at most $\kappa(H)$, the cardinality of R.

Now let S be the subgroup of $(\mathbb{Q}/\mathbb{Z})^r$ generated by the elements of both $S^{(1)}$ and $S^{(2)}$. Then the cardinality of S satisfies

$$\operatorname{card}(S) \leqq (\kappa(H))^2. \tag{3.14}$$

Let B be the subgroup of $\mathbb{Z}^r$ consisting of all $\beta = (b_1, \ldots, b_r)$ such that

$$b_1 \sigma_1 + \ldots + b_r \sigma_r \equiv 0 \pmod{\mathbb{Z}} \tag{3.15}$$

for all $(\sigma_1, \ldots, \sigma_r)$ in S. By writing S as a direct sum of cyclic groups of orders $k_1, \ldots, k_m$, say, and appealing to Lemma 9 (p. 98) of [5], we see easily that B has finite index at most $k_1 \ldots k_m = \operatorname{card}(S)$ in $\mathbb{Z}^r$ (in fact a general theory of duality between $\mathbb{Z}^r$ and $(\mathbb{Q}/\mathbb{Z})^r$ shows that this index must be exactly $\operatorname{card}(S)$).

Next it follows by applying (3.15) to (3.13) that

$$b_1 \omega^{(e)}(\rho(\tau_1)) + \ldots + b_r \omega^{(e)}(\rho(\tau_r)) \equiv 0 \pmod{\mathbb{Z}}$$

for $e=1,2$ and all $(b_1, \ldots, b_r)$ in B. Consequently

$$b_1 \rho(\tau_1) + \ldots + b_r \rho(\tau_r) = 0 \tag{3.16}$$

in E_∞. Thus if we define A_0' as the subset of A_0 consisting of elements of the form

$$\tau' = b_1 \tau_1 + \ldots + b_r \tau_r$$

as $(b_1, \ldots, b_r)$ runs over all elements of B, we see from (3.12) and (3.15) that $A'_0 \subseteq \mathscr{A}(H)$. Furthermore it is clear that $[A_0 : A'_0] = [\mathbb{Z}^r : B]$, which does not exceed card (S). We conclude from (3.7) and (3.14) that

$$\det(\mathscr{A}(H)) \leqq \det(A'_0) = [A_0 : A'_0] \det(A_0) \leqq (\kappa(H))^2 \, \delta(H),$$

and this completes the proof of the lemma.

5. Degree Estimates

We start with the following simple observation.

Lemma 6. *Let A be a subgroup of $\mathbb{Z}^n$ of rank n. Then for any ρ in* $\mathrm{Hom}(A, E_\infty)$ *we have*

$$\operatorname{card}(\mathscr{H}_\rho(A)) = (\det(A))^2.$$

Proof. By Theorem I (p. 11) of [5] we can find generators

$$\tau_i = (t_{i1}, \ldots, t_{in}) \qquad (1 \leqq i \leqq n)$$

of A such that $t_{ij} = 0$ for $i < j$, and then we have

$$\det(A) = |t_{11} \ldots t_{nn}|.$$

Thus $\mathscr{H}_\rho(A)$ is the set of h in E^n satisfying

$$t_{i1}\,\pi_1(h) + \ldots + t_{ii}\,\pi_i(h) = \rho(\tau_i) \qquad (1 \leqq i \leqq n). \tag{3.17}$$

Now for any positive integer t and any f in E there are exactly t^2 points e on E with $te = f$. Hence by (3.17) with $i = 1$ the number of possibilities for $\pi_1(h)$ is exactly t_{11}^2. If $n = 1$ this completes the proof. Otherwise, if $n > 1$, we show by induction on r that the first r equations of (3.17) give exactly $t_{11}^2 \ldots t_{rr}^2$ possibilities for the element $(\pi_1(h), \ldots, \pi_r(h))$ on E^r. This is already done for $r = 1$. So assume it true for some r with $1 \leqq r < n$. Then (3.17) with $i = r+1$ shows that for each choice of $(\pi_1(h), \ldots, \pi_r(h))$ there are exactly $t_{r+1,r+1}^2$ possibilities for $\pi_{r+1}(h)$. This does the induction step, and as

$$t_{11}^2 \ldots t_{nn}^2 = (\det(A))^2$$

the case $r = n$ completes the proof of the lemma.

The next result supplies the crucial estimate for $\delta(H)$ and $\kappa(H)$ when the equations defining the algebraic subgroup H are known. Recall that E^n is embedded in $\mathbb{P}_N$ with $N = 3^n - 1$.

Lemma 7. *For integers r, D with $1 \leqq r \leqq n$ and $D \geqq 1$ let H be an algebraic subgroup of E^n, of dimension $n - r$, which is defined in E^n by homogeneous polynomials of degrees at most D. Then*

$$(\delta(H))^2 \, \kappa(H) \leqq 3^N D^r.$$

Proof. Without loss of generality, we may suppose that the associated primitive group A_0 of H given by Lemma 3 is generated by elements $\tau_i=(t_{i1}, \ldots, t_{in})$ $(1 \leqq i \leqq r)$ with

$$\delta(H)=\det(A_0)=|\det_{1\leqq i,j\leqq r}(t_{ij})|. \tag{3.18}$$

Let E^r denote the product of the first r factors of E^n, embedded in $\mathbb{P}_N$ in the obvious way. The inequality of the present lemma will come from comparing estimates for the cardinality of $H \cap E^r$.

We use the decomposition $H=\bigcup \mathscr{H}_\rho(A_0)$ of Lemma 3, and we start by showing that

$$\operatorname{card}(\mathscr{H}_\rho(A_0)\cap E^r)=(\delta(H))^2 \tag{3.19}$$

for all ρ in R. If $r=n$ this is obvious from Lemma 6 and the definition of $\delta(H)$. Otherwise, if $r<n$, write $\varepsilon_1, \ldots, \varepsilon_n$ for the standard basis elements of $\mathbb{Z}^n$ and let A be the subgroup of $\mathbb{Z}^n$ generated by A_0 and $\varepsilon_{r+1}, \ldots, \varepsilon_n$. By the non-vanishing of (3.18), A has rank n and

$$\det(A)=\det(A_0)=\delta(H). \tag{3.20}$$

Now we can extend each element ρ of R in $\operatorname{Hom}(A_0, E_\infty)$ to an element of $\operatorname{Hom}(A, E_\infty)$ by defining $\rho(\varepsilon_i)=0$ $(r+1\leqq i\leqq n)$. Then clearly $\mathscr{H}_\rho(A_0)\cap E^r=\mathscr{H}_\rho(A)$, and now (3.19) once again follows from Lemma 6 and (3.20).

Since the $\mathscr{H}_\rho(A_0)$ are disjoint, we conclude that

$$\operatorname{card}(H\cap E^r)=(\delta(H))^2\,\kappa(H). \tag{3.21}$$

On the other hand, the subgroup E^r is defined in E^n by equations of degree at most 1. Hence it is defined in $\mathbb{P}_N$ by equations of degree at most 3; that is, there exist homogeneous polynomials $Q_1, \ldots, Q_l$, of degrees at most 3, whose set of common zeroes in $\mathbb{P}_N$ is exactly E^r. Thus the homogeneous ideal $\mathfrak{J}_0=(Q_1, \ldots, Q_l)$ has rank $r_0=N-r$. Furthermore, by hypothesis there are homogeneous polynomials $P_1, \ldots, P_k$, of degrees at most D, such that H is the intersection of E^n with their set of common zeroes in $\mathbb{P}_N$. It follows that the ideal $\mathfrak{J}=(\mathfrak{J}_0, \ldots, P_k)$ has rank N, and its associated variety in $\mathbb{P}_N$ is $H\cap E^r$. Hence

$$\operatorname{card}(H\cap E^r)\leqq\deg\mathfrak{J}, \tag{3.22}$$

where $\deg\mathfrak{J}$ is the degree of $\mathfrak{J}$. But by Theorem II of Chapter 2 with $s=N$ we deduce

$$\deg\mathfrak{J}\leqq 3^{N-t}D'^t=3^N(D'/3)^t,$$

where $D'=\max(3, D)$ and

$$t=\min(N-r_0, k)\leqq N-r_0=r.$$

Thus

$$\deg\mathfrak{J}\leqq 3^N(D'/3)^r\leqq 3^N D^r,$$

and the lemma now follows on comparing this with (3.21) and (3.22).

6. Proof of Theorem III

This is a rapid deduction from the results proved so far. Let S be an algebraic subset of E^n, of dimension at most $n-r$, which is defined in E^n by homogeneous polynomials of degrees at most D. Suppose S contains an algebraic subgroup H of E^n. By the Proposition of Chapter 1 with $a=2$, we see that there exists an algebraic subgroup H' of E^n, with

$$H \subseteq H' \subseteq S, \tag{3.23}$$

which is itself defined by homogeneous polynomials of degrees at most $2D$. Hence if $n-s$ is the dimension of H', we have $r \leqq s \leqq n$. Put $\delta=\delta(H')$ and $\kappa=\kappa(H')$. Then Lemma 7 gives

$$\delta^2 \kappa \leqq 3^N (2D)^s. \tag{3.24}$$

Also from Lemma 5 the annihilator $A'=\mathscr{A}(H')$ satisfies

$$\det(A') \leqq \delta\kappa^2. \tag{3.25}$$

Finally by Lemma 4 the group A' contains elements $\tau_1, \ldots, \tau_s$, linearly independent over $\mathbb{Z}$, with

$$|\tau_i| \leqq s(\det(A'))^{1/(s+1-i)} \quad (1 \leqq i \leqq s). \tag{3.26}$$

So by (3.24) and (3.25) we deduce

$$\det(A') \leqq \delta\kappa^2 \leqq (\delta^2 \kappa)^2 \leqq 3^{2N}(2D)^{2s}$$

and thus by (3.26)

$$|\tau_i| \leqq n \cdot 3^{2N} 2^{2n} D^{2s/(s+1-i)} \leqq 2^{5N} D^{2s/(s+1-i)} \quad (1 \leqq i \leqq s).$$

Now we have

$$2s/(s+1-i) \leqq 2r/(r+1-i) \quad (1 \leqq i \leqq r);$$

hence (3.27) yields the required estimates for $\tau_1, \ldots, \tau_r$. Moreover these elements lie in $A'=\mathscr{A}(H')$, and hence also in $\mathscr{A}(H)$, by (3.23). This completes the proof of Theorem III.

The Corollary is easily deduced. We apply the Theorem to the algebraic subset S of E^n defined by the single equation resulting from the polynomial

$$(z_1 \ldots z_n)^D P(x_1/z_1, \ldots, x_n/z_n),$$

which can be expressed as a homogeneous polynomial Q in $X_0, \ldots, X_N$ of degree at most D. Thus here $r=1$. For each $(w_1, \ldots, w_n)$ in the group W there is an element γ of E^n with $\pi_i(\gamma)=\psi(w_i)$ $(1 \leqq i \leqq n)$, and the totality of such elements form a subgroup Γ of E^n (not necessarily algebraic, of course). Then by hypothesis Q vanishes on Γ, and therefore also on the Zariski closure H of Γ in E^n. So $H \subseteq S$, and Theorem III now supplies $\tau \neq 0$ in $\mathbb{Z}^n$ with $|\tau| \leqq 2^{5N} D^2$ such that $\tau(h)=0$ for all h in H. In particular $\tau(\gamma)=0$ for all γ in Γ, and putting $\tau=(t_1, \ldots, t_n)$ we see that $t_1 w_1 + \ldots + t_n w_n$ lies in the lattice $\mathscr{L}$ for all $(w_1, \ldots, w_n)$ in W. This completes the proof of the Corollary.

Chapter 4. An Effective Version of Hilbert's Nullstellensatz

1. Introduction

We give here an explicit version of the Nullstellensatz for polynomials with algebraic coefficients. Let K be an algebraic number field with ring of integers $\mathcal{O}$, and for $n \geqq 1$ put $\mathfrak{A} = \mathcal{O}[x_1, \ldots, x_n]$. For a polynomial P in $\mathfrak{A}$ we define the size $\|P\|$ as the maximum of the absolute values of all the conjugates of its coefficients. It is sometimes convenient to speak also of the logarithmic size $\log \|P\|$. For brevity we write $N = 2^{n-1}$.

Theorem IV. *For integers $k \geqq 1$, $d \geqq 1$ let $P_1, \ldots, P_k$ and Q be polynomials in $\mathfrak{A}$ of total degrees at most d such that Q vanishes at all common zeroes (if any) of $P_1, \ldots, P_k$ in $\mathbb{C}^n$. Then there is a positive integer $e \leqq (8d)^{2N}$ and polynomials $A_1, \ldots, A_k$ in $\mathfrak{A}$ of total degrees at most $(8d)^{2N+1}$, such that*

$$aQ^e = A_1 P_1 + \ldots + A_k P_k$$

for some non-zero element a of $\mathcal{O}$. Furthermore, if for some real $h \geqq 0$ the logarithmic sizes of $P_1, \ldots, P_k$ and Q are at most h, then the logarithmic sizes of $A_1, \ldots, A_k$ and a are at most $(8d)^{4N-1}(h + 8d \log 8d)$.

The proof proceeds by a series of lemmas. We also need a Proposition due essentially to Hermann [11]; in fact this whole chapter is based on the methods of [11] (see also Seidenberg's commentary [22]). But as Lazard [13] has pointed out, there are minor errors in [11], some of which remain uncorrected in [22]. We have therefore thought it advisable to prove the Proposition *ab initio.*

The Corollary to the Proposition (Section 4) shows that in the so-called Weak Nullstellensatz, with $Q = 1$, the exponent $2N+1$ in the estimate for the degrees of $A_1, \ldots, A_k$ may be replaced by $N = 2^{n-1}$. One might conjecture that this exponent could further be reduced to n, and it can be shown that this would then be best possible. However, if $n \geqq 3$, it seems rather difficult to get any exponent whatsoever less than 2^{n-1}.

It is convenient to record here two simple inequalities involving sizes. Let Q be a polynomial in $\mathfrak{A}$ of total degree at most d. Then for any polynomial A in $\mathfrak{A}$ we have

$$\|AQ\| \leqq C_d \|A\| \|Q\|,$$

where C_d denotes the number of monomials of total degree at most d. As $C_d \leqq (d+1)^n$, we deduce that

$$\|AQ\| \leqq (d+1)^n \|A\| \|Q\|. \tag{4.1}$$

Also for real $M > 0$ let μ_{ij} $(1 \leqq i, j \leqq n)$ be integers of $\mathcal{O}$ with sizes at most M and non-zero determinant, and define new variables $y_1, \ldots, y_n$ by

$$x_i = \mu_{i1} y_1 + \ldots + \mu_{in} y_n \qquad (1 \leqq i \leqq n).$$

Then if $Q(x_1, \ldots, x_n)$ is expressed as a polynomial $P(y_1, \ldots, y_n)$ in $y_1, \ldots, y_n$ we have

$$\|P\| \leqq C'_d (nM)^d \|Q\|,$$

where C'_d is the number of monomials of total degree exactly d. As $C'_d \leqq (d+1)^{n-1}$, we deduce

$$\|P\| \leqq (d+1)^{n-1} (nM)^d \|Q\|. \tag{4.2}$$

2. Preliminaries

Our first result is a refinement of Lemma 5 (p. 285) of [4]; the use there of the Vandermonde determinant was somewhat wasteful. For the time being we work over a polynomial ring $\mathfrak{R} = K[x_1, \ldots, x_n]$, where K is any subfield of $\mathbb{C}$.

Lemma 1. *For $k \geqq 1$, $m \geqq 1$ let $P_1, \ldots, P_k$ be polynomials in $\mathfrak{R}$ and let $\mathfrak{I}_1, \ldots, \mathfrak{I}_m$ be ideals of $\mathfrak{R}$ such that for each j with $1 \leqq j \leqq m$ not all of $P_1, \ldots, P_k$ lie in $\mathfrak{I}_j$. Then there are rational integers $\lambda_1, \ldots, \lambda_k$, of absolute values at most m, such that $Q = \lambda_1 P_1 + \ldots + \lambda_k P_k$ does not lie in any of $\mathfrak{I}_1, \ldots, \mathfrak{I}_m$.*

Proof. Fix j with $1 \leqq j \leqq m$, and let V_j be the vector space over K of all elements $(\lambda_1, \ldots, \lambda_k)$ of K^k such that $\lambda_1 P_1 + \ldots + \lambda_k P_k$ lies in $\mathfrak{I}_j$. Our hypotheses imply that V_j is a proper subspace of K^k, and therefore there is a non-zero linear form vanishing on V_j. That is to say, there are complex numbers $\alpha_{j1}, \ldots, \alpha_{jk}$, not all zero, such that

$$\alpha_{j1} \lambda_1 + \ldots + \alpha_{jk} \lambda_k = 0 \tag{4.3}$$

for all $(\lambda_1, \ldots, \lambda_k)$ in V_j. For independent variables $u_1, \ldots, u_k$ consider the polynomial

$$A(u_1, \ldots, u_k) = \prod_{j=1}^{m} (\alpha_{j1} u_1 + \ldots + \alpha_{jk} u_k).$$

This is not identically zero, and its degree in each variable is at most m. A standard argument now shows that there are rational integers $\lambda_1, \ldots, \lambda_k$ with $0 \leqq \lambda_i \leqq m$ $(1 \leqq i \leqq k)$ such that $A(\lambda_1, \ldots, \lambda_k) \neq 0$. For these integers (4.3) is impossible for all j with $1 \leqq j \leqq m$, so the resulting element $(\lambda_1, \ldots, \lambda_k)$ is not in any of the subspaces $V_1, \ldots, V_m$. Hence $Q = \lambda_1 P_1 + \ldots + \lambda_k P_k$ has the desired property. This proves the lemma.

Note that the argument of [4] would have given an estimate of order $(km)^k$ for the coefficients.

Lemma 2. *For integers $k \geqq 1$, $d \geqq 1$ let $P_1, \ldots, P_k$ be polynomials in $\mathfrak{R}$ of total degrees at most d with no common zero in $\mathbb{C}^n$. Then there exist rational integers λ_{ij} $(1 \leqq i \leqq n+1,\ 1 \leqq j \leqq k)$, of absolute values at most d^n, such that the polynomials $Q_1, \ldots, Q_{n+1}$ defined by*

$$Q_i = \lambda_{i1} P_1 + \ldots + \lambda_{ik} P_k \quad (1 \leqq i \leqq n+1) \tag{4.4}$$

have no common zero in $\mathbb{C}^n$.

Proof. We consider for each r with $1 \leqq r \leqq n+1$ the following statement.

(I_r): There exist rational integers λ_{ij} $(1 \leqq i \leqq r,\ 1 \leqq j \leqq k)$, of absolute values at most d^n, such that the ideal $\mathfrak{J}_r = (Q_1, \ldots, Q_r)$ generated by the polynomials (4.4) for $1 \leqq i \leqq r$ has the following property. If $r \leqq n$ then either $\mathfrak{J}_r = \mathfrak{R}$ or $\mathfrak{J}_r$ is a proper ideal of rank r and degree at most d^r in the sense of [3] (see also [15]); whereas if $r = n+1$ then $\mathfrak{J}_{n+1} = \mathfrak{R}$.

We shall prove (I_r) by induction on r. For $r=1$ the validity of (I_1) is clear, as at least one of $P_1, \ldots, P_k$ is non-zero, and we take this as Q_1. Now suppose (I_r) is true for some r with $1 \leqq r \leqq n$; we shall deduce (I_{r+1}) by constructing Q_{r+1}. There are two possibilities. Firstly, if $\mathfrak{J}_r = \mathfrak{R}$ then we take $\lambda_{r+1,j} = 0$ $(1 \leqq j \leqq k)$ so that $Q_{r+1} = 0$; then plainly $\mathfrak{J}_{r+1} = (\mathfrak{J}_r, Q_{r+1}) = \mathfrak{R}$. Secondly, if $\mathfrak{J}_r$ is a proper ideal of $\mathfrak{R}$ of rank r and degree at most d^r, then Macaulay's Theorem shows that $\mathfrak{J}_r$ is unmixed, and so by Lemma 2 (p. 85) of [15] it has $m \leqq d^r$ prime components, which we denote by $\mathfrak{P}_1, \ldots, \mathfrak{P}_m$. For each h with $1 \leqq h \leqq m$, not all of $P_1, \ldots, P_k$ can lie in the proper ideal $\mathfrak{P}_h$, otherwise this would give them a common zero in $\mathbb{C}^n$. Hence by Lemma 1 above there are rational integers $\lambda_{r+1,j}$ $(1 \leqq j \leqq k)$, with absolute values at most $m \leqq d^r \leqq d^n$, such that Q_{r+1} in (4.4) does not lie in any of $\mathfrak{P}_1, \ldots, \mathfrak{P}_m$.

Now if $\mathfrak{J}_{r+1} = (\mathfrak{J}_r, Q_{r+1}) = \mathfrak{R}$ we have verified the statement (I_{r+1}). In particular this must hold for $r=n$, by Lemma 2 of [15]. So henceforth we may assume $n \geqq 2$, $r < n$, and $\mathfrak{J}_{r+1} \neq \mathfrak{R}$. But then Lemma 2 of [15] shows that $\mathfrak{J}_{r+1}$ has rank $r+1$ and degree at most d^{r+1}. This completes the induction step.

Finally, the present lemma is an immediate consequence of the statement (I_{n+1}).

3. Linear Equations

The next two results are due essentially to Hermann [11], and they concern solutions of linear equations over $\mathfrak{R} = K[x_1, \ldots, x_n]$. As in the previous section, K is here an arbitrary subfield of $\mathbb{C}$. For integers $p \geqq 1$, $q \geqq 1$, $d \geqq 1$ we consider a system

$$A_{j1} X_1 + \ldots + A_{jp} X_p = B_j \quad (1 \leqq j \leqq q) \tag{4.5}$$

of q equations in p unknowns $X_1, \ldots, X_p$, where the A_{ji} $(1 \leqq i \leqq p,\ 1 \leqq j \leqq q)$ and the B_j $(1 \leqq j \leqq q)$ are polynomials in $\mathfrak{R}$ of total degrees at most d.

Lemma 3. *Suppose the system* (4.5) *has a solution with $X_1, \ldots, X_p$ in $\mathfrak{R}$. Then there is a polynomial ring $\mathfrak{R}'$ with the property that there exists a solution of* (4.5) *of the form*

$$X_i = X_{i0} + X_{i1} x_n + \ldots + X_{ie} x_n^e \quad (1 \leqq i \leqq p) \tag{4.6}$$

for some $e \leqq qd$ and some $X_{i0}, \ldots, X_{ie}$ in $\mathfrak{R}'$. Furthermore if $n=1$ we may take $\mathfrak{R}' = K$, while if $n \geqq 2$ we may take $\mathfrak{R}' = K[x_1', \ldots, x_{n-1}']$ for new variables $x_1', \ldots, x_{n-1}'$ defined by

$$x_i = x_i' + \mu_i x_n \quad (1 \leqq i \leqq n-1), \tag{4.7}$$

where $\mu_1, \ldots, \mu_{n-1}$ are rational integers of absolute values at most qd.

Proof. Let r be the rank of the set of linear forms on the left-hand side of (4.5). If $r=0$ the lemma is trivial, as the system vanishes identically. Also if $r=p$ the solution of (4.5) is unique, and Cramer's Rule immediately gives the required estimates (with $\mu_1=\ldots=\mu_{n-1}=0$ if $n\geqq 2$), because $q\geqq r=p$. Hence we may suppose that

$$1\leqq r\leqq \min(p-1, q).$$

The system (4.5) is then equivalent to (possibly after permuting the unknowns)

$$\Delta X_i=\Delta_i+\Delta_{i,r+1}X_{r+1}+\ldots+\Delta_{ip}X_p \quad (1\leqq i\leqq r), \tag{4.8}$$

where $\Delta\neq 0$, Δ_i $(1\leqq i\leqq r)$ and Δ_{ij} $(1\leqq i\leqq r,\ r+1\leqq j\leqq p)$ are certain signed minors of order r, and hence polynomials in $\mathfrak{R}$ of total degrees at most rd. Let $h\leqq rd$ be the exact total degree of Δ. If $h=0$, then Δ is constant, and there is a solution of (4.5) with $X_{r+1}=\ldots=X_p=0$ which evidently has the required properties (again with $\mu_1=\ldots=\mu_{n-1}=0$ if $n\geqq 2$). So henceforth we can assume $h\geqq 1$. Denote by Δ_0 the homogeneous part of Δ of degree h. If $n=1$ we now define $\mathfrak{R}'=K$. But if $n\geqq 2$ the polynomial $\Delta_0(x_1, \ldots, x_{n-1}, 1)$ is not identically zero, and so there exist integers $\mu_1, \ldots, \mu_{n-1}$ with

$$0\leqq \mu_i\leqq h\leqq rd\leqq qd \quad (1\leqq i\leqq n-1)$$

such that

$$\delta=\Delta_0(\mu_1, \ldots, \mu_{n-1}, 1)\neq 0.$$

Define $x'_1, \ldots, x'_{n-1}$ by (4.7), and write $\mathfrak{R}'=K[x'_1, \ldots, x'_{n-1}]$. Then for either definition of $\mathfrak{R}'$ we see that the difference $\Delta(x_1, \ldots, x_n)-\delta x_n^h$ can be expressed as a polynomial of degree strictly less than h in x_n with coefficients in $\mathfrak{R}'$.

Now suppose $X_1, \ldots, X_p$ in $\mathfrak{R}$ satisfy (4.5). Since the leading coefficient of Δ in $\mathfrak{R}'[x_n]$ is a unit in $\mathfrak{R}'$, the Euclidean Algorithm in this ring leads easily to the existence of polynomials $Y_{r+1}, \ldots, Y_p$ in $\mathfrak{R}$ such that

$$X'_j=X_j-Y_j\Delta \quad (r+1\leqq j\leqq p)$$

are also polynomials of degree strictly less than h in x_n with coefficients in $\mathfrak{R}'$. Defining

$$X'_i=X_i-\Delta_{i,r+1}Y_{r+1}-\ldots-\Delta_{ip}Y_p \quad (1\leqq i\leqq r),$$

we see from (4.8) that $X'_1, \ldots, X'_p$ also provide a solution of (4.5). Also from (4.8) with $X_1, \ldots, X_p$ replaced by $X'_1, \ldots, X'_p$ it is clear that when we write $X'_1, \ldots, X'_r$ as polynomials in x_n with coefficients in $\mathfrak{R}'$, their degrees in x_n are strictly less than $(h+rd)-h=rd$. Thus we obtain the required expressions (4.6) with $e<rd\leqq qd$. This completes the proof of the lemma.

We now introduce a sequence of polynomials $M_i(u, v)$ $(i\geqq 1)$ in two variables u, v as follows. We put $M_1(u, v)=uv$ and we define inductively

$$M_{i+1}(u, v)=M_i(u(uv+v+1), v) \quad (i\geqq 1). \tag{4.9}$$

Proposition. *Suppose the system* (4.5) *has a solution with $X_1, \ldots, X_p$ in $\mathfrak{R}$. Then there are rational integers μ_{rs} $(1\leqq r,s\leqq n)$, with unit determinant and absolute values at most $M_n(q, d)$, such that if we define new variables $y_1, \ldots, y_n$ by*

$$x_r = \mu_{r1} y_1 + \ldots + \mu_{rn} y_n \quad (1 \leqq r \leqq n) \tag{4.10}$$

then (4.5) *has a solution* $X_1, \ldots, X_p$ *in* $\mathfrak{R}$ *with*

$$X_i(x_1, \ldots, x_n) = Y_i(y_1, \ldots, y_n) \quad (1 \leqq i \leqq p),$$

where the degree of Y_i *in* y_s *is at most*

$$M_{n+1-s}(q, d) \quad (1 \leqq i \leqq p, \ 1 \leqq s \leqq n).$$

Proof. We use induction on n. If $n=1$ the result follows immediately from Lemma 3, as $M_1(q, d) = qd$. Now assume the Proposition holds for polynomials in $n-1$ variables, for some $n \geqq 2$. To prove it for the system (4.5), we introduce the variables $x'_1, \ldots, x'_{n-1}$ defined by (4.7) and we write

$$A_{ji} = A_{ji0} + A_{ji1} x_n + \ldots + A_{jid} x_n^d \quad (1 \leqq i \leqq p, \ 1 \leqq j \leqq q)$$
$$B_j = B_{j0} + B_{j1} x_n + \ldots + B_{jd} x_n^d \quad (1 \leqq j \leqq q)$$

for polynomials

$$A_{jit}, B_{jt} (1 \leqq i \leqq p, \ 1 \leqq j \leqq q, \ 0 \leqq t \leqq d)$$

in $\mathfrak{R}' = K[x'_1, \ldots, x'_{n-1}]$. Substituting these and (4.6) into (4.5), and equating coefficients of x_n^t for $0 \leqq t \leqq d+e$, we obtain a new system of linear equations over $\mathfrak{R}'$ for the variables X_{ij} $(1 \leqq i \leqq p, \ 0 \leqq j \leqq e)$. As $d+e \leqq (q+1)d$, the number of equations is at most

$$q' = q(qd + d + 1) \tag{4.11}$$

and their coefficients are polynomials in $\mathfrak{R}'$ of total degrees at most $d' = d$. By Lemma 3 these equations have a solution over $\mathfrak{R}'$ derived from the original solution of (4.5), and so we can apply the inductive hypothesis of the Proposition. Using (4.9) and (4.11), we deduce that there are rational integers μ'_{rs} $(1 \leqq r, s \leqq n-1)$, with unit determinant and absolute values at most $M_{n-1}(q', d') = M_n(q, d)$, such that if

$$x'_r = \mu'_{r1} y'_1 + \ldots + \mu'_{r,n-1} y'_{n-1} \quad (1 \leqq r \leqq n-1)$$

there is a solution with

$$X_{ij}(x'_1, \ldots, x'_{n-1}) = Y_{ij}(y'_1, \ldots, y'_{n-1}) \quad (1 \leqq i \leqq p, \ 0 \leqq j \leqq e), \tag{4.12}$$

where the degree of Y_{ij} in y'_s is at most

$$M_{n-s}(q', d') = M_{n+1-s}(q, d) \quad (1 \leqq i \leqq p, \ 0 \leqq j \leqq e, \ 1 \leqq s \leqq n-1).$$

We now observe that by (4.7) we have

$$x_r = \mu'_{r1} y'_1 + \ldots + \mu'_{r,n-1} y'_{n-1} + \mu_r x_n \quad (1 \leqq r \leqq n-1).$$

These are of the required form (4.10) if we take $y_r = y'_r$ $(1 \leqq r \leqq n-1)$ and $y_n = x_n$. Finally by substituting (4.12) into (4.6) we find the desired solution of (4.5). This completes the proof of the Proposition by induction.

In the next section we shall need estimates for the functions $M_m(q, d)$ for $q \geqq 1$, $d \geqq 1$. For $m \geqq 1$ it is easy to verify by induction on j that

$$M_m(q, d) \leqq M_{m-j}((2^{2^j}-1)(qd)^{2^j}/d, d) \qquad (0 \leqq j < m),$$

and the case $j = m-1$ gives

$$M_m(q, d) \leqq (2^{2^{m-1}}-1)(qd)^{2^{m-1}} < (2qd)^{2^{m-1}}. \tag{4.13}$$

Corollary (Weak Nullstellensatz). *For integers $k \geqq 1$, $d \geqq 1$ let $P_1, \ldots, P_k$ be polynomials in $\mathfrak{R}$ of total degrees at most d with no common zero in $\mathbb{C}^n$. Then there are polynomials $A_1, \ldots, A_k$ in $\mathfrak{R}$, of total degrees at most $2(2d)^N$, such that*

$$1 = A_1 P_1 + \ldots + A_k P_k.$$

Proof. This is immediate from the Proposition together with the estimate (4.13) for $q = 1$.

Now let K be an algebraic number field. We need a simple lemma on the sizes of solutions of systems of linear equations analogous to (4.5) over the ring of integers $\mathcal{O}$ of K. For integers $p \geqq 1$, $q \geqq 1$ and real $A \geqq 1$ we consider the system

$$a_{j1} x_1 + \ldots + a_{jp} x_p = 0 \qquad (1 \leqq j \leqq q), \tag{4.14}$$

where the a_{ij} $(1 \leqq i \leqq p,\ 1 \leqq j \leqq q)$ are elements of $\mathcal{O}$ of sizes at most A.

Lemma 4. *For an integer t with $1 \leqq t \leqq p$ suppose that the system* (4.14) *has a solution with $x_1, \ldots, x_p$ in $\mathcal{O}$ such that $x_t \neq 0$. Then it has a solution with $x_1, \ldots, x_p$ in $\mathcal{O}$ of sizes at most $((p-1)A)^{p-1}$ and $x_t \neq 0$.*

Proof. Let r be the rank of the system (4.14); as in the proof of Lemma 3 we may assume that $1 \leqq r \leqq p-1$ and that the system is equivalent to

$$\delta x_i = \delta_{i,r+1} x_{r+1} + \ldots + \delta_{ip} x_p \qquad (1 \leqq i \leqq r), \tag{4.15}$$

where $\delta \neq 0$ and δ_{ij} $(1 \leqq i \leqq r,\ r+1 \leqq j \leqq p)$ are certain signed minors of order r. Hence they are integers of $\mathcal{O}$ of sizes at most $(rA)^r \leqq ((p-1)A)^{p-1}$. We now distinguish two cases. Firstly suppose $1 \leqq t \leqq r$. Then $\delta_{ts} \neq 0$ for some s with $r+1 \leqq s \leqq p$, otherwise (4.15) would imply that $x_t = 0$ for all solutions of (4.14). We may now exhibit the required solution of (4.14), by taking $x_s = \delta$ and

$$x_j = 0 \ \ (r+1 \leqq j \leqq p, j \neq s), \qquad x_i = \delta_{is} \qquad (1 \leqq i \leqq r).$$

Secondly suppose $r+1 \leqq t \leqq p$. Then the required solution has $x_t = \delta$ and

$$x_j = 0 \ \ (r+1 \leqq j \leqq p, j \neq t), \qquad x_i = \delta_{it} \qquad (1 \leqq i \leqq r).$$

This proves the lemma.

4. Proof of Theorem IV

Suppose that for some $k \geqq 1$, $d \geqq 1$ we have polynomials $P_1, \ldots, P_k$ and Q in $\mathfrak{A} = \mathcal{O}[x_1, \ldots, x_n]$, of total degrees at most d, such that Q vanishes whenever $P_1, \ldots, P_k$ vanish together in $\mathbb{C}^n$. Clearly it suffices to prove the Theorem with $P_1, \ldots, P_k$ replaced by a maximal subset linearly independent over K. Thus we may without loss of generality suppose that already $P_1, \ldots, P_k$ are linearly independent over K, and this implies that

$$k \leqq (d+1)^n \leqq (2d)^n.$$

Also we may assume $Q \neq 0$, otherwise the Theorem is trivial.

We introduce a new variable x_{n+1} (the Rabinowitsch trick) and a new polynomial

$$P_{k+1}(x_1, \ldots, x_{n+1}) = 1 - x_{n+1}\, Q(x_1, \ldots, x_n). \tag{4.16}$$

Then the polynomials $P_1, \ldots, P_{k+1}$ have no common zero and their total degrees are at most $d+1 \leqq 2d$. Hence by Lemma 2 we can find rational integers λ_{ij} $(1 \leqq i \leqq n+2,\ 1 \leqq j \leqq k+1)$, of absolute values at most $(2d)^{n+1}$, such that the polynomials

$$Q_i = \lambda_{i1} P_1 + \ldots + \lambda_{i,k+1} P_{k+1} \qquad (1 \leqq i \leqq n+2) \tag{4.17}$$

have no common zero. The Nullstellensatz over the field K now shows that there exist polynomials $X_1, \ldots, X_{n+2}$ in $K[x_1, \ldots, x_{n+1}]$ such that

$$1 = X_1 Q_1 + \ldots + X_{n+2} Q_{n+2}.$$

We apply the Proposition to this system, with $p = n+2$ and $q = 1$, and we use the inequalities (4.13). We deduce the existence of rational integers μ_{rs} $(1 \leqq r, s \leqq n+1)$, with unit determinant and absolute values at most $(4d)^{2N}$, such that if we define the new variables $y_1, \ldots, y_{n+1}$ by

$$x_r = \mu_{r1} y_1 + \ldots + \mu_{r,n+1} y_{n+1} \qquad (1 \leqq r \leqq n+1) \tag{4.18}$$

then there are polynomials $Y_1, \ldots, Y_{n+2}$ in $K[y_1, \ldots, y_{n+1}]$ with

$$\begin{aligned} 1 = {} & Y_1(y_1, \ldots, y_{n+1})\, Q_1(x_1, \ldots, x_{n+1}) + \ldots \\ & + Y_{n+2}(y_1, \ldots, y_{n+1})\, Q_{n+2}(x_1, \ldots, x_{n+1}), \end{aligned} \tag{4.19}$$

where the degree of Y_i in y_s is strictly less than

$$d_s = (4d)^{2^{n+1-s}} \qquad (1 \leqq i \leqq n+2,\ 1 \leqq s \leqq n+1).$$

Define also polynomials $R_1, \ldots, R_{n+2}$ by (4.18) and

$$R_i(y_1, \ldots, y_{n+1}) = Q_i(x_1, \ldots, x_{n+1}) \qquad (1 \leqq i \leqq n+2); \tag{4.20}$$

by (4.17) these lie in $\mathcal{O}[y_1, \ldots, y_{n+1}]$. Then on clearing denominators in (4.19) we obtain an integer $a \neq 0$ of $\mathcal{O}$ and polynomials $C_1, \ldots, C_{n+2}$ in $\mathcal{O}[y_1, \ldots, y_{n+1}]$ with

$$a = C_1(y_1, \dots, y_{n+1}) R_1(y_1, \dots, y_{n+1}) + \dots$$
$$+ C_{n+2}(y_1, \dots, y_{n+1}) R_{n+2}(y_1, \dots, y_{n+1}), \tag{4.21}$$

where the degree of C_i in y_s is strictly less than d_s ($1 \leqq i \leqq n+2$, $1 \leqq s \leqq n+1$).

We now regard (4.21) as a system of linear equations in a and the coefficients of $C_1, \dots, C_{n+2}$, and we apply Lemma 4 to this system with the condition $a \neq 0$. By the estimates for the separate degrees of $C_1, \dots, C_{n+2}$, the number of unknowns in (4.21) is

$$p = 1 + (n+2) d_1 \dots d_{n+1} = 1 + (n+2)(4d)^{4N-1}. \tag{4.22}$$

Furthermore the coefficients in (4.21) are just the coefficients of $R_1, \dots, R_{n+2}$, and we proceed to estimate these.

Let $H = e^h \geqq 1$; this is an upper bound for the sizes of the original polynomials $P_1, \dots, P_k, Q$. Then by (4.16) also $\|P_{k+1}\| \leqq H$, and so by (4.17) we see that

$$\|Q_i\| \leqq (k+1)(2d)^{n+1} H \leqq (2d)^{4n} H \qquad (1 \leqq i \leqq n+2), \tag{4.23}$$

since $k \leqq (2d)^n$. Because the polynomial Q_i has total degree at most $d+1$ and the coefficients μ_{rs} in (4.18) have absolute values at most $M \leqq (4d)^{2N}$, we find from (4.20), (4.23) and (4.2) that

$$\|R_i\| \leqq (d+2)^n ((n+1) M)^{d+1} \|Q_i\| \leqq (4d)^{8Nd} H \qquad (1 \leqq i \leqq n+2).$$

It follows from Lemma 4 that the sizes of $C_1, \dots, C_{n+2}$, as well as the size of $a \neq 0$, can be chosen not to exceed $\mathscr{C}$, where by (4.22)

$$\log \mathscr{C} \leqq (p-1) \log((p-1)(4d)^{8Nd} H) \leqq (p-1) \log((8d)^{12Nd} H)$$
$$\leqq 12Nd(p-1) \log(8d) + (p-1) h. \tag{4.24}$$

Next, since the transformation (4.18) has unit determinant, we can write

$$C_i(y_1, \dots, y_{n+1}) = B_i(x_1, \dots, x_{n+1}) \qquad (1 \leqq i \leqq n+2) \tag{4.25}$$

for polynomials $B_1, \dots, B_{n+2}$ in $\mathcal{O}[x_1, \dots, x_{n+1}]$. As the total degree of each C_i is at most e, where

$$e + 1 = d_1 + \dots + d_{n+1} + 1 \leqq 2(4d)^{2N} < (8d)^{2N}, \tag{4.26}$$

this is also an upper bound for the degree of each B_i ($1 \leqq i \leqq n+2$). To estimate their sizes, we note that the transformation inverse to (4.18) has coefficients of absolute values at most $M' \leqq n^n (4d)^{2nN} \leqq (8d)^{2nN}$. Hence we find from (4.2) that

$$\|B_i\| \leqq (e+1)^n ((n+1) M')^e \|C_i\| \leqq (8d)^{4nNe} \|C_i\| \qquad (1 \leqq i \leqq n+2),$$

and so $B_1, \dots, B_{n+2}$ have sizes at most $\mathscr{B}$, where

$$\log \mathscr{B} \leqq \log \mathscr{C} + 4nNe \log(8d).$$

By (4.24) this gives

$$\log \mathscr{B} \leqq (p-1)h + \{12Nd(p-1) + 4nNe\} \log(8d). \tag{4.27}$$

And from (4.20), (4.21), (4.25) we see that

$$a = B_1(x_1, \ldots, x_{n+1})\, Q_1(x_1, \ldots, x_{n+1}) + \ldots$$
$$+ B_{n+2}(x_1, \ldots, x_{n+1})\, Q_{n+2}(x_1, \ldots, x_{n+1}).$$

Now (4.17) leads to

$$a = A'_1(x_1, \ldots, x_{n+1})\, P_1(x_1, \ldots, x_n) + \ldots$$
$$+ A'_{k+1}(x_1, \ldots, x_{n+1})\, P_{k+1}(x_1, \ldots, x_{n+1}), \tag{4.28}$$

where $A'_1, \ldots, A'_{k+1}$ have total degrees at most e and their sizes are at most

$$\mathscr{A}' \leqq (n+2)(2d)^{n+1}\, \mathscr{B} \leqq (4d)^{2n}\, \mathscr{B}. \tag{4.29}$$

Finally we set $x_{n+1} = (Q(x_1, \ldots, x_n))^{-1}$ in (4.28) and we multiply by Q^e. Then $P_{k+1} = 1 - x_{n+1}\, Q$ becomes identically zero, and we get

$$aQ^e = A_1(x_1, \ldots, x_n)\, P_1(x_1, \ldots, x_n) + \ldots$$
$$+ A_k(x_1, \ldots, x_n)\, P_k(x_1, \ldots, x_n), \tag{4.30}$$

where

$$A_i(x_1, \ldots, x_n) = Q^e\, A'_i(x_1, \ldots, x_n, Q^{-1}) \quad (1 \leqq i \leqq k).$$

Thus on writing

$$A'_i = A'_{i0} + A'_{i1}\, x_{n+1} + \ldots + A'_{ie}\, x^e_{n+1} \qquad (1 \leqq i \leqq k)$$

for A'_{ij} $(1 \leqq i \leqq k,\ 0 \leqq j \leqq e)$ in $\mathfrak{A} = \mathcal{O}[x_1, \ldots, x_n]$, we see that

$$A_i = Q^e\, A'_{i0} + Q^{e-1}\, A'_{i1} + \ldots + A'_{ie} \qquad (1 \leqq i \leqq k). \tag{4.31}$$

Clearly the total degrees of $A_1, \ldots, A_k$ are at most

$$(d+1)e \leqq (d+1)(8d)^{2N} \leqq (8d)^{2N+1}. \tag{4.32}$$

Also by repeatedly using the inequality (4.1) and summing, we find that

$$\|A_i\| \leqq (e+1)((d+1)^n H)^e\, \mathscr{A}' \leqq (4d)^{ne}\, H^e\, \mathscr{A}' \qquad (1 \leqq i \leqq k).$$

Therefore from (4.26) and (4.29) we conclude that $A_1, \ldots, A_k$ have sizes at most $\mathscr{A}$, where

$$\log \mathscr{A} \leqq \log \mathscr{B} + n(e+2) \log(4d) + eh.$$

Thus (4.27) gives

$$\log \mathscr{A} \leqq (p-1+e)h + \{12Nd(p-1) + 7nNe\} \log(8d). \tag{4.33}$$

But by (4.22) and (4.26) we have

$$p-1+e \leqq (n+2)(4d)^{4N-1}+(8d)^{2N}$$

which is at most $(8d)^{4N-1}$. The same inequalities lead to

$$12Nd(p-1)+7nNe \leqq 12Nd(n+2)(4d)^{4N-1}+7nN(8d)^{2N}$$

which is readily seen not to exceed $(8d)^{4N}$. Hence from (4.33) we deduce

$$\log \mathscr{A} \leqq (8d)^{4N-1}(h+8d\log(8d))$$

as required. Together with the estimate (4.32) for the total degree of $A_1, \ldots, A_k$ and the estimate (4.26) for e, this completes the proof of Theorem IV.

Chapter 5. Proof of the Main Theorem

1. Generalised Addition Formulae

As in Chapter 3 we have to work with a group variety E^n where E is the elliptic curve associated with the Weierstrass elliptic function $\wp(w)$. For the convenience of the reader we recall the embedding of E in projective space $\mathbb{P}_2$. Let $\sigma(w)$ be the corresponding Weierstrass sigma function. Fix a complex number w_0 not in the period lattice $\mathscr{L}$ of $\wp(w)$, and put

$$h(w)=(\sigma(w+w_0))^3, \quad f(w)=h(w)\wp(w+w_0), \quad g(w)=h(w)\wp'(w+w_0).$$

Then E is the locus of the points $\psi(w)$ in $\mathbb{P}_2$ with projective coordinates $f(w)$, $g(w)$, $h(w)$ as w ranges over $\mathbb{C}$. As before, we work with the addition law on E induced by the usual addition law on $\mathbb{C}$. This gives the group variety employed in [17] and Chapter 3 with shifted origin $\psi(0)$. It is not difficult to see that E is defined over the field

$$K=\mathbb{Q}(g_2, g_3, \wp(w_0), \wp'(w_0)),$$

where g_2, g_3 are the invariants of $\wp(w)$. From now on we assume that g_2, g_3, $\wp(w_0)$, $\wp'(w_0)$ are algebraic numbers, so that K is an algebraic number field. For an element α (usually an integer) of K we define its size as the maximum of the absolute values of its conjugates, and its logarithmic size accordingly.

For an integer $m \geqq 1$ we denote by $\mathbb{Z}^m$ the additive group of all $\sigma=(s_1, \ldots, s_m)$ for integers $s_1, \ldots, s_m$, and we write

$$|\sigma|=\max(|s_1|, \ldots, |s_m|).$$

The following result provides us with formulae allowing us to evaluate linear combinations of m elements on E. It is important for us to have such a detailed version which adequately reflects the nature of the corresponding morphism $E^m \rightarrow E$ of projective varieties.

Lemma 1. *For each $m \geqq 1$ there exist constants a_m, c_m, depending only on g_2, g_3, w_0 and m, such that for each σ in $\mathbb{Z}^m$ and each integer p with $0 \leqq p \leqq a_m$ there are polynomials $F_\sigma^{(p)}$, $G_\sigma^{(p)}$, $H_\sigma^{(p)}$, homogeneous of degree at most $c_m(1+|\sigma|^2)$ in each of the sets of variables x_j, y_j, z_j $(1 \leqq j \leqq m)$, whose coefficients are integers of K of logarithmic sizes at most $c_m(1+|\sigma|^2)$, with the following properties. Suppose for each j with $1 \leqq j \leqq m$ we have a complex number w_j, and numbers ξ_j, η_j, ζ_j which are either all zero or projective coordinates of $\psi(w_j)$. Then for all p with $0 \leqq p \leqq a_m$ the numbers*

$$
\begin{aligned}
&F_\sigma^{(p)}(\xi_1, \eta_1, \zeta_1, \ldots, \xi_m, \eta_m, \zeta_m)\\
&G_\sigma^{(p)}(\xi_1, \eta_1, \zeta_1, \ldots, \xi_m, \eta_m, \zeta_m)\\
&H_\sigma^{(p)}(\xi_1, \eta_1, \zeta_1, \ldots, \xi_m, \eta_m, \zeta_m)
\end{aligned}
\tag{5.1}
$$

are either all zero or projective coordinates of $\psi(s_1 w_1 + \ldots + s_m w_m)$. Furthermore, if ξ_j, η_j, ζ_j are not all zero for each j with $1 \leqq j \leqq m$, there exists at least one p with $0 \leqq p \leqq a_m$ such that the numbers (5.1) *are not all zero.*

Proof. Although this result resembles Lemma 7 (p. 176) of [25], we cannot apply this lemma directly, because not all of the conditions of Serre's appendix to [25] are satisfied for our particular embedding of E. Namely, the divisor D of Proposition 3 (p. 194) is not symmetric, as it corresponds to the point w_0 of $\mathbb{C}$, which is not in the period lattice. Moreover, even if we could apply this lemma, it would not show that the number a_m could be chosen independently of σ. In fact this defect is not crucial but it would be very inconvenient, so we prefer to give a proof which does not appeal to [25]. In what follows c, c', c'' will denote constants depending only on g_2 and g_3.

We start by proving that for each integer s there are polynomials

$$F_s(x, y, z), \qquad G_s(x, y, z), \qquad H_s(x, y, z), \tag{5.2}$$

homogeneous of degree exactly s^2, whose coefficients are integers of K of logarithmic sizes at most $c(1+s^2)$, such that

$$
\begin{aligned}
F_s(f(w), g(w), h(w)) &= C_s f(s(w+w_0)-w_0)\\
G_s(f(w), g(w), h(w)) &= C_s g(s(w+w_0)-w_0)\\
H_s(f(w), g(w), h(w)) &= C_s h(s(w+w_0)-w_0)
\end{aligned}
\tag{5.3}
$$

identically in w, for some constant $C_s \neq 0$. These imply the multiplication formulae predicted by Corollary 2 (p. 195) of [25] for the elliptic curve E with the conventional group law. However, [25] gives no estimates for their coefficients, so we give a direct proof.

All this is obvious if $s=0$ or ± 1; so henceforth assume $|s| \geqq 2$. Now the function

$$\Psi_s = \Psi_s(w) = \sigma(s(w+w_0))/(\sigma(w+w_0))^{s^2}$$

is a polynomial in $\wp = \wp(w+w_0)$ and $\wp' = \wp'(w+w_0)$ of weight s^2-1 if $\wp$, $\wp'$ are assigned weights 2, 3 respectively. Furthermore, well-known estimates show that the coefficients of this polynomial are polynomials of degree at most $c's^2$ in $\frac{1}{4}g_2$, $\frac{1}{4}g_3$ whose coefficients are themselves rational integers of absolute values at most c'^{s^2}.

Next, by differentiation we obtain similar estimates for the functions

$$\begin{aligned} s^3\Psi_s^3\wp(s(w+w_0)) &= s^3\Psi_s^3\wp - s\Psi_s^2\Psi_s'' + s\Psi_s\Psi_s'^2 \\ s^3\Psi_s^3\wp'(s(w+w_0)) &= s^2\Psi_s^3\wp' - \Psi_s^2\Psi_s''' + 3\Psi_s\Psi_s'\Psi_s'' - 2\Psi_s'^3 \end{aligned} \tag{5.4}$$

together with $s^3\Psi_s^3$, which have weights at most $3s^2$. A simple induction using the identity

$$\wp^3 = \tfrac{1}{4}g_2\wp + (\tfrac{1}{4}g_3 + \tfrac{1}{4}\wp'^2)$$

shows that for each $k \geqq 3$ we may write

$$4^k\wp^k = P_k(\wp')\wp^2 + Q_k(\wp')\wp + R_k(\wp'),$$

where P_k, Q_k, R_k are polynomials of degrees at most $2(k-2)/3$, $2(k-1)/3$, $2k/3$ respectively, whose coefficients are polynomials in $\tfrac{1}{4}g_2$, $\tfrac{1}{4}g_3$ of degrees at most $c''k$ whose coefficients are themselves rational integers of absolute values at most c''^k. It follows without difficulty that, after multiplying by 4^{3s^2}, the functions (5.4), together with $s^3\Psi_s^3$, may be written as linear combinations of

$$1, \wp', \ldots, \wp'^{s^2}, \wp, \wp\wp', \ldots, \wp(\wp')^{s^2-1}, \wp^2, \wp^2\wp', \ldots, \wp^2(\wp')^{s^2-2}$$

with coefficients satisfying similar estimates. Now on multiplying by $d_s\sigma^{3s^2}$, where $\sigma=\sigma(w+w_0)$ and d_s is a suitable denominator of powers of $\tfrac{1}{4}g_2$, $\tfrac{1}{4}g_3$, we obtain the desired polynomials F_s, G_s, H_s in $\sigma^3\wp$, $\sigma^3\wp'$, σ^3, with $C_s = s^3 4^{3s^2} d_s$.

At this point we remark that if for some complex number w the numbers ξ, η, ζ are either all zero or projective coordinates of $\psi(w)$, then the numbers

$$F_s(\xi,\eta,\zeta), \qquad G_s(\xi,\eta,\zeta), \qquad H_s(\xi,\eta,\zeta) \tag{5.5}$$

are themselves either all zero or projective coordinates of $\psi(s(w+w_0)-w_0)$. This is clear by homogeneity, as we have

$$\xi = \lambda f(w), \qquad \eta = \lambda g(w), \qquad \zeta = \lambda h(w)$$

for some λ, and then the numbers (5.5) are multiples of the left-hand sides of (5.3) by λ^{s^2} (which is to be interpreted as 1 if $\lambda = s = 0$).

Next we write down a complete set of addition laws on E, again with respect to the conventional group law. As in the general discussion in Lemma 1 (p. 492) of [17], we obtain a fixed system of bihomogeneous polynomials

$$\begin{aligned} &F^{(r)}(x,y,z,x',y',z') \\ &G^{(r)}(x,y,z,x',y',z') \\ &H^{(r)}(x,y,z,x',y',z') \qquad (1 \leqq r \leqq k) \end{aligned}$$

(independent of w_0) with the following properties. Suppose for any complex numbers w, w' that ξ, η, ζ are projective coordinates of $\psi(w)$ and ξ', η', ζ' are projective coordinates of $\psi(w')$. Then for any r with $1 \leqq r \leqq k$ the numbers

$$\begin{aligned} &F^{(r)}(\xi,\eta,\zeta,\xi',\eta',\zeta') \\ &G^{(r)}(\xi,\eta,\zeta,\xi',\eta',\zeta') \\ &H^{(r)}(\xi,\eta,\zeta,\xi',\eta',\zeta') \end{aligned} \tag{5.6}$$

are either all zero or projective coordinates of $\psi(w+w'+w_0)$. Furthermore, the latter holds for at least one value of r. Also the homogeneity remark above enables us to extend this to cover the case when ξ, η, ζ or ξ', η', ζ' are all zero; however, then the numbers (5.6) could well be all zero for all r.

Now we can build up the required polynomials $F_\sigma^{(p)}$, $G_\sigma^{(p)}$, $H_\sigma^{(p)}$. Let d be the smallest positive integer such that $d\wp(w_0)$, $d\wp'(w_0)$ are algebraic integers, and write $s_0=1-s_1-\ldots-s_m$. Let X_0, Y_0, Z_0 denote the polynomials (5.2) for $s=s_0$ evaluated at $x=d\wp(w_0)$, $y=d\wp'(w_0)$, $z=d$. Also for $1\leqq j\leqq m$ let X_j, Y_j, Z_j denote the polynomials (5.2) for $s=s_j$ when x, y, z are replaced by the variables x_j, y_j, z_j. Take $a_m=k^m-1$ and set up a $(1-1)$ correspondence between integers p with $0\leqq p\leqq a_m$ and elements $(r_1, \ldots, r_m)$ of $\mathbb{Z}^m$ with $1\leqq r_1, \ldots, r_m\leqq k$. For such an element define X'_j, Y'_j, Z'_j $(0\leqq j\leqq m)$ inductively by

$$X'_0=X_0, \qquad Y'_0=Y_0, \qquad Z'_0=Z_0$$

and

$$\begin{aligned} X'_j&=F^{(r_j)}(X_j, Y_j, Z_j, X'_{j-1}, Y'_{j-1}, Z'_{j-1}) \\ Y'_j&=G^{(r_j)}(X_j, Y_j, Z_j, X'_{j-1}, Y'_{j-1}, Z'_{j-1}) \\ Z'_j&=H^{(r_j)}(X_j, Y_j, Z_j, X'_{j-1}, Y'_{j-1}, Z'_{j-1}). \quad (1\leqq j\leqq m) \end{aligned} \tag{5.7}$$

We proceed to verify that X'_m, Y'_m, Z'_m can be taken as the polynomials $F_\sigma^{(p)}$, $G_\sigma^{(p)}$, $H_\sigma^{(p)}$ of Lemma 1 for the corresponding integer p.

To start with, there is no difficulty in obtaining the desired estimates for the degrees and coefficients of X'_m, Y'_m, Z'_m. Next, suppose that for each j with $1\leqq j\leqq m$ we have a complex number w_j, and numbers ξ_j, η_j, ζ_j which are either all zero or projective coordinates of $\psi(w_j)$. Then X_0, Y_0, Z_0 are projective coordinates of $\psi(s_0 w_0-w_0)$, and the values of X_j, Y_j, Z_j at

$$x_j=\xi_j, \qquad y_j=\eta_j, \qquad z_j=\zeta_j \qquad (1\leqq j\leqq m) \tag{5.8}$$

are either all zero or projective coordinates of

$$\psi(s_j(w_j+w_0)-w_0) \qquad (1\leqq j\leqq m).$$

By induction it follows easily from (5.7) that the values of X'_j, Y'_j, Z'_j at (5.8) are either all zero or projective coordinates of $\psi(w'_j)$, where

$$w'_j=s_0 w_0+s_1(w_1+w_0)+\ldots+s_j(w_j+w_0)-w_0.$$

Since $s_0=1-s_1-\ldots-s_m$, we see that

$$w'_m=s_1 w_1+\ldots+s_m w_m,$$

and so the numbers (5.1) are either all zero or projective coordinates of $\psi(s_1 w_1+\ldots+s_m w_m)$ as desired.

Finally suppose that ξ_j, η_j, ζ_j are not all zero for each value of j with $1\leqq j\leqq m$. Then X_j, Y_j, Z_j at (5.8) must be projective coordinates of $\psi(s_j(w_j+w_0)-w_0)$, by (5.3) $(1\leqq j\leqq m)$. Now we can choose inductively integers $r_1, \ldots, r_m$ in (5.7) such that X'_j, Y'_j, Z'_j at (5.8) are projective coordinates of $\psi(w'_j)$ $(1\leqq j\leqq m)$. This gives a corresponding choice of p such that the numbers (5.1) are not all zero, and thereby completes the proof of Lemma 1.

2. The Basic Zero Estimate

The main result of this section is a deduction from the results of Chapter 1 and Chapter 3. For $n \geqq 1$ let E^n denote the product of n copies of E, embedded as usual in projective space $\mathbb{P}_N$ with $N=3^n-1$. For $1 \leqq i \leqq n$ let π_i be the projection from E^n to its i-th factor, and, given $\tau=(t_1, \ldots, t_n)$ in $\mathbb{Z}^n$ and g in E^n write

$$\tau(g)=t_1 \pi_1(g)+\ldots+t_n \pi_n(g).$$

For $m \geqq 1$ let $\gamma_1, \ldots, \gamma_m$ be elements of E^n, and for $\sigma=(s_1, \ldots, s_m)$ in $\mathbb{Z}^m$ let

$$\Psi(\sigma)=s_1 \gamma_1+\ldots+s_m \gamma_m.$$

As usual, for $S \geqq 0$ we denote by $\mathbb{Z}^m(S)$ the set of such σ in $\mathbb{Z}^m$ with $0 \leqq s_1, \ldots, s_m \leqq S$. Henceforth we assume $m \geqq 3$, and we write

$$\theta=n/(m-2).$$

Lemma 2. *Suppose for some integer $D \geqq 1$ there is a homogeneous polynomial of degree at most D that vanishes on the set $\Psi(\mathbb{Z}^m(n(3^N D)^\theta))$ but not on all of E^n. Then there exist $\sigma_1, \sigma_2, \sigma_3$ in $\mathbb{Z}^m$, linearly independent over $\mathbb{Z}$, and $\tau \neq 0$ in $\mathbb{Z}^n$, with*

$$|\tau| \leqq 2^{5N}(3^N D)^2, \qquad |\sigma_j| \leqq (3^N D)^\theta \qquad (j=1,2,3) \tag{5.9}$$

such that

$$\tau(\Psi(\sigma_j))=0 \qquad (j=1,2,3).$$

Proof. First we apply the main result of Chapter 1 with $G=E^n$ and $\theta=n/(m-2) \geqq n/m$, taking the constant c as 3^{-N}. The resulting integers k, r then satisfy

$$1 \leqq k \leqq m, \qquad 1 \leqq r \leqq n, \qquad nk+r(m-2)>mn.$$

It follows in particular that

$$nk>(m-2)(n-r)+2n \geqq 2n$$

and so $k \geqq 3$. Also there is a subgroup Z of $\mathbb{Z}^m$ of rank at least k and an algebraic subgroup H of G of dimension at most $n-r$ such that $\Psi(Z) \subseteq H$. Further Z contains k elements $\sigma_1, \ldots, \sigma_k$ linearly independent over $\mathbb{Z}$ with

$$|\sigma_j| \leqq (3^N D)^{r/(m+1-j)} \qquad (1 \leqq j \leqq k), \tag{5.10}$$

and H lies in an algebraic subset S of E^n, of dimension at most $n-r$, which is defined in E^n by the vanishing of homogeneous polynomials of degrees at most $3^N D$.

To get more information on H we apply the main result of Chapter 3. It follows that there exist elements $\tau_1, \ldots, \tau_r$ of $\mathbb{Z}^n$, linearly independent over $\mathbb{Z}$ with

$$|\tau_i| \leqq 2^{5N}(3^N D)^{2r/(r+1-i)} \qquad (1 \leqq i \leqq r) \tag{5.11}$$

such that

$$\tau_i(h)=0 \qquad (1 \leqq i \leqq r)$$

for all h in H. We deduce that

$$\tau_i(\Psi(\sigma_j))=0 \qquad (1 \leqq i \leqq r,\ 1 \leqq j \leqq k).$$

The desired result now comes from taking only $\tau=\tau_1$ and $\sigma_1, \sigma_2, \sigma_3$, and the estimates (5.9) are consequences of (5.10) and (5.11).

3. Ideal-theoretic Interpretation

Our task in this section is to formulate Lemma 2 in terms of ideals, using the polynomials $F_\sigma^{(p)}$, $G_\sigma^{(p)}$, $H_\sigma^{(p)}$ constructed in Lemma 1. We work with $2mn$ indeterminates

$$x_{ij}, y_{ij} \qquad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$$

and the associated polynomial ring

$$\mathfrak{R}=\mathbb{C}[x_{11}, \ldots, x_{nm}, y_{11}, \ldots, y_{nm}].$$

For an integer $D \geqq 1$ and a non-zero polynomial $P=P(x_1, \ldots, x_n)$ in $\mathbb{C}[x_1, \ldots, x_n]$ of degree at most D in each of the independent variables $x_1, \ldots, x_n$ we define an ideal $\mathfrak{I}_P=\mathfrak{I}_P(D)$ of $\mathfrak{R}$ as follows. Recall the number a_m of Lemma 1, and for $\pi=(p_1, \ldots, p_n)$ in $\mathbb{Z}^n(a_m)$ and σ in $\mathbb{Z}^m$ write

$$P_\sigma^{(\pi)}=(Z_1 \ldots Z_n)^D P(X_1/Z_1, \ldots, X_n/Z_n) \tag{5.12}$$

where, after expanding, we make the substitutions

$$\begin{aligned} X_i&=F_\sigma^{(p_i)}(x_{i1}, y_{i1}, 1, \ldots, x_{im}, y_{im}, 1)\\ Y_i&=G_\sigma^{(p_i)}(x_{i1}, y_{i1}, 1, \ldots, x_{im}, y_{im}, 1)\\ Z_i&=H_\sigma^{(p_i)}(x_{i1}, y_{i1}, 1, \ldots, x_{im}, y_{im}, 1). \quad (1 \leqq i \leqq n) \end{aligned} \tag{5.13}$$

Then $P_\sigma^{(\pi)}$ is a polynomial in $\mathfrak{R}$, which strictly speaking depends also on D. Next let c be the smallest positive integer such that cg_2, cg_3, $c\wp(w_0)$, $c\wp'(w_0)$ are algebraic integers, and write

$$C_{ij}=c\{y_{ij}^2-(4x_{ij}^3-g_2 x_{ij}-g_3)\} \qquad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m). \tag{5.14}$$

Then we define $\mathfrak{I}_P$ as the ideal of $\mathfrak{R}$ generated by the polynomials C_{ij} $(1 \leqq i \leqq n, 1 \leqq j \leqq m)$ and $P_\sigma^{(\pi)}$ for all π in $\mathbb{Z}^n(a_m)$ and all σ in $\mathbb{Z}^m(n(3^N D)^\theta)$.

Next we define a second ideal $\mathfrak{J}=\mathfrak{J}(D)$ of $\mathfrak{R}$ that still depends on D but is independent of P. For this we shall need to use Lemma 1 with m replaced by n. For an integer q with $0 \leqq q \leqq a_n$, $\pi=(p_1, \ldots, p_n)$ in $\mathbb{Z}^n(a_m)$, and σ in $\mathbb{Z}^m$, τ in $\mathbb{Z}^n$ define

$$\begin{aligned} F_{\tau,\sigma}^{(q,\pi)}&=F_\tau^{(q)}(X_1, Y_1, Z_1, \ldots, X_n, Y_n, Z_n)\\ G_{\tau,\sigma}^{(q,\pi)}&=G_\tau^{(q)}(X_1, Y_1, Z_1, \ldots, X_n, Y_n, Z_n)\\ H_{\tau,\sigma}^{(q,\pi)}&=H_\tau^{(q)}(X_1, Y_1, Z_1, \ldots, X_n, Y_n, Z_n), \end{aligned} \tag{5.15}$$

where X_i, Y_i, Z_i $(1 \leqq i \leqq n)$ are given by (5.13). Let $\mathfrak{K}(\tau, \sigma)$ be the ideal of $\mathfrak{R}$ generated by the polynomials C_{ij} $(1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$ together with the polynomials (5.15) for all q with $0 \leqq q \leqq a_n$ and all π in $\mathbb{Z}^n(a_m)$.

Further define the polynomials

$$\begin{aligned} X_{\tau,\sigma}^{(q,\pi)} &= c\{F_{\tau,\sigma}^{(q,\pi)} - \wp(w_0) H_{\tau,\sigma}^{(q,\pi)}\} \\ Y_{\tau,\sigma}^{(q,\pi)} &= c\{G_{\tau,\sigma}^{(q,\pi)} - \wp'(w_0) H_{\tau,\sigma}^{(q,\pi)}\} \\ Z_{\tau,\sigma}^{(q,\pi)} &= c\{\wp(w_0) G_{\tau,\sigma}^{(q,\pi)} - \wp'(w_0) F_{\tau,\sigma}^{(q,\pi)}\} \end{aligned} \tag{5.16}$$

and let $\mathfrak{J}(\tau, \sigma)$ be the ideal of $\mathfrak{R}$ generated by these polynomials for all q with $0 \leqq q \leqq a_n$ and all π in $\mathbb{Z}^n(a_m)$. Then for τ in $\mathbb{Z}^n$ and $\sigma_1, \sigma_2, \sigma_3$ in $\mathbb{Z}^m$ let

$$\mathfrak{J}(\tau, \sigma_1, \sigma_2, \sigma_3) = (\mathfrak{J}(\tau, \sigma_1), \mathfrak{J}(\tau, \sigma_2), \mathfrak{J}(\tau, \sigma_3)) \tag{5.17}$$

be the ideal generated by the elements of the ideals $\mathfrak{J}(\tau, \sigma_1)$, $\mathfrak{J}(\tau, \sigma_2)$, $\mathfrak{J}(\tau, \sigma_3)$.

Finally (and this is where the dependence on D comes in) let $\mathscr{X} = \mathscr{X}(D)$ be the set of quadruples $(\tau, \sigma_1, \sigma_2, \sigma_3)$ satisfying the conditions of Lemma 2; that is, $\sigma_1, \sigma_2, \sigma_3$ are elements of $\mathbb{Z}^m$, linearly independent over $\mathbb{Z}$, and $\tau \neq 0$ is in $\mathbb{Z}^n$ with

$$|\tau| \leqq 2^{5N}(3^N D)^2, \qquad |\sigma_j| \leqq (3^N D)^\theta \qquad (j = 1, 2, 3).$$

Let $\mathfrak{J} = \mathfrak{J}(D)$ be the product of the ideals $\mathfrak{J}(\tau, \sigma_1, \sigma_2, \sigma_3)$ taken over all $(\tau, \sigma_1, \sigma_2, \sigma_3)$ in $\mathscr{X}$. The considerations of the preceding two sections may now be concisely expressed as follows.

Lemma 3. *For any integer $D \geqq 1$, any σ in $\mathbb{Z}^m$, and any τ in $\mathbb{Z}^n$, we have $\mathfrak{K}(\tau, \sigma) = \mathfrak{R}$. Furthermore for any non-zero polynomial P of degree at most D in each variable there exists a positive integer e such that $\mathfrak{J}^e \subseteq \mathfrak{J}_P$.*

Proof. We use a method of generic points (compare also [16]). We select in an arbitrary way complex numbers ξ_{ij}, η_{ij} $(1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$ such that the polynomials C_{ij} in (5.14) vanish after the substitutions

$$x_{ij} = \xi_{ij}, \qquad y_{ij} = \eta_{ij} \qquad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m). \tag{5.18}$$

To prove $\mathfrak{K}(\tau, \sigma) = \mathfrak{R}$, it suffices, by Hilbert's Nullstellensatz, to exhibit at least one of the remaining generators (5.15) of $\mathfrak{K}(\tau, \sigma)$ that does not vanish at (5.18).

Now the vanishing of C_{ij} at (5.18) implies that there exist complex numbers w_{ij} such that

$$\xi_{ij} = \wp(w_{ij} + w_0), \qquad \eta_{ij} = \wp'(w_{ij} + w_0) \qquad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m). \tag{5.19}$$

We now fix i and we apply Lemma 1 with $\sigma = (s_1, \ldots, s_m)$ to the numbers $w_{i1}, \ldots, w_{im}$. We deduce the existence of an integer p_i with $0 \leqq p_i \leqq a_m$ such that the polynomials (5.13) evaluated at (5.18) are projective coordinates of the point $\psi(w_i)$ on E, where

$$w_i = s_1 w_{i1} + \ldots + s_m w_{im} \qquad (1 \leqq i \leqq n). \tag{5.20}$$

Next we apply Lemma 1, with m replaced by n and σ by $\tau=(t_1, \ldots, t_n)$, to the points $w_1, \ldots, w_n$. We deduce the existence of q with $0 \leqq q \leqq a_n$ such that the polynomials (5.15) evaluated at (5.18) are projective coordinates of $\psi(t_1 w_1 + \ldots + t_n w_n)$. In particular at least one of these values is non-zero, and this suffices to show that $\mathfrak{K}(\tau, \sigma) = \mathfrak{R}$.

Next, with ξ_{ij}, η_{ij} chosen arbitrarily as above, in order to prove that $\mathfrak{J}^e \subseteq \mathfrak{J}_P$ for some positive integer e it suffices again by the Nullstellensatz to show that if in addition the remaining generators (5.12) of $\mathfrak{J}_P$ vanish at (5.18), then so do all the elements of $\mathfrak{J}$. But let $\sigma = (s_1, \ldots, s_m)$ be an arbitrary element of $\mathbb{Z}^m(n(3^N D)^\theta)$, and let $p_1, \ldots, p_n$ be the integers supplied by the argument above. If $\pi = (p_1, \ldots, p_n)$ then the polynomial (5.12) lies in $\mathfrak{J}_P$, and so by hypothesis it vanishes at (5.18). Now the polynomial P gives rise in the usual way to an associated homogeneous polynomial of degree D in the $N+1$ homogeneous variables of $\mathbb{P}_N$. If we define the points $\gamma_1, \ldots, \gamma_m$ of E^n by

$$\pi_i(\gamma_j) = \psi(w_{ij}) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m), \tag{5.21}$$

then in the notation of Section 3 we have

$$\pi_i(\Psi(\sigma)) = \psi(w_i) \quad (1 \leqq i \leqq n).$$

Consequently the associated homogeneous polynomial vanishes on the set $\Psi(\mathbb{Z}^m(n(3^N D)^\theta))$, but, as $P \neq 0$, not on all of E^n. Thus we have reached the situation of Lemma 2 for the points $\gamma_1, \ldots, \gamma_m$ defined by (5.21).

It follows that there exists a quadruplet $(\tau, \sigma_1, \sigma_2, \sigma_3)$ in the set $\mathscr{X}$ such that

$$\tau(\Psi(\sigma)) = 0 \tag{5.22}$$

for $\sigma = \sigma_1, \sigma_2, \sigma_3$. We proceed to show that (5.22) implies that all the generators (5.16) of the ideal $\mathfrak{J}(\tau, \sigma)$ vanish at (5.18).

For let q be any integer with $0 \leqq q \leqq a_n$, let $\pi = (p_1, \ldots, p_n)$ be any element of $\mathbb{Z}^n(a_m)$, and write as usual $\tau = (t_1, \ldots, t_n)$, $\sigma = (s_1, \ldots, s_m)$. By Lemma 1 the values of the polynomials (5.13) at (5.18) are either all zero or projective coordinates of $\psi(w_i)$, where w_i is defined by (5.20) $(1 \leqq i \leqq n)$. Then Lemma 1 with m replaced by n shows that the values of the polynomials (5.15) at (5.18) are either all zero or projective coordinates of the point

$$\psi(t_1 w_1 + \ldots + t_n w_n) = \tau(\Psi(\sigma)).$$

But by (5.22) the numbers $\wp(w_0)$, $\wp'(w_0)$, 1 are also projective coordinates of this point. Thus the values of the polynomials (5.16) at (5.18) must be zero. As q, π were arbitrary, this shows that all the generators (5.16) of $\mathfrak{J}(\tau, \sigma)$ vanish at (5.18).

Finally, since (5.22) holds for $\sigma = \sigma_1, \sigma_2, \sigma_3$, we deduce from the definition (5.17) that all polynomials of $\mathfrak{J}(\tau, \sigma_1, \sigma_2, \sigma_3)$ vanish at (5.18). As this ideal is a factor of $\mathfrak{J}$, it follows at last that all polynomials of $\mathfrak{J}$ also vanish at (5.18). This completes the proof of Lemma 3.

4. Reduction to Fewer Variables

We would now like to use an explicit version of the Nullstellensatz to convert the assertions of Lemma 3 into a system of identities. But if we work over the polynomial ring $\mathfrak{R}$ our estimates would necessarily get too large for the application we have in mind, essentially because of the large number of variables involved. Thus in this section we proceed to introduce a polynomial ring in fewer variables. This corresponds to the possibility that the numbers in our Main Theorem generate a field of low transcendence degree.

So for an integer $k \geqq 1$ we use new variables $z_1, \ldots, z_k$, and we suppose given a collection of polynomials $X_{ij}(z_1, \ldots, z_k)$, $Y_{ij}(z_1, \ldots, z_k)$ $(1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$ in $\mathbb{C}[z_1, \ldots, z_k]$. These define a $\mathbb{C}$-algebra homomorphism from $\mathfrak{R}$ to $\mathbb{C}[z_1, \ldots, z_k]$ by specifying the images of x_{ij}, y_{ij} as

$$\bar{x}_{ij} = X_{ij}(z_1, \ldots, z_k), \qquad \bar{y}_{ij} = Y_{ij}(z_1, \ldots, z_k) \qquad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m) \tag{5.23}$$

respectively. In general for P in $\mathfrak{R}$ we denote its image in $\mathbb{C}[z_1, \ldots, z_k]$ by $\bar{P}$, and for an ideal $\mathfrak{I}$ of $\mathfrak{R}$ we write $\bar{\mathfrak{I}}$ for the ideal of $\mathbb{C}[z_1, \ldots, z_k]$ generated by all $\bar{P}$ as P runs over $\mathfrak{I}$. In particular $\bar{\mathfrak{R}} = \mathbb{C}[z_1, \ldots, z_k]$ itself. Let $d \geqq 1$ be an upper bound for the total degrees of the polynomials in (5.23).

We recall the ideals $\mathfrak{I}_P$, $\mathfrak{I}$, $\mathfrak{K}(\tau, \sigma)$ of the preceding section; the number of ideals $\mathfrak{I}(\tau, \sigma_1, \sigma_2, \sigma_3)$ multiplied together to give $\mathfrak{I}$ is the cardinality of $\mathscr{X}$, which is of order at least D^n, as is evident from (5.9). The next lemma shows how to improve this substantially by passing from $\mathfrak{R}$ to $\bar{\mathfrak{R}}$.

Lemma 4. *For any integer $D \geqq 1$, any σ in $\mathbb{Z}^m$, and any τ in $\mathbb{Z}^n$, we have $\bar{\mathfrak{K}}(\tau, \sigma) = \bar{\mathfrak{R}}$. Furthermore, there exists a constant c, depending only on m, n and d, with the following property. For any non-zero polynomial P of degree at most D in each variable there exists a positive integer e_0 and a subset $\mathscr{X}_0$ of $\mathscr{X}$ of cardinality at most $cD^{k(1+2\theta)}$, such that if $\mathfrak{I}_0$ is the product of the ideals $\mathfrak{I}(\tau, \sigma_1, \sigma_2, \sigma_3)$ taken over all $(\tau, \sigma_1, \sigma_2, \sigma_3)$ in $\mathscr{X}_0$, we have*

$$\bar{\mathfrak{I}}_0^{e_0} \subseteq \bar{\mathfrak{I}}_P. \tag{5.24}$$

Proof. It is clear from Lemma 3 that $\bar{\mathfrak{K}}(\tau, \sigma) = \bar{\mathfrak{R}}$. In a similar way we deduce that

$$\bar{\mathfrak{I}}^e \subseteq \bar{\mathfrak{I}}_P. \tag{5.25}$$

Now the ideal $\bar{\mathfrak{I}}_P$ of $\bar{\mathfrak{R}}$ is generated by polynomials of total degrees at most $E \leqq c' D^{1+2\theta}$ with c' depending only on m, n and d. This is clear from the equations (5.12) in conjunction with the estimates of Lemma 1. Now if already $\bar{\mathfrak{I}}_P = \bar{\mathfrak{R}}$ then what we are trying to prove is obvious; otherwise the Corollary of Chapter 2 shows that the number l of isolated prime components of $\bar{\mathfrak{I}}_P$ satisfies

$$l \leqq E + E^2 + \ldots + E^k \leqq kE^k \leqq cD^{k(1+2\theta)}.$$

Let $\mathfrak{P}_1, \ldots, \mathfrak{P}_l$ denote these components. Then for each i with $1 \leqq i \leqq l$ we have from (5.25)

$$\bar{\mathfrak{I}}^e \subseteq \bar{\mathfrak{I}}_P \subseteq \mathfrak{P}_i.$$

Since $\bar{\mathfrak{J}}$ is the product over the images in $\bar{\mathfrak{R}}$ of the ideals $\mathfrak{J}(\tau, \sigma_1, \sigma_2, \sigma_3)$ as $(\tau, \sigma_1, \sigma_2, \sigma_3)$ runs over all elements of $\mathscr{X}$, it follows that at least one of these ideals, call it $\mathfrak{J}_i$, satisfies $\bar{\mathfrak{J}}_i \subseteq \mathfrak{P}_i$. If $h \leqq l$ is the number of different ideals among $\mathfrak{J}_1, \ldots, \mathfrak{J}_l$ obtained in this way, we may assume that these are $\mathfrak{J}_1, \ldots, \mathfrak{J}_h$. Thus we obtain a subset $\mathscr{X}_0$ of $\mathscr{X}$ of cardinality h such that the corresponding product $\mathfrak{J}_0$ is $\mathfrak{J}_1 \ldots \mathfrak{J}_h$. We proceed to verify (5.24).

Choose any primary decomposition of $\bar{\mathfrak{J}}_P$, and let e_0 be the largest exponent of any primary component appearing, whether isolated or embedded. Let $\mathfrak{P}$ be an arbitrary prime component of $\bar{\mathfrak{J}}_P$, and let $\mathfrak{Q}$ be the corresponding primary component in the decomposition chosen above. Then we have $\mathfrak{P}_i \subseteq \mathfrak{P}$ for some i with $1 \leqq i \leqq l$, with equality if and only if $\mathfrak{P}$ is itself isolated. There exists j with $1 \leqq j \leqq h$ such that $\bar{\mathfrak{J}}_j \subseteq \mathfrak{P}_i$, and since the exponent of $\mathfrak{Q}$ is at most e_0, we deduce that

$$\bar{\mathfrak{J}}_0^{e_0} = (\bar{\mathfrak{J}}_1 \ldots \bar{\mathfrak{J}}_h)^{e_0} \subseteq \bar{\mathfrak{J}}_j^{e_0} \subseteq \mathfrak{P}_i^{e_0} \subseteq \mathfrak{P}^{e_0} \subseteq \mathfrak{Q}.$$

This holds for all primary components $\mathfrak{Q}$ of $\bar{\mathfrak{J}}_P$, and (5.24) follows.

5. The Analytic Construction

For the rest of this chapter we shall assume that the Main Theorem is false, and we shall eventually deduce a contradiction. Thus we have complex numbers $u_1, \ldots, u_n$ and $v_1, \ldots, v_m$ and a real number $\kappa \geqq 0$ satisfying

$$|t_1 u_1 + \ldots + t_n u_n| \geqq \exp(-T^\kappa), \qquad |s_1 v_1 + \ldots + s_m v_m| \geqq \exp(-S^\kappa) \tag{5.26}$$

for all integers $t_1, \ldots, t_n, s_1, \ldots, s_m$ with

$$T = \max(|t_1|, \ldots, |t_n|), \qquad S = \max(|s_1|, \ldots, |s_m|)$$

sufficiently large. We assume

$$mn \geqq \{2^{k+1}(k+7) + 4\kappa\}(m+2n) \tag{5.27}$$

for some integer $k \geqq 1$, but that all the finite numbers among the $\wp(u_i v_j)$ $(1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$ lie in a field of transcendence degree at most $k-1$ over $\mathbb{Q}$. By (5.27) we have $m \geqq 3$, and therefore we can use the results of the previous sections.

Recall that an algebraic point of $\wp(w)$ is a complex number w_0 such that either w_0 is in the period lattice $\mathscr{L}$ or $\wp(w_0)$, $\wp'(w_0)$ are algebraic numbers. Since the set of algebraic points is well-known to be an additive group of unbounded rank, we can find a w_0, not in $\mathscr{L}$, such that for all $\tau = (t_1, \ldots, t_n)$ in $\mathbb{Z}^n$ and $\sigma = (s_1, \ldots, s_m)$ in $\mathbb{Z}^m$ the numbers $\wp(w + w_0)$, $\wp'(w + w_0)$ are defined, where

$$w = (t_1 u_1 + \ldots + t_n u_n)(s_1 v_1 + \ldots + s_m v_m).$$

In particular, the numbers

$$\wp(u_i v_j + w_0), \qquad \wp'(u_i v_j + w_0) \qquad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$$

together with $g_2, g_3, \wp(w_0), \wp'(w_0)$ lie in a field K^* of transcendence degree at most $k-1$ over $\mathbb{Q}$. Let $\mathcal{O}$ denote the ring of integers of the algebraic number field K generated over $\mathbb{Q}$ by $g_2, g_3, \wp(w_0), \wp'(w_0)$. It is not difficult to show that there exist complex numbers $\zeta_1, \ldots, \zeta_k$ such that

$$\wp(u_i v_j + w_0) = X_{ij}(\zeta_1, \ldots, \zeta_k),$$
$$\wp'(u_i v_j + w_0) = Y_{ij}(\zeta_1, \ldots, \zeta_k) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m), \tag{5.28}$$

for polynomials X_{ij}, Y_{ij} $(1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$ in $\mathcal{O}[z_1, \ldots, z_k]$. For if $k=1$ we can find ζ in K^* such that $K^* = K(\zeta)$; and in this case we put $\mathcal{O}' = \mathcal{O}$, while if $k \geqq 2$ we can find $\zeta_1, \ldots, \zeta_{k-1}, \zeta$ in K^* such that $K^* = K(\zeta_1, \ldots, \zeta_{k-1}, \zeta)$ is a finite extension of $K(\zeta_1, \ldots, \zeta_{k-1})$; and in this case we put $\mathcal{O}' = \mathcal{O}[\zeta_1, \ldots, \zeta_{k-1}]$. Then we define $\zeta_k = \zeta/\delta$, where δ is any non-zero element of $\mathcal{O}'$ such that

$$(\delta/\zeta)\,\wp(u_i v_j + w_0), \quad (\delta/\zeta)\,\wp'(u_i v_j + w_0) \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$$

lie in $\mathcal{O}'[\zeta]$. Then (5.28) is true by virtue of the identities $(\zeta/\delta)\,\zeta^r = \delta^r \zeta_k^{r+1}$.

We start by constructing the polynomial P to which we shall eventually apply the results of the previous sections. Following an idea of Philippon (see [19]) we choose the coefficients in K rather than in K^*. Henceforth we use positive constants $c_1, c_2, \ldots$ which depend only on $m, n, u_1, \ldots, u_n, v_1, \ldots, v_m$, g_2, g_3, w_0 and $\zeta_1, \ldots, \zeta_k$. We also introduce an integer parameter $D \geqq 1$ which is sufficiently large with respect to all these quantities. We recall the polynomials $P_\sigma^{(\pi)}$ of $\mathfrak{R}$ of Section 3, and the homomorphism from $\mathfrak{R}$ to $\bar{\mathfrak{R}}$ corresponding to the polynomials X_{ij}, Y_{ij} defined by (5.28). We put

$$\theta = n/(m-2), \qquad \phi = (n-k)/(m+2k)$$

and

$$S_0 = D^\phi (\log D)^{-1}.$$

Lemma 5. *There exists a non-zero polynomial $P(x_1, \ldots, x_n)$, of degree at most D in each variable, whose coefficients are in $\mathcal{O}$ of logarithmic sizes at most $c_1 D S_0^2$, such that*

$$\bar{P}_\sigma^{(\pi)}(\zeta_1, \ldots, \zeta_k) = 0$$

for all π in $\mathbb{Z}^n(a_m)$ and all σ in $\mathbb{Z}^m(S_0)$.

Proof. We shall in fact ensure that

$$\bar{P}_\sigma^{(\pi)}(z_1, \ldots, z_k) = 0 \tag{5.29}$$

identically in $z_1, \ldots, z_k$ for all π in $\mathbb{Z}^n(a_m)$ and all σ in $\mathbb{Z}^m(S_0)$. From the formulae (5.12) and (5.13) and the estimates of Lemma 1 it is clear that if P has degree at most D in each variable, then the polynomials $\bar{P}_\sigma^{(\pi)}$ have total degrees at most $c_2 D S_0^2$. Furthermore their coefficients are linear forms in the coefficients of P. Thus each condition (5.29) leads to the vanishing of at most $c_3 (D S_0^2)^k$ such linear forms. So the total number of linear equations to be solved is at most

$$c_4 (D S_0^2)^k S_0^m = c_4 D^n (\log D)^{-m-2k}.$$

On the other hand, there are at least $c_5 D^n$ coefficients of P at our disposal, so the equations are certainly solvable. An estimate for the sizes of these coefficients now follows from any standard version of Siegel's Box Principle for linear forms. For Lemma 1 shows that the coefficients in the linear equations are in $\mathcal{O}$ of logarithmic sizes at most $c_6 DS_0^2$. This completes the proof of the lemma.

To apply Lemma 4 we must extend the range of σ. Accordingly let

$$S = n(3^N D)^\theta.$$

Lemma 6. *We have*

$$|\bar{P}_\sigma^{(\pi)}(\zeta_1, \ldots, \zeta_k)| \leqq \exp(-S_0^m)$$

for all π in $\mathbb{Z}^n(a_m)$ and all σ in $\mathbb{Z}^m(S)$.

Proof. It is convenient to start by observing that the real-valued function

$$\mu(w) = \max(|f(w)|, |g(w)|, |h(w)|)$$

satisfies

$$c_7^{-1-|w|^2} \leqq \mu(w) \leqq c_7^{1+|w|^2}, \tag{5.30}$$

as is easily proved by considering period parallelograms and functional equations (see for example Lemma 3 (p. 27) of [1]).

Now by applying Lemma 1 with $w_j = u_i v_j$ $(1 \leqq j \leqq m)$ for each i with $1 \leqq i \leqq n$ and using (5.12) and (5.13) together with the definitions (5.23), (5.28) of our basic homomorphism, it is readily seen from Lemma 5 that the function

$$\lambda(w) = P(\wp(u_1 w + w_0), \ldots, \wp(u_n w + w_0)) \tag{5.31}$$

vanishes at

$$w = s_1 v_1 + \ldots + s_m v_m = v(\sigma) \tag{5.32}$$

for all $\sigma = (s_1, \ldots, s_m)$ in $\mathbb{Z}^m(S_0)$. Hence the entire function

$$\Lambda(w) = \lambda(w) \prod_{i=1}^{n} (\sigma(u_i w + w_0))^{3D} \tag{5.33}$$

also has at least S_0^m zeroes, and use of the maximum modulus principle on the circles $|w| = c_8 S$ and $|w| = c_8^2 S$ for sufficiently large c_8 leads in the usual way to the estimate

$$|\Lambda(v(\sigma))| \leqq M \exp(-3S_0^m) \tag{5.34}$$

for all σ in $\mathbb{Z}^m(S)$. Here M is the maximum modulus of $\lambda(w)$ on the larger circle. But by Lemma 5 and (5.30) we have $M \leqq \exp(c_9 DS^2)$. Now the inequality

$$m\phi > 1 + 2\theta \tag{5.35}$$

is easily checked from (5.27) (see also (5.39) later on), and we deduce from (5.34) and (5.35) that

$$|\Lambda(v(\sigma))| \leqq \exp(-2S_0^m). \tag{5.36}$$

To verify the estimates of Lemma 6, let $\pi=(p_1, \ldots, p_n)$ be in $\mathbb{Z}^n(a_m)$, let σ be in $\mathbb{Z}^m(S)$, and write for brevity

$$F_i=\bar{X}_i(\zeta_1, \ldots, \zeta_k), \quad G_i=\bar{Y}_i(\zeta_1, \ldots, \zeta_k), \quad H_i=\bar{Z}_i(\zeta_1, \ldots, \zeta_k) \quad (1 \leqq i \leqq n),$$

where X_i, Y_i, Z_i are the polynomials in (5.13). Write also

$$f_i=f(u_i v(\sigma)), \quad g_i=g(u_i v(\sigma)), \quad h_i=h(u_i v(\sigma)) \quad (1 \leqq i \leqq n);$$

these are projective coordinates of $\psi(u_i v(\sigma))$. In particular $h_i \neq 0$ $(1 \leqq i \leqq n)$. Now Lemma 1 and (5.23), (5.28) show that for each i with $1 \leqq i \leqq n$ either $F_i=G_i=H_i=0$ or these numbers are also projective coordinates of $\psi(u_i v(\sigma))$. If the former possibility holds for some i then clearly $\bar{P}_\sigma^{(\pi)}(\zeta_1, \ldots, \zeta_k)=0$ by (5.12), so in this case we are done. Otherwise $H_i \neq 0$ $(1 \leqq i \leqq n)$ because $h_i \neq 0$ $(1 \leqq i \leqq n)$, and then (5.12) gives

$$\bar{P}_\sigma^{(\pi)}(\zeta_1, \ldots, \zeta_k)=(H_1 \ldots H_n)^D P(F_1/H_1, \ldots, F_n/H_n). \tag{5.37}$$

On the other hand, by (5.31) and (5.33) we have

$$\Lambda(v(\sigma))=(h_1 \ldots h_n)^D P(f_1/h_1, \ldots, f_n/h_n). \tag{5.38}$$

Hence we need upper bounds for the ratios H_i/h_i $(1 \leqq i \leqq n)$. But $|H_i/h_i|=M_i/\mu_i$, where

$$M_i=\max(|F_i|, |G_i|, |H_i|), \quad \mu_i=\max(|f_i|, |g_i|, |h_i|) \quad (1 \leqq i \leqq n).$$

Now $M_i \leqq \exp(c_{10} S^2)$ $(1 \leqq i \leqq n)$ by the estimates of Lemma 1, and $\mu_i=\mu(u_i v(\sigma))$ $(1 \leqq i \leqq n)$, which by (5.30) is at least $\exp(-c_{11} S^2)$. Hence from (5.37) and (5.38) we deduce

$$|\bar{P}_\sigma^{(\pi)}(\zeta_1, \ldots, \zeta_k)| \leqq \exp(c_{12} D S^2) |\Lambda(v(\sigma))|.$$

Finally by (5.36) and the observation (5.35) this leads to the desired inequality of Lemma 6.

6. Completion of the Proof

Before we proceed further it is convenient here to record some consequences of the basic inequality (5.27). Temporarily writing $\theta_0=n/m$, $\lambda=1/(1+2\theta)$, $\lambda_0=1/(1+2\theta_0)$, we easily verify that

$$m\phi\lambda_0+k>mn/(m+2n)$$

and

$$m\phi\lambda_0-m\phi\lambda<n(\lambda_0-\lambda)<2n(\theta-\theta_0)\lambda_0^2.$$

This last expression is

$$4mn^2/\{(m-2)(m+2n)^2\} \leqq 3,$$

so these together with (5.27) yield

$$m\phi>\{2^{k+1}(k+7)+4\kappa-k-3\}(1+2\theta). \tag{5.39}$$

Thus putting

$$\psi = 2^k(k+4)(1+2\theta), \qquad \chi = m\phi - \psi$$

we have

$$m\phi + (k+3)(1+2\theta) > 2\psi \tag{5.40}$$

and

$$\chi > 4\kappa(1+2\theta). \tag{5.41}$$

Finally writing

$$\rho = 2^{k+3}(1+2\theta)$$

we see also that

$$\chi > \rho. \tag{5.42}$$

We now recall the ideals $\mathfrak{J}_p$, $\mathfrak{J}(\tau, \sigma)$, $\mathfrak{K}(\tau, \sigma)$ defined in Section 3 by means of explicitly given generators. For each such ideal we shall refer to the corresponding system of generators as the distinguished basis, even though there might well be a simpler basis available (for example, Lemma 3 shows that each ideal $\mathfrak{K}(\tau, \sigma)$ has the single generator 1). We also refer to the distinguished basis of $\mathfrak{J}(\tau, \sigma_1, \sigma_2, \sigma_3)$ via the definition (5.17); this consists just of the union of the elements of the distinguished bases of $\mathfrak{J}(\tau, \sigma_1)$, $\mathfrak{J}(\tau, \sigma_2)$ and $\mathfrak{J}(\tau, \sigma_3)$. Finally if $\mathfrak{J}$ is any ideal of $\mathfrak{R}$ with a distinguished basis defined in this way, then the images of the elements of this basis under the homomorphism (5.23) are elements of a basis in $\bar{\mathfrak{R}}$ of $\bar{\mathfrak{J}}$, and we shall refer to this basis as the distinguished basis of $\bar{\mathfrak{J}}$.

Next we recall the ideal $\mathfrak{J}_0$ of Lemma 4. The following result uses the Nullstellensatz of Chapter 4 to transfer the inequalities of Lemma 6 to this ideal via the results of Lemma 4.

Lemma 7. *For any τ in $\mathbb{Z}^n$ and any σ in $\mathbb{Z}^m$ with*

$$|\tau| \leqq 2^{5N}(3^N D)^2, \qquad |\sigma| \leqq (3^N D)^\theta \tag{5.43}$$

we have

$$|R(\zeta_1, \ldots, \zeta_k)| \geqq \exp(-c_{13} D^\rho \log D) \tag{5.44}$$

for some element R of the distinguished basis of $\bar{\mathfrak{K}}(\tau, \sigma)$. Furthermore there exists a quadruplet $(\tau, \sigma_1, \sigma_2, \sigma_3)$ in $\mathscr{X}_0$ such that

$$|Q(\zeta_1, \ldots, \zeta_k)| \leqq \exp(-D^\chi) \tag{5.45}$$

for all elements Q of the distinguished basis of $\bar{\mathfrak{J}}(\tau, \sigma_1, \sigma_2, \sigma_3)$.

Proof. Denote by $R_1, \ldots, R_r$ the elements of the distinguished basis of $\bar{\mathfrak{K}}(\tau, \sigma)$. Since this ideal is $\bar{\mathfrak{R}}$ by Lemma 4, the unit polynomial 1 satisfies the hypotheses of the Nullstellensatz with respect to $R_1, \ldots, R_r$. As these latter polynomials come from (5.14) and (5.15), we have $r \leqq D^{c_{14}}$. Also, by Lemma 1 and (5.43), their total degrees are at most $c_{14} D^{4+2\theta}$ and their coefficients are in $\mathcal{O}$ of logarithmic sizes at most $c_{14} D^{4+2\theta}$. Hence, on noting that $4+2\theta \leqq 4(1+2\theta)$, we find that there exist polynomials $B_1, \ldots, B_r$ in $\bar{\mathfrak{R}} = \mathbb{C}[z_1, \ldots, z_k]$, of total degrees at most $c_{15} D^\rho \log D$, with coefficients in $\mathcal{O}$ of logarithmic sizes at most $c_{15} D^\rho \log D$, such that

$$\beta = B_1 R_1 + \ldots + B_r R_r \tag{5.46}$$

for some integer $\beta \neq 0$ of $\mathcal{O}$ whose logarithmic size is also at most $c_{15} D^\rho \log D$.

We now substitute $z_1 = \zeta_1, \ldots, z_k = \zeta_k$ into (5.46). Since

$$|\beta| \geqq \exp(-c_{16} D^\rho \log D),$$
$$|B_i(\zeta_1, \ldots, \zeta_k)| \leqq \exp(c_{17} D^\rho \log D) \quad (1 \leqq i \leqq r),$$

the inequality (5.44) follows for $R = R_i$ for some integer i with $1 \leqq i \leqq r$.

Next, to prove (5.45) it is convenient to argue by contradiction. Let h denote the cardinality of $\mathscr{X}_0$ and write $\mathfrak{J}_1, \ldots, \mathfrak{J}_h$ for the ideals $\mathfrak{J}(\tau, \sigma_1, \sigma_2, \sigma_3)$ as $(\tau, \sigma_1, \sigma_2, \sigma_3)$ runs over $\mathscr{X}_0$. Suppose on the contrary that for each i with $1 \leqq i \leqq h$ there exists Q_i in the distinguished basis of $\bar{\mathfrak{J}}_i$ such that

$$|Q_i(\zeta_1, \ldots, \zeta_k)| > \exp(-D^\chi). \tag{5.47}$$

Then the product $Q = Q_1 \ldots Q_h$ satisfies

$$|Q(\zeta_1, \ldots, \zeta_k)| > \exp(-h D^\chi). \tag{5.48}$$

On the other hand, let $E = [D^{(k+3)(1+2\theta)}]$ and denote by $P_1, \ldots, P_p$ the E-th powers of the elements of the distinguished basis of $\bar{\mathfrak{J}}_P$. There are two kinds of basis element, corresponding to (5.12) and (5.14). However, by (5.23) and (5.28) we have

$$\bar{C}_{ij}(\zeta_1, \ldots, \zeta_k) = 0 \quad (1 \leqq i \leqq n,\ 1 \leqq j \leqq m). \tag{5.49}$$

Hence the estimates of Lemma 6 show that

$$|P_j(\zeta_1, \ldots, \zeta_k)| \leqq \exp(-E S_0^m) \quad (1 \leqq j \leqq p) \tag{5.50}$$

for all the basis elements.

Now $Q = Q_1 \ldots Q_h$ lies in $\bar{\mathfrak{J}}_1 \ldots \bar{\mathfrak{J}}_h = \bar{\mathfrak{J}}_0$ and so by Lemma 4 it satisfies the hypotheses of the Nullstellensatz with respect to $P_1, \ldots, P_p$. As before, we have $p \leqq D^{c_{18}}$. Also, since $h \leqq c_{19} D^{k(1+2\theta)}$ by Lemma 4, calculations as above show that Q has total degree at most $c_{20} D^{(k+4)(1+2\theta)}$ and its coefficients are in $\mathcal{O}$ of logarithmic sizes at most $c_{20} D^{(k+4)(1+2\theta)}$. Moreover Lemma 1 shows that the images of the polynomials (5.12) have total degrees at most $c_{21} D S^2$ and their coefficients are in $\mathcal{O}$ of logarithmic sizes at most $c_{21} D S^2$. Hence, on noting that $E \leqq D^{(k+3)(1+2\theta)}$, we see that all the polynomials $Q, P_1, \ldots, P_p$ have total degrees at most $c_{22} D^{(k+4)(1+2\theta)}$ and coefficients in $\mathcal{O}$ of logarithmic sizes at most $c_{22} D^{(k+4)(1+2\theta)}$. Thus we find that there exists a positive integer $e \leqq c_{23} D^\psi$ and polynomials $A_1, \ldots, A_p$ of $\bar{\mathfrak{R}} = \mathbb{C}[z_1, \ldots, z_k]$ of total degrees at most $c_{23} D^{2\psi} \log D$, whose coefficients are in $\mathcal{O}$ of logarithmic sizes at most $c_{23} D^{2\psi} \log D$, such that

$$\alpha Q^e = A_1 P_1 + \ldots + A_p P_p \tag{5.51}$$

for some integer $\alpha \neq 0$ of $\mathcal{O}$ whose logarithmic size is also at most $c_{23} D^{2\psi} \log D$.

As before, we substitute $z_1 = \zeta_1, \ldots, z_k = \zeta_k$ into (5.51) using (5.50). Since

$$|\alpha| \geqq \exp(-c_{24} D^{2\psi} \log D),$$
$$|A_j(\zeta_1, \ldots, \zeta_k)| \leqq \exp(c_{25} D^{2\psi} \log D) \quad (1 \leqq j \leqq p),$$

the basic inequality (5.40) shows at first that

$$|Q(\zeta_1, \ldots, \zeta_k)|^e \leqq \exp(-\tfrac{1}{2} E S_0^m).$$

Then our upper bound for e gives

$$|Q(\zeta_1, \ldots, \zeta_k)| \leqq \exp(-c_{26} E S_0^m D^{-\psi}).$$

But this contradicts (5.48) as $h \leqq c_{27} D^{k(1+2\theta)}$ by Lemma 4, and thereby completes the proof of Lemma 7.

For $\tau = (t_1, \ldots, t_n)$ in $\mathbb{Z}^n$ and $\sigma = (s_1, \ldots, s_m)$ in $\mathbb{Z}^m$ write

$$u(\tau) = t_1 u_1 + \ldots + t_n u_n, \qquad v(\sigma) = s_1 v_1 + \ldots + s_m v_m.$$

Let $(\tau, \sigma_1, \sigma_2, \sigma_3)$ be the quadruplet of Lemma 7. Then $\sigma_1, \sigma_2, \sigma_3$ are linearly independent over $\mathbb{Z}$, $\tau \neq 0$, and we have

$$|\tau| \leqq c_{28} D^2, \qquad |\sigma_j| \leqq c_{28} D^\theta \qquad (j = 1, 2, 3). \tag{5.52}$$

We deduce the following as a consequence of Lemma 7.

Lemma 8. *There exist elements $\omega_1, \omega_2, \omega_3$ of the period lattice $\mathscr{L}$ such that*
$$|u(\tau) v(\sigma_j) - \omega_j| \leqq \exp(-\tfrac{1}{3} D^\chi) \qquad (j = 1, 2, 3).$$

Proof. Fix σ as one of $\sigma_1, \sigma_2, \sigma_3$; then τ, σ satisfy the conditions (5.43). The polynomial R in (5.44) is the image of one of the generators (5.14) and (5.15); but by (5.49) it cannot be any of the $\bar{C}_{ij}$ $(1 \leqq i \leqq n,\ 1 \leqq j \leqq m)$. Hence there exists q with $0 \leqq q \leqq a_n$ and π in $\mathbb{Z}^n(a_m)$ such that R is the image of one of the polynomials in (5.15). For brevity let

$$F = \bar{F}_{\tau,\sigma}^{(q,\pi)}(\zeta_1, \ldots, \zeta_k), \qquad G = \bar{G}_{\tau,\sigma}^{(q,\pi)}(\zeta_1, \ldots, \zeta_k), \qquad H = \bar{H}_{\tau,\sigma}^{(q,\pi)}(\zeta_1, \ldots, \zeta_k) \tag{5.53}$$

denote their values at $z_1 = \zeta_1, \ldots, z_k = \zeta_k$. Then we deduce from (5.44) that

$$\max(|F|, |G|, |H|) \geqq \delta \tag{5.54}$$

with

$$\delta = \exp(-c_{13} D^\rho \log D).$$

But for these values of q, π the images of the polynomials (5.16) are elements of the distinguished basis of $\mathfrak{J}(\tau, \sigma)$. Hence by (5.17) they are elements of the distinguished basis of $\mathfrak{J}(\tau, \sigma_1, \sigma_2, \sigma_3)$; consequently Lemma 7 holds with Q as any of these images. Writing for brevity

$$\wp = \wp(w_0), \qquad \wp' = \wp'(w_0), \tag{5.55}$$

we deduce from (5.45) with $Q = \bar{X}_{\tau,\sigma}^{(q,\pi)}$ and $Q = \bar{Y}_{\tau,\sigma}^{(q,\pi)}$ that

$$\max(|F - \wp H|, |G - \wp' H|) \leqq \varepsilon \tag{5.56}$$

with

$$\varepsilon = \exp(-D^\chi).$$

Now (5.56) gives

$$|F| \leqq \varepsilon + |\wp| \, |H| \leqq \varepsilon + c_{29} |H|$$
$$|G| \leqq \varepsilon + |\wp'| \, |H| \leqq \varepsilon + c_{29} |H|,$$

and trivially we also have

$$|H| \leqq \varepsilon + c_{29} |H|.$$

Then by (5.54)

$$\delta \leqq \varepsilon + c_{29} |H|.$$

Therefore

$$|H| \geqq c_{30}(\delta - \varepsilon),$$

and since $h = (\sigma(w_0))^3 \neq 0$ we deduce using the inequality (5.42) that $|H| \geqq c_{31} \delta$. In particular $H \neq 0$, and now dividing (5.56) by H gives

$$\max(|F/H - \wp|, |G/H - \wp'|) \leqq c_{32} \varepsilon/\delta, \tag{5.57}$$

and again by (5.42) this is at most $\exp(-\frac{3}{4} D^{\chi})$, say.

Now if $\pi = (p_1, \ldots, p_n)$ then by Lemma 1 for each i with $1 \leqq i \leqq n$ the numbers $\bar{X}_i(\zeta_1, \ldots, \zeta_k)$, $\bar{Y}_i(\zeta_1, \ldots, \zeta_k)$, $\bar{Z}_i(\zeta_1, \ldots, \zeta_k)$ corresponding to (5.13) are either all zero or projective coordinates of $\psi(u_i v(\sigma))$, by (5.23) and (5.28). Hence Lemma 1 with m replaced by n shows that F, G, H are either all zero or projective coordinates of $\psi(u(\tau) v(\sigma))$. But $H \neq 0$, and it follows that

$$F/H = \wp(u(\tau) v(\sigma) + w_0), \qquad G/H = \wp'(u(\tau) v(\sigma) + w_0). \tag{5.58}$$

It is well-known that the difference $\wp(w + w_0) - \wp(w_0)$ has a zero of order at most 2 at any given point. Hence (5.55), (5.58) and the inequalities (5.57), together with standard properties of elliptic functions, yield the existence of a period ω of $\mathscr{L}$ such that

$$|u(\tau) v(\sigma) - \omega| \leqq \exp(-\tfrac{1}{3} D^{\chi}).$$

Taking $\sigma = \sigma_1, \sigma_2, \sigma_3$, we deduce Lemma 8 at once.

Now we finish the proof with a simple application of the Box Principle. We note by Lemma 8 and (5.52) that

$$|\omega_j| \leqq c_{33} D^{2+\theta} \qquad (j = 1, 2, 3). \tag{5.59}$$

Thus, since $\mathscr{L}$ is discrete, we can find rational integers r_1, r_2, r_3, not all zero, of absolute values at most $c_{34} D^{2(2+\theta)}$, such that

$$r_1 \omega_1 + r_2 \omega_2 + r_3 \omega_3 = 0.$$

From Lemma 8 we find that

$$|u(\tau) v(\sigma)| \leqq \exp(-\tfrac{1}{4} D^{\chi}), \tag{5.60}$$

where

$$\sigma = r_1 \sigma_1 + r_2 \sigma_2 + r_3 \sigma_3. \tag{5.61}$$

But $\tau \neq 0$; therefore by (5.26)

$$|u(\tau)| \geqq \exp(-c_{35}|\tau|^{\kappa}) \geqq \exp(-c_{36} D^{2\kappa}).$$

Similarly since $\sigma_1, \sigma_2, \sigma_3$ are linearly independent over $\mathbb{Z}$, we also have $\sigma \neq 0$, and so by (5.26) and (5.61)

$$|v(\sigma)| \geqq \exp(-c_{37}|\sigma|^{\kappa}) \geqq \exp(-c_{38} D^{\kappa(4+3\theta)}).$$

It follows that

$$|u(\tau)\, v(\sigma)| \geqq \exp(-c_{39} D^{4\kappa(1+2\theta)}). \tag{5.62}$$

Now we obtain our final contradiction by comparing (5.60) and (5.62) and appealing to the inequality (5.41). This completes the proof of the Main Theorem.

References

1. Anderson, M., Masser, D.W.: Lower bounds for heights on elliptic curves. Math. Z. **174**, 23–34 (1980)
2. Borel, A.: Linear algebraic groups. New York: Benjamin 1969
3. Brownawell, W.D.: On the orders of zero of certain functions. Mém. Soc. Math. France **2**, 5–20 (1980)
4. Brownawell, W.D., Masser, D.W.: Multiplicity estimates for analytic functions II. Duke Math. J. **47**, 273–295 (1980)
5. Cassels, J.W.S.: An introduction to the geometry of numbers. Berlin-Göttingen-Heidelberg: Springer 1959
6. Chudnovsky, G.V.: Analytic methods in diophantine approximation. Kiev Preprints I.M. 74.8 and 74.9 (1974)
7. Chudnovsky, G.V.: Algebraic independance of values of the exponential and elliptic functions, Proceedings of the International Congress of Math. Vol. 1, pp. 339–350. Helsinki 1978
8. Chudnovsky, G.V.: Indépendance algébrique dans la méthode de Gelfond-Schneider. C.R. Acad. Sci. Paris **291 A**, 365–368 (1980)
9. Gelfond, A.O.: Transcendental and algebraic numbers. New York: Dover 1960
10. Hartshorne, R.: Algebraic geometry. Berlin-Heidelberg-New York: Springer 1977
11. Hermann, G.: Die Frage der endlich vielen Schritte in der Theorie der Polynomideale. Math. Ann. **95**, 736–788 (1926)
12. Kolchin, E.R.: Algebraic groups and algebraic dependence. Amer. J. Math. **90**, 1151–1164 (1968)
13. Lazard, D.: Algèbre linéaire sur $K[X_1, \ldots, X_n]$ et élimination. Bull. Soc. Math. France **105**, 165–190 (1977)
14. Masser, D.W.: Small values of the quadratic part of the Néron-Tate height, Progress in Math. Vol. 12, pp. 213–222. Boston-Basel-Stuttgart: Birkhäuser 1981
15. Masser, D.W.: On polynomials and exponential polynomials in several complex variables. Invent. Math. **63**, 81–95 (1981)
16. Masser, D.W.: A vanishing theorem for power series. Invent. Math. **67**, 275–296 (1982)
17. Masser, D.W., Wüstholz, G.: Zero estimates on group varieties I. Invent. Math. **64**, 489–516 (1981)
18. Masser, D.W., Wüstholz, G.: Algebraic independence properties of values of elliptic functions, to appear in the Proceedings of the Exeter Journées Arithmétiques 1980 (ed. J.V. Armìtage), L.M.S. Lecture Nótes No. 56, Cambridge 1982
19. Philippon, P.: Indépendance algébrique de valeurs de fonctions exponentielles p-adiques. J. reine angew. Math. **329**, 42–51 (1981)

20. Ramachandra, K.: Contributions to the theory of transcendental numbers I, II. Acta Arith. **14**, 65–88 (1968)
21. Reyssat, E.: Un critère d'indépendance algébrique. J. reine angew. Math. **329**, 66–81 (1981)
22. Seidenberg, A.: Constructions in algebra. Trans. Amer. Math. Soc. **197**, 273–313 (1974)
23. Tijdeman, R.: On the number of zeros of general exponential polynomials. Indag. Math. **33**, 1–7 (1971)
24. Tijdeman, R.: An auxiliary result in the theory of transcendental numbers. J. Number Th. **5**, 80–94 (1973)
25. Waldschmidt, M.: Nombres transcendants et groupes algébriques. Astérisque **69–70** (1979)
26. Zariski, O., Samuel, P.: Commutative algebra Vol. II. Berlin-Heidelberg-New York: Springer 1968

Oblatum 3-VII-1982

Invent. math. 72, 465–489 (1983)

Inventiones mathematicae

Ensembles analytiques complexes définis comme ensembles de densité et contrôles de croissance

P. Lelong

Université Paris VI, Tour 45–46, 4, Place Jussieu, F-75230-Paris, Cedex 05, France
et 9, Place de Rungis, F-75013-Paris, France

1. Introduction

1.1. Sur une variété analytique complexe G de dimension pure n, on notera $H(G)$ l'algèbre des fonctions holomorphes. On notera $M_+^p(G)$ le cône des courants positifs fermés qui sont de dimension p, $0 \leqq p \leqq n-1$, c'est-à-dire représentés par une forme différentielle de bidegré $(n-p, n-p)$ et $PS(G)$ le cône des fonctions plurisousharmoniques sur G. A $V \in PS(G)$ on associe le courant

$$T_V = \frac{i}{\pi} \partial \bar{\partial} V = \frac{1}{2\pi} d d^c V \quad (d = \partial + \bar{\partial};\ d^c = i(\bar{\partial} - \partial)). \tag{1}$$

D'autre part on note $\mathcal{N}(G)$ la famille des sous-ensembles analytiques de G. Une propriété d'un courant $T \in M_+^p(G)$ est de posséder (cf. [6]) en tout point $x \in G$ un nombre $v(x, T) \geqq 0$, *fini*, qui par rapport à des coordonnées locales a le sens d'une densité en dimension réelle $2p$ pour la mesure positive σ_T, trace de T. Ce nombre est invariant par les isomorphismes analytiques et est donc une propriété géométrique de T en x (cf. Y.T. Siu [12]). En particulier si l'on a $T = [X]$, courant d'intégration sur $X \in \mathcal{N}(G)$, de dimension pure p, le nombre $v(x, T)$ est la multiplicité de x sur X et celle de l'idéal O_X en x ([4]). Un résultat local important de Y.T. Siu [12] énonce que les *ensembles de densité* $E(c, T)$ définis sur G par

$$E(c, T) = [x \in G;\ v(x, T) \geqq c], \qquad c > 0,\ T \in M_+^p(G) \tag{2}$$

sont analytiques, de dimension (en général non homogène) maxima $p = \dim T$.

On considérera ici le cas où G est soit C^n, soit une variété de Stein munie d'un plongement h dans C^{2n+1} et notée alors (G, h). Un ensemble $X \in \mathcal{N}(G)$ admet une représentation globale par des équations; soit

$$X = \bigcap F_j^{-1}(0),\ 1 \leqq j \leqq n+1, \quad \text{et} \quad F_j \in H(G).$$

Offprint requests to: The author at his second address

En posant

$$V(x)=\sum_{1}^{n+1}|F_j(x)|^2 \tag{3}$$

on a $V\in PS(G)$ et, par un calcul facile, on obtient

$$v(x,T_V)=\min_j[\omega_j(x)]$$

où $\omega_j(x)$ est l'ordre en x de la fonction F_j. On en déduit [7]: *sur une variété de Stein G, la famille $\mathcal{N}(G)$ des sous-ensembles analytiques coïncide avec celle des ensembles de densité $E(1,T_V)$, où V parcourt le cône $PS(G)$.*

Il existe ainsi une application surjective du cône $PS(G)$ sur $\mathcal{N}(G)$. On notera que les représentations globales de $X\in\mathcal{N}(G)$ soit par $X=E(1,[X])$ (représentation intrinsèque), soit par $X=E(1,T_V)$ où V a la forme (3) sont des représentations de X du type $X=E(c,T)$. De telles représentations globales suggèrent des problèmes de contrôle de croissance qu'on étudiera pour $G=C^n$, et G de Stein, $G=(G,h)$ grâce à l'image h. L'indicatrice de croissance de $T\in M_+^p(C^n)$ sera la fonction $v_T(r)=(\tau_{2p}r^{2p})^{-1}\sigma_T(r)$, où $\sigma_T(r)$ est la mesure σ_T portée par $\|z\|\leqq r$ et $\tau_{2p}=\pi^p[p!]^{-1}$ est la mesure de la boule unité de C^p. Etant donné $T\in M_+^p(C^n)$ existe-t-il une majoration de la croissance de $E^k(c,T)\in\mathcal{N}(C^n)$, $0\leqq k\leqq p$, composante de dimension k de l'ensemble analytique $E(c,T)$, quand $c>0$ et $v_T(r)$ sont donnés?

On notera que ce type de problème se présente en théorie des nombres (cf. le mémoire [2] de E. Bombieri): une méthode classique depuis C. Siegel de condensation de zéros A de fonctions holomorphes conduit en effet à des problèmes d'ensembles de densité pour des courants qui sont des limites faibles de courants d'intégration $[A]$ sur A.

1.2. Dans $\mathcal{N}(G)$ on distinguera la famille $\mathcal{N}_i(G)$ des ensembles irréductibles dans G qu'on appellera *cycles*. Soit Z un cycle: l'analyticité des $E(c,T)$ entraîne que $v(x,T)$ ait une valeur constante $N_T(Z)\geqq 0$, quand x varie sur Z, sauf sur un sous-ensemble analytique strict $X\subset Z$, où l'on a

$$v(x,T)>N(Z).$$

On est conduit alors à énoncer:

Définition 1.1. On appelle valence d'un cycle $Z\in\mathcal{N}_i(G)$ relativement à un courant positif fermé T le nombre

$$N_T(Z)=\inf_{x\in Z}v(x,T)$$

On a $N_T(Z)=0$ si $\dim Z>\dim T$; on a encore

$$N_T(z)=[\inf c, c>0;\ Z\in E(c,T)].$$

Au §2 on précisera la notion de contrôle de croissance.

Au §3 l'étude des séries $\sum_q T_q$, où l'on a $T_q \in M_+^p(G)$, puis celle des séries $S^k = \sum_\alpha N_T(Z_\alpha^k)[Z_\alpha^k]$ de courants d'intégration Z_α^k de dimension k, affectés des valences $N_T(Z_\alpha^k)$, $0 \leqq k \leqq n-1$, conduit pour $k = \dim T = p$ à une décomposition $T = S^p + T'$ où T' et S^p sont dans $M_+^p(G)$; il en découle une majoration précise de la composante $E^p(c, T)$ de $E(c, T)$ qui est de la dimension maxima.

Au §4 des exemples montrent que les séries S^k pour $k \leqq p-1$ peuvent diverger: quand c décroissant traverse une valeur $c_0 > 0$, une composante $E^q(c, T)$ pour $q \leqq p-1$ peut en croissant consister d'un nombre non borne' de cycles Z_0^q sur un compact et ceux-ci sont absorbés par un cycle $Z^s \in E(c_0, T)$ d'une dimension $s > q$ quant c traverse c_0. On est conduit à l'étude de cycles extrémaux pour T.

Au §5 le résultat obtenu concerne tout l'ensemble $E(c, T)$ et est inattendu, bien qu'il généralise un énoncé de E. Bombieri obtenu dans [2] pour $T = T_V$, $V \in PS(C^n)$ étant du type minimal de croissance défini par $M_V(r) = \sup_{\|z\| \leqq r} V(z) \sim a \log r$. Il existe une hypersurface $Y_c = F_1^{-1}(0)$, où $F_1 \in H(C^n)$ a une croissance contrôlée; on a $M_1(r) \leqq nc^{-1} M_V(r+\eta) + C(n, \eta) \log(1+r)$ pour tout $\eta > 0$, où $M_1(r) = \sup \log |F_1(z)|$ pour $\|z\| \leqq r$. Il en résulte non pas un contrôle strict de Y_c, et de l'indicatrice $\omega(r)$ de Y_c, mais un *contrôle asymptotique*, la majoration de $\omega(r)$ faisant intervenir une constante additive dépendant de Y_c. Un contrôle strict existe toutefois dans le passage intermédiaire de T à un courant T_V avec conservation du nombre v. Au §6, on fait une hypothèse (S_γ) dite de stabilité: il existe $\gamma > 0$ tel qu'on ait $E(c', T) = E(c, T)$ pour $0 \leqq c - \gamma < c' \leqq c$. On obtient alors un contrôle de croissance pour un système d'équations définissant $E(c, T)$, en utilisant une transformation introduite par C.O. Kiselman [5] qui abaisse $v(x, T)$ d'une même quantité en tous points.

Le §7 étudie la classe des $T \in M_+^p(C^n)$ qui ont une indicatrice $v_T(r)$ bornée pour $r \to +\infty$, courants qu'on peut appeler «algébriques»: leurs ensembles de densité $E(c, T)$ sont algébriques et les résultats précédents donnent des limitation des degrés.

2. Contrôles de croissance et espaces de Fréchet à indicatrices

2.1. Nous étudierons des applications ψ: $E_1 \to E_2$, *non nécessairement linéaires* d'espaces de Fréchet. Sur un tel espace E la topologie est déterminée par une suite croissante $\{p_q(x)\}$ de semi-normes, q parcourant les entiers. Mais elle peut être définie aussi bien par la donnée d'une application $E \times R^+ \to R^+$, soit $(x, r) \to p(x, r)$ telle qu'on ait $p(x, r_q) = p_q(x)$, $\lim r_q = +\infty$ et p croissante de r pour $r > 0$. Si $p(x, r)$ est continu de (x, r), on dira que p est *une indicatrice sur E*.

Exemples. a) $E = H(C^n)$; pour $F \in H(C^n)$ on définit $p(F, r) = \sup |F(z)|$ pour $\|z\| \leqq r$.

b) $E = \mathscr{M}(C^n)$, espace des mesures sur C^n. On définit $\mathscr{M}(C^n) = \lim_{\mathscr{M}(G_q)}$ où G_q est un recouvrement de C^n par des boules compactes. Sur l'espace de Fréchet $\mathscr{M}(C^n)$ on utilisera l'indicatrice $\|\mu\|(r)$, norme de $\mu \in \mathscr{M}(C^n)$ sur le compact $\|z\| \leqq r$.

Définition 2.1. a) On dira que l'application ψ: $E_1 \to E_2$ d'espaces de Fréchet se fait avec contrôle de croissance si tout ensemble borné $B \subset E_1$ a une image $\psi(B)$ bornée dans E_2. Si E_1 et E_2 sont à indicatrices, p_1 et p_2, ceci équivaut à dire: à toute fonction croissante $g_1(r)$, $0 < g_1(r) < \infty$, correspond $g_2(r)$ de même nature telle qu'on ait pour $r \geqq 1$ et tout $x \in E_1$:

$$p_2[\psi(x), r] \leqq g_2(r) \quad \text{dès qu'on a} \quad p_1(x, r) \leqq g_1(r) \quad \text{pour tout} \quad r \geqq 1. \tag{2.1}$$

b) On dira que le contrôle est seulement asymptotique si l'on a (2.1) pour $r > r_0(x)$, $0 < r_0(x) < \infty$.

Remarque 2.2. 1) Le contrôle asymptotique sera obtenu parfois sous la forme

$$p_2[\psi(x), r] \leqq g_2(r) + C(x) \quad \text{pour tout} \quad r \geqq 1 \tag{2.2}$$

où la constante C ne dépend que de $x \in E_1$ et non du paramètre r, et où $\lim_{r = +\infty} g_2(r) = \infty$.

2) Si on a des applications $E_1 \to E_2 \to E_3$ le contrôle et le contrôle asymptotique se conservent dans la composition des applications.

3) Dans la suite l'application ψ ne sera en général étudiée que sur un sous-ensemble de E_1.

Exemple. L'application ψ de $L^1_{\text{loc}}(C^n)$ dans $\mathcal{M}(C^n)$ définie par $V \to \frac{1}{2\pi} \Delta V$ sur le cône $PS(C^n)$, sous-ensemble fermé de $L^1_{\text{loc}}(C^n)$: elle envoie un borné de $PS(C^n)$ dans un ensemble de mesures (positives) borné, ψ se fait donc avec contrôle.

2.2. Dans un domaine U de coordonnées locales z_j, $\bar{z}_j$ d'une variété analytique complexe de dimension n, on notera $\beta = \frac{1}{2} \partial \bar{\partial} \|z\|^2$; $\beta_p = [p!]^{-1} \beta^p$ et $\beta_n = dv_n$ la forme volume. On note $\mathscr{C}_0(U)$ les fonctions continues à support compact dans U, et $\mathscr{C}_0^{p,q}(U)$ les formes de type (p, q) des dz_i, $d\bar{z}_j$ à coefficients dans $\mathscr{C}_0(U)$. A un sous-espace L^p de C^n on associe canoniquement une (p, q) forme différentielle $\tau(L^p)$. Si π: $C^n \to L^p$ est la projection orthogonale, on pose $\tau(L^p) = \pi^* \beta_p$ (cf. [8, p. 59]). Un courant T est dit positif (cf. [6]) de dimension p, s'il est de bidegré $(n-p, n-p)$ et si pour tout système $\alpha_1, \ldots, \alpha_p$ de formes $(1, 0)$ à coefficients dans $\mathbb{C}$, le produit $T \wedge i\alpha_1 \wedge \alpha_1 \wedge \ldots \wedge i\alpha_p \wedge \bar{\alpha}_p$ est une mesure positive-ou encore: si pour tout L^p de la grassmanienne $G_{n,p}$, la mesure $\mu(T, L^p) = T \wedge \tau(L^p)$ est positive.

Pour un choix convenable d'un système $\{L^p_s\}$, $s = 1, \ldots, N$, où $N = (C^p_n)^2$, les $\tau(L^p_s)$ forment une base sur $\mathbb{C}$ des (p, p) formes constantes (cf. [8]). En posant $\mu_s = \mu(T, L^p_s)$, on obtient une expression de $T(\varphi)$, où $\varphi \in \mathscr{C}_0^{p,p}(U)$ a des coefficients $\varphi_{I,J}(x)$ continus:

$$T(\varphi) = \sum_{s, I, J} C_s^{I,J} \mu_s(\varphi_{I,J}), \qquad C_s^{I,J} \in \mathbb{C} \tag{2.3}$$

les constantes complexes $C_s^{I,J}$ ne dépendent ni de T ni de φ. La mesure trace $\sigma_T = T \wedge \beta_p$ est positive et *majore les* μ_s; T est continu d'ordre zéro et il existe des coefficients $C_{n,p}$, $C'_{n,p}$ tels que si l'on pose $|\varphi|(x) = \sup_{I,J} |\varphi_{I,J}(x)|$, on ait pour $\varphi \in \mathscr{C}^{p,p}(U)$:

$$|T(\varphi)| \leqq C_{n,p} \sigma(|\varphi|). \tag{2.4}$$

2.3. L'espace des courants $(n-p,\ n-p)$, continus d'ordre nul sur G sera noté $\mathscr{M}^{p,p}(G)$; il s'identifie à $[\mathscr{M}(G)]^N$.

Dans la suite G sera supposée *réunion dénombrable de compacts* et on notera U_s un recouvrement dénombrable de G par des compacts, assez fin pour que tout U_s appartienne à un domaine de coordonnées locales. Alors dire que B est un ensemble borné de courants positifs dans $\mathscr{M}^{p,p}(G)$ équivaut à dire que l'on a sur chaque U_s une majoration des traces σ_T pour $T \in B$. Si $G = C^n$, on aura une majoration $\sigma_T(r) < g(r) < \infty$, pour tout $r > 0$.

Si T appartient au cône $M_+^p(G)$ des courants positifs fermés, le quotient $\nu(x, T, r) = (\tau_{2p} r^{2p})^{-1} \sigma(x, T, r)$, où $\sigma(x, T, r)$ est la masse de σ_T portée par la boule compacte $B(x, r)$, est fonction croissante de r; $\nu(x, T) = \lim_{r=0} \nu(x, T, r)$ (nombre de Lelong de T en x dans [12, 15]) est une fonction définie sur G, indépendamment des coordonnées locales, et est semi-continue supérieurement; on a $\nu(x, T) \geqq 0$; l'ensemble $[x \in G;\ \nu(x, T) > 0]$ sera dit *l'ensemble de densité de T sur G*: il est une réunion dénombrable des sous-ensembles analytiques $\bigcup_n E\left(\frac{1}{n}, T\right)$ quand G est réunion dénombrable de compacts.

On a posé d'autre part dans C^n, $\nu_T(r) = \nu(0, T, r)$, indicatrice qui se substitue à $\sigma_T(r)$ pour l'étude du contrôle de croissance sur $M_+^p(C^n)$.

Exemple. Soient $E_1 = H(C^n)$, $E_2 = L^1_{\mathrm{loc}}(C^n)$, $E_3 = \mathscr{M}^{p,p}(C^n)$. L'application ψ_1: $F \to \log|F|$ dans le cône $PS(C^n) \subset E_2$ est définie et continue sur $H(C^n) - \{0\}$ (cf. pour les détails [11]). Un borné fermé B qui ne contient pas la constante $F = 0$ a pour image un compact, donc un borné sur $PS(C^n)$; l'application ψ_1: $H(C^n) - \{0\} \to PS(C^n)$ se fait donc avec un contrôle asymptotique du type (2.2), la constante $C(F)$ étant telle que $C(F) \geqq a$ exclut $F = 0$. D'autre part $V \to \frac{1}{2\pi} dd^c V$ est continu de $PS(C^n)$, muni de la topologie induite par $L^1_{\mathrm{loc}}(C^n)$, dans $\mathscr{M}^{n-1,n-1}(C^n)$. Finalement l'application composée $F \to \frac{1}{2\pi} dd^c \log|F|$ se fait avec un contrôle de croissance du type (2.2) qu'on explicite facilement.

3. Séries $S = \Sigma N_T(Z_\alpha)\,[Z_\alpha]$ de courants d'intégration avec valence

3.1. On suppose toujours la variété analytique complexe G munie d'un recouvrement dénombrable $\{U_s\}$, chaque U_s étant un domaine strictement compact sur une carte de coordonnées locales utilisée pour le calcul de la mesure trace σ_T. De ce qui précède résulte:

Proposition 3.1. *Une série formelle $S = \sum_\alpha T_\alpha$, $\alpha \in A$, de courants positifs de dimension p sur G, $0 \leqq p \leqq n-1$, converge dans $\mathscr{M}^{p,p}(G)$ si et seulement si la série $S' \doteq \sum_\alpha \sigma_\alpha$ des mesures traces est convergente sur chaque U_s, ou encore si l'on a $S'_s = \sum_\alpha \sigma_\alpha(U_s) < \infty$ pour tout s. La série formelle S se réduit alors à une série $\sum T_q$, $q \in \mathbb{N}$, les T_α étant nuls sur G, sauf pour une infinité dénombrable d'indices α.*

On remarque que la convergence faible de S, c'est-à-dire sur chaque $\varphi\in\mathscr{C}_0^{p,p}(G)$ équivaut ici à $S'_s<\infty$ et à la convergence en norme dans $\mathscr{M}^{p,p}(G)$.

3.2. Si l'on a de plus $T_q\in M_+^p(G)$, les fonctions $\nu(x,T_q)$ sont définies et semi-continues supérieurement sur G et l'on a:

Proposition 3.2. *Soit $S=\sum_q T_q$ une série de courants positifs fermés de dimension p sur la variété G dénombrable à l'infini. Alors si S converge, la série de fonctions*

$$S''(x)=\sum_q \nu(x,T_q) \tag{3.1}$$

converge uniformément sur tout compact $K\subset G$. Il en est de même de la série numérique

$$\sum(K)=\sum_q \sup_{x\in K}\nu(x,T_q)<\infty \tag{3.2}$$

et l'on a pour le courant $S\in M_+^p(G)$:

$$\nu(x,S)=\sum_q \nu(x,T_q). \tag{3.3}$$

On utilisera la propriété suivante de $\nu(x,T)$:

Lemme 3.3. *Soit $D\Subset U_s$, la trace σ_T étant calculée sur U_s. Il existe une constante C ne dépendant que de la configuration (D,U_s) telle qu'on ait:*

$$\nu(x,T)\leqq C\sigma_T(U_s) \qquad \textit{pour tout } x\in D. \tag{3.4}$$

Soit $a=\operatorname{dist}(D,bU_s)>0$ en coordonnées locales. En exprimant que pour tout $x\in D$ et tout $r<a$ la masse $\sigma(x,T,r)$ de σ_T est majorée par $\sigma_T(U_s)$, on a

$$\nu(x,T)\,\tau_{2p}r^{2p}\leqq\sigma(x,T,r)\leqq\sigma_T(U_s) \tag{3.5}$$

qui établit (3.4) avec $C=(\tau_{2p}a^{2p})^{-1}$. On a aussi d'après (3.5):

$$\sup_{x\in D}\nu(x,T)\leqq C\sigma_T(U_s) \tag{3.6}$$

ce qui établit la convergence uniforme de (3.1) sur D, donc sur tout compact de G. De même (3.6) établit la convergence de $\sum(D)$ et, par un recouvrement fini de K par des domaines tels que D, on établit la convergence de (3.2).

Remarque 3.4. a) Si l'on a $\dim Z=k>p=\dim T$, on a $N_T(Z)=0$, car $N_T(Z)=c>0$ entraîne $Z\subset E(c,T)$ alors que $E(c,T)$ a une dimension maxima au plus p d'après le théorème de Siu.

b) Si $\dim Z=k<p=\dim T$, on a $\mathbb{1}_Z T=0$ pour la restriction du courant T au cycle Z. D'après (2,4) il suffit de montrer $\mathbb{1}_Z\sigma_T=0$ sur un compact. On en donne ici une démonstration comme conséquence de l'énoncé suivant plus général où Z est remplacé par le support Y non nécessairement analytique de certaines mesures.

c) Si $\dim Z=p=\dim T$, on a $\mathbb{1}_Z T=N_T(Z)[Z]$, cf. [12, p. 89 et 111].

Proposition 3.5. *Soit μ une mesure positive dans un domaine borné G de C^n. On suppose que pour $0\leqq r<r_0$, la masse $\mu(x,r)$ de μ portée par une boule compacte*

$B(x,r)\subset G$ a la propriété suivante: il existe $m>0$, tel que $\ell(x,r)=(\tau_m r^m)^{-1}\mu(x,r)$ est fonction croissante de r et pour tout $x\in K\Subset G$ vérifie $\lim_{r=0}\ell(x,r)\geqq C(K)>0$. Alors si l'on a $Y=\operatorname{supp}\mu$, pour tout courant T positif fermé de dimension (complexe) $p>\frac{m}{2}$, on a $\mathbb{1}_Y T=0$.

La proposition 3.5 entraîne b) car si Z est un cycle et si $\dim Z=k<p$, la mesure $\mu=\sigma_{[Z]}$, aire de Z vérifie la condition indiquée avec $C(K)=1$, d'après le théorème de P. Thie (cf. [15]), pour $m=2k<2p$. La proposition 3.5 résulte d'une construction élémentaire. Soit $G_a\Subset G$ le compact des points $x\in G$ tels qu'on ait $d(x,\partial G)\geqq a>0$. La fonction $v(x,T,r)=(\tau_{2p}r^{2p})^{-1}\sigma(x,T)$ étant croissante de r (cf. [6]) on a pour $x\in G_{2a}$, et $0<r<a$ la majoration

$$v(x,T,r)\leqq v(x,T,a)\leqq\sigma_T(G_a)(\tau_{2p}a^{2p})^{-1}=c_1 \tag{3.7}$$

où les c_s ne dépendent pas de r. Pour r assez petit ($0\leqq r\leqq a(\sqrt{8n})^{-1}$ suffit), on construit un pavage régulier $P_1=\{A_j\}$ de $R^{2n}=C^n$ par des pavés de coté r, soit, (x^j,y^j) étant le centre de A_j:

$$A_j=\left[z_s=x_s+iy_s;\ x_s^j-\frac{r}{2}<x_s\leqq x_s^j+\frac{r}{2},\ y_s^j-\frac{r}{2}<y_s\leqq y_s^j+\frac{r}{2}\right],\qquad 1\leqq s\leqq n.$$

On associe à A_j le pavé A'_j de coté $3r$, de même centre obtenu en ajoutant à A_j les A_t contigus à A_j. Le degré du recouvrement $\{A'_j\}$ est au plus 3^{2n}. On ne considère que les pavés A_j (notés $\tilde{A}_j$) pour lesquels on a $\tilde{A}_j\cap Y\cap G_{2a}\neq\emptyset$; pour $r<a(\sqrt{8n})^{-1}$ le pavé $\tilde{A}'_j$ associé est dans G_a.

L'hypothèse sur μ utilisée pour $K=G_a$ permet alors de majorer en fonction de $\mu(G_a)$ le nombre $N(r)$ des $\tilde{A}_j$. On a

$$N(r)\leqq 3^{2n}\mu(G_a)[C(G_a)\tau_m r^m]^{-1}=c_2r^{-m}.$$

D'autre part chaque $\tilde{A}_j$ est contenu dans une boule $B(x,r\sqrt{2n})\,C\,G_a$ et d'après (3.7) porte une masse σ_T majorée par c_3r^{2p}, où $c_3=c_1(2n)^p\tau_{2p}$, le courant T étant de dimension p. Finalement on a, pour le recouvrement de $Y\cap G_{2a}$ par les $\tilde{A}_j$:

$$\sigma_T(\bigcup_j A_j)\leqq N(r)\,c_3r^{2p}\leqq c_4r^{2p-m}.$$

D'où en faisant tendre r vers zéro pour $m<2p$:

$$|\mathbb{1}_Y T|_{G_a}=|\mathbb{1}_Y\sigma_T|_{G_a}=0.$$

Définition 3.6. Soient T et T' deux courants positifs de même dimension k; on dira que T' majore $T(T\leqq T')$ si $T'-T$ est un courant positif.

Théorème 3.7. *Soit T un courant positif fermé sur G et $\dim T=p$. Alors*

a) Les cycles Z^p sur G qui vérifient $\dim Z=p$ et $N_T(Z)>0$ forment une famille dénombrable Z_q^p.

b) La série de courants d'intégration à valence sur les Z_q^p, soit

$$S^p = \sum_q N_T(Z_q^p)\,[Z_q^p]$$

converge. On a $S^p \in M_+^p(G)$ *et* $S^p \leqq T$ *sur* G.

c) Pour le courant résiduel $R = T - S^p \in M_+^p(G)$, *les cycles* Z *vérifiant* $N_R(Z) > 0$ *sont de dimension au plus* $p-1$.

Soit Z_α^p, $\alpha \in A$, une indexation des cycles de dimension p sur G et soit $H \subset A$ une partie finie des indices. On note $X(H) = \bigcup_{\alpha \in H} Z_\alpha^p$. L'ensemble $X(H)$ est un ensemble analytique complexe dans G. Les intersections $Z_\alpha \cap Z_\rho$ étant de dimension $p-1$ au plus, on a pour toute forme $\varphi \in \mathscr{C}^{pp}(G)$:

$$\sum_{\alpha \in H} N_T(Z_\alpha^p)\,[Z_\alpha^p](\varphi) = \int_{X(H)} v(x, T)\,\varphi = T_H(\varphi). \tag{3.8}$$

On a $T_H \in M_+^p(G)$. Le courant $R_H = T - T_H$, restriction de T à l'ouvert connexe $G \setminus X(H)$ appartient à $M_+^p(G)$. On a donc

$$T_H \leqq T \quad \text{et} \quad \sigma_H \leqq \sigma_T.$$

La seconde majoration, appliquée à la mesure trace de T_H, et la proposition 3.1 entraînent la convergence de la série (3.8) sur tout U_s car T_H y a une majoration qui est indépendante de $H \subset A$. La convergence de (3.6) entraîne alors que la famille des cycles Z^p de dimension p qui vérifient $N_T(Z^p) > 0$ soit dénombrable, on la note $\{Z_q^p\}$. De plus la série

$$S^p = \sum_q N(Z_q^p)\,[Z_q^p]$$

converge; on a $S^p \in M_+^p(G)$ et le courant limite forte:

$$R = \lim_H R_H = \lim_{m=\infty} \left\{ T - \sum_1^m N(Z_q^p)\,[Z_q^p] \right\}$$

est positif fermé. Montrons que si Z_0^p est un cycle de dimension p dans G, on a $N_R(Z_0^p) = 0$; il suffit pour cela de prouver l'existence de $x_0 \in Z_0^p$ vérifiant $v(x_0, R) = 0$. On a $v(x_0, T) = v(x_0, S^p) + v(x_0, R)$. On a $v(x_0, T) = N_T(Z_0^p) = c \geqq 0$ pour tout $x \in Z_0^p$ sauf sur un sous-ensemble $A_1 = \bigcup_m [Z_0^p \cap X_m]$, où $X_m = E\left(c_0 + \frac{1}{m}, T\right)$ est un ensemble analytique: A_1 est donc une réunion dénombrable de sous-ensembles analytiques stricts de Z_0^p. D'autre part on a $v(x, S^p) = \sum_q v(x, T_q)$ où l'on a posé $T_q = N_T(Z_q^p)\,[Z_q^p]$. Dans $S^p = \sum_q T_q$ isolons le terme $N_T(Z_0^p)\,[Z_0^p]$ qu'on note T_0; on a pour $x \in Z_0^p$ et $q \neq 0$:

$$v(x, T_q) = 0 \qquad \text{pour } x \notin Z_0^p \cap Z_q^p = X_q'.$$

On a donc $v(x, S^p) = v(x, T_0)$ sauf pour x appartenant à l'ensemble $\bigcup_q X_q'$ sur Z_0^p. On a alors $v(x, S^p) = N_T(Z_0^p) = c$ sauf sur une réunion dénombrable A_2 de sous-ensembles analytiques stricts sur Z_0^p. Finalement en dehors d'un tel sous-ensemble, on a sur Z_0^p: $v(x, T) = v(x, S^p) = N_T(Z_0^p)$, d'où $v(x, R) = 0$, ce qui établit l'existence de x_0 et $N_R(Z_0^p) = 0$.

3.4. Notons pour $c>0$, $E(c,T)=\bigcup_k E^k(c,T)$ une décomposition dans G de l'ensemble de densité en ses composantes des dimensions k, $0 \leqq k \leqq p = \dim T$. Le résultat précédent permet de majorer $E^p(c,T)$. On obtient aisément:

Corollaire 3.8. *Pour tout compact $K \subset G$, d'une carte locale, l'aire σ de $E^p(c,T)$ vérifie $\sigma(K) \leqq c^{-1}\, \sigma_T(K)$, où σ_T est la trace de T.*

D'autre part soit $X = E^p(c,T)$ et $x \in X$. On a

$$v(x,T) \geqq v(x,S^p) = \sum_q N(z_q^p)\, v(x,Z_q^p) \geqq c \sum_q{}' v(x,Z_q^p) = c\, p_x(X)$$

la somme $\sum'$ étant relative aux seuls cycles contenus dans $E^p(c,T)$. D'où:

Corollaire 3.9. *Si pour $x \in E^p(c,T) = X$, on note $p_x(X)$ la multiplicité de x sur X, on a en tout point de $E^p(c,T)$ dans G la majoration*

$$p_x(X) \leqq c^{-1}\, v(x,T)$$

du degré de la singularité en x sur l'ensemble analytique $E^p(c,T)$.

En particulier pour V plurisousharmonique dans G et $T = T_V = \frac{1}{2\pi} dd^c V$, on aura

Corollaire 3.10. *Soit $V \in PS(G)$. Alors:*

1) *Sur toute carte locale de G, l'aire de $E^{n-1}(c,T_V)$ composante de codimension 1 de l'ensemble de densité $E(c,T_V)$, est majorée par la mesure $\frac{1}{2\pi c} \Delta V$.*

2) *La multiplicité d'un point x sur $E^{n-1}(c,T_V)$, soit p_x est majorée par $c p_x \leqq v(x,T_V) = \lim_{r=0} (\log r)^{-1}\, \lambda(V,x,r)$ où λ est la moyenne de V sur la sphère de centre x, de rayon r.*

Corollaire 3.11. *Soit G une variété de Stein vérifiant $H^2(G,R)=0$. Sur G toute fonction $V \in PS(G)$ admet une décomposition, unique à l'addition près d'une fonction pluriharmonique, sous la forme $V = U + W$, où l'on a $U \in PS(G)$, $W \in PS(G)$ et où T_U n'a que des cycles de densité de codimension un, tandis que ceux de T_W sont de codimension au moins 2.*

On résout $\frac{1}{2\pi} dd^c U = S^{n-1} = \sum N(Z_q^{n-1})[Z_q^{n-1}]$, les valences étant relatives à T_V. Puis on résout $\frac{1}{2\pi} dd^c W = T_V - S^{n-1} = R$.

On a alors pour $\varphi = V - U - W$ l'équation $dd^c \varphi = 0$ qui montre que φ est pluriharmonique. D'où la décomposition $V = U + (W + \varphi)$ qui possède les propriétés énoncées d'après le théorème 3.7.

3.5. *Contrôle de croissance de $E^p(c,T)$ pour $G = C^n$ et $p = \dim T$.*

Comme indicatrice de croissance d'un courant $T \in M_+^p(C^n)$, on utilisera (cf. §1.1) le quotient croissant $v_T(r) = (\tau_{2p} r^{2p})^{-1}\, \sigma_T(r)$, où $\sigma_T(r)$ est la masse de σ_T portée par $\|z\| \leqq r$. On note $\omega_p(r)$ l'indicatrice du courant $S^p = \sum_q N_T(Z_q^p)[Z_q^p]$ et $\omega_p(c,r)$ celle de l'ensemble analytique $E^p(c,T)$, l'indicatrice d'un tel ensemble étant par définition celle du courant d'intégration. De ce qui précède résulte.

Théorème 3.12. *L'indicatrice de S^p, soit $\omega_p(r)$ est contrôlée par celle de T. On a*

$$\omega_p(r) \leqq \nu_T(r). \tag{3.9}$$

Corollaire 3.13. *L'indicatrice de l'ensemble analytique $X = E^p(c, T)$ dans C^n est contrôlée par c et $\nu_T(r)$. On a*

$$\omega_p(c, r) \leqq c^{-1}\, \nu_T(r) \tag{3.10}$$

A partir du Corollaire (3.9) on verrait que la multiplicité $p_x(X)$ de $x \in X = E^p(c, T)$ est contrôlée en fonction de $\|x\| = r$ et de $\nu_T(r)$; ce résultat, précisons le, est obtenu seulement pour les singularités de la *composante de dimension maxima de l'ensemble de densité*, c'est-à-dire pour celle de codimension 1 si l'on a $T = \frac{1}{2\pi} dd^c V$, et $V \in PS(C^n)$.

A partir du théorème 3.12, on obtient en utilisant les résultats de [10] pour $p = n-1$ ou ceux de [13] pour p quelconque un contrôle de croissance pour un système F_j de fonctions entières telles que $E^p(c, T) = \bigcap F_j^{-1}(0)$, $1 \leqq j \leqq n+1$.

Corollaire 3.14. *L'ensemble $E^p(c, T)$ peut être défini en annulant $n+1$ fonctions entières F_j telles que si l'on pose $M_j(r) = \sup \log |F_j(z)|$ pour $\|z\| \leqq r$, on ait pour $\varepsilon > 0$ donné:*

$$c M_j(r) \leqq C(\varepsilon, r_0) \log^2 r\, \nu_T(r + \varepsilon r) \qquad \textit{pour } r > r_0 \tag{3.11}$$

où l'on choisit $r_0 > 1$ puis $C(\varepsilon, r_0)$ fonction de ε, r_0 et des dimensions. Pour $p = n-1$, on aura $X = E^{n-1}(c, T) = F^{-1}(0)$ où $F \in H(C^n)$ peut être choisie, si $0 \notin E^{n-1}(c, T)$, de manière à vérifier une majoration

$$c M_F(r) \leqq A(n, q)\, r^q \left[\int_0^r t^{-q-1}\, \nu_T(t)\, dt + r \int_r^\infty t^{-q-2}\, \nu_T(t)\, dt \right]. \tag{3.12}$$

L'entier q, genre du courant T est le plus petit entier pour lequel on a $\int_1^\infty t^{-q-2}\, \nu_T(t)\, dt < \infty$.

3.6. *Cas d'une variété de Stein G plongée dans un C^m.*

On identifiera G supposée toujours de dimension n, à son image $h(G) \subset \mathbb{C}^m$ obtenue par le plongement h dans $\mathbb{C}^m$, $m \geqq 2n+1$. Au voisinage de chaque point de G, on prendra comme coordonnées locales $z_1, \ldots, z_n$ qui sont n des coordonnées $\xi_1, \ldots, \xi_m$ de C^m. On utilisera un recouvrement assez fin de C^m, soit $\{U_j'\}$, pour que, sur G, les $G \cap U_j'$ soient des domaines de coordonnées locales $\xi_{i_1}, \ldots, \xi_{i_n}$. D'autre part on utilisera sur G la métrique induite par C^m.

Proposition 3.15. *Le plongement h: $G \to \mathbb{C}^m$ induit une application $T \to h_* T$ de $M_+^p(G)$ dans $M_+^p(\mathbb{C}^m)$ qui conserve le nombre $\nu(x, T)$. On a $\nu[h(x), h_* T] = \nu(x, T)$ pour tout $x \in G$.*

En effet, soit $\sum \alpha_j(\xi) = 1$ une partition continue de l'unité dans C^m, subordonnée au recouvrement $\{U_j'\}$, et soit $\varphi \in \mathscr{C}_0^{p,p}(\mathbb{C}^n)$. On définit dans $\mathbb{C}^m$ l'image $T' = h_* T$ par

$$T'(\varphi)=\sum_j T'(\alpha_j\varphi)=\sum_j T[h^*(\alpha_j\varphi)]$$

où h^* consiste sur un U'_j à annuler les ξ_s qui ne sont pas des coordonnées locales. Le courant T' est fermé car on a

$$\begin{aligned} bT'(\varphi)&=\sum_j bT'(\alpha_j\varphi)=\sum_j T'[d(\alpha_j\varphi)]=\sum_j T[h^*d(\alpha_j\varphi)]\\ &=\sum_j T[dh^*(\alpha_j\varphi)]=0.\end{aligned}$$

On vérifie aisément que $T'=h_*T$ est positif et qu'on a $\operatorname{supp} T'\subset h(G)$ dans $\mathbb{C}^m$. Soit σ' la mesure-trace de T' dans $\mathbb{C}^m$. La trace de la forme $\beta_p(d\xi)$ sur $h(G)\cap U'_j$ est $\beta_p(dz)$. Soit $\xi^0\in h(G)$: la boule $B(\xi^0,R)$ de $\mathbb{C}^m$ a pour trace sur $h(G)$ la boule de centre ξ^0, de rayon R dans l'espace des coordonnées nées locales z_s. On a donc pour $f\in\mathscr{C}_0(U'_j)$:

$$\sigma'(f)=\int_{C^m} T'\wedge f\beta_p(d\xi)=\int_G T[h^*(f\beta_p)]=\int T\wedge\beta_p(dz)\,h^*f=\sigma(f)$$

qui entraîne $\nu(\xi^0,T)=\nu(\xi^0,T')$ et l'énoncé.

Les ensembles $K(r)=h^{-1}[B(0,r)]$, sections de G par les boules $B(0,r)$ de C^m forment une famille de compacts d'épuisement sur G; si l'on remplace h par un autre plongement h', on modifie la famille $K(r)$, mais les topologies limites projectives qui définissent le contrôle de croissance ne sont pas modifiées; seules les indicatrices le sont.

On a alors:

Théorème 3.16. *Soit G une variété de Stein et T un courant positif fermé sur G. On calcule $\sigma_T(r)$ et $\nu_T(r)$ sur l'image $T'=h_*T$, donnée dans C^m par un plongement h de G dans C^m, l'origine de C^m étant prise en dehors de $h(G)$. On opère de même pour le courant $\sum_q N(Z_q^p)[Z_q^p]=S^p$ relatif aux cycles à densité de T qui sont de la dimension $p=\dim T$, et l'on calcule l'indicatrice $\omega_p(r)$ de S^p sur l'image h_*S^p par rapport aux boules de C^m. Alors on a $\omega_p(r)\leqq\nu_T(r)$. L'indicatrice de l'ensemble $E^p(c,T)\subset G$ vérifie le contrôle* (3.10), *$\omega_p(c,r)$ étant calculée sur l'image dans C^m.*

Ainsi les propriétés de contrôle de croissance s'étendent aux variétés de Stein, leur expression explicite faisant seule intervenir le plongement.

4. Exemples Stabilité à gauche des ensembles $E(c, T)$

4.1. Des exemples simples montrent l'instabilité à droite (c'est-à-dire pour c positif croissant) des ensembles de densité $E(c,T)$. Tout d'abord il est clair que la décomposition de $E(c,T)$ en cycles ne fait apparaître que certains des cycles Z vérifiant $N_T(Z)>0$.

Définition 4.1. On dira qu'un cycle Z sur une variété analytique G est extrémal par rapport à un courant $T\in M_+^p(G)$ si l'on a $N_T(Z)>0$ et si il n'existe pas de cycle Z' contenant Z (auquel cas on a $\dim Z'>\dim Z$) qui vérifie $N_T(Z')=N_T(Z)$.

Une décomposition de $E(c, T)$ en cycles de diverses dimensions k, $0 \leqq k \leqq p = \dim T$, est une réunion (finie sur tout compact) de cycles extrémaux; réciproquement tout cycle qui figure dans une telle décomposition pour un $c > 0$ est extrémal. Tout cycle Z pour lequel on a $\dim Z = \dim T$ et $N_T(Z) > 0$ est extrémal, car la dimension de $E(c, T)$ est au plus p.

Théorème 4.2. *En général:*

$-1°$) *Les séries* $S^k = \sum_q N_T(Z_q^k)[Z_q^k]$ *relatives aux cycles extrémaux de dimension k ne convergent pas, même sur les compacts, pour* $k < \dim T$.

$-2°$) *L'ensemble* $E'(c, T) = [x \in G;\ v(x, T) > c]$, $c > 0$, *n'est pas analytique.*

$-3°$) *Même sur les compacts, il n'existe pas de majoration de* $E^k(c, T)$ *pour* $0 \leqq k \leqq p-1$.

Il suffit de donner un contre-exemple pour $p = n-1$, dans $\mathbb{C}^n(z_1, \ldots, z_n)$ pour $T = T_V$ et $V \in PS(\mathbb{C}^n)$. On pose pour $n \geqq 3$:

$$V = \log|z_1| + \sum_{q=1}^{\infty} 2^{-q} \log|z_2 - \lambda_q| \tag{4.1}$$

où $\{\lambda_q\}$ est une suite dense sur le disque unité de $\mathbb{C}$.

Posons $X_q = [z \in \mathbb{C}^n;\ z_2 = \lambda_q]$ et soit Y défini dans $\mathbb{C}^n$ par $z_1 = 0$. L'ensemble de densité du courant $T_V = \frac{i}{\pi} \partial\bar{\partial} V$ est défini par $v(x, T_V) > 0$ et se compose de Y et des X_q. On a $v(z, T_V) = 1$ pour $[y \in Y, y \notin \bigcup_q X_q]$. On a $v(z, T_V) = 1 + 2^{-q}$ pour $z \in Y \cap X_q = Z_q^{n-2}$. Ainsi $E'(1, T_V) = [z \in \mathbb{C}^n;\ v(z, T_V) > 1]$ n'est pas un ensemble analytique dans C^n; il contient sur la boule $\|z\| \leqq 2$ *une infinité de cycles* Z_q^{n-2} *qui seront absorbés par le cycle Y de dimension supérieure quand c décroissant traverse la valeur* 1. La série $\sum (1 + 2^{-q})[Z_q^{n-2}]$ relative aux cycles extrémaux de codimension 2 diverge; la série des mesures traces n'est pas bornée sur le compact $\|z\| \leqq 2$, ce qui établit les conclusions négatives de l'énoncé. On voit de plus qu'il existe sur le compact $\|z\| \leqq 2$ une infinité de cycles extrémaux Z vérifiant $N_T(Z) \geqq c_0 > 0$.

4.2. **Définition 4.3.** On appellera valence propre relative au courant T sur G, notée $N'_T(Z)$, d'un cycle Z le nombre:

$$N'_T(Z) = N_T(Z) - \sup_j N_T(Z_j)$$

le sup étant pris pour les cycles $Z_j \in \mathfrak{N}_i(G)$ contenant Z (on a alors $\dim Z_j > \dim Z$).

On a $N'_T(Z) = 0$ si Z n'est pas extrémal. Avant d'établir la réciproque, on montrera:

Théorème 4.4. *Soit G une variété analytique complexe et $T \in M_+^p(G)$. Pour tout domaine $D \Subset G$, d'adhérence $\bar{D}$ compacte dans G et tout $c > 0$, il existe $\gamma = \gamma(c, D) > 0$, tel qu'on ait:*

$$E(c, T) \cap D = E(c', T) \cap D \text{ pour tout } c' \text{ vérifiant } c - \gamma < c' \leqq c.$$

Soit D_1 un domaine vérifiant $D \Subset D_1 \Subset G$. La fonction $v(x, T)$ est semi-continue supérieurement et $E(c, T) \cap D_1 = [x \in D_1;\ v(x, T) \geqq c]$, $c > 0$, est son ensemble de niveau; sur un compact $K \subset D_1$ qui ne coupe pas $E(c, T)$, *la fonction* $v(x, T)$ a un maximum strictement inférieur à c. Si l'énoncé est en défaut, il existe donc une suite infinie $S_0 = \{x_q^0\}$ dont les points vérifient $x_q^0 \in D \setminus E(c, T)$, et pour laquelle les $v(x_q^0, T) < c$ sont une suite croissante de nombres positifs tendant vers c, tendis que x_q^0 tend vers un point $\xi \in E(c, T) \cap \bar{D}$. On va montrer que ceci est impossible, en formant des suites extraites de S_0, soit $S_s = \{x_q^s\}$ infinies qui ont les mêmes propriétés mais appartiennent à des ensembles analytiques dont la dimension est strictement décroissant. On obtiendra dans D_1 un ensemble de dimension zéro contenant une telle suite sur $\bar{D}$, d'où contradiction.

Considérons d'abord le point $x_1^0 \in S_0$; on a $v(x_1^0, T) = c_1 < c$ puisque $x_1^0 \in D$ n'appartient pas à $E(c, T)$. Dans D_1, $E(c_1, T)$ est analytique et contient S_0, mais seule une partie finie de S_0 peut appartenir à la composante $E^0(c_1, T)$ de dimension zéro, sur $\bar{D}$. Il existe donc un cycle dans $D_1 \cap E(c, T)$ soit Z_1 de dimension $k \geqq 1$ qui contient une sous-suite infinie, soit $S_1 = \{x_q^1\}$ de S_0. On a $N_T(Z_1) < c$, car Z_1 n'appartient pas à $E(c, T)$. On choisit alors c_2 vérifiant $N_T(Z_1) < c_2 < c$. Comme plus haut, $E(c_2, T)$ contient un cycle, soit Z_2 porteur d'une sous-suite infinie de S_1, soit S_2. On a $Z_2 \neq Z_1$ car on a $N_T(Z_1) < c_2$ et $N_T(Z_2) \geqq c_2$. Ainsi S_2 appartient à $Z_1 \cap Z_2$ dont la dimension maxima est au plus $k-1$. Il existe donc un cycle Z_2' de dimension au plus $k-1$, qui contient une sous-suite infinie de S_2, soit $S_3 = \{x_q^3\}$ sur laquelle $v(x_q^3, T)$ est fonction croissante et tend vers c, tandis que $\lim\limits_q x_q^3 = \xi$. On choisit maintenant c_3 vérifiant $N_T(Z_2') < c_3 < c$ et on obtient dans $E(c_3, T)$ un nouveau cycle Z_3 contenant une partie infinie de S_3. On a $\dim[Z_2' \cap Z_3] \leqq k-2$, et cette intersection contient donc un cycle Z_3' qui porte une sous-suite infinie, soit S_4 de S_3, et pour lequel on a $\dim Z_3' \leqq k-2$.

Finalement après un nombre fini d'extractions on obtient une suite infinie $\{x_n^r\} = S_r$ de points dont une partie infinie appartient à un ensemble analytique de $E(c_{r+1}, T)$, $c_{r+1} < c$, et plus précisément à sa composante $E^0(c_{r+1}, T)$ de dimension zéro. On aboutit ainsi à une contradiction, cette composante étant constituée d'un nombre fini de points sur le compact $\bar{D}$.

Corollaire 4.5. *Si Z est un cycle extrémal, on a $N_T'(Z) > 0$.*

En effet on a $N_T(Z) = c > 0$ et d'après l'énoncé précédent, il existe $\gamma > 0$, un point $x_0 \in Z$ et un voisinage compact U de x_0 dans G tels que on ait $E(c', T) \cap U = E(c, T) \cap U$ pour $c - \gamma < c' \leqq c$.

Un cycle Z' qui contient Z vérifie $N_T(Z') < N_T(Z)$ si Z est extrémal dans G. Il existe dans $U \cap Z'$ un point x où l'on a $v(x, T) < c$. On a alors aussi $x \notin E(c', T)$ qui entraine $N_T(Z') < c'$, pour tout c' de l'intervalle $c - \gamma < c' \leqq c$. Finalement on a $N_T(Z') \leqq c - \gamma$, et $N_T'(z) \geqq \gamma > 0$ d'après la définition 4.3. pour tout cycle Z' qui contient Z. *Il y a donc identité entre les cycles extrémaux sur G et les cycles pour lesquels on a $N_T'(Z) > 0$.*

Définition 4.6. On dira que $c > 0$ est valeur de stabilité sur G pour $E(c, T)$ s'il existe un intervalle $\gamma > 0$, $0 < \gamma \leqq c_c$ tel que l'on ait $E(c', T) = E(c, T)$ sur G pour $c - \gamma < c' \leqq c$.

Le théorème 4.4. montre que les $E(c,T)$ sont *localement stables à gauche* (c'est-à-dire pour c décroissant); l'intervalle de stabilité $\gamma(c,K)$, pour un compact K donné pouvant tendre vers zéro quand c décroissant tend vers une valeur $c_0>0$ (cf. l'exemple donné en 4.2. où l'on a $c_0=1$).

Remarque 4.7. On verra plus loin (Remarque 4.10) que pour $k<p$ les séries $S'^k \leqq \sum_q N'_T(Z^k_q)[Z^k_q]$ relatives aux cycles extrémaux de dimension inférieure à $p=\dim T$, munis des valences propres ne sont pas contrôlées par la seule donnée de l'indicatrice de T. L'énoncé suivant généralise cependant une propriété des cycles Z^p de dimension maxima.

Théorème 4.8. *Soit G une variété analytique complexe et $T\in M^p_+(G)$. Si G est dénombrable à l'infini, alors sur G les cycles extrémaux relatifs au courant T sont en infinité dénombrable.*

Cet énoncé sera une conséquence de l'énoncé suivant, plus précis:

Théorème 4.9. *Soit $D\Subset G$ un domaine relativement compact sur la variété G où est défini un courant T positif fermé.*

1°) *Le nombre $v(x,T)$ prend sur D son maximum α fini, $\alpha\geqq 0$. L'intervalle $]0,\alpha]$ est recouvert par une famille dénombrable $\mathscr{F}=\{I_k\}$ d'intervalles $c'_k<c\leqq c''_k$ fermés a droite, ouverts à gauche, de longueur $\ell(I_k)>0$, deux à deux sans point commun et l'ensemble $L\subset]0,\alpha]$ des extrémités des I_k est un ensemble dénombrable fermé sur $]0,\alpha]$.*

2°) *Quand c parcourt un segment I_k, l'ensemble $E(c,T)\cap D$ ne varie pas. Si Z est un cycle extrémal dans D, on a $N_T(Z)\subset L$. Réciproquement à chaque valeur numérique $c\in L$ (ensemble des valeurs critiques de T dans D), il correspond une famille finie non vide de cycles extrémaux rencontrant D pour lesquels on a $N_T(Z)=c$.*

Montrons d'abord que l'énoncé 4.9. entraîne 4.8. Si Z est un cycle extrémal dans G, la section $Z\cap D$, pour $D\Subset G$, est formée d'une famille finie de cycles et chacun d'eux est extrémal dans D (conséquence immédiate de la définition 4.1.). D'après le théorème 4.9., l'ensemble L des valeurs $N_T(Z)$, pour $Z\subset D$, est dénombrable; et à $c\in L$ ne correspond qu'un nombre fini de cycles extrémaux dans D. La collection des cycles extrémaux de G qui intersectent D est donc dénombrable; l'énoncé 4.8. en résulte, G étant dénombrable à l'infini.

Montrons le théorème 4.9: d'après 4.4., et $D\Subset G$, l'ensemble L des $c>0$ pour lesquels on a $E(c,T)\cap D=\emptyset$ est un segment $]\alpha,+\infty]$ ouvert à gauche; il est non vide car $v(x,T)$ est borné sur $\bar{D}$; son complémentaire $W=]0,\alpha]$ est fermé à droite et $v(x,T)$ prend donc sur l'ouvert D sa valeur maxima α. Soit $c_0\in W$: l'ensemble $E(c,T)\cap D$ étant fonction monotone de $c>0$, l'ensemble des $c>0$ pour lesquels on a $E(c,T)\cap D=E(c_0,T)\cap D$ est un segment $I(c_0)$ de R^+; il est ouvert à gauche et fermé à droite d'après le théorème 4.4. On forme ainsi une collection de segments $I(c)$ qui recouvrent W. Deux d'entre eux, $I(c')$, $I(c'')$, ou bien coïncident, ou bien sont sans point commun. En effet $c_1\in I(c')\cap I(c'')$ entraîne $E(c,T)\cap D=E(c_1,T)\cap D$ sur $I(c')\cup I(c'')$. D'où: $I(c')=I(c_1)=I(c'')$. Finalement les $I(c)$ ont une longueur totale α; chacun est de longueur non nulle; ils forment donc une collection dénombrable I_k. Soit L l'ensemble dénombrable des extrémités droites des I_k; c'est un ensemble fermé sur W, car si ξ est limite d'une suite $c_m \in L$, on a $0 \leqq \xi \leqq \alpha$ et si l'on a $\xi > 0$, le segment $I(\xi)$ n'a pas ξ comme point intérieur, ξ est donc extrémité droite de $I(\xi)$, ce qui entraîne $\xi \in L$. Ainsi les $\mathring{I}_k$ sont les segments ouverts contigus à un ensemble fermé sur W. Les exemples plus haut montrent que l'ensemble L' des points limites de L comprend en général $c=0$ en plus de L.

Par contruction, $E(c, T) \cap D$ demeure constant sur un segment I_k et L est l'ensemble des discontinuités de $E(c, T)$. Si Z est un cycle extrémal dans D, on a $0 < N_T(Z) \leqq \alpha$ et de plus $E(c', T) \cap D$ ne contient plus Z pour $c' > N_T(Z)$, donc c' n'est pas intérieur à un segment I_k, ce qui établit $N_T(Z) \subset L$. Inversement si l'on a $c_0 \in L$, pour $c > c_0$, l'ensemble $E(c, T) \cap D$ est un vrai sous-ensemble analytique de $E(c_0, T)$. Il existe donc dans la décomposition (M) de $E(c_0, T)$ en composantes irréductibles dans D des cycles Z (nécessairement en nombre fini, d'après $D \Subset G$), qui n'appartiennent pas à $E(c', T)$ pour $c' > c_0$. On a alors $N_T(Z) = c_0$ pour cette famille finie, non vide, de cycles. Ils sont d'autre part extrémaux puisqu'ils figurent dans (M), ce qui achève d'établir le théorème 4.9.

Remarque 4.10. 1°) Pour tout $c > 0$, les composantes de $E(c, T)$ irréductibles dans G sont des cycles extrémaux; ceux-ci apparaissent (ou disparaissent) quant c traverse une valeur $c \in L$.

2°) Si l'on a une exhaustion de G par des D_n croissants, on a $L(D_1) \subset L(D_2)$ pour $D_1 \subset D_2 \Subset G$ et $L(G) = \lim L(D_n)$.

3°) Le caractère extrémal d'un cycle Z est local; il en est de même de sa valence propre $N'_T(Z)$: soit ω un ouvert non vide de G et $Z \cap \omega \neq \emptyset$; les divers cycles Z'_j de la décomposition de $Z \cap \omega$ vérifient $N'_T(Z'_j) = N'_T(Z)$.

4°) Même si on remplace $N_T(Z)$ par $N'_T(Z)$, la série S^k du théorème 4.2. pour $k < p = \dim T$ n'est pas contrôlée par la croissance de T. Exemple: $G = C^2(z_1, z_2)$ et $T = \dfrac{i}{\pi} \partial \bar{\partial} V$, où l'on a $V = \frac{1}{2} \log[|F_1|^2 + |F_2|^2]$; on suppose que F_1 et F_2 sont deux fonctions entières dans C^2 n'ayant que des zéros communs isolés $\{a_j\} \in C^2$. Alors pour les courant T les seuls cycles Z à valence $N_T(Z) > 0$ sont les points a_j, et pour ces cycles on a $N_T(a_j) = N'_T(a_j)$. Un exemple classique [3] de M. Cornalba et B. Schiffmann montre qu'il n'existe pas de contrôle du nombre $n(r)$ des a_j dans la boule $B(0, r)$ par la seule croissance de $|F_1|^2 + |F_2|^2$; dans ce cas il n'existe donc pas un contrôle de la série S^k, pour $k=0$, relative à T, par l'indicatrice de croissance du courant T dans C^n.

5. Existence d'une hypersurface $Y_c = F_1^{-1}(0)$ contenant tout l'ensemble $E(c, T)$ avec contrôle asymptotique

5.1. La démonstration comportera ici deux parties. D'abord on passera du courant donné $T \in M^p_+(C^n)$ à un courant $\theta = \dfrac{1}{2\pi} dd^c V$, $V \in PS(C^n)$ tels qu'on ait $\nu(x, \theta) = \nu(x, T)$ pour tout x. Cette opération $M^p_+(C^n) \to M^{n-1}_+(C^n)$ est faite avec contrôle de croissance; elle utilise une propriété simple que nous avions donnée dans [7], puis la belle construction faite par H. Skoda [13] de V par des potentiels. On précisera ici quelques points implicites dans [13], afin de montrer l'existence d'un vrai contrôle dans cette première partie. La seconde qui fait passer de θ et de V à Y_c et F_1 est une simple application du théorème (noté ici H.B.S.) de Hörmander-Bombieri et de la résolution du $\bar{\partial}$ dans L^2; on l'utilisera sous la forme précise donnée par H. Skoda dans [14].

5.2. Montrons d'abord:

Théorème 5.1. *Soit $T \in M_+^p(\mathbb{C}^n)$ et R_0 donné, $R_0 > 0$. Il existe $V \in PS(C^n)$ qui vérifie les conditions suivantes, où $\lambda(V, x, r)$ désigne la moyenne de V sur la sphère de centre x, de rayon r:*

a) *On a $v(x, T_V) = v(x, T)$ pour tout $x \in C^n$.*

b) *Il existe $C(p) > 0$ et une minoration:*

$$\lambda(V, 0, \tfrac{1}{2}) \geqq -C(p)\, v_T(6). \tag{5.1}$$

c) *$M_V(r) = \sup V(z)$ pour $\|z\| \leqq r$ vérifie une majoration*

$$M_V(r) \leqq g(r) \tag{5.2}$$

où $g(r)$ est calculable (sous diverses formes) à partir de $v_T(r)$ et de n, p.

Corollaire 5.2. *Soit $T \in M_+^p(\mathbb{C}^n)$: il existe un courant $\theta \in M_+^{n-1}(\mathbb{C}^n)$ qui vérifie $v(x, \theta) = v(x, T)$ avec contrôle de croissance dans le passage $M_+^p(\mathbb{C}^n) \to M_+^{n-1}(\mathbb{C}^n)$. On a $\theta = T_V$ où V est donné par l'énoncé précédent.*

Montrons que le théorème 5.1. entraîne le Corollaire. On a pour $k > 1$ et $r > \frac{1}{2}$:

$$\begin{aligned} v_\theta(r) &= \frac{\partial}{\partial \log r} \lambda(V, 0, r) \leqq (\log k)^{-1} [\lambda(V, 0, kr) - \lambda(V, 0, \tfrac{1}{2})] \\ &\leqq (\log k)^{-1} [g(kr) + C(p)\, v_T(6)]. \end{aligned} \tag{5.3}$$

Ainsi si $v_T(r)$ est majoreé, $v_\theta(r)$ l'est aussi, par une fonction calculable à partir de $v_T(r)$.

5.3. La démonstration du théorème 5.1. repose alors sur deux propriétés;

a) Soit σ_T la mesure trace de T et soit $\eta(z) \in \mathcal{D}(C^n)$, $0 < \eta(z) \leqq 1$, et $\eta(y) = 1$ dans la boule $\|z\| \leqq R$. On pose

$$U_\eta(z) = -\omega_{2p}^{-1} \int d\sigma_T(a) \|a - z\|^{-2p} \eta(a). \tag{5.4}$$

Alors la fonction R^{2n}-sousharmonique $U_\eta(z)$ vérifie en tout point x intérieur à $B(0, R)$

$$v(x, T) = \lim_{r=0} (\log r)^{-1} \lambda(U_\eta, x, r) \tag{5.5}$$

où l'on suppose $0 < r < \inf(1, R - \|x\|)$. Pour la démonstration, on renvoie à [7, p. 138]. Il est clair qu'on ne modifie pas la limite (5.5) en ajoutant à U_η une fonction continue au voisinage du point x.

b) On considère avec H. Skoda [13] auquel on renvoie pour les détails une partition de l'unité ($\mathscr{C}^\infty$), soit $\Sigma b_j(z) = 1$. Soit $a(z) \in \mathcal{D}(C^n)$, $0 \leqq a(z) \leqq 1$, vérifiant $a(z) = 1$ pour $\|z\| \leqq 1$ et $a(z) = 0$ pour $\|z\| \geqq 1 + \varepsilon$, $0 < \varepsilon < 1$. On pose pour j entier

$j \geqq 1$: $a_j(z) = a(zj^{-1})$, puis $b_j = a_j - a_{j-1}$ pour $j \geqq 2$ et $b_1 = a_1$. On choisit alors $\eta_j(z) = a[(j + 2\varepsilon j)^{-1} z]$ et enfin:

$$U_j(x) = -\omega_{2p}^{-1} \int_{C^n} \|z - a\|^{-2p} \eta_j(a)\, d\sigma_T(a), \tag{5.6}$$

$$U = \Sigma b_j U_j. \tag{5.7}$$

Le noyau $\ell(a,z) = -\omega_{2p}^{-1} \|a - z\|^{-2p}$ est R^{2n}-sousharmonique et l'on a donc pour sa moyenne prise sur $\|z\| = \rho$:

$$\lambda[\ell(a,z), 0, \rho] \geqq \omega_{2p}^{-1} \sup[-\rho^{-2p}, -\|a\|^{-2p}]. \tag{5.8}$$

Dans la boule $\|z\| < 1$, $U(x)$ est égal à $U_1(x)$ augmenté d'un potentiel de masses a pour lesquelles on a $\|a\| \leqq 1 + 5\varepsilon < 6$. On en déduit à partir de (5.8), pour la moyenne de U sur $\|z\| = 1/2$ une minoration

$$\lambda(U, 0, \tfrac{1}{2}) \geqq -\omega_{2p}^{-1} \left[2^{2p} \sigma_T(\tfrac{1}{2}) + \int_{1/2}^{6} t^{-2p} d\sigma_T(t) \right]$$

ou en remplaçant $\sigma_T(t) = \tau_{2p} t^{2p}$ après intégration par parties:

$$\lambda(U, 0, \tfrac{1}{2}) \geqq -C(p)\, v_T(6) \tag{5.9}$$

où $C(p) = \dfrac{1}{2p} + \log 12$ si $p > 0$. Si $p = 0$ on utilise le noyau $\log \|a - z\|$ et on remplace $\dfrac{1}{2p}$ par 1.

Le potentiel (5.7) vérifie (5.5) car tout $x \in C^n$ est centre d'une boule dans laquelle $U(z)$ ne diffère du potentiel de σ_T que par une fonction ($\mathscr{C}^\infty$). On construit alors, comme dans [13], $W \in PS(C^n) \cap \mathscr{C}^\infty$ vérifiant $0 = W(0) \leqq \lambda(W, 0, \tfrac{1}{2})$ et telle qu'on ait $V = U + W \in PS(C^n)$. Alors V vérifie a) de l'énoncé. D'autre part on a (au choix) les contrôles suivants qui précisent $g(r)$ et c), cf. [13]:

$$M_V(r) \leqq C(\varepsilon) \log^2(1 + r)\, v_T(r + \varepsilon r) \text{ pour tout } \varepsilon > 0 \quad \text{et } r > r_0(\varepsilon) \tag{5.10}$$

$$M_V(r) \leqq C(\varepsilon, d)(1 + r)^m \int_1^{1+r} v_T(t + \varepsilon t) t^{-m-1} dt$$

pour tout $\varepsilon > 0$, $m > 0$.

5.4. Il y a donc contrôle dans le passage de $T \in M_+^p(\mathbb{C}^n)$ à $\theta = \dfrac{1}{2\pi} dd^c V$ grâce à la minoration (5.1) et à la majoration (5.2). On peut l'obtenir plus directement si $v_T(r)$ est à croissance lente, le genre $q \geqq 0$ étant le plus petit entier pour lequel l'intégrale $\int_1^\infty v_T(t) t^{-q-2} dt$ converge.

Proposition 5.3. *Si l'indicatrice $\nu_T(r)$ du courant $T \in M_+^p(C^n)$ est de genre* 0 *ou* 1, *il existe un noyau symétrique $K(t, r)$ tel qu'on ait*

$$\nu_\theta(r) = r \int_0^\infty t^{2p} \nu_T(t) K(t, r) \, dt. \tag{5.11}$$

Faisons la démonstration pour $q=0$ en notant $h_p(t,r)$ la moyenne du noyau $-\omega_{2p}^{-1} \|a-z\|^{-2p}$ pour z parcourant la sphère $\|z\|=r$; $h_p(t,r)$ est fonction symétrique de $\|a\|=t$, $\|z\|=r$. On a

$$h_p(t,r) = C_{n,p} \int_0^\pi (r^2 + t^2 - 2tr\cos\varphi)^{-p} \sin^{2n-1}\varphi \, d\varphi \tag{5.12}$$

où $C_{n,p} = \omega_{2p}^{-1} \omega_{2n}^{-1} \omega_{2n-1}$ et $\omega_m = 2\pi^{m/2} \left[\Gamma\left(\frac{m}{2}\right)\right]^{-1}$.

On a si $\nu_T(r)$ est de genre $q=0$, (cf. [1], [10], [13]):

$$V(z) = \omega_{2p}^{-1} \int_{C^n} [(1 + \|a\|^{2p})^{-1} - \|a-z\|^{-2p}] \, d\sigma_T(a) \tag{5.13}$$

et $V \in PS(C^n)$ vérifie $\nu(x, T_V) = \nu(x, T)$. La moyenne $\lambda(V, 0, r)$ s'écrit

$$\lambda(V, 0, r) = \int_0^\infty t^{-1} \nu_T(t) \left[\frac{t^{4p}}{(1+t^{2p})^2} + \frac{t^{2p+1}}{2p} \frac{\partial h_p}{\partial t}(t,r)\right] dt. \tag{5.14}$$

On a d'autre part $\dfrac{\partial}{\partial \log r} \lambda(V, 0, r) = \nu_\theta(r)$ indicatrice du courant θ qu'on cherche associé à V. Il en résulte en utilisant (5.13) l'expression (5.11) où le noyau $K(t, r)$ a la valeur

$$K(t, r) = \frac{1}{2p} \frac{\partial^2 h_p}{\partial t \, \partial r}(t, r). \tag{5.15}$$

On vérifie que pour $p \leqq n-2$, $\dfrac{\partial^2 h_p}{\partial t \, \partial r}(t, r)$ est continu de t pour r fixé, $r > 0$, $0 < t < \infty$. Pour $p = 0$, on remplace dans (5.15) p par l'unité et la quantité à intégrer dans (5.12) par $\log\|a-z\| - \log(1+\|a\|)$.

Pour $q=1$ on obtient encore $V \in PS(C^n)$ donné par un potentiel où le crochet de (5.13) a un terme supplémentaire. Ce terme disparait toutefois dans le calcul de $\lambda(V, 0, r)$ et (5.11) demeure valable avec la valeur (5.15) de $K(t, r)$. On a d'autre part pous $r < r_0$, une majoration $K(t, r) = O\left(\dfrac{1}{t^{2p+3}}\right)$ qui assure la convergence de (5.11).

Remarque 5.4. Soit $T \in M_+^p(C^n)$: alors l'ensemble des $x \in C^n$ où l'intégrale $\int_0^\varepsilon t^{-1} \nu(x, T, t) \, dt$ diverge est pluripolaire.

En effet c'est èvident pour $T = T_V$, car la divergence de l'intégrale équivaut à $V(x) = -\infty$. La propriété locale est évidemment valable sur une variété G.

Les résultats obtenus concernant le contrôle de croissance dans le passage de $T \in M_+^p(C^n)$ à $\theta = T_V$ avec conservation du nombre $\nu(x, T)$ s'étendent comme plus haut à une variété de Stein.

5.4. Rappelons l'énoncé suivant
(H.B.S.) – Soient $\varphi \in PS(\mathbb{C}^n)$, $\varepsilon > 0$ et soit $x_0 \in \mathbb{C}^n$ un point au voisinage duquel $e^{-\varphi}$ est sommable. On pose:

$$\psi(x) = \varphi(x) + (n+\varepsilon)\log(1 + \|x\|^2). \tag{5.16}$$

Alors il existe $F \in H(C^n)$ *qui vérifie* $F(x_0) = 1$ *et*

$$\|F\|_\varphi^2 = \int_{C^n} |F(x)|^2 e^{-\psi(x)} dv < \infty. \tag{5.17}$$

Rappelons aussi [13]: l'ensemble A_φ des points au voisinage desquels $e^{-\varphi}$ n'est pas sommable est un sous-ensemble analytique etl'on a

$$E(2n, T_\varphi) \subset A_\varphi \subset E(2, T_\varphi) \subset [x; \varphi(x) = -\infty]. \tag{5.18}$$

On a évidemment $A_\varphi = A_\psi$.

Proposition 5.5. *Soint* $V \in PS(\mathbb{C}^n)$ *et* $c > 0$ *donnés. Pour tout* $\varepsilon > 0$, *il existe une hypersurface* $Y_c = F_1^{-1}(0)$, *où* $F_1 \in H(C^n)$, *qui contient l'ensemble de densité* $E(c, T_V)$ *et vérifie* $\|F_1\|_\psi = 1$ *selon* (5.2), *avec*

$$\psi(x) = \frac{2n}{c} V(x) + (n+\varepsilon)\log(1 + \|x\|^2) \tag{5.19}$$

et $M_1(r) = \sup |F_1(x)|$ *pour* $\|x\| \leqq r$ *vérifie au choix l'une des majorations suivantes:*

a) *pour tout* $\varepsilon' > 0$, *on a*

$$M_1(r) \leqq \frac{n}{c} M_V(r + \varepsilon') + (n+\varepsilon)\log(1 + r) + C(n, \varepsilon') \tag{5.20}$$

où $C(n, \varepsilon') = n\varepsilon' - n\log\varepsilon' - \frac{1}{2}\log\tau_{2n}$.

b) *pour tout* $\alpha > 1$, *on a pour* $r > 1$:

$$M_1(r) \leqq \frac{n}{c} M_V(\alpha r) + \varepsilon \log r + C(n, \varepsilon, \alpha) \tag{5.21}$$

pour $C(n, \varepsilon, \alpha) = \frac{1}{2}(n+\varepsilon)\log 2 - n\log(\alpha - 1) - \frac{1}{2}\log\tau_{2n}$.

On obtient donc un contrôle pour F_1. La démonstration de la proposition 5.5 s'appuie sur (H.B.S.): on a choisi le coefficient $\frac{2n}{c}$ de $V(z)$ de manière à obtenir $\nu(z, T_\psi) \geqq 2n$ pour $z \in E(c, T_V)$, ce qui oblige dans (5.17) la fonction entiére F_1 à s'annuler sur cet ensemble. De la convergence de (5.17) on déduit par application du lemme de Schwarz la majoration

$$|F_1(z)| \leqq g(z)\|F_1\|_\psi \tag{5.22}$$

avec $\log g(z)=\frac{1}{2}M_\psi(\|x\|+a)-n\log a-\frac{1}{2}\log\tau_{2n}$. Quitte à multiplier F_1 par une constante on peut supposer $\|F_1\|_\psi=1$, ce qui conduit aux majorations (5.20) et (5.21) en posant soit $a=\varepsilon'$, soit $r+a=\alpha r$; la dernière convient aux croissances minimales c'est-à-dire du type $M_V(r)\sim a\log, r$. On obtient finalement la majoration de M_1(r) en substituant à $M_V(r)$ l'une au choix des majorations (5.10), ou celle plus précise qu'on déduira de la Proposition 5.3 si $v_T(r)$ est de genre 0 ou 1. Donnons seulement un Corollaire.

Corollaire 5.6. *Si $T\in M_+^p(C^n)$ a une indicatrice d'ordre fini $\rho\geqq 0$, il en est de même de F_1 et du courant d'intégration $[Y_c]$, le type (maximum, moyen, nul) étant le même si ρ n'est pas entier.*

Remarque 5.7. Les majorations de $M_1(r)$ donnent des majorations de l'indicatrice $\omega_c(r)$ de Y_c, d'après $\omega_c(r)=\frac{\partial}{\partial\log r}\lambda(\log|F_1|,0,r)$. On a pour tout $k>1$:

$$\omega_c(r)\leqq(\log k)^{-1}[M_1(kr)-\lambda(\log|F_1|,0,1)]$$

On obtient un contrôle asymptotique du type indiqué à la Remarque 2.2 du §2.1.

Remarque 5.8. Soit $T=T_V$. On a donné d'autre part au §3 un vrai contrôle de la composante de codimension un de $E(c,T)$ alors qu'on n'obtient ici qu'un contrôle *asymptotique* pour Y_c de codimension 1 *qui contient tout* $E(c,T)$. Un problème ouvert se ramène à ceci: existe-t-il une suite bornée $T_q\in M_+^{n-1}(\mathbb{C}^n)$, pour laquelle la présence des composantes de «petite» dimension de $E(c,T_q)$ entraîne l'impossibilité d'obtenir un plongement $E(c,T_q)\subset Y_{c,q}$ avec $[Y_{c,q}]\subset M_+^{n-1}(C^n)$ demeurant dans un ensemble borné (donc d'indicatrice $\omega_c(r)$ majorée pour tout $r>0$ indépendamment de q).

6. Utilisation d'une hypothèse de stabilité

L'exemple donné au §4 prouve la non existence d'un contrôle même asymptotique des composantes $E^k(c,T)$ de $E(c,T)$, en fonction de la seule donné $[c,v_T(r)]$ lorsqu'on a $k<\dim T$. Toutefois on montrera qu'il existe un contrôle d'un système d'équations définissant $E(c,T)$, quand on fait une hypothèse (St_γ) exprimant la stabilité à gauche de $E(c,T)$ non plus localement mais sur tout C^n.

(St_γ). *Il existe γ, $0\leqq\gamma<c$ tel que pour tout c', vérifiant $0\leqq c-\gamma<c'\leqq c$, on ait $E(c',T)=E(c,T)$.*

On traitera le cas $T=T_V$, $V\in PS(C^n)$, le cas général d'une variété de Stein et d'un courant de dimension $p<n-1$ se ramenant comme on l'a vu à celui-là. On utilisera une transformation dans le cône $PS(\mathbb{C}^n)$ introduite par C.O. Kiselman [5]: $V\to V_\delta$ définie par

$$V_\delta(z)=\inf_{r>0}[\lambda(V,z,r)-\delta\log r],\qquad \delta\geqq 0. \tag{6.1}$$

Elle abaisse en tout point x le nombre $\nu(x, T_V)$. Soit T et T_δ relatifs à V et à V_δ. On a:

$$\nu(x, T_\delta) = \sup[\nu(x, T) - \delta, 0]. \tag{6.2}$$

On choisit $\mu > 0$, $\delta > 0$ de manière que:

a) $\exp[-\mu V_\delta]$ soit non sommable au voisinage de tout point $x \in E(c, T_V)$.

b) que cette fonction soit sommable au voisinage de tout point $x' \notin E(c, T_V)$, ce qui conduit à réaliser $\nu(x', \mu T_\delta) < 2$

c) on cherche de plus à minimiser la croissance de μV_δ. On a pour $\|x\| \leqq r$, d'après (6.1):

$$\mu V_\delta(x) \leqq \mu \lambda(V, x, r_0) - \delta \log r_0 \leqq \mu M_V(r + r_0) + C(r_0, \delta). \tag{6.3}$$

La condition a) est vérifiée dès qu'on a $\mu(c - \delta) \geqq 2n$ et on minimise μ en prenant l'égalité. Si l'on a $\gamma = c$ (stabilité totale à gauche), on prendra $\delta = 0$, $\mu = 2nc^{-1}$. Si l'on a $\gamma < c$, il existe a, $0 < a < 1$ vérifiant $\delta = c - \dfrac{n}{n-1+a}\gamma > 0$ et l'on prend cette valeur qui conduit à $\mu = 2(n-1+a)\gamma^{-1}$. On a pour $x \in E(c, T_V)$ et $W = \mu V_\delta$: $\nu(x, T_W \geqq 2n)$, tandis qu'on a $\nu(x, T_W) \leqq \mu(c - \delta - \gamma) = 2(1-a) < 2$ pour $x \notin E(c, T_V)$. On utilise alors la fonction plurisousharmonique poids:

$$\varphi(z) = 2\gamma^{-1}(n-1+a)\, V_\delta(z) + (n+\varepsilon) \log(1 + \|z\|^2)$$

et l'espace de Hilbert $H_\varphi \subset H(C^n)$ de fonctions entières F défini par

$$\|F\|_\varphi^2 = \int_{C^n} |F(z)|^2 e^{-\varphi(z)} d v(z) < \infty.$$

Sur H_φ la forme linéaire $\hat{x}(F) = F(x)$, $x \in C^n$, est continue d'après (5.22). Alors ayant déterminé $F_1 \in H(C^n)$ nulle sur $E(c, T_V)$ comme au §5, on considère dans l'ensemble analytique $X_1 = F_1^{-1}(0)$ les composantes $X_{1,s} \not\subset E(c, T_V)$; on choisit un point $x_{1,s} \in X_{1,s}$, $x_{1,s} \notin E(c, T_V)$. Les hyperplans $\hat{x}_{1,s}(F) = 0$ sont fermés dans H_φ; leur réunion est maigre; il existe donc $F_2 \in H_\varphi$ qui vérifie $F(x_{1,s}) \neq 0$ pour tout s et $\|F_2\|_\varphi = 1$. On a alors $E(c, V) \subset [F_1^{-1}(0) \cap F_2^{-1}(0)] = X_2$. On continue en éliminant de X_2 les composantes de codimension 2 étrangères à $E(c, T_V)$. On obtient ainsi $F_1, \ldots, F_{n+1}$ dans H_φ, vérifiant $\|F_k\|_\varphi = 1$ et dont $E(c, T_V)$ est exactement l'ensemble des zéros communs. D'où:

Théorème 6.1. *Soit $T = T_V \in M_+^{n-1}(C^n)$ et $c > 0$; si T et c vérifient l'hypothèse de stabilité (St_γ) pour un $\gamma > 0$, $0 < \gamma \leqq c$, alors l'ensemble de densité $E(c, T)$ est représentable par $n+1$ équations $F_j = 0$ où F_1 est obtenu avec le contrôle donné à la proposition 5.5, et où pour $2 \leqq j \leqq n+1$, F_j vérifie le même contrôle si $\gamma = c$. Si l'on a $0 < \gamma < c$, les F_j pour $j \geqq 2$ vérifient un contrôle obtenu en remplaçant dans (5.19) et (5.20) le coefficient $\dfrac{n}{c}$ par $\dfrac{n-1+a}{\gamma}$ où l'on prend a le plus petit possible vérifiant $0 < a < 1$ et $c(1-a) < n(c-\gamma)$.*

Remarque 6.2. a) On n'obtient donc avec l'hypothèse de stabilité qu'un contrôle des $F_j \in H(C^n)$ définissant $E(c, T_V)$: l'exemple donné dans [3] montre la non existence d'un contrôle asymptotique de $E(c, T_V)$, même avec l'hypothèse (St_1).

b) Le contrôle de F_1 n'utilise pas l'hypothèse de stabilité; mais on l'utilisera pour majorer F_1 si $\frac{n-1}{\gamma}$ est inférieur à $\frac{n}{c}$.

7. Cas des courants «algébriques» ou d'indicatrice bornée

7.1. Soit $T \in M_+^p(C^n)$ d'indicatrice bornée, et $\lim_{r=\infty} \nu_T(r) = a > 0$. Alors la proposition 5.3 et (5.13) donnent V de la classe de croissance minimale:

$$\lim_{r=\infty} M_V(r)(\log r)^{-1} = a; \quad \lim \nu_\theta(r) = a \tag{7.1}$$

où le courant $\theta = T_V \in M_+^{n-1}(C^n)$ vérifie $\nu(x, \theta) = \nu(x, T)$. On note $S_a(C^n)$ la classe des fonctions $V \in PS(C^n)$ qui vérifient la majoration (7.1) pour $M_V(r)$. Pour une étude de ces classes cf. [9]. Si l'on a $V_1 \in S_a(C^n)$ et $V_2 \in S_b(C^n)$ on a $V_1 + V_2 \in S_{a+b}(C^n)$. Si l'on a $V \in S_a(C^n)$ on a aussie

$$a = \lim_{r=\infty} (\log r)^{-1} \lambda(V, 0, r) \tag{7.2}$$

On notera $S_{ah}(C^n)$ la sous-classe vérifiant $V(uz) = a \log|u| + V(z)$, pour tout $x \in C^n$ et tout $u \in \mathbb{C}$.

A $V \in S_a(C^n)$, on associe $\tilde{V} \in S_{ah}(C^{n+1})$ selon

$$\begin{aligned} \tilde{V}(z_0, z_1, \ldots, z_n) &= a \log|z_0| + V(z_1 z_0^{-1}, \ldots, z_n z_0^{-1}) \\ V(z_1, \ldots, z_n) &= \tilde{V}(1, z_1, \ldots, z_n). \end{aligned} \tag{7.3}$$

$\tilde{V}$ majoré au voisinage des points vérifiant $z_0 = 0$, s'étend (avec valeur $-\infty$ aux points où l'on a $z_0 = 0$) en une fonction $\tilde{V} \in PS(C^{n+1})$. A $x \in C^n$ faisons correspondre d'autre part une droite complexe de C^{n+1}, soit $d(x)$, par l'origine, pointée à l'origine: pour $x = (x_1, \ldots, x_n) \in C^n$, $d(x)$ est parcourue, quand x_0 parcourt $\mathbb{C} - \{0\}$, par le point $x' = (x_0, x_1 x_0, \ldots, x_n x_0)$ dans $\mathbb{C}^{n+1}$.

Par tout point x' de C^{n+1} vérifiant $x'_0 \neq 0$, il passe une et une seule droite $d(x)$. Dans la suite de ce paragraphe, on écrit $\nu(x, V)$ au lieu de $\nu(x, T_V)$.

Proposition 7.1. *Soit V plurisousharmonique dans une boule B de centre 0 de C^n et soit l une courbe $x_k = \varphi_k(u)$ aboutissant à l'origine où l'on a $\varphi_k(0) = 0$, $\Sigma|\varphi'_k(0)|^2 \neq 0$, les φ_k étant holomorphes pour $|u| < 1$. On définit à l'origine le nombre $\nu(0, V, \ell)$ relatif au courant T_V et à la courbe ℓ par*

$$\nu(0, V, \ell) = \lim_{r=0} (\log r)^{-1} \frac{1}{2\pi} \int_0^{2\pi} V \circ \varphi(re^{i\theta}) \, d\theta. \tag{7.4}$$

On pose $\nu(0, V, \ell) = +\infty$ si ℓ appartient à l'ensemble $V(x) = -\infty$. On note $\nu(0, V, y)$ si ℓ est la droite $x_k = y_k u$, pour $y \in C^n - \{0\}$. Alors on a

$$\nu(0, V, \ell) \geqq \nu(0, V) = \inf_y \nu(0, V, y).$$

De plus les y vérifiant $\nu(0, V, y) > \nu(0, V)$ forment un cône pluripolaire dans C^n.

La propriété résulte de [7], cf. p. 143: on peut supposer $V<0$. Alors si dans (7.4) on prend $\varphi_k = y_k u$, on a $v(0, V, u y) = v(0, V, y)$ pour tout $u \in \mathbb{C}$. Dans le second membre de (7.4), soit $S_r(y)$, les $-S_r(y)$ sont une famille de fonctions négatives, plurisousharmoniques, dont la régularisée supérieure est une constante α. L'ensemble $v(x, V, y) < \alpha$ est un cône pluripolaire, on a donc $\alpha = v(0, V)$. Si ℓ est une courbe de l'énoncé, soit $\varphi_1'(0) \neq 0$; on exprime $u = \psi(x_1)$ et on définit ℓ par les équations $z_k = x_k - \varphi_k \circ \psi(x_1) = 0, 2 \leqq k \leqq n$ et $z_1 = x_1$. Alors $x \to z$ est un isomorphisme analytique, l'image de ℓ est l'axe des z_1. On a donc $v(0, V, \ell) \geqq v(0, V)$ ce qui établit l'énoncé. On a alors:

Proposition 7.2. *Soit $\tilde{V} \in S_{ah}(C^{n+1})$ et $d(x)$ la droite pointée à l'origine dans C^{n+1}, transformée de $x \in C^n$, parcourue par $x' = (x_0, x_1 x_0, \ldots, x_n x_0)$ pour $x_0 \neq 0$, x_0 variable. Alors $v(x', \tilde{V})$ est constant quand x' parcourt $d(x)$ et l'on a*

$$v(x', \tilde{V}) = v(x, V); \qquad x' \in d(x). \tag{7.2}$$

Il suffit d'établir (7.2). On calcule $v(x', \tilde{V}) = \inf_\beta v(x', \tilde{V}, \beta)$ pour $\beta = (\beta_0, \ldots, \beta_n) \in C^{n+1}$, c'est-à-dire sur les droites de C^{n+1} issues de x', de direction β et d'équations $z_0' = x_0' + \beta_0 u, \ldots, z_n' = x_n' + \beta_n u$; on a

$$\tilde{V}(z_0', \ldots, z_n') = V \circ z(u) + a \log |x_0' + \beta_0 u|$$

où $z(u)$ est la courbe ℓ: $z_k = (x_k' + \beta_k u)(x_0' + \beta_0 u)^{-1}$, $1 \leqq k \leqq n$. On a $v(x', \tilde{V}) = \inf_\beta v(x', \tilde{V}, \beta) = \inf_\ell v(x, V, \ell) \geqq v(x, V)$.

Mais pour $\beta_0 = 0$, les courbes $\ell(\beta)$ ont pour image toutes les droites issues dans C^n de x. On a donc l'égalité (7.2). Montrons alors:

Théorème 7.3. *Si la fonction plurisousharmonique V appartient à une classe $S_a(C^n)$ de croissance minimale, alors les ensembles de densité $E(c, T_V)$ sont algébriques pour tout $c>0$. Ils sont vides pour $c>a$.*

En effet à $x \in E(c, T_V)$ correspond par $x \to d(x)$ un cône pointé dans C^{n+1} qui est l'ensemble de densité relatif au courant associé à $\tilde{V}$. Cet ensemble est analytique. D'autre part il est conique, invariant par $x' \to ux$, $u \neq 0$, $u \in \mathbb{C}$. C'est donc un ensemble algébrique et $E(c, T_V)$ qui en est la section par $x_0 = 1$ est algébrique. On a $v(0, \tilde{V}) = a$ et $v(x, V) \leqq a$ pour tout x d'après le comportement de $\lambda(V, x, r)$ pour $r \to \infty$.

Remarque 7.4. Réciproquement à partir des polynomes $P_j(z) = 0$ qui définissent un ensemble algébrique A, on a en posant $V = \frac{1}{2} \log \Sigma |P_j|^2 \in S_a(C^n)$, une représentation $A = E(1, T_V)$ pour $a = \sup_j [\text{degré } P_j]$. *Ainsi les ensembles de densité $E(c, T_V)$ pour V parcourant une classe $S_a(C^n)$ - par exemple pour $a=1$ - sont exactement les ensembles algébriques de C^n.*

7.2. Soit $V \in S_a(C^n)$. Alors on a pour $T = T_V$: $\lim v_T(r) = a$; d'où d'après le §3:

Corollaire 7.5. *Soit $V \in S_a(C^n)$: les cycles Z_s de codimension 1, vérifient*

$$\sum N(Z_s) [\text{degré } Z_s] \leqq a \tag{7.3}$$

où $N(Z_s)$ est la valence de Z_s.

Plus généralement, on a: en utilisant (5.13) et les théorèmes 7.3 et 3.7:

Théorème 7.6. *Soit* $T \in M_+^p(C^n)$ *d'indicatrice bornée par* $a>0$. *Alors pour tout* $c>0$, *l'ensemble* $E(c,T)$ *est algébrique. Les cycles à densité de la dimension maxima* $p=\dim T$, *vérifient* (7.3).

Corollaire 7.7. *La partie de dimension p de l'ensemble* $E(c,T)$ *pour* $T \in M_+^p(C^n)$, *a un degré au plus* $\frac{a}{c}$.

Du §5 on déduit:

Corollaire 7.8. *Soit* $T \in M_+^p(C^n)$ *d'indicatrice* $v_T(r)$ *majorée par* $a>0$. *Alors l'ensemble algébrique* $E(c,T)$ *est contenu dans une hypersurface* $Y_c = F_1^{-1}(0)$ *où* F_1 *est un polynome de degré au plus* $\frac{na}{c}$.

De même, comme conséquence du §6, on aura:

Corollaire 7.9. *Soit* $V \in S_a(C^n)$ *et* $c>0$ *une valeur pour laquelle on suppose:* $E(c', T_V) = E(c, T_V)$ *pour* $c-\gamma < c' \leqq c$. *Alors l'ensemble* $E(c, T_V)$ *peut être déterminé en annulant* $n+1$ *polynomes* P_j, *où l'on a degré* $P_1 \leqq nac^{-1}$, *et pour* $2 \leqq j \leqq n+1$ *degré* $P_j \leqq \frac{(n-1+\varepsilon)a}{\gamma}$ *pour tout* ε *vérifiant* $0<\varepsilon<1$ *et* $\varepsilon > 1-n\left(1-\frac{\gamma}{c}\right)$.

On a le même résultat pour l'ensemble $E(c,T)$ si l'on suppose $T \in M_+^p(C^n)$, et $\lim\limits_{r=\infty} v_T(r) = a$.

On notera que pour le type de croissance minimal $M_V(r) \sim a \log r$, les contrôles asymptotiques sont des contrôles véritables. Il est facile de voir que c'est là une propriété réservée à ce type de croissance.

References

1. Avanissian, V.: Fonctions entières de p variables et fonctions plurisousharmoniques à croissance très lente. Journal d'Analyse, Jérusalem **9**, 347–364 (1962)
2. Bombieri, E.: Algebraic values of meromorphic maps. Inv. Math. **10**, 267–287 (1970)
3. Cornalba, M., Shiffman, B.: A counterexample to the transcendental Bezout problem. Ann. of Math. **96**, 402–406 (1972)
4. Draper, R.N.: Intersection theory in analytic geometry. Math. Ann. **180**, 175–204 (1969)
5. Kiselman, C.O.: Densité des fonctions plurisousharmoniques. Bull. Soc. Math. de France **107**, 295–304 (1979)
6. Lelong, P.: Intégration sur un ensemble analytique complexe. Bull. Soc. Math. de France **85**, 239–362 (1957)
7. Lelong, P.: Sur la structure des courants positifs fermés. Lecture Notes in Math., vol. 578, pp. 136–156. Berlin-Heidelberg-New York: Springer 1976; Note aux C.R. Acad. Sciences, Paris **283**, 449–452 (1976)
8. Lelong, P.: Plurisubharmonic functions and positive differential forms. Gordon and Breach 1968
9. Lelong, P.: Fonctions entières de type exponentiel dans C^n. Ann. Inst. Fourier, **16**, pp. 269–318, cf. Annexe: Note sur une classe de fonctions plurisousharmoniques de croissance minimale
10. Lelong, P.: Fonctions entières et fonctions plurisousharmoniques d'ordre fini dans C^n. Journal d'Analyse Math. de Jérusalem **12**, 365–407 (1964)
11. Lelong, P.: Topologies sur les courants positifs fermés. Lecture Notes in Math., vol. 275, pp. 27–70. Berlin-Heidelberg-New York: Springer 1971

12. Siu, Y.T.: Analyticity of sets associated to Lelong numbers. Invent. Math. **27**, 53–156 (1974)
13. Skoda, H.: Sous ensembles analytiques d'ordre fini ou infini dans C^n. Bull. Soc. Math. France **100**, 353–408 (1972)
14. Skoda, H.: Estimation L^2 pour l'opérateur $\bar{\partial}$ et applications arithmétiques. Lecture Notes in Math., vol. 578, pp. 314–323. Berlin-Heidelberg-New York: Springer 1976
15. Thie, P.: The Lelong number in a point of a complex analytic set. Math. Ann. **172**, 269–312 (1967)

Oblatum 6-VI-1981 & 5-VIII-1982 & 26-I-1983

Invent. math. 72, 491–506 (1983)

Inventiones mathematicae

On a Theorem of Castelnuovo, and the Equations Defining Space Curves

L. Gruson[1], R. Lazarsfeld[2]*, and C. Peskine[3]

[1] Départment de Mathématiques, Université de Lille 1, F-59655 Villeneuve d'Ascq, France
[2] Department of Mathematics, Harvard University, Cambridge, MA 02138, USA
[3] Mathematisk Institutt, Universitetet I Oslo, P.B. 1053 – Blindern, Oslo 3, Norway

Contents

Introduction

Consider a reduced irreducible curve

$$X \subseteq \mathbb{P}^r \qquad (r \geqq 3)$$

of degree d, not contained in any hyperplane. For a given integer $n \geqq 0$, it is natural to ask whether X enjoys one or more of the following properties:

(A_n) The line bundle $\mathcal{O}_X(n)$ is non-special.

(B_n) Hypersurfaces of degree n trace out a *complete* linear system on X, i.e., the homomorphism

$$H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(n)) \to H^0(X, \mathcal{O}_X(n))$$

is surjective.

(C_n) X is cut out in $\mathbb{P}^r$ by hypersurfaces of degree n, and the homogeneous ideal of X is generated in degrees $\geqq n$ by its component of degree n.

It is of course classical that each of these conditions is satisfied for all sufficiently large n. But one would like to have an *explicit* bound on how large n must be, and to understand the extremal examples.

* Partially supported by the NSF and the AMS.

The fundamental work in this direction was carried out by Castelnuovo [3]. For property (A), using the argument leading to his celebrated bound on the genus of a space curve, Castelnuovo obtained complete results (cf. [20], or [6]; when $r=3$ a considerable refinement was stated earlier by Halphen [9], and recently proved by the first and third authors [7] and by Hartshorne [10]). The present paper is chiefly concerned with property (B). Castelnuovo proved that for smooth curves, at least, (B_n) holds for $n \geqq d-2$, and he suggested that an irrational curve satisfies (B_{d-3}). Examples show that Castelnuovo's estimate is optimal for curves in $\mathbb{P}^3$. But given $X \subseteq \mathbb{P}^r$ it is natural to expect that property (B_n) holds in fact for $n \geqq d+1-r$. Our purpose here is to complete Castelnuovo's theorem by showing that this is indeed the case, and by describing the examples on the boundary.

Our main result is the following

Theorem. *Let $X \subseteq \mathbb{P}^r$ be a reduced irreducible curve of degree d, not contained in any hyperplane. Then:*

(i) *Property (B_n) is satisfied for all $n \geqq d+1-r$.*

(ii) *Property (B_{d-r}) fails if and only if X is smooth and rational, and either $d=r+1$, or $d>r+1$ and X has a $(d+2-r)$-secant line.*

Note that it is not required in (i) that X be smooth. The first and third authors had previously extended the estimate given by Castelnuovo to possibly singular curves. They had also proved Castelnuovo's assertion that any extremal example in $\mathbb{P}^3$ is rational (unpublished). Special cases of (i) were proved by Jongmans [12] via a reduction to Castelnuovo's results, and by Meadows [13] for certain rational curves. A modern exposition of Castelnuovo's theorem is given by Szpiro in his notes [20].

It is known by work of Mumford [14], who attributes the idea to Castelnuovo, that if a curve satisfies (A_{n_0-2}) and (B_{n_0-1}) for some $n_0 \geqq 0$, then (C_n) holds for all $n \geqq n_0$. It follows from Castelnuovo's results for property (A), or from the proofs below, that (A_{d-r}) always holds. The cases where (A_{d-1-r}) or (B_{d-r}) fail are easily analyzed separately, and we obtain the

Corollary. *Let $X \subseteq \mathbb{P}^r$ $(r \geqq 3)$ be a reduced, irreducible curve of degree d, not contained in any hyperplane. Then:*

(i) *Property (C_n) holds for all $n \geqq d+2-r$.*

(ii) *Property (C_{d+1-r}) fails if and only if X is a smooth rational curve having a $(d+2-r)$-secant line.*

The equations defining space curves have been studied by several authors, notably Mumford [15], Saint-Donat [18], and Arbarello-Sernese [1]. Our viewpoint differs somewhat from theirs, however, in that we are forced to deal with embeddings defined by possibly incomplete linear systems.

Unlike Castelnuovo's arguments, which are geometric in nature, our proofs are essentially cohomological. They depend on a simple but somewhat surprising technique. Roughly speaking, the method is to "resolve" the ideal sheaf $\mathscr{I}_{X/\mathbb{P}^r}$ (or something related) by a complex with generally non-trivial homology supported on X. For example, to prove the first statement of the Theorem, we use a Beilinson-type construction to express X as the locus where a matrix of linear forms drops rank, and then take the corresponding Eagon-Northcott complex. By allowing non-trivial homology, one arrives at complexes much simpler than those that would be required for an actual resolution of the ideal sheaf. Happily, as long as the complex is exact away from a one-dimensional set, one is still able to read off the desired vanishings. We hope that this technique may find other applications in the study of space curves.

An explicit estimate for the regularity of an arbitrary ideal sheaf on $\mathbb{P}^r$ has been obtained by Gotzmann [5] and generalized by Bayer [2] to any coherent sheaf. Gotzmann's bound depends only on the Hilbert polynomial of the scheme in question, however, and as one would expect the numbers that it gives are generally far from optimal for reduced curves.

Our exposition proceeds in three stages. First we establish the regularity assertions (i) of the Theorem and Corollary (§1). We next show (§2) by a similar but independent argument that an *irrational* curve satisfies (B_{d-r}) and (C_{d+1-r}), so that any extremal example must be rational. For rational curves, a refinement of the proof in §1 then gives the classification statements (ii) (§3). Strictly speaking, the first assertions of the Theorem and its Corollary are consequences of the enumeration of the extremal examples, and it would have been possible to organize the presentation in such a way as to avoid at least the last lemma of §1. However the regularity results (i) are substantially easier to prove than the classification statements (ii), and it seemed to us worthwhile to treat them directly. Finally we discuss in §4 some open problems.

We have greatly benefitted from the suggestions and encouragement of several people, especially E. Arbarello, D. Eisenbud, J. Harris, A.P. Rao, and K. Vilonen.

§0. Notation and Conventions

(0.1) We work over an algebraically closed field k of arbitrary characteristic.

(0.2) Unless otherwise stated, a *curve* is a *reduced* and *irreducible*, but possibly singular, projective variety of dimension one. Recall that a curve $X \subseteq \mathbb{P}^r$ is *non-degenerate* if it is not contained in a hyperplane. Given a variety $V \subseteq \mathbb{P}^r$, we denote by $\mathscr{I}_V$ the ideal sheaf of V in $\mathbb{P}^r$; if $X \subseteq V$ is a subvariety, we indicate the ideal sheaf of X in V by $\mathscr{I}_{X/V}$. For a coherent sheaf $\mathscr{F}$ on V, $\mathscr{F}(n)$ denotes as usual $\mathscr{F} \otimes \mathcal{O}_{\mathbb{P}^r}(n)$, and following common usage we let $h^i(V, \mathscr{F}) = \dim H^i(V, \mathscr{F})$.

(0.3) Consider a generically surjective homomorphism

$$u\colon E \to F$$

of vector bundles of ranks e and f on a smooth variety. Associated to u are several *Eagon-Northcott complexes* (cf. [4, 19, or 8]), of which we shall need the following two:

(0.4) $$0 \to \Lambda^e E \otimes S^{e-f}(F)^* \to \ldots \to \Lambda^{f+1} E \otimes F^* \to \Lambda^f E \xrightarrow{\Lambda^f u} \Lambda^f F \to 0$$

(0.5) $$0 \to \Lambda^e E \otimes S^{e-f-1}(F)^* \to \ldots \to \Lambda^{f+2} E \otimes F^* \to \Lambda^{f+1} E$$
$$\to E \otimes \Lambda^f F \xrightarrow{u \otimes 1} F \otimes \Lambda^f F \to 0.$$

The basic fact for our purposes is that *these complexes are exact away from the support of* coker u ([4, 19, 8]). (In general they aren't acyclic unless Supp(coker u) has the expected codimension $e-f+1$.)

§1. A Regularity Theorem

It will be convenient to phrase our results in terms of the regularity of the ideal sheaf. Recall ([14], Lecture 14) that a coherent sheaf $\mathscr{F}$ on $\mathbb{P}^r$ is said to be *n-regular* if $H^i(\mathbb{P}^r, \mathscr{F}(n-i))=0$ for $i>0$. The usefulness of this concept lies in the fact that if $\mathscr{F}$ is n-regular, then:

(i) $\mathscr{F}(n)$ is generated by its global sections,

(ii) the maps

$$H^0(\mathscr{F}(n)) \otimes H^0(\mathcal{O}_{\mathbb{P}^r}(l)) \to H^0(\mathscr{F}(n+l))$$

are surjective for $l \geqq 0$, and

(iii) $\mathscr{F}$ is $(n+1)$-regular.

([14], loc. cit.). We will say that a curve $X \subseteq \mathbb{P}^r$ is n-regular if its ideal sheaf $\mathscr{I}_X$ is, and n-irregular otherwise. Thus for $n \geqq 0$, X is n-regular if and only if properties (A_{n-2}) and (B_{n-1}) of the Introduction are satisfied, in which case (C_n) holds by facts (i) and (ii), as do $(A_{n'-2})$, $(B_{n'-1})$ and $(C_{n'})$ for any $n' \geqq n$ by (iii). Hence the first assertions of the results stated in the Introduction are consequences of

Theorem 1.1. *Let $X \subseteq \mathbb{P}^r$ be a (reduced and irreducible) non-degenerate curve of degree d. Then X is $(d+2-r)$-regular.*

The key to Theorem 1.1 is the following result, which allows one to estimate the regularity of a curve in terms of a vanishing on its normalization.

Proposition 1.2. *Let $X \subseteq \mathbb{P}^r$ be a reduced (but possibly reducible) curve, with normalization $\tilde{X}$, and let p: $\tilde{X} \to \mathbb{P}^r$ denote the natural map. Set*

$$M = p^* \Omega^1_{\mathbb{P}^r}(1),$$

and suppose that A is a line bundle on $\tilde{X}$ such that

$$H^1(\tilde{X}, \Lambda^2 M \otimes A) = 0.$$

Then $X \subseteq \mathbb{P}^r$ is $h^0(\tilde{X}, A)$-regular.

Proof. The first step is to show that there is an exact sequence

$$(1.3)\qquad H^0(\tilde{X}, M\otimes A)\otimes_k \mathcal{O}_{\mathbb{P}^r}(-1) \xrightarrow{\;u\;} H^0(\tilde{X}, A)\otimes_k \mathcal{O}_{\mathbb{P}^r} \to p_* A \to 0$$

of sheaves on $\mathbb{P}^r$. This follows readily from an evident variation of the Beilinson spectral sequence (cf. [16], II.3.1). For later reference, and for the benefit of the reader not versed in such matters, we may give an elementary and self-contained (but equivalent) derivation as follows. Setting $\mathcal{O}_{\tilde{X}}(1) = p^*\mathcal{O}_{\mathbb{P}^r}(1)$, the morphism $p: \tilde{X} \to \mathbb{P}^r$ is determined by a subspace $V \subseteq H^0(\tilde{X}, \mathcal{O}_{\tilde{X}}(1))$ of dimension $r+1$. This gives rise to an exact sequence

$$(1.4)\qquad 0 \to M \to V_{\tilde{X}} \to \mathcal{O}_{\tilde{X}}(1) \to 0$$

of vector bundles on $\tilde{X}$, where $V_{\tilde{X}} = V\otimes_k \mathcal{O}_{\tilde{X}}$. Denote by π and f the projections of $\tilde{X}\times\mathbb{P}^r$ onto $\tilde{X}$ and $\mathbb{P}^r$ respectively. On $\tilde{X}\times\mathbb{P}^r$ there are vector bundle homomorphisms

$$\begin{tikzcd}
0 \arrow[r] & \pi^* M \arrow[r] \arrow[dr, "s"'] & \pi^* V_{\tilde{X}} \arrow[d] \\
 & & f^*\mathcal{O}_{\mathbb{P}^r}(1),
\end{tikzcd}$$

and the graph $\Gamma \subseteq \tilde{X}\times\mathbb{P}^r$ of p is defined scheme-theoretically by the vanishing of s. $\tilde{X}\times\mathbb{P}^r$ being smooth, one obtains a Koszul resolution of $\mathcal{O}_\Gamma$, and twisting by $\pi^* A$ gives the following resolution of $\mathcal{O}_\Gamma\otimes\pi^* A$:

$$(1.5)\qquad \begin{tikzcd}[column sep=small]
 & & & 0 \arrow[dr] & & 0 & & \\
 & & & & \mathscr{F}_0 \arrow[ur] \arrow[dr] & & & \\
\dots \arrow[r] & \pi^*(\Lambda^2 M\otimes A)\otimes f^*\mathcal{O}_{\mathbb{P}^r}(-2) \arrow[rr] \arrow[dr] & & \pi^*(M\otimes A)\otimes f^*\mathcal{O}_{\mathbb{P}^r}(-1) \arrow[rr] \arrow[ur] & & \pi^* A \arrow[r] & \mathcal{O}_\Gamma\otimes\pi^* A \arrow[r] & 0. \\
 & & \mathscr{F}_1 \arrow[ur] \arrow[dr] & & & & & \\
 & 0 \arrow[ur] & & 0 & & & &
\end{tikzcd}$$

The sheaves $\mathscr{F}_0$ and $\mathscr{F}_1$ are defined as indicated. By Künneth,

$$R^q f_*(\pi^*(\Lambda^p M\otimes A)\otimes f^*\mathcal{O}_{\mathbb{P}^r}(-p)) = H^q(\tilde{X}, \Lambda^p M\otimes A)\otimes_k \mathcal{O}_{\mathbb{P}^r}(-p).$$

Therefore

$$R^1 f_* \mathscr{F}_1 = 0$$

thanks to the vanishing of $H^1(\tilde{X}, \Lambda^2 M\otimes A)$ and the fact that $R^1 f_*$ is right exact since f has fibre dimension one. Taking direct images in (1.5) one then finds the exact sequence

$$f_*(\pi^*(M\otimes A)\otimes f^*\mathcal{O}_{\mathbb{P}^r}(-1)) \xrightarrow{u} f_*(\pi^*A) \to f_*(\mathcal{O}_\Gamma\otimes\pi^*A)$$

$$\xrightarrow{\delta} R^1f_*(\pi^*(M\otimes A)\otimes f^*\mathcal{O}_{\mathbb{P}^r}(-1))$$

of sheaves on $\mathbb{P}^r$. But $f_*(\mathcal{O}_\Gamma\otimes\pi^*A)=p_*A$ is a torsion $\mathcal{O}_{\mathbb{P}^r}$-module, whereas $R^1f_*(\pi^*(M\otimes A)\otimes f^*\mathcal{O}_{\mathbb{P}^r}(-1))$ is locally free, and so $\delta=0$. Thus one arrives finally at the exact sequence (1.3). We set

$$n_0=h^0(\tilde{X},A).$$

Let $\mathcal{J}\subseteq\mathcal{O}_{\mathbb{P}^r}$ denote the zeroth Fitting ideal sheaf of p_*A computed from (1.3), i.e., the image of $\Lambda^{n_0}u$. Since p_*A is supported on X, the subscheme defined by $\mathcal{J}$ coincides set-theoretically with X, and hence

$$\mathcal{J}\subseteq\mathcal{I}_X$$

because X is reduced. Moreover as the Fitting ideals of a module are independent of the presentation used to compute them, $\mathcal{I}_X/\mathcal{J}$ is supported in the finitely many points of $\mathbb{P}^r$ at which p_*A fails to be locally isomorphic to $\mathcal{O}_X$ (i.e., the singular points of X). Thus $H^i(\mathbb{P}^r,\mathcal{I}_X(m))$ is a quotient of $H^i(\mathbb{P}^r,\mathcal{J}(m))$ for all $i>0$ and $m\in\mathbb{Z}$, so it suffices to prove that $\mathcal{J}$ is n_0-regular.

Consider to this end the Eagon-Northcott complex (0.4) constructed from u. It takes the form

$$\ldots\to\mathcal{O}_{\mathbb{P}^r}^{M_{r-1}}(-n_0+1-r)\to\ldots\to\mathcal{O}_{\mathbb{P}^r}^{M_1}(-n_0-1)\to\mathcal{O}_{\mathbb{P}^r}^{M_0}(-n_0)\xrightarrow{\varepsilon}\mathcal{J}\to 0,$$

where $\varepsilon=\Lambda^{n_0}u$ is surjective. This complex is exact off X, and we observe that twisting by $\mathcal{O}_{\mathbb{P}^r}(n_0-m)$ for some $1\leqq m\leqq r$ kills the higher cohomology of the first $r+1-m$ locally free terms from the right. Thus the following Lemma applies to prove the n_0-regularity of $\mathcal{J}$. ■

Lemma 1.6. *Let*

$$\mathcal{L}.:\ \ldots\to\mathcal{L}_{r-1}\to\ldots\to\mathcal{L}_1\to\mathcal{L}_0\xrightarrow{\varepsilon}\mathcal{J}\to 0$$

be a complex of coherent sheaves on a projective variety V of dimension r, with ε surjective. Assume that

(a) *$\mathcal{L}.$ is exact away from a set of dimension $\leqq 1$,*

and

(b) *For a given integer $1\leqq m\leqq r$, one has*

$$H^i(V,\mathcal{L}_0)=\ldots=H^i(V,\mathcal{L}_{r-m})=0\quad \textit{for } i>0.$$

Then $H^i(V,\mathcal{J})=0$ for $i\geqq m$.

Proof. Hypothesis (a) guarantees that the homology sheaves $\mathcal{H}_j$ of $\mathcal{L}.$ have vanishing cohomology in degrees $\geqq 2$:

$$H^i(V, \mathscr{K}_j) = 0 \quad \text{for } i \geqq 2 \quad \text{and all} \quad j.$$

With this in mind, the Lemma is most simply proved by chopping $\mathscr{L}.$ into short exact sequences in the usual way, and chasing through the resulting diagram. ■

Remark. Note for later reference that Proposition 1.2 remains valid for any line bundle A for which the sheaf $R^1 f_* \mathscr{F}_1$ occurring in the proof is supported on a set of dimension $\leqq 0$. For then one still has $\mathscr{J} \subseteq \mathscr{I}_X$, while $\mathscr{I}_X/\mathscr{J}$ and coker u continue to be supported in sets of dimensions zero and one respectively.

Theorem 1.1 now follows from

Lemma 1.7. *Let $\tilde{X}$ be a smooth irreducible curve of genus g, and let $p: \tilde{X} \to \mathbb{P}^r$ be a morphism of degree d (i.e., $\deg p^* \mathcal{O}_{\mathbb{P}^r}(1) = d$). Assume that $\tilde{X}$ does not map into a hyperplane, and set $M = p^* \Omega^1_{\mathbb{P}^r}(1)$. Then there exists a line bundle A on $\tilde{X}$, with $h^0(\tilde{X}, A) = d + 2 - r$, such that $H^1(\tilde{X}, \Lambda^2 M \otimes A) = 0$.*

Proof. We assert that M admits a filtration

$$(1.8) \qquad M = F^1 \supseteq F^2 \supseteq \ldots \supseteq F^r \supseteq F^{r+1} = 0$$

by vector bundles such that each of the quotients $L_i = F^i/F^{i+1}$ is a line bundle of strictly negative degree. Indeed, $H^0(\tilde{X}, M) = 0$ by (1.4) since $\tilde{X}$ does not map to a hyperplane, and in particular no non-zero sub-bundle of M is trivial. The existence of the desired filtration is then a consequence of the elementary observation that if a non-trivial vector bundle F on a smooth curve is a sub-bundle of a trivial bundle, then F has a line bundle quotient of negative degree.

In order that $H^1(\tilde{X}, \Lambda^2 M \otimes A) = 0$, it suffices that

$$(*) \qquad H^1(\tilde{X}, L_i \otimes L_j \otimes A) = 0 \quad \text{for all } 1 \leqq i < j \leqq r.$$

Since M has degree $-d$ and all the L_i have degree <0,

$$\deg(L_i \otimes L_j) \geqq r - 2 - d$$

for any $1 \leqq i < j \leqq r$. But a generic line bundle of degree $\geqq g-1$ is non-special, so $(*)$ will be satisfied if A is a sufficiently general line bundle of degree $g + d + 1 - r$. Moreover $p(\tilde{X})$ being non-degenerate one has $d \geqq r$, so we may suppose in addition that $H^1(\tilde{X}, A) = 0$, in which case $h^0(\tilde{X}, A) = d + 2 - r$. ■

Remarks. (1) If $X \subseteq \mathbb{P}^r$ is reduced but possibly reducible, a variant of Theorem 1.1 holds. Specifically, suppose that X has irreducible components X_i of degree d_i, and that X_i spans a $\mathbb{P}^{r_i} \subseteq \mathbb{P}^r$. Set

$$m_i = \begin{cases} d_i + 2 - r_i & \text{if } \; d_i \geqq 2 \\ 1 \;\; \text{if } \; d_i = 1 & \text{(i.e., if } X_i \text{ is a line).} \end{cases}$$

Applying a slight modification of (1.7) component by component on $\tilde{X}$, one finds from Proposition 1.2 that X is (Σm_i)-regular.

For example, suppose that $X \subseteq \mathbb{P}^r$ consists of d straight lines. Then X is d-regular. In general this is optimal, for if there is a line $L \nsubseteq X$ meeting each of the components of X at distinct points, then X is not cut out by hypersurfaces of degree $d-1$ (compare §3). By contrast, when X is the union of d *generic* lines, Hartshorne and Hirschowitz [11] have shown that the map

$$H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(n)) \to H^0(X, \mathcal{O}_X(n))$$

has maximal rank for all n, so that generically one has a much better regularity result.

(2) In the situation of Lemma 1.7 it is amusing to vary the filtration (1.8) while twisting by a fixed line bundle. Suppose, for example, that $X \subseteq \mathbb{P}^r$ is a smooth non-degenerate curve of degree d and genus g. At least in characteristic zero, so that not every secant line is multi-secant, one sees that for almost all choices of $r-1$ points $P_1, \ldots, P_{r-1} \in X$ there is a filtration (1.8) with

$$F^i/F^{i+1} = \begin{cases} \mathcal{O}_X(-P_i) & \text{if } 1 \leqq i \leqq r-1 \\ \mathcal{O}_X(-1) \otimes \mathcal{O}_X(\Sigma P_j) & \text{if } i = r \end{cases}$$

(project from the P_i). In particular, by taking the P_i sufficiently generally, it follows that $H^1(X, M(1)) = 0$ provided that $r-1 \geqq g$. This in turn implies that if $r \geqq g+1$, then the natural map

$$(*) \qquad H^1(\mathbb{P}^r, \mathcal{I}_X(1)) \otimes H^0(\mathbb{P}^r, \mathcal{O}_{\mathbb{P}^r}(n)) \to H^1(\mathbb{P}^r, \mathcal{I}_X(n+1))$$

is *surjective* for $n \geqq 0$. So for example if $X \subseteq \mathbb{P}^3$ has genus two or less, then the Horrocks-Hartshorne-Rao module $\oplus H^1(\mathbb{P}^3, \mathcal{I}_X(n))$ of X (cf. [17]) is generated by its degree one piece. Similarly, if $X \subseteq \mathbb{P}^r$ is defined by the *complete* linear system associated to a line bundle of degree $d \geqq 2g+1$, then $r = d-g \geqq g+1$, and so we recover from $(*)$ Mumford's theorem [15] that X is projectively normal.

§ 2. A Rationality Theorem

We begin in this section the classification of all curves for which Theorem 1.1 is optimal, i.e., all $(d+1-r)$-irregular curves. The final list is summarized in a table at the end of §3. Our immediate goal is to show that with one exception such a curve must be rational:

Theorem 2.1. *Let*

$$X \subseteq \mathbb{P}^r \quad (r \geqq 3)$$

be a (reduced, irreducible) non-degenerate curve of degree d and geometric genus g. If $g \geqq 1$, *then X is* $(d+1-r)$*-regular unless it is an elliptic normal curve.*

Remark. If X is elliptic normal, so that $d = r+1$, then evidently (A_0) fails. On the other hand, X is projectively normal, and property (C_2) holds provided that $r \geqq 3$ (cf. (2.3) below).

Unfortunately, there are a few examples where there does not exist a line bundle A with $h^0(A) \leqq d+1-r$ satisfying the hypotheses of Proposition 1.2 (e.g. an elliptic quintic in $\mathbb{P}^3$). It is simplest for Theorem 2.1 to argue directly. As in § 1, let $\tilde{X}$ be the normalization of X, so that $\tilde{X}$ has genus $g \geqq 1$. Denote by p: $\tilde{X} \to \mathbb{P}^r$ the natural map, and set $\mathcal{O}_{\tilde{X}}(1) = p^* \mathcal{O}_{\mathbb{P}^r}(1)$.

Lemma 2.2. *There exists a line bundle A on $\tilde{X}$ of degree $g-1$ such that $h^0(A) = 1$. For any such line bundle, and any $n > 0$,*

$$h^0(A(-n)) = 0 \quad \textit{and} \quad h^1(A(n)) = 0.$$

Proof. A sufficiently general effective divisor of any degree $\leqq g$ corresponds to line bundle A with $h^0(A) = 1$. A non-zero section of $A(-n)$ would give rise to an inclusion $0 \to \mathcal{O}_{\tilde{X}}(n) \to A$, but since $h^0(\mathcal{O}_{\tilde{X}}(n)) > 1$ when $n > 0$ this is absurd. When A has degree $g-1$, $h^0(\Omega^1_{\tilde{X}} \otimes A^*) = 1$ and the same argument shows that $h^1(A(n)) = 0$ for positive n. ■

Fix a line bundle A as in the Lemma, and set

$$E = \bigoplus_{n \in \mathbb{Z}} H^0(\mathbb{P}^r, p_* A(n)) \qquad [= \bigoplus_{n \in \mathbb{Z}} H^0(\tilde{X}, A \otimes \mathcal{O}_{\tilde{X}}(n))],$$

so that E is in the natural way a graded module over the homogeneous coordinate ring $S = k[T_0, \ldots, T_r]$.

Lemma 2.3. *E admits a minimal free resolution of length $r-1$ having the following numerical type:*

$$\begin{aligned} 0 \to S(-r-1) \oplus S^{d-r-1}(-r) \to S^{n_{r-2}}(-r+1) \to \ldots \\ \ldots \to S^{n_1}(-2) \to S^{d-r-1}(-1) \oplus S \to E \to 0. \end{aligned}$$

Proof. E is a Cohen-Macaulay module of dimension two, so in any event has a free resolution of length $r-1$.

By Lemma 2.2, E has one generator, say e, in degree zero and none in negative degrees. Since $X \subseteq \mathbb{P}^r$ is non-degenerate, there are no relations of linear dependence among the elements $T_0 \cdot e, \ldots, T_r \cdot e \in E_1$. Hence by Lemma 2.2 and Riemann-Roch, we require $d-r-1$ new generators in degree one. Similarly, it follows from duality that $\operatorname{Ext}^{r-1}_S(E, S(-r-1))$ vanishes in negative degrees, has one generator in degree zero, and $d-r-1$ in degree one. Therefore a minimal resolution of E must be of the form

$$\begin{aligned} 0 \to S(-r-1) \oplus S^{d-r-1}(-r) \oplus \bigoplus_{j_{r-1}} S(-c_{r-1, j_{r-1}}) \to \bigoplus_{j_{r-2}} S(-c_{r-2, j_{r-2}}) \to \ldots \\ \ldots \to \bigoplus_{j_1} S(-c_{1, j_1}) \to \bigoplus_{j_0} S(-c_{0, j_0}) \oplus S^{d-r-1}(-1) \oplus S \to E \to 0 \end{aligned}$$

for suitable integers c_{k, j_k}. Furthermore, by minimality the numbers

$$u_k = \min_{j_k} \{c_{k, j_k}\} \quad \text{and} \quad v_k = \max_{j_k} \{c_{k, j_k}\}$$

are strictly increasing for $1 \leqq k \leqq r-2$. But in view of the remarks above, the first module of syzygies of E vanishes below degree two, i.e., $u_1 \geqq 2$, and similarly $v_{r-2} \leqq r-1$. Hence $u_k = v_k = k+1$ for $1 \leqq k \leqq r-2$, and by the same token the terms $\bigoplus S(-c_{0,j_0})$ and $\bigoplus S(-c_{r-1,j_{r-1}})$ do not actually appear. ■

Proof of Theorem 2.1. If X is elliptic normal there is nothing to prove. Assuming that X is neither rational nor elliptic normal, we have $d-r \geqq 2$.

Consider (for $r \geqq 3$) the resolution of $p_* A$ obtained from Lemma 2.3 by sheafifying:

$$\ldots \to \mathcal{O}_{\mathbb{P}^r}^{n_1}(-2) \xrightarrow{v} \mathcal{O}_{\mathbb{P}^r}^{d-r-1}(-1) \oplus \mathcal{O}_{\mathbb{P}^r} \to p_* A \to 0.$$

The non-zero section of $p_* A$ gives rise to a homomorphism $0 \to \mathcal{O}_X \xrightarrow{s} p_* A$, and if u denotes the composition of v with the projection $\mathcal{O}_{\mathbb{P}^r}^{d-r-1}(-1) \oplus \mathcal{O}_{\mathbb{P}^r} \to \mathcal{O}_{\mathbb{P}^r}^{d-r-1}(-1)$, then one obtains the following commutative diagram of exact sequences of sheaves on $\mathbb{P}^r$:

$$\begin{tikzcd}
 & & & 0 \arrow[d] & 0 \arrow[d] & \\
 & 0 \arrow[d] & & \mathcal{O}_{\mathbb{P}^r} \arrow[r] \arrow[d] & \mathcal{O}_X \arrow[r] \arrow[d, "s"] & 0 \\
0 \arrow[r] & K \arrow[r] \arrow[d] & \mathcal{O}_{\mathbb{P}^r}^{n_1}(-2) \arrow[r, "v"] \arrow[d, equal] & \mathcal{O}_{\mathbb{P}^r}^{d-r-1}(-1) \oplus \mathcal{O}_{\mathbb{P}^r} \arrow[r] \arrow[d] & p_* A \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & N \arrow[r] & \mathcal{O}_{\mathbb{P}^r}^{n_1}(-2) \arrow[r, "u"] & \mathcal{O}_{\mathbb{P}^r}^{d-r-1}(-1) \arrow[r] \arrow[d] & \operatorname{coker} s \arrow[r] \arrow[d] & 0. \\
 & & & 0 & 0 &
\end{tikzcd}$$

Here K and N are of course defined as the kernels of v and u respectively. Note that $\operatorname{coker} u = \operatorname{coker} s$ is supported in a finite set.

The snake lemma shows that $N/K \cong \mathcal{I}_X$. Hence to prove the $(d+1-r)$-regularity of $\mathcal{I}_X$, it suffices to establish:

(i) K is $(d+2-r)$-regular;

and

(ii) N is $(d+1-r)$-regular.

Assertion (i) is clear: $H^1(\mathbb{P}^r, K(n)) = H^2(\mathbb{P}^r, K(n)) = 0$ for all $n \in \mathbb{Z}$ by construction, $H^3(\mathbb{P}^r, K(d-1-r)) = H^1(\mathbb{P}^r, p_* A(d-1-r)) = 0$ by Lemma 2.2 since $d-1-r>0$, and the remaining vanishings follow from the fact that $p_* A$ is supported on a curve.

Turning to (ii), consider the Eagon-Northcott complex (0.5) constructed from u. Twisting by $\mathcal{O}_{\mathbb{P}^r}(d-r-1)$, it takes the form

$$(*) \quad \ldots \to \mathcal{O}_{\mathbb{P}^r}^{M_{r-1}}(-d) \to \ldots \to \mathcal{O}_{\mathbb{P}^r}^{M_0}(-d-1+r) \xrightarrow{\varepsilon} \mathcal{O}_{\mathbb{P}^r}^{n_1}(-2) \xrightarrow{u} \mathcal{O}_{\mathbb{P}^r}^{d-r-1}(-1).$$

Thus $\operatorname{Im}\varepsilon$ is a subsheaf of N, and $N/\operatorname{Im}\varepsilon$ – like all the homology of $(*)$ – is supported on the zero-dimensional set $\operatorname{Supp}(\operatorname{coker} u)$. So it suffices to verify the $(d+1-r)$-regularity of $\operatorname{Im}\varepsilon$. But this follows from Lemma 1.6. ■

Remarks. (1) If $X \subseteq \mathbb{P}^r$ is rational (i.e., $\tilde{X} = \mathbb{P}^1$) but singular, then X satisfies (B_{d-r}), and is $(d+1-r)$-regular provided that $d > r+1$. Indeed, we may construct a partial desingularization $X' \to X$ of X such that X' has a single simple node or cusp. Thus X' has arithmetic genus one, and $\omega_{X'} = \mathcal{O}_{X'}$. Then the previous arguments go through with $A = \mathcal{O}_{X'}$ and $E = \bigoplus_{n \in \mathbb{Z}} H^0(\mathbb{P}^r, p'_* A(n))$, where $p'\colon X' \to \mathbb{P}^r$ denotes the natural map.

(2) A similar approach can be used to prove the $(d-1)$-regularity of a rational curve $X \subseteq \mathbb{P}^r$ (resolve $p_* \mathcal{O}_{\mathbb{P}^1}(d-2)$, as in § 1).

§ 3. The Existence of Secant Lines

The present section is devoted to the analysis of $(d+1-r)$-irregular rational curves, completing the proof of the results stated in the Introduction.

We begin with some remarks on secants. Let $X \subseteq \mathbb{P}^r$ $(r \geqq 3)$ be a non-degenerate curve of degree d. One says that a linear space $L \subset \mathbb{P}^r$ is *n-secant* to X if

$$\dim_k(\mathcal{O}_{\mathbb{P}^r}/\mathcal{I}_X + \mathcal{I}_L) \geqq n.$$

If X has an n-secant *line*, then evidently X cannot be cut out by hypersurfaces of degree $n-1$. In particular, any non-degenerate $X \subseteq \mathbb{P}^r$ with a $(d+2-r)$-secant line L is $(d+1-r)$-irregular. Such a curve is necessarily rational (clearly) and smooth (e.g., by Remark 1 at the end of § 2). Examples exist for any $d \geqq r \geqq 3$.

Our object now is to show that (almost) all $(d+1-r)$-irregular rational curves arise in this manner:

Theorem 3.1. *Let $X \subseteq \mathbb{P}^r$ $(r \geqq 3)$ be a non-degenerate curve of degree d, and assume that X is rational (i.e., that its normalization $\tilde{X}$ is $\mathbb{P}^1$). Then X fails to be $(d+1-r)$-regular if and only if either:*

(i) $d = r$, *i.e., X is a rational normal curve,*

(ii) $d = r+1$,

or

(iii) $d > r+1$, *and X has a $(d+2-r)$-secant line.*

Remark. The case $d = r+1$ is exceptional. If X is smooth, then clearly property (B_1) fails whether or not X has a trisecant line. On the other hand, if X is singular, then (B_1) holds but $h^1(X, \mathcal{O}_X) \neq 0$, i.e., (A_0) fails. We will see, however, that (C_2) is satisfied unless X has a trisecant.

Proof. In view of what has already been said, it remains only to prove the existence of a $(d+2-r)$-secant line when $d > r+1$ and X is $(d+1-r)$-irregular. Moreover by the first remark at the end of the previous section we may assume that X is smooth. We use the notation introduced in the proof of

Proposition 1.2: in particular, $p\colon \mathbb{P}^1=\tilde{X}\to\mathbb{P}^r$ denotes the evident map, and $M=p^*\Omega^1_{\mathbb{P}^r}(1)$.

We assert to begin with that the decomposition of M into a direct sum of line bundles must take one of the following forms:

$$M=\mathcal{O}_{\mathbb{P}^1}^{r-2}(-1)\oplus\mathcal{O}_{\mathbb{P}^1}(-a)\oplus\mathcal{O}_{\mathbb{P}^1}(-b)\qquad(a,b\geqq 2),\tag{1}$$

or

$$M=\mathcal{O}_{\mathbb{P}^1}^{r-1}(-1)\oplus\mathcal{O}_{\mathbb{P}^1}(-d-1+r).\tag{2}$$

Indeed, X being non-degenerate, all the summands of M have negative degree. On the other hand, $H^1(\mathbb{P}^1,\Lambda^2 M\otimes\mathcal{O}_{\mathbb{P}^1}(d-r))\neq 0$ by Proposition 1.2 since X is $(d+1-r)$-irregular. Recalling that M has degree $-d$, it follows that M must contain $\mathcal{O}_{\mathbb{P}^1}(-1)$ as a summand at least $r-2$ times, as claimed. We treat the two cases (1) and (2) separately.

Case (1). Set $A=\mathcal{O}_{\mathbb{P}^1}(d-r)$ and consider the diagram (1.5) arising in the proof of Proposition 1.2. Taking direct images as in that proof, and using the decomposition (1) of M, one finds the following diagram of sheaves on $\mathbb{P}^r$, whose top row is exact:

$$(*)\qquad\begin{tikzcd}
H^1(\mathbb{P}^1,\Lambda^3 M\otimes A)\otimes_k\mathcal{O}_{\mathbb{P}^r}(-3) \arrow[r, "v"] & H^1(\mathbb{P}^1,\Lambda^2 M\otimes A)\otimes_k\mathcal{O}_{\mathbb{P}^r}(-2) \arrow[r] & R^1f_*\mathcal{F}_1 \arrow[r] & 0 \\
H^1(\mathbb{P}^1,\mathcal{O}_{\mathbb{P}}^{r-2}(-a-b-1)\otimes A)\otimes_k\mathcal{O}_{\mathbb{P}^r}(-3) \arrow[u, hook] \arrow[ur, "w"'] & & &
\end{tikzcd}$$

Observe that $R^1f_*\mathcal{F}_1$ must be supported on a set of dimension at least one; otherwise, as noted following the proof of Lemma 1.6, the arguments of § 1 would apply to give the $h^0(\mathbb{P}^1,A)$-regularity of X. The goal now is to show that

$$L\underset{\text{def}}{=}\operatorname{Supp}(R^1f_*\mathcal{F}_1)$$

is a $(d+2-r)$-secant line to X.

To this end we analyze the map w in $(*)$. Use the decomposition (1) of M and (1.4) to construct the commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & \mathcal{O}_{\mathbb{P}^1}^{r-2}(-1) \arrow[r] \arrow[d, hook] & V_{\mathbb{P}^1} \arrow[r] \arrow[d, equal] & E \arrow[r] \arrow[d, two heads] & 0 \\
0 \arrow[r] & M \arrow[r] & V_{\mathbb{P}^1} \arrow[r] & \mathcal{O}_{\mathbb{P}^1}(d) \arrow[r] & 0
\end{tikzcd}$$

defining a vector bundle E of rank three and degree $r-2$ on $\mathbb{P}^1$. Then $\mathbb{P}(E)$ $[=\operatorname{Proj}(\operatorname{Sym}(E))]$ embeds naturally in $\mathbb{P}(V_{\mathbb{P}^1})=\mathbb{P}^1\times\mathbb{P}^r$, and the graph Γ of p sits in $\mathbb{P}(E)$ via the quotient $E\to\mathcal{O}_{\mathbb{P}^1}(d)\to 0$. Comparing the Koszul complex (1.5) with the evident Koszul resolution of $\mathcal{O}_{\mathbb{P}(E)}$ on $\mathbb{P}^1\times\mathbb{P}^r$, and bearing in mind the splitting (1) of M, one finds that

$$\operatorname{coker} w=R^1f_*(\mathcal{O}_{\mathbb{P}(E)}\otimes N),$$

where N is a line bundle on $\mathbb{P}^1\times\mathbb{P}^r$. In particular, $\operatorname{coker} w$ is supported in the locus on $\mathbb{P}^r$ over which the projection $\mathbb{P}(E)\to\mathbb{P}^r$ fails to be finite. This is in any event a linear space, of dimension one less than the number of trivial summands of E. But E is non-trivial for reasons of degree (when $r \geqq 3$), and so $\operatorname{Supp}(\operatorname{coker} w)$ has dimension $\leqq 1$. On the other hand, $L \subseteq \operatorname{Supp}(\operatorname{coker} w)$ and $\dim L \geqq 1$. Therefore $L = \operatorname{Supp}(\operatorname{coker} w)$ is a line, and $E = \mathcal{O}_{\mathbb{P}^1}^2 \oplus \mathcal{O}_{\mathbb{P}^1}(r-2)$; L is the image of the divisor $Y = \mathbb{P}(\mathcal{O} \oplus \mathcal{O}) \subseteq \mathbb{P}(E)$ under projection to $\mathbb{P}^r$:

$$\begin{tikzcd}
\mathbb{P}^1 = \Gamma \arrow[d] & \subseteq \mathbb{P}(E) \arrow[d] \arrow[dr, phantom, "\square"] & \supseteq \mathbb{P}(\mathcal{O} \oplus \mathcal{O}) = Y \arrow[d] \\
X & \subseteq \mathbb{P}^r & \supseteq L
\end{tikzcd}$$

But Y intersects Γ in a divisor of degree $d+2-r$. Since X is smooth by assumption, it follows that L is $(d+2-r)$-secant to X, as desired.[1]

Case (2). Much as above, define E to be the cokernel of the composition

$$\mathcal{O}_{\mathbb{P}^1}^{r-1}(-1) \hookrightarrow M \hookrightarrow V_{\mathbb{P}^1} = \mathcal{O}_{\mathbb{P}^1}^{r+1},$$

so that E is a vector bundle of rank two and degree $r-1$. As before, $\mathbb{P}(E)$ embeds in $\mathbb{P}^1 \times \mathbb{P}^r$, and the graph Γ of p sits in $\mathbb{P}(E)$. Denote by $S \subseteq \mathbb{P}^r$ the rational normal scroll arising as the image of the projection $\mathbb{P}(E) \to \mathbb{P}^r$, and let $t\colon \mathbb{P}(E) \to \mathbb{P}^1$ be the bundle map. Note that $\int c_1(\mathcal{O}_{\mathbb{P}(E)}(1))^2 = r-1$, and that $\Gamma \subseteq \mathbb{P}(E)$ is the divisor of a section of $t^* \mathcal{O}_{\mathbb{P}^1}(d+1-r) \otimes \mathcal{O}_{\mathbb{P}(E)}(1)$.

Consider the decomposition of E as a sum of line bundles:

$$E = \mathcal{O}_{\mathbb{P}^1}(a) \oplus \mathcal{O}_{\mathbb{P}^1}(b)$$

with

$$0 \leqq a \leqq b, \qquad a+b = r-1.$$

If $a=0$, then S is a cone and X is singular when $d \geqq r+1$. So we may assume that $a \geqq 1$, in which case $S \cong \mathbb{P}(E)$ and X is smooth. The vanishing of $H^1(X, \mathcal{O}_X(d-r-1))$ is then automatic if $d \geqq r+1$, and we may suppose that $H^1(\mathbb{P}^r, \mathcal{I}_X(d-r)) \neq 0$. Recalling that S is projectively normal, it follows that $H^1(S, \mathcal{I}_{X/S}(d-r)) \neq 0$. But $\mathcal{I}_{X/S}(d-r) = t^* \mathcal{O}_{\mathbb{P}^1}(r-1-d) \otimes \mathcal{O}_{\mathbb{P}(E)}(d-r-1)$, and so

$$t_*(\mathcal{I}_{X/S}(d-r)) = \operatorname{Sym}^{d-r-1}(\mathcal{O}_{\mathbb{P}^1}(a) \oplus \mathcal{O}_{\mathbb{P}^1}(b)) \otimes \mathcal{O}_{\mathbb{P}^1}(r-1-d).$$

Then $\mathcal{I}_{X/S}(d-r)$ has non-vanishing H^1 if and only if

$$(d-r-1)(a-1) \leqq 0.$$

Provided that $d > r+1$, this forces $a=1$. But when $a=1$, the line bundle $t^* \mathcal{O}_{\mathbb{P}^1}(2-r) \otimes \mathcal{O}_{\mathbb{P}(E)}(1)$ has a section whose divisor L is a line in $\mathbb{P}^r$. Computing intersection numbers on S, one has

$$\#(\Gamma \cdot L) = d+2-r,$$

so L is $(d+2-r)$-secant to X, and we are done. ∎

---

[1] One may verify that in Case (1),

$$h^1(\mathbb{P}^r, \mathcal{I}_X(d-r)) = 1$$

The theorem stated in the Introduction is now proved.

Remarks. (1) In case (2), the computation just completed shows that if $h^1(\mathbb{P}^r, \mathscr{I}_X(d-r)) \neq 0$, then $h^1(\mathbb{P}^r, \mathscr{I}_X(d-r)) = 1$ unless $r=3$, in which case X lies on the smooth quadric $\mathbb{P}(\mathcal{O}_{\mathbb{P}^1}(1) \oplus \mathcal{O}_{\mathbb{P}^1}(1)) = S$ and $h^1(\mathbb{P}^r, \mathscr{I}_X(d-r)) = d-3$.

(2) For the Corollary in the Introduction, it remains to analyze property (C) for a rational curve $X \subseteq \mathbb{P}^r$ of degree $d=r+1$. Keeping the notation of the previous proof, we are necessarily in case (2), and there are three possibilities:

(i) $a=0$ ($\Leftrightarrow X$ is singular)
(ii) $a=1$ ($\Leftrightarrow X$ has a trisecant line)
(iii) $a>1$ ($\Leftrightarrow X$ is smooth and has no trisecant).

We assert that (C_2) holds in cases (i) and (iii); it obviously fails when X has a trisecant line. In fact, since the scroll S is itself 2-regular, it is equivalent to verify that the homomorphism

$$H^0(S, \mathscr{I}_{X/S}(2)) \otimes H^0(S, \mathcal{O}_S(n)) \to H^0(S, \mathscr{I}_{X/S}(n+2))$$

is surjective for $n \geqq 0$. One has

$$H^0(S, \mathscr{I}_{X/S}(n)) = H^0(\mathbb{P}^1, \mathrm{Sym}^{n-1}(\mathcal{O}_{\mathbb{P}^1}(a) \oplus \mathcal{O}_{\mathbb{P}^1}(b)) \otimes \mathcal{O}_{\mathbb{P}^1}(-2))$$
$$H^0(S, \mathcal{O}_S(n)) = H^0(\mathbb{P}^1, \mathrm{Sym}^n(\mathcal{O}_{\mathbb{P}^1}(a) \oplus \mathcal{O}_{\mathbb{P}^1}(b))),$$

and writing R for the graded polynomial ring in two variables, the question is in turn equivalent to the surjectivity of the evident map

$$(R_{a-2} \oplus R_{b-2}) \otimes \mathrm{Sym}^n(R_a \oplus R_b) \to \bigoplus_{l=0}^{n+1} R_{(n+1-l)a+lb-2}.$$

But this is clearly surjective if $a=0$ or $a>1$ (whereas when $a=1$ the $l=0$ term on the right is not in the image).

It is amusing to tabulate the data we have collected. The chart on the preceeding page shows the various possibilities for a $(d+1-r)$-irregular non-degenerate curve of degree d in $\mathbb{P}^r$.

Table of All Non-degenerate $(d+1-r)$-Irregular Curves $X \subseteq \mathbb{P}^r$ $(r \geqq 3)$ of Degree d

			C_{d+1-r}	B_{d-r}	A_{d-1-r}
$d=r$: X rational normal			No	0	$d-1$
$d=r+1$: X elliptic normal or rational					
X elliptic normal			Yes	0	1
X rational, singular			Yes	0	1
X rational, smooth		$\exists$ tri-sec line	No	1	0
		$\nexists$ tri-sec line	Yes	1	0
$d>r+1$: X rational, smooth, with a $(d+2-r)$-secant line					
$r=3$	$X \subseteq$ smooth quadric		No	$d-3$	0
	$X \nsubseteq$ smooth quadric		No	1	0
$r \geqq 4$			No	1	0

The first column of data indicates whether or not property (C_{d+1-r}) is satisfied. The entries in the columns B_{d-r} and A_{d-1-r} give respectively the dimensions of the groups $H^1(\mathbb{P}^r, \mathscr{I}_X(d-r))$ and $H^2(\mathbb{P}^r, \mathscr{I}_X(d-1-r))$ (these being the superabundances measuring the failure of the corresponding property)

§ 4. Open Problems

We discuss in conclusion some open questions.

(1) The fact that the extremal examples in the main theorem and its corollary are smooth rational curves suggests that one should have progressively stronger regularity estimates as the curve $X \subseteq \mathbb{P}^r$ becomes in some sense increasingly complex. The most naive hope might be for an estimate in terms of the genus g of X, but even for curves on a smooth quadric surface there is no non-decreasing function $f(g)$, going to infinity with g, for which property (B_n) holds for $n \geqq d+1-r-f(g)$. A more promising invariant is the integer

$$e(X) \underset{\text{def}}{=} \max\{n \mid H^1(X, \mathcal{O}_X(n)) \neq 0\},$$

so that for example $e(X) = -1$ if and only if X is smooth and rational.

Conjecture. *For a non-degenerate curve $X \subseteq \mathbb{P}^r$ of degree d, property (B_n) is satisfied for*

$$n \geqq d-r-e(X).$$

When $r=3$ the conjecture has been proved by the first and third authors (to appear), to whom it is due. The conjecture would imply Theorems 1.1 and 2.1.

(2) The thrust of our classification results is that the failure of a curve $X \subseteq \mathbb{P}^r$ to satisfy (B_{d-r}) is accounted for by the presence of a $(d+2-r)$-secant line. We may ask whether the same phenomenon persists for curves for which (B_n) fails provided that n is not too small. Specifically, one might venture the

Conjecture. *For $n \geqq \frac{2d}{3} - (r-3)$, (B_n) fails (essentially) if and only if X has an $(n+2)$-secant line.*

The equivocation is to allow for the possibility of minor exceptions such as the situation with rational curves of degree $d=r+1$. The conjecture has been verified for $X \subseteq \mathbb{P}^3$ and $n \geqq d-4$. We refer the reader to the survey [21] for further discussion.

(3) Practically nothing is known in the way of reasonably sharp explicit regularity results for varieties of dimension two or more. It would be premature to suggest any conjectures, but a number of possibilities present themselves. As Eisenbud among others has remarked, an extremely optimistic hope might be that if $X \subseteq \mathbb{P}^r$ is a reduced, irreducible, non-degenerate variety of dimension m and degree d, then perhaps X satisfies (B_n) for $n \geqq d+m-r$, and hence is $(d+m+1-r)$-regular. This is checked for certain projections of rational scrolls by Meadows [13], and it would be instructive to work out other examples. Even substantially weaker results and estimates would be of interest. For example,. it is elementary that a variety $X \subseteq \mathbb{P}^r$ as above is cut out by hypersurfaces of degree d, at least if it is smooth (cf. [15], proof of Theorem 1). Is X in fact d-regular? Another interesting problem, suggested by Rao, is to generalize Castelnuovo's geometric argument [3] to varieties of dimension two or more. Finally, in thinking about how the techniques above might generalize,

one is led to ask the following question: given a homomorphism

$$u\colon \mathcal{O}_{\mathbb{P}^r}(-1)^N \to \mathcal{O}_{\mathbb{P}^r}^a,$$

is $\mathcal{J} = F^0(u) \subseteq \mathcal{O}_{\mathbb{P}^r}$ a-regular? (This would imply Eisenbud's estimate for smooth varieties ruled over a curve.)

References

1. Arbarello, E., Sernesi, E.: Petri's approach to the study of the ideal associated to a special divisor. Invent. Math. **49**, 99–119 (1978)
2. Bayer, D.: Thesis, Harvard University, June 1982
3. Castelnuovo, G.: Sui multipli di una serie lineare di gruppi di punti appartenente ad una curva algebrica. Rend. Circ. Mat. Palermo **7**, 89–110 (1893)
4. Eagon, J., Northcott, D.: Ideals defined by matrices and a certain complex associated with them. Proc. Royal Soc. A **269**, 188–204 (1962)
5. Gotzmann, G.: Eine Bedingung für die Flachheit und das Hilberpolynom eines graduierten Ringes. Math. Z. **158**, 61–70 (1978)
6. Griffiths, P., Harris, J.: Principles of algebraic geometry. New York: Wiley Interscience 1978
7. Gruson, L., Peskine, C.: Genre des courbes de l'espace projectif. In: Olson, L. (ed.), Algebraic geometry, proceedings, Tromsø, Norway 1977. Lecture Notes in Math. vol. 687, Berlin-Heidelberg-New York: Springer 1978
8. Gruson, L., Peskine, C.: Courbes de l'espace projectif: variétés de sécantes. To appear in the proceedings of the 1981 Nice conference on enumerative and classical geometry
9. Halphen, G.: Mémoire sur la classification des courbes gauches algébriques. J. Ec. Polyt. **52**, 1–200 (1882)
10. Hartshorne, R.: Stable reflexive sheaves. II. Invent. Math. **66**, 165–190 (1982)
11. Hartshorne, R., Hirschowitz, A.: Droites en position générale dans l'espace projectif, to appear.
12. Jongmans, F.: Contributions à la théorie des variétés algébriques. Mém. Soc. Royale Science de Liège **7** (1947)
13. Meadows, C.: Linear systems cut out by quadrics on projections of varieties. Journal of Alg. (to appear).
14. Mumford, D.: Lectures on curves on an algebraic surface. Ann. of Math. Studies **59** (1966)
15. Mumford, D.: Varieties defined by quadratic equations, Corso C.I.M.E. 1969 (Questions on algebraic varieties). (Rome: Cremonese 1970), pp. 30–100
16. Okonek, C., Schneider, M., Spindler, H.: Vector bundles on complex projective space. Progress in Math. vol. 3. Birkhäuser 1980
17. Rao, A.P.: Liaison among curves in $\mathbb{P}^3$. Invent. Math. **50**, 205–217 (1979)
18. Saint-Donat, B.: On Petri's analysis of the linear system of quadrics through a canonical curve. math. Ann. **206**, 157–175 (1973)
19. Szpiro, L.: Travaux de Kempf, Kleiman, Laksov sur les diviseurs exceptionnels. Sém. Bourbaki, June 1972, exp. no. 417
20. Szpiro, L.: Equations defining space curves. Tata Institute Lectures on Math., no. 62, 1979
21. Peskine, C.: to appear in the proceedings of the 1982 conference at Ravello

Oblatum 19-VII-1982

Note Added in Proof

We would like to call the reader's attention to very interesting work by M. Green (Koszul cohomology and the geometry of projective varieties, to appear), which among other things generalizes and clarifiés some of the results of [15].

Invent. math. 72, 507 (1983)

Inventiones mathematicae

Erratum

Three Conjectures of Combinatorial Topology

M. Gutierrez

Invent. math. **67**, 357–362 (1982)

There are serious errors in the proofs of Lemmas 1 and 6, on which the entire paper depends. The results of this article, therefore, are not proved and should be considered open problems.